



ISSN: 2301-5373
E-ISSN: 2654-5101

Volume 11 • Number 1 • August 2022

JELIKU

Jurnal Elektronik Ilmu Komputer Udayana

Informatics Study Program

Faculty of Mathematics and Natural Sciences

Udayana University

Table of Contents

Analisis Sentimen Ulasan E-Commerce Pakaian Berdasarkan Kategori dengan Algoritma Convolutional Neural Network

I Made Adi Susilayasa, Anak Agung Istri Eka Karyawati, Luh Gede Astuti, Luh Arida Ayu Rahning Putri, I Gede Arta Wibawa, I Komang Ari Mogi1-12

Rancang Bangun Sistem E-Tendering Kerajinan Khas Bali Dengan Metode Hybrid Collaborative Filtering

I Gusti Agung Gede Arya Kadyanan, I Wayan Rudi Eri Astawan, Ida Bagus Made Mahendra, Anak Agung Istri Eka Karyawati, I Dewa Made Bayu Atmaja Darmawan, I Gusti Ngurah Anom Cahyadi Putra13-22

Pengelompokan Pelanggan Toko Kerajinan Menggunakan K-Means dengan Model RFM dan LRFM

I Kadek Ari Surya, Made Agung Raharja, I Komang Ari Mogi, Agus Muliantara, I Gede Arta Wibawa, I Gusti Ngurah Anom Cahyadi Putra22-32

Aplikasi Website Pengamanan File Dokumen Menggunakan Kriptografi RSA

Made Ari Widiarsana, I Gusti Ngurah Anom Cahyadi Putra, I Ketut Gede Suhartana, Luh Gede Astuti, I Putu Gede Hendra Suputra, I Wayan Supriana.....33-44

Video Steganography Encryption on Cloud Storage for Securing Digital Image

Chrisna Joshua Sergio Prasetyo, I Putu Gede Hendra Suputra, Luh Arida Ayu Rahning Putri, I Made Widiartha, I Ketut Gede Suhartana, Anak Agung Istri Ngurah Eka Karyawati.....45-54

Implementasi Port Knocking Dalam Mengamankan Jaringan VPN Pada Sistem E-learning

I Ketut Prawira Adhisastra, I Komang Ari Mogi, Cokorda Rai Adi Pramatha, I Gede Santi Astawa, Made Agung Raharja, Ngurah Agus Sanjaya ER55-62

Implementation of Genetic Algorithm in Determining Class Schedules Based on User Needs

Agus Prayogo, I Gede Santi Astawa, I Gusti Agung Gede Arya adyanan, Ngurah Agus Sanjaya ER, Ida Bagus Made Mahendra, Ida Bagus Gede Dwidasmaras63-74

Penerapan Metode MFCC dan Naive Bayes untuk Deteksi Suara Paru-Paru

Ni Made Ayu Suandewi, I Gede Arta Wibawa, I Gusti Ngurah Anom Cahyadi Putra, I Komang Ari Mogi, Ngurah Agus Sanjaya ER, Cokorda Rai Adi Pramatha75-82

Implementasi Double Frequency Modulation dan FFT dalam Sintesis Suara Rindik

I Komang Kumara Saduadnyana, I Ketut Gede Suhartana, Luh Arida Ayu Rahning Putri, I Gusti Ngurah Anom Cahyadi Putra, I Komang Ari Mogi, Ida Bagus Made Mahendra.....93-100

Sistem Monitoring Tanaman Hidroponik Berbasis Internet of Things menggunakan Restful API

Karel Leo Rivaldo, I Komang ari mogi, I Putu Gede Hendra Suputra, Ngurah Agus Sanjaya, Dewa Made Bayu Atmaja Darmawan, Ida Bagus Gede Dwidasmara 101-110

Penerapan Metode Adaptive Neuro Fuzzy Inference System (ANFIS) Dengan Membership function Tipe Gaussian dan Generalized Bell Dalam Prediksi Harga Tertinggi Saham

I Putu Sedana Wijaya, Made Agung Raharja, Luh Arida Ayu Rahning Putri, I Putu Gede Hendra Suputra, Ida Bagus Made Mahendra, I Gede Santi Astawa 111-120

Analisis Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Multinomial Naïve Bayes dan Query Expansion Ranking

Yuma Anugrah Virya Gunawan, Ngurah Agus Sanjaya ER, Ida Bagus Made Mahendra, I Made Widiartha, I Gusti Ngurah Anom Cahyadi Putra, I Gusti Agung Gede Arya Kadyanan.....121-128

Implementasi Algoritma Gronsfeld Cipher dan Steganografi End Of File Untuk Pengamanan Data

Kadek Vincky Sedana, I Komang Ari Mogi, Ida Bagus Gede Dwidasmara, I Gede Arta Wibawa, Cokorda Rai Adi Pramatha, Luh Gede Astuti 129-136

Aplikasi Identifikasi Nada Darbuka Dengan Onset Detection, MFCC, Dan KNN

Hairul Lana, Ngurah Agus Sanjaya, I Dewa Made Bayu Atmaja, Anak Agung Istri Ngurah Eka Karyawati, I Made Widiartha, I Gusti Agung Gede Arya Kadyanan..... 137-146

Classification of Typhus and Dengue Fever Using the Pseudo Nearest Neighbor Algorithm I Dewa Ngurah Tri Hendrawan, Ida Bagus Gede Dwidasmara, I Gusti Agung Gede Arya Kadyanan, I Putu Gede Hendra Suputra, Anak Agung Istri Ngurah Eka Karyawati, Ida Bagus Made Mahendra	147-156
Implementasi Metode Convolutional Neural Network Untuk Pengenalan Pola Motif Kain Tenun Rote Ndao Berbasis Android Derwi Rainord Nalle, Luh Gede Astuti, I Gede Santi Astawa, Luh Arida Ayu Rahning Putri, AAIN Eka Karyawati, I Wayan Supriana.....	157-166
Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur Farin Istighfarizky, Ngurah Agus Sanjaya ER, I Made Widiartha, Luh Gede Astuti, I Gusti Ngurah Anom Cahyadi Putra, I Ketut Gede Suhartana	167-176
Aplikasi Pencarian Informasi Al-Quran dengan Ontologi Web Semantik Safira, I Wayan Santiyasa, I Gede Arta Wibawa, Ngurah Agus Sanjaya ER, Ida Bagus Gede Dwidasmara, Luh Gede Astuti	177-186
Implementasi Metode K-Nearest Neighbor Dalam Mengklasifikasikan Jenis Suara Berdasarkan Jangkauan Vokal Ni Made Putri Wahyuni, Luh Arida Ayu Rahning Putri, I Gusti Ngurah Anom Cahyadi Putra, Dewa Made Bayu Atmaja Darmawan, Made Agung Raharja, Agus Muliantara.....	187-194
Semantics-Based Learning Media System Web Basic Mathematics Junior High School Samson Cornelius Gele Yowe, I Gede Santi Astawa, I Ketut Gede Suhartana, I Wayan Santiyasa, Ida Bagus Gede Dwidasmara, I Gede Arta Wibawa.....	195-204
Application Of Learning Vector Quantization Algorithm In Educational Games Learning To Write Aksara Bali Agus Wisnawa, I Gede Arta Wibawa, Ida Bagus Gede Dwidasmara, Made Agung Raharja, I Komang Ari Mogi, Luh Arida Ayu Rahning Putri	205-215

This page is intentionally left blank.

SUSUNAN DEWAN REDAKSI
JURNAL ELEKTRONIK ILMU KOMPUTER UDAYANA (JELIKU)

Penanggung Jawab :

Dra. Ni Luh Watiniasih M.Sc., Ph.D.

Redaktur :

Gst. Ayu Vida Mastrika Giri, S.Kom., M.Cs.

Penyunting/Editor :

Agus Muliantara, S.Kom., M.Kom

I Komang Ari Mogi, S.Kom., M.Kom

I Gusti Agung Gede Arya Kadyanan, S.Kom., M.Kom

Dr. Anak Agung Istri Ngurah Eka Karyawati, S.Si., M.Eng

Disain Grafis :

I Gede Yogananda Adi Baskara

I Gusti Agung Ayu Gita Pradnyaswari Mantara

Fotografer :

I Kadek Agus Candra Widnyana

I Komang Dwipayoga

Sekretariat :

Ni Ketut Alit Widiastuti, S.Kom.

Anak Agung Raka Darmawan, S.Kom.

I Putu Herryawan, S.Kom.

This page is intentionally left blank.

Analisis Sentimen Ulasan *E-Commerce* Pakaian Berdasarkan Kategori dengan Algoritma *Convolutional Neural Network*

I Made Adi Susilayasa¹, AAIN Eka Karyawati², Luh Gede Astuti³, Luh Arida Ayu Rahning Putri⁴, I Gede Arta Wibawa⁵, I Komang Ari Mogi⁶

^aInformatics departement, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹3523adisusilayasa@gmail.com

²eka.karyawati@unud.ac.id

³lg.astuti@cs.unud.ac.id

⁴rahningputri@unud.ac.id

⁵gede.arta@cs.unud.ac.id

⁶arimogi@cs.unud.ac.id

Abstract

Almost everyone looks at reviews before deciding to buy an item in e-commerce. Consumers say that online reviews influence their purchasing decisions. Based on these data, consumers need sentiment reviews to make a decision to choose a product/service. However, the results of the sentiment analysis are still less specific, so the review classification process is carried out based on the review category. Sentiment classification process based on clothing category is carried out using the Convolutional neural network method. The amount of data used is 3384 data with 3 categories. The category classification model shows good performance. When evaluated with testing data (unseen data), the accuracy value is 88%, the precision value is 88%, recall is 88% and the f1-score is 88%. For the sentiment classification model with the bottoms category, the resulting accuracy value is 80%, precision is 81%, recall is 80%, and f1-score is 79%. For the sentiment classification model with the dresses category, the accuracy value is 81%, precision is 81%, recall is 81%, and f1-score is 81%. For sentiment classification with the tops category the resulting accuracy value is 77%, precision is 77%, recall is 77%, and f1-score is 77%.

Keywords: e-commerce, review, sentiment, NLP, CNN

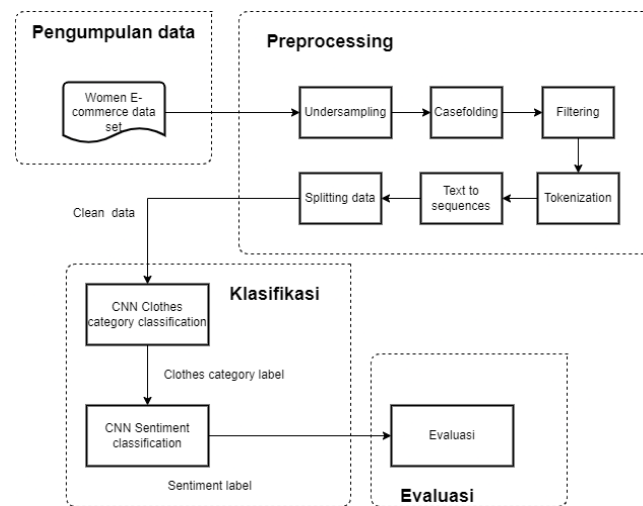
1. Pendahuluan

Beberapa pengguna e-commerce dalam menentukan membeli suatu produk umumnya melihat ulasan sebelum memutuskan untuk membeli barang. tepatnya 91% seseorang yang berusia 18-34 tahun memercayai ulasan online seperti halnya rekomendasi pribadi, dan 93% konsumen mengatakan bahwa ulasan online mempengaruhi keputusan pembelian mereka [1]. Berdasarkan data tersebut, konsumen membutuhkan ulasan sentimen guna membuat keputusan untuk memilih suatu produk/jasa. Namun hasil analisis sentiment yang dihasilkan masih kurang spesifik, maka dilakukan proses klasifikasi ulasan tersebut berdasarkan kategori ulasan. Sehingga analisis sentimen yang dihasilkan lebih spesifik. CNN biasa digunakan untuk mendeteksi dan mengenali objek pada sebuah gambar. CNN juga dapat diaplikasikan sebagai Natural Language Processing dengan mengubah representasi dari sebuah kalimat menjadi vektor. Implementasi CNN dalam NLP dapat menghasilkan performa yang baik. Seperti pada penelitian Rahman [2] hasil evaluasi untuk data opini publik pada Instagram mendapatkan hasil rata-rata precision 95.80%, recall 88.12%, f1-score 91.62%.

Dataset E-Commerce pakaian yang digunakan terbagi menjadi dua sentimen yaitu positif dan negatif. Pada penelitian ini melakukan analisis sentimen berdasarkan kategori yang terdapat di dalam dataset seperti tops, bottoms, jackets, dresses, dan intimates. Penelitian ini dilakukan dengan menggunakan dataset berbahasa Inggris yang diperoleh dari Kaggle dengan total keseluruhan data sebanyak 23.485 data dengan perbandingan data positif dan negatif 19314 data dengan label positif dan 4172 data dengan label negatif.

2. Metode Penelitian

Pada bagian menggambarkan secara umum bagaimana nantinya pelaksanaan penelitian akan dilakukan. Dimulai dari pengumpulan data yang didapatkan dari website Kaggle. Dataset berbahasa Inggris yang diperoleh dari Kaggle dengan total keseluruhan data sebanyak 23.485 data dengan perbandingan data positif dan negatif 19314 data dengan label positif dan 4172 data dengan label negatif. Data tersebut kemudian melalui *preprocessing*. Pada tahap preprocessing terdapat beberapa tahap seperti *undersampling*, *case folding*, *tokenization*, *filtering*, *text to sequences*, dan *splitting*. Tahap selanjutnya dilanjutkan dengan implementasi arsitektur CNN dan *tuning hyperparameter batch size*, *epoch*, *learning rate* dengan menggunakan metode *k-fold cross validation*. Tahap selanjutnya merupakan evaluasi dengan menggunakan *confusion matrix*.



Gambar 1. Alur Penelitian

Secara umum alur penelitian berdasarkan Gambar 1 akan dilakukan dengan berbagai tahapan yaitu pengumpulan data, preprocessing, *word embedding*, klasifikasi, dan evaluasi.

2.1. Pengumpulan data

Data yang digunakan pada penelitian ini adalah data sekunder. Data tersebut didapatkan dari website Kaggle. Dataset tersebut berjumlah sebanyak 23486 dengan 10 *data feature* namun hanya 3 *data feature* saja yang digunakan.

Tabel 1. Contoh dataset

Review Text	Positive FeedbackCount	Department Name
Absolutely wonderful - silky and sexy and comf...	0	Intimate
Love this dress! it's sooo pretty. i happene...	4	Dresses
Thisshirt is very flattering to all due to th...	6	Tops

Tabel 2. Distribusi Data

Kategori pakaian	Label data	Jumlah data
Tops	Positif	564
	Negatif	564
Dresses	Positif	564
	Negatif	564
Bottoms	Positif	564

	Negatif	564
--	---------	-----

Ulasan yang terdapat pada data telah dilabeli dengan label 0 untuk merepresentasikan produk tersebut tidak direkomendasikan dan 1 apabila produk tersebut direkomendasikan. Label pada masing-masing aspek tersebut memiliki sentimen positif dan negatif. Pada Tabel 1 merupakan contoh data yang digunakan pada penelitian ini yang berisikan data judul film, *reviewfilm*, dan nilai sentimen dari *review*. Setelah melalui preprocessing distribusi data menjadi seperti pada Tabel 2

2.2. Preprocessing

Sebelum *dataset* diolah menggunakan metode *Convolutional Neural Network*, *dataset* melalui tahapan *pre-processing* yang mana pada tahapan *pre-processing* terdapat beberapa proses didalamnya seperti *tokenization*, *case folding*, *text to sequences*, dan *filtering*. Tahap ini dilakukan untuk memastikan data yang akan digunakan telah bersih, sehingga dapat mencegah *error* atau turunnya performa mesin [3]. Setelah melewati tahapan *pre-processing* data tersebut dikelompokkan berdasarkan pada kategori pada masing-masing data. Kemudian representasi data tersebut diubah menjadi vektor dengan menggunakan algoritma *word embedding*.

a. Under sampling

Proses ini merupakan proses untuk normalisasi jumlah data menjadi seimbang. Metode ini bekerja dengan mengurangi data mayoritas secara acak, sehingga jumlah data sama dengan data minoritas. Setelah melalui proses ini data yang digunakan sebanyak 3384 data dengan 3 kelas (*tops*, *bottoms*, *dresses*)

b. Case folding

Pada tahapan *case folding* berfungsi untuk mengubah keseluruhan data teks yang ada dalam *dataset* menjadi huruf kecil sehingga keseluruhan teks memiliki bentuk standar[3]

c. Filtering

Pada tahap ini dilakukan proses menghilangkan simbol-simbol, *stopwords*, dan angka pada kalimat. Sehingga tidak terdapat simbol, angka, dan *stopwords* dalam data yang digunakan.

d. Tokenization

Pada tahapan *tokenization* berfungsi untuk merubah kalimat menjadi kata atau bisa juga disebut token. Kemudian pada tahap ini juga dilakukan pembuatan korpus kata berdasarkan kata-kata yang dihasilkan pada tahap ini[3]

e. Text to sequences

Tahap ini dilakukan untuk merubah representasi kata dalam kalimat menjadi angka. Kata yang dirubah digantikan dengan index kata dalam korpus kata[4]

f. Splitting dataset

Proses ini dilakukan dengan menggunakan metode *holdout*. Untuk pembagian *dataset* yaitu 90% data digunakan pada proses *training* dan 10% data digunakan untuk *testing*.

2.3. Word Embedding

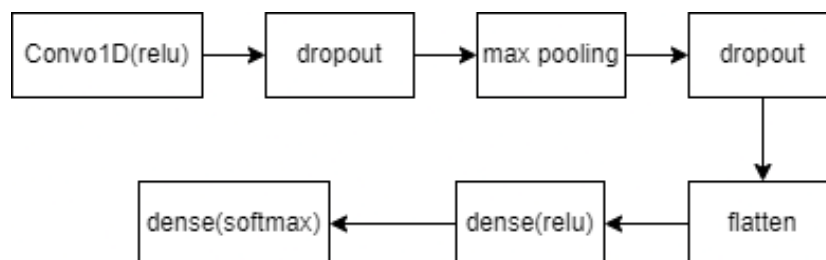
Setiap kata diubah menjadi vektor yang merepresentasikan sebuah titik pada *space* dengan dimensi tertentu. Kata-kata yang memiliki arti serupa memiliki representasi vektor dengan posisi yang tidak berjauhan antara satu dengan lainnya pada vektor *space* tersebut [5]

2.4 Convolutional Neural Network (CNN)

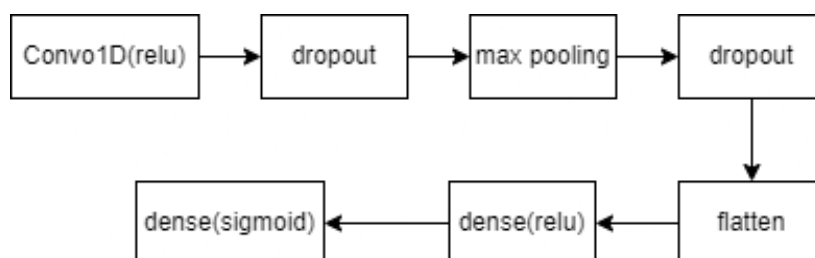
Convolutional Neural Network merupakan perkembangan dari *Artificial Neural Network* dengan penambahan *layer* pada *hidden layer*. CNN umumnya yang digunakan untuk proses klasifikasi gambar. CNN dapat diaplikasikan dalam *natural language processing* dengan mengubah representasi dari sebuah kalimat menjadi *matrix*. Dengan konfigurasi *Hyperparameter* dan vektor statis CNN dengan konfigurasi sederhana dapat menghasilkan hasil yang baik pada beberapa percobaan yang berbeda [6]. Arsitektur dari CNN yang akan diimplementasikan oleh penulis diambil berdasarkan referensi pada penelitian [2] Namun dengan sedikit perubahan pada jumlah *units* dan penambahan *layer dropout* setelah *layer konvolusi* dan *layer max pooling*. Setelah melalui proses *word embedding* data dimasukkan ke dalam lapisan *konvolusi*. Pada penelitian ini penulis akan menggunakan *konvolusi 1D*. Pemilihan *konvolusi 1D* dipilih dikarenakan ketika diaplikasikan dengan menggunakan teks sebagai data input, *konvolusi 1D* merupakan pilihan yang biasanya digunakan [7]

Sebuah kernel konvolusi (filter) berukuran tertentu digeser ke seluruh vektor tersebut. Pada setiap pergeseran menghasilkan *dot product*. Dengan menggeserkan filter tersebut ke seluruh vektor maka dihasilkan sebuah feature map. Setelah melalui lapisan konvolusi fitur aktivasi yang digunakan adalah ReLu. Output dari fungsi ReLu kemudian di masukkan ke dalam pooling layer. Untuk arsitektur yang digunakan pada penelitian ini dapat dilihat pada Gambar 2 dan Gambar 3. Arsitektur dari CNN yang akan diimplementasikan oleh penulis diambil berdasarkan referensi pada penelitian [2] dengan sedikit perubahan pada jumlah units dan penambahan layer dropout setelah layer konvolusi dan layer max pooling.

Setelah melalui proses word embedding data dimasukkan ke dalam lapisan konvolusi. Pada lapisan konvolusi terdapat proses perkalian *matrix* yang menghasilkan *feature map*. Setelah melalui lapisan konvolusi fitur aktivasi yang digunakan adalah ReLu. Output dari fungsi ReLu kemudian di masukkan ke dalam pooling layer. Bagian berikutnya merupakan lapisan max pooling layer, lapisan ini digunakan untuk mereduksi input secara spasial dengan menggunakan operasi down-sampling. Penulis menggunakan metode max pooling layer. *Dropout* ditambahkan setelah lapisan konvolusi dan lapisan *max pooling*. Penambahan parameter *dropout* bertujuan untuk mengurangi persentase terjadinya *overfitting* dan *underfitting*, selain itu juga mempercepat proses *learning*. Rentang nilai yang digunakan pada parameter *dropout* berkisar dari 0 sampai 1. Pada penelitian ini parameter *dropout* ditentukan secara eksplisit dengan rentang angka yang sesuai dengan data yang penulis miliki, untuk memberikan akurasi yang maksimal dari model tersebut dan dilakukan secara repetitif hingga ditemukan *Hyperparameter* terbaik. Lapisan *flattening* ditambahkan untuk memastikan dimensi dari vektor yang dihasilkan sebelumnya sesuai dengan dimensi vektor yang dibutuhkan pada layer selanjutnya. Pada lapisan selanjutnya terdapat lapisan *fully-connected* dengan jumlah units sebanyak 500 dan berdasarkan referensi yang diambil pada penelitian [2]. Pada output layer terdapat lapisan *fully-connected* dengan jumlah *neuron* tiga untuk mesin klasifikasi kategori dan fungsi aktivasi *softmax*. Sedangkan untuk mesin klasifikasi sentimen, pada output layer jumlah *neuron* sebanyak dua dan fungsi aktivasi *sigmoid*.



Gambar 2. Arsitektur mesin klasifikasi kategori



Gambar 3. Arsitektur mesin klasifikasi sentimen

3. Hasil dan Pembahasan

3.1. Analisa Parameter

Hyperparameter yang akan diuji *batch size*, *learning rate*, dan *epoch* dengan menggunakan *k-fold cross validation* jumlah *fold* yang digunakan sebanyak lima. *K-fold cross validation* bekerja dengan membagi sampel data secara acak dan mengelompokkan data menjadi data train dan data validasi sebanyak nilai *k* pada *k-fold* [8]. Jumlah *fold* yang digunakan sebanyak lima. Proses *tuning*

hyperparameter dilakukan secara berurut dimulai dari *Tuning batch size* kemudian *learning rate* dan terakhir *epoch*.

Tabel 3. *Hyperparameter* untuk mesin klasifikasi kategori

No	Parameter	Ukuran
1	<i>Jumlah filter</i>	100
2	<i>Kemel size</i>	2
3	<i>Batch size</i>	16, 32, 64, 128
4	<i>Learning rate</i>	0.01, 0.001, 0.0001
5	<i>Epoch</i>	3, 5, 10, 15

Tabel 4. *Hyperparameter* untuk mesin klasifikasi sentimen

No	Parameter	Ukuran
1	<i>Jumlah filter</i>	100
2	<i>Kemel size</i>	2
3	<i>Batch size</i>	8, 16, 32, 64
4	<i>Learning rate</i>	0.01, 0.001, 0.0001
5	<i>Epoch</i>	3, 5, 10, 15

Untuk nilai parameter yang akan diuji dapat dilihat pada Tabel 3 dan Tabel 4. Terdapat perbedaan pada nilai yang diuji pada parameter *batch size*. Hal tersebut karena pada mesin klasifikasi sentimen berdasarkan kategori jumlah data yang tersedia lebih sedikit daripada data untuk mesin klasifikasi kategori.

a. *Tuning parameter batch size*

Proses ini dilakukan dengan memanfaatkan metode *5-fold cross validation* untuk mendapatkan performa terbaik dari mesin klasifikasi kategori dan mesin klasifikasi sentimen pada tiap kategori. Hasil tuning parameter *batch size* dapat dilihat pada Tabel 5 untuk klasifikasi kategori dan Tabel 6 untuk klasifikasi sentimen pada tiap kategori. Nilai yang diambil merupakan nilai rata-rata akurasi validasi tiap *5-fold* pada masing-masing nilai yang diuji. Berdasarkan Tabel 5 nilai parameter terbaik *batch size* untuk mesin klasifikasi kategori sebesar 64. Untuk mesin klasifikasi sentimen berdasarkan Tabel 6, Tabel 7, dan Tabel 8 sebesar 16.

Tabel 5. *Tuning batch size* klasifikasi kategori

<i>Learning Rate</i>	<i>Epoch</i>	<i>Batch Size</i>	<i>5-Fold cross validation</i>					<i>Validation Accuracy</i>
			<i>Fold-1</i>	<i>Fold-2</i>	<i>Fold-3</i>	<i>Fold-4</i>	<i>Fold-5</i>	
0.001	5	16	87,50%	91,00%	84,70%	86,20%	86,70%	87,30%
0.001	5	32	86,70%	91,50%	86,20%	86,20%	87,00%	87,50%
0.001	5	64	86,50%	90,40%	88,70%	87,80%	85,90%	87,90%
0.001	5	128	85,00%	90,60%	85,40%	87,00%	87,20%	87,10%

Tabel 6. *Tuning batch size* klasifikasi sentimen *bottoms*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy Bottoms
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	5	8	82,30%	78,80%	81,30%	77,30%	78,30%	79,60%
0.001	5	16	82,30%	79,30%	85,20%	82,30%	88,10%	83,40%
0.001	5	32	81,80%	85,20%	84,20%	82,30%	75,90%	81,90%
0.001	5	64	82,80%	84,70%	80,30%	77,30%	82,30%	81,50%

Tabel 7. *Tuning batch size* klasifikasi sentimen *dresses*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy Dresses
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	5	8	76,80%	65,50%	75,90%	86,70%	82,80%	77,30%
0.001	5	16	81,80%	80,30%	80,80%	84,20%	83,70%	82,20%
0.001	5	32	79,30%	77,30%	76,80%	85,70%	82,80%	80,40%
0.001	5	64	78,30%	83,30%	62,60%	80,80%	76,80%	76,40%

Tabel 8. *Tuning batch size* klasifikasi sentimen *tops*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy Tops
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	5	8	82,80%	81,30%	80,30%	82,30%	82,80%	81,90%
0.001	5	16	81,80%	83,30%	80,80%	83,30%	87,70%	83,30%
0.001	5	32	82,80%	82,80%	81,30%	83,30%	84,20%	82,90%
0.001	5	64	84,70%	77,80%	76,80%	72,40%	81,80%	78,70%

b. *Tuning parameter learning rate*

Hasil tuning parameter *learning rate* untuk mesin klasifikasi kategori dapat dilihat pada Tabel 9 dengan nilai terbaik 0.001 sedangkan untuk nilai terbaik pada mesin klasifikasi sentimen berdasarkan Tabel 10, Tabel 11, dan Tabel 12 sebesar 0.001

Tabel 9. *Tuning learning rate* klasifikasi kategori

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.01	5	64	83,40%	90,30%	83,70%	83,90%	82,40%	84,80%

0.001	5	64	86,70%	91,10%	87,00%	85,90%	87,30%	87,60%
0.0001	5	64	68,90%	59,90%	73,60%	71,10%	67,50%	68,20%

Tabel 10. Tuning learning rate klasifikasi sentimen bottoms

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy bottoms
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.01	5	64	64,00%	79,30%	77,30%	79,30%	78,30%	75,70%
0.001	5	64	82,80%	86,70%	82,30%	84,20%	84,20%	84,00%
0.0001	5	64	61,60%	65,00%	63,50%	62,00%	59,60%	62,40%

Tabel 11. Tuning learning rate klasifikasi sentimen dresses

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy dresses
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.01	5	64	77,80%	75,90%	72,40%	75,40%	73,90%	75,10%
0.001	5	64	77,80%	78,80%	76,80%	82,30%	80,80%	79,30%
0.0001	5	64	64,00%	48,80%	57,60%	57,10%	72,90%	60,00%

Tabel 12. Tuning learning rate klasifikasi sentimen tops

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy tops
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.01	5	64	77,80%	65,50%	80,80%	73,40%	78,80%	75,30%
0.001	5	64	82,80%	81,30%	80,80%	79,30%	85,20%	81,90%
0.0001	5	64	61,00%	61,00%	57,60%	56,16%	59,60%	59,10%

c. Tuning parameter epoch

Proses *tuning parameter epoch* dilakukan sama seperti proses *tuning parameter batch size* dan *learning rate*. Berdasarkan Tabel 13 nilai *epoch* terbaik dihasilkan ketika *epoch* sebesar 5. Berdasarkan Tabel 14 dan Tabel 16 untuk nilai *epoch* terbaik pada mesin klasifikasi sentiman pada kategori *bottoms* dan *tops* sebesar 10 sedangkan untuk kategori *dresses* berdasarkan Tabel 15 nilai *epoch* terbaik sebesar 5.

Tabel 13. Tuning epoch klasifikasi kategori

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	3	64	87,80%	89,30%	85,90%	86,20%	87,50%	87,40%
0.001	5	64	87,80%	91,10%	87,40%	86,40%	88,00%	88,10%
0.001	10	64	86,90%	91,60%	88,90%	85,60%	87,00%	88,00%

0.001	15	64	88,20%	90,50%	87,40%	86,70%	87,20%	87,90%
-------	----	----	--------	--------	--------	--------	--------	--------

Tabel 14. *Tuning epoch* klasifikasi sentimen *bottoms*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	3	16	82,80%	79,30%	78,80%	74,90%	81,80%	79,50%
0.001	5	16	82,80%	79,80%	81,30%	74,40%	85,70%	80,80%
0.001	10	16	82,80%	83,30%	80,80%	82,30%	82,80%	82,40%
0.001	15	16	83,30%	81,30%	81,30%	75,90%	82,30%	80,80%

Tabel 15. *Tuning epoch* klasifikasi sentimen *dresses*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	3	16	83,70%	78,30%	77,30%	74,40%	83,20%	79,40%
0.001	5	16	83,30%	81,30%	78,80%	79,30%	84,70%	81,50%
0.001	10	16	82,80%	80,30%	76,80%	79,30%	85,70%	80,90%
0.001	15	16	80,80%	78,30%	80,80%	79,80%	81,80%	80,30%

Tabel 16. *Tuning epoch* klasifikasi sentimen *tops*

Learning Rate	Epoch	Batch Size	5-Fold cross validation					Validation Accuracy
			Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	
0.001	3	16	73,40%	80,80%	77,80%	82,30%	84,20%	79,70%
0.001	5	16	82,30%	81,30%	75,90%	87,20%	74,90%	80,00%
0.001	10	16	81,80%	80,30%	77,80%	81,80%	81,80%	80,70%
0.001	15	16	82,30%	78,30%	78,30%	75,40%	81,80%	79,80%

Kombinasi *hyperparameter* terbaik untuk mesin klasifikasi kategori didapatkan dengan nilai *epoch* 5, *learning rate* 0.001, dan *batch size* 64. Untuk mesin klasifikasi sentimen kategori *bottoms* kombinasi *hyperparameter* dengan performa terbaik didapatkan dengan nilai *batch size* 16, *learning rate* 0.001, dan *epoch* 10. Untuk mesin klasifikasi sentimen kategori *dresses* kombinasi *hyperparameter* dengan performa terbaik didapatkan dengan nilai *batch size* 16, *learning rate* 0.001, dan *epoch* 5. Untuk mesin klasifikasi sentimen kategori *tops* kombinasi *hyperparameter* dengan performa terbaik didapatkan dengan nilai *batch size* 16, *learning rate* 0.001, dan *epoch* 10.

3.2. Evaluasi

Evaluasi dari mesin dengan parameter terbaik dilakukan dengan menggunakan metode *confusion matrix*, karena mesin *deep learning* yang dilatih merupakan mesin klasifikasi [9]. Terdapat beberapa *metrics* yang dijadikan patokan seperti nilai *recall*, *precision*, *f1-score*. Pengukuran ketiganya nilai

tersebut dilakukan untuk mendapatkan informasi secara detail bagaimana performa dari mesin yang dilatih.

a. Mesin klasifikasi kategori

Hasil evaluasi model terbaik untuk klasifikasi kategori pakaian dengan tuning *Hyperparameterbatch size* 64, *learning rate* 0.001, dan *epoch* 5. Berdasarkan pada Tabel 17 rata-rata nilai recall, precision, dan f-score yang dihasilkan yaitu 88%.

Tabel 17. Evaluasi mesin klasifikasi kategori

	precision	recall	f1-score	Jumlah data
Bottoms	0.87	0.92	0.90	113
Dresses	0.96	0.82	0.89	113
Tops	0.82	0.89	0.86	113
accuracy			0.88	339
average	0.88	0.88	0.88	339

b. Mesin klasifikasi sentimen

Hasil evaluasi model klasifikasi sentimen kategori bottoms dengan tuning *Hyperparameterbatch size* 16, *learning rate* 0.001, dan *epoch* 10 dapat dilihat pada Tabel 18 mesin diuji dengan menggunakan data testing menunjukkan performa yang kurang dibandingkan mesin klasifikasi sentimen pada kategori lainnya. Rata-rata nilai recall yang dihasilkan 81%, precision 80%, dan f1-score 79%.

Tabel 18. Evaluasi mesin klasifikasi sentimen *bottoms*

	precision	recall	f1-score	Jumlah data
negative	0.74	0.91	0.82	56
positive	0.89	0.68	0.77	57
accuracy			0.80	113
average	0.81	0.80	0.79	113

Hasil evaluasi mesin klasifikasi *dresses* dengan tuning *Hyperparameterbatch size* 16, *learning rate* 0.001, dan *epoch* 5 pada Tabel 19 menunjukkan performa yang lebih baik dibandingkan dengan hasil evaluasi pada model klasifikasi sentimen kategori bottoms. Nilai rata-rata *precision*, *recall*, dan *f1-score* yang dihasilkan sama yaitu dengan tingkat persentase sebesar 81%.

Tabel 19. Evaluasi mesin klasifikasi sentimen *tops*

	precision	recall	f1-score	Jumlah data
negative	0.81	0.79	0.80	56
positive	0.80	0.82	0.81	57

accuracy			0.81	113
average	0.81	0.81	0.81	113

Berdasarkan pada Tabel 20 hasil evaluasi model klasifikasi *tops* dengan tuning hyperparameter batch size 16, learning rate 0.001, dan epoch 10 pada Tabel 20 menunjukkan performa yang baik. Nilai rata-rata precision, recall, f1-score, dan accuracy yang dihasilkan sama yaitu dengan tingkat persentase sebesar 77%.

Tabel 20. Evaluasi mesin klasifikasi sentimen *tops*

	precision	recall	f1-score	Jumlah data
negative	0.77	0.77	0.77	56
positive	0.77	0.77	0.77	57
accuracy			0.77	113
average	0.77	0.77	0.77	113

4. Kesimpulan

Mesin klasifikasi kategori ulasan produk pakaian performa terbaik didapatkan dengan kombinasi hyperparameter batch size sebesar 64, learning rate sebesar 0.001, epoch sebesar 5 dengan akurasi validasi yang dihasilkan sebesar 88,1%. Evaluasi mesin dengan data *testing* menunjukkan *performa* yang baik dengan akurasi, *precision*, *recall* dan *f1-score* yang dihasilkan cukup tinggi ketika dievaluasi dengan *data testing (unseen data)*. Nilai akurasi yang dihasilkan sebesar 88%, nilai *precision* yang dihasilkan sebesar 88%, *recall* 88% dan *f1-score* sebesar 88%.

Performa terbaik pada mesin klasifikasi sentimen dengan kategori *bottoms* didapatkan dengan kombinasi hyperparameter batch size sebesar 16, learning rate sebesar 0.001, dan epoch sebesar 10 dengan akurasi validasi sebesar 82,36%. Pada mesin klasifikasi sentimen kategori *dresses* performa terbaik didapatkan dengan kombinasi hyperparameter batch size sebesar 16, learning rate sebesar 0.001, dan epoch sebesar 5 dengan akurasi validasi sebesar 81,5%. Untuk mesin klasifikasi sentimen kategori *tops* performa terbaik didapatkan dengan kombinasi hyperparameter batch size sebesar 16, learning rate sebesar 0.001, dan epoch sebesar 10 dengan akurasi validasi 80,7%. Hasil evaluasi mesin klasifikasi sentimen pada tiap-tiap kategori menunjukkan performa yang baik. Dimana mesin klasifikasi sentimen dengan kategori *bottoms*, nilai akurasi yang dihasilkan sebesar 80%, *precision* 81%, *recall* 80%, dan *f1-score* 79%. Pada kategori *dresses* nilai akurasi yang dihasilkan sebesar 81%, *precision* 81%, *recall* 81%, dan *f1-score* 81%. Pada kategori *tops* nilai akurasi yang dihasilkan sebesar 77%, *precision* 77%, *recall* 77%, dan *f1-score* 77%.

Referensi

- [1] D. Kaemingk, "Online reviews statistics to know in 2021," 2020. <https://www.qualtrics.com/blog/online-review-stats/> (accessed Jun. 12, 2021).
- [2] M. A. Rahman *et al.*, "Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network," *J. Intell. Syst. Comput.*, pp. 1–9, 2020.
- [3] K. D. Yonatha Wijaya and A. A. I. N. E. Karyawati, "The Effects of Different Kernels in SVM Sentiment Analysis on Mass Social Distancing," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 2, p. 161, 2020, doi: 10.24843/jlk.2020.v09.i02.p01.
- [4] M. T. Ari Bangsa, S. Priyanta, and Y. Suyanto, "Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 2, p. 123, 2020, doi: 10.22146/ijccs.51646.
- [5] J. W. Gotama Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*. 2020.

- [6] K. Yoon, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014, doi: 10.3115/v1/d14-1181.
- [7] Y. Goldberg, *Yoav Goldberg Book - Neural Network Methods in Natural Language Processing*. 2017.
- [8] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with *K-fold* cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Renew. Energy*, vol. 115, pp. 411–422, Jan. 2018, doi: 10.1016/j.renene.2017.08.061.
- [9] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps*. 2019.

This page is intentionally left blank

Rancang Bangun Sistem E-Tendering Kerajinan Khas Bali Dengan Metode Hybrid Collaborative Filtering

I Gusti Agung Gede Arya Kadyanan^{a1}, I Wayan Rudi Eri Astawan^{a2}, Ida Bagus Made Mahendra^{b3},
Anak Agung Istri Ngurah Eka Karyawati^{b4}, I Dewa Made Bayu Atmaja Darmawan^{b5}, I Gusti Ngurah
Anom Cahyadi Putra^{b6}

^aProgram Studi Teknik Informatika, Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu
Pengetahuan Alam Universitas Udayana
Jalan Raya Kampus Unud, Badung, 08361, Bali, Indonesia

¹ gungde@unud.ac.id

² rudiastawan43@gmail.com

³ ibm.mahendra@unud.ac.id

⁴ eka.karyawati@unud.ac.id

⁵ dewabayu@unud.ac.id

⁶ anom.cp@unud.ac.id

Abstrak

Bali merupakan salah satu daerah di Indonesia yang mempunyai banyak kerajinan. Banyak peminat yang menjadikan kerajinan khas Bali sebagai souvenir maupun digunakan sendiri. Banyaknya peminat belum diimbangi dengan informasi lokasi pemesanan kerajinan khas Bali. *E-tendering* akan memudahkan pembeli dalam melakukan pemesanan kerajinan khas Bali. *E-tendering* merupakan cara pemilihan penyedia barang/jasa yang diikuti oleh semua penyedia barang/jasa yang terdaftar pada sistem pengadaan elektronik. Untuk mendapatkan hasil yang sesuai berdasarkan spesifikasi pengguna digunakan sistem rekomendasi. Sistem rekomendasi adalah sistem yang memberikan rekomendasi kepada pengguna dalam menemukan tempat pembuatan kerajinan khas Bali berdasarkan kriteria serta pengguna sebelumnya. Sistem rekomendasi ini dibangun menggunakan metode ICHM (*item-based clustering hybrid method*) dan algoritma *slope one*. Sistem ini akan memberikan rekomendasi berupa tempat pemesanan kerajinan khas Bali berdasarkan *rating* item dan konten item. Pengujian menggunakan MAE (*Mean Average Error*) yang mendapatkan nilai kurang dari 1,000. Semakin rendah nilai MAE maka nilai rekomendasi semakin akurat.

Kata Kunci: ICHM, Kerajinan Khas Bali, Rekomendasi, Slope One

1. Pendahuluan

Bali merupakan salah satu daerah di Indonesia yang mempunyai banyak jenis kerajinan. Salah satunya adalah kerajinan khas Bali yang memiliki keunikan dan keagungannya sendiri sesuai dengan kebutuhan dan kegunaannya, serta banyak mengandung unsur-unsur budaya Bali [4]. Jenis-jenis kerajinan khas Bali diantaranya lukisan khas Bali, kerajinan emas dan perak, topeng, kain endek dan songket, gamelan, anyaman dan patung khas Bali. Kerajinan khas Bali menjadi kerajinan yang dicari dan sangat digemari karena kerajinan khas Bali begitu unik dan menjadi ciri khas yang dapat dijadikan oleh-oleh atau souvenir bagi wisatawan yang berkunjung ke Bali dan juga sebagai pilihan souvenir untuk acara-acara perusahaan yang mengadakan acara *gathering* perusahaan atau pertemuan kantor di Bali. Banyaknya peminat dari kerajinan khas Bali tidak diimbangi dengan informasi pemesanan produk kerajinan, banyak calon pembeli yang masih kesulitan mencari tempat yang bisa memesan produk kerajinan khas Bali. Selama ini proses pemesanan kerajinan khas Bali yang masih dilakukan secara konvensional, calon pembeli harus datang langsung ke tempat pengerajin atau pelaku usaha kerajinan untuk mendapatkan informasi mengenai kerajinan dan memesan kerajinan secara langsung.

Dari permasalahan di atas maka diperlukan sebuah *e-tendering* (Sistem Pengadaan Secara Elektronik) yang akan memudahkan konsumen dalam melakukan pemesanan kerajinan khas Bali. Dalam Kamus Besar Bahasa Indonesia [10] menyatakan bahwa tender merupakan kegiatan atau tawaran untuk mengajukan harga, melakukan pemborongan pekerjaan, atau menyediakan barang. Sistem *e-tendering* dibangun pada penelitian ini menggunakan *Hybrid Collaborative Filtering* yang akan

membuat sistem *e-tendering* ini memiliki keunikan karena memberikan beberapa rekomendasi pengerajin atau pelaku usaha yang bisa mengerjakan kerajinan yang sudah di tenderkan sebelumnya. *E-tendering* ini akan menyediakan fitur tender yang dapat digunakan pembuat jasa kerajinan untuk mendapatkan proyek kerajinan, dimana para pembuat kerajinan hanya akan diminta memberikan penawaran harga ke tender proyek yang diinginkan. Dalam proses rekomendasi ini akan menggunakan beberapa kriteria untuk mendukung proses rekomendasi diantaranya adalah harga, *rating*, dan histori proyek tender. Dimana hasil rekomendasi yang akan diberikan berdasarkan kriteria yang dimasukkan oleh para pelaku usaha. Semetara nilai *rating* akan didapatkan saat sebuah proyek selesai dilakukan maka dari sistem akan meminta nilai kepuasan dari pengguna.

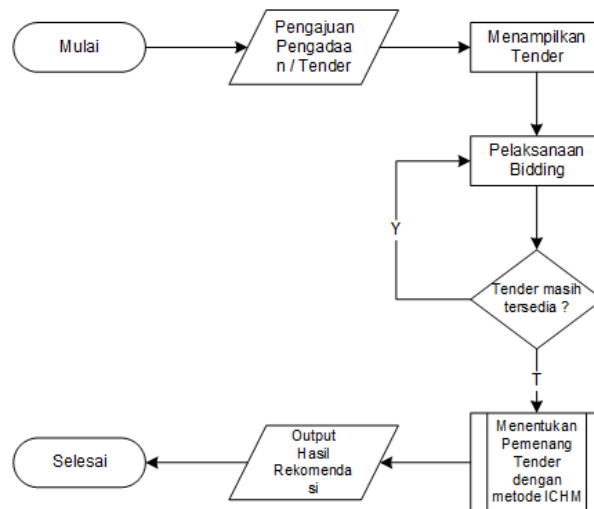
2. Metode Penelitian

2.1 Data Penelitian

Pada tahapan pengumpulan data menggunakan metode studi pustaka dari beberapa referensi diantaranya buku, jurnal, hasil penelitian, skripsi atau tesis yang berhubungan dengan penelitian yang dilakukan. Hasil dari tahapan pengumpulan data mendapatkan data tempat kerajinan khas Bali sebanyak 50 data.

2.2 Desain Sistem

Desain Sistem *e-tendering* kerajinan khas Bali akan diilustrasikan dalam bentuk *flowchart* :



Pada *flowchart* dapat dilihat alur dari sistem *e-tendering* kerajinan khas Bali. Pertama dilakukan pengajuan sebuah pengadaan pada sistem, kemudian sistem akan menampilkan tender diteruskan dengan proses *bidding* atau penawaran harga. proses *bidding* ini hanya bisa dilakukan saat tenggang waktu pengadaan masih tersedia, jika waktu pengadaan sudah berakhir maka data *bidding* akan diproses menggunakan metode ICHM untuk mendapatkan hasil rekomendasi pemenang tender.

2.3 Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah model aplikasi yang dapat memberikan sebuah rekomendasi dari keinginan pelanggan berdasarkan hasil observasi yang dilakukan [3]. Metode rekomendasi, dibagi dalam beberapa klasifikasi yaitu *content-based recommendation*, *collaborative based recommendation* dan *hybrid based recommendation* [5].

2.4 Tender dan E-tendering

- Tender adalah suatu rangkaian proses kegiatan penawaran untuk menyeleksi, mendapatkan, menetapkan serta menunjukan suatu perusahaan yang paling pantas untuk mengerjakan suatu paket pekerjaan [1].
- E-tendering* adalah suatu tata cara pemilihan penyedia barang/jasa yang dilaksanakan secara terbuka dan dapat diikuti oleh semua penyedia yang terdaftar pada Sistem Pengadaan Secara Elektronik dengan memberikan satu kali penawaran dalam waktu yang telah ditentukan [6].

2.5 Hybrid Collaborative Filtering Recommendation

Metode *hybrid collaborative filtering* secara umum dapat diartikan bahwa pendekatan *hybrid recommendations* menggabungkan lebih dari satu metode yang ada pada *recommender system*, kombinasi yang ada pada teknik ini misalnya dengan menggabungkan metode *content based* dengan *collaborative filtering* [7]. Terdapat beberapa cara yang dapat dilakukan untuk penggabungan dalam metode *hybrid collaborative filtering* yaitu Penggabungan secara Linier (*Linear Combination*), Penggabungan secara Sekuensial (*Sequential Combination*), Peggabungan secara *Item-based Clustering Hybrid Method (ICHM)* [8].

2.6 Item-Based Clustering Hybrid Method

Item-Based Clustering Hybrid Method (ICHM) merupakan salah satu metode berbasis *hybrid* yang digunakan untuk menggabungkan pendekatan *Content-Based Filtering* dan *Collaborative Filtering* [9].

- a. Implementasikan algoritma *clustering* pada konten item. Kemudian hitung nilai peluang setiap *item* ke setiap klaster untuk membangun matrik *group-rating*. Algoritma yang digunakan adalah algoritma *K-Means Clustering*.

$$Pro(j, k) = 1 - \frac{CS(j, k)}{\max CS(i, k)} \quad (1)$$

Dimana :

$Pro(j, k)$: Merupakan peluang item j untuk menjadi bagian dari klaster k

$CS(j, k)$: Merupakan *counter-similarity* antara item j dan klaster k , menggunakan persamaan *Euclidean Distance*.

$\max CS(i, k)$: Nilai *similarity* terbesar antara sebuah item dan klaster k .

- b. Perhitungan *similarity* dilakukan pada matriks *group-rating* dan matriks *item-rating*, lalu hasilnya digabungkan untuk perhitungan prediksi. Persamaan *pearson correlation-based similarity* merupakan persamaan berbasis korelasi digunakan untuk menghitung *similarity item-rating*, dengan persamaan sebagai berikut.

$$sim(i, j) = \frac{\sum_{n=1}^m (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{n=1}^m (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{n=1}^m (R_{u,j} - \bar{R}_j)^2}} \quad (2)$$

Dimana :

$sim(i, j)$: Nilai *similarity* antara item i dan item j

m : Jumlah total *user* yang *me-rating* item i dan item j

\bar{R}_i dan \bar{R}_j : Rating rata – rata pada item i dan item j

$R_{u,i}$ dan $R_{u,j}$: *User* u *me-rating* item i dan item j

- c. Persamaan *Adjusted Cosine Similarity* digunakan untuk menghitung *similarity group-rating*, dengan persamaan sebagai berikut.

$$sim(k, l) = \frac{\sum_{u=1}^m (R_{u,k} - \bar{R}_u)(R_{u,l} - \bar{R}_u)}{\sqrt{\sum_{u=1}^m (R_{u,k} - \bar{R}_u)^2} \sqrt{\sum_{u=1}^m (R_{u,l} - \bar{R}_u)^2}} \quad (3)$$

Dimana :

$sim(k, l)$: Nilai *similarity* antara item k dan item l

\bar{R}_u : Rata – rata nilai klaster u

$R_{u,k}$ dan $R_{u,l}$: Nilai klaster u dengan item k atau item l

- d. Menghitung prediksi *rating* untuk suatu item dibagi menjadi dua berdasarkan kasus atau kondisi, yaitu non *cold-start problem* dan *cold-start problem*.

1. Non cold-start problem

Non cold-start problem merupakan kondisi item yang sudah mendapatkan *rating* dari beberapa *user*. Kondisi ini menggunakan metode *weighted average of deviation* yang didapat dari rata – rata item yang telah di-*rating*, dengan persamaan sebagai berikut.

$$P_{u,k} = \bar{R}_k + \frac{\sum_{i=1}^n (R_{u,i} - \bar{R}_i) \times \text{sim}(k, i)}{\sum_{i=1}^n |\text{sim}(k, i)|} \quad (4)$$

Dimana :

- $P_{u,k}$: Prediksi *rating* item k untuk *user* u.
- n : Jumlah *rated item* *user* u.
- $R_{u,i}$: *Rating* dari *user* u untuk item i.
- \bar{R}_k dan \bar{R}_i : *Rating* rata – rata untuk item k dan item i.
- $\text{sim}(k, i)$: *Similarity* antara item k dengan seluruh *rated item active user*.

2. Cold-start problem

Cold-start problem merupakan kondisi item baru masuk kedalam sistem dan belum mendapat *rating* sama sekali oleh *user*. Metode yang digunakan adalah *weighted sum*, dengan persamaan sebagai berikut.

$$P_{u,k} = \frac{\sum_{i=1}^n R_{u,i} \times \text{sim}(k, i)}{\sum_{i=1}^n |\text{sim}(k, i)|} \quad (5)$$

Dimana :

- $P_{u,k}$: Prediksi *rating* item k untuk *user* u.
- n : Jumlah *rated item* *user* u.
- $R_{u,i}$: *Rating* dari *user* u untuk item i.
- $\text{sim}(k, i)$: Nilai *similarity* antara item k dengan seluruh *rated item active user*

1.6 Desain Evaluasi Sistem

Pada tahap pengujian sistem merupakan tahap untuk melakukan perancangan skenario untuk pengujian sistem. Pada penelitian ini akan menggunakan empat jenis pengujian sistem yaitu *White Box Testing*, *Black Box Testing*, *Stress Testing* dan *Accuracy Testing*.

a. Black Box Testing

Black box testing merupakan salah satu pengujian sistem yang biasanya berupa perancangan *test case* yang menunjukkan apakah data yang dimasukan sesuai dengan keluaran yang diharapkan oleh kebutuhan pengguna.

b. Stress Testing

Stress testing merupakan pengujian yang bertujuan untuk melihat kemampuan sistem dalam menangani kebutuhan dari sumber daya yang tidak normal dan untuk menangani jumlah maksimum yang bisa di-*handle* oleh sistem.

c. Accuracy Testing

Mean Average Error atau MAE merupakan salah satu metode standar statistika dalam pengujian suatu tingkat akurasi dari sebuah sistem rekomendasi [2]. MAE melakukan perhitungan kesalahan antara *rating* yang sebenarnya dari pengguna (p_i) dan *rating* dari prediksi yang merupakan hasil dari perhitungan yang dilakukan oleh sistem (q_i). Semakin kecil nilai yang didapat dari hasil perhitungan MAE maka prediksi yang dihasilkan semakin akurat.

$$MAE = \sum_{i=1}^n \frac{|p_i - q_i|}{N} \quad (6)$$

d. White Box Testing

White box testing merupakan salah satu pengujian sistem yang digunakan untuk melakukan validasi desain, keputusan, asumsi dan menemukan kesalahan pemrograman dan kesalahan implementasi dalam perangkat lunak. Metode yang digunakan dalam pengujian *white box* adalah metode basis *path*. Berikut rancangan pengujian *white box* dengan menggunakan basis *path*, yaitu sebagai berikut:

1. Mengubah *pseudocode* metode *simplex* menjadi *flowchart* kemudian menjadi *flow graph* sehingga menggambarkan alur dari algoritma.

2. Dari *flow graph*, dapat dihitung *cyclomatic complexity*, yakni :

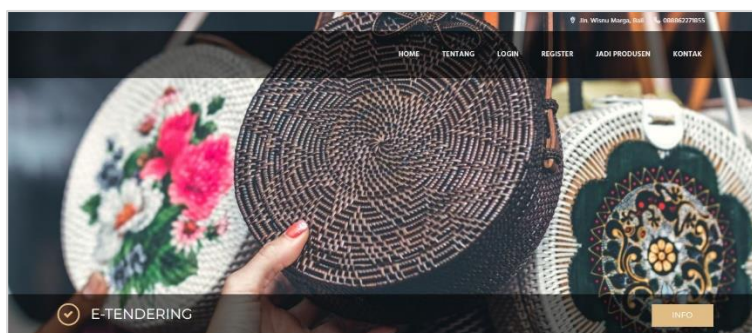
$$V(G) = \text{Jumlah Busur} - \text{Jumlah simpul} + 2 \quad (7)$$

Berdasarkan jumlah *Cyclomatic Complexity* maka dapat dibuat jalur (*path*) pengujian sesuai *flow graph*.

3. Hasil dan Pembahasan

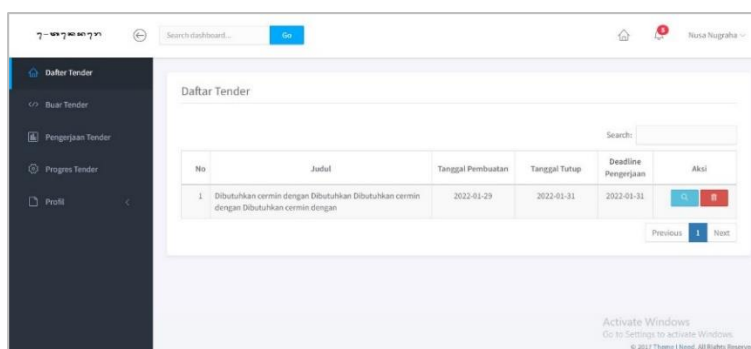
3.1 Implementasi Sistem

Sistem *e-tendering* kerajinan khas Bali dibangun dengan metode pengembangan sistem SDLC (*System Development Life Cycle*). Sistem *e-tendering* ini dibangun dengan tiga sisi *users* yaitu Produsen, Konsumen, dan Admin. Pada sisi Produsen, sistem dirancang untuk memberikan kemudahan mendapatkan sebuah proyek kerajinan dengan mengikuti tender terbuka dengan memberikan penawaran serta kualitas terbaik. Kemudian pada sisi Konsumen, sistem dirancang untuk memberikan kemudahan dalam membuat kerajinan, serta mendapatkan harga dan kualitas yang paling sesuai. Sedangkan pada sisi admin, sistem dirancang untuk dapat mengelola keseluruhan data produsen, konsumen, tender, dan proses lamaran dalam *e-tendering* kerajinan. Untuk pengiriman hasil kerajinan bisa dikomunikasikan lebih lanjut oleh produsen dan konsumen.



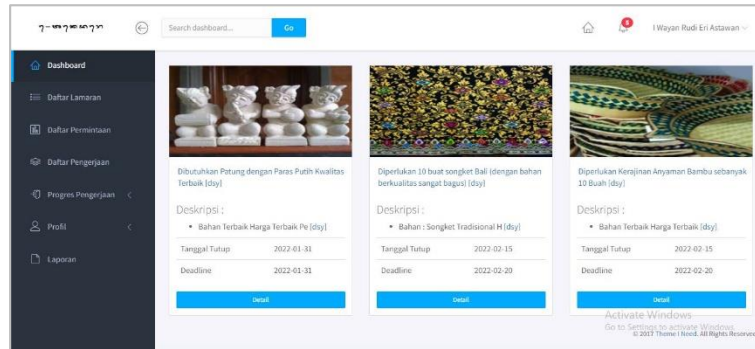
Gambar 1. Tampilan Awal Sistem

Pada tampilan awal sistem terdapat beberapa fitur yang dapat digunakan oleh *users*, diantaranya fitur *login*, *register*, *contact*, daftar tender yang terbaru yang baru didaftarkan ke sistem, daftar tender yang sudah memasuki tahap pengerjaan dan beberapa informasi lainnya.



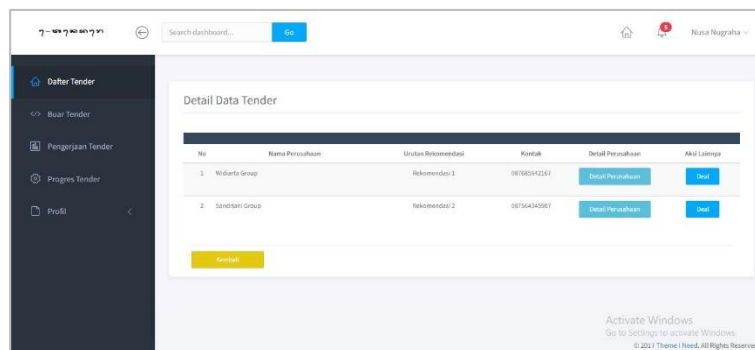
Gambar 2. Tampilan Pendaftaran Pengadaan/Tender

Pada menu tender ini sistem menyediakan beberapa menu yang dapat digunakan oleh pengguna yaitu daftar tender, buat tender, pengerjaan tender, progres tender. Pengguna dapat mengajukan pengadaan kerajinan khas Bali dengan mendaftar terlebih dahulu pada sistem dengan memberikan beberapa data mengenai tender yang akan diajukan diantaranya judul tender, deskripsi tender, tanggal tutup, *deadline* pengerjaan, dan gambar dari tender yang ingin diajukan.



Gambar 3. Tampilan Pengadaan/Tender yang Sedang Berjalan

Merupakan halaman yang digunakan oleh para pengerajin (produsen) untuk melihat daftar pengadaan yang sedang berjalan. Pada halaman ini terdapat beberapa fitur yang dapat digunakan dalam sistem, diantaranya adalah fitur *bidding* (penawaran harga), menerima pemesanan dan melakukan pengerjaan pemesanan. Untuk mengikuti proses *bidding* produsen pertama harus mengakses halaman *dashboard* dimana ditampilkannya jenis-jenis pengadaan yang dapat diajukan penawaran. Setelah memilih sebuah tender, produsen dapat mengajukan penawaran harga dan juga kelengkapan yang lainnya. Untuk hasil pemenang tender akan diumumkan setelah tanggal tender berakhir yang akan ditampilkan pada halaman konsumen. Jika konsumen ada tertarik pada salah satu produsen maka konsumen akan mengajukan pengerjaan tender, dan produsen dapat mengerjakan tender yang sudah disepakati tersebut.



Gambar 4. Tampilan Pemenang Tender (Rekomendasi)

Hasil rekomendasi pemenang tender dapat dilihat pengguna (konsumen) pada menu daftar tender dengan memilih salah satu tender. Hasil rekomendasi ini hanya dapat dilihat ketika tanggal tender sudah berakhir dikarenakan ada sebuah kondisi yang mengatur jika tanggal tender sudah berakhir baru metode *ichm* akan dijalankan. Metode *ICHM* akan memproses dua data *input* yaitu *rating* dan konten, data *input rating* akan diproses nilai kemiripannya dengan menggunakan *adjust cosine similarity*, sementara data konten akan di proses menggunakan *correlation based similarity*. Setelah kedua proses tersebut selesai akan dilakukan perhitungan *total user similarity* menggunakan proses *linear combination* kemudian dilanjutkan dengan menghitung nilai prediksi *cold start problem* dan *non cold start problem*. Hasil prediksi akan diurutkan secara *descending* dan ditampilkan pada halaman konsumen.

3.2 Evaluasi Sistem

Pada bagian ini dijelaskan pengujian yang diterapkan pada sistem *e-tendering* kerajinan khas Bali.

a. Pengujian *Black Box*

Pengujian *blackbox* digunakan untuk menguji fungsionalitas dari sistem *e-tendering* kerajinan khas Bali. Pengujian ini dilakukan untuk memastikan *input* dan *output* dari sistem sudah sesuai dengan apa yang diharapkan. Berikut merupakan hasil pengujian *blackbox* dari sistem *e-tendering* kerajinan khas Bali.

Tabel 1. Hasil Pengujian *Blackbox Testing* Sistem *E-tendering*

ID	Deskripsi Pengujian	Yang Diharapkan	Hasil Pengujian	Kesimpulan
	Pengujian Pada Proses <i>Login</i>	Sistem melakukan proses <i>login</i> ke dalam sistem.	Pengguna berhasil <i>login</i> kedalam sistem	Sesuai
	Pengujian pada proses register pengguna	Sistem memberikan <i>form</i> pendaftaran dan menyimpan data pengguna pada <i>database</i>	Sistem mampu memberikan <i>form</i> pendaftaran pada pengguna serta dapat menyimpan data pengguna ke dalam <i>database</i>	Sesuai
	Pengujian pada proses pendaftaran pengadaan/tender	Sistem memberikan <i>form</i> pendafaran serta menyimpan data pengadaan pada <i>database</i>	Sistem dapat menampilkan <i>form</i> pendaftaran, setelah selesai memasukkan data, data dapat tersimpan di <i>database</i>	Sesuai
	Menampilkan data pengadaan yang ditenderkan	Sistem menampilkan daftar pengadaan yang akan ditenderkan	Sistem dapat menampilkan daftar pengadaan yang akan ditenderkan.	Sesuai
	Melakukan proses <i>bidding</i>	Sistem menangani proses <i>bidding</i>	Sistem dapat menangani proses <i>bidding</i> , dari memberikan <i>form bidding</i> sampai dengan menyimpan data <i>bidding</i> ke <i>database</i>	Sesuai
	Memberikan pemenang tender (rekomendasi)	Sistem memberikan beberapa rekomendasi pengerajin kerajinan khas Bali	Sistem dapat memberikan daftar pengerajin yang masuk rekomendasi yang paling sesuai berdasarkan metode ICHM	Sesuai

Berdasarkan hasil pengujian *blackbox testing* pada sistem *e-tendering* dapat disimpulkan bahwa hasil pengujian adalah sesuai harapan.

b. Pengujian *White Box*

Pada pengujian *white box* ini merupakan pengujian yang dilakukan untuk menguji dan menganalisis kode program bilamana terjadi kesalahan atau tidak di sebut dengan pengujian *white box*. Pengujian *white box* ini dilakukan dengan melihat *pure* kode tanpa melihat tampilan *interface* dari halaman sistem. Di dalam metode ICHM yang digunakan pada aplikasi *e-tendering* terdapat beberapa proses yang diujikan dengan *whitebox testing* diantaranya : fungsi *pearson similarity*, fungsi *adjust cosine similarity*, fungsi kombinasi linier, fungsi *coldstart item*. Karena kemiripan proses yang terjadi, hanya akan di jelaskan proses *cold start item*.

Tabel 2. Flowgraph Fungsi ColdStartItem

1	function weightedSum(ratingItem, linearSim,	
2	indexUser)	
3	coldStart = [];	
4	prediksiPembilang = 0;	
5	prediksiPenyebut = 0;	
6	for all kVal, k in ratingItem do	
7	for all iVal, i in ratingItem do	
8	prediksiPembilang += \$iVal * linearSim[\$k][\$i];	
9	prediksiPenyebut += abs(\$linearSim[\$k][\$i]);	
10	end for	
11	ws = (\$prediksiPenyebut != 0) ? prediksiPembilang / \$prediksiPenyebut : 0	
12	prediksiPembilang = 0 prediksiPenyebut = 0 ColdStart[] = ws end for return coldStart end function	

Perhitungan *cyclomatic complexity* dari fungsi Kombinasi Linier adalah sebagai berikut:

$$V(G) = E - N + 2$$

$$V(G) = 8 - 7 + 2$$

$$V(G) = 3$$

Berdasarkan hasil dari perhitungan *cyclomatic complexity* diatas didapatkan jalur pengujian sebagai berikut:

$$\text{Jalur 1} = 1 - 2 - 3 - 4 - 5 - 6, 7 - 8$$

$$\text{Jalur 2} = 1 - 2 - 3 - 4 - 5 - 3 - 4 - 5 - 6, 7 - 8$$

$$\text{Jalur 3} = 1 - 2 - 3 - 4 - 5 - 6, 7 - 2 - 3 - 4 - 5 - 6, 7 - 8$$

c. Stress Testing

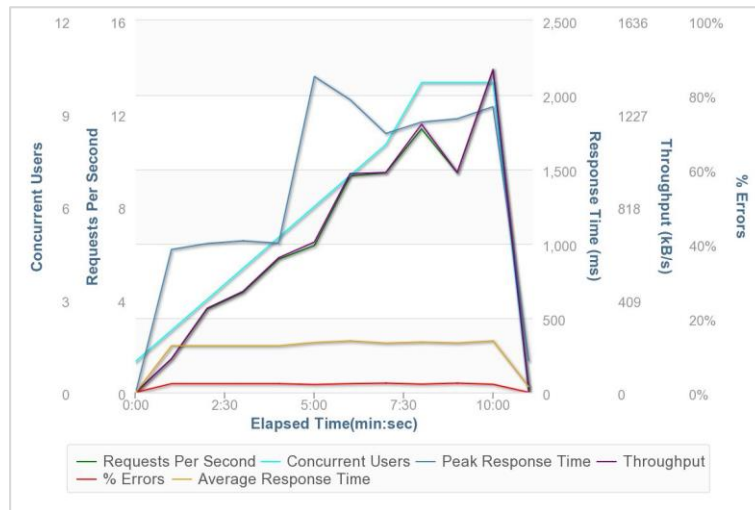
Tool yang digunakan pada *stess testing* sistem *e-tendering* kerajinan khas Bali adalah sebuah *website online* dengan alamat url LoadStrom.com.

Tabel 3. Hasil statistik dari *stress testing*

Total Requests	4,494
Total Errors	103 (2.29%)
Peak RPS	13.87
Average RPS	7.49
Peak Response Time (ms)	2,124
Average Response Time (ms)	333
Total Data Transferred	463.18
Peak Throughput (kB/s)	1,421
Average Throughput (kB/s)	771.96

Sistem e-tendering kerajinan khas Bali diberikan beban berupa 10 *user* yang akan melakukan pengaksesan ke *dashboard* sistem selama 10 menit secara bersamaan. Setiap user melakukan pengaksesan terhadap website tersebut secara terus menerus dengan interval 1 menit. *Average Response Time* menunjukkan rata-rata semua waktu *response* setiap permintaan untuk interval tertentu atau diseluruh durasi uji. *Peak Response Time* menunjukkan *response* paling lambat di antara semua permintaan untuk interval tertentu atau diseluruh durasi uji. *Total errors* menunjukkan jumlah kesalahan pada semua permintaan untuk interval tertentu, pada durasi total ditampilkan semua hitungan untuk semua kesalahan yang terjadi selama proses *stress test*.

Average RPS menunjukkan jumlah total permintaan untuk interval tertentu dan membagi totalnya dengan jumlah detik dalam interval. *Throughput* menunjukkan total byte yang diterima untuk interval tertentu dan membagi jumlah detik dalam interval, kemudian diubah menjadi nilai kilobyte untuk mendapatkan satuan kB/s. Melalui pengujian tersebut mendapatkan 4494 total request yang terjadi pada sistem dengan waktu server merespon *request* dari *user* yaitu selama 7.49 ms . Rata-rata kecepatan transper data yang dimiliki sistem sebesar 771.96.



Gambar 5. Grafik *Stress Testing* pada sistem *E-tendering*

d. *Accuracy Testing*

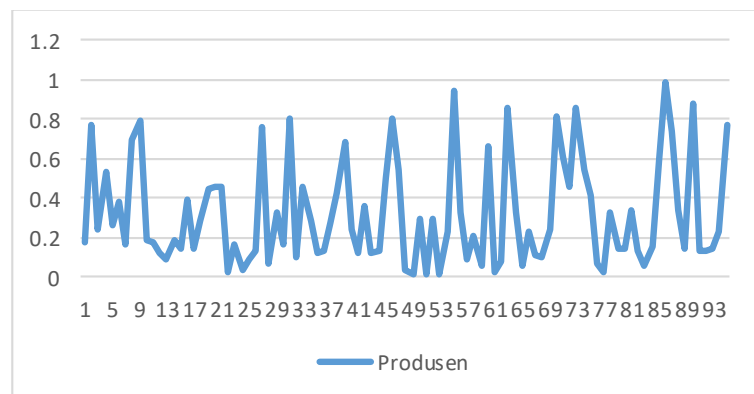
Pengujian akurasi pada sistem *e-tendering* kerajinan khas Bali menggunakan persamaan *MAE (Mean Absolute Error)*. Penggunaan persamaan ini bertujuan untuk menguji tingkat akurasi dari hasil rekomendasi yang dihasilkan oleh metode *ICHM*. Proses dari perhitungan pengujian *MAE* menggunakan hasil rekomendasi dari metode *ICHM* dan *rating* yang diberikan oleh pengguna pada sistem. Perhitungan *MAE* menggunakan persamaan sebagai berikut :

$$MAE (Produsen 1) = \sum_{i=1}^n \frac{|p_i - q_i|}{N}$$

$$MAE (Produsen 1) = \frac{|4 - 2.84 + 1 - 3.52 + 1 - 3.47 + \dots + 3 - 2.85|}{10}$$

$$MAE (Produsen 1) = 0,78$$

Perhitungan di atas merupakan perhitungan produsen dari produsen satu dengan yang mendapatkan nilai *MAE* sebesar 0,78. Berikut merupakan grafik lengkap dari dari hasil perhitungan *MAE* pada masing2 Produsen :



Gambar 6. Grafik Hasil Perhitungan *MAE*

Berdasarkan hasil pengujian dengan perhitungan MAE (*mean absolute error*) mendapatkan hasil rekomendasi yang paling akurat pada produsen ke 49, 51 dan 52 dengan nilai MAE paling rendah sebesar 0,01. Disisi lain hasil rekomendasi yang kurang akurat dari hasil perhitungan MAE pada produsen ke 86 dengan nilai paling tinggi sebesar 0.98.

4. Kesimpulan

- Berdasarkan pada penelitian yang telah dilakukan, maka dapat diambil kesimpulan meliputi :
- a. Sistem *e-tendering* produk kerajinan khas Bali yang sudah berhasil dibangun sudah mampu mengimplementasikan metode ICHM (*Item-based Clustering Hybrid Method*) dan *algoritma Slope One*.
 - b. Aplikasi sistem *e-tendering* kerajinan khas Bali dapat membantu peminat kerajinan khas Bali dengan memberikan pilihan tempat pemesanan dan memudahkan dalam pemesanan produk kerajinan khas Bali.
 - c. Produsen kerajinan yang baru masuk ke sistem atau mengalami *cold start problem* (belum memiliki rating) dapat muncul sebagai rekomendasi karena pada proses menentukan rekomendasi tidak hanya menggunakan *rating item* sebagai acuan tetapi menggunakan kombinasi *rating item* dan konten item.

References

- [1] D. J. Surjawan dan G. Apriyanti, "Sistem Lelang Tender dan Dealing Online Berbasis Web", *Jurnal Sistem Informasi*, Vol.7, No.1, hal 1-16, 2017
- [2] F. Masruri dan W.F. Mahmudy, "Personalisasi Web E-Commerce Menggunakan Recommender System dengan Metode Item-Based Collaborative Filtering", *Kursor*, 1-12, 2017
- [3] G. Adomavicius dan A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 734-749.
- [4] I. W. Tista, *Kerajinan Bali dan Budaya Bali*, Bandung: Graha Ilmu, 2017.
- [5] L. McGinty dan B. Stave, analysis of criticizing and preference based feed back in conversation on recommender system, *Adaptive selection*, 2018
- [6] M. C. Vivek dan S. Sahana, "Socio-economic characteristics of the farmers following e-tendering system for Arecanut in Karnataka", *Indian Research Journal of Extension Education*, 21, 117-125, 2021
- [7] N. Amir, "Implementasi Algoritma Cluster pada Pencarian dan Rekomendasi Tempat Kuliner dengan metode ICHM (Item Based Clustering Hybrid Method) dan Berbasis LBS di Yogyakarta", Universitas Pembangunan Nasional, 2018
- [8] R. K. Paramitha, "Rancang Bangun Sistem Rekomendasi Pemilihan Sepeda Menggunakan Metode Item-Based Clustering Hybrid", 2021
- [9] T. Theris, "Rancang Bangun Sistem Rekomendasi Pemilihan Mobil Menggunakan Metode Item-Based Clustering Hybrid", Universitas Multimedia Nusantara, 2020
- [10] Peraturan Kepala LKPP Nomor 1 Tahun 2018 tentang Etendering

Pengelompokan Pelanggan Toko Kerajinan Menggunakan K-Means dengan Model RFM dan LRFM

I Kadek Ari Surya^{a1}, Made Agung Raharja^{a2}, I Komang Ari Mogi^{a3}, Agus Muliantara^{a4}, I Gede Arta Wibawa^{a5}, I Gusti Ngurah Anom Cahyadi Putra^{a6}

^aDepartment of Informatics, Faculty of Mathematics and Natural Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia

¹ikd.arisurya@gmail.com

²made.agung@unud.ac.id

³arimogi@unud.ac.id

⁴muliantara@unud.ac.id

⁵gede.arta@unud.ac.id

⁶anom.cp@unud.ac.id

Abstract

Customer groups in a shop are important to identify to determine sales strategy. The research was conducted at one of the handicraft shops in Bali which was affected by intense market competition. The purpose of this research is to get the best customer cluster and data model. The stages in this research start from preprocessing customer data to generate Recency, Frequency, Monetary (RFM) and Length, Recency, Frequency, Monetary (LRFM) models, then clustering with K-Means and evaluating cluster quality with Silhouette Coefficient (SC). The results showed the RFM model produces an SC value of 0.545 (medium structure) when $k = 2$ and the LRFM model with the largest SC value at $k = 3$ with a value of 0.415 (weak structure). SC values in both models tend to increase as the percentage of data increases. Cluster 1 has 817 members with the last transaction taking a long time, but has below average orders and monetary. Cluster 2 has 158 members with the last transaction at the most recent time, and with above average orders and monetary.

Keywords: Clustering, Customer, Craft, K-Means, RFM, LRFM, Silhouette Coefficient

1. Pendahuluan

Mengenali karakteristik kelompok pelanggan dalam suatu bidang bisnis menjadi hal yang penting untuk dapat mempertahankan pelanggan. Karakteristik pelanggan dapat dijadikan acuan dalam menentukan strategi pemasaran dalam menjaga aliran bisnis tetap terjaga pada suatu perusahaan. Hal ini perlu dipertimbangkan oleh berbagai bidang usaha mengingat ketatnya persaingan pasar pada era sekarang.

Industri kerajinan merupakan bidang usaha dengan perkembangan inovasi yang begitu pesat dan persaingan pasar yang ketat. Bidang industri kerajinan terhitung sebagai salah satu bidang yang berperan besar terhadap perekonomian masyarakat. Pemulihan ekonomi daerah harus didukung oleh peran aktif masyarakat dan pemerintah agar dapat menghasilkan kondisi perekonomian yang lebih baik [1]. Bali merupakan daerah di Indonesia yang aktif memproduksi produk kerajinan. Namun beberapa tahun belakangan ini dari data Dinas Perdagangan dan Perindustrian (Disperindag) Provinsi Bali pada tahun 2019 terjadi penurunan permintaan export sebesar 13% dan tahun 2020 sebesar 29.44% [2] [3]. Kondisi ini berdampak pada industri kecil lainnya salah satunya toko kerajinan oleh2bali.com. Toko ini memiliki pencatatan data pelanggan dan order yang dapat diolah dengan data mining yakni teknik *clustering* untuk dapat membantu mengenali karakteristik kelompok pelanggan.

Salah satu metode yang populer digunakan untuk mengelompokkan data adalah K-Means. Penelitian pada kasus *customer segmentation* menggunakan perbandingan K-Means, K-Medoids, dan DBSCAN menunjukkan hasil bahwa algoritma K-Means memiliki akurasi paling baik [4]. Metode K-Means sering digunakan karena bersifat dinamis pada data tersebar serta mudah diterapkan dan diinterpretasikan

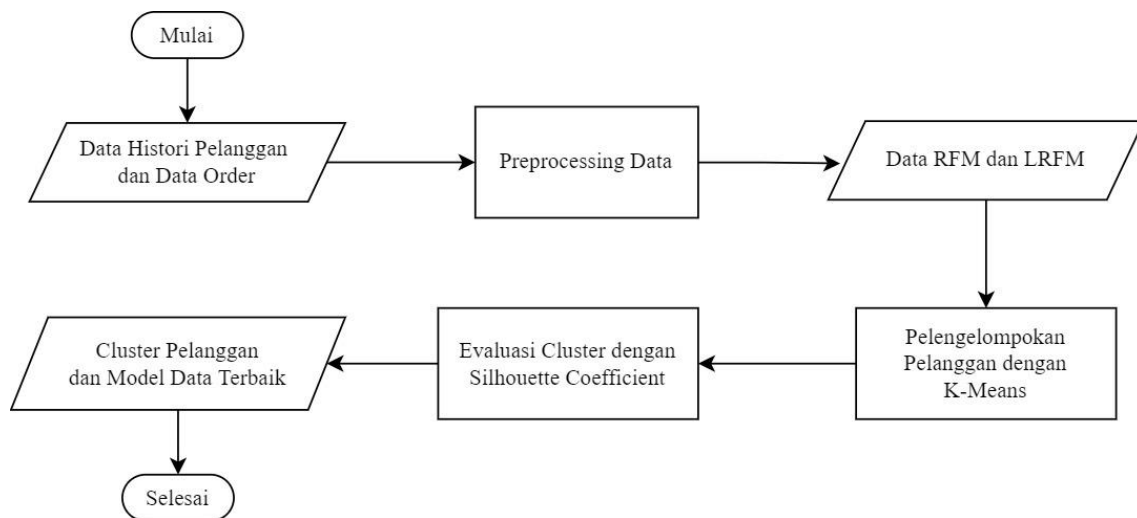
[5]. Dalam mengimplementasikan teknik *clustering* menggunakan metode tertentu seperti K-Means dapat menerapkan beberapa model data.

RFM adalah model data yang dapat digunakan untuk menganalisis perilaku pelanggan, seperti seberapa baru pelanggan melakukan pembelian (*Recency*), seberapa sering pelanggan membeli (*Frequency*), dan berapa banyak uang yang dihabiskan pelanggan dalam melakukan transaksi (*Monetary*) [4]. Model RFM telah banyak digunakan dalam dunia pemasaran terutama untuk segmentasi pasar [6]. Selain model RFM terdapat juga model LRFM. Model LRFM adalah model yang berasal dari pengembangan RFM dengan menambahkan variabel *Length*. Model LRFM merupakan metode analisis nilai pelanggan yang diterapkan pada segmentasi pelanggan [7]. Model LRFM membedakan antara pelanggan yang mempunyai hubungan jangka panjang atau jangka pendek dalam suatu perusahaan [8].

Berdasarkan penjelasan di atas pada penelitian ini bertujuan untuk mengetahui cluster pelanggan dengan menggunakan metode K-Means dan mengetahui model data terbaik dengan membandingkan hasil clustering dengan model data RFM dan LRFM pada proses clustering pelanggan di toko kerajinan tempat studi kasus.

2. Metode Penelitian

Proses penelitian dimulai dari pengumpulan data, *preprocessing* data, proses pengelompokan pelanggan menggunakan K-Means, dan perbandingan perhitungan *silhouette coefficient* digunakan dalam menentukan kualitas cluster hasil dari model RFM dan LRFM. Tahapan penelitian pada dilihat pada *flowchart* yang ditunjukkan pada Gambar 1.



Gambar 1. Desain Umum Penelitian

2.1 Pengumpulan Data

Data pada penelitian menggunakan data sekunder yang diperoleh dari toko kerajinan oleh2bali.com yang berlokasi di Denpasar, Bali. Data yang didapatkan adalah data histori pelanggan dan data order dari tahun 2014-2020. Terdapat 975 data pelanggan dan 2463 data order. Format data berupa *comma separated values* (.csv). Data histori pelanggan terdiri dari atribut *id_customer*, *name*, *last_active*, *email*, *orders*, *total_spend*, *aov*, *country*, *city*, *region*. Sedangkan data order terdiri atribut *id_order*, *id_customer*, *name*, *status*, *date*, *total*. Berikut ini adalah atribut dari data yang yang didapatkan yang ditunjukkan pada Tabel 1 dan Tabel 2:

Tabel 1. Deskripsi Data Histori Pelanggan

Nama Field	Tipe Data	Keterangan
<i>id_customer</i>	<i>string</i>	Id unik dari pelanggan.
<i>name</i>	<i>string</i>	Nama pelanggan.

<i>last_active</i>	<i>datetime</i>	Tanggal terakhir pelanggan melakukan pemesanan barang.
<i>email</i>	<i>string</i>	Alamat email pelanggan
<i>order</i>	<i>integer</i>	Total order yang sudah terbayarkan oleh pelanggan.
<i>total_spend</i>	<i>integer</i>	Total uang yang dihabiskan oleh pelanggan untuk total order produk.
<i>aov</i>	<i>integer</i>	<i>Average Order Value</i> adalah rata-rata pembelian pelanggan.
<i>country</i>	<i>string</i>	Negara asal dari pelanggan.
<i>city</i>	<i>string</i>	Kota asal dari pelanggan.
<i>region</i>	<i>string</i>	Provinsi asal dari pelanggan.

Table 2. Deskripsi Data Order

Nama Field	Tipe Data	Keterangan
<i>id_order</i>	<i>string</i>	Id unik dari setiap order.
<i>id_customer</i>	<i>string</i>	Id unik dari customer.
<i>name</i>	<i>string</i>	Nama pelanggan yang melakukan order.
<i>status</i>	<i>string</i>	Status order yang menunjukkan sudah completed atau belum completed order.
<i>date</i>	<i>date</i>	Tanggal order dilakukan.
<i>total</i>	<i>integer</i>	Total uang yang harus dibayarkan untuk melakukan setiap order.

2.2 Preprocessing Data

Preprocessing data merupakan proses perbaikan untuk mengolah data mentah menjadi format data yang dibutuhkan pada penelitian. Pada penelitian ini *preprocessing* data dimulai dari melakukan *data cleaning*, *data selection* dan *data transformation*. Pada *data cleaning* dilakukan penghapusan *record* data yang memiliki nilai *null* atau *missing value*. Kemudian pada *data selection* dipilih beberapa atribut data yang dibutuhkan pada penelitian terutama untuk memperoleh model data RFM dan LRFM. Berikut ini adalah penjelasan mengenai atribut *recency* (R), *frequency* (F), *monetary* (M), dan *length* (L).

a. *Recency* (R)

Recency merupakan jarak waktu antara waktu penelitian dengan waktu terakhir pelanggan melakukan transaksi. Pelanggan yang paling baru memiliki nilai *recency* yang kecil dan berpotensi dalam melakukan pembelian lagi jika dibandingkan pelanggan dengan waktu terakhir transaksi yang cukup lama. *Recency* dapat dicari dengan menggunakan persamaan 1.

$$R_{(i)} = TR - TL_{(i)} \quad (1)$$

Dimana:

$R_{(i)}$: *recency* pada pelanggan i

TR : tanggal penelitian

$TL_{(i)}$: tanggal terakhir pelanggan i melakukan transaksi

b. *Frequency* (F)

Frequency merupakan total jumlah transaksi yang dilakukan pada jangka waktu yang ditentukan. Nilai *frequency* yang tinggi menandakan tingginya tingkat loyalitas pelanggan. *Frequency* dapat dicari dengan menggunakan persamaan 2.

$$F_{(i)} = \sum_{t=TA}^{t=TR} fr_{(t)} \quad (2)$$

Dimana:

$F_{(i)}$: *frequency* pada pelanggan i

t : waktu dalam tanggal

TA : tanggal awal yang ditentukan

TR : tanggal penelitian

$fr_{(t)}$: jumlah transaksi yang dilakukan pelanggan i pada setiap waktu t

c. *Monetary* (M)

Monetary merupakan nilai total uang yang dihabiskan pelanggan pada periode tertentu dalam melakukan transaksi. Nilai *monetary* yang tinggi menunjukkan bahwa kontribusi pelanggan terbilang besar terhadap perusahaan. *Monetary* dapat dicari dengan menggunakan persamaan 3.

$$M_{(i)} = \sum_{t=TA}^{t=TR} mr_{(t)} \quad (3)$$

Dimana:

$M_{(i)}$: *monetary* pada pelanggan i

t : waktu dalam tanggal

TA : tanggal awal yang ditentukan

TR : tanggal penelitian

$mr_{(t)}$: jumlah uang yang dibayarkan pelanggan i untuk setiap transaksi pada waktu t

d. *Length* (L)

Length menyatakan jarak antara waktu transaksi terakhir pada periode tertentu dengan waktu awal pelanggan melakukan transaksi. *Length* dapat dicari dengan menggunakan persamaan 4.

$$L_{(i)} = TL_{(i)} - TF_{(i)} \quad (4)$$

Dimana:

$L_{(i)}$: *length* pada pelanggan i

$TL_{(i)}$: tanggal terakhir pelanggan i melakukan transaksi

$TF_{(i)}$: tanggal pertama kali pelanggan i melakukan transaksi

Berdasarkan penjelasan di atas model data RFM dan LRFM dapat dihitung dari ketersediaan data penelitian. Data *recency* didapatkan dari tanggal penelitian dikurangi dengan tanggal *last_active*, data *frequency* bisa didapatkan dari atribut *order* dan *monetary* bisa didapatkan dari atribut *total_spend*. Sedangkan untuk atribut data LRFM (*length*, *recency*, *frequency*, *monetary*) yakni atribut *length* nilainya dapat dicari dari atribut *date* pada dataset order saat pertama kali pelanggan melakukan transaksi dikurangi dengan tanggal *last_active* pada dataset histori pelanggan.

Preprocessing dilanjutkan pada proses *data transformation*. Transformasi data dilakukan dengan menyamakan rentang data dari 0 – 1 dengan metode *Min Max Normalization*. Berikut ini merupakan hasil normalisasi data yang ditunjukkan pada Tabel 3 dan Tabel 4.

Tabel 3. Normalisasi Data RFM

Id	R	F	M
1	0.00000	1.00000	0.28903
2	0.00590	0.33333	0.10008
3	0.01417	0.16667	0.02802
973	0.89335	0.00000	0.14812
974	0.90437	0.16667	0.28423
975	0.91814	0.33333	0.02482

Tabel 4. Normalisasi Data LRFM

Id	L	R	F	M
1	0.70769	0.00000	1.00000	0.28903
2	0.14615	0.00590	0.33333	0.10008
3	0.31154	0.01417	0.16667	0.02802
973	0.00000	0.89335	0.00000	0.14812
974	0.11923	0.90437	0.16667	0.28423
975	0.20769	0.91814	0.33333	0.02482

2.3 Pengelompokan Pelanggan dengan K-Means

K-Means merupakan metode *clustering* non hirarki dengan membagi data menjadi satu atau lebih kelompok data (cluster) [9]. Metode K-Means memiliki konsep yang sederhana dan umum digunakan. Selain itu K-Means dapat melakukan pengelompokan data dalam jumlah besar dengan waktu komputasi yang terbilang cepat serta efisien. Berikutnya dijelaskan mengenai tahapan dari metode *clustering* menggunakan algoritma K-Means:

- Tentukan jumlah cluster sebanyak k cluster dan dataset yang akan dikelompokkan.
- Dipilih *initial centroid* sebanyak k data. Pada penelitian ini *initial centroid* dipilih secara random.
- Mencari jarak pada setiap objek data dengan masing-masing *centroid*. Perhitungan jarak dapat dilakukan menggunakan persamaan ukuran jarak *Euclidean Distance* dengan persamaan:

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n d(x_i - y_i)^2} \quad (5)$$

- Mengalokasikan masing-masing objek data ke cluster dengan jarak paling kecil atau minimum.
- Tentukan centroid cluster yang baru dengan menghitung rata-rata dari semua data yang terdapat di dalam cluster tersebut.
- Lakukan langkah c, d, dan e sampai tidak ada perubahan objek data dalam suatu cluster.

2.4 Evaluasi Cluster dengan Silhouette Coefficient

Silhouette Coefficient diperkenalkan oleh Rousseeuw pada tahun 1987. *Silhouette Coefficient* digunakan untuk mengetahui kualitas dari cluster data yang dihasilkan oleh algoritma *clustering* dengan mengkombinasikan konsep *cohesion* dan *separation*. Tahap perhitungannya sebagai berikut [10]:

- Menghitung rata-rata jarak suatu objek i ke semua objek selain objek tersebut dalam suatu cluster yang dilambangkan dengan a(i) :

$$a_i = \frac{1}{|A|-1} \sum_j \epsilon_{A,j \neq i} d(i,j) \quad (6)$$

Dengan A adalah cluster tempat objek tersebut berada, d(i,j) merupakan jarak antar objek pada cluster yang sama.

- Kemudian cari rata-rata jarak dari objek data tersebut dengan objek lain dalam cluster yang berbeda, kemudian pilih nilai rata-rata jarak yang terkecil :

$$b(i) = (D(i, C)) \quad (7)$$

- Setelah nilai a(i) dan b(i) didapatkan, selanjutnya dapat dihitung *silhouette coefficient* dari objek ke i :

$$s(i) = \frac{b(i)-a(i)}{(a(i),b(i))} \quad (8)$$

Berdasarkan kategori nilai rata-rata *silhouette coefficient* (SC) menurut Rousseeuw, hasil perhitungan nilai SC yang nilainya mendekati 1, menandakan bahwa kualitas cluster yang dihasilkan semakin bagus. Kategori nilai SC berdasarkan pendapat Rousseeuw dapat dilihat sebagai berikut:

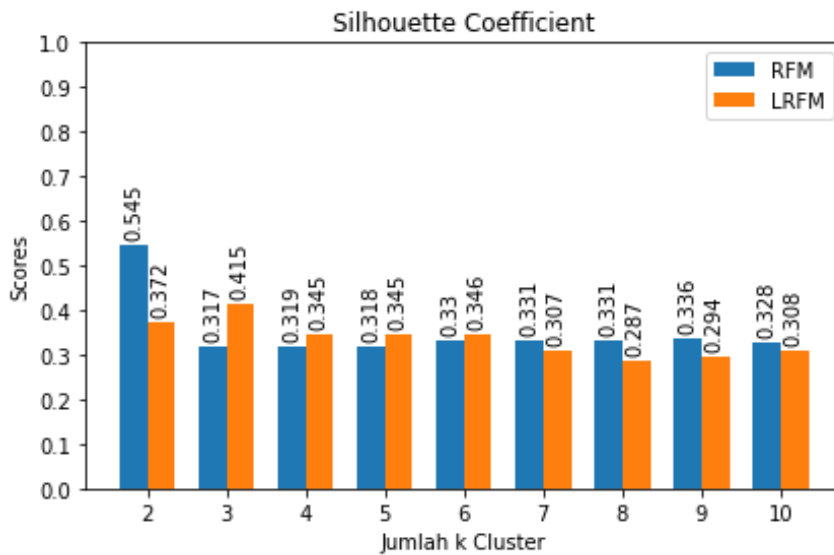
- SC ≤ 0,25 (*no structure*)
- 0,25 < SC ≤ 0,5 (*weak structure*)
- 0,5 < SC ≤ 0,7 (*medium structure*)
- 0,7 < SC ≤ 1 (*strong structure*)

3. Hasil dan Pembahasan

Pada bagian ini menjelaskan mengenai perbandingan *silhouette coefficient* yang dihasilkan dari model data RFM dan LRFM. Selain itu juga dijelaskan mengenai hasil pengelompokan terbaik dari setiap model data beserta karakteristiknya.

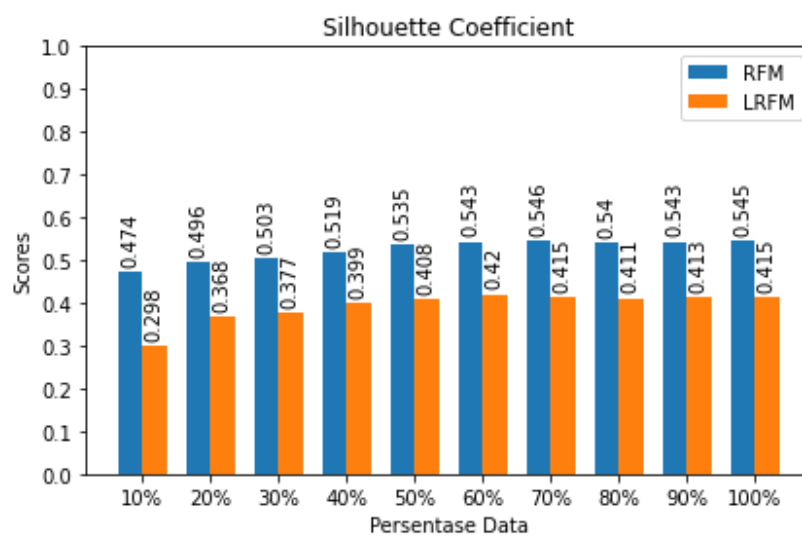
3.1. Perbandingan Silhouette Coefficient

Perbandingan *silhouette coefficient* (SC) model RFM dan LRFM bertujuan untuk mengetahui model data yang menghasilkan hasil *clustering* terbaik. Proses *clustering* dilakukan pada jumlah cluster $k = 2$ sampai $k = 10$ untuk mengetahui nilai SC terbaik pada setiap model.



Gambar 2. *Silhouette Coefficient* pada Model RFM dan LRFM

Pada Gambar 2 terlihat nilai SC terbesar dihasilkan oleh model RFM pada saat jumlah $k = 2$ dengan nilai 0.545. Seiring dengan pertambahan jumlah k cluster nilai SC pada model RFM mengalami penurunan saat $k = 3$ dan perubahan nilai SC setelahnya tidak terlalu besar. Sedangkan model LRFM hanya menghasilkan nilai SC terbesar 0.415 pada saat jumlah $k = 3$. Berikutnya juga dilakukan perbandingan nilai SC pada variasi persentase data.



Gambar 3. *Silhouette Coefficient* pada Variasi Persentase Data Model RFM dan LRFM

Perbandingan SC juga dilakukan menggunakan variasi persentase data RFM dan LRFM dari 10% hingga 100% dengan jumlah k cluster terbaik pada masing-masing model. Pengujian ini dilakukan

untuk mengenali pengaruh dari penggunaan variasi persentase data terhadap nilai SC yang dihasilkan. Dari grafik yang ditampilkan pada Gambar 3 terlihat bahwa model data RFM selalu menghasilkan skor SC lebih besar dibandingkan dengan model data LRFM. Seiring dengan bertambahnya variasi persentase data skor terbesar SC yang dihasilkan pada model data RFM dan LRFM sama-sama cenderung meningkat walaupun tidak terlalu besar.

Tabel 5. Keanggotaan Cluster pada Model RFM Berdasarkan *Silhouette Coefficient*

Cluster Pelanggan	<i>Strong Structure</i>	<i>Medium Structure</i>	<i>Weak Structure</i>	<i>No Structure</i>	Total
Cluster 1	504	267	30	16	817
Cluster 2	0	97	40	21	158
Total	504	364	70	37	

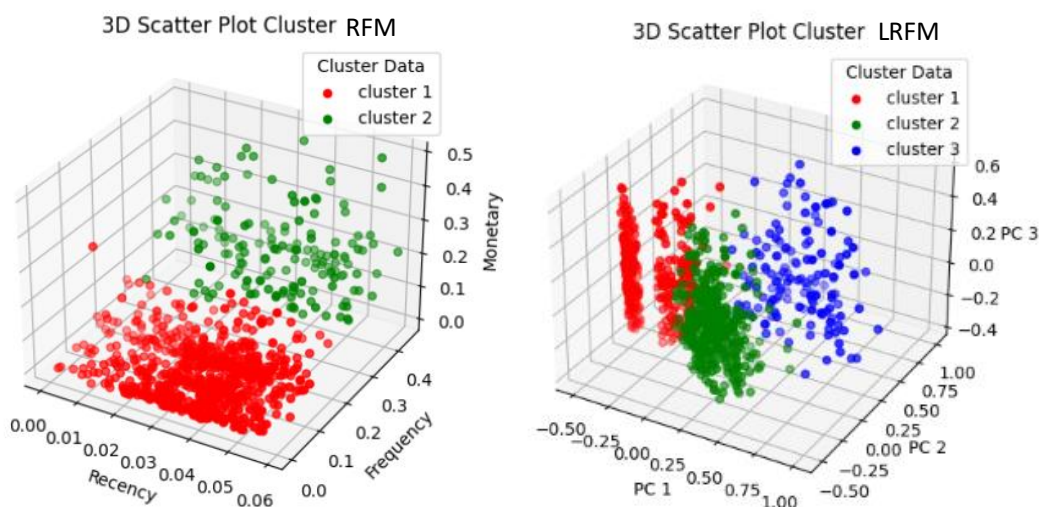
Tabel 6. Keanggotaan Cluster pada Model LRFM Berdasarkan *Silhouette Coefficient*

Cluster Pelanggan	<i>Strong Structure</i>	<i>Medium Structure</i>	<i>Weak Structure</i>	<i>No Structure</i>	Total
Cluster 1	0	237	133	69	439
Cluster 2	0	172	154	65	391
Cluster 3	0	0	79	66	145
Total	0	409	366	200	

Pada Tabel 5 dan Tabel 6 merupakan keanggotaan cluster pada model RFM dan LRFM berdasarkan pada nilai SC terbaik pada masing-masing model data. Model RFM menghasilkan cluster pelanggan yang lebih baik dibandingkan menggunakan model LRFM pada toko kerajinan tempat studi kasus karena nilai SC terbesar yang dihasilkan model RFM sebesar 0.545. Selain itu dari struktur keanggotaan cluster, model RFM memiliki *strong structure* sebanyak 504 anggota dan hanya memiliki *weak structure* sebanyak 37 anggota. Dibandingkan dengan hasil clustering terbaik LRFM tidak memiliki anggota *strong structure* tetapi banyak memiliki *no structure* yakni 200 anggota.

3.2. Perbandingan Hasil Cluster Terbaik

Berikut ini merupakan visualisasi dari hasil *clustering* terbaik pada setiap model data. Visualisasi ditampilkan melalui *scatterplot*.



Gambar 4. *Scatterplot* Cluster Pelanggan dengan Model RFM dan LRFM

Pada Gambar 4 visualisasi cluster pelanggan pada model RFM ditampilkan melalui *3D scatterplot* dengan sumbu *recency*, *frequency*, dan *monetary*. Pada model RFM pengelompokan pelanggan menjadi dua cluster juga terlihat cukup baik karena tidak banyak data anggota cluster yang tumpang

tindih. Sedangkan pada model LRFM memiliki empat atribut data dan untuk dapat menampilkan cluster pelanggan pada *3D scatterplot* data model LRFM perlu dilakukan reduksi dimensi. Reduksi dimensi yang digunakan adalah *Principal Component Analysis* (PCA). Visualisasi cluster pelanggan model LRFM juga terlihat cukup baik namun pengelompokan data pelanggan pada model RFM terlihat lebih baik dibandingkan LRFM. Sehingga berdasarkan pengujian sebelumnya dan visualisasi cluster hasil *clustering* terbaik dihasilkan oleh model RFM. Masing-masing cluster terbaik yang dihasilkan model RFM memiliki karakteristiknya masing-masing.

a. Cluster 1

Cluster 1 memiliki total 817 pelanggan. Berdasarkan *centroid* Cluster 1 pelanggan melakukan pembelian terakhir dalam waktu yang sudah lama, dan memiliki intensitas order dan *monetary* yang di bawah rata-rata. Kemudian berdasarkan nilai aktual dari setiap atribut data pada Cluster 1 adalah pelanggan memiliki rata-rata nilai *recency* sebesar 1492, *frequency* sebesar 2, dan *monetary* sebesar Rp. 65.712. Pelanggan mancanegara dari cluster ini sebanyak 1.5% yang berasal dari Canada, Germany, Prancis, Hongkong, Amerika, Singapore, Malaysia dan Jepang. Berdasarkan tiga wilayah *region* paling banyak berasal dari daerah Jawa Barat, DKI Jakarta, Jawa Timur dengan dominan waktu order pada bulan Oktober, Agustus, Juli yang tepatnya pada pertengahan bulan sekitar jam 11.00 hingga 14.00 siang.

b. Cluster 2

Cluster 2 memiliki total 158 pelanggan. Berdasarkan *centroid* Cluster 2 pelanggan melakukan pembelian terakhir pada waktu terkini, dan memiliki intensitas order dan *monetary* yang di atas rata-rata. Kemudian berdasarkan nilai aktual dari setiap atribut data pada Cluster 2 adalah pelanggan memiliki rata-rata nilai *recency* sebesar 1490, *frequency* sebesar 6, dan *monetary* sebesar Rp. 277.943. Pelanggan mancanegara dari cluster ini sebanyak 3.5% yang berasal dari Malaysia, Canada, dan Hungary. Berdasarkan tiga wilayah *region* paling banyak berasal dari daerah DKI Jakarta, Jawa Barat, Banten dengan dominan waktu order pada bulan Oktober, November, Desember yang tepatnya pada pertengahan bulan sekitar jam 8.00 hingga 10.00 pagi.

4. Kesimpulan

Kesimpulan dari penelitian ini mendapatkan hasil bahwa model RFM menghasilkan kelompok pelanggan terbaik pada toko kerajinan tempat studi kasus dengan nilai *silhouette coefficient* sebesar 0.545 (*medium structure*) saat $k = 2$ dibandingkan dengan model LRFM yang hanya 0.415 saat $k = 3$. Nilai *silhouette coefficient* yang dihasilkan pada setiap model cenderung meningkat ketika terjadi penambahan persentase data. Kemudian terdapat dua kelompok pelanggan terbaik yakni Cluster 1 memiliki anggota sebanyak 817 dengan melakukan pembelian terakhir dalam waktu yang cukup lama, dan memiliki intensitas order dan *monetary* yang di bawah rata-rata. Cluster 2 memiliki anggota sebanyak 158 dengan melakukan transaksi terakhir pada waktu terkini, dan memiliki intensitas order dan *monetary* yang di atas rata-rata.

Referensi

- [1] M. A. Raharja, I. D. M. B. A. Darmawan, D. P. E. Nilakusumawati, and I. W. Supriana, "Analysis of membership function in implementation of adaptive neuro fuzzy inference system (ANFIS) method for inflation prediction," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012005.
- [2] Disperindag Bali, "Laporan Kinerja Instansi Pemerintah (LKjIP) Dinas Perdagangan dan Perindustrian Provinsi Bali 2019," *Dinas Perdagangan dan Perindustrian Provinsi Bali*, 2019, [Online]. Available: <https://disperindag.baliprov.go.id/lkjip-2019-dinas-perindustrian-dan-perdagangan-provinsi-bali/>
- [3] Disperindag Bali, "Laporan Kinerja Instansi Pemerintah (LKjIP) Dinas Perdagangan dan Perindustrian Provinsi Bali 2020," *Dinas Perdagangan dan Perindustrian Provinsi Bali*, 2020, [Online]. Available: <https://disperindag.baliprov.go.id/wp-content/uploads/2021/03/LKJiP-Disperindag-2020.pdf>
- [4] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
- [5] A. A. A. Rospricilia, Tita Ayu, Syurfah Ayu Ithriah, "Segmentasi Pelanggan Menggunakan

- Metode K-Means Clustering Berdasarkan Model RFM,” *J. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 699–708, 2020.
- [6] B. E. Adiana, I. Soesanti, and A. E. Permanasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik Clustering,” *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.
- [7] S. Monalisa, P. Nadya, and R. Novita, “Analysis for Customer Lifetime Value Categorization with RFM Model,” in *Procedia Computer Science*, 2019, vol. 161, pp. 834–840. doi: 10.1016/j.procs.2019.11.190.
- [8] R. Rahmadiani, A. Dhini, and E. Laoh, “Estimating Customer Lifetime Value using LRFM Model in Pharmaceutical and Medical Device Distribution Company,” in *2020 International Conference on ICT for Smart Society (ICISS)*, 2020, pp. 1–5. doi: 10.1109/ICISS50791.2020.9307592.
- [9] M. A. Raharja and I. W. Supriana, “Analisis Klasifikasi Tingkat Kesehatan Lembaga Perkreditan Desa (Lpd) Menggunakan Metode K-Means Clustering,” *J. Teknol. Inf. dan Komput.*, vol. 5, no. 1, pp. 83–90, 2019, doi: 10.36002/jutik.v5i1.641.

This page is intentionally left blank

Aplikasi Website Pengamanan File Dokumen Menggunakan Kriptografi RSA

I Made Ari Widiarsana^{a1}, I Gusti Ngurah Anom Cahyadi Putra^{a2}
I Ketut Gede Suhartana^{a3}, Luh Gede Astuti^{a4}, I Putu Gede Hendra Suputra^{a5} I Wayan Supriana^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia

¹awidiarsana.aw@gmail.com

²anom.cp@unud.ac.id

³ikg.suhartana@unud.ac.id

⁴lg.astuti@unud.ac.id

⁵hendra.suputra@unud.ac.id

⁶wayan.supriana@unud.ac.id

Abstract

Data security is something that needs to be considered in maintaining the confidentiality of information, especially those that only contain information that can be known by the authorized party. There are still many cases of data leakage that occur in Indonesia, especially in documents. Documents can be secured using cryptographic techniques. One of the well-known cryptography is RSA Cryptography. The security of RSA cryptography lies in the difficulty of factoring large numbers into prime factors. In previous research conducted by (Azhar & Yuliany, 2019)[1] they could only encrypt files with the .pdf extension and also the images contained in the document were not successfully encrypted. The implementation of RSA cryptography will be made using the python programming language based on the website. The system created has a success rate of 100% in encrypting documents for each document, and in decrypting it has a success rate of 85% to 96%.

Keywords: Data Security, Cryptography, RSA, Website, Django, Python

1. Pendahuluan

Pada saat ini penggunaan internet sudah seperti kebutuhan primer di masyarakat dunia. Teknologi internet semakin hari semakin kian berkembang. Keamanan data merupakan hal yang perlu diperhatikan dalam menjaga kerahasiaan informasi, terutama yang hanya memuat informasi yang dapat diketahui oleh pihak yang berwenang. Ada risiko penyadapan saat mengirim data atau informasi tanpa keamanan, dan mudah bagi pihak yang tidak berwenang untuk menemukan informasi yang terkandung di dalamnya. Keamanan dilakukan dengan melindungi keamanan file untuk menghindari masalah dengan pihak jahat yang dapat membuka kerahasiaan informasi. Pencarian celah keamanan adalah proses menyembunyikan atau menyamarkan informasi dengan cara yang tidak jatuh ke pihak lain yang tidak berkepentingan. Pada saat ini di Indonesia masih banyak kasus kebocoran data. Pada awal Mei 2020, sebanyak 91 data pengguna dari Tokopedia bocor dan dijual di situs gelap (dark web). Pada 21 Mei 2020 Daftar Pemilih Tetap (DPT) Pemilu 2014 yang dalam bentuk file pdf bocor di situs dan forum komunitas hacker. Data yang dihimpun mencakup sejumlah informasi sensitif, seperti nama lengkap, nomor kartu keluarga, Nomor Induk Kependudukan (NIK), tempat dan tanggal lahir, alamat rumah, serta beberapa data pribadi lainnya [2]. Pada tahun 2021 kembali lagi terjadi kasus kebocoran data pribadi yang kemungkinan adalah data dari BPJS Kesehatan yang telah di *upload* di internet. Data yang bocor tersebut ditemukan dalam bentuk *file excel* yang sudah di *upload* ke dalam komunitas *hacker*. Diperkirakan data yang bocor sebanyak 20 juta data yang meliputi NIK, nomor HP, alamat, alamat email, Nomor Pokok Wajib Pajak (NPWP), tempat tanggal lahir, jenis kelamin, jumlah tanggungan, serta foto [3]

Ilmu yang mempelajari penyandian atau pengkodean terhadap suatu berkas disebut dengan Kriptografi [4]. Sehingga data yang bersifat pribadi seperti NIK, NPWP, Nomor HP yang tercatat dalam dokumen harus di lindungi agar tidak disalahgunakan oleh pihak-pihak yang tidak berkepentingan [5]. Algoritma kriptografi yang paling terkenal yaitu algoritma RSA (Riverst Shamir Adleman). RSA ditemukan pada tahun 1976 oleh tiga peneliti MIT (Massachusetts Institute of Technology), Ron Rivest, Adi Shamir, dan Len Adleman. Pada penelitian yang dilakukan oleh Rakhmat Kurniawan pada tahun 2017 juga menggunakan algoritma RSA dalam pengamanan *file* dokumen, akan tetapi pada penelitian tersebut hanya terbatas pada enkripsi *text* dan juga kunci yang digunakan untuk enkripsi dan dekripsi sudah ditetapkan diawal (tidak ada pembangkitan kunci RSA)[6]. Keamanan algoritma RSA adalah sulitnya memfaktorkan bilangan besar menjadi faktor prima. Proses dari pembangkitan kunci RSA adalah :

- Tentukan dua bilangan prima sembarang p dan q
- Hitung nilai n dengan $p \times q$. Disarankan bahwa nilai p dan q tidak sama, karena jika $p = q$ maka nilai $n = p^2$, maka nanti nilai p dapat diperoleh dengan akar kuadrat dari n
- Hitung nilai dari $\phi(n) = (p - 1)(q - 1)$
- Pilih kunci publik e yang relatif prima dengan nilai $\phi(n)$
- Pembangkitan kunci privat menggunakan persamaan $e \times d = 1 \pmod{\phi(n)}$.

Yang akan mendapatkan nilai kunci publik adalah (e, n) dan kunci privat adalah (d, n) . Proses dari enkripsi RSA adalah $c = m^e \pmod n$. Dan untuk proses dekripsi dari algoritma RSA adalah $m = c^d \pmod n$ [7]. Tingkat keamanan dari sebuah pembangkitan kunci RSA adalah semakin panjang kunci yang dibangkitkan (nilai p dan q besar) maka semakin kuat pula kunci yang dibangkitkan.

2. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif, dan menggunakan metode *waterfall* untuk metode pengembangannya. Menggunakan metode ini karena pengembangan dilakukan secara berurutan. Sehingga hasil yang didapatkan akan lebih optimal

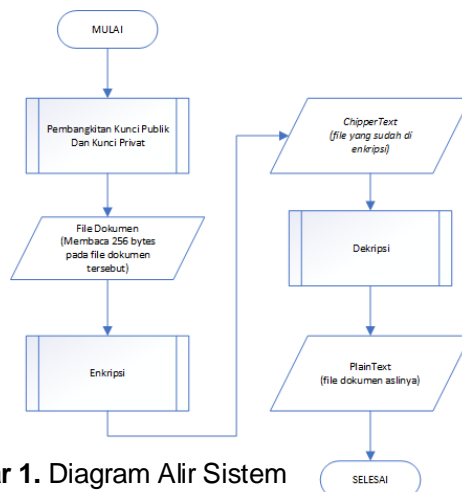
2.1. Gambaran Umum Aplikasi

Implementasi Kriptografi RSA dalam Pengamanan *File* Dokumen dalam paper ini memiliki tiga kategori utama yaitu enkripsi dokumen, dekripsi dokumen dan juga pembangkitan kunci RSA.

- Pada proses pembangkitan kunci *user* diminta untuk memilih dua buah bilangan prima yang nilainya tidak boleh sama dan juga minimal 3 digit angka.
- Pada enkripsi dokumen, pengguna akan *mengupload* dokumen berupa *.docx*, *.pptx*, *.xlsx*, atau *.pdf* dan juga diminta untuk memasukkan kunci publik dari proses pembangkitan kunci RSA. Sistem akan melakukan enkripsi terhadap nilai dari *hexabytes* pada dokumen tersebut.
- Pada dekripsi dokumen, pengguna akan *mengupload* dokumen berupa *.docx*, *.pptx*, *.xlsx*, atau *.pdf* dan juga diminta untuk memasukkan kunci privat dari proses pembangkitan kunci RSA. Sistem akan melakukan dekripsi terhadap nilai dari *hexabytes* pada dokumen tersebut menggunakan algoritma RSA.

2.2. Desain Aplikasi

Aplikasi yang dibuat menggunakan bahasa pemrograman python, dengan framework django dan tampilan antarmuka menggunakan html, css, jquery, dan juga bootstrap. Berikut adalah *flowchart* yang digunakan dalam Implementasi Kriptografi RSA dalam Pengamanan File Dokumen di paper ini.

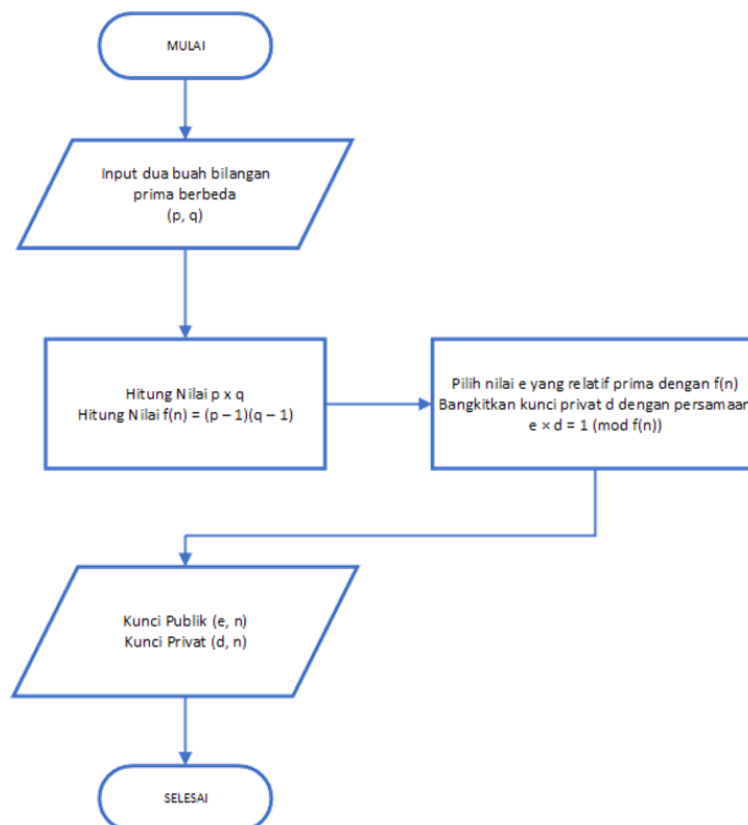


Gambar 1. Diagram Alir Sistem

Penjelasan diagram alir sistem:

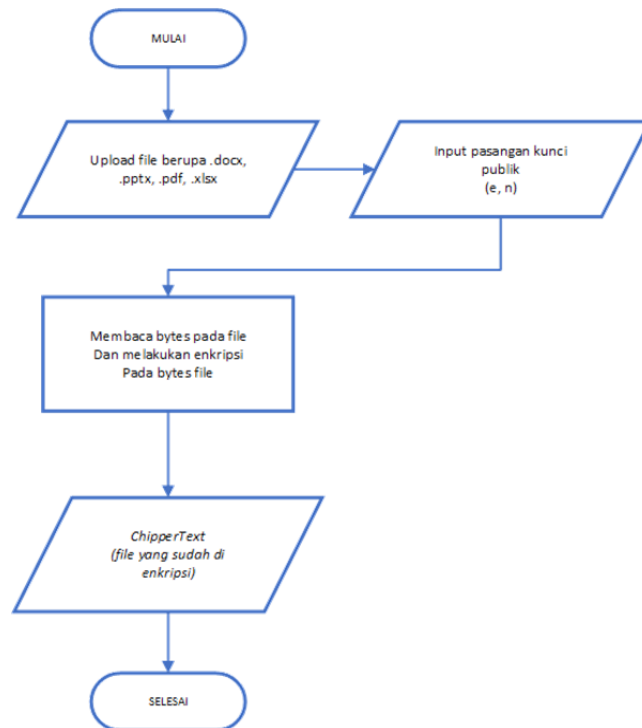
- Mulai
- Membaca file dokumen yang di upload ke sistem, dan membaca 256 bytes di dalam file dokumen tersebut.
- Melakukan proses pembangkitan kunci RSA, disini user diminta untuk memilih dua buah bilangan prima untuk proses pembangkitan kunci RSA
- Melakukan enkripsi 256 bytes dokumen yang sudah di upload menggunakan kunci publik yang sudah di berikan
- Menghasilkan output file chipertext yang dapat di unduh namun tidak dapat dibuka jika belum dilakukan dekripsi terhadap file tersebut
- Proses dekripsi yang dimana nantinya user mengupload file chipertext dan menggunakan kunci privat untuk dekripsinya.
- Akan menghasilkan output berupa plaintext atau file dokumen sama seperti aslinya (sebelum di enkripsi)
- Selesai.

Dalam pembangkitan kunci RSA diperlukan dua buah bilangan prima berbeda p dan q



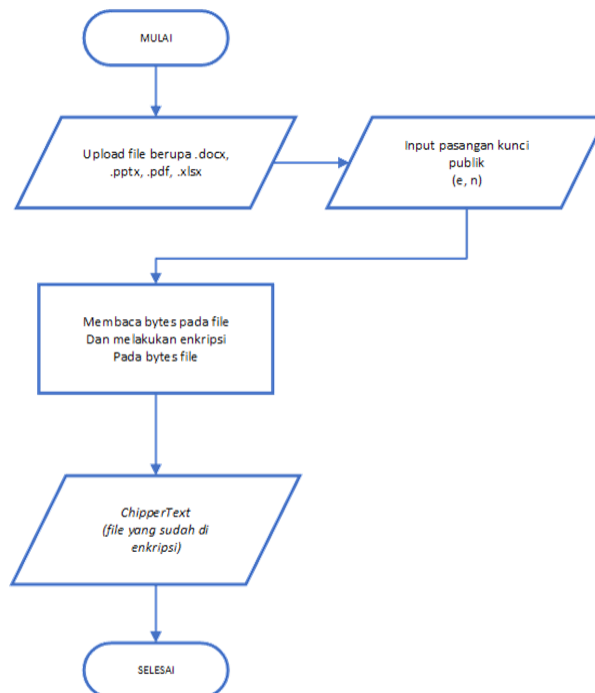
Gambar 2. Diagram Alir Pembangkitan Kunci RSA

Dalam proses enkripsi data plaintext (m) akan di enkripsi menggunakan pasangan kunci publik (e, n).



Gambar 3. Diagram Alir Enkripsi RSA

Dalam proses dekripsi data chippertext akan di dekripsi menggunakan pasangan kunci privat (d, n)



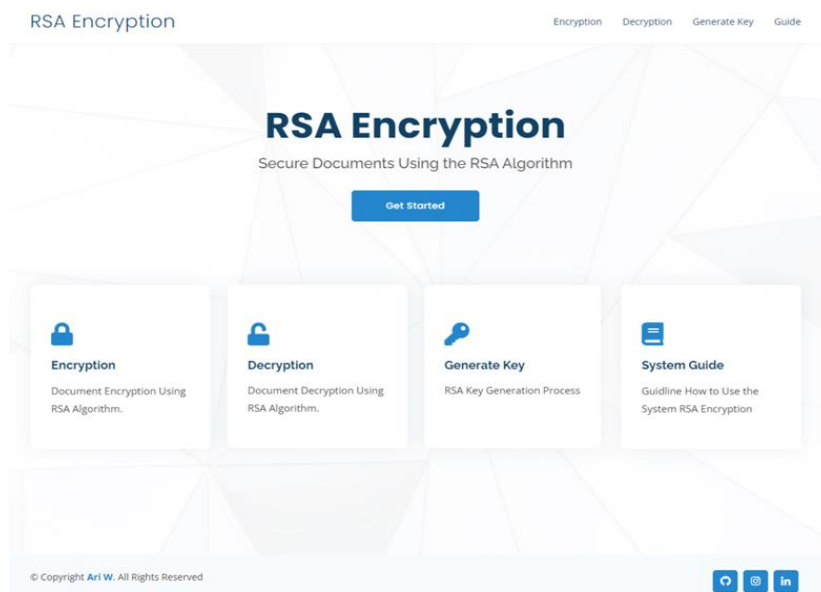
Gambar 3. Diagram Alir Dekripsi RSA

3. Hasil dan Pembahasan

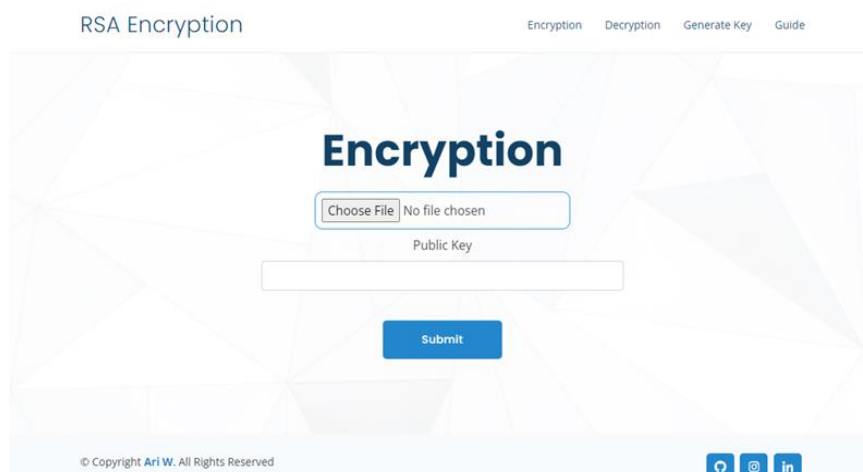
3.1. Tampilan Awal Aplikasi

Pada halaman awal menampilkan 4 buah fitur utama yaitu untuk enkripsi, dekripsi, pembangkitan kunci dan juga panduan penggunaan sistem seperti pada gambar 4.

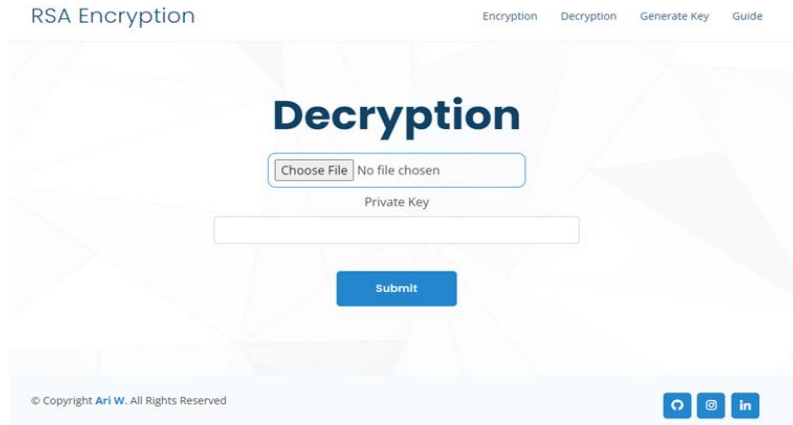
- Pada halaman enkripsi, di halaman ini pengguna mengunggah sebuah file dokumen berupa .docx, .xlsx, .pptx, atau .pdf. Serta pada halaman ini pengguna memasukkan kunci publik, seperti pada gambar 5.
- Pada halaman dekripsi pengguna akan mengunggah file dokumen yang telah di enkripsi dan memasukkan nilai dari kunci privat nya, seperti pada gambar 6.
- Pada halaman pembangkitan kunci RSA, pengguna akan diminta untuk memilih dua buah bilangan prima yang nilainya lebih besar dari 100 (minimal 3 digit angka), dan juga mengisi alamat email, dikarenakan nantinya hasil dari pembangkitan kunci akan dikirimkan melalui email seperti pada gambar 7.
- Pada halaman panduan penggunaan, pengguna dapat melihat panduan dari penggunaan sistem. Penjelasan bagaimana cara melakukan enkripsi seperti gambar 8



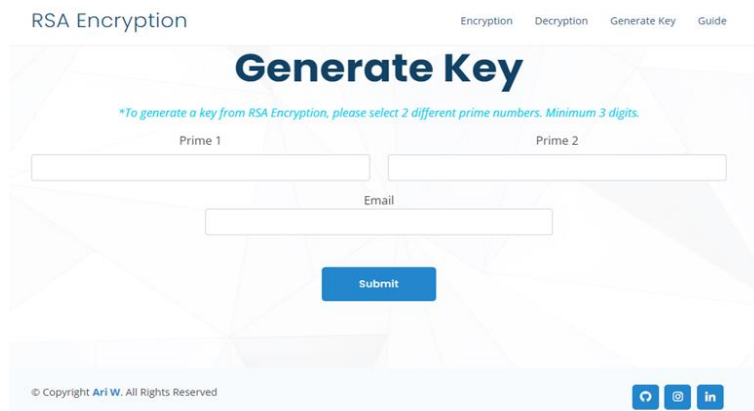
Gambar 4. Tampilan Home



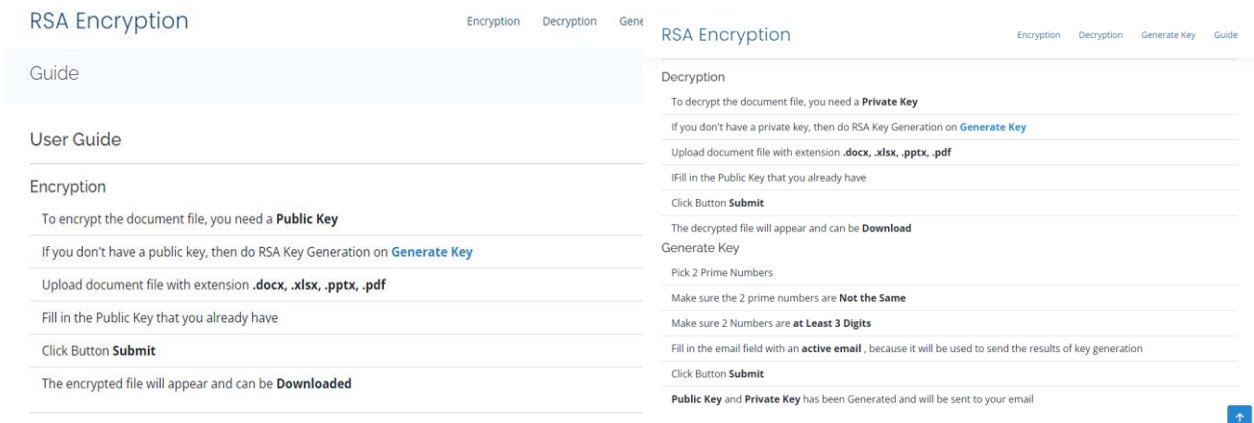
Gambar 5. Tampilan Halaman Enkripsi



Gambar 6. Tampilan Halaman Dekripsi



Gambar 7. Tampilan Halaman Pembangkitan Kunci



Gambar 8. Tampilan Halaman Panduan Penggunaan Sistem

3.2. Pengujian Sistem

Metode pengujian sistem menggunakan metode blackbox dan juga metode *brute-force* untuk mengetahui tingkat keamanan sistem. Hasil dari pengujian black box:

Tabel 1. Pengujian Enkripsi RSA

No.	Skenario Pengujian	Uji Kasus	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1	File dokumen dan kunci publik tidak di isikan , klik tombol Save	<i>File</i> dokumen: (tidak ada) Kunci publik : (tidak ada)	Sistem akan menampilkan pesan error, bahwa <i>field</i> tersebut harus di isi	Sesuai Harapan	Valid
2	File dokumen yang di <i>upload</i> valid dan kunci publik tidak di isi kemudian klik tombol Save	<i>File</i> dokumen: (valid) Kunci publik : (kosong)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
3	File dokumen yang di <i>upload</i> kosong dan kunci publik valid kemudian klik tombol Save	<i>File</i> dokumen: (kosong) Kunci publik : (valid)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
4	File dokumen yang di <i>upload</i> tidak valid (<i>upload file</i> bukan berekstensi <i>.docx, .pdf, .xlsx, atau .pptx</i>)	<i>File</i> dokumen : (<i>file</i> berekstensi <i>.txt</i>) Kunci publik : (valid)	Sistem akan menampilkan pesan error " <i>Please enter a value with a valid extension.</i> "	Sesuai Harapan	Valid
5	<i>File</i> dokumen valid dan kunci publik tidak sesuai format	<i>File</i> dokumen ; (valid) Kunci publik : (abcd)	Sistem akan menampilkan pesan error " Kunci Publik tidak sesuai format."	Sesuai Harapan	Valid
6	<i>File</i> dokumen valid dan kunci publik valid	<i>File</i> dokumen : (valid) Kunci publik : (valid)	Sistem akan menampilkan <i>pop up</i> dari <i>file</i> yang telah di enkripsi dengan nama acak	Sesuai Harapan	Valid

7	Ketika <i>pop up</i> hasil enkripsi tampil dan tombol <i>download</i> di klik	Tombol <i>download</i> di tekan	<i>File</i> hasil yang telah di enkripsi akan langsung terunduh di <i>device user</i>	Sesuai Harapan	Valid
8	Ketika <i>pop up</i> hasil enkripsi tampil dan tombol <i>upload drive</i> di tekan	Tombol <i>upload drive</i> di tekan	Sistem akan menampilkan <i>pop up</i> untuk meminta pengguna <i>login</i> ke akun <i>google</i> . Ketika berhasil <i>login</i> , <i>file</i> akan ter- <i>upload</i> ke <i>drive</i> pengguna	Sesuai Harapan	Valid

Tabel 2. Pengujian Dekripsi RSA

No.	Skenario Pengujian	Test Case	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1	File dokumen dan kunci privat tidak di berikan kemudian klik tombol Save	<i>File</i> dokumen: (tidak ada) Kunci privat : (tidak ada)	Sistem akan menampilkan pesan error, bahwa <i>field</i> tersebut harus di isi	Sesuai Harapan	Valid
2	File dokumen yang di <i>upload</i> valid dan kunci privat tidak di isi kemudian klik tombol Save	<i>File</i> dokumen: (valid) Kunci privat : (kosong)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
3	File dokumen yang di <i>upload</i> kosong dan kunci privat valid kemudian klik tombol Save	<i>File</i> dokumen: (kosong) Kunci privat : (valid)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
4	File dokumen yang di <i>upload</i> tidak valid (<i>upload file</i> bukan berekstensi <i>.docx</i> , <i>.pdf</i> , <i>.xlsx</i> , atau <i>.pptx</i>)	<i>File</i> dokumen : (<i>file</i> berekstensi <i>.txt</i>) Kunci privat : (valid)	Sistem akan menampilkan pesan error " <i>Please enter a value with a valid extension.</i> "	Sesuai Harapan	Valid
5	<i>File</i> dokumen valid dan kunci privat tidak sesuai format	<i>File</i> dokumen ; (valid) Kunci privat : (abcd)	Sistem akan menampilkan pesan error "Kunci Privat tidak sesuai format."	Sesuai Harapan	Valid
6	<i>File</i> dokumen valid dan kunci privat valid	<i>File</i> dokumen : (valid) Kunci privat : (valid)	Sistem akan menampilkan <i>pop up</i> dari <i>file</i> yang telah di dekripsi dengan nama dekripsiFile	Sesuai Harapan	Valid

7	Ketika <i>pop up</i> hasil dekripsi tampil dan tombol <i>download</i> di klik	Tombol <i>download</i> di tekan	<i>File</i> hasil yang telah di dekripsi akan langsung terunduh di <i>device user</i>	Sesuai Harapan	Valid
8	Ketika <i>pop up</i> hasil dekripsi tampil dan tombol <i>upload drive</i> di tekan	Tombol <i>upload drive</i> di tekan	Sistem akan menampilkan <i>pop up</i> untuk meminta pengguna <i>login</i> ke akun <i>google</i> . Ketika berhasil <i>login</i> , <i>file</i> akan ter- <i>upload</i> ke <i>drive</i> pengguna	Sesuai Harapan	Valid

Tabel 3. Pengujian Pembangkitan Kunci

No.	Skenario Pengujian	Test Case	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1	Bilangan 1 dan Bilangan 2 dikosongkan kemudian tombol Save di tekan	Bilangan 1: (tidak ada) Bilangan 2: (tidak ada)	Sistem akan menampilkan pesan error, bahwa <i>field</i> tersebut harus di isi	Sesuai Harapan	Valid
2	Bilangan 1 valid, dan Bilangan 2 tidak di isi	Bilangan 1: (103) Bilangan 2: (kosong)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
3	Bilangan 1 tidak di isi dan Bilangan 2 valid	Bilangan 1: (kosong) Bilangan 2: (743)	Sistem akan menunjukkan pesan error " <i>This field is required.</i> "	Sesuai Harapan	Valid
4	Bilangan 1 atau Bilangan 2 di isi kurang dari 3 digit	Bilangan 1: (11) Bilangan 2: (13)	Sistem akan menampilkan pesan error " <i>Please enter at least 3 characters.</i> "	Sesuai Harapan	Valid
5	Bilangan 1 atau Bilangan 2 di isi dengan huruf	Bilangan 1: (abc) Bilangan 2: (xyz)	Sistem akan menampilkan pesan error " <i>Please enter a valid number.</i> "	Sesuai Harapan	Valid
6	Bilangan 1 atau Bilangan 2 di isi dengan bilangan bukan prima	Bilangan 1: (120) Bilangan 2: (200)	Sistem akan menampilkan pesan error "Bilangan harus prima."	Sesuai Harapan	Valid
7	Bilangan 1 dan Bilangan 2 di isi dengan bilangan prima (valid)	Bilangan 1: (139) Bilangan 2: (743)	Sistem akan memunculkan <i>field</i> Kunci publik dan kunci privat yang bisa di <i>copy paste</i> .	Sesuai Harapan	Valid

			Sistem juga akan menampilkan <i>file .txt</i> yang bisa di <i>download</i> oleh pengguna		
8	Sistem menampilkan hasil <i>generate key RSA</i>	<i>File .txt</i> di tekan	<i>File .txt</i> yang di dalamnya berisi nilai kunci publik dan kunci privat akan ter- <i>download</i> ke <i>device user</i>	Sesuai Harapan	Valid

Pengujian BruteForce

Untuk menguji tingkat keamanan RSA metode yang umum digunakan adalah metode brute-force. Brute-force merupakan melakukan pengecekan list kunci yang ada pada kunci sebenarnya hingga menemukan kunci yang sama [8]. Dari uji coba brute-force didapatkan:

Tabel 4. Brute-force RSA

Prima 1	Prima 2	Kunci Privat	Waktu
101	431	36997 43531	1120.90 detik
101	881	1741 88981	9400.82 detik
431	773	273197 333163	183600.87 detik

Semakin besar nilai dari hasil pembangkitan kunci RSA maka semakin kuat juga nilai dari kunci tersebut, dikarenakan waktu yang dibutuhkan untuk menemukan kunci tersebut juga sangat lama

4. Kesimpulan

Sistem pengamanan dokumen menggunakan algoritma RSA sudah berjalan sesuai fungsionalitas. Dalam pengimplementasiannya sistem yang dibuat akan mengenkripsi nilai dari hexabytes dari file dokumen sehingga ketika dokumen tersebut sudah berhasil di enkripsi dokumen tersebut tidak dapat dibuka. Lama waktu yang dibutuhkan untuk proses enkripsi dan dekripsi ditentukan oleh besaran *file* dokumen dan juga besaran kunci publik dan kunci privatnya, semakin besar maka semakin lama juga waktu yang dibutuhkan. Hasil dari penelitian ini diharapkan mampu membantu orang-orang dalam melindungi *file* dokumen agar aman dipergunakan.

References

- [1] J. K. Azhar and S. Yuliany, "Implementasi Algoritma RSA (Rivest, Shamir danAdleman) untuk Enkripsi dan Dekripsi File .pdf," no. December, 2019.
- [2] C. Stephanie, "7 Kasus Kebocoran Data yang Terjadi Sepanjang 2020," *Kompas.com*, 2021. .
- [3] A. Rizal, "Kebocoran Data BPJS Kesehatan Buktikan Lemahnya Perlindungan Data Pribadi," *Infokomputer.grid.id*, 2021. .
- [4] R. Firmansyah, "Implementasi Keamanan Pesan Teks Menggunakan Kriptografi Algoritma Rsa Dengan Metode Waterfall Berbasis Java," *Joutica*, vol. 4, no. 1, p. 174, 2019, doi: 10.30736/jti.v4i1.265.

- [5] W. Djafar, "Perlindungan Data Pribadi di Indonesia: Lanskap, Urgensi, dan Kebutuhan Pembaruan," *J. Becoss*, vol. 1, no. 1, pp. 147–154, 2019.
- [6] R. Kurniawan, "Rancang Bangun Aplikasi Pengaman Isi File Dokumen Dengan RSA," *J. Ilmu Komput. dan Inform.*, vol. 01, no. November, pp. 46–52, 2017.
- [7] I. R. Munir, "Algoritma RSA dan ElGamal," *Kriptografi*, p. 13, 2010.
- [8] K. Berlin and S. . Henakaran, "An Overview of Cryptanalysis of RSA Public key System," *Int. J. Eng. Technol.*, vol. 9, no. 5, pp. 3575–3579, 2017, doi: 10.21817/ijet/2017/v9i5/170905312.

This page is intentionally left blank

Video Steganography Encryption on Cloud Storage for Securing Digital Image

Chrisna Joshua Sergio Prasetyo^{a1}, I Putu Gede Hendra Suputra^{a2}, Luh Arida Ayu Rahning Putri^{a3}, I Made Widiartha^{a4}, I Ketut Gede Suhartana^{a5}, Anak Agung Istri Ngurah Eka Karyawati^{a6}

^aInformatics Department, Faculty of Mathematics and Natural Sciences, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹sergio.febraro@gmail.com

²hendra.suputra@unud.ac.id

³rahningputri@unud.ac.id

⁴madewidiartha@unud.ac.id

⁵ikg.suhartana@unud.ac.id

⁶eka.karyawati@unud.ac.id

Abstract

Cloud storage is a data storage service in cloud computing that allows stored data to be shared and accessed via the internet. Cloud storage is usually used to store personal data such as files, photos, or videos with so that these data can be accessed anywhere via the internet without the need to use physical storage media. However, cases of data leaks in cloud storage still occur which causes personal data stored in cloud storage to be accessed by other people who do not have access.

The Client-Side Steganography Encryption on Cloud Storage Application was developed using the Modified Least Significant Bit (LSB) method and Advanced Encryption Standard (AES) algorithm. This desktop-based application was developed to protect personal data of digital images embedded in a video so that unauthorized parties cannot view the data. This application is developed to be a data security solution on cloud storage to prevent theft of personal data by non-existent parties.

From the test results, the developed application can receive input, process the input, and produce the desired output. The image from the extraction process from video also does not change at all in terms of visual or visible. The results obtained from this test is the PSNR value with an average of 36.395 dB. Good PSNR value is above 30 dB and indicates that the quality of the extracted image is good and also indicates that the developed application can protect digital images embedded in videos.

Keywords: *Steganography, Cryptography, Digital Image, Video, Cloud Storage*

1. Introduction

Cloud computing is one of the technologies that is growing rapidly. One of the *cloud computing* services is *cloud storage*. Cloud storage is a data storage service that commonly used to store personal files such as images, videos, documents, and any other files.

Most of the cloud storage services is already use their own security protocol to maintain their data in it. But there are several cases of data breaching in *cloud storage* which made the data stored in *cloud storage* can be accessed by unauthorized person. Gmail, Google's email service, tops the list for phising (27.8%) and keylogger (29.8%) cases [1].

This Google's email service is synchronized with their cloud storage, Google Drive. With many cases of data breaching [1], so an additional security is needed to secure data that stored in Google Drive. This data breaching risk can be prevented by using steganography and cryptography techniques.

Least Significant Bit (LSB) Modification method provides security protection by hiding an encrypted message into a container media. Least Significant Bit Modification method provide good protection if combined with Advanced Encryption Standard (AES) algorithm [2].

Advanced Encryption Standard (AES) algorithm turns the message into unreadable files and can be decrypted using the same key that used in the encryption process. This encryption-decryption processing time depends on the size of the data [3].

Based on the problem, the authors intend to develop Video Steganography Encryption on Cloud Storage application using Least Significant Bit Modification Method combined with Advanced Encryption Standard Algorithm encryption. This desktop-based application developed with the aim of protecting personal data that stored in cloud storage. This application is developed to be a data security solution on cloud storage to prevent data breaching by unauthorized parties.

2. Research Methods

2.1 Advanced Encryption Standard (AES) Algorithm

In general, cryptography can be interpreted as the science and art of encryption which aims to maintain the security and confidentiality of data. Cryptography supports the needs of two aspects of information security, secrecy (protection of the confidentiality of information data) and authenticity (protection against counterfeiting and unwanted alteration of information). The process of scrambling the message is called encryption and when tidying the scrambled message, it is called decryption. The initial message that has not been scrambled or that has been tidied up is called plaintext, while messages that have been scrambled are called ciphertext [4].

Advanced Encryption Standard (AES) is a symmetric ciphertext block that can encrypt and decrypt information. This algorithm works using cryptography key 128, 192, or 256 bits. This different type of keys will affect total number of rounds used in this algorithm [5].

There are four transformations in encryption using AES algorithm [3]. These transformations are SubBytes (change every byte state with the byte in the S-Box table), ShiftRows (shifting bit process, where the leftmost bit will be moved to the rightmost bit (bit rotation)), MixColumns (operate every element that is in a column in state), and AddRoundKey ((XOR between states with RoundKey). The four transformations will be repeated in 10, 12, or 14 rounds, depending on the type of key used (128, 192, and 256 bits). The encryption flow for this application can be seen on Figure 1.

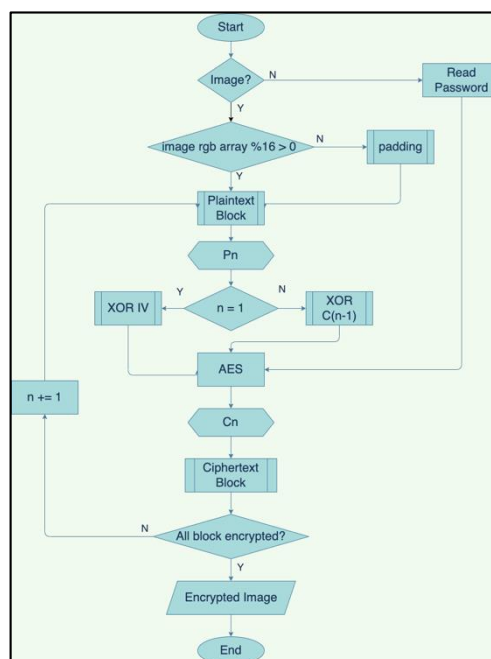


Figure 1. AES Encryption Flowchart

Decryption using AES algorithm used the inverse transformation of the AES Encryption. These transformations are InvSubBytes (change every byte state with the byte in the Inverse S-Box table), InvShiftRows (shifting bit process, where the rightmost bit will be moved to the leftmost bit (bit rotation)), InvMixColumns (each column in the state is multiplied by the multiplication matrix in AES). AddRoundKey is a self-inverse transformation on condition that uses the same key. The decryption flow for this application can be seen on Figure 2.

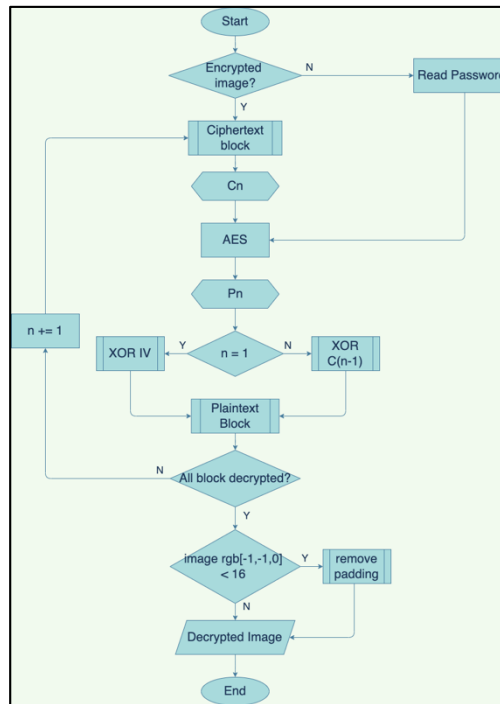


Figure 2. AES Decryption Flowchart

2.2 Least Significant Bit (LSB) Modification Method

Steganography is the art of hiding messages in digital media in such a way that other people do not realize that there is a message in the media. In the field of computer security, steganography is used to hide confidential data, when encryption cannot be done or at the same time as encryption [6].

The Least Significant Bit (LSB) is the least significant bit in a binary data. The LSB bit is located to the right of the binary number sequence. The opposite of LSB is MSB (Most Significant Bit) which is the most important part in a binary sequence that lies to the left of the binary number sequence.

The arrangement of bits in a byte that describes the suitable bit to replace is bit LSB, because this replacement only changes the byte value one higher or one lower than the previous value. Suppose the bytes in the image represent the color certain, then the change in the LSB bit does not change the color significantly and the changes are undetectable by the human eye [7].

For this application, the embedding process flow starts with user input in the form of images, videos, and passwords. Then the image resolution will be resized according to the video resolution. The video input then extracted for each frame. The first to third frames will be used for the image insertion process that has been encrypted with AES encryption. The four Most Significant Bit (MSB) images will be inserted into the first four Least Significant Bit (LSB) frames, four Least Significant Bit (LSB) images will be inserted into the second four Least Significant Bit (LSB) frames, and the third frame will be used for inserts the image resolution in string format after adjusting the video resolution.

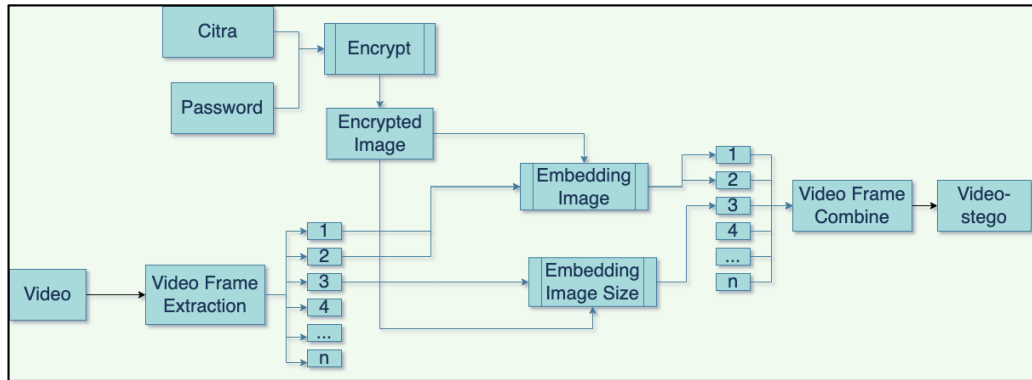


Figure 3. LSB Embedding Flowchart

The extraction process flow starts from entering the user in the form of stego-video and password. The video input is then extracted for each frame. The first to the third frames will be used for the image extraction process from stego-video. The first four Most Significant Bit (MSB) images will be taken from the first four Least Significant Bit (LSB) frames, four Least Significant Bit (LSB) images will be taken from the second four Least Significant Bit (LSB) frames. The results of the extraction of these two frames are then recorded and combined into a single unit, then the data is converted into a digital image.

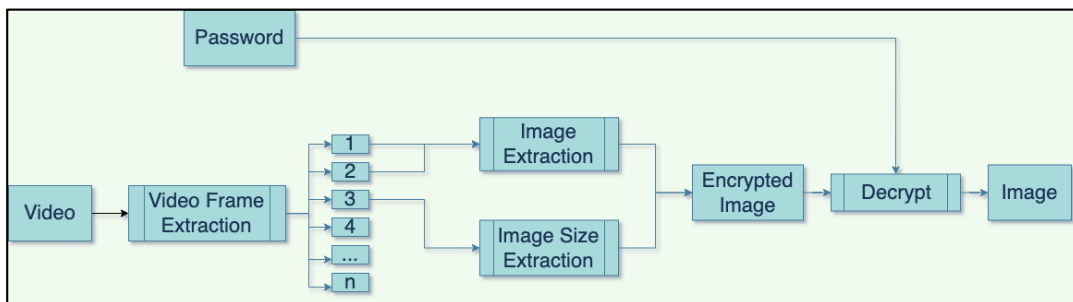


Figure 4. LSB Extraction Flowchart

2.3 Waterfall Method

Waterfall method provides sequential software life-flow approach starting from analysis, design, coding, and testing [8]. The development of this application will be using the waterfall method. Analysis stage will determine the capabilities that must be possessed by the system to meet what user needs.

1. The programming language used in the development of this application is Python.
2. This application is a desktop-based application with a minimum version of Python is 3.8 or newer.
3. This application is required to install several Python libraries used in this application before used.

In the design stage, the requirements needed in the development of this application are as follows.

- a. Front Page
 On the front page of this application, two options will be displayed, embedding and extraction, where the user can choose according to what the user wants. If the user selects embedding it will go to the embedding process page, whereas if the user selects extraction, it will go to the extraction process page.
- b. Embedding Page
 On the embedding process page, users are required to input a digital image file with .JPEG extension and a video file with .MP4 extension. Users are also required to enter a password that will be used in the digital image encryption process. After that, the user can click submit button and it will enter into the encryption and embedding process.
- c. Embedding Result Page

On the embedding results page, it will be displayed that the embedding process has been successful. After that, the user can upload the stego-video to Google Drive using the Google Drive API or save it in local storage.

d. Extraction Page

On the extraction process page, the user is required to input a stego-video file in that has a digital image in it. Users are also required to enter the same password during the insertion process which will be used in the digital image decryption process. After that, the user can click submit button and will enter into the decryption and extraction process.

e. Extraction Result Page

On the extraction results page, the extracted digital image will be displayed. Users can download the extracted digital image to be stored on their device.

3. Result and Discussion

The implementation of the development of the application will be made based on the design stage that translates the design into a program.

3.1 Coding

The implementation of this system is that the application can encrypt digital images, embed digital images into video, extract digital images from video, and decrypt encrypted images.

a. Application Interface

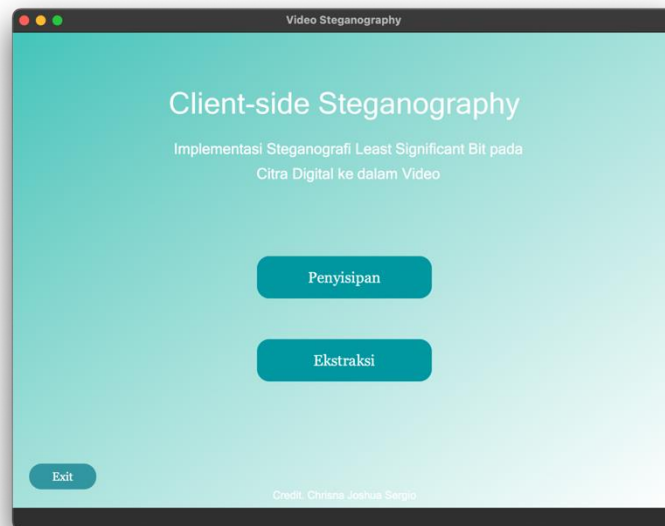


Figure 5. Main Menu Page

The main page display contains the main menus in the application. There are two menus in this page, Embedding Menu, which leads to the embedding page and Extraction Menu, which leads to the extraction page. On this page, there is also an exit button that can be used to exit the application.

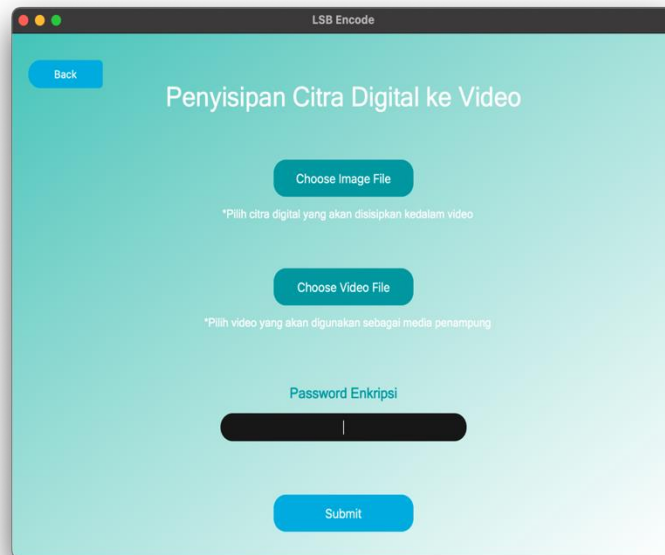


Figure 6. Embedding Page

The embedding page related to the process of embedding a digital image into a video. There are three inputs that must be entered by the user to carry out the embedding process. There are four buttons and one text field on this page. The choose image file button is used to handle image input, the choose video file button is used to handle video input, the back button is used to return to the main page, the text field is used to handle password input, and submit button is used to process all the inputs.

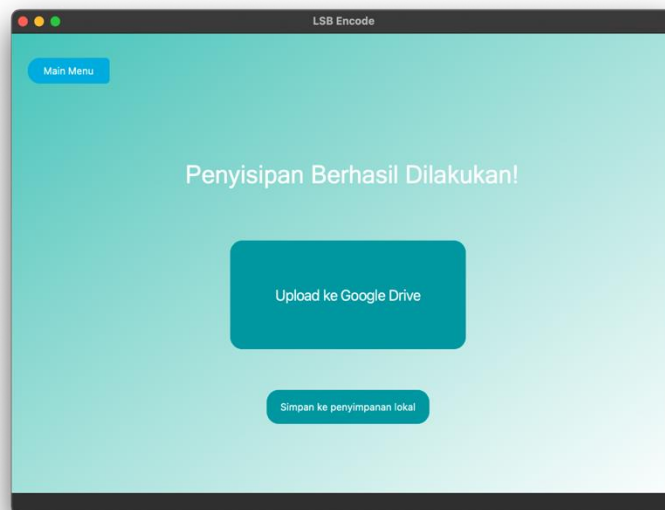


Figure 7. Embedding Result Page

The embedding result page indicates that the process of embedding a digital image into a video has been successfully carried out. On this page, there are three buttons that can be used by the user. The save to local storage button is used to save the stego-video to the user's local storage, the upload to Google Drive button is used to upload the stego-video to the user's Google Drive via Google Drive API, and the main menu button is used to return to the main page.

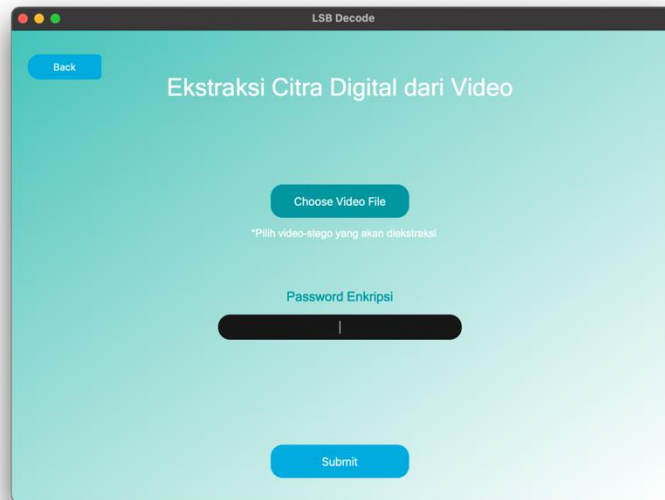


Figure 8. Extraction Page

Extraction page related to the process of extracting digital images from stego-video. There are two inputs that must be entered by the user to carry out the extraction process. There are three buttons and one text field on this page. The choose video file button functions to handle video input that has been embedded with an image, the back button is used to return to the main page, the text field is used to handle password input, and submit button is used to perform the extraction process from both inputs.




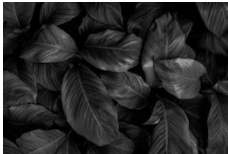


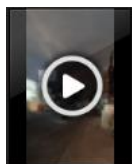
Figure 9. Extraction Result Page

The extraction result page indicates that the digital image extraction process from stego-video has been successfully carried out. On this page, users can see a preview of the successfully extracted image. This image preview display uses a fixed size and is not the original size of the extracted image. There are two buttons on this page. The save to local storage button is used to save the extracted image to the user's local storage and the main menu button is used to return to the main page.

3.2 Testing

The research data that will be used for testing this system is 3 digital image files and 2 video files. These data can be seen in Table 1.

Table 1. Research Data

No.	Filename	Resolution	File Size	Preview
1	Citra1.jpeg	528x324	49 KB	
2	Citra2.jpeg	626x418	61 KB	
3	Citra3.jpeg	4608×2963	2,2 MB	
4	Video1.mp4	640×360	784 KB	
5	Video2.mp4	720x1280	4,2 MB	

Testing the quality of the extraction results is carried out to find out how the quality of the image after it is extracted from stego-video. The test is carried out using the Peak Signal Noise Ratio (PSNR) method by comparing the original image and the extracted image. Peak Signal to Noise Ratio (PSNR) is the ratio between the maximum value of the measured signal and the amount of noise that affects the signal [9]. PSNR is measured in decibels. The results of the extraction quality test can be seen in Table 2.

Table 2. Extraction Quality Test Results

No.	Video File (.mp4)	Image File (.jpeg)	MSE	PSNR (dB)
1	Video1	Citra1	5.774328466691948	40.516
		Citra2	20.01752914951989	35.117
		Citra3	34.879654431216935	32.705
2	Video2	Citra1	5.367029898379105	40.833
		Citra2	15.276360013860014	36.291
		Citra3	33.308293336533076	32.905
Average				36.395

A good image is an image that has a small MSE value. The smaller the MSE value, the more similar an image is to the original image where the value of each pixel location is the same. The smaller the MSE value, the higher the PSNR value. In other words, the image is said to be good if the PSNR value is above 30 dB [10].

Table 4.3 shows the results of the extraction quality using the PSNR method. The table displays information in the form of MSE and PSNR values from the comparison of the original image and the extracted image. The highest PSNR value from this test is 40.516 dB in Video1 and Citra1 data, while the lowest PSNR value from this test is 32.705 dB in Video1 and Citra3 data. From this test, the average PSNR value is 36.395 dB and shows that the quality of the extracted image is said to be good. From this test, it can be concluded that the reduction in the extracted image is not significant because the original image and the extracted image still look the same visually. The quality of the extraction results is influenced by the resolution of the image along with the resolution of the video used as a media container. The MSE value is quite large because the image frame size is adjusted to the video frame size.

4. Conclusion

The quality of extraction with steganography technique using Modified Least Significant Bit and Advanced Encryption Standard methods can be said to be good. This is evidenced by the PSNR value in testing the quality of the extraction results. The average PSNR value obtained is 36.395 dB. Good PSNR value is above 30 dB and indicates that the quality of the insertion and extraction results is good. Based on the previous statement, the application has succeeded reaching the goal by securing digital image in video and extracting it from the stego-video without any data lost.

This application can be developed in terms of maintaining the quality of the extracted image against treatments such as compression, to reduce the stego-video file size and the extracted image file size. It can also be developed to run on various operating systems, so users with a base other than desktop can use this application.

References

- [1] K. Thomas et al., "Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials," in Proceedings of the ACM Conference on Computer and Communications Security, Oct. 2017, pp. 1421–1434. doi: 10.1145/3133956.3134067.
- [2] E. Nirmala, "Penerapan Steganografi File Gambar Menggunakan Metode Least Significant Bit (LSB) dan Algoritma Kriptografi Advanced Encryption Standard (AES) Berbasis Android," Jurnal Informatika Universitas Pamulang, vol. 5, no. 1, 2020, [Online]. Available: <http://openjournal.unpam.ac.id/index.php/informatika36>

- [3] F. Muharram, H. Azis, and A. R. Manga, "Analisis Algoritma pada Proses Enkripsi dan Dekripsi File Menggunakan Advanced Encryption Standard," *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 2, 2018.
- [4] H. Mukhtar, *Kriptografi untuk Keamanan Data*. Yogyakarta: Deepublish, 2018.
- [5] D. Darwis, R. Prabowo, and N. Hotimah, "Kombinasi Gifshuffle, Enkripsi AES dan Kompresi Data Huffman untuk Meningkatkan Keamanan Data," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, p. 389, Oct. 2018, doi: 10.25126/jtiik.201854727.
- [6] R. D. Kusuma, M. Syafaat, and A. Setiawan, "Aplikasi Steganografi Video dengan Metode Least Significant Bit (LSB) untuk Alat Bantu pada Foto Film Militer," *Jurnal Elkasista*, vol. 1, 2020.
- [7] U. A. Anti, A. H. Kridalaksana, and D. M. Khairina, "Steganografi pada Video Menggunakan Metode Least Significant Bit (LSB) dan End of File (EOF)," *Jurnal Informatika Mulawarman*, vol. 12, no. 2, 2017.
- [8] G. W. Sasmito, "Penerapan Metode Waterfall Pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal," vol. 2, no. 1, 2017, [Online]. Available: <http://www.tegalkab.go.id>,
- [9] G. W. Bhaudhayana and I. M. Widiartha, "Implementasi Algoritma Kriptografi AES 256 dan Metode Steganografi LSB pada Gambar Bitmap," *Jurnal Ilmiah Ilmu Komputer Universitas Udayana*, vol. 8, no. 2, 2015.
- [10] G. Badshah, S. C. Liew, J. M. Zain, and M. Ali, "Watermark Compression in Medical Image Watermarking Using Lempel-Ziv-Welch (LZW) Lossless Compression Technique," *Journal of Digital Imaging*, vol. 29, no. 2, pp. 216–225, Apr. 2016, doi: 10.1007/s10278-015-9822-4.

Implementasi Port Knocking Dalam Mengamankan Jaringan VPN Pada Sistem E-Learning

I Ketut Prawira Adhisastra¹, I Komang Ari Mogi², Cokorda Rai Adi Pramatha³, I Gede Santi Astawa⁴,
Made Agung Raharja⁵, Ngurah Agus Sanjaya ER⁶

³Informatics departement, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹prawira.adhisastra@yahoo.com

²arimogi@unud.ac.id

³cokorda@unud.ac.id

⁴santi.astawa@unud.ac.id

⁵made.agung@unud.ac.id

⁶agus_sanjaya@unud.ac.id

Abstract

universities that have started using e-learning as a learning medium. However, the system on e-learning is vulnerable to attacks such as identity theft, and authentication on e-learning systems. One way to secure data on a network is to implement a Virtual Private Network (VPN) on the network which can make a network private. That way, the security on the port on the E-learning server will be guaranteed. But attacks on servers more often occur on port 23 (telnet) because telnet does not encrypt the connection process. One of the attacks that is often used is a brute force attack. Based on these problems, research has been carried out to secure port 23 (telnet) and port 80 (e-learning website) on the e-learning server from brute force attacks by implementing the port knocking method. Port 23 and port 80 on the e-learning server that uses the port knocking method are always closed so that the client must do port knocking before being able to use port 23 and port 80 on the e-learning server. By implementing port knocking on the e-learning server, brute force attacks can be avoided because the attacked port is closed.

Keywords: VPN, Port Knocking, E-Learning, Security, Digital

1. Pendahuluan

E-Learning adalah sistem pendidikan yang menggunakan aplikasi elektronik untuk mendukung pengembangan kegiatan belajar mengajar dengan media internet[1]. E-learning memungkinkan perolehan pengetahuan dan informasi melalui perangkat yang terhubung dengan internet seperti handphone, komputer, tablet, dan lain-lainnya. Namun, karena sistem e-learning merupakan jaringan yang terbuka pasti rentan terhadap berbagai serangan keamanan. Sistem e-learning rentan terhadap serangan seperti pencurian identitas, dan otentikasi pada sistem e-learning[2]. Peretas merupakan ancaman besar bagi kehidupan di jaman sekarang, dengan segala jenis kecanggihan teknologi yang ada, peretas dapat menggunakan metode tertentu untuk menyerang suatu jaringan internet. Salah satu cara untuk dapat mengamankan suatu data pada jaringan tertentu adalah dengan menggunakan Virtual Private Network (VPN). Virtual Private Network (VPN) adalah sebuah teknologi komunikasi yang memungkinkan untuk dapat terkoneksi ke jaringan publik dan menggunakannya untuk bergabung dengan jaringan local[3].

Dengan menggunakan VPN pada sistem e-learning maka keamanan untuk sistem e-learning akan terjamin. Virtual Private Network (VPN) merupakan sebuah teknologi komunikasi yang dapat memungkinkan untuk terkoneksi ke jaringan publik dan dapat menggunakannya untuk bergabung dengan jaringan lokal. Dengan menggunakan cara tersebut maka akan didapatkan hak dan pengaturan yang sama seperti menggunakan koneksi atau jaringan LAN, walaupun sebenarnya menggunakan jaringan publik. Dengan begitu, keamanan pada bagian port-nya pada server E-learning akan terjamin. Karena port adalah sebuah mekanisme komunikasi perangkat satu dengan

perangkat lainnya. Oleh karena itu port 23 (telnet) juga sangat perlu diamankan, telnet berfungsi untuk seorang network administrator untuk bisa melakukan remote untuk login ke dalam suatu komputer. Jika port ini terus terbuka maka peretas akan menemukan celah untuk mencuri informasi dan data yang ada. Maka dari itu port yang terbuka tersebut perlu di tutup jika tidak ada komunikasi agar peretas tidak dapat masuk ke dalam port tersebut.

Maka dari itu penelitian ini menambahkan sebuah metode untuk mengamankan suatu jaringan VPN dengan menggunakan metode Port knocking. Metode port knocking merupakan sebuah teknik pertahanan yang digunakan untuk mencegah penyerang melakukan scanning untuk mendapatkan suatu informasi tentang kelemahan servis yang berpotensi dieksploitasi. Port knocking merupakan sebuah metode yang memberikan kekuasaan kepada user berdasarkan firewall untuk melakukan suatu komunikasi melalui port yang tertutup[4]. Port knocking bekerja seperti halnya brankas dengan kunci kombinasi angka putar. pada brankas tersebut, Anda diharuskan memutar kunci kombinasi beberapa kali hingga tepat seperti yang ditentukan[5]. Menggunakan metode port knocking dalam mengamankan suatu jaringan yang memiliki tujuan untuk memberikan tambahan lapisan keamanan yang ringan pada jaringan komputer yang berjalan melalui port yang tertutup.

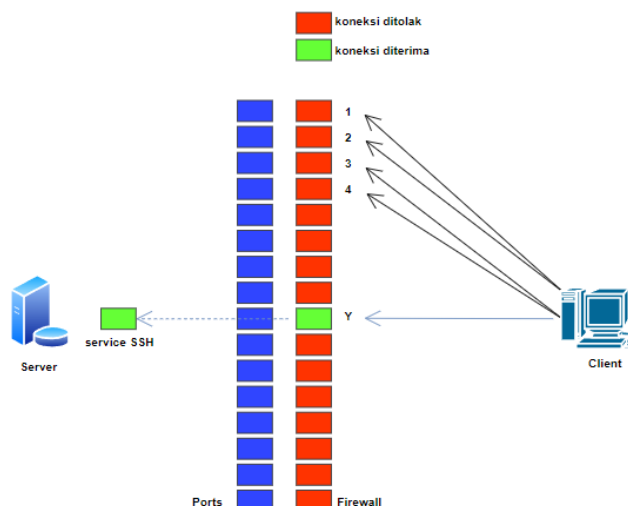
Pada penelitian ini, penulis ingin meningkatkan keamanan jaringan virtual private network pada sistem E-learning dengan mengimplementasikan metode port knocking. Metode port knocking akan digunakan untuk menutup dan membuka akses masuk kedalam port tertentu dan metode ini terintegrasi bersama firewall. Untuk mendapatkan akses penuh ke dalam port, hanya *user* yang sudah memiliki akses kedalam port knocking yang dapat mengakses langsung, seperti admin atau client.

2. Metode Penelitian

Metode port knocking akan menggunakan algoritma PCG-XSH-RR untuk melakukan pengacakan pada jumlah dan nomor port yang berfungsi untuk membangkitkan sebuah deretan nomor *port* yang harus diketuk oleh client untuk melakukan koneksi terhadap sistem e-learning. Jaringan VPN pada sistem E-Learning akan diamankan dengan menggunakan metode port knocking untuk mengatasi serangan brute force.

2.1. Port Knocking

Port knocking merupakan metode yang digunakan untuk dapat mengamankan jaringan internet agar tidak mudah diretas atau diserang. Cara kerja *port knocking* ini adalah dengan membuka dan menutup akses ke port tertentu dan metode ini terintegrasi Bersama dengan firewall. Untuk dapat memperoleh akses penuh ke dalam port, hanya *user* yang sudah memiliki akses yang dapat mengakses langsung, seperti admin atau client.



Gambar 2.1 Proses Port Knocking

Metode *Port knocking* dapat diilustrasikan seperti pada gambar 2.1 di atas. Apabila *client* ingin mengakses service SSH pada server melalui port Y, maka *client* terlebih dahulu harus mengirimkan paket SYN menuju port 1,2,3,4 agar dapat mengakses port Y [7]. Namun metode ini masih memiliki beberapa kelemahan dalam mengatasi *port*. Penelitian ini menambahkan algoritma PCG-XSH-RR untuk melakukan pengacakan bilangan pada port.

2.2. PCG-XSH-RR

Permuted Congruential Generator-Xorshift-Random Rotation (PCG-XSH-RR) merupakan salah satu algoritma pengacakan bilangan yang dikembangkan oleh Melissa E. O'Neill pada tahun 2014 [8]. PCG memiliki keunggulan dari algoritma random lainnya, yaitu lebih cepat dalam waktu komputasi, banyak memiliki variasi, menggunakan memori yang kecil, seragam, sederhana, mudah dipahami dan sulit untuk di prediksi. Algoritma PCG-XSH-RR adalah varian yang paling disarankan dalam penggunaan algoritma PCG karena mempunyai state 64-bit dan output 32-bit selain itu algoritma ini juga memiliki tujuan untuk dapat membuat pembangkit bilangan acak yang baik untuk dapat digunakan dalam berbagai hal [7].

Permuted Congruential Generator-Xorshift-Random Rotation (PCG-XSH-RR) yang berfungsi untuk membangkitkan bilangan acak sebagai jumlah port dan nomor port yang harus diketuk untuk melakukan pengetukan pada metode port knocking. Pembangkitan bilangan acak ini dilakukan pada komputer server yang kemudian akan dikirimkan ke client untuk melakukan pengetukan. Bilangan acak ini dilakukan dengan menghasilkan 2 objek yang berupa jumlah port dan nomor port. Pengacakan terhadap jumlah port memiliki range 1 sampai 100, kemudian untuk nomor port memiliki range 1 sampai 65000. Berdasarkan penjelasan di atas objek dari algoritma ini adalah jumlah port dan nomor port yang akan digunakan sebagai rule untuk melakukan port knocking dari komputer client ke server.

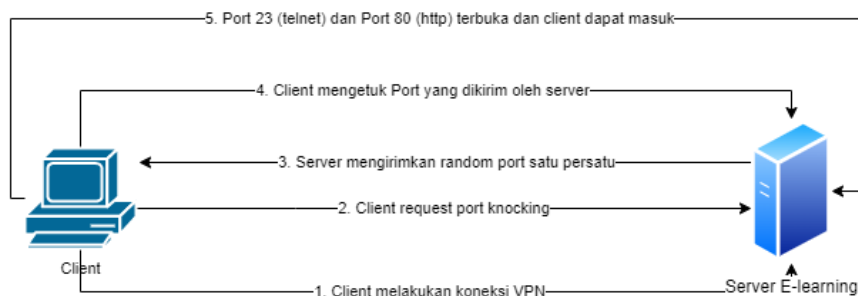
Table 1. Ilustrasi port nomor yang digunakan

190	394	874	8740	543
-----	-----	-----	------	-----

Table 1. adalah ilustrasi dari pembangkitan bilangan acak untuk port yang menggunakan algoritma PCG-XSH-RR, pembangkitan bilangan acak ini dilakukan sebanyak dua kali untuk menentukan jumlah port dan port berapa yang digunakan sebagai rule untuk melakukan pengetukan pada metode port knocking. Berdasarkan tabel 1 didapatkan port sejumlah 5 dengan hasil port yang random. Port tersebut nantinya akan digunakan untuk melakukan pengetukan sebelum port tujuan (23 & 80) dapat terbuka. Sebelumnya port 23 & 80 dalam keadaan tertutup sebelum dilakukan pengetukan dan akan terbuka setelah dilakukan pengetukan yang sesuai dengan rule port dari hasil pembangkitan bilangan acak.

2.3. Bagan Rancangan Sistem

Berikut adalah bagan dari rancangan sistem pada penelitian ini



Gambar 2.2 Bagan dari Rancangan Sistem

Gambar 2.2 merupakan bagan dari rancangan sistem pada penelitian ini yang terdapat dua komputer yaitu, komputer server e-learning dan komputer client. Kedua komputer tersebut memiliki fungsi dan tugas yang berbeda, berikut adalah fungsi dan tugas dari kedua komputer :

1. Server

Komputer server e-learning memiliki fungsi untuk melakukan pembangkitan bilangan acak dengan algoritma PCG-XSH-RR dan untuk menerima request dari client. Request yang dikirim oleh client berupa request untuk membuka port 23 (telnet) dan port 80 yang akan menampilkan halaman website e-learning jika client menginputkan IP dari server e-learning pada browser. Kemudian server akan mengirimkan nomor port dari hasil pembangkitan bilangan acak yang telah dilakukan secara bertahap untuk di ketuk oleh client.

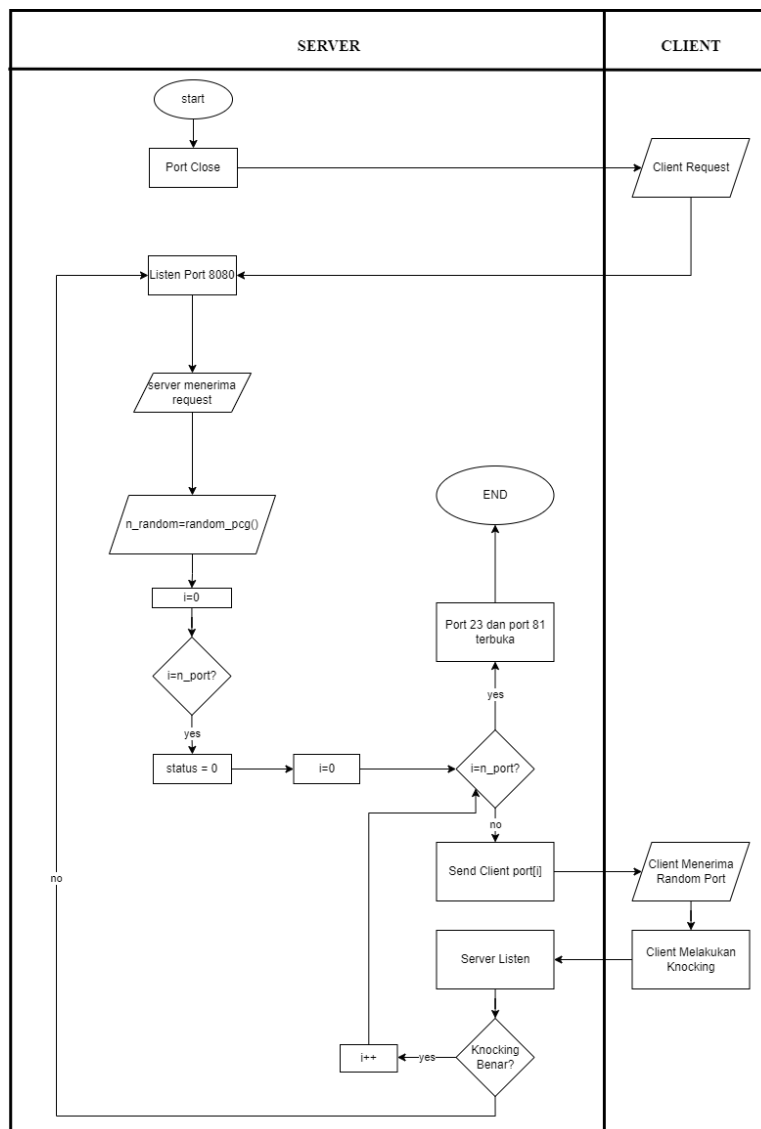
2. Client

Komputer client memiliki fungsi melakukan request ke komputer server untuk dapat membuka port 23 dan port 80. Sebelum dapat membuka port 23 dan 80 client harus melakukan request untuk mendapatkan nomor port dari server, kemudian client akan melakukan pengetukan apabila sudah mendapatkan nomor port yang diacak oleh server secara bertahap. Setelah proses pengetukan selesai dilakukan dengan benar, maka port 23 dan 80 dapat terbuka dan dapat diakses oleh client.

2.7. Flowchart Proses Open Port dan Close Port

Pada proses ini akan dijelaskan bagaimana client dapat membuka port yang tertutup pada server dan client penutup port pada server apabila sudah selesai melakukan koneksi.

1. Proses open port

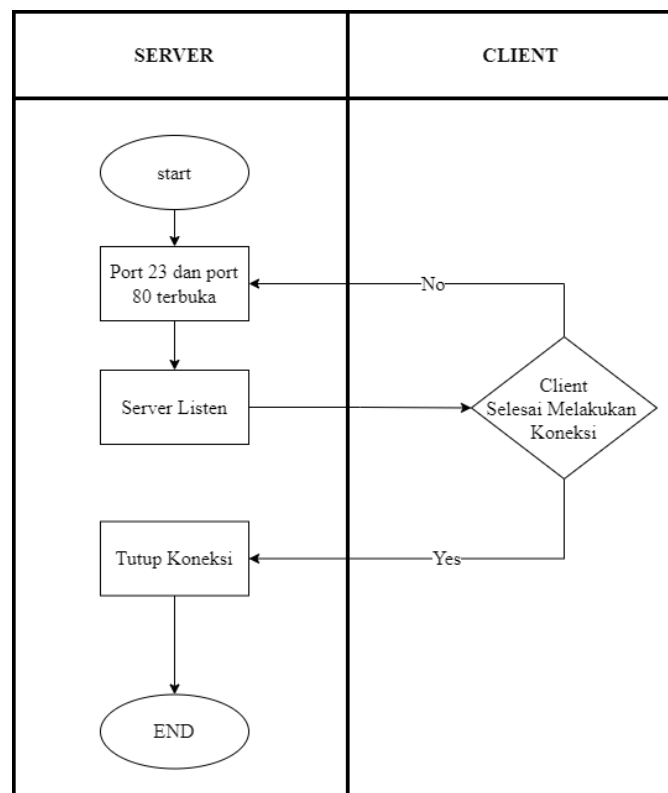


Gambar 2.3 Flowchart Proses Open Port

Berikut penjelasan untuk gambar 2.3 Flowchart Proses Open Port :

1. Pada proses pertama keadaan *port* masih tertutup untuk koneksi terhadap *client* baru.
2. *Client* melakukan *request* dan akan didengar oleh *port* 8080, dimana keadaan *port* 8080 selalu dalam keadaan *listen*. Fungsi dari *port* 8080 disini adalah untuk selalu *listening* terhadap *request* dari *client* yang berupa *request port knocking*.
3. *Server* menerima *request* dari *client* dan akan dilakukan pengujian terhadap *request* tersebut.
4. *request* dari *client* merupakan *port knocking* maka akan dilakukan pengacakan dengan algoritma PCG-XSH-RR.
5. Variabel *n_port* digunakan untuk menampung nilai terhadap *port* yang telah diacak.
6. Variabel *i* akan bernilai 0.
7. Diuji apakah *i* bernilai sama dengan *n_port* atau tidak.
8. Apabila tidak maka akan dilakukan proses terhadap variabel *port[i]* yang menampung *port* yang diacak. Kemudian nilai *i* akan selalu ditambah 1, kemudian pengujian kembali dilakukan terhadap nilai *i*.
9. Apabila iya maka variabel status akan bernilai 0 atau false.
10. Kemudian *i* bernilai 0.
11. Dilakukan pengujian kembali apakah *i* sama dengan *port* yang telah *dirandom*.
12. Apabila tidak maka *port* yang telah diacak akan dikirim satu per satu ke *client*. *Client* akan menerima *port* dan melakukan pengetukan. Dalam keadaan tersebut *server* dalam keadaan *listen* untuk menerima ketukan. Kemudian akan dilakukan pengujian apakah pengetukan sudah benar atau tidak. Apa bila iya maka nilai *i* akan bertambah 1 dan pengujian terhadap nilai *i* apakah sama dengan *port* yang diacak akan dilakukan kembali. Apabila tidak maka proses akan kembali menuju keadaan dimana *port* 8080 hanya dalam keadaan *listen* untuk menerima *request* kembali dari *client* berupa *request port knocking*.
13. Apabila iya maka *port telnet* akan terbuka dan proses selesai.

2. Proses close port



Gambar 2.4 Flowchart Proses Close port

Berikut penjelasan untuk gambar 2.4 *flowchart* proses *close port*:

1. Pada proses pertama *port 23* dan *port 80* sudah dalam keadaan terbuka dan *server* masih dalam keadaan *listen*.
2. Kemudian dilakukan pengujian apakah *client* sudah selesai melakukan koneksi atau tidak.
3. Apabila tidak maka *port 23* dan *port 80* masih dalam keadaan terbuka dan apabila iya maka koneksi akan ditutup.
4. Proses selesai

3. Hasil dan Pembahasan

Pada bagian hasil dan pembahasan ini, penulis melakukan pengujian serangan brute force terhadap jaringan pada sistem e-learning dengan menggunakan metode port knocking dan tidak menggunakan metode port knocking. Dengan tujuan seberapa pengaruh penggunaan metode port knocking untuk mengamankan jaringan vpn pada system e-learning.

pada penelitian ini menggunakan tools Ncrack 0.4ALPHA. pengujian ini akan mencoba untuk masuk ke server melalui port 23 (telnet) dengan menggunakan ncrack. Pada proses serangn brute force attack ini akan digunakan wordlist yang berisi username dan password. Pada wordlist ini berisikan list kata – kata yang berbeda pada masing-masing file.

```
root@client-VirtualBox:/home/client/Documents/wordlist# ncrack -U username2.list
-P password2.list 192.168.1.11:23

Starting Ncrack 0.7 ( http://ncrack.org ) at 2022-04-12 11:41 WITA

Discovered credentials for telnet on 192.168.1.11 23/tcp:
192.168.1.11 23/tcp telnet: 'server' '12345'

Ncrack done: 1 service scanned in 45.01 seconds.

Ncrack finished.
root@client-VirtualBox:/home/client/Documents/wordlist#
```

Gambar 3.1 Brute Force Attack pada Server

Pada gambar 3.1 telah dilakukan brute force attack menuju ke server e-learning yang tidak menggunakan metode port knocking, brute force attack menyerang melalui port 23 untuk mendapatkan username dan password. Serangan berhasil dilakukan dan ncrack berhasil di mendapatkan username dan password melalui port 23 (telnet).

```
root@client-VirtualBox:/home/client/Documents/wordlist# ncrack -U username2.list
-P password2.list 192.168.1.11:23

Starting Ncrack 0.7 ( http://ncrack.org ) at 2022-04-12 11:54 WITA

Ncrack done: 1 service scanned in 3.00 seconds.

Ncrack finished.
root@client-VirtualBox:/home/client/Documents/wordlist#
```

Gambar 3.2 Brute Force Attack Pada Server Dengan Port knocking

Pada gambar 3.2 telah dilakukan brute force attack menuju ke server e-learning yang menggunakan metode port knocking, brute force attack menyerang melalui port 23 untuk mendapatkan username dan password. Serangan tidak berhasil dilakukan dan ncrack gagal mendapatkan username dan password melalui port 23 (telnet). Hal tersebut dapat terjadi karena metode port knocking membuat port 23 yang awalnya terbuka menjadi tertutup, sehingga hanya client yang sudah memiliki akses saja yang dapat membuka port 23.

Tabel 3.1 Pengujian Brute Force Attack

JUMLAH WORDLIST	TANPA METODE PORT KNOCKING		DENGAN METODE PORT KNOCKING	
	WAKTU (DALAM DETIK)	HASIL BRUTE FORCE ATTACK	WAKTU (DALAM DETIK)	HASIL BRUTE FORCE ATTACK
1	33	SUKSES	3	TIDAK SUKSES
2	45	SUKSES	3	TIDAK SUKSES
3	51	SUKSES	3	TIDAK SUKSES
4	66	SUKSES	3	TIDAK SUKSES
5	69	SUKSES	3	TIDAK SUKSES
6	75	SUKSES	3	TIDAK SUKSES
7	81	SUKSES	3	TIDAK SUKSES
8	90	SUKSES	3	TIDAK SUKSES
9	144	SUKSES	3	TIDAK SUKSES
10	146	SUKSES	3	TIDAK SUKSES

Pada Tabel 3.1 merupakan hasil dari pengujian untuk Brute Force Attack yang dilakukan sebanyak 10 kali dengan wordlist yang di dalamnya terdapat kata-kata yang berbeda di setiap file wordlist. Dilakukan 10 kali pengujian berdasarkan pada perhitungan confidence interval yang didapatkan yaitu dengan nilai kepercayaan 95% bahwa 10 kali sudah mampu memberikan hasil yang valid.

4. Kesimpulan

Berdasarkan pada penelitian yang telah dilakukan, metode port knocking dapat mengatasi serangan pada jaringan VPN dengan membuat port 23 (telnet) dan port 80 (web e-learning) dalam keadaan filtered atau tertutup, sehingga hanya client atau admin yang telah memiliki akses yang dapat mengakses port tersebut. Port 23 (telnet) dan port 80 (web e-learning) dapat terjamin keamanannya dari serangan brute force yang menggunakan tool ncrack.

Referensi

- [1] I. G. So and F. Kurniawan, "Perancangan E-Learning Berbasis Internet pada Sekolah SMK Negeri 13 Jakarta," *Binus Bus. Rev.*, vol. 1, no. 2, p. 394, 2010, doi: 10.21512/bbr.v1i2.1085.
- [2] S. K. Chung, O. C. Yee, M. M. Singh, and R. Hassan, "SQL injections attack and session hijacking on e-learning systems," *I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc.*, no. I4ct, pp. 338–342, 2014, doi: 10.1109/I4CT.2014.6914201.
- [3] I. Afrianto and E. B. Setiawan, "Kajian virtual private network (vpn) sebagai sistem pengamanan data pada jaringan komputer (studi kasus jaringan komputer unikom)," *Maj. Ilm. UNIKOM*, vol. 12, no. 1, pp. 43–52, 2015, doi: 10.34010/miu.v12i1.34.
- [4] Edy Haryanto1), Widyawan2), dan Dani Adhipta 3) 1, "Meningkatkan keamanan port knocking dengan kombinasi special features icmp, source port, dan tunneling," pp. 187–194, 2016.
- [5] E. Sutanta, "Analisis Keamanan Sistem Aplikasi (Study Kasus Pada Aplikasi E-Learning Di IST AKPRIND Yogyakarta)," *Pros. Semin. Nas. SNAST2008, IST AKPRIND Yogyakarta, ISSN 1979-911X*, pp. 243–253, 2008, [Online]. Available: <http://snast.akprind.ac.id/>.
- [6] P. Riska, P. Sugiartawan, and I. Wiratama, "Sistem Keamanan Jaringan Komputer Dan Data Dengan Menggunakan Metode Port Knocking," *J. Sist. Inf. dan Komput. Terap. Indones.*, vol. 1, no. 2, pp. 53–64, 2018, doi: 10.33173/jsikti.12.
- [7] I. M. A. D. Putra and I. K. Ari Mogi, "The Data Communication Security Design on IoT Based Systems with the Port Knocking Method," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, p. 387, 2020, doi: 10.24843/jlk.2020.v08.i04.p04.
- [8] M. E. O'neill, "A PCG: A Family of Simple Fast Space-Efficient Statistically Good Algorithms for Random Number Generation," *ACM Trans. Math. Softw. V*, vol. V, no. 212, pp. 1–46, 2017, [Online]. Available: <http://www.pcg-random.org/pdf/toms-oneill-pcg-family.pdf>.

Implementation of Genetic Algorithm in Determining Class Schedules Based on User Needs

Agus Prayogo^{a1}, I Gede Santi Astawa^{a2}, I Gusti Agung Gede Arya Kadyanan^{a3}, Ida Bagus Gede Dwidasmara^{a4}, Dr. Ngurah Agus Sanjaya^{a5}, Ida Bagus Made Mahendra^{a6}

^aInformatics Department, Faculty of Mathematics and Natural Sciences, University of Udayana South Kuta, Badung, Bali, Indonesia

¹aguspravoqo99@gmail.com

²ibm.mahendra@unud.ac.id

³agus_sanjaya@unud.ac.id

⁴santi.astawa@unud.ac.id

⁵dwidasmara@unud.ac.id

⁶gungde@unud.ac.id

Abstract

Education is an important thing in life. In practice, education consists of teachers and students. A teacher teaches a subject to students. Students are taught by several or more teachers. There is a provision in which a class of students can only receive one subject at a time. Teachers can also only teach one subject at a time. This requires a system that is used to regulate so that these provisions can be fulfilled without ignoring other provisions. The system used to manage these problems is a subject scheduling system. This system regulates the class placement of students with the subjects they study. In manual implementation, the process is very inefficient in terms of time, human resources and thought power. The method that can be used to solve this problem is the genetic algorithm. A genetic algorithm is a heuristic method or procedure inspired by the natural selection process. Genetic algorithms are generally used to produce high-quality solutions to problems in the form of optimization and search by relying on biologically inspired operators such as mutation, crossbreeding, and selection

Keywords: Education, Scheduling, Genetic Algorithm, Optimization,

1. Introduction

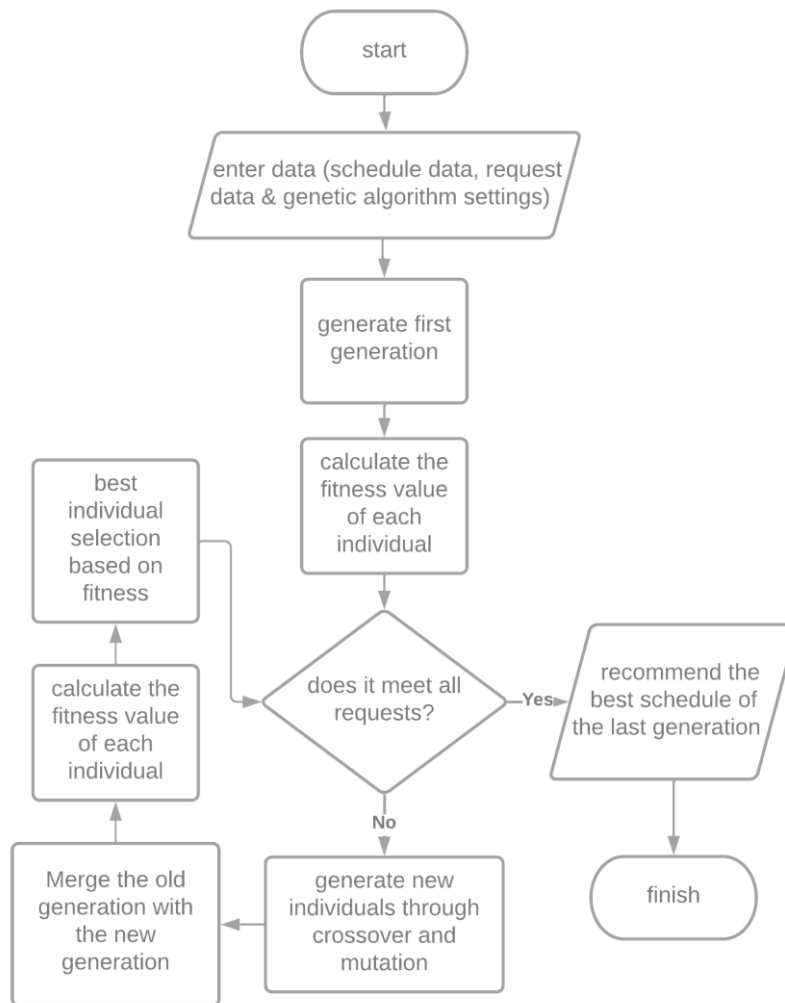
Education is an important thing in life. There are several institutions that are trusted in creating quality education, such as schools and colleges. The education is organized in a structured and complex system. The system consists of smaller systems, where the small system consists of various components that work according to their respective roles and capacities. These components complement each other, so the system will not be able to work properly if one of them is disturbed. In practice, education consists of teachers and students. A teacher teaches a subject to students. Students are taught by several or more teachers. There is a provision in which a class of students can only receive one subject at a time. Teachers can also only teach one subject at a time. This requires a system that is used to regulate so that these provisions can be fulfilled without ignoring other provisions. The system used to manage these problems is a subject scheduling system. This system regulates the class placement of students with the subjects they study. This system has problems, such as colliding schedules, teachers who cannot teach at a certain time, and so on. In manual implementation, the process is very inefficient in terms of time, human resources and thought power. If a predetermined schedule cannot meet all applicable provisions, then the schedule cannot be used in teaching activities. The schedule needs to be reworked. For a small scale, creating a schedule manually is quite easy. However, for a larger scale that is generally used in large institutions that require greater flexibility, it requires system that can solve these problems quickly and precisely. For that, a method is needed to build the system. The method that can be used to solve this problem is the genetic algorithm. A genetic

algorithm is a heuristic method or procedure inspired by the natural selection process. Genetic algorithms are generally used to produce high-quality solutions to problems in the form of optimization and search by relying on biologically inspired operators such as mutation, crossbreeding, and selection. The genetic algorithm has several variables that can affect the optimal level of the results obtained. Some of these variables are the percentage of mutations and the percentage of crossover. This research is expected to find the best percentage of mutations and percentage of crossover in order to get the most optimal results. Several research have succeeded in using genetic algorithms in making scheduling systems. One of them is [1] research with the title "Implementation of Genetic Algorithms in Web-Based Lecture Scheduling Applications by Adopting the Waterfall Model". At the conclusion of it, the system he created was successful in carrying out the scheduling process quickly without any conflicting schedules. The application can also reset the schedule if there are lecturers who cannot teach at certain times. However, the system cannot add requests given by system entities. Thus, the system only relies on changes to the final schedule. In this research, the author adds a new feature where users can add new conditions to the system where the system must meet these conditions in order to get the desired results.

2. Research Methods

2.1 Genetic Algorithm

The following is the design of the Genetic Algorithm method in Determining Lecture Schedules Based on User Needs



The following is a description of the flow chart above.

1. In the first step, the user enters the data namely;
 - a. Schedule data, namely data containing a list of teaching classes, a list of teachers, a list of courses, a list of activities, a list of classrooms, working days, and the number of teaching hours each day
 - b. Request data
 - c. Data on the number of individuals per generation.
2. Then, the system will generate the first schedule of a number of individuals per generation.
3. The fitness of that generation will be calculated to determine the best schedule. Fitness is obtained from 2 steps, namely;
 - a. Counting the number of activities that collide with each other, where a teacher or a teaching class is assigned to 2 or more rooms at the same time. The fewer the number of collisions, the better.
 - b. Calculation of the match of requests with those obtained in the schedule. The more requests a schedule fulfills, the better.

Implementation of Genetic Algorithm in Determining Class Schedules Based on User Needs

4. If from that generation there is not a single schedule that can fulfill all requests, then go to step 5. If so, go to step 9.
5. Production of new generations from previous generations through mutation and crossover processes.
6. From the new generation, combine with the existing generation to become a larger population.
7. Calculate the fitness of the large population as done in step 3
8. Prune the population by taking only the best individuals of the initial generation. Go back to step 4.
9. Show the best choice of schedule recommendations.

2.2 Scheduling

Scheduling comes from the word schedule which gets the suffix pen which means time division based on a work order arrangement plan or list or activity table or activity plan with a detailed implementation time division. Scheduling is a very important problem in an educational institution [1].

3. Result and Discussion

3.1. Database Schema

the following is the schematic of the base system

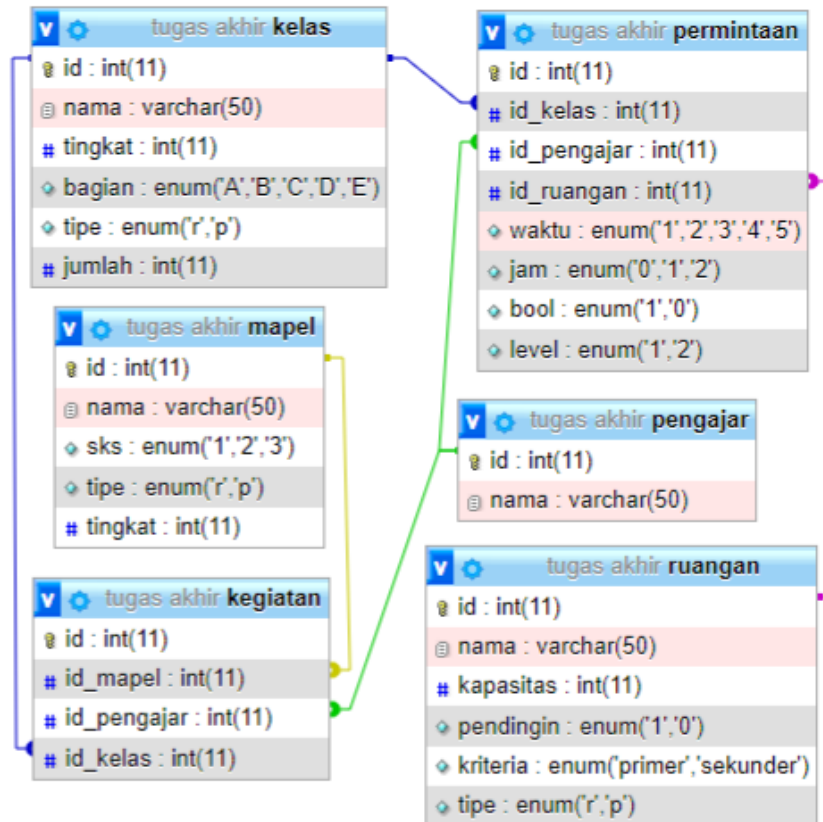


Figure 3. system database schema

system database consists of 6 tables, namely:

- *kelas*, that is, student class
- *permintaan*, the requests from teachers and student classes
- *mapel*, that is lecture subjects
- *pengajar*, that is lecturer
- *kegiatan*, namely the relationship between lecturers, subjects, and classes
- *kelas*, namely the classroom where the activity is carried out

3.2. System Interface Design

The system for determining the lecture schedule is implemented on a web-based basis. The following is the design of the system interface.

Daftar Relasi Kegiatan Ajar-Mengajar

crud data

- kelas
- pengajar
- mapel
- ruangan
- kegiatan
- permintaan

tingkat	nama mapel	SKS	pengajar			
			Kelas A	Kelas B	Kelas C	Kelas D
19	Ilmu Sosial dan Budaya Dasar	2	MIPA1	MIPA1	MIPA1	MIPA1
	Pancasila	2	MIPA2	MIPA2	MIPA2	MIPA2
	Kalkulus II	3	Supriana	Supriana	Supriana	Ari Mogi
	Statistika Dasar	3	Santiyasa	Santiyasa	Santiyasa	Santiyasa
	Struktur Data	3	Eka Karyawati	Eka Karyawati	Widiartha	Widiartha
	Matematika Diskrit II	3	Gede Santi	Gede Santi	Vida	Vida
	Aljabar Linear Elementer	3	Arida	Arida	Arida	Arida
18	Etika	2	MIPA3	MIPA3	MIPA3	MIPA3
	Rekayasa Perangkat Lunak	3	IBM Mahendra	IBM Mahendra	Agung Kadyanan	Agung Kadyanan
	Pemrograman Berorientasi Obyek	3	Hendra	Hendra	Hendra	Gede Arta
	Riset Operasi	3	Agung Raharja	Agung Raharja	Agung Raharja	Widiartha
	Basis Data Lanjut	3	IBG Dwidasmara	IBG Dwidasmara	IBG Dwidasmara	Agus Sanjaya
	Organisasi dan Arsitektur Komputer	3	Suhartana	Suhartana	Suhartana	Suhartana
	Komunikasi Data dan Jaringan Komputer	3	Bayu	Bayu	Bayu	Ari Mogi
17	Tata Tulis Karya Ilmiah	2	Anom Cahyadi	Anom Cahyadi	Anom Cahyadi	-
	Kewirausahaan	2	Gede Santi	Gede Santi	Gede Santi	-
	Etika Profesi	2	Cokorda Rai	Cokorda Rai	Cokorda Rai	-

Figure 4. System interface main page design

Figure 4 is the main view of the system. On this page the user can see all the data systematically, where the data is presented in the form of a relationship between one data and another to make it easier for the user to manage the data.

There is also the system can save the table that was successfully formed in .xls format using Javascript programming. Here is a snippet of the program code syntax.

Table 1. The syntax snippet of the function exports a table in the form of .xls

```

- function exportTableToExcel(tableID, filename = ''){
-     var downloadLink;
-
-     var dataType = 'application/vnd.ms-excel';
-
-     var tableSelect = document.getElementById(tableID);
-     var tableHTML = tableSelect.outerHTML.replace(/ /g,
- '%20');
-
-     // Specify file name
-     filename = filename?filename+'.xls':'excel_data.xls';
-
-     // Create download link element
-     downloadLink = document.createElement("a");
-
-     document.body.appendChild(downloadLink);

```

3.3. Genetic Algorithm Implementation

1. Initialization

The first step in the genetic algorithm is to generate a number of individuals that are used as the first generation. The types of data used in constructing the individual are activity data and room data.

2. Evaluation

From a number of individuals in the first generation, the fitness value of each generation is calculated. The fitness value is the value to determine the level of goodness of the individual. In this study, the researcher determined that the higher the fitness value of an individual, the better the individual.

3. Reproduction

Reproduction is done in 2 ways, namely crossover and mutation. The crossover method used is Partially Mapped Crossover (PMX). PMX is a 2-point crossover with the addition that no chromosome in an individual can exist more than once.

4. Reevaluation

New individuals resulting from the reproduction process will be calculated to obtain a fitness value to continue in the elimination process.

5. Elimination

The elimination method used is the Roulette Wheel. In this process, individuals who have good fitness values will have a greater chance of being selected as individuals in the next generation.

3.4 Testing

Here are the results of the tests that have been carried out. The graphs below display the fitness value of the best individual in a generation in 4 experiments on the system. In the graph, the vertical axis indicates the value of the best individual in a generation, while the horizontal is the order of the generations. Each experiment was carried out with repetitions of 100 generations.

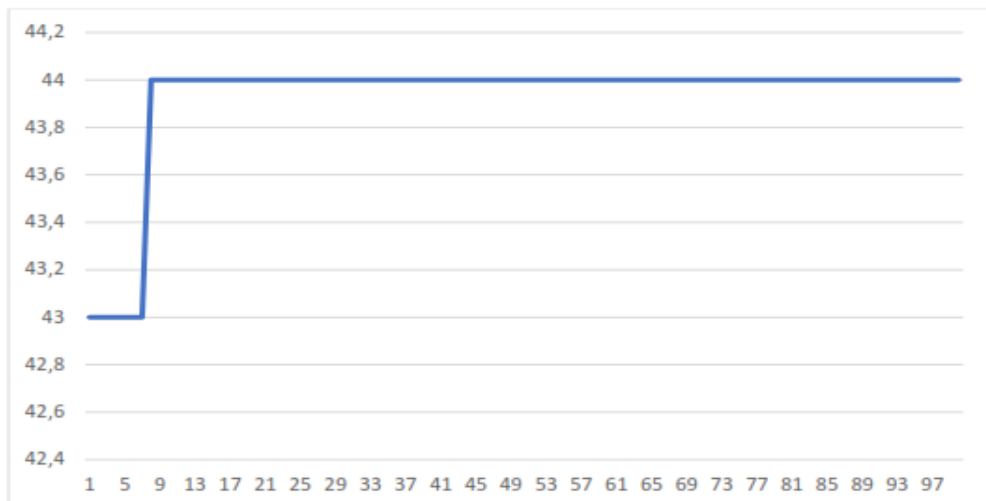


Figure 5. The graph of the best individual change on the first try

From figure 5 above, first generation has greatest value of 43 constantly until 7th generation, then increased to 44 at 8th to 100th generation.

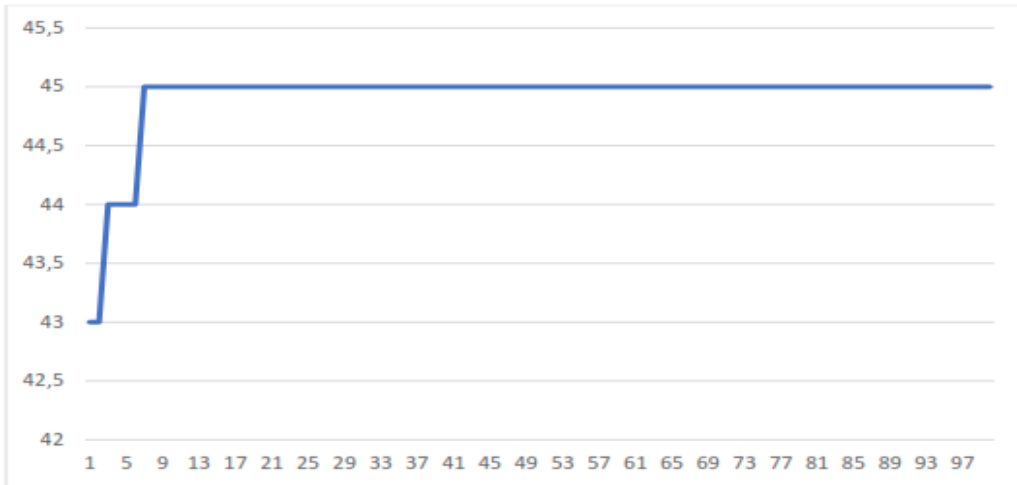


Figure 6. The graph of the best individual change on the second tries

On the next figure, first generation has greatest value of 43 to 2nd generation, then increased to 44 at 3rd to 6th generation, then increased again to 45 at 7th to 100th generation.

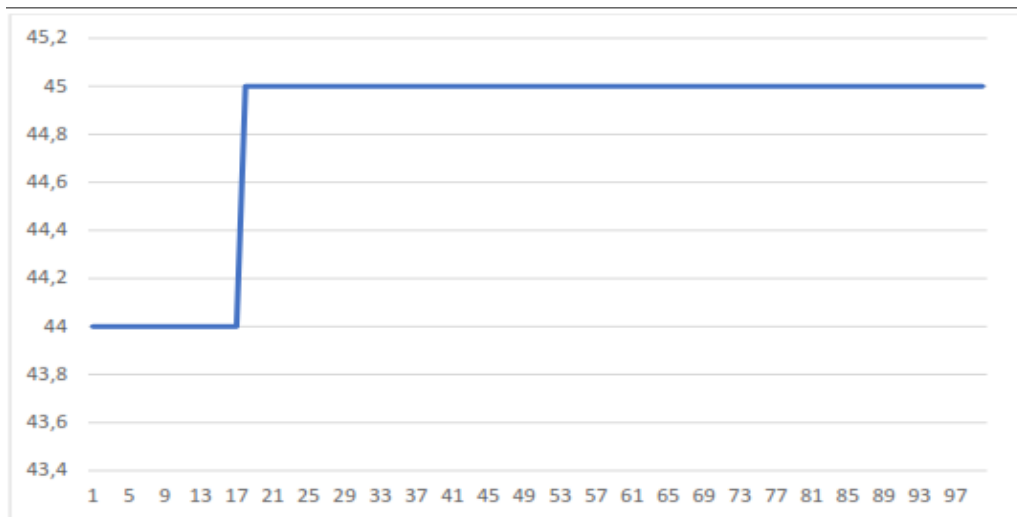


Figure 7. The graph of the best individual change on the third tries

Next one, the first generation has greatest value of 44 to 17th generation, then increased to 45 at 18th to 100th generation.

Implementation of Genetic Algorithm in Determining Class Schedules Based on User Needs

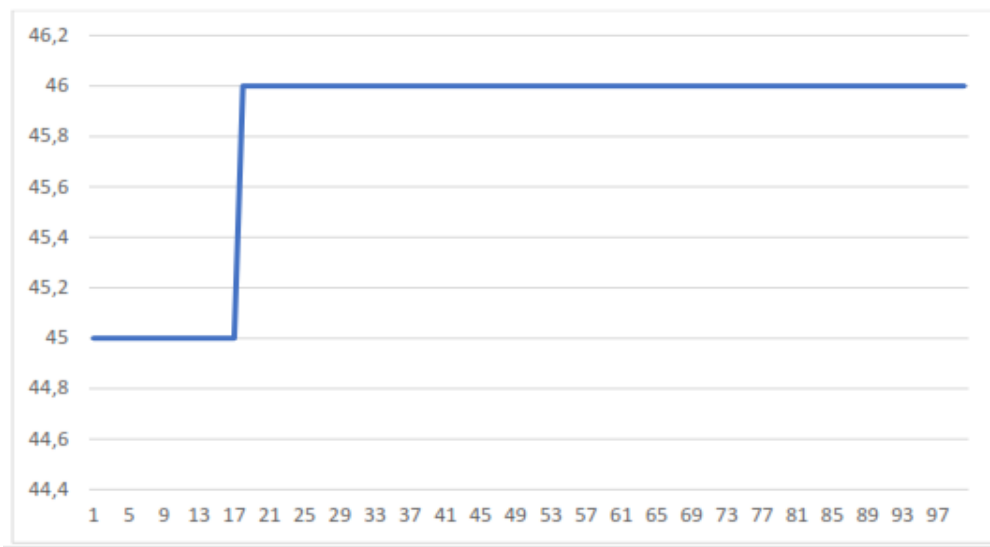


Figure 8. The graph of the best individual change on the 4th tries

Lastly, the first generation has greatest value of 45 to 17th generation, then increased to 46 at 18th to 100th generation.

From the graphs above, it can be concluded that the average increase in the fitness value of the best individuals in each generation is:

- trial 1 : 0.010101
- Trial 2 : 0.020202
- Trial 3 : 0.010101
- Trial 4 : 0.010101

For the calculation of accuracy, the 5th experiment was carried out with all the same parameters as the previous experiment. In the 100th iteration, the individual with the best fitness value is 45. The individual has a total of 12 activities that have requests, with 7 of them being fulfilled, and 96 possible collisions with no collisions. From the data obtained, the accuracy can be calculated as follows.

$$\begin{aligned} & \text{accuration} \\ &= \left(0.5 \frac{\text{the total number of collisions that occurred} - \text{number of collisions}}{\text{Total number of collisions that can occur}} \right) \\ &+ \left(0.5 \frac{\text{the number of activities that meet the demand}}{\text{the total number of activities that have a demand}} \right) \\ &= \left(0.5 \frac{96 - 0}{96} \right) + \left(0.5 \frac{7}{12} \right) = \left(0.5 \frac{96 - 0}{96} \right) + \left(0.5 \frac{7}{12} \right) = 0.79166 \end{aligned}$$

4. Conclusion

From the research that has been done, the conclusions obtained is, the genetic algorithm has succeeded in optimizing the predetermined schedule. However, it never got the best results even though it had been done for 100 generations due to the shortcomings of the genetic algorithm. It takes at least 18 iterations to get a constant value with increments as high as 1-2 fitness values. The system has an accuracy of 0.7966 with a value of 1 as a form of perfect accuracy.

References

- [1] Ahmat Josi. *Implementasi Algoritma Genetika Pada Aplikasi Penjadwalan Perkuliahan Berbasis Web Dengan Mengadopsi Model Waterfall (Studi Kasus STMIK Prabumulih)*. 2017
- [2] I Made Suwija Putra. *Penerapan Algoritma Genetika Dan Implementasi Dalam Matlab*. 2018
- [3] Tanweer Alam, et al., *Genetic Algorithm: Reviews, Implementations and Applications*. 2020
- [4] Listanto Tri Utomo, Farid Wahyudi. *Implementation Of Genetic Algorithm in Optimization of Subject Scheduling*. 2020
- [5] David Kristiadi, Rudy Hartanto. *Genetic Algorithm for Lecturing Schedule Optimization (Case Study: University of Boyolali)*. 2019

This page is intentionally left blank

Penerapan Metode MFCC dan *Naive Bayes* untuk Deteksi Suara Paru-Paru

Ni Made Ayu Suandewi ^{a1}, I Gede Arta Wibawa ^{a2}, I Gusti Ngurah Anom Cahyadi ^{a3},
I Komang Ari Mogi ^{a4}, Ngurah Agus Sanjaya ER ^{a5}, Cokorda Rai Adi Pramatha ^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Badung, Bali, Indonesia

¹suandewi72@gmail.com

²gede.arta@unud.ac.id

³anom.cp@unud.ac.id

⁴arimogi@unud.ac.id

⁵agus_sanjaya@unud.ac.id

⁶cokorda@unud.ac.id

Abstract

Lung disease is an unpleasant illness that can be dangerous if not treated properly. This is because lung disease can infect others. The lungs are an important part of the human organ that distributes oxygen throughout the body, so this lung disease needs to be treated with proper procedures. Lung disease problems can be solved using an expert system. Expert systems can help doctors work to provide an early diagnosis of ear diseases. The method used in this study is the MFCC method, which provides compelling information about lung disease and provides treatment solutions based on the symptoms of each existing disease. This system is performed by calculating the symptom weights of the disease, which is obtained from expert experience, and produces the optimum value in the form of the maximum value. The highest value data provides the results of the disease diagnosis at the level of confidence in the form of percentage values. Based on the test results, the goodness of fit between the system diagnostics using the system validation test method and the expert diagnostic results is 90%. The results of the tests performed show that the system is operating normally according to its capabilities.

KeyWords: MFCC, Lung Health, Machine Learning

1. Pendahuluan

Paru-paru merupakan salah satu organ tubuh yang sangat vital bagi manusia yang berperan dalam sistem pernapasan karena dapat memenuhi kebutuhan oksigen tubuh. Jika terjadi sebuah gangguan pada paru – paru dapat menyebabkan kecacatan hingga kematian [1]. Banyak orang di seluruh dunia menderita kerusakan paru-paru dan sering mengalami kecacatan, yang bahkan dapat menyebabkan kematian pasiennya. Diperkirakan ratusan ribu hingga jutaan orang di seluruh dunia menderita penyakit paru-paru setiap tahun, menyebabkan 19% kematian dan 15% cacat seumur hidup di seluruh dunia (FIRS2010).

Penemuan mengenai gangguan pada paru-paru biasanya dimiliki oleh dokter dengan melakukan diagnosa apakah paru-paru seseorang normal atau abnormal[2]. Untuk mendapatkan perawatan yang tepat, Anda perlu tahu persis apa yang dilakukan paru-paru Anda. Orang dengan penyakit saluran napas kronis membutuhkan perangkat yang dapat mendeteksi kondisi paru-paru secara real time. Penelitian ini berfokus pada kajian dan uji coba metode naive bayesian yang dimodifikasi dalam menentukan kondisi paru-paru seseorang.

Penelitian sebelumnya mengenai gangguan paru-paru menggunakan metode *Convolutional Neural Network* (NFCC) untuk melakukan klasifikasi suara paru-paru untuk mengetahui sebuah paru-paru apakah dalam keadaan normal atau tidak. Klasifikasi suara paru-paru dikelompokkan dalam 3 jenis suara yaitu normal, crackel, wheeze, dan crackel-wheeze/both. Hasil akhir yang

Penerapan Metode MFCC dan Naive Bayes untuk Deteksi Suara Paru-Paru didapatkan untuk akurasi dari masing-masing class sebesar 74% dari seluruh dataset yang diuji [2].

Penelitian lainnya juga membahas mengenai pengenalan suara paru-paru dengan metode MFCC. Pada penelitian ini dilakukan klasifikasi suara paru-paru normal dan abnormal dengan MFCC sebagai ekstraksi ciri dan *Backpropagation* sebagai metode pengelompokannya dengan menghitung *Coefficient Cepstral* dari suara paru-paru. Hasil akhir yang didapatkan untuk hasil klasifikasi dengan akurasi yang diperoleh sebesar 93,97% untuk data latih dan 92,66% untuk data uji [1].

Langkah-langkah yang dilakukan dalam penelitian ini adalah pengolahan data paru, ekstraksi ciri, dan klasifikasi. Langkah yang digunakan untuk mendeteksi kondisi paru-paru pada penelitian ini adalah file audio yang pertama. wav melakukan langkah pra-pemrosesan yang terdiri dari normalisasi data, pengurangan kebisingan dan penghilangan keheningan. Data tersebut terdiri dari ekstraksi fitur seperti Premphasize, frame blocking, windowing, fast folia conversion, mel frequency wrapping, mel frequency wrapping, dan DCT (discrete cosine transform). Mendapatkan nilai koefisien yang diproses oleh proses klasifikasi naive bay yang dimodifikasi oleh proses ini.

2. Metode Penelitian

2.1 Tinjauan Pustaka

Untuk mendukung penelitian yang dilakukan, melakukan kajian terhadap penelitian sebelumnya sangat penting dilakukan dan merupakan kebutuhan bagi peneliti yang menjadikan penelitian sebelumnya sebuah acuan untuk kedepannya. Salah satu penelitian mengungkapkan bahwa pengenalan suara paru-paru dapat membantu dokter dalam mengenali suara paru-paru normal dengan penelitian yang dilakukan menggunakan metode LPC dan JST di mana suara paru-paru dianggap sebagai speech dan menghitung koefisien LPC. Akurasi yang didapatkan dengan JST yaitu 100% [3].

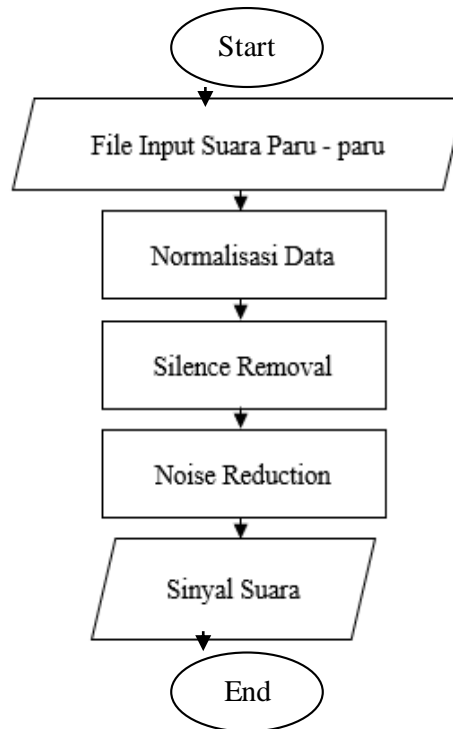
Penelitian lain juga dilakukan mengenai analisa kelainan suara paru-paru menggunakan media android sebagai media implementasi yang merupakan pembaharuan dari teknik auskultasi yang dilakukan oleh dokter dengan hasilnya adalah menampilkan suara paru-paru saat ekspirasi dan inspirasi [4].

2.1 Dataset

Pada penelitian ini, data yang digunakan adalah data sekunder yang didapatkan melalui website ICBHI dengan jumlah data latih sebanyak 3000 data dan data uji sebanyak 300 data dengan pembagian kelas sebanyak 4 kelas yakni wheeze, Tracheal, Vesicular, dan Crackle.

2.2 Preprocessing

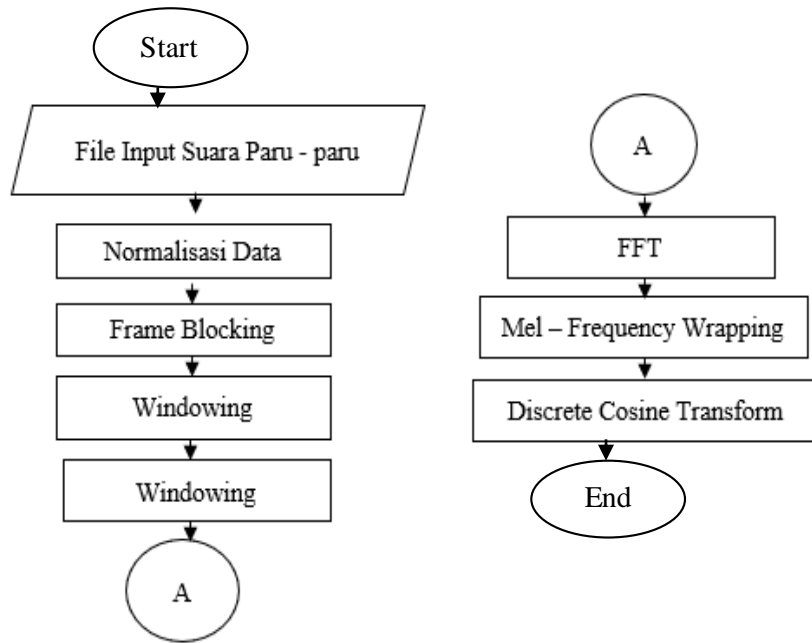
Processing adalah proses yang bertujuan memperbaiki rekaman suara input agar menghasilkan sinyal suararekaman yang baik, pada penelitian ini proses preprocessing terbagi menjadi 3 tahap yaitu normalisasi data, silence removal dan noise reduction. pada tahap normalisasi merupakan melakukan penyetaraan amplitude setiap sinyal yang terekam. setelah dilakukan normalisasi, setiap sampel rekaman suara yang terbentuk pasti mengandung daeran silence. oleh karena itu tahap kedua pada proses preprocessing yaitu silence removal dimana sinyal suara yang merupakan silence akan dihilangkan. selanjutnya tahap noise reduction merupakan proses untuk mengurangi besar noise pada sinyal suara input. hasil dari proses preprocessing ini adalah sinyal suara yang tidak memiliki noise dan silence. flowchart proses preprocessing ini dapat dilihat pada gambar 1.



Gambar 1. Gambar Proses Preprocessing

2.3 Ekstraksi Fitur

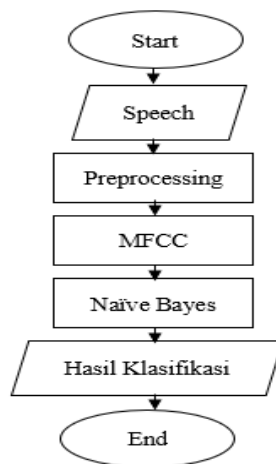
Setelah pretreatment untuk ekstraksi ciri, digunakan metode MFCC. Ini dilakukan dalam beberapa tahap: pre-emphasis, frame blocking, windowing, dan Fast Fourier Transform (FFT). Pembungkus frekuensi mel, dan transformasi kosinus diskrit DCT. Pada tahap pertama, pre-emphasis dilakukan untuk meningkatkan frekuensi tinggi untuk meningkatkan akurasi ketika ekstraksi fitur diperoleh. Selanjutnya, tahap frame blocking dimana sinyal audio tersegmentasi menjadi beberapa frame. Pada tahap ini, efek sinyal terputus-putus terjadi. Tahap pemrosesan jendela dijalankan untuk meminimalkan diskontinuitas sinyal pada awal dan akhir setiap frame. Kemudian dilanjutkan ke tahap FFT. Tahap ini menerjemahkan dari domain waktu ke domain frekuensi setiap N frame sampel. Setelah data dikonversi ke domain frekuensi, langkah selanjutnya adalah pembungkusan frekuensi mel. Di sini, sinyal diplot terhadap spektrum mel, meniru pendengaran manusia. Selanjutnya, langkah terakhir adalah transformasi kosinus diskrit. Ini adalah tahap mengubah spektrum mel menjadi domain waktu yang digunakan untuk mendapatkan koefisien. Koefisien cepstrum frekuensi Mel (MFCC). Proses ekstraksi fitur dapat dilihat pada gambar 2 untuk memperoleh Discrete Courier Transform.



Gambar 2 . Proses Ekstraksi MFCC suara data latih paru paru

2.4 Klasifikasi

Pada proses klasifikasi Naïve Bayes yang dimodifikasi, data fitur yang dihasilkan oleh proses ekstraksi fitur akan menjadi nilai input untuk proses klasifikasi ini. Klasifikasi Naïve Bayes memiliki dua proses tambahan: proses validitas data-ke-data. Bobot voting data sampel audio yang digunakan dalam proses pelatihan meningkatkan tingkat keberhasilan sistem ini dalam klasifikasi. Sebaliknya, semakin sedikit data sampel audio yang digunakan dalam proses pelatihan, semakin rendah tingkat kesalahan atau kesalahan klasifikasi. Kemudian normalkan data latih dan uji untuk mengurangi jarak antar data. Kemudian hitung jarak terakreditasi antara data pelatihan dan validasi setiap data pelatihan. Selanjutnya menghitung nilai validitas yang diperoleh dari perhitungan jarak adhesi antar data latih. Kemudian menghitung jarak koneksi antara data latih dengan data uji dan nilai yang diolah dengan proses pembobotan. Memilih. Setelah mendapatkan nilai dari perhitungan jarak inklusi antara data latih dan data uji, maka dilakukan proses voting bobot. Proses pembobotan diambil dari perhitungan jarak inklusi antara data latih dengan data uji dan validitas data latih, yaitu nilai maksimum yang diperoleh untuk menentukan kelas klasifikasi. Proses klasifikasi dapat dilihat pada gambar 3.



Gambar 3. Diagram alir proses klasifikasi

2.5 Cara kerja system

Bagian pengujian ini dilakukan dengan menggunakan K-Fold Cross Validation. Di sini, semua data yang digunakan dibagi menjadi beberapa bagian untuk pengujian dan beberapa untuk pelatihan. Proses ini dijalankan beberapa kali dengan menggabungkan dataset yang berbeda. Tes dijalankan menggunakan validasi silang 5 kali lipat. Ini berarti menjalankan eksperimen 5 kali secara bertahap. Pada percobaan pertama, kami membuat partisi pertama untuk data pengujian dan partisi lainnya untuk data pelatihan. Pada percobaan kedua, partisi kedua digunakan sebagai data uji dan partisi lainnya digunakan sebagai data latih. Pada percobaan ketiga, kami membuat partisi ketiga untuk data uji dan partisi lain untuk data pelatihan. Dari kelima hasil eksperimen tersebut, nilai evaluasi kinerja model dicatat dengan menggunakan matriks konfusi yang dapat menentukan jumlah data yang teridentifikasi dengan benar dan menghitung keakuratan sistem yang dibuat. Untuk mencoba atau menguji metode klasifikasi saat melakukan pembelajaran mesin, Anda memerlukan sistem validasi yang digunakan untuk memvalidasi data yang diprediksi oleh mesin. Hitung apakah prediksi tersebut benar atau salah, dan seberapa akurat prediksi tersebut dari model dan kumpulan data yang ada (Chandra et al, 2020). Pada tahap ini digunakan metode confusion matrix untuk menghitung tingkat akurasi.

Table 1. Confusion Matrix

TrueLabel	Kelas hasil matriks	
	positif	negatif
positif	TP	FN
negatif	FP	TN

Dalam menentukan akurasi terdapat 3 rumus yang digunakan adalah akurasi, precision, dan recall dimana dapat dilihat dibawah persamaan ini

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (1)$$

$$Akurasi = \frac{TP}{TP + PP} \times 100\% \quad (2)$$

$$Akurasi = \frac{TP}{TP + PN} \times 100\% \quad (3)$$

Dimana:

TP(True Positives) = jumlah objek positif yang benar diklasifikasikan

TN(True Negatif) = Jumlah objek negative yang salah diklasifikasikan

FP(False Positives) = jumlah objek negative yang benar diklasifikasikan

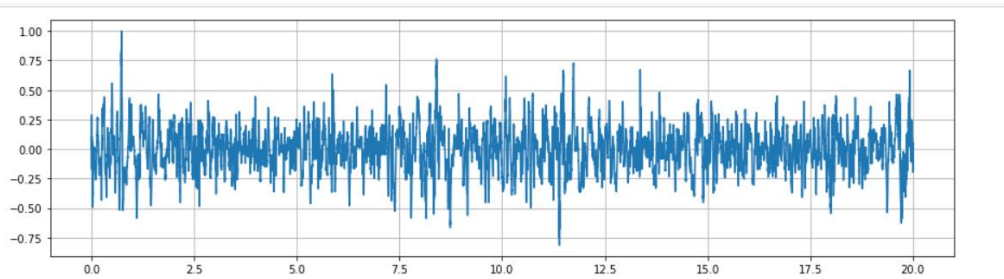
FN(False Negatives) = jumlah objek positif yang salah diklasifikasikan

3. Hasil Dan Pembahasan

Sebanyak 80% dari total data digunakan pada tahap pelatihan dan validasi. Beberapa Aplikasi yang telah dirancang dan diimplementasikan berbasis sistem yang memungkinkan dalam menampilkan data yang berbentuk audio, dimana frontend serta backend utama aplikasi ini dikembangkan berbasis web dengan penggunaan Bahasa pemrograman python, Penggunaan Bahasa pemrograman python dalam pembuatan sistem kemudian dilakukan ekstraksi data yang bertujuan untuk mempermudah dalam mengatasi noise pada inputan audio data suara paru paru yang di ambil melalui web penyedia data suara yang merupakan web free service. Pembuatan sistem juga menggunakan algoritma MFCC, Mel Frequency Cepstrum Coefficients digunakan sebagai fitur dari klip suara. MFCC banyak digunakan dalam sistem pengenalan suara. Mereka

Penerapan Metode MFCC dan Naive Bayes untuk Deteksi Suara Paru-Paru

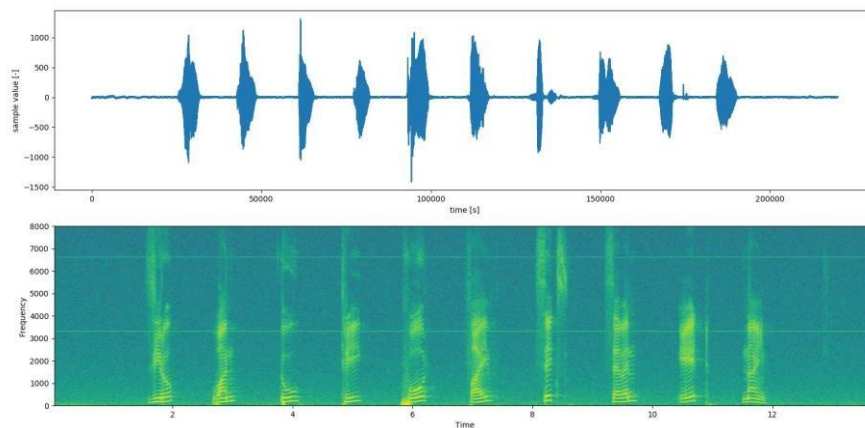
juga digunakan secara luas dalam pekerjaan sebelumnya pada deteksi suara pernapasan tambahan karena mereka menyediakan ukuran spektrum daya jangka pendek dari sinyal domain waktu. Baik konten frekuensi maupun waktu penting untuk membedakan antara suara tambahan yang berbeda, karena suara tambahan yang berbeda dapat ada dalam satu klip pada periode waktu yang berbeda dan durasinya berbeda. Oleh karena itu, MFCC sangat membantu dalam menangkap perubahan konten frekuensi sinyal dari waktu ke waktu. Frekuensi ditempatkan pada skala mel, yang merupakan skala frekuensi nonlinier yang jaraknya dianggap sama oleh sistem pendengaran manusia. Keluaran MFCC berupa vektor fitur 2 dimensi (waktu dan frekuensi), yang kemudian diratakan menjadi larik satu dimensi sebelum diproses lebih lanjut. Proses penghitungan MFCC dapat dilihat pada gambar 4.



Gambar 4. Tampilan MFCC

Berdasarkan gambar 4, audio paru-paru memiliki durasi 20 detik dengan amplitude maksimal dari suara paru-paru adalah 1.00. Hasil MFCC memudahkan membaca fitur yang ada pada file audio dan memudahkan dalam melakukan audio windowing untuk mencari suara paru-paru tanpa adanya suara tambahan lainnya. Hasil dari penelitian ini adalah hasil pengujian akurasi MFCC terhadap penentuan hasil prediksi paru-paru. Berdasarkan hasil yang didapatkan, dari proses pengujian akurasi MFCC diperoleh dua buah analisa yang dirangkum dari proses pengujian serta dari hasil pengujian yang didapatkan, berikut merupakan analisa yang dirangkum dari pengujian pada penelitian ini :

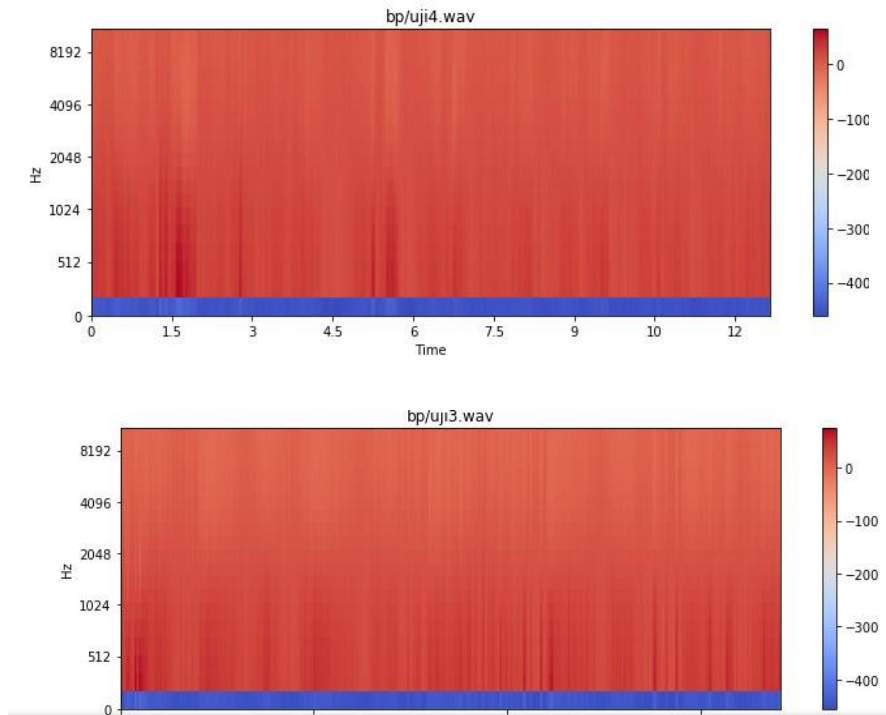
Pertama dari observasi awal terhadap interaksi audio selama proses pengambilan data dan pengujian, kelima responden anak menunjukkan interaksi yang berbeda – beda , dimana masing – masing anak tersebut menunjukkan tingkat antusias dan terutama rasa ingin tahu yang tinggi ketika diminta menggunakan aplikasi saat proses pengambilan data maupun pengujian, dimana secara keseluruhan kelima audio tersebut memberikan respons yang positif selama kedua proses tersebut dan terutama terhadap sistem.



Gambar 5. Tampilan Audio Windowing

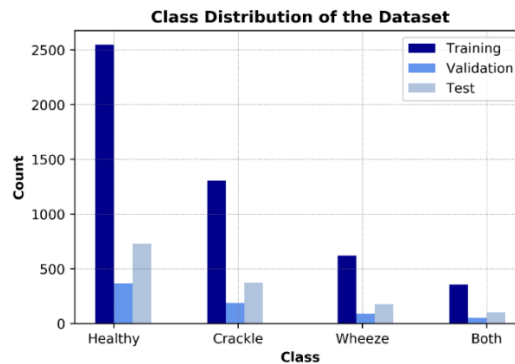
Pada gambar 5, merupakan hasil dari proses windowing yang hanya memetakan suara paru-paru dengan melihat amplitude pada proses MFCC sebelumnya dan pada proses windowing

merupakan tahapan untuk mengidentifikasi apakah inputan suara sudah dapat di input pada sistem atau masih perlu adanya tahapan lain, tahapan ini bertujuan untuk mengetahui inputan audio melalui pengecekan audio suara paru-paru serta hasil dari proses windowing.



Gambar 6. Iterasi MFCC

Kedua, berdasarkan hasil pengujian tingkat akurasi yang diperoleh, jika rata-rata tingkat akurasi audio adalah 88% secara keseluruhan, maka kesamaan minimum yang digunakan adalah pengenalan karakter yang diperoleh dalam penelitian ini. ketepatan. Berdasarkan kedua analisis tersebut, dapat disimpulkan bahwa hasil yang diperoleh adalah metode MFCC yang melebihi tingkat akurasi yang diharapkan yang dapat dilihat pada gambar 7.



Gambar 7. Hasil Pengujian

Pada gambar 7, merupakan tampilan hasil dari pengujian di mana setiap class dilakukan data training dan data testing yang menunjukkan bahwa class healthy memiliki data sebanyak 500 hingga 1000 data suara paru-paru termasuk ke dalam class healthy kurang dari 500 termasuk class crackel. Sehingga hasil akurasi dapat dikatakan sangat baik dengan nilai akurasi yang

diperoleh sebesar 88%.

4. Kesimpulan

Sistem prediksi suara paru-paru menggunakan MFCC dan Naïve Bayes dalam melakukan proses dan output keadaan paru-paru menggunakan audio sebagai media melakukan prediksi dapat dilakukan dengan baik melalui beberapa proses iterasi dan proses normalisasi. Akurasi sistem prediksi paru-paru menggunakan MFCC dan Naïve bayes menggunakan metode uji dan data latih yang telah di sediakan.

REFERENSI

- [1] F. Syafria, A. Buono, and B. P. Silalahi, "Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier," *J. Ilmu Komput. dan Agri-Informatika*, vol. 3, no. 1, p. 27, 2017, doi: 10.29244/jika.3.1.27-36.
- [2] I. W. Hasanain and A. Rizal, "Klasifikasi Suara Paru-Paru Menggunakan Convolutional Neural Network (CNN)," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3218–3223, 2021.
- [3] A. Rizal, L. Anggraeni, and V. Suryani, "Pengenalan Suara Paru-Paru Normal Menggunakan LPC dan Jaringan Syaraf Tiruan Back-Propagation," *Preceeding Int. Semin. Electr. Power, Electron. Commun. Control. Informatics (EECCIS 2006)*, pp. 6–10, 2006.
- [4] D. Kurniawan, "Rancang Bangun Alat Deteksi Suara Paru-Paru Untuk Menganalisa Kelainan Paru-Paru Berbasis Android," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 2, no. 2, pp. 156–168, 2017, doi: 10.21831/elinvo.v2i2.17309.
- [5] D. Ilmu and F. Ugm, "Klasifikasi Suara Paru-Paru Berdasarkan Ciri MFCC," no. x, pp. 1–12, 2012, doi: 10.22146/ijeis.xxxx.
- [6] F. Ramadhana, F. Fauziah, and W. Winarsih, "Aplikasi Sistem Pakar untuk Mendiagnosa Penyakit ISPA menggunakan Metode Naive Bayes Berbasis Website," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 4, no. 3, p. 320, 2020, doi: 10.30998/string.v4i3.5441.
- [7] A. Wijayanto and P. J. R, "Deteksi Kelainan Parenkim Paru Berdasarkan Power Spectra Density Suara Paru Dengan Metode Welch," vol. 2011, no. 1es, pp. 978–979, 2011.

Pengembangan Perangkat Lunak Pembelajaran Penulisan Aksara Bali Menggunakan SDLC Untuk Anak-Anak

I Gusti Ayu Nyoman Tri Wahyuni Mahayani^{a1}, I Gede Arta Wibawa^{a2}, I Gusti Ngurah Anom Cahyadi Putra^{a3} I Wayan Santiyasa^{a4},
AAIN Eka Karyawati^{a5}, Ngurah Agus Sanjaya ER^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bukit Jimbaran, Badung, Bali, Indonesia

¹tri.wahyuni7540@gmail.com

²gede.arta@unud.ac.id

³anom.cp@unud.ac.id

⁴santiyasa@unud.ac.id

⁵eka.karyawati@unud.ac.id

⁶agus_sanjaya@unud.ac.id

Abstract

Balinese script is a traditional script originating from Bali. To maintain sustainability, formal education in schools through Balinese language subjects is one of many ways to maintain it. Apart from studying at school, it would be even better if Balinese script learning could be done outside of school by appealing to the concept of learning. One way to create interesting learning is through application games. Making a Balinese script writing application aims to help students learn Balinese script by implementing the Software Development Life Cycle (SDLC) and Black Box testing of making Balinese script writing applications. Making the application is expected to be one of the media that can assist students in learning Balinese script.

Keywords: Balinese Script, Software Development Life Cycle (SDLC), Children, Software, Game

1. Pendahuluan

Aksara Bali adalah Aksara tradisional masyarakat Bali yang masih digunakan sampai sekarang. Upaya pelestarian dan pembinaan bahasa Bali telah dilakukan sejak jaman dahulu, seperti upaya pedoman pemasangan Aksara Bali dan pengejaan bahasa Bali dengan huruf latin [8]. Selain itu pelestarian Aksara Bali juga bisa dilakukan dan diterapkan pada anak-anak agar mereka tidak melupakan budaya yang telah lama ada dan telah diwariskan.

Perkembangan teknologi dimasa ini memiliki dampak positif dan negative, dimana bila dilihat dari dampak negative adanya teknologi membuat anak-anak melupakan budaya tradisional yang ada dari permasalahan tersebut maka dilakukan suatu tindakan yaitu membangun perangkat lunak untuk melatih anak-anak dalam menulis Aksara Bali dan menquji kegunaan perangkat lunak yang dibangun. Aplikasi yang dibuat akan berupa media permainan atau *game* edukasi Aksara Bali, permainan atau game edukasi berperan sebagai bentuk perangkat lunak interaktif yang mampu memacu imajinasi pada anak. [9].

Penelitian yang dilakukan bertujuan untuk melatih anak-anak dalam menulis aksara bali. Metode yang dilakukan adalah *Software Development Life Cycle* (SDLC) serta pengujian yang dilakukan menggunakan uji *black box* dan uji kegunaan perangkat lunak

2. Metode Penelitian

Metode siklus hidup pengembangan perangkat lunak (SDLC) adalah metode tradisional yang berguna untuk mengembangkan, memelihara, dan menggunakan informasi. Langkah-langkah dalam metode ini adalah analisis kebutuhan, perancangan, implementasi, pengujian [1].

2.1. Analisis Kebutuhan

Adapun analisis kebutuhan pada penelitian ini yaitu terkait kebutuhan pengguna yang dikerjakan pada aplikasi [10]. Kebutuhan tersebut berisikan tentang menu-menu yang ada pada aplikasi dimana menu-menu tersebut mampu memenuhi kebutuhan yang pengguna inginkan. Dalam analisis kebutuhan penelitian dilakukan pencarian responden untuk aplikasi dengan hasil responden yang didapat adalah anak-anak usia 7-12 tahun yang mendapatkan pelajaran menulis Aksara Bali di sekolah dasar. Berikut adalah analisis kebutuhan berupa menu-menu yang ada pada aplikasi. Tabel 1 berisikan Informasi kebutuhan dari pengguna dan sistem.

Tabel 1. Tabel Kebutuhan Pengguna

No	Analisis Kebutuhan	
	Anak-Anak	Aplikasi
1	Anak-anak berperan sebagai pengguna, dimana mereka bisa mengakses aplikasi, melihat tampilan atau desain aplikasi, serta memilih menu-menu yang ada	Memberikan fasilitas untuk belajar sambil bermain kepada anak-anak
2	Anak-anak dapat belajar dan bermain <i>game</i> Aksara Bali guna meningkatkan kemampuan dalam menulis Aksara Bali serta melestarikan budaya yang telah ada.	Memberikan cara belajar baru yang lebih interaktif sehingga anak-anak tidak cepat bosan
3	Anak-anak dapat semakin berminat untuk belajar menulis Aksara Bali.	Memberikan pengalaman bagi anak-anak agar tidak hanya mengetahui tentang teknologi namun bisa ikut melestarikan budaya yang telah lama ada.

2.2. Desain (Perancangan)

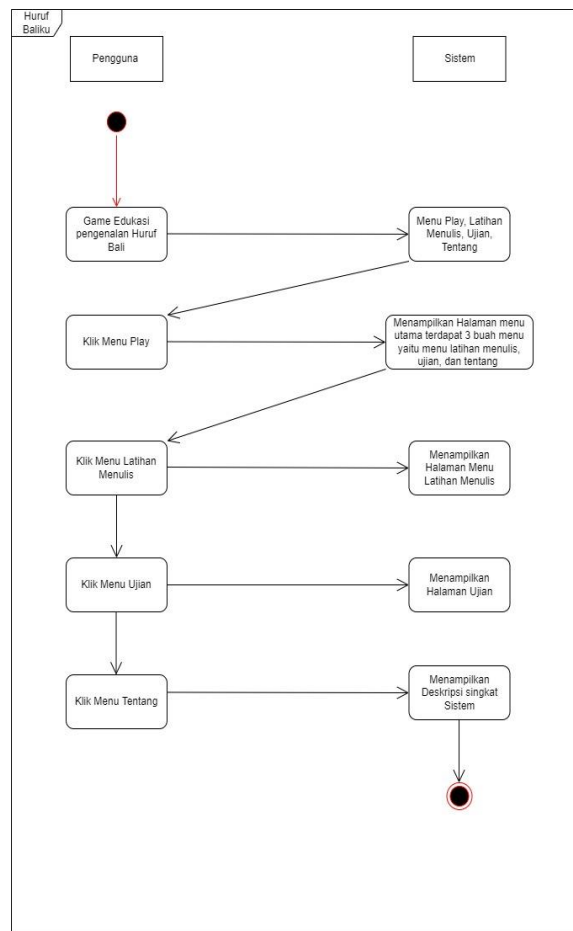
Dalam tahap perancangan dilakukan perancangan model UML (*Unified Modeling Language*), pembuatan desain *use case diagram* dan perancangan antarmuka (*interface*) aplikasi.

2.2.1. Model UML (*Unified Modeling Language*)

Bahasa Pemodelan Terpadu adalah bahasa grafis untuk melihat, mendefinisikan, membangun dan mendokumentasikan sistem perangkat lunak intensif [5]. Adapun bagian UML dari pembuatan sistem yaitu *activity diagram* dan *use case diagram*.

2.2.2. Activity Diagram

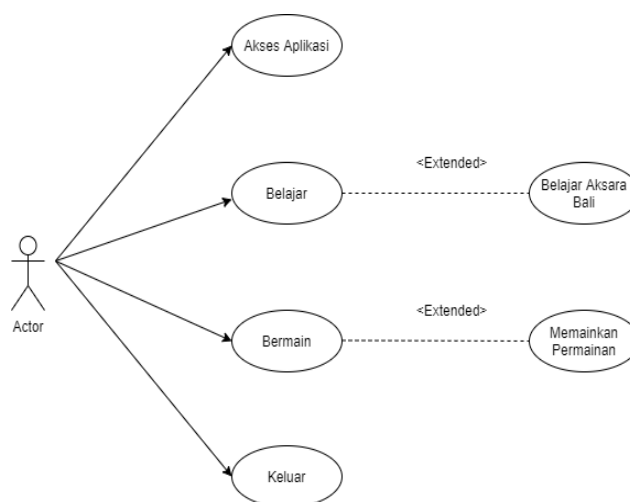
Diagram aktivitas adalah bagan yang menggambarkan kerangka kerja yang direncanakan dan bagaimana opsi perkembangan terjadi dan selesai [6]. Gambar 1 berisikan informasi diagram aktivitas aplikasi.



Gambar 1. Activity Diagram Aplikasi

2.2.3. Use Case Diagram

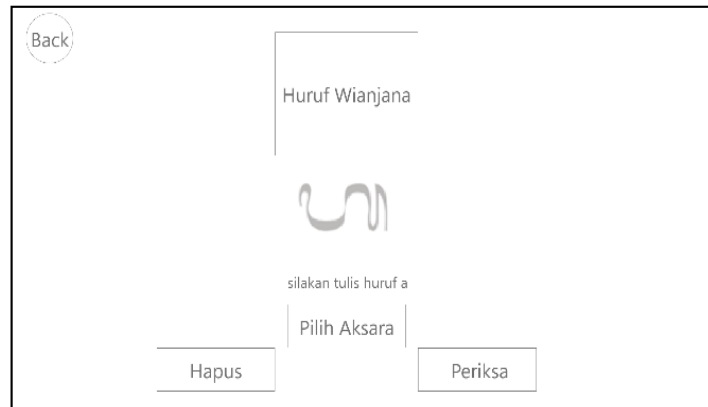
Kasus penggunaan berfungsi untuk mengimplementasikan interaksi apa saja yang dilakukan oleh pengguna dengan aplikasi yang dibuat [7]. Berikut diagram kasus penggunaan dari interaksi pengguna dan aplikasi. Gambar 2 berisikan informasi kasus penggunaan aplikasi.



Gambar 2. Use Case Diagram Aplikasi

2.2.4. Antarmuka (Interface) Aplikasi

Rancangan *Interface* yang dibuat akan ditunjukkan sebagai gambaran tentang bagaimana hasil akhir tampilan dari aplikasi. Gambar 3 merupakan salah satu desain antarmuka dari aplikasi.



Gambar 3. Halaman Latihan Menulis

2.3. Implementasi

Implementasi user interface dilakukan dengan melakukan proses coding sesuai dengan rancangan yang dibangun dengan memanfaatkan HTML, CSS, Javascript, dan Python

2.4. Pengujian

Pengujian yang dilakukan pada aplikasi dilakukan dengan dua skenario pengujian, yaitu tes *black box*, dan uji penggunaan perangkat lunak. Tes *black box* untuk menguji fungsionalitas perangkat lunak, sedangkan uji penggunaan dilakukan untuk menguji efektifitas perangkat lunak dalam membantu pengguna dalam berlatih menulis bali.

2.4.1. Uji Black Box

Uji *Black Box* bertujuan untuk menguji kesesuaian fungsi yang disediakan oleh system dengan spesifikasi kebutuhan pengguna. Uji *Black box* juga bertujuan untuk membuktikan apakah perangkat lunak yang dikerjakan telah mampu memenuhi kebutuhan pengguna [2]. Dari spesifikasi kebutuhan pengguna jika diterapkan ke dalam item menu diperoleh seperti tabel 2. Tabel 2 berisikan informasi kebutuhan user dalam menu tawaran pada sistem.

Tabel 2. Spesifikasi Kebutuhan Pengguna

No	Kebutuhan Pengguna
1	Terdapat menu play dan tentang saat memulai aplikasi
2	Terdapat menu Latihan Menulis Aksara Bali
3	Terdapat menu Pilih Aksara Bali
4	Terdapat menu Hapus
5	Terdapat menu Periksa
6	Terdapat menu back
7	Terdapat menu Ujian
8	Terdapat menu Tentang

2.4.2. Uji Penggunaan Perangkat Lunak

Uji Penggunaan perangkat lunak menggunakan metode *Pre Test-Post test*. *Pre test* dilakukan setelah perangkat lunak selesai dikerjakan. Pengguna langsung diarahkan dalam menu ujian untuk memperoleh nilai *pre-test* [3]. Selanjutnya pengguna diberikan waktu untuk berlatih menggunakan perangkat lunak selama 3 hari. Hari ke-4 dilakukan menu ujian *post-test* terhadap pengguna. Hasil uji *pre-test post test* akan dibandingkan untuk melihat perkembangan masing-masing pengguna [4]. Tabel 3 berisikan informasi skenario *uji pre-test* dan *post test*.

Tabel 3. Uji Kegunaan Perangkat Lunak

No	Pengguna	Skor	
		Pre-test	Post-test
1	1		
2	2		
:	:		
:	:		
:	:		
n	n		

Pada tahapan *pre-test* dan *post-test*, 10 soal akan dikeluarkan oleh perangkat lunak secara random. Skor diperoleh dengan menghitung nilai penulisan aksara yang dinilai benar oleh perangkat lunak nilai skor *pre-test* akan dibandingkan dengan nilai *post-test* masing-masing pengguna.

3. Hasil dan Pembahasan

Pada bagian ini ditampilkan hasil tampilan implementasi aplikasi, hasil uji *black box*, dan hasil uji penggunaan perangkat lunak.

3.1. Hasil Tampilan Implementasi Perangkat Lunak



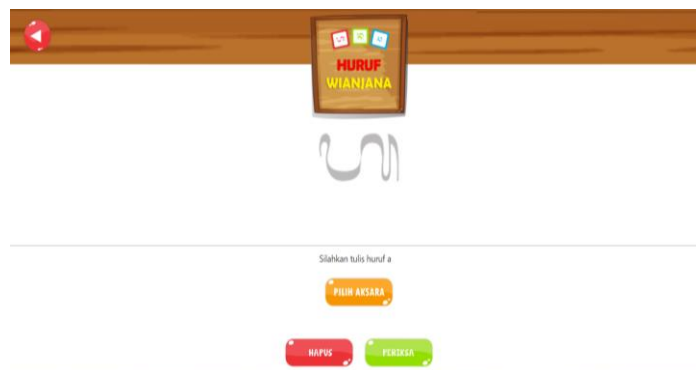
Gambar 4. Halaman Utama

Pada gambar 4 terdapat judul serta latar aplikasi yang dibuat berwarna agar memberikan kesan menarik pada permainan saat diakses. Selain itu terdapat 2 tombol yaitu tombol play dan tentang yang memiliki fungsi seperti yang sudah dijelaskan pada bagian sebelumnya dirancangan desain aplikasi.



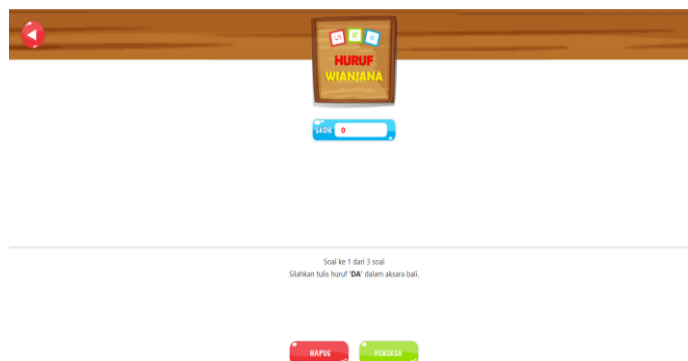
Gambar 5. Halaman Menu Utama

Pada gambar 5 terdapat 3 tombol yaitu tombol latihan menulis, ujian, dan tentang. Tombol Latihan Menulis akan membawa pengguna ke halaman latihan menulis, tombol ujian akan membawa pengguna ke halaman ujian, dan tombol tentang akan menampilkan halaman tentang aplikasi. Selain itu pada saat pengguna telah masuk ke halaman menu utama ini, pengguna bisa mendengar audio dari aplikasi dimana dengan adanya audio atau latar musik pada aplikasi akan membuat aplikasi menjadi lebih menyenangkan untuk digunakan dan diakses.



Gambar 6. Halaman Latihan Menulis

Pada gambar 6 terdapat 4 tombol yaitu pilih aksara, hapus, periksa, dan back. Tombol pilih aksara berfungsi untuk menampilkan aksara apa yang ingin dipelajari atau ditulis oleh pengguna dimana aksaranya terdiri dari 18 huruf yaitu aksara wianjana.



Gambar 7. Halaman Ujian

pada gambar 7 ada 3 tombol yang sama dengan halaman latihan menulis, dan perbedaan terletak pada papan skor untuk menampilkan nilai yang didapat pengguna, perbedaan selanjutnya tidak ada tombol pilih aksara karena pada halaman ujian pengguna akan diberikan 3 soal tulis aksara secara acak sehingga pengguna tidak dapat menentukan sendiri

aksara yang harus di tulis, dan perbedaan yang terakhir adalah tidak ada pola aksara untuk menguji ingatan pengguna saat belajar menulis di halaman latihan menulis. Berikut tampilan saat pengguna melakukan ujian tulis aksara.



Gambar 8. Halaman Tentang

Pada halaman tentang pengguna akan diberikan tampilan seperti pada gambar 8 dimana pada tampilan tersebut berisi informasi aplikasi.

3.2. Hasil Uji Black Box

Tabel 4. Hasil Pengujian Black Box

No	Pengguna	Terdapat menu play	Terdapat menu Latihan Menulis	Terdapat menu Pilih Aksara	Terdapat Menu Hapus	Terdapat menu Periksa	Terdapat menu back	Terdapat menu Ujian	Terdapat menu Tentang	
1	1	1	1	1	1	1	1	1	1	
2	2	1	1	1	1	1	1	1	1	
3	3	1	1	1	1	1	1	1	1	
4	4	1	1	1	1	1	1	1	1	
5	5	1	1	1	1	1	1	1	1	
6	6	1	1	1	1	1	1	1	1	
7	7	1	1	1	1	1	1	1	1	
8	8	1	1	1	1	1	1	1	1	
9	9	1	1	1	1	1	1	1	1	
10	10	1	1	1	1	1	1	1	1	
	Total	10	10	10	10	10	10	10	10	
	Akurasi item spesifikasi	100%	100%	100%	100%	100%	100%	100%	100%	
	Akurasi Total								100%*8/8=	100%

Dari tabel 4 terlihat bahwa semua menu/fungsi tersedia atau tanpa nilai 0 sehingga akurasi menjadi 100%. Hal ini menandakan bahwa perangkat lunak telah memenuhi semua kebutuhan pengguna.

3.3. Hasil Uji Penggunaan Perangkat Lunak

Pada bagian ini akan diberikan perbandingan skor pengguna yang mengalami perubahan menjadi lebih baik setelah menggunakan aplikasi secara maksimal. Penggunaan di uji coba oleh anak-anak sekolah dasar yang sudah menjadi responden.

Tabel 5. Hasil Perbandingan Skor Anak

No	Pengguna	Skor	
		pre-test	post-test
1	1	4	8
2	2	5	9
3	3	4	7
4	4	5	7
5	5	7	10
6	6	5	9
7	7	5	9
8	8	7	9
9	9	6	8
10	10	6	9
rata-rata		5,4	8,5

Dari tabel 5 dapat dilihat skor rata-rata pre-test adalah 5,4 sedangkan skor rata-rata adalah 8,5. Jika dilihat dari jangkauan penilaian SKS perkuliahan maka skor rata-rata pre-test tergolong ke nilai kurang, sedangkan hasil *post-test* tergolong ke dalam nilai sangat baik. Hal ini menandakan adanya peningkatan nilai dari kurang ke sangat baik.

4. Kesimpulan

Perangkat lunak telah dikerjakan dengan pemenuhan kebutuhan pengguna bernilai 100% yang berarti semua fungsi yang dibutuhkan user telah terpenuhi. Penggunaan perangkat lunak dapat memberikan peningkatan nilai dari kurang ke sangat baik, sehingga perangkat lunak yang dikerjakan dapat membantu pembelajaran siswa tentang menulis Aksara Bali

Daftar Pustaka

- [1] A. Z. Arif, "Rancang Bangun Aplikasi IT Support Berbasis Android Menggunakan Metode SDLC (System Development Life Cycle) PT.Adhi Karya (Persero) Tbk., Departemen EPC", *RESTIKOM: Riset Teknik Informatika dan Komputer*, vol. 2, no. 2, 2020.
- [2] T. S. Jaya, "Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis (Studi Kasus: Kantor Digital Politeknik Negeri Lampung)", *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, vol.03, no.02, 2018.
- [3] S. Rahayu, M. Iqbal, dan R. D. A. Budiman, "Efektivitas Media Pembelajaran Matematika Berbasis Web Dan Game Edukasi Terhadap Peningkatan Hasil Belajar Siswa SMP", *Jurnal Pendidikan Informatika dan Sains*, vol. 10(2), p. 177-184, 2021.
- [4] Kamal, Sukirman, N. Alamsyah, dan Sirwan, "Penerapan Aplikasi *Multism* Pada Mata Kuliah Dasar Elektronika", *Jurnal Dinamika Informatika*, vol. 11, no. 1, 2022.
- [5] M. H. D. Adiantika, I. G. A. Wibawa, I. G. N. A. C. Putra, "Game Edukasi Pengenalan Huruf Latin Menggunakan Template Matching Untuk Anak Usia Dini", *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 9, no. 1, 2020.
- [6] D. E. R. Putra dan R. Fidiwati, "Perancangan Aplikasi Berbasis Android Game Edukasi Pengenalan Aksara Batak-Indonesia-Inggris", *Duniailmu.org*, vol. 1(1), 2021.
- [7] F. R. Doni, F. Putriana, A. M. Lukman, dan B. Sudrajat, "Aplikasi Game Edukasi Membaca Huruf "ALFA : LABIRIN" Berbasis Android", *Evolusi : Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022.
- [8] M. S. Dewi, M. W. A. Kesiman, dan I. M. G. Sunarya, "Aplikasi Pembelajaran Pengenal Aksara Bali Menggunakan Metode Template Matching", *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 3, no. 1, 2014.
- [9] N. K. A. Wirdiani, A. A. K. A. Cahyawan W, D. P. Atmaja, "Aplikasi Game Edukasi Pasang Pageh Aksara Bali Berbasis Android", *Scientific Journal Of Informatics*, vol. 2, no. 2, 2015.
- [10] R. H. Kusumodestoni, B. B. Wahono, G. Sudiryanto, dan F. Shobah, "Penerapan Metode Waterfall Pada Aplikasi Pengenalan Huruf Hijaiyah Berbasis Android Pada Paud Nabata", *INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi*, vol. 24, no. 1, 2022.

This page is intentionally left blank

Implementasi *Double Frequency Modulation* dan FFT dalam Sintesis Suara Rindik

I Komang Kumara Saduadnyana^{a1}, I Ketut Gede Suhartana^{a2}, Luh Arida Ayu Rahning Putri^{a3}, I Gusti Ngurah Anom Cahyadi Putra^{a4}, I Komang Ari Mogi^{a5}, Ida Bagus Made Mahendra^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia

¹saduadnyanakumara@gmail.com

²ikg.suhartana@unud.ac.id

³rahningputri@unud.ac.id

⁴anom.cp@unud.ac.id

⁵arimogi@unud.ac.id

⁶ibm.mahendra@unud.ac.id

Abstract

The modern era has changed a lot of people's lives, for example traditional culture is slowly being abandoned which is being replaced by the ease of technology which is growing to affect people's lifestyles. One of them is gamelan culture in the Bali area, namely gamelan rindik. In the past, the rindik gamelan was usually used as an entertainment musical instrument played by village farmers, then for the joged roof performance and later adapted as a welcome musical instrument in hotels. However, along with the development of technology, making this gamelan less desirable because people, especially the younger generation, are more interested in modern musical instruments which are more attractive than traditional gamelan such as the rindik. Because of this, the author is interested in making a digital form of gamelan rindik by synthesizing the sound of gamelan rindik and then presenting it through a media website that can be accessed on any platform. This can be an introduction to the younger generation about gamelan rindik with a more attractive visual. Synthesizing requires a sound processing technique with a certain method commonly used, namely Double Frequency Modulation. However, before doing the synthesis, another method is needed to obtain the features that will be used during the synthesis, one of which is the frequency feature of the sound produced by the rindik gamelan. To get the frequency using the Fast Fourier Transform method. The research was conducted by analyzing 11 blades on the rindik and then the synthesis process was carried out. From the test results, it was found that the method used was successful in synthesizing using three datasets with each accuracy for dataset 1 having an accuracy of 36%, for dataset 2 having an accuracy of 100% and dataset 3 having an accuracy of 45%. This result was obtained by recording the sound of the rindik gamelan on its 11 blades.

Keywords: DFM, Time Domain, Frequency Domain, Rindik, Fundamental Frequency

1. Pendahuluan

Musik tradisional adalah musik yang hadir dan hidup di lingkungan masyarakat secara turun temurun dan kemudian dipertahankan sebagai sarana hiburan. Namun seiring berkembangnya zaman, musik tradisional sulit mendapatkan tempatnya di hati para pendengarnya serta jarang dinikmati generasi muda yang dipengaruhi oleh pengaruh asing[1]. Selain zaman dan teknologi yang mempengaruhi turunnya minat dengan musik tradisional juga disebabkan karena kurangnya informasi mengenai budaya seni musik tradisional yang kita miliki dan menganggap musik tradisional sudah ketinggalan zaman[1]. Generasi muda yang seharusnya melestarikan musik tradisional sangat jarang memiliki ketertarikan jika dibandingkan dengan musik barat dengan tampilan yang lebih menarik dari

segi komposisi dan penvaiiannya. Beberapa penelitian yang membahas mengenai pelestarian musik tradisional. Salah satu gamelan yang termasuk dalam musik tradisional yakni gamelan rindik. Gamelan rindik merupakan gamelan yang berasal dari Bali yang populer. Gamelan rindik terbuat dari bambu dengan jumlah bilah yang dimiliki gamelan rindik sebanyak 11 bilah. Namun generasi muda tidak memiliki minat untuk sekedar memainkan atau mempelajari gamelan rindik yang dikarenakan gaya hidup yang memberikan dampak signifikan. Untuk meningkatkan minat dengan gamelan tradisional Bali khususnya rindik, diperlukan sebuah digitalisasi gamelan rindik yang dituangkan dalam media elektronik[2].

Beberapa penelitian yang pernah dilakukan terkait penggunaan teknik digitalisasi suara gamelan yaitu dengan menggunakan sintesis suara serta menggunakannya sebagai kajian guna menjadi sebuah acuan dalam melakukan penelitian ini. Dalam kajian yang dilakukan ditemukan beberapa artikel yang memiliki keterkaitan dengan penelitian ini. Salah satu penelitian yang dilakukan tahun 2021 terkait sintesis suara gamelan gerantang dengan memanfaatkan metode DFM (*Double Frequency Modulation*). Penelitian ini didasarkan pada pengukuran suara pada saat pembuatan gamelan masih menggunakan cara konvensional dengan menggunakan indera pendengaran. Untuk memudahkan hal itu, maka dilakukan sintesis suara yang nantinya frekuensi yang dihasilkan dapat digunakan sebagai acuan dalam pembuatan gamelan gerantang kedepannya[3]. Penelitian lainnya yang dilakukan pada tahun 2016, di mana sintesis suara dilakukan pada gamelan rindik dengan metode Mod FM yaitu metode yang merupakan pembaharuan dari metode FM (*Frequency Modulation*). Penelitian ini dilakukan untuk mendigitalkan kebudayaan Bali yang salah satunya menjadi sebuah ikon yaitu gamelan rindik[2]. Penggunaan metode FM untuk melakukan sintesis pernah dilakukan pada gamelan Terompong dengan latar belakang dilakukannya penelitian di mana objek yang diteliti memiliki dimensi yang cukup besar sehingga untuk memainkannya harus berada di suatu tempat. Sehingga dibuatkan sebuah bentuk digitalisasi agar mudah memainkannya[4]. Selain itu, beberapa penelitian yang menggunakan DFM mengenai MIMO Radar dengan usulan sebuah frekuensi yang termodulasi ganda dengan bentuk gelombang modulasi untuk radar *multiple-input multiple-output*[5]. Penerapan metode DFM juga pernah dilakukan dengan mengangkat permasalahan *reflectometry* domain dikorelasikan dengan optik Brillouin menggunakan DFM dan modulasi fase[6].

Dalam melakukan digitalisasi gamelan rindik diperlukan ekstraksi fitur dari audio untuk mendapatkan frekuensi yang akan memudahkan dalam tahap analisis sebelum memasuki proses sintesis. Untuk ekstraksi fitur pada sebuah audio menggunakan metode Fast Fourier Transform (FFT) dengan melakukan sebuah transformasi sinyal menjadi sinyal frekuensi[7]. Proses transformasi dilakukan untuk mengubah sinyal analog pada audio menjadi sinyal digital untuk memudahkan proses analisis. Proses ekstraksi fitur pada sebuah audio menghasilkan sebuah bilangan kompleks yang kemudian dari bilangan kompleks tersebut akan didapatkan frekuensi yang digunakan dalam proses sintesis suara gamelan rindik. Untuk meninjau keselarasan daripada hasil sintesis dengan audio original dengan membandingkan frekuensi yang dihasilkan dari proses sintesis. Hasil akhir dari penelitian ini berupa sistem yang dikembangkan dengan metode prototype. Dengan menggunakan metode ini akan mempercepat proses daripada pengembangan sistem[8].

2. Metode Penelitian

2.1. Dataset

Pengumpulan data yang digunakan untuk penelitian ini menggunakan software Audacity sebagai alat perekam untuk merekam suara gamelan rindik. Suara rindik yang diambil sebanyak 1 ketuk di setiap bilahnya dengan durasi ± 1 detik dengan format audio yang digunakan adalah .wav. Pada penelitian ini, digunakan 3 dataset untuk mencari frekuensi minimum dan maksimum dari masing – masing bilah. Tiga dataset tersebut didapatkan dari 3 alat yang berbeda dengan dua daerah yang berbeda. Untuk dataset 1 dan 2 didapatkan di daerah desa Pemogan, Denpasar Selatan dan dataset 3 didapatkan di daerah Bangli, desa Abuan. Sehingga jumlah keseluruhan data berjumlah 33 audio. Data tersebut akan diproses dengan metode FFT untuk mendapatkan fitur berupa frekuensi dengan mengambil nilai absolut dari bilangan imajiner yang dihasilkan.

2.2. Analisis Suara

FFT (*Fast Fourier Transform*) merupakan sebuah metode yang biasanya digunakan dalam pemrosesan suara dengan mengubah bentuk domain dari domain waktu (*time domain*) menjadi domain frekuensi (*frequency domain*)[9]. Perubahan domain diperlukan untuk memudahkan peneliti dalam menganalisis

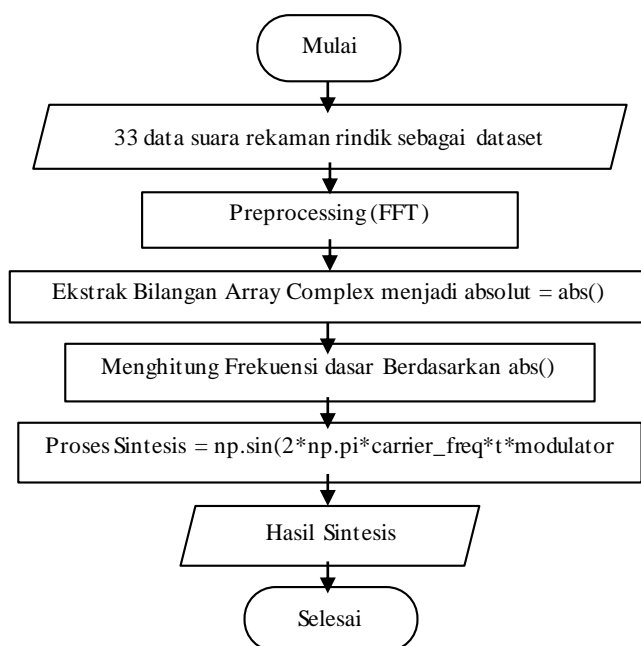
frekuensi dasar yang terdapat pada gamelan rindik berdasarkan amplitude yang tertinggi. Pencarian frekuensi dasar dengan FFT berdasarkan persamaan (1) sebagai berikut:

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \quad (1)$$

Keterangan:

$S(f)$ = Sinyal dalam domain frekuensi
 $s(t)$ = Sinyal dalam domain waktu
 $s(t)e^{-j2\pi ft}$ = konstanta sinyal
 f = frekuensi
 t = waktu

Setelah mendapatkan fitur audio dengan FFT, kemudian dilakukan proses sintesis. Sebelum memasuki proses sintesis ini, yang menjadi input daripada sistem yakni frekuensi dasar, bungkus sinyal (*envelope*) dan penghalusan sinyal. Selain itu juga terdapat perbandingan ratio untuk frekuensi pemodulasi[10] sebagai inputan dengan susunan fm1:fm2:fc di mana fm1 merupakan frekuensi pemodulasi pertama, pemodulasi kedua dan frekuensi pembawa. Setelah semua proses dilakukan, maka dapat dilakukan pengujian sistem untuk proses sintesis yang dihasilkan. Untuk proses sintesis dengan DFM dapat dilihat pada gambar 1.



Gambar 1. Flowchart Proses Sintesis dengan DFM

Pada gambar 1, proses sintesis dimulai dengan menginputkan 33 dataset suara rindik. Kemudian proses selanjutnya adalah mencari fitur dari audio input menggunakan FFT. Dari hasil FFT dilakukan ekstrak bilangan array dengan mengambil nilai absolutnya dan bilangan tersebut diproses kembali untuk mendapatkan frekuensi dasar dari setiap bilah. Setelah mendapatkan fitur audio berupa frekuensi dasar, kemudian dilakukan proses sintesis menggunakan formula pada metode DFM. Dari proses tersebut kemudian menghasilkan hasil suara sintesis setiap bilah pada rindik.

3. Hasil dan Diskusi

Hasil dari penelitian ini adalah nilai frekuensi dari setiap bilah pada rindik. Penelitian ini menggunakan tiga dataset yang kemudian hasil dari sintesis sistem akan dibandingkan dengan frekuensi sebelum sintesis apakah frekuensi yang dihasilkan dari proses sintesis berada dalam rentang nilai minimal atau maksimal dari nilai frekuensi setiap bilah.

3.1. Hasil Analisis Suara

Tahap proses analisis suara dilakukan untuk mendapatkan fitur dari sebuah audio. Proses analisis suara dilakukan sebelum melakukan pengembangan sistem yang nantinya akan menggunakan fitur suara dalam menguji keselarasan. Proses ini melibatkan FFT yang menghasilkan bilangan kompleks yang kemudian diambil nilai absolutnya untuk mendapatkan nilai *real* dari frekuensi. Karena hasil dari FFT adalah bilangan array yang terdiri dari bilangan absolut dan imajiner. Sehingga bilangan absolut ini diperlukan untuk mendapatkan nilai frekuensi yang sebenarnya. Berikut tabel hasil ekstraksi fitur dari setiap bilah.

Tabel 1. Hasil Ekstraksi Fitur

Dataset	Frekuensi Bilah										
	1	2	3	4	5	6	7	8	9	10	11
1	151,4	171	210,7	237,1	265,2	315,7	342,7	418	467,8	525,9	628,7
2	165,8	190,9	219,8	251,5	286,7	336,2	367,5	439,7	501,7	554,1	665
3	180,8	204,5	241,5	217,5	309,9	365,6	406,7	487,2	540,7	606,6	725,9
Max	180,8	204,5	241,5	251,5	309,9	365,5	406,7	487,2	540,7	606,6	725,9
Min	151,4	171	210,7	217,5	265,2	315,7	342,7	418	467,8	525,9	628,7

Pada tabel 1, merupakan hasil proses ekstraksi fitur setiap bilangannya berdasarkan bilangan kompleks yang diekstrak untuk mendapatkan nilai absolut. Berdasarkan tabel untuk masing – masing bilah terdapat frekuensi minimal dan maksimal yang menjadi acuan untuk menguji keselarasan hasil dari sintesis. Jika hasil dari sintesis per bilah berada pada nilai minimum dan maksimum maka hasil sintesis dapat dikatakan selaras dan jika tidak berada dalam rentang tersebut maka dapat dikatakan tidak selaras.

3.2. Sintesis Suara dengan DFM

Pada tahap ini merupakan tahapan yang dilakukan untuk menyintesis suara gamelan rindik dengan metode DFM. Untuk input dari metode ini terdapat 3 buah inputan yakni frekuensi dasar, sinyal pembungkus dan perbandingan sinyal pemodulasi antara fm1:fm2:fc. Proses sintesis ini dilakukan menggunakan sistem yang dikembangkan dengan menggunakan bahasa pemrograman python dengan framework Streamlit. Berikut beberapa fitur yang terdapat pada sistem yang dibangun.

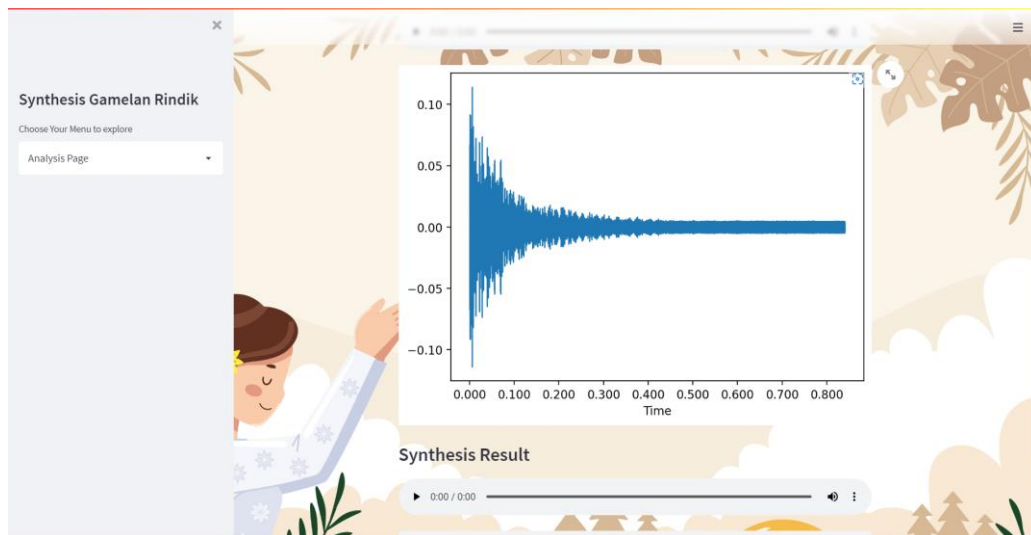
1. Menu Home



Gambar 2. Menu Home Page

Pada gambar 2, menampilkan tampilan awal dari menu home pada sistem di mana menu home memiliki tampilan berupa pengenalan mengenai gamelan rindik dan informasi seputar gamelan rindik. Bagaimana sejarahnya, bentuk dari gamelan rindik dan tampilan daripada gamelan rindik itu sendiri. Pada menu ini terdapat sebuah gambar dan deskripsi terkait gamelan rindik serta pada sebelah kiri terdapat sidebar untuk memilih menu. Terdapat 3 pilihan menu yakni home page, menu analysis page dan play sound rindik. Menu home akan muncul ketika sistem pertama kali diakses dengan alamat localhost:8501. Sistem ini masih bersifat prototype dengan mengakses sistem secara lokal.

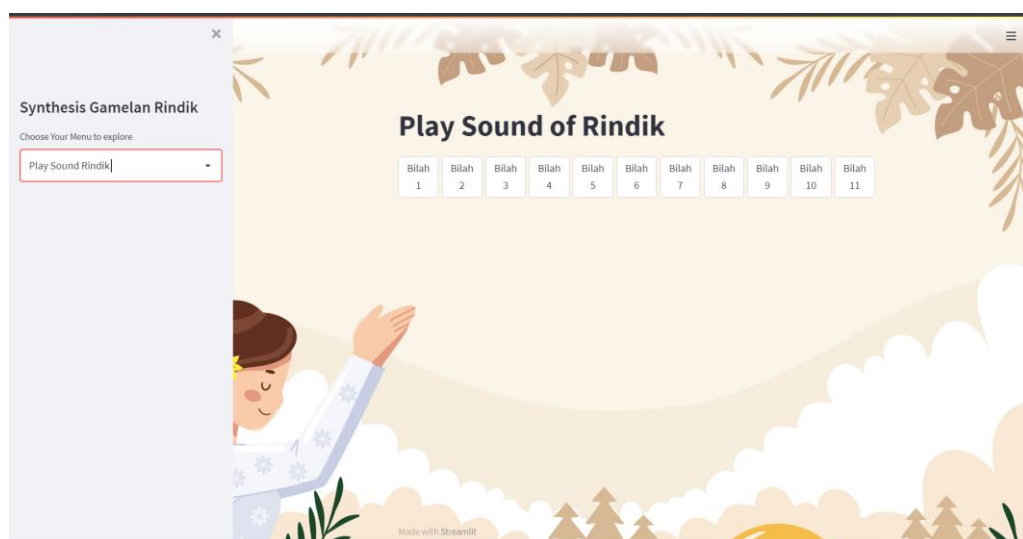
2. Menu Analysis



Gambar 3. Menu Analysis

Pada gambar 3, merupakan tampilan dari Menu Analysis yang menampilkan hasil analisis dan hasil plot domain waktu dari hasil sintesis maupun audio original yang diinputkan dan dapat memutar suara audio original dengan suara yang telah disintesis serta menampilkan apakah hasil dari sintesis selaras atau tidak.

3. Menu Play Rindik Sound



Gambar 4. Menu Play Sound Rindik

Menu akan menampilkan representasi dari gamelan rindik melalui sebuah tombol yang memiliki label 1 – 11. Jika tombol ditekan akan mengeluarkan suara gamelan rindik namun suara yang telah dilakukan sintesis.

3.3. Hasil dan Pengujian

Pada bab ini membahas hasil akhir dan pengujian yang dilakukan oleh sistem. Sintesis suara gamelan rindik dengan metode DFM dapat dilakukan berdasarkan hasil analisis frekuensi, semakin tinggi suara yang dihasilkan, maka frekuensi yang dihasilkan juga semakin tinggi sesuai dengan bentuk dan ukuran daripada gamelan rindik. Pengujian dilakukan dengan menguji frekuensi hasil sintesis apakah terdapat dalam rentang maksimum dan minimum di setiap masing – masing bilah. Berikut tabel untuk hasil pengujian yang dilakukan.

Tabel 2. Hasil Pengujian Dataset 1

Hasil Frekuensi Sebelum dan Sesudah Sintesis pada Dataset 1			
Bilah	Sebelum Sintesis	Setelah Sintesis	Keterangan
1	Frequency: 151 Hz	Frequency: 151 Hz	Tidak Selaras
2	Frequency: 170 Hz	Frequency: 169 Hz	Tidak Selaras
3	Frequency: 209 Hz	Frequency: 208 Hz	Tidak Selaras
4	Frequency: 238 Hz	Frequency: 237 Hz	Selaras
5	Frequency: 264 Hz	Frequency: 262 Hz	Tidak Selaras
6	Frequency: 313 Hz	Frequency: 319 Hz	Selaras
7	Frequency: 341 Hz	Frequency: 337 Hz	Tidak Selaras
8	Frequency: 416 Hz	Frequency: 419 Hz	Selaras
9	Frequency: 461 Hz	Frequency: 469 Hz	Selaras
10	Frequency: 517 Hz	Frequency: 517 Hz	Tidak Selaras
11	Frequency: 637 Hz	Frequency: 628 Hz	Tidak Selaras
Akurasi			$4/11 * 100\% = 36\%$

Berdasarkan hasil tabel 2, dapat dilihat bahwa untuk dataset 1, memiliki akurasi 36%. Hal ini dipengaruhi oleh kualitas dari objek rindik yang diteliti. Di mana hal ini berlaku suara yang mati pada bilah. Artinya, tali pengait antar bilah perlu diperbaiki untuk menghasilkan suara kembali dan hal ini mempengaruhi proses pengujian yang dilakukan. Selain karena pengaruh dari objek yang diteliti juga hasil akurasi dapat dipengaruhi dengan tempat/lokasi dilakukannya perekaman suara bilah rindik. Jika dilakukan di tempat terbuka maka kecenderungan untuk mendapatkan *noise* baik dari suara sekitar, suara kendaraan maupun suara lainnya dapat mempengaruhi proses daripada perekaman untuk mendapatkan suara rindik serta hal lain yang dapat mempengaruhi yakni device yang digunakan untuk melakukan perekaman. Dalam hal ini, peneliti menggunakan device dalam keadaan baik dan hanya saja hasil akurasi dipengaruhi oleh keadaan objek yang sudah harus dibenahi kembali.

Tabel 3. Hasil Pengujian Dataset 2

Hasil Frekuensi Sebelum dan Sesudah Sintesis pada Dataset 1			
Bilah	Sebelum Sintesis	Setelah Sintesis	Keterangan
1	Frequency: 165 Hz	Frequency: 163 Hz	Selaras
2	Frequency: 190 Hz	Frequency: 188 Hz	Selaras
3	Frequency: 223 Hz	Frequency: 216 Hz	Selaras
4	Frequency: 251 Hz	Frequency: 251 Hz	Selaras
5	Frequency: 286 Hz	Frequency: 283 Hz	Selaras
6	Frequency: 337 Hz	Frequency: 336 Hz	Selaras
7	Frequency: 368 Hz	Frequency: 373 Hz	Selaras
8	Frequency: 442 Hz	Frequency: 437 Hz	Selaras
9	Frequency: 502 Hz	Frequency: 500 Hz	Selaras
10	Frequency: 558 Hz	Frequency: 551 Hz	Selaras
11	Frequency: 669 Hz	Frequency: 669 Hz	Selaras
Akurasi			$11/11 * 100\% = 100\%$

Berdasarkan tabel 3, hasil sintesis yang dilakukan pada setiap bilah dengan hasil selaras pada hasil setelah sintesis berada pada nilai minimum dan maksimum dari nilai frekuensi setiap bilahnya. Sehingga akurasi menjadi 100% untuk dataset 2. Hasil ini baik dikarenakan rindik yang digunakan memiliki kondisi yang baik, di mana tidak ada tali pengait yang tidak kencang dan memiliki suara dan bungkus sinyal yang baik sehingga mendapatkan akurasi 100%.

Tabel 4. Hasil Pengujian Dataset 3

Hasil Frekuensi Sebelum dan Sesudah Sintesis pada Dataset 1			
Bilah	Sebelum Sintesis	Setelah Sintesis	Keterangan
1	Frequency: 180 Hz	Frequency: 180 Hz	Selaras
2	Frequency: 206 Hz	Frequency: 206 Hz	Selaras
3	Frequency: 242 Hz	Frequency: 240 Hz	Selaras
4	Frequency: 274 Hz	Frequency: 272 Hz	Tidak Selaras
5	Frequency: 309 Hz	Frequency: 311 Hz	Tidak Selaras
6	Frequency: 365 Hz	Frequency: 367 Hz	Selaras
7	Frequency: 408 Hz	Frequency: 409 Hz	Tidak Selaras
8	Frequency: 486 Hz	Frequency: 490 Hz	Tidak Selaras
9	Frequency: 543 Hz	Frequency: 546 Hz	Selaras
10	Frequency: 608 Hz	Frequency: 608 Hz	Tidak Selaras
11	Frequency: 727 Hz	Frequency: 730 Hz	Tidak Selaras
Akurasi			$5/11 * 100\% = 45\%$

Berdasarkan tabel 4, untuk dataset 3 memiliki akurasi 45%. Hal ini dikarenakan keadaan seperti pada rindik untuk dataset 1. Terdapat beberapa bilah rindik yang tidak memiliki suara yang baik serta beberapa bilah menyatu dengan bilah lainnya. Sehingga mempengaruhi akurasi yang dihasilkan. Untuk kasus yang tidak menyatu dikarenakan tali pengait yang sudah tidak bagus dalam menyangga bilah sehingga menghasilkan suara yang berbeda dan juga suara yang dihasilkan tidak baik. Sehingga berpengaruh terhadap hasil sintesis. Selain itu juga terdapat beberapa bilah pada saat diketuk menghasilkan suara mati seperti halnya pada dataset 1. Suara mati ini mengeluarkan suara yang tidak baik, dalam artian bahwa gema yang dihasilkan tidak bagus dan menghasilkan frekuensi dibawah dari nilai minimum. Hal tersebut dapat diatasi dengan mengganti dengan bilah yang baru dan dilakukan penyetulan ulang untuk didapatkan suara gamelan rindik dengan hasil yang terbaik. Dengan begitu maka frekuensi yang dihasilkan akan sesuai dengan nilai maksimum dan minimum berdasarkan proses analisis yang dilakukan dengan metode FFT.

4. Kesimpulan

Dari hasil penelitian yang dilakukan dapat diambil sebuah kesimpulan sebagai berikut:

- Proses sintesis suara gamelan rindik dengan menggunakan metode DFM telah berhasil dilakukan dengan menggunakan input frekuensi pembawa, frekuensi pemodulasi pertama dan sinyal pembungkus.
- Sistem sintesis suara yang dibangun berhasil dikembangkan dengan input berupa audio file dari salah satu bilah yang ingin disintesis.
- Akurasi yang dihasilkan dalam sintesis suara gamelan rindik berdasarkan nilai frekuensi minimum dan maksimum pada proses preprocessing. Akurasi yang didapatkan berdasarkan tiga dataset yang digunakan yaitu dataset 1 memiliki akurasi 36%, dataset 2 memiliki akurasi 100% dan dataset 3 memiliki akurasi 45% dan hasil tersebut berdasarkan suara audio dari objek yang diteliti.

References

- [1] Mas'ud Abid, "Menumbuhkan Minat Generasi Muda Untuk Mempelajari Musik Tradisional," *Pros. Semin. Nas. Pendidik. Progr. Pascasarj. Univ. PGRI Palembang*, vol. 2, 2019.
- [2] I. M. Widiartha, "Rindik Voice Synthesis Using Modified Frequency Modulation As Bali Cultural Preservation Efforts," *Kursor*, vol. 8, no. 3, p. 151, 2017, doi: 10.28961/kursor.v8i3.90.
- [3] M. A. Raharja and I. D. M. B. A. Darmawan, "Sintesis Suara Gamelan Gerantang Bali Menggunakan Metode Double Frequency Modulation (Dfm)," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 4, no. 2, pp. 119–126, 2021, doi: 10.31598/jurnalresistor.v4i2.760.
- [4] I. G. E. W. Pratama and L. A. A. R. Putri, "Terompong Basic Tone Synthesis with Frequency

- Modulation Method,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 3, p. 429, 2021, doi: 10.24843/jlk.2021.v09.i03.p14.
- [5] C. Gao, K. C. Teh, and A. Liu, “Double-Modulated Frequency Modulation Waveforms for MIMO Radar,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 2024–2028, 2016, doi: 10.1109/LGRS.2016.2621740.
- [6] O. Matsuoka, M. Kishi, and K. Hotate, “Brillouin optical correlation domain reflectometry with double frequency modulation and phase modulation,” *23rd Int. Conf. Opt. Fibre Sensors*, vol. 9157, p. 91576G, 2014, doi: 10.1117/12.2059432.
- [7] D. Tiara, “Fast Fourier Transform (FFT) Dalam Transformasi Sinyal Frekuensi Suara Sebagai Upaya Perolehan Average Energy (AE) Musik,” *J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 14, no. 1, pp. 28–35, 2014.
- [8] P. Yoko, R. Adwiya, and W. Nugraha, “Penerapan Metode Prototype dalam Perancangan Aplikasi SIPINJAM Berbasis Website pada Credit Union Canaga Antutn,” *J. Merpati*, vol. 7, no. 3, pp. 212–223, 2019, [Online]. Available: <http://jurnal.univbinainsan.ac.id/index.php/jusim/article/download/331/228>.
- [9] G. A. Ambaye, “Time and Frequency Domain Analysis of Signals: A Review,” *Int. J. Eng. Res. Technol.*, vol. 9, no. 12, pp. 271–276, 2020, [Online]. Available: www.ijert.org.
- [10] D. Pamungkas, S. R. Kurniawan, and B. F. Simamora, “Perbandingan Antara Domain Waktu dan Frekuensi untuk Pengenalan Sinyal EMG,” *J. Rekayasa Elektr.*, vol. 17, no. 1, pp. 36–41, 2021, doi: 10.17529/jre.v17i1.16844.

Sistem Monitoring Tanaman Hidroponik Berbasis Internet of Things menggunakan Restful API

Karel Leo Rivaldo ^{a1}, I Komang Ari Mogi ^{a2}, I Putu Gede Hendra Suputra^{a3}, Ngurah Agus Sanjaya^{a4}, I Dewa Made Bayu Atmaja Darmawan^{a5}, Ida Bagus Gede Dwidasmara^{a6}

^aProgram Studi Informatika, Universitas Udayana
Kuta Selatan, Badung, Bali, Indonesia

¹leokarel12@gmail.com

²arimogi@unud.ac.id

³hendra.suputra@unud.ac.id

⁴agus_sanjaya@unud.ac.id

⁵dewabayu@unud.ac.id

⁶dwidasmara@unud.ac.id

Abstract

Agricultural technology is one of the important things in today's era. one of the Indonesian government programs made is a roadmap with the name Making Indonesia 4.0. in the agricultural roadmap program one of the important technology applications is hydroponic agriculture, in the implementation of hydroponic plants a good form of data communication is needed, therefore in this study the author uses a form of data communication with Restful Architecture API as communication in the hydroponic agriculture monitoring system. In this study, two tests were carried out, namely the test was carried out by testing the system in sending sensor data and the second. System testing in reading and controlling the pH sensor value with parameter intervals of 1 second, 5 seconds, 10 seconds, 20 seconds, and 30 seconds. The results obtained in this study are 30 seconds is the best time in the process of sending data one interval from the microcontroller to the system.

Keywords: Technology, Agriculture, Hydroponic, Monitoring, Data Communication, Restfull API

1. Pendahuluan

Pada saat ini perkembangan teknologi sudah semakin berkembang dan sudah mencapai era revolusi industri 4.0. Kemajuan teknologi di Indonesia sudah disambut baik oleh pemerintah dengan dibuatnya roadmap dengan nama "Making Indonesia 4.0". sektor pertanian merupakan salah satu sektor penting dalam penghasil bahan baku untuk sebuah produk yang siap dipasarkan dalam program roadmap. Sektor pertanian di era saat ini sudah dapat berjalan berdampingan dengan teknologi yang ada, salah satu penerapan teknologi di bidang pertanian adalah sistem monitoring tanaman hidroponik. (1) Tanaman hidroponik merupakan salah satu metode dalam budidaya pertanian menggunakan media air sebagai media utama dalam pengganti tanah. Keuntungan dalam menggunakan metode hidroponik adalah petani tidak memerlukan lahan yang luas dalam pelaksanaannya. (2) Di dalam melakukan pertanian hidroponik factor kondisi suhu dan ph air merupakan hal yang perlu diperhatikan, kondisi suhu dan ph air akan mempengaruhi tingkat keberhasilan dari tanaman hidroponik. Dalam melakukan pelaksanaan pertanian hidroponik diperlukan suatu teknologi yang bisa melakukan monitoring suhu dan ph air untuk mendapatkan hasil yang baik dalam penerapannya, oleh karena itu sistem monitoring berbasis IoT dapat membantu petani dalam meningkatkan keberhasilan dari tanaman hidroponik tersendiri.(3)

Dalam penerapan sistem monitoring tanaman hidroponik berbasis Internet of things diperlukan suatu komunikasi data antara mikrokontroler dengan sistem. Komunikasi data merupakan suatu proses pengiriman paket data dari sumber ke tujuan, komunikasi data dapat berjalan dengan baik jika mengacu pada aturan yang telah dibuat. (4) Salah satu bentuk arsitektur komunikasi data adalah Restful api. Restful api merupakan suatu arsitektur komunikasi data yang menggunakan protokol http untuk melakukan pertukaran data dalam mengembangkan suatu aplikasi. Konsep kerja dari Restful API adalah Rest client akan melakukan request data pada sumber daya yaitu Rest server, dan tugas dari Rest server akan menyediakan data yang di request oleh Rest client, pada tahapan ini setiap data

sumber daya dibedakan melalui sebuah global ID atau universal Resource Identifier. Dengan melihat cara kerja dari arsitektur Restful Api peneliti dapat mengembangkan sebuah sistem monitoring tanaman hidroponik dengan dua buah basis platform yaitu website dan mobile. Dari sistem tersebut pengguna dapat memonitoring kondisi suhu dan ph air secara real time. Pada penelitian ini peneliti akan menguji efektivitas dari arsitektur Restful api dalam melakukan pengiriman data nilai sensor. Pengujian dilakukan dengan menentukan waktu optimal bagi arsitektur Restful api untuk mengirim data dari Rest sever ke Rest client.

2. Metodologi Penelitian

Pada penelitian ini peneliti menggunakan Model prototyping yang memiliki definisi model perancangan sistem dengan metode siklus hidup yang direpresentasikan dalam konsep model bekerja. Dalam pengimplementasian Model prototyping akan di jelaskan pada sub bahasan berikut.

2.1. Pengumpulan Data kebutuhan

Pengumpulan data kebutuhan pada penelitian ini berfokus pada pengembangan dan pendefinisian objek dalam perangkat lunak dan perangkat keras untuk mengidentifikasi masalah di dalamnya. Dalam sistem monitoring hidroponik memiliki dua buah tahapan yaitu pertama data diperoleh dengan menginstruksikan Mikrokontroler untuk mengambil data di setiap sensor Sensor suhu dan pH sebagai pengukur suhu dan pH air, data tersebut akan dikirimkan ke dalam server dengan menggunakan jaringan local, pada proses ini data akan dikirim dalam format JSON. Tahapan kedua adalah data dari nilai suhu dan ph air tanaman hidroponik yang di representasikan melalui Gambar 1.

TABEL PH DAN SUHU UNTUK SAYURAN BUAH		
NAMA HERB	PH	SUHU
Cabe	6.0 – 6.5	23-26°C
Kacang Polong	6.0 – 7.0	
Okra	6.5	
Tomat	6.0 – 6.5	
Terong	6.0	
Timun	5.5	
Timun Jepang	6.0	

Gambar 1. Nilai suhu dan pH Tanaman Hidroponik

Sumber : www.hidroponikpedia.com

2.2. Membangun Prototyping

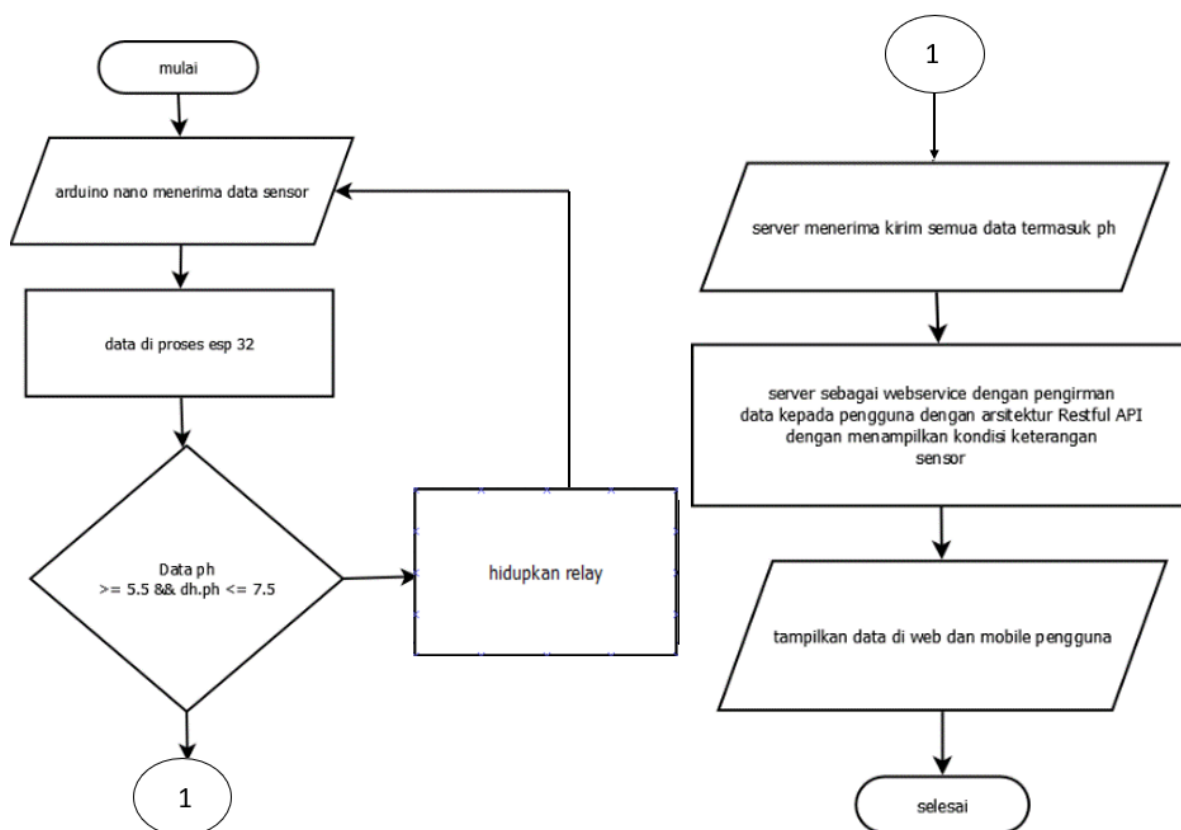
Pada tahapan ini merupakan tahapan perancangan sistem sementara yang pada penyajiannya berfokus pada pelanggan. Proses ini meliputi pada perancangan flowchart dan perancangan fitur, berikut adalah bagian dari tahapan prototyping :

1. Perancangan Flowchart Sistem

Perancangan flowchart sistem digunakan untuk menggambarkan alur kerja dari sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful api, pada tahapan ini sistem komunikasi data antara mikrokontroler hingga ke sistem menggunakan arsitektur Resful api dengan lima buah metode Get, Post, Put, dan Delete. Berikut gambaran flowchart sistem yang ditunjukkan pada Gambar 2.

2. Perancangan Fitur Sistem

Pada perancangan sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful api, pengguna dapat melihat nilai besaran sensor suhu dan ph air secara real time dengan interval yang telah di tentukan oleh peneliti melalui pengujian dari sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful api pada fitur ini pengguna juga dapat melihat kondisi dari tanaman hidroponik yang didapatkan melalui hasil perhitungan logika fuzzy di dalam sistem.



Gambar 2. Flowchart Sistem

2.3. Evaluasi Prototyping

Dalam tahapan evaluasi prototyping peneliti sebagai pengembang sistem akan dilakukan pembacaan kembali perancangan sistem pada tahapan evaluasi prototyping meliputi proses pengumpulan data dan pembangunan prototyping, hasil dari evaluasi sistem menunjukkan mikrokontroler dapat menjalankan tugasnya dengan baik, dengan pengejlasan sebagai berikut.

- Mikrokontroler mampu mengirimkan data sensor ke dalam sistem.
- Sistem server mampu mengambil data yang di kirimkan oleh mikrokontroler dan melakukan proses algoritma fuzzy dari nilai besaran sensor suhu dan ph air.
- Sistem server mampu memberikan informasi mengenai nilai besaran sensor ke dalam sistem pengguna dengan baik menggunakan bentuk komunikasi data dari arsitektur Resfull api dengan metode Get, Post, Put, dan Delete.

2.4. Mengkodekan Sistem

Pada tahapan ini peneliti akan melakukan pengkodean sistem, dalam tahapan ini pengkodean dilakukan dengan dua buah tahapan utama yaitu :

- Pengkodean terhadap mikrokontroler
Pada tahapan ini peneliti akan melakukan pengkodean terhadap mikrokontroler dengan melakukan kalibrasi terhadap sensor suhu dan sensor ph air yang digunakan dalam penelitian ini, tahapan ini mikrokontroler akan melakukan post data ke dalam sistem untuk mengirimkan data nilai besaran sensor untuk diproses di dalam sistem
- Pengkodean terhadap sistem
Pada tahapan ini peneliti akan melakukan pengkodean terhadap sistem sejumlah tiga buah yaitu sistem server, client web, dan client mobile. Pada sisi server berfungsi untuk mengambil data dari mikrokontroler dan berisi mengenai informasi tanaman hidroponik, pada sisi server memiliki fitur lihat data dari nilai besaran sensor dengan bentuk visualisai sejumlah dua buah yaitu grafik dan table, pada sisi server data dari mikrokontroler akan diubah ke dalam format JSON dan akan diproses menggunakan algoritma Fuzzy untuk menentukan kondisi tanaman

hidroponik agar tersampaikan kepada pengguna. Kedua adalah sisi client web dimana dalam sisi client web berisi mengenai nilai dari sensor yang di kirimkan oleh server dengan jumlah data 10 buah data terbaru, data tersebut akan di tampilkan dan bentuk visualisasi sejumlah dua buah yaitu grafik dan table dari besaran nilai sensor. Ketiga adalah sisi client mobile dimana dalam tahapan ini berisi mengenai nilai besaran sensor terbaru sejumlah satu buah. Dalam sisi client mobile data akan ditampilkan dalam bentuk angka nilai sensor suhu dan ph air. Dalam tahapan ini menggunakan beberapa metode dari arsitektur Restful api yaitu Get, Post, Put, dan Delete untuk membangun komunikais data antara mikrokontroler hingga ke server.

2.5. Menguji Sistem

Pada tahapan pengujian sistem akan dilakukan dengan dua buah tahapan yaitu pertama pengujian sistem dalam mengirimkan data sensor dan yang kedua adalah pengujian sistem dalam membaca dan mengontrol nilai sensor pH dengan interval parameter 1 detik, 5 detik, 10 detik, 20 detik, dan 30 detik.

2.6. Evaluasi Sistem

Pada tahapan ini dilakukan evaluasi mengenai tahapan pengkodean sistem dan pengujian sistem monitoring tanaman hidroponik dengan memberikan suatu visualisai data dalam bentuk tabel untuk mengetahui waktu yang optimal dalam satu buah proses interval pengiriman data sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful API.

2.7. Menggunakan Sistem

Pada tahapan ini perangkat lunak sudah masuk dalam kategori tahap selesai, sehingga perangkat lunak sudah bisa digunakan dengan baik sesuai dengan kondisi waktu yang optimal dalam penerapan sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful API.

3. Hasil dan Pembahasan

3.1. Tampilan Server

Pada halaman server memiliki 6 buah fitur utama yaitu halaman home, sistem, data, tambah produk, edit data produk, dan halaman hapus data.



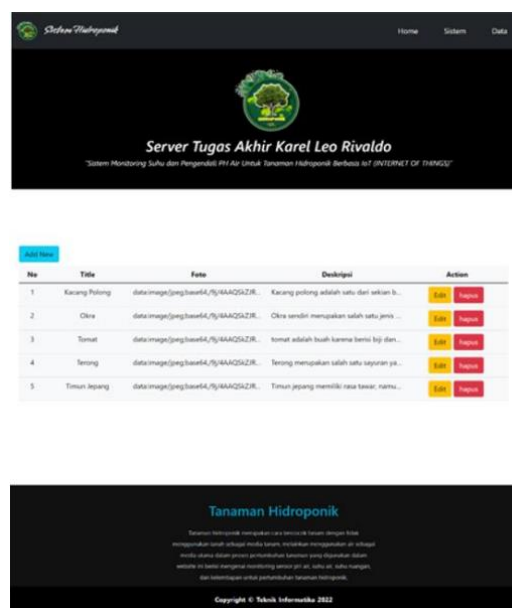
Gambar 3. halaman Home

- Pada halaman home merupakan halaman awal website yang berisi mengenai informasi produk yang akan diimplementasikan ke dalam tanaman hidroponik yang ditunjukkan pada Gambar 3.
- Pada halaman sistem merupakan halaman utama dalam penerapan sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful API, dalam halaman ini berisi mengenai data beserta kondisi sensor pH Air, dan sensor suhu air yang

- didapatkan melalui pembacaan dari mikrokontroler dan data sensor dikirimkan melalui bentuk komunikasi data restful API yang ditunjukkan pada Gambar 4.
- c. Pada halaman data
Pada halaman data berisi mengenai kumpulan data konten informasi produk dalam halaman ini pengguna dapat menambahkan, mengedit, dan menghapus data yang ditunjukkan pada Gambar 5.
 - d. Pada halaman tambah data
Pada halaman tambah data berisi mengenai form untuk menambahkan informasi data produk dengan 3 buah inputan yaitu title, foto dan deskripsi yang ditunjukkan pada Gambar 6.
 - e. Pada halaman edit data produk
Pada halaman edit data produk berisi mengenai form untuk mengubah data informasi produk yang telah di masukan pengguna, dalam halaman edit data menggunakan method patch dalam konsep restfull API yang ditunjukkan pada Gambar 7.
 - f. Pada halaman hapus data
Pada halaman hapus data digunakan untuk menghapus data yang sudah di masukan dalam sistem.



Gambar 4. Halaman Sistem



Gambar 5. Halaman Data

Sistem Hidroponik Home Sistem Data

Server Tugas Akhir Karel Leo Rivaldo
Sistem Monitoring Suhu dan Pengendali PH Air Untuk Tanaman Hidroponik Berbasis IoT (INTERNET OF THINGS)

Title
Masukkan Title

Foto
Choose File No file chosen

Deskripsi
Masukkan Deskripsi

Simpan

Tanaman Hidroponik

Tanaman hidroponik merupakan cara bercocok tanam dengan tidak menggunakan tanah sebagai media tanam, melainkan menggunakan air sebagai media utama dalam proses pertumbuhan tanaman yang dipelihara dalam wadah-wadah dengan menggunakan monitoring sensor pH dan suhu air, suhu ruangan, dan kelembapan untuk pertumbuhan tanaman hidroponik.

Copyright © Teknik Informatika 2022

Gambar 6. Halaman Tambah Data

Sistem Hidroponik Home Sistem Data

Server Tugas Akhir Karel Leo Rivaldo
Sistem Monitoring Suhu dan Pengendali PH Air Untuk Tanaman Hidroponik Berbasis IoT (INTERNET OF THINGS)

Title
Kerang Polong

Foto
Choose File No file chosen

Deskripsi
Kerang polong adalah satu dari sekian banyak jenis kacang-kacangan dengan bentuk folis bulat, berukuran kecil, dan berwarna hijau lusen. Kerang polong

Simpan

Tanaman Hidroponik

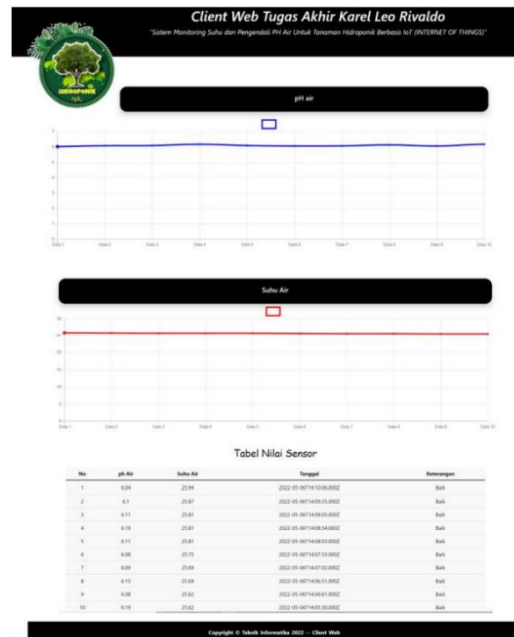
Tanaman hidroponik merupakan cara bercocok tanam dengan tidak menggunakan tanah sebagai media tanam, melainkan menggunakan air sebagai media utama dalam proses pertumbuhan tanaman yang dipelihara dalam wadah-wadah dengan menggunakan monitoring sensor pH dan suhu air, suhu ruangan, dan kelembapan untuk pertumbuhan tanaman hidroponik.

Copyright © Teknik Informatika 2022

Gambar 7. Halaman Edit data

3.2. Tampilan Client Web

Pada tampilan Client Web berisi mengenai data sensor sebanyak 10 buah data sensor ph dan suhu air terbaru yang dikirimkan oleh mikrontroler kepada server. Pada sistem client web menggunakan method Get dalam konsep restfull API yang ditunjukkan pada Gambar 8.



Gambar 8. Halaman Client Web

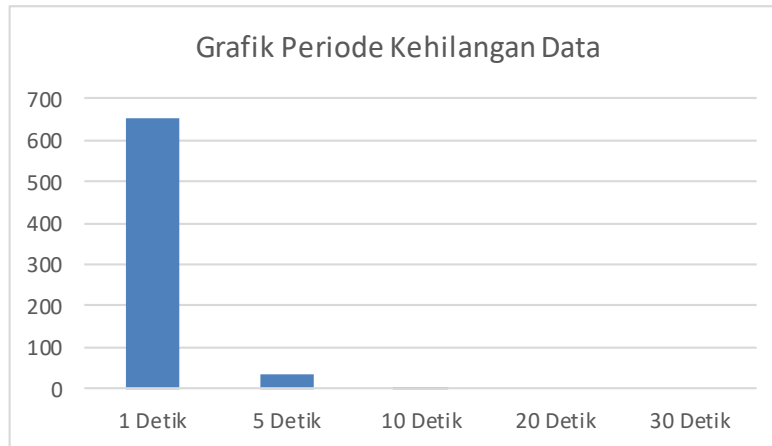
3.3. Tampilan Client Mobile

Pada tampilan client mobile berisi mengenai 1 buah data sensor ph dan suhu air terbaru yang dikirimkan oleh mikrontroler, dalam sisi client mobile menggunakan framework dari react native dan expo sebagai aplikasi penghubung di dalam mobile client. Sisi mobile dapat dijalankan melalui platform sistem operasi android dan IOS, pada sisi client mobile menggunakan method Get dalam konsep restfull API yang ditunjukkan pada Gambar 9.



Gambar 9. Halaman Client Mobile

3.4. Hasil Pengujian pengiriman data sensor

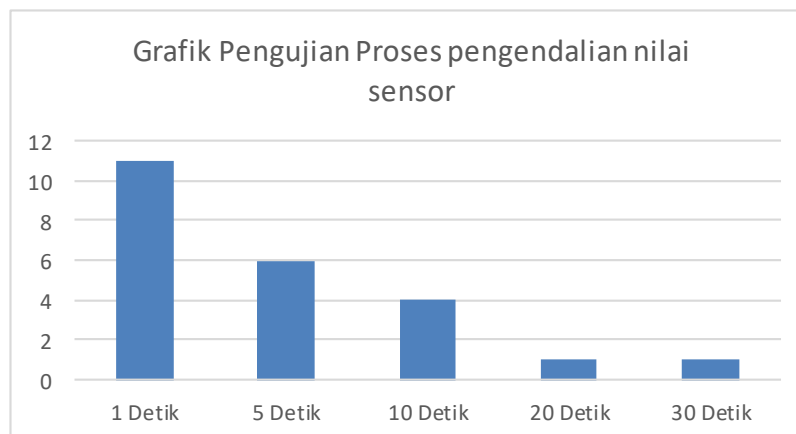


Gambar 10. Pengujian pengiriman data sensor

Hasil pengujian dalam pengiriman data sensor dilakukan pada dengan interval waktu 1, 5, 10, 20, 30 detik dalam satu periode pengambilan data. Hasil yang keseluruhan pengujian sistem dalam pengiriman data sensor yang dapat dilihat pada Gambar 10. Hasil pengiriman terbaik didapatkan pada interval waktu 30 detik per periode, dengan jumlah data yang dihasilkan per periode sebanyak 60 data dengan keterlambatan pengiriman maksimal 6 detik.

3.5. Hasil pengujian pengendalian nilai pH air

Pengujian pengendali nilai pH dilakukan dengan interval waktu pemeriksaan pH air setiap 1, 5, 10, 20, 30 detik dalam satu periode pengambilan data. Mekanisme pengendalian pH air dilakukan dengan penambahan cairan asam atau basa, terhadap air yang akan dikontrol oleh sistem. Pemberian cairan akan dikendalikan sepenuhnya oleh sistem. Intensitas pemberian cairan asam dan basa dapat dilihat pada Gambar 11. Hasil pengujian pengendali nilai pH air menunjukkan semakin kecil interval waktu yang diterapkan, maka semakin tinggi kadar cairan yang akan tercampur dalam air. Hal tersebut membuktikan bahwa Restful API mampu menerima hasil dari setiap pemeriksaan nilai pH air, sehingga sistem dapat melakukan pengendalian terhadap air.



Gambar 11. Hasil pengujian pengendalian nilai pH air

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan sistem sistem monitoring tanaman hidroponik berbasis internet of things menggunakan Restful API berjalan dengan baik dengan waktu optimal pengiriman data yaitu 30 detik dalam satu periode interval. Waktu 30 detik menunjukkan data yang dikirimkan oleh mikrokontroler ke dalam sistem di terima secara utuh sebanyak 60 buah data dan pada pengujian pengendalian nilai sensor waktu 30 detik dalam satu periode interval pengiriman data hanya memerlukan satu buah proses dalam pembacaan nilai sensor agar kondisi tanaman menjadi baik. Tujuan penelitian digunakan untuk meningkatkan hasil dari pertanian tanaman hidroponik yang menjadi masalah bagi petani. Hasil dari penelitian ini diharapkan dapat membantu petani dalam memonitoring sistem pertanian hidroponik guna memperoleh suatu hasil yang baik dan waktu yang optimal di dalam pengimplementasian sistem monitoring suhu dan ph air tanaman hidroponik berbasis internet of things dengan menggunakan arsitektur restfull api.

References

- [1] Alam, R. L., & Nasuha, A. (2020). Alat Pengontrol Ph Air dan Monitoring Lingkungan Tanaman Hidroponik Menggunakan Fuzzy Logic Berbasis Internet Of Things. *Elinvo (Electronics, Informatics, and Vocational Education)*, 5(1), 11–20. <https://doi.org/10.21831/elinvo.v5i1.34587>
- [2] Bayu. (2016, November). Tabel PPM dan pH Nutrisi Hidroponik - HIDROPONIKPEDIA. <http://hidroponikpedia.com/tabel-ppm-dan-phnutrisi-hidroponik/>
- [3] Dani, A. W. (2020). Optimalisasi Pertumbuhan Pada Sayuran Hidroponik Nutrient Film Technique Dengan Metode Fuzzy Logic Berbasis Internet of Things. *Jurnal Teknologi Elektro*, 11(1), 1. <https://doi.org/10.22441/jte.2020.v11i1.001>
- [4] Megawati, D., Masykuroh, K., & Kurnianto, D. (2020). Rancang Bangun Sistem Monitoring PH dan Suhu Air pada Akuaponik Berbasis Internet of Thing (IoT). *TELKA - Telekomunikasi Elektronika Komputasi Dan Kontrol*, 6(2), 124–137. <https://doi.org/10.15575/telka.v6n2.124-137>
- [5] Mulasari, S. A. (2019). Penerapan Teknologi Tepat Guna (Penanam Hidroponik Menggunakan Media Tanam) Bagi Masyarakat Sosrowijayan Yogyakarta. *Jurnal Pemberdayaan: Publikasi Hasil Pengabdian Kepada Masyarakat*, 2(3), 425. <https://doi.org/10.12928/jp.v2i3.418>
- [6] Prayitno. (2017). Sistem Monitoring Suhu, Kelembaban, dan Pengendali Penyiraman Tanaman Hidroponik menggunakan Blynk Android. *Circulation Research*, 1(10), 292–297. <https://doi.org/10.1161/CIRCRESAHA.112.270033>

This page is intentionally left blank

Penerapan Metode Adaptive Neuro Fuzzy Inference System (ANFIS) Dengan *Membership function* Tipe *Gaussian* dan *Generalized Bell* Dalam Prediksi Harga Tertinggi Saham

I Putu Sedana Wijaya^{a1}, Made Agung Raharja^{a2}, Luh Arida Ayu Rahning Putri^{a3}, I Putu Gede Hendra Suputra^{a4}, Ida Bagus Made Mahendra^{a5}, I Gede Santi Astawa^{a6}

^aInformatics Department, Faculty of Math and Sciences, Udayana University
Jalan Raya Kampus Unud, Jimbaran, Bali, 80361, Indonesia

¹sedanawjy@gmail.com

²made.agung@unud.ac.id

³rahningputri@unud.ac.id

⁴hendra.suputra@unud.ac.id

⁵bm.mahendra@unud.ac.id

⁶santi.astawa@unud.ac.id

Abstract

Many people who have capital are currently buying up shares in the stock market in the hope that the stock price will rise when the Covid-19 pandemic ends. As someone who wants to try investing in the stock market, you must be able to estimate the profits and losses from buying shares. One way that can help consideration in making decisions to buy and sell shares is to make predictions. There are many algorithms that can be used in prediction, one of which is the Adaptive Neuro Fuzzy Inference System (ANFIS) method which is a combination of the Fuzzy Logic algorithm and Artificial Neural Networks. The application of the ANFIS method requires a good ANFIS structure by selecting the right number and type of membership functions. In this study the Gaussian and Gbell type membership functions are used because they have the advantage of allowing subtle changes and can accommodate inaccuracies in measurements so that they match the pattern of historical data that moves smoothly at one time. In this study, it was found that the gaussian type has better accuracy than the gbell type by 97.87% to predict the highest stock price of Tencent Holdings Limited and the gbell type has a better accuracy than the gaussian type of 97.8% to predict the highest price of Take-Two shares. interactive.

Keywords: Prediction, ANFIS, Membership function, Stock, Data.

Abstrak

Banyak kalangan yang memiliki modal saat ini beramai-ramai memborong saham di *stock market* dengan harapan harga saham tersebut akan naik saat pandemi Covid-19 berakhir. Sebagai seseorang yang ingin mencoba berinvestasi di *stock market* harus mampu memperkirakan untung dan rugi dari pembelian saham. Salah satu cara yang dapat membantu pertimbangan dalam pengambilan keputusan membeli dan menjual saham adalah melakukan prediksi. Terdapat banyak algoritma yang dapat digunakan dalam prediksi salah satunya adalah metode *Adaptive Neuro Fuzzy Inference System* (ANFIS). Penaplikasian metode ANFIS memerlukan struktur ANFIS yang baik dengan pemilihan jumlah dan tipe *membership function* yang tepat. Pada penelitian ini *membership function* tipe *gaussian* dan *gbell* digunakan karena memiliki kelebihan yaitu memungkinkan perubahan halus dan dapat mengakomodasi ketidaktepatan dalam pengukuran sehingga cocok dengan pola data histori yang bergerak secara halus di satu waktu. Pada penelitian ini diperoleh bahwa tipe *gaussian* memiliki akurasi yang lebih baik dibandingkan tipe *gbell* sebesar 97.87% untuk memprediksi harga tertinggi saham Tencent Holdings Limited dan tipe *gbell* memiliki akurasi yang lebih baik dibandingkan tipe *gaussian* sebesar 97,8% untuk memprediksi harga tertinggi saham Take-Two Interactive.

Kata kunci: Prediksi, ANFIS, *Membership function*, Saham, Data.

1. Pendahuluan

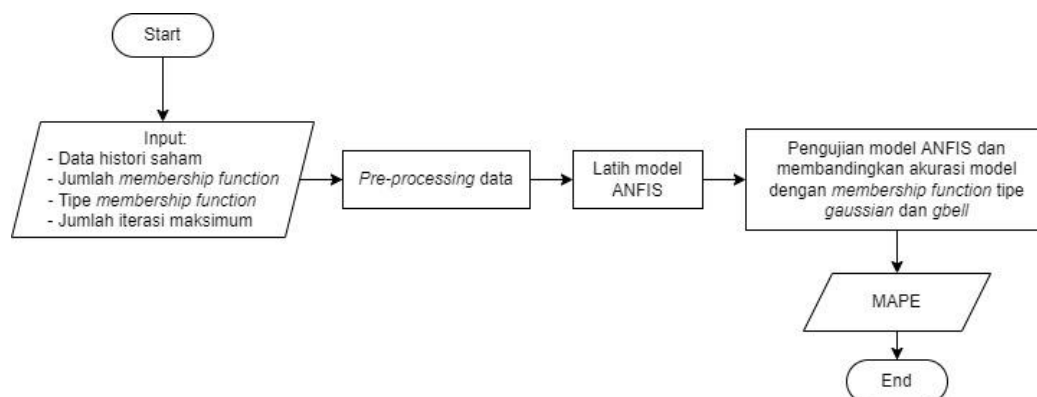
Saham merupakan suatu tanda penyertaan kepemilikan seseorang atau badan dalam suatu perusahaan terbatas [1]. Banyak kalangan yang memiliki modal saat ini beramai-ramai memborong saham di *stock market* dengan harapan harga saham tersebut akan naik saat pandemi Covid-19 berakhir. Salah satu bidang bisnis yang harga sahamnya diharapkan akan terus meningkat saat pandemi adalah bisnis *video games*. Sebagai seseorang yang ingin mencoba berinvestasi harus mampu memperkirakan untung dan rugi dari pembelian saham, harus siap dengan resiko yang terjadi akibat fluktuatifnya pasar saham. Salah satu cara yang dapat membantu pertimbangan dalam pengambilan keputusan membeli dan menjual saham adalah melakukan prediksi.

Menurut [1], prediksi adalah proses untuk memperkirakan berapa kebutuhan di masa yang akan datang yang meliputi kebutuhan dalam ukuran kuantitas, kualitas, waktu dan lokasi yang dibutuhkan dalam rangka memenuhi permintaan barang ataupun jasa. Terdapat banyak metode yang dapat digunakan dalam prediksi salah satunya adalah metode *Adaptive Neuro Fuzzy Inference System* (ANFIS) yang merupakan penggabungan dari metode Logika Fuzzy dan Jaringan Syaraf Tiruan [2]. ANFIS merupakan metode *hybrid* yang berdasarkan menutupi kelemahan dari masing-masing metode tersebut. ANFIS (*Adaptive Neuro-Fuzzy Inference System*) dapat memberikan solusi untuk model sistem prediktif yang mampu beradaptasi dengan baik untuk menangani sistem yang kompleks, nonlinier dan berubah waktu melalui metode pembelajaran terhadap data numerik sistem [3].

Metode ANFIS memerlukan struktur yang baik berdasarkan pemilihan jumlah dan tipe *membership function* yang tepat. *Membership function* merupakan fungsi yang memetakan elemen suatu himpunan ke nilai keanggotaan fuzzy pada interval 0 hingga 1. Pemilihan tipe *membership function* bergantung pada pola data yang digunakan, sehingga dapat mempengaruhi akurasi model ANFIS yang dibangun. Pada penelitian ini digunakan *membership function* dengan tipe *gaussian* dan *generalized bell (gbell)*. Kedua tipe ini memiliki kelebihan yaitu memungkinkan perubahan halus dan dapat mengakomodasi ketidaktepatan dalam pengukuran sehingga cocok dengan pola data historis yang bergerak secara halus di satu waktu [4]. Oleh itu kedua tipe *membership function* ini digunakan pada metode ANFIS untuk melakukan prediksi harga tertinggi saham. Adapun beberapa batasan lainnya pada penelitian ini adalah jumlah iterasi maksimum sebesar 5000, toleransi *error* sebesar 10^{-5} , dan data historis saham yang digunakan adalah data perusahaan yang bergerak di bidang *video games* yaitu Tencent Holdings Limited dan Take-Two Interactive Software, Inc.

2. Metode Penelitian

Pada penelitian ini dilakukan prediksi harga tertinggi saham pada hari berikutnya berdasarkan pola data yaitu harga pembukaan saham, harga terendah saham, dan harga penutupan saham di masa lampau dengan proses *pre-processing* data seperti normalisasi terhadap data, membagi data menjadi 2 yaitu data latih dan data uji, dan pemilihan *setting* parameter seperti jumlah *membership function*, tipe *membership function*, serta jumlah iterasi yang digunakan. Tahapan berikutnya adalah melatih model ANFIS dengan menggunakan metode ANFIS yang dibangun dengan bantuan *library* Tensorflow pada bahasa pemrograman Python hingga toleransi *error Mean Squared Error* (MSE) berada di bawah 10^{-5} atau iterasi telah maksimum. Kurva dari kedua *membership function* akan dicetak setelah proses metode ANFIS selesai. Tahap terakhir adalah tahap pengujian untuk mengukur kinerja model ANFIS yang dibangun menggunakan metode uji *Mean Absolute Percentage Error* (MAPE) dan hasil pengujian akan dicetak pada sistem. Proses pada penelitian ini terdapat pada Gambar 1.



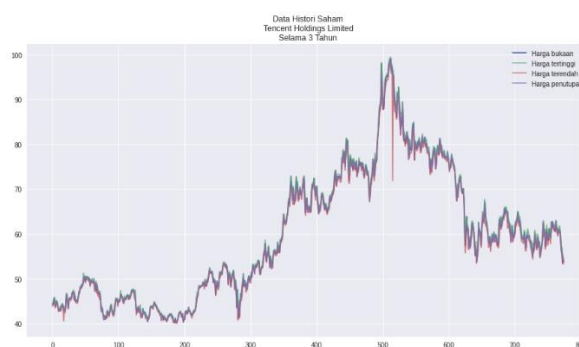
Gambar 1. Flowchart Penelitian ANFIS

2.1. Data Penelitian

Dari situs Yahoo Finance, peneliti memperoleh data histori saham Tencent Holdings Limited dan Take-Two Interactive Software, Inc dengan rentang waktu selama 3 tahun dari tanggal 1 Februari 2019 hingga 1 Maret 2022 sebanyak 745 dengan tren grafik terlihat pada Gambar 2 dan 3. Terdapat 7 variabel dalam dataset ini dengan masing-masing penjelasan terdapat pada Tabel 1.

Tabel 1. Parameter Data Penelitian

Parameter	Deskripsi
<i>Date</i>	Tanggal transaksi saham (per hari)
<i>Open</i>	Harga pembukaan saham di hari tersebut
<i>High</i>	Harga tertinggi saham di hari tersebut
<i>Low</i>	Harga terendah saham di hari tersebut
<i>Close</i>	Harga penutupan saham di hari tersebut
<i>Volume</i>	Jumlah saham yang diperdagangkan di hari tersebut
<i>Adjusted Close</i>	Harga penutupan saham yang disesuaikan di hari tersebut



Gambar 2. Grafik Data Saham Tencent Holdings Limited.



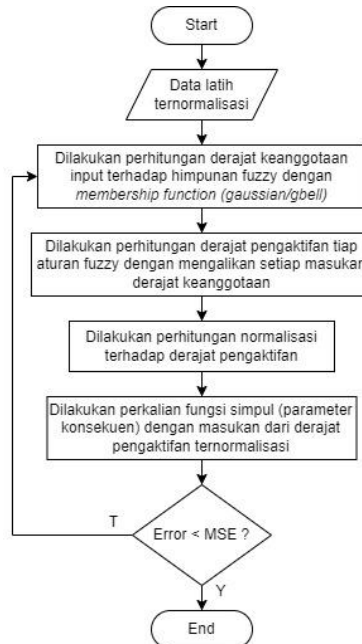
Gambar 3. Grafik Data Saham Take-Two Interactive Software, Inc.

2.2. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS merupakan arsitektur yang secara fungsional sama dengan logika fuzzy dan juga sama dengan jaringan syaraf tiruan fungsi radial dengan sedikit batasan tertentu [6]. ANFIS merupakan penggabungan dua metode yaitu antara Logika Fuzzy dan Jaringan Saraf Tiruan. Jaringan Saraf Tiruan adalah struktur jaringan di mana keseluruhan perilaku input-output ditentukan oleh serangkaian parameter yang dimodifikasi. Sedangkan sistem logika fuzzy bersifat adaptif, berarti bahwa sistem fuzzy dapat disesuaikan dengan kondisi sesuai. Logika fuzzy memiliki kelebihan dalam memodelkan aspek kualitatif dari pengetahuan manusia dan proses pengambilan keputusan dengan menerapkan basis aturan (*rules*). Sistem inferensi fuzzy dapat dilatih dengan metode propagasi balik berdasarkan pasangan data *input-output* menggunakan arsitektur jaringan saraf [3]. Kombinasi sistem inferensi fuzzy dengan jaringan saraf ini disebut dengan neuro-fuzzy.

Efektifitas dari ANFIS dengan penggabungan dua metode logika fuzzy dan jaringan saraf tiruan telah diuji oleh banyak penelitian seperti memodelkan dua dimensi fungsi *sinc*, memodelkan fungsi tiga input linear yang digunakan sebagai permasalahan *benchmark* untuk pendekatan fuzzy, menjelaskan bagaimana cara mengidentifikasi komponen non-linear dalam sebuah kontrol sistem yang *online*, hingga melakukan prediksi data *time series* dan hasilnya telah dibandingkan secara ekstensif dengan

pendekatan *connectionist* dan metode statistik konvensional [2]. ANFIS memiliki lima tahapan yang direpresentasikan pada layer di strukturnya yang dibangun menyerupai struktur jaringan saraf tiruan dengan fungsi logika fuzzy yang bergantung pada model fuzzy yang digunakan seperti Sugeno, Tsukasamoto, atau Mamdani. Input yang diproses pada ANFIS akan difuzzifikasi terlebih dahulu pada *layer* pertama dengan menggunakan *membership function*. Terdapat banyak tipe *membership function*, beberapa yang populer diantaranya adalah tipe triangular, *trapezoidal*, *gaussian*, dan *generalized bell*. Adapun flowchart metode ANFIS ditunjukkan pada Gambar 4.



Gambar 4. Flowchart Metode ANFIS

Berikut langkah-langkah yang ada pada metode ANFIS:

1). Semua simpul di *layer* pertama adalah simpul adaptif (parameter dapat berubah) dengan fungsi simpul:

$$O_{1,i} = \mu_{A_i}(x), \quad \text{untuk } i = 1, 2 \quad (1)$$

Dengan x dan y adalah masukan pada simpul i . Sedangkan A_i adalah *membership function* dari masing-masing simpul. Simpul $O_{1,i}$ menyatakan derajat keanggotaan setiap masukan terhadap himpunan fuzzy A dan B . *Membership function* yang digunakan pada penelitian ini adalah *gaussian* dan *generalized bell*.

2). Semua simpul di *layer* kedua adalah simpul non adaptif (parameter tetap). Luaran dari simpul ini adalah hasil perkalian dari setiap sinyal masukan yang datang dari *layer* 1 dengan fungsi perkalian:

$$O_{2,i} = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \quad i = 1, 2. \quad (2)$$

Setiap luaran dari simpul ini menyatakan derajat pengaktifan (*firing strength*) tiap aturan fuzzy. Banyaknya simpul pada *layer* ini menunjukkan banyaknya aturan yang dibentuk. Fungsi perkalian yang digunakan adalah interpretasi dari operator AND.

3). Semua simpul di *layer* ketiga adalah simpul non adaptif (parameter tetap) yang memiliki luaran fungsi derajat pengaktifan (*firing strength*) ternormalisasi yaitu rasio luaran simpul ke- i pada *layer* sebelumnya, dengan bentuk simpul:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \quad (3)$$

Apabila dibentuk lebih dari 2 aturan, maka fungsi dapat diperluas dengan membagi w_i dengan jumlah total w untuk seluruh aturan.

4). Semua simpul di *layer* keempat adalah simpul adaptif (parameter dapat berubah) dengan fungsi simpul:

$$O_{4,i} = f_i = \bar{w}_i \bar{a}_i z_i, \quad i = 1, 2. \quad (4)$$

Dengan parameter konsekuen a_i dan z_i masing-masing sebagai area dan pusat dari *membership function* konsekuen.

5). Semua simpul di *layer* kelima memiliki fungsi untuk menjumlahkan semua masukan yang datang dari *layer* keempat dengan fungsi simpul:

$$O_{5,i} = \sum f_i = \sum \bar{w}_i \bar{a}_i z_i, \quad i = 1, 2. \quad (5)$$

2.3. Mean Squared Error (MSE)

MSE merupakan salah satu metode untuk mencari rata-rata kesalahan kuadrat di antara nilai hasil prediksi dengan nilai aktual. MSE umumnya digunakan untuk mengecek estimasi *error* pada kasus prediksi. Semakin mendekati nilai nol, nilai MSE semakin baik karena menunjukkan bahwa hasil prediksi sesuai dengan data aktual dan model yang dibangun dapat memprediksi dengan baik (Das, Jiang and Rao, 2004). Nilai MSE dapat dihitung dengan persamaan berikut.

$$MSE = \frac{\sum_{t=1}^n (At - Ft)^2}{n} \quad (6)$$

Keterangan:

- 1) At = Nilai aktual
- 2) Ft = Nilai hasil prediksi
- 3) n = Banyaknya data

2.4. Mean Absolute Percentage Error (MAPE)

MAPE merupakan indikator persentase kesalahan antara data aktual dengan data peramalan. Nilai MAPE dapat dihitung dengan persamaan 2.6.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (7)$$

Keterangan:

- 1) Y_t = nilai aktual pada periode t
- 2) \hat{Y}_t = nilai prediksi pada periode t

Model peramalan akan sangat baik bila menghasilkan nilai MAPE yang kurang dari 10% dan akan bernilai buruk bila berada diatas 50% [7]. Kriteria nilai MAPE ditunjukkan pada Tabel 2.1.

Tabel 2. Kriteria Nilai MAPE [8]

Nilai MAPE	Kriteria
< 10%	Sangat Baik
10-20%	Baik
20-50%	Cukup
> 50%	Buruk

3. Hasil dan pembahasan

3.1. Analisa Data

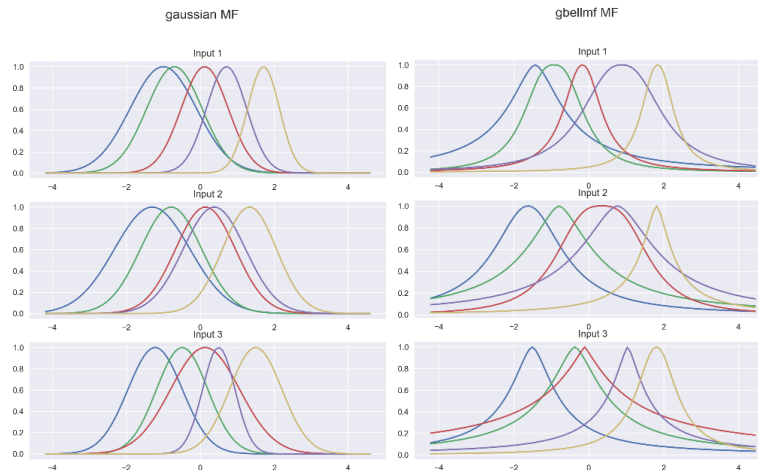
Pada penelitian ini digunakan data *time series* dalam kurun waktu kurang lebih 3 tahun dari tanggal 1 Februari 2019 hingga 1 Maret 2022 untuk masing-masing perusahaan Tencent Holdings Limited dan Take-Two Interactive Software, Inc. Setiap data saham dari masing-masing perusahaan memiliki 745 data. Dapat dilihat bahwa tren dari kedua kurva di atas mengalami peningkatan selama 3 tahun terakhir. Parameter yang akan digunakan pada data training adalah harga bukaan (open), harga terendah (low), dan harga penutupan (close) saham serta keluaran prediksi yang dihasilkan adalah harga tertinggi (high) saham. Data dari parameter yang digunakan akan dilakukan normalisasi terlebih dahulu. Dari 775 data tersebut selanjutnya dibagi menjadi dua bagian, yaitu:

- a. Data training: data yang akan dilatih oleh model ANFIS berjumlah 70% atau 512 data.
- b. Data testing: data yang akan diuji dengan metode uji MAPE berjumlah 30% atau 232 data.

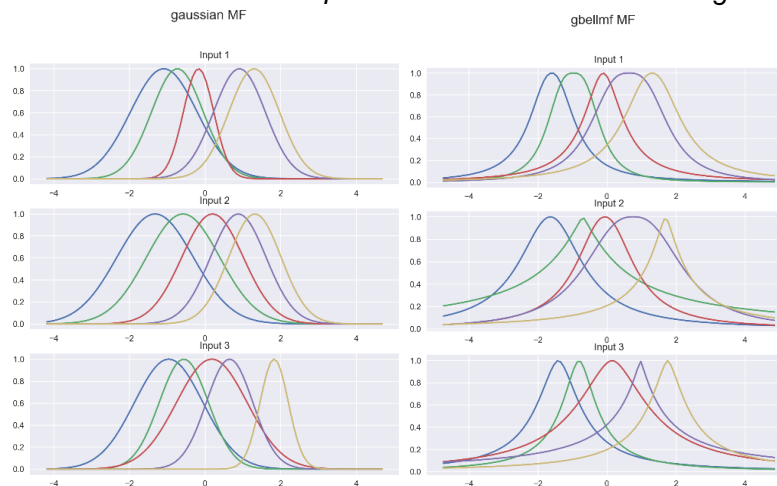
3.2. Permodelan ANFIS

3.2.1 Fuzzifikasi dan Set *Membership Function* untuk Penentuan Harga Saham

Pada *layer* pertama ANFIS didapatkan kurva dari masing-masing *membership function* pada Gambar 5 dan Gambar 6 berikut.



Gambar 5. Kurva membership function ANFIS Tencent Holdings Limited



Gambar 6. Kurva membership function ANFIS Take-Two Interactive Software, Inc.

Sumbu x pada kurva di atas menunjukkan jarak dari masing – masing tipe *membership function*, sedangkan sumbu y menunjukkan nilai *membership function*. Terlihat bahwa terdapat lima line dengan warna dan representasi dari setiap nilai *membership function* yaitu biru untuk sangat rendah, *orange* untuk rendah, merah untuk normal, ungu untuk tinggi, kuning untuk sangat tinggi. Pada Gambar 5 kurva *gbell* memiliki lebar pada titik tengah (a) pada line merah (normal) yang terlihat seperti bentuk kerucut pada input 3 (harga penutupan saham) tidak merepresentasikan bentuk lonceng *gbell* dengan baik, sedangkan kurva *gaussian* mampu merepresentasikan bentuk lonceng dengan lebih baik dengan lebar atau sigma tidak beda jauh antar nilai *membership function*. Sebaliknya pada Gambar 6 kurva *gbell* dan kurva *gaussian* dapat merepresentasikan bentuk lonceng dengan baik dengan perbedaan yang tidak terlalu signifikan.

3.2.2. Rule Evaluation Fuzzifikasi untuk Penentuan Harga Saham

Pada *layer* kedua digunakan aturan IF ... THEN ... dengan penghubung AND sebagai *rule evaluation*. Terdapat 125 *rules* pada penelitian ini karena masing-masing input yaitu harga bukaan (open), harga terendah (low), dan harga penutupan (close) memiliki 5 fuzzy value sehingga aturan menjadi 5³. Adapun penjabaran *rules* tersebut disajikan pada Tabel 3 berikut.

Tabel 3. Rules penentuan harga tertinggi saham

No.	Rules
1	IF (open is very low) AND (low is very low) AND (close is very low) THEN (high is very low)
2	IF (open is very low) AND (low is very low) AND (close is low) THEN (high is low)
3	IF (open is very low) AND (low is very low) AND (close is normal) THEN (high is normal)
4	IF (open is very low) AND (low is very low) AND (close is high) THEN (high is high)
...	
122	IF (open is very high) AND (low is very high) AND (close is low) THEN (high is low)
123	IF (open is very high) AND (low is very high) AND (close is normal) THEN (high is normal)

124	IF (open is very high) AND (low is very high) AND (close is high) THEN (high is high)
125	IF (open is very high) AND (low is very high) AND (close is very high) THEN (high is very high)

3.2.3. Training ANFIS

Dengan menggunakan model ANFIS yang telah dibangun, akan dibandingkan *performance* dari dua *membership function* yaitu *gaussian* dengan *gbell* pada ANFIS memprediksi harga tertinggi saham dari Tencent Holdings Limited dan Take-Two Interactive Software, Inc.

Tabel 4. Training model ANFIS Tencent Holdings Limited

Iterasi ke-	<i>gaussian</i>		<i>gbell</i>	
	loss	val_loss	loss	val_loss
1	0.024707066	0.001807548	0.010431395	0.001716427
...				
2057	4.15E-05	4.63E-05	3.93E-05	0.000163461
2058	3.72E-05	4.30E-05	4.79E-05	2.29E-05
...				
2950	3.32E-05	4.95E-05	4.60E-05	2.25E-05
2951	4.12E-05	6.34E-05	4.58E-05	2.55E-05
...				
5000	4.04E-05	6.45E-05	4.41E-05	3.14E-05
Rata-rata Error	4.5326E-05	5.1539E-05	4.87034E-05	3.08903E-05

Dilihat pada proses training ANFIS pada Tabel 4 dengan dua *membership function* dari data saham Tencent Holdings Limited, dapat diketahui bahwa struktur ANFIS yang dibangun dapat digunakan untuk memprediksi harga tertinggi (*high*) saham. Hal ini karena hasil error atau loss yang didapatkan setelah 5000 epoch di bawah toleransi 10^{-5} . Dapat diketahui juga bahwa *membership function gaussian* lebih baik dibandingkan *membership function gbell*. Hal ini karena *average training error* dengan tipe *gaussian* lebih rendah yaitu sebesar 4.5326E-05 dibandingkan tipe *gbell* sebesar 4.87034E-05.

Tabel 5. Training model ANFIS Take-Two Interactive Software, Inc

Iterasi ke-	<i>gaussian</i>		<i>gbell</i>	
	loss	val_loss	loss	val_loss
1	0.002878178	0.00090732	0.065913163	0.001820721
...				
4338	9.68E-05	0.000123259	0.000101379	0.000117335
4339	0.000106599	0.000115294	0.000104057	0.000108402
...				
4985	0.000105062	0.00010888	9.41E-05	0.000133573
4986	0.000105024	9.89E-05	0.000102787	0.000101113
...				
5000	0.000101365	0.000140406	9.98E-05	0.000119895
Rata-rata Error	0.000117181	0.000112768	0.000107349	0.000120323

Sedangkan hasil *training* ANFIS dari data Take-Two Interactive Software, Inc. pada Tabel 5 dapat diketahui bahwa model menghasilkan *average error* yang lebih besar yaitu sebesar 0.000117181 dengan menggunakan *membership function gaussian* dibandingkan dengan *average error* sebesar 0.000107349 menggunakan *membership function gbell*. Hal ini berarti bahwa *membership function gbell* dapat memprediksi harga saham pada perusahaan Take-Two Interactive Software, Inc. dengan lebih baik dibandingkan *membership function gaussian*. Dilihat dari tabel 5 di atas, diketahui juga bahwa hasil training menggunakan *membership function gbell* memiliki *error* minimum sebesar 9.67674E-05 pada iterasi ke 4338. Ini juga menunjukkan bahwa model ANFIS dapat memprediksi harga tertinggi saham karena error menyentuh toleransi 10^{-5} dengan jumlah iterasi maksimum hingga 5000.

3.3. Hasil Analisa

Untuk mengukur kinerja dari model ANFIS yang telah dibangun untuk memprediksi harga tertinggi saham, maka diperlukan testing. Testing dilakukan dengan membandingkan data *testing* dengan hasil prediksi harga tertinggi saham berjumlah 232 data. Hasil prediksi keseluruhan data testing Tencent Holdings Limited dapat dilihat pada Tabel 6.

Tabel 6. *Testing data Tencent Holdings Limited*

Data Testing	Harga Aktual	gaussian		gbell	
		Prediksi	Error	Aktual	Error
1	76.75	78.23654	1.48654	78.28708	1.53708
2	76.790001	76.18868	0.601321	76.34705	0.442951
3	76.769997	75.62179	1.148207	75.88286	0.887137
4	76.110001	76.62867	0.518669	76.77756	0.667559
...					
229	50.5	49.678722	0.821278	49.730736	0.769264
230	50.16	50.455944	0.295944	50.49502	0.33502
231	48.43	49.65043	1.22043	49.71058	1.28058
232	47.700001	48.26065	0.560649	48.332005	0.632004

Pada prediksi harga tertinggi saham Tencent Holdings Limited dengan kedua tipe *membership function* dapat dilihat bahwa error yang terjadi tidak begitu signifikan. Namun terdapat perbedaan *error* seperti pada prediksi data harga tertinggi saham ke 230 hingga 232 antara *membership function* tipe *gaussian* dan tipe *Gbell* dapat dilihat bahwa tipe *gaussian* memiliki error yang lebih kecil dibandingkan tipe *Gbell*. Hal ini ditunjukkan dari perbedaan error seperti pada data ke 232 sebesar 0.560649 untuk tipe *gaussian* dan 0.632004 untuk tipe *gbell*.

Tabel 7. *Testing data Take-Two Interactive Software, Inc*

Data Testing	Harga Aktual	gaussian MF		Gbell MF	
		Prediksi	Error	Aktual	Error
1	171.429993	168.87624	2.553753	168.79947	2.630523
2	169.25	168.17776	1.07224	168.18433	1.06567
3	167.479996	165.56448	1.915516	165.82451	1.655486
4	165.490005	164.2888	1.201205	164.38129	1.108715
...					
229	155.839996	150.2841	5.555896	150.86278	4.977216
230	157.089996	150.62344	6.466556	151.78915	5.300846
231	152	150.88594	1.11406	151.31923	0.68077
232	147.899994	147.17052	0.729474	147.35936	0.540634

Sedangkan pada hasil prediksi dengan data Take-Two Interactive Software, Inc dapat dilihat pada Tabel 7. Error hasil prediksi dengan *membership function gaussian* dan *membership function gbell* menunjukkan bahwa *error* yang terjadi tidak begitu signifikan dan relatif kecil. Namun terdapat pula perbedaan *error* seperti pada prediksi 5 data harga tertinggi saham terakhir, dapat dilihat bahwa tipe *gbell* memiliki error yang lebih kecil dibandingkan tipe *gaussian*. Hal ini ditunjukkan pada data ke 232 dari perbedaan jumlah *error* 0.540634 untuk tipe *gbell* dan 0.729474 untuk tipe *gaussian*. Dari kedua hasil prediksi harga tertinggi saham di perusahaan Tencent Holdings Limited dan Take-Two Interactive Software, Inc., diperoleh perhitungan untuk nilai uji error MAPE sebagai berikut.

Tabel 8. *Uji MAPE hasil prediksi model ANFIS*

Data	Nilai Uji MAPE	
	Tipe gaussian MF	Tipe Gbell/MF
Tencent Holdings Limited	0.02129 (2.13%)	0.02151 (2.15%)
Take-Two Interactive Software, Inc.	0.02276 (2.28%)	0.02196 (2.2%)

Dilihat pada tabel di atas, maka dapat ditunjukkan bahwa akurasi model ANFIS untuk data Tencent Holdings Limited lebih baik menggunakan *membership function* tipe *gaussian* dengan tingkat akurasi sebesar 97.87% sedangkan akurasi model ANFIS untuk data Take-Two Interactive Software, Inc. lebih baik menggunakan *membership function* tipe *gbell* dengan tingkat akurasi sebesar 97.8%.

4. Kesimpulan

Pada penelitian ini dilakukan prediksi harga tertinggi saham berdasarkan data dari perusahaan Tencent Holdings Limited dan perusahaan Take-Two Interactive Software, Inc sebanyak 745 data histori saham yang diperoleh dari situs Yahoo Finance menggunakan metode *Adaptive Neuro Fuzzy Inference System* (ANFIS). Pada model ANFIS yang dibangun digunakan dua *membership function* yaitu tipe *gaussian* dan *gbell*. Pengukuran kinerja metode CART penelitian ini menggunakan metode *Mean Absolute Percentage Error* (MAPE). Dari penelitian ini diperoleh bahwa kinerja dari model ANFIS dengan tipe *gaussian* lebih baik dibandingkan tipe *gbell* dengan tingkat akurasi 97.87% untuk memprediksi harga tertinggi saham Tencent Holdings Limited, sedangkan tipe *gbell* memiliki tingkat akurasi sebesar 97.8% yang lebih baik dibandingkan tipe *gaussian* untuk memprediksi harga tertinggi saham Take-Two Interactive Software, Inc. Terlihat bahwa tipe *gaussian* dapat menghasilkan keputusan yang lebih tepat untuk prediksi harga tertinggi saham perusahaan Tencent Holdings Limited dan tipe *gbell* menghasilkan keputusan yang lebih tepat untuk prediksi harga tertinggi saham perusahaan Take-Two Interactive Software, Inc.

Referensi

- [1] S. Sitohang and V. Karnadi, "Prediksi Harga Saham Menggunakan Generalize Fuzzy Inference System (GENFIS3)," no. September, 2020.
- [2] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. United States of America: Prentice-Hall, Inc., 1997.
- [3] M. A. Raharja, I. D. M. B. A. Darmawan, D. P. E. Nilakusumawati, and I. W. Supriana, "Analysis of membership function in implementation of adaptive neuro fuzzy inference system (ANFIS) method for inflation prediction," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012005.
- [4] Y. Tan, C. Shuai, L. Jiao, and L. Shen, "An adaptive neuro-fuzzy inference system (ANFIS) approach for measuring country sustainability performance," *Environ. Impact Assess. Rev.*, vol. 65, no. October 2016, pp. 29–40, 2017, doi: 10.1016/j.eiar.2017.04.004.
- [5] A. Hartanti, "Peramalan Harga Saham Pada Lima Emiten Terbaik Versi Forbes Tahun 2012 Menggunakan Fuzzy Model," 2014.
- [6] M. I. Azhar and W. F. Mahmudy, "Prediksi Curah Hujan Menggunakan Metode Adaptive Neuro Fuzzy Inference System (ANFIS)," vol. 2, no. 11, 2018.
- [7] P. C. Chang, Y. W. Wang, and C. H. Liu, "The development of a weighted evolving fuzzy neural network for PCB sales forecasting," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 86–96, 2007, doi: 10.1016/j.eswa.2005.11.021.
- [8] I. M. Satriawan, I. M. Mataram, and A. A. N. Amrita, "Peramalan beban listrik jangka pendek menggunakan metode ANFIS di Gardu Induk Nusa Dua Bali," *Spektrum*, vol. 7, no. 1, pp. 83–89, 2020.
- [9] O. Noureldeen, "Design of robust intelligent protection technique for large -scale grid-connected wind farm," 2018.
- [10] M. Lutfi, "Prediksi Harga Terendah Dan Harga Tertinggi Dengan Menggunakan Metode Anfis Untuk Analisa Teknikal Pada Forex Market," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 1, no. 3, pp. 261–268, 2019, doi: 10.35746/jtim.v1i3.40.

Analisis Sentimen Ulasan Aplikasi Transportasi *Online* Menggunakan *Multinomial Naïve Bayes* dan *Query Expansion Ranking*

Yuma Anugrah Virya Gunawan^{a1}, Ngurah Agus Sanjaya ER^{a2}, Ida Bagus Made Mahendra^{a3}, I Made Widiartha^{a4}, I Gusti Ngurah Anom Cahyadi Putra^{a5}, I Gusti Agung Gede Arya Kadyanan^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana
Bali, Indonesia

¹yumagunawan22@gmail.com

²agus_sanjaya@unud.ac.id

³ibm.mahendra@unud.ac.id

⁴madewidiartha@unud.ac.id

⁵anom.cp@unud.ac.id

⁶gungde@unud.ac.id

Abstract

The rapid development of the transportation industry in recent years has led to a new innovation in the field of transportation, namely the application of online transportation services. To facilitate the translation of user satisfaction, in addition to users being able to provide reviews, the Google Play Store uses a rating system consisting of a rating of 1 to 5. However, users often do not provide a rating that is in accordance with the review so that this is not enough to determine the sentiment of the review. This research is focused on evaluating the performance of the features selection using Query Expansion Ranking on the Multinomial Naïve Bayes method in the problem of sentiment analysis on the two of most popular online transportation service applications in Indonesia, namely Gojek and Grab. From the results of the performance evaluation using k-fold cross validation, it was found that the best feature selection ratio was 20% with the best performance in terms of precision.

Keywords: Sentiment Analysis, Feature Selection, Multinomial Naïve Bayes, TF – IDF, Query Expansion Ranking

1. Pendahuluan

Perkembangan industri transportasi yang pesat dalam beberapa tahun terakhir membawa sebuah inovasi baru di bidang transportasi, yaitu adanya layanan transportasi *online* dengan menggunakan teknologi berbasis aplikasi. Aplikasi layanan transportasi *online* ini memberikan kemudahan bagi para penggunanya untuk memesan layanan seperti mengantar orang, makanan, atau barang hanya dengan menggunakan *smartphone* [1].

Dalam hal distribusi aplikasi, khususnya di sistem operasi *android*, salah satu wadah bagi pengguna untuk memperoleh aplikasi dari pengembang adalah melalui *Google Play Store*. Di samping itu, pengguna juga dapat menuliskan dan membagikan ulasan terkait pengalaman terhadap pemakaian suatu aplikasi. Untuk memudahkan penerjemahan kepuasan pengguna, banyak *platform* seperti *Google Play Store* memakai sistem rating yang terdiri dari rating 1 hingga 5, namun seringkali pengguna tidak memberikan rating yang sesuai dengan ulasannya sehingga hal tersebut belum cukup untuk menentukan sentimen dari ulasan tersebut. Oleh karena itu, untuk memudahkan pengelolaan dan analisis data ulasan, perlu dilakukan analisis sentimen. Pendekatan analisis sentimen dapat menyediakan analisis yang mendalam mengenai sentimen, pendapat, dan ekspresi seseorang dalam bentuk data tekstual [2].

Dalam klasifikasi sentimen, ada satu hal yang paling penting yaitu proses pelabelan atau *labeling*. Proses pelabelan adalah proses pemberian label pada dataset ulasan. Proses pelabelan yang paling umum adalah pelabelan manual dengan bantuan ahli bahasa. Namun, hal ini sulit dilakukan dengan jumlah dataset yang banyak karena akan memakan waktu yang lama dan memerlukan tim ahli yang

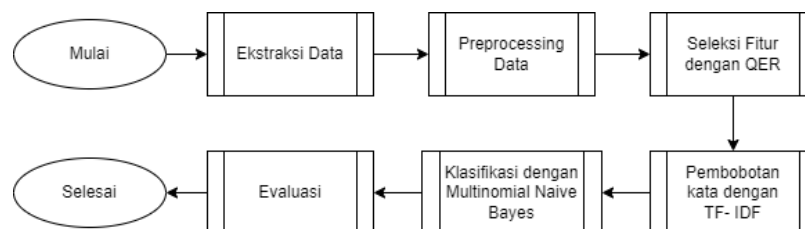
banyak. Penelitian ini menggunakan skema pelabelan *average labeling* karena dapat diterapkan pada ukuran dataset yang besar dengan waktu yang singkat [3].

Salah satu cara meningkatkan kualitas fitur dataset adalah dengan proses seleksi fitur. Metode seleksi fitur yang digunakan dalam penelitian ini adalah metode *Query Expansion Ranking* [4]. Sedangkan, untuk klasifikasi teks menggunakan salah satu metode yang banyak digunakan pada pendekatan klasifikasi dan analisis sentimen yakni varian *multinomial* dari metode *Naïve Bayes* yaitu *Multinomial Naïve Bayes* [5]. Implementasi *query expansion* sendiri sudah banyak diterapkan pada berbagai penelitian, salah satunya pada penelitian *set labelling* [6]. Penelitian oleh [7] menunjukkan bahwa metode seleksi fitur *Query Expansion Ranking* juga menghasilkan performa yang baik pada metode KNN.

Penelitian ini bertujuan untuk menemukan performa rasio seleksi fitur terbaik dari metode *Query Expansion Ranking* pada metode klasifikasi *Multinomial Naïve Bayes* dalam permasalahan analisis sentimen terhadap ulasan aplikasi layanan transportasi *online*. Performa diukur berdasarkan hasil *accuracy*, *precision*, *recall*, dan *f1 – score*.

2. Metode Penelitian

Pada bagian ini akan dijelaskan gambaran secara umum mengenai alur penelitian yang dilakukan oleh penulis. Pertama adalah proses ekstraksi data dari sumber yakni situs web *Google Play Store*, kemudian dari data yang diekstraksi dihasilkan *raw data*. Dataset hasil ekstraksi akan melalui proses *text preprocessing* agar menjadi *clean data*. Data yang sudah bersih akan melalui proses seleksi fitur menggunakan metode *Query Expansion Ranking* yang berkaitan dengan skenario pengujian yang akan dijalankan yakni untuk mencari rasio persentase seleksi fitur terbaik dengan penggunaan seluruh fitur sebagai *baseline*. Selanjutnya, dilakukan pembobotan kata dengan TF – IDF yang menghasilkan vektor fitur untuk digunakan dalam tahap klasifikasi. Tahap klasifikasi dilakukan dengan metode *Multinomial Naïve Bayes*. Hasil klasifikasi dapat dievaluasi berdasarkan nilai *accuracy*, *recall*, *precision*, dan *F – 1 score*. Diagram alir penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

2.1. Ekstraksi Data

Ekstraksi data dilakukan menggunakan teknik *web scraping* dengan bahasa pemrograman Python dan *library google-play-scrapers*. Data yang diambil merupakan ulasan dari *region* Indonesia yang berbahasa Indonesia dari dua aplikasi layanan transportasi *online* yakni *Gojek* dan *Grab*. Dalam penelitian ini, penulis mengambil 1000 ulasan dengan masing – masing 500 data untuk tiap kelas positif dan negatif dari tiap aplikasi sehingga total menggunakan 2000 ulasan. Adapun skema pelabelan yang digunakan yaitu *average labeling* dimana berdasarkan pada rating kedua aplikasi saat proses pengambilan data, maka ulasan dengan rating 1 – 4 dilabelkan sebagai negatif dan rating 5 sebagai positif. Hasil ekstraksi ulasan ditunjukkan pada Gambar 2.

```

{'reviewId': 'gp:A0qpT0EOMv80XG85eNdMPiyow-CzPjKfW0wy2wh1RZRjDlX7s8pr242DEsK3P3Aeww5mQQZHec1weAti7Psdmg',
'userName': 'Syahdan N',
'userImage': 'https://play-lh.googleusercontent.com/a-/A0h14GjBvZh43xfIYY8DHmoFLt0CUFWQTwhYJXIC65ASDA',
'content': 'Tolong ditambahkan menu atau pilihan penghapusan akun', 'score': 4,
'thumbsUpCount': 0, 'reviewCreatedVersion': '4.16.1',
'at': datetime.datetime(2021, 4, 6, 9, 27, 46), 'replyContent': None, 'repliedAt': None}
  
```

Gambar 2. Hasil Ekstraksi Ulasan

2.2. Text Preprocessing

Dokumen hasil ekstraksi data yang sudah disimpan akan melalui proses *preprocessing* seperti yang ditunjukkan pada Tabel 1. Tahap *preprocessing* pada penelitian ini terdiri dari *case folding*, *tokenizing*, *word handling*, *filtering*, dan *stemming*.

Case folding adalah tahapan teknik pembersihan yang dilakukan untuk menyisakan kata dengan karakter huruf kecil. Oleh karena itu, dilakukan penghilangan angka, tanda baca, emotikon dan lain sebagainya serta pengubahan semua karakter ke huruf kecil.

Tokenizing merupakan tahapan pemecahan *string* kalimat menjadi *token* yang berupa kata.

Word handling merupakan tahapan yang dilakukan untuk menangani kata singkatan atau salah ejaan (*typo*) dengan cara mengganti kata pada ulasan menjadi kata yang sesuai dengan *dictionary* yang disediakan. Algoritma penanganan kata menggunakan algoritma dari Peter Norvig¹.

Filtering adalah tahapan untuk menghilangkan kata yang sering muncul namun tidak memiliki relevansi dengan teks atau dapat dikatakan kurang bermakna. Untuk kamus kata dan *stopwords* yang digunakan merupakan kombinasi dari *list stopwords* yang disediakan oleh *library Sastraw* dan media internet².

Stemming adalah tahapan untuk mengubah kata menjadi bentuk dasar dengan menghilangkan semua imbuhan.

Tabel 1. Proses *Text Preprocessing*

No	Proses	Input	Output
1	<i>Case Folding</i>	Sava sanqt kcewa pada aolikasi ini, pelayanan sring terlambat belakangan ini.. Harapannya bisa lebih baik lagi ke depannya,.. 🙏🙏🙏	sava sanqt kcewa pada aolikasi ini pelayanan sring terlambat belakangan ini harapannya bisa lebih baik lagi ke depannya
2	<i>Tokenizing</i>	sava sanqt kcewa pada aolikasi ini pelayanan sring terlambat belakangan ini harapannya bisa lebih baik lagi ke depannya	'saya', 'sangat', 'kecewa', 'pada', 'aolikasi', 'ini', 'pelayanan', 'sring', 'terlambat', 'belakangan', 'ini', 'harapannya', 'bisa', 'lebih', 'baik', 'lagi', 'ke', 'depannya'
3	<i>Word Handling</i>	'saya', 'sangat', 'kecewa', 'pada', 'aolikasi', 'ini', 'pelayanan', 'sring', 'terlambat', 'belakangan', 'ini', 'harapannya', 'bisa', 'lebih', 'baik', 'lagi', 'ke', 'depannya'	'saya', 'sangat', 'kecewa', 'pada', 'aplikasi', 'ini', 'pelayanan', 'sering', 'terlambat', 'belakangan', 'ini', 'harapannya', 'bisa', 'lebih', 'baik', 'lagi', 'ke', 'depannya'
4	<i>Filtering</i>	'saya', 'sangat', 'kecewa', 'pada', 'aplikasi', 'ini', 'pelayanan', 'sering', 'terlambat', 'belakangan', 'ini', 'harapannya', 'bisa', 'lebih', 'baik', 'lagi', 'ke', 'depannya'	'kecewa', 'aplikasi', 'pelayanan', 'terlambat', 'harapannya', 'lebih', 'baik', 'depannya'
5	<i>Stemming</i>	kecewa, 'aplikasi', 'pelayanan', 'terlambat', 'harapannya', 'lebih', 'baik', 'depannya'	kecewa, 'aplikasi', 'layan', 'lambat', 'harap', 'lebih', 'baik', 'depan'

2.3. Seleksi Fitur dengan *Query Expansion Ranking*

Dalam penelitian ini, penulis menggunakan metode seleksi fitur *Query Expansion Ranking* (QER). Pada metode QER, skor tiap fitur akan dihitung, kemudian fitur dengan skor terendah akhirnya dipilih dan digunakan dalam proses klasifikasi. Jika skor fitur rendah, perbedaan antara probabilitas kelas positif dan negatif menjadi besar. Hal ini menunjukkan bahwa fitur bersifat lebih unik pada setiap kelas dan lebih bermakna bagi proses klasifikasi [4]. Adapun langkah perhitungannya sebagai berikut:

- a. Hitung nilai probabilitas fitur *f* pada dokumen kelas positif.

$$pf = \frac{df^+ + 0.5}{n^+ + 1.0} \quad (1)$$

- b. Hitung nilai probabilitas fitur *f* pada dokumen kelas negatif.

$$qf = \frac{df^- + 0.5}{n^- + 0.5} \quad (2)$$

¹ <http://norvig.com/spell-correct.html>

² <http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/>

c. Hitung skor QER untuk fitur f

$$score_f = \frac{|pf+qf|}{|pf-qf|} \quad (3)$$

Keterangan :

- df^+ : jumlah dokumen kelas positif yang memuat fitur f
- n^+ : jumlah dokumen kelas positif
- df^- : jumlah dokumen kelas negatif yang memuat fitur f
- n^- : jumlah dokumen kelas negatif
- pf : nilai probabilitas fitur f pada dokumen kelas positif
- qf : nilai probabilitas fitur f pada dokumen kelas negatif
- $score_f$: nilai skor QER fitur f

2.4. Pembobotan Kata dengan *Term Frequency – Inverse Document Frequency* (TF – IDF)

TF - IDF adalah metode pembobotan kata yang digunakan untuk menentukan pentingnya sebuah kata dalam sebuah dokumen. Ada dua komponen dalam perhitungan nilai TF – IDF, yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF menentukan pentingnya suatu kata relatif terhadap kemunculannya dalam suatu dokumen, sedangkan IDF menentukan suatu kata penting dalam suatu dokumen jika tidak sering muncul di dokumen lain [8]. Adapun rumus perhitungan sesuai penelitian oleh [9]:

$$a_{ij} = tf_{ij} * \log \left(\frac{N}{n_j} \right) \quad (4)$$

Dengan keterangan :

- a_{ij} : nilai bobot *term j* pada dokumen i
- tf_{ij} : jumlah kemunculan kata atau nilai *term frequency term j* pada dokumen i
- N : jumlah dokumen pada dataset
- n_j : jumlah dokumen yang memuat *term j*

2.5. Klasifikasi dengan *Multinomial Naïve Bayes*

Multinomial Naïve Bayes merupakan implementasi algoritma *Naïve Bayes* yang umumnya digunakan dalam pemrosesan teks dengan mengikuti prinsip distribusi multinomial [10]. Penggunaan model distribusi multinomial menunjukkan bahwa vektor fitur dalam suatu dokumen dibentuk dari frekuensi kemunculan setiap kata pada dokumen tersebut. Adapun algoritma perhitungannya [11]:

a. Untuk tiap fitur pada vektor data latih, hitung nilai kemungkinan kemunculan fitur i pada kelas y .

$$\theta_{yi} = \frac{N_{yi+a}}{N_y+an} \quad (5)$$

Keterangan:

- θ_{yi} : nilai kemungkinan kemunculan *term i* pada kelas y
- N_{yi} : jumlah kemunculan *term i* pada kelas y
- N_y : jumlah kemunculan seluruh fitur pada kelas y
- a : nilai *Laplace smoothing parameter*
- n : total fitur dalam koleksi dokumen

b. Hitung nilai *prior probability* pada setiap kelas y

$$P_y = \frac{D_y}{D} \quad (6)$$

Keterangan:

- P_y : nilai *prior probability* pada kelas y
- D_y : jumlah dokumen pada kelas y
- D : jumlah seluruh dokumen

c. Tentukan hasil prediksi kelas pada data uji

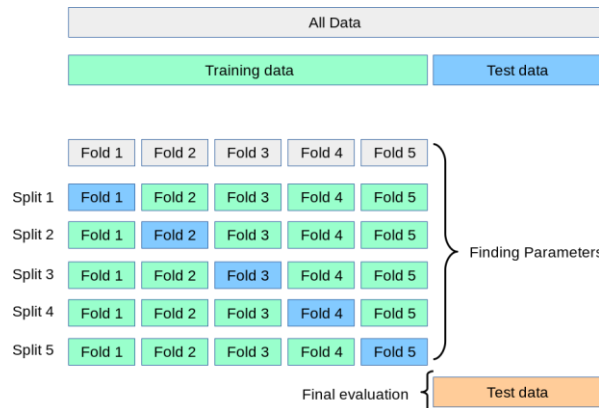
$$\hat{c} = \operatorname{argmax} P_y \prod (\theta_{yi})^{M_{yi}} \quad (7)$$

Keterangan:

- \hat{c} : kelas hasil prediksi
- M_{yi} : jumlah kemunculan *term i* pada kelas y dalam dokumen uji

2.6. Skenario Pengujian dan Evaluasi

Pengujian dalam penelitian ini menggunakan *K – Fold Cross – Validation* yang ditunjukkan pada Gambar 3 [12]. Dataset ulasan akan dibagi menjadi data latih dan data uji terlebih dahulu dengan porsi 80 : 20. Kemudian data latih akan dilakukan iterasi sebanyak *k* bagian atau *folds* yang membagi tiap bagiannya menjadi data latih dan data validasi dengan porsi satu bagian menjadi data validasi dan bagian lainnya menjadi data latih secara bergantian untuk setiap iterasinya. Pengujian ini dilakukan untuk mendapatkan parameter rasio seleksi fitur terbaik yakni dengan melihat rata – rata nilai metrik (*accuracy, precision, recall, dan f1 – score*) dari keseluruhan iterasi terhadap rasio seleksi fitur yang digunakan. Parameter rasio seleksi fitur terbaik akan dibandingkan performanya pada data uji dengan tanpa penggunaan seleksi fitur atau penggunaan fitur penuh sebagai metode *baseline*.



Gambar 3. *K – Fold Cross Validation*

Evaluasi bertujuan untuk mengukur performa pada sistem yang dibangun. Performa sistem klasifikasi dapat dievaluasi dengan membandingkan kelas yang diprediksi dengan kelas sebenarnya. Informasi ini dapat dituangkan dalam tabel yang disebut *confusion matrix* seperti yang ditunjukkan pada Tabel 2. Dari perhitungan pada *confusion matrix*, diperoleh nilai *accuracy, precision, recall, dan f1 – score*.

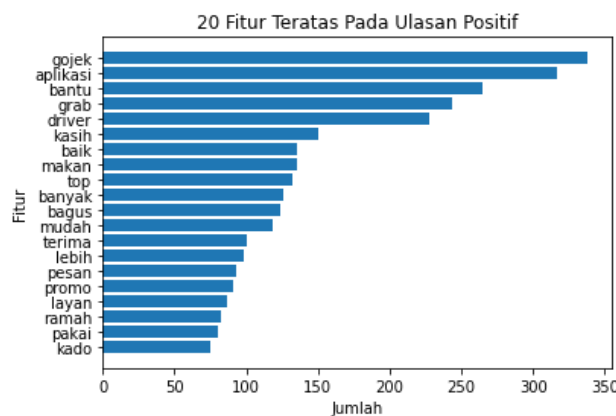
Tabel 2. *Confusion Matrix*

	Actual Positive	Actual Negative
Predicted Positive	<i>True Positive</i>	<i>False Positive</i>
Predicted Negative	<i>False Negative</i>	<i>True Negative</i>

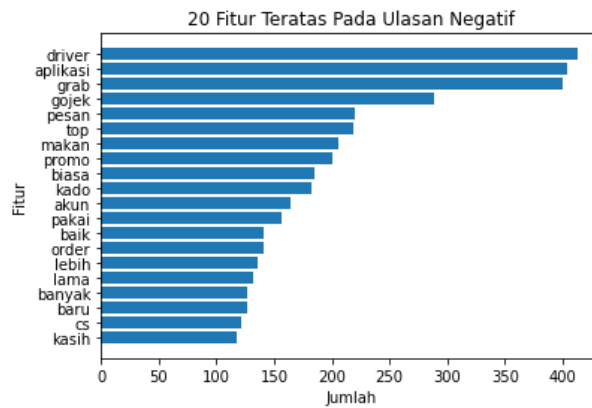
3. Hasil dan Pembahasan

3.1. Hasil Eksplorasi Data

Pada tahap ini, dilakukan eksplorasi data untuk mengetahui informasi sebaran fitur pada dataset yang digunakan. Salah satu langkah yang dapat dilakukan untuk mengetahui sebaran fitur adalah dengan melihat jumlah fitur terbanyak. Gambar 4 dan Gambar 5 menunjukkan 20 fitur teratas pada kelas positif dan negatif secara berurutan.



Gambar 4. 20 Fitur Teratas Pada Ulasan Kelas Positif



Gambar 5. 20 Fitur Teratas Pada Ulasan Kelas Negatif

Dengan melihat 20 fitur teratas, dapat memberikan gambaran lebih jelas mengenai jumlah tiap fitur pada tiap kelas. Dari kedua hasil visualisasi terlihat bahwa terdapat banyak kesamaan fitur yang sering muncul pada kedua kelas yakni kata “*gojek*”, “*grab*”, “*driver*”, dan “*aplikasi*”. Dimana kata “*gojek*” dan “*grab*” merupakan nama aplikasi yang merupakan sumber dari dataset yang digunakan serta kata “*driver*” dan “*aplikasi*” yang masih berkaitan dengan aplikasi layanan transportasi *online*. Selain itu, juga terdapat fitur yang dominan pada kelas tertentu seperti kata “*bantu*” dan “*bagus*” yang dominan pada kelas positif serta kata “*promo*”, “*akun*” dan “*kecewa*” yang dominan pada kelas negatif. Dalam visualisasi juga ditunjukkan bahwa terdapat ketimpangan pada jumlah fitur, dimana kata “*driver*” pada kelas positif hanya berjumlah sekitar 200, sedangkan pada kelas negatif berjumlah sekitar 400. Hal itu berarti dalam dataset yang digunakan lebih banyak kata “*driver*” dengan konotasi negatif. Begitu pula untuk kata “*grab*”.

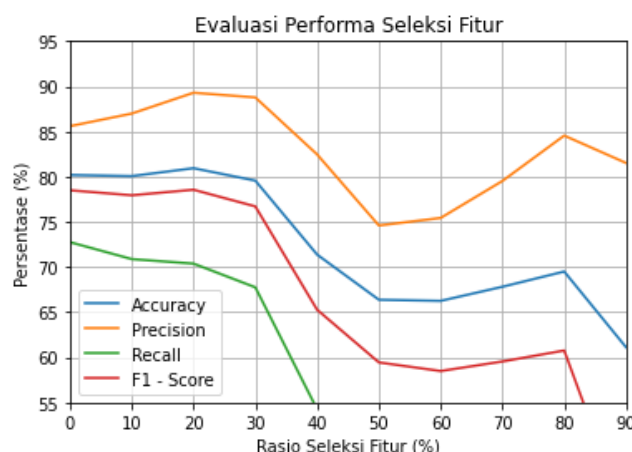
3.2. Hasil Evaluasi Performa Seleksi Fitur

Pengujian dilakukan untuk mencari parameter rasio seleksi fitur terbaik menggunakan *cross validation* dengan 10 *folds*. Dari 1600 data sebagai data latih, dibagi lagi menjadi data latih dan data validasi pada proses *cross validation*. Performa rasio seleksi fitur diukur dari rata – rata nilai metrik *accuracy*, *precision*, *recall* dan *f1 – score*. Adapun hasil evaluasi performa seleksi fitur dapat dilihat pada Tabel 3.

Tabel 3. Evaluasi Performa Seleksi Fitur

Rasio Seleksi Fitur	Rata – Rata			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – score</i>
0	0.80188	0.85591	0.7275	0.78492
10	0.80063	0.86985	0.70875	0.7794
20	0.81	0.89514	0.70625	0.78559
30	0.79688	0.8765	0.6775	0.76812
40	0.71438	0.82535	0.53875	0.6507
50	0.66188	0.74384	0.4925	0.59155
60	0.65686	0.75234	0.4775	0.58318
70	0.67812	0.79471	0.47625	0.59446
80	0.69312	0.84773	0.47625	0.60801
90	0.61063	0.81064	0.28625	0.41826

Untuk memudahkan melihat perubahan tren kenaikan dan penurunan nilai metrik evaluasi, dibentuk visualisasi berupa grafik yang ditunjukkan pada Gambar 6. Dimana 0% berarti tanpa penggunaan seleksi fitur atau penggunaan seluruh fitur pada dokumen hingga 90% pengurangan fitur. Dari grafik tersebut diketahui bahwa *precision*, *accuracy*, dan *f1 – score* terdapat kenaikan tren paling tinggi mencapai 20% penggunaan seleksi fitur, kemudian tren menurun setelahnya. Adapun nilai *recall* terdapat penurunan dari metode *baseline*. Kenaikan dan penurunan dari metode *baseline* terlihat tidak terlalu signifikan pada segi *accuracy*, *f1 – score*, dan *recall*. Namun, dari segi *precision* perubahan hampir mencapai 5% pada penggunaan rasio seleksi fitur 20%.



Gambar 6. Tren Evaluasi Performa Seleksi Fitur

Penurunan tren setelah rasio penggunaan seleksi fitur sebanyak 20% menunjukkan adanya pengurangan fitur yang terlalu banyak jumlahnya sehingga menyebabkan fitur yang memiliki relevansi dengan data validasi juga ikut dihilangkan. Kenaikan yang cukup signifikan dari segi *precision* terjadi karena nilai *false positive* mengalami penurunan dimana berarti pada rasio seleksi fitur 20%, jumlah ulasan dengan kelas negatif yang diberi label positif oleh model sudah berkurang dibandingkan pada metode *baseline*. Namun, di sisi lain penurunan *recall* disebabkan oleh peningkatan nilai *false negative* yang berarti model lebih banyak memberikan label negatif pada ulasan positif dibandingkan metode *baseline*. Nilai *f1-score* hampir tidak terlihat perubahan karena kenaikan dan penurunan nilai *precision* dan *recall* yang sebanding. Sedangkan kenaikan akurasi yang tidak terlalu signifikan disebabkan oleh sedikit peningkatan pada jumlah prediksi yang benar dari kedua kelas. Seleksi fitur dengan rasio 20% dinilai paling optimal dibandingkan rasio lainnya, maka akan dipilih untuk dibandingkan performanya dengan metode *baseline* pada data uji yang belum dikenali pada proses *cross validation* sejumlah 400 data. Evaluasi performa model pada data uji dapat dilihat pada Tabel 4.

Tabel 4. Evaluasi Performa pada Data Uji

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Tanpa Seleksi Fitur	0.7575	0.81988	0.66	0.73130
Rasio Seleksi Fitur Terbaik (20%)	0.75	0.85211	0.605	0.70760

Berdasarkan Tabel 4 dari hasil evaluasi performa pada data uji didapatkan nilai metode *baseline* atau tanpa penggunaan seleksi fitur untuk *accuracy*, *precision*, *recall*, dan *f1-score* secara berturut – turut adalah 0.7575, 0.81988, 0.66 dan 0.73130. Sedangkan untuk penggunaan rasio seleksi fitur terbaik yakni sebesar 20% didapatkan 0.75, 0.85211, 0.605, 0.70760. Dapat diketahui bahwa dalam hal performa pada data uji, metode *baseline* lebih baik dari segi *accuracy*, *recall*, dan *f1-score*. Sedangkan penggunaan seleksi fitur sedikit lebih baik dari sisi *precision*.

4. Kesimpulan

Dari hasil pengujian dan evaluasi pada penelitian ini diketahui bahwa penggunaan rasio seleksi fitur paling optimal yaitu pada rasio pengurangan 20% fitur. Hasil evaluasi performa pada data uji menunjukkan bahwa penggunaan rasio seleksi fitur sebanyak 20% memperoleh nilai *accuracy*, *precision*, *recall*, dan *f1-score* secara berturut – turut sebesar 0.75, 0.85211, 0.605, dan 0.70760. Sedangkan pada metode *baseline* memperoleh nilai 0.7575, 0.81988, 0.66 dan 0.73130. Dalam hal performa pada data uji, penggunaan seleksi fitur lebih unggul dari segi *precision*, sedangkan metode *baseline* sedikit lebih baik dari segi *accuracy*, *recall*, dan *f1-score*.

Adapun dari pengujian pada variasi rasio seleksi fitur yang berbeda, diketahui bahwa performa dari segi *precision*, *accuracy*, dan *f1-score* terdapat kenaikan tren hingga paling tinggi mencapai 20%, kemudian tren menurun setelahnya. Adapun nilai *recall* terdapat penurunan tren. Peningkatan nilai *precision* dan penurunan nilai *recall* tersebut dipengaruhi oleh penurunan nilai *false positive* dan

kenaikan nilai *false negative*, sehingga terdapat kecenderungan model lebih banyak kepada pemberian label negatif.

Daftar Pustaka

- [1] I. Farida, A. Tarmizi, and Y. November, "Analisis Pengaruh Bauran Pemasaran 7P Terhadap Kepuasan Pelanggan Pengguna Gojek Online," *J. Ris. Manaj. dan Bisnis Fak. Ekon. UNIAT*, vol. 1, no. 1, pp. 31–40, 2016, doi: 10.36226/jrmb.v1i1.8.
- [2] L. Zhang and B. Liu, "Sentiment Analysis and Opinion Mining", *Encyclopedia of Machine Learning and Data Mining*, pp. 1152-1161, 2017. Available: 10.1007/978-1-4899-7687-1_907 [Accessed 12 July 2022].
- [3] V. O. Tama, Y. Sibaroni, and Adiwijaya, "Labeling Analysis in the Classification of Product Review Sentiments by using Multinomial Naive Bayes Algorithm," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012036.
- [4] T. Parlar, S. A. Özel, and F. Song, "QER: a new feature selection method for sentiment analysis," *Human-centric Comput. Inf. Sci.*, vol. 8, no. 1, pp. 1–19, 2018, doi: 10.1186/s13673-018-0135-8.
- [5] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, "Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 2, no. 1, pp. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.
- [6] E. R. Ngurah Agus Sanjaya, T. Abdessalem, J. Read, and S. Bressan, "Set labelling using multi-label classification," *ACM Int. Conf. Proceeding Ser.*, pp. 216–220, 2018, doi: 10.1145/3282373.3282391.
- [7] N. D. Mentari, M. A. Fauzi, and L. Muflikhah, "Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 8, pp. 2739–2743, 2018.
- [8] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document Preprocessing with TF-IDF to Improve the Polarity Classification Performance of Unstructured Sentiment Analysis," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 235–242, 2020, doi: 10.22219/kinetik.v5i3.1066.
- [9] M. Liu and J. Yang, "An improvement of TFIDF weighting in text categorization," *Int. Conf. Comput. Technol. Sci.*, vol. 47, no. Iccts, pp. 44–47, 2012, doi: 10.7763/IPCSIT.2012.V47.9.
- [10] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.
- [11] N. Rezaeian and G. Novikova, "Persian text classification using naive bayes algorithms and support vector machine algorithm," *Indones. J. Electr. Eng. Informatics*, vol. 8, no. 1, pp. 178–188, 2020, doi: 10.11591/ijeii.v8i1.1696.
- [12] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python Pedregosa, Varoquaux, Gramfort *et al.*," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: <http://scikit-learn.org>.

Implementasi Algoritma Gronsfeld Cipher dan Steganografi End Of File Untuk Pengamanan Data

Kadek Vincky Sedana^{a1}, I Komang Ari Mogi^{a2}, Ida Bagus Gede Dwidasmara^{a3}, I Gede Arta Wibawa^{a4},
Cokorda Rai Adi Pramatha^{a5}, Luh Gede Astuti^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana,
Badung, Bali, Indonesia

¹vinckysedana@gmail.com

²arimogi@unud.ac.id

³dwidasmara@unud.ac.id

⁴gede.arta@unud.ac.id

⁵cokorda@unud.ac.id

⁶lg.astuti@unud.ac.id

Abstract

Information is a collection of data or facts that are organized or processed in a certain way so that they have meaning for both recipients and users. Data and information are very important things to keep their security or confidentiality so that unauthorized parties cannot find out the data or information. This Data Security Application was developed using the Gronsfeld Cipher algorithm and the End Of File Steganography method. This desktop-based application was created with the aim of increasing data security on confidential information so that unauthorized parties cannot find out the contents of the data. This application is expected to be a solution to prevent the theft of important information by unauthorized parties. From the test results, the developed application can receive input, process input and produce output as expected. Testing the quality of the embedded image gets an average PSNR value of 75,543 dB which can be said to be good because an image can be said to be good if it has a PSNR value above 40 dB, this also shows that the developed application can protect the data embedded in the digital image.

Keywords: *Cryptography, Steganography, Gronsfeld Cipher, End Of File, Data Security*

1. Pendahuluan

Pesatnya perkembangan teknologi informasi mempengaruhi semua aspek kehidupan. Salah satunya yaitu dalam pengamanan informasi yang bersifat rahasia. Data atau informasi adalah hal yang sangat penting untuk dijaga keamanan atau kerahasiannya agar pihak lain yang tidak berkepentingan tidak bisa mengetahui data maupun informasi tersebut. Salah satu cara untuk mengamankan data adalah menggunakan teknik kriptografi yaitu ilmu yang mempelajari teknik-teknik matematika yang berhubungan dengan aspek keamanan informasi seperti kerahasiaan, integritas data, serta otentikasi [1]. Implementasinya dapat menggunakan banyak metode enkripsi dari enkripsi klasik hingga enkripsi modern.

Gronsfeld Cipher adalah salah satu algoritma kriptografi yang dapat digunakan. *Gronsfeld Cipher* adalah algoritma enkripsi yang menggunakan kunci numerik dan tabel dalam proses enkripsi dan dekripsi, jika teks yang akan dienkripsi atau didekripsi lebih panjang dari kunci yang digunakan, kunci yang digunakan akan diulang dari kiri. Teks yang telah disandikan akan menjadi teracak dan tidak memiliki makna, namun penggunaan teknik kriptografi belum cukup untuk melindungi data, karena teks terenkripsi masih ditampilkan walaupun dalam bentuk simbol yang tidak beraturan [2]. Jika teks yang telah disandikan didapatkan oleh orang lain tentu akan menjadi sebuah ancaman karena teks yang telah disandikan tersebut dapat dianalisa oleh orang yang mengerti tentang kriptografi atau bahkan hanya sekedar mengacak pesan tersebut. Oleh karena itu, diperlukan teknik lain yang dapat digabungkan dengan teknik kriptografi.

Steganografi adalah salah satu teknik yang dapat digabungkan dengan kriptografi, ini adalah teknik menyembunyikan data rahasia pada suatu media sehingga keberadaan data rahasia tidak dapat dideteksi oleh orang lain [3]. Cara ini digunakan untuk menyembunyikan teks yang telah disandikan agar tidak menimbulkan kecurigaan terhadap orang lain, karena secara visual tidak ada yang berubah

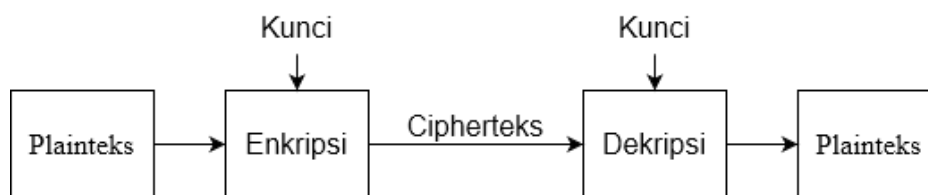
dari media yang ditumpanginya. Salah satu metode steganografi yang dapat digunakan adalah metode *End Of File* (EOF). Metode *End Of File* menyisipkan data terenkripsi ke dalam nilai akhir file gambar, yang hanya memperbesar ukuran file dan memiliki garis tambahan di akhir file gambar sehingga kualitasnya tidak banyak berubah dari citra aslinya.

Berdasarkan permasalahan tersebut, penulis bermaksud untuk mengembangkan sebuah aplikasi yang dapat mengamankan data dengan menggabungkan teknik kriptografi menggunakan algoritma *Gronsfield Cipher* dan teknik steganografi *End Of File* dengan cara informasi atau pesan rahasia dalam bentuk teks dienkripsi terlebih dahulu menggunakan teknik kriptografi, kemudian file yang telah dienkripsi tersebut akan disisipkan ke dalam sebuah gambar dengan menggunakan teknik steganografi. Aplikasi ini dikembangkan untuk menjadi solusi meningkatkan keamanan data pada informasi-informasi rahasia sehingga pihak yang tidak berkepentingan tidak dapat mengetahui isi dari data yang disimpan.

2. Metode Penelitian

2.1 Kriptografi Algoritma *Gronsfield Cipher*

Kriptografi adalah ilmu yang mempelajari teknik matematika yang berkaitan dengan aspek keamanan informasi seperti kerahasiaan, integritas data dan otentikasi [3]. Kriptografi terdiri dari dua proses utama: yaitu enkripsi dan dekripsi. Dalam kriptografi, pesan asli disebut *plaintext*, sedangkan pesan yang telah dienkripsi atau disandikan disebut *ciphertext*.



Gambar 1. Skema Proses Enkripsi dan Dekripsi

Pada Gambar 1 ditunjukkan skema dari proses enkripsi dan dekripsi, di mana pada proses enkripsi, plaintexts akan dienkripsi menggunakan kunci yang dimasukkan dan akan mendapatkan ciphertexts, begitu juga pada proses dekripsi, ciphertexts yang dimiliki akan didekripsi menggunakan kunci yang sesuai dan akan menghasilkan plaintexts atau pesan asli.

Gronsfield Cipher adalah algoritma kriptografi yang menggunakan suatu kunci numerik dan tabel dalam proses enkripsi dan dekripsi, jika teks yang akan dienkripsi atau dekripsi lebih panjang dari kunci yang digunakan maka kunci yang digunakan akan diulang dari kiri, kunci yang digunakan juga biasanya cukup pendek misalnya 1324. Caranya adalah dengan mengubah huruf menjadi bilangan desimal, maka plaintexts hanya terdiri dari susunan angka, bukan huruf. Kemudian enkripsi menggunakan prinsip yang sama dengan Algoritma *Vigenère* yaitu menggunakan tabel yang hanya berukuran 10x10. Rumus enkripsi *Gronsfield Cipher*.

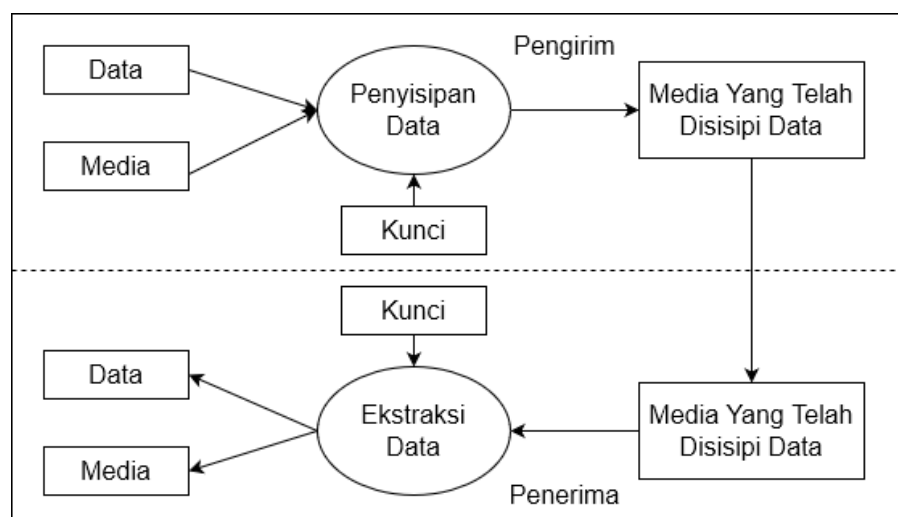
$$C_i = (P_i + K_i) \bmod 26 \quad (1)$$

Rumus dekripsi *Gronsfield Cipher*.

$$P_i = (C_i - K_i) \bmod 26 \quad (2)$$

2.2 Steganografi Metode *End Of File*

Steganografi adalah suatu metode penyembunyian suatu data atau informasi ke dalam sebuah media atau data digital lain yang tidak diduga oleh orang pada umumnya sehingga orang yang melihatnya tidak curiga [4]. Dalam steganografi, informasi atau pesan yang akan disembunyikan disebut dengan *hiddenteks* atau *hidden object*, lalu media yang akan digunakan sebagai penampung pesan disebut *coverteks* atau *cover object*, lalu media yang telah disisipkan pesan disebut *stegoteks* atau *stego object* dan yang terakhir kunci yang digunakan untuk menyisipkan atau mengekstrak pesan disebut *stegokey*



Gambar 2. Skema Proses Steganografi

Pada Gambar 2 ditunjukkan skema proses penyisipan dan ekstraksi. Di mana pada proses penyisipan, data akan disisipkan ke dalam suatu media menggunakan kunci dan akan menghasilkan sebuah *stego-file*. Sedangkan pada proses ekstraksi, *stego-file* yang dimiliki akan diekstraksi menggunakan kunci yang sesuai dan akan menghasilkan data dan media asli sebelum penyisipan.

Metode *End Of File* (EOF) merupakan salah satu teknik yang digunakan dalam steganografi. Teknik ini digunakan dengan cara menambahkan data atau pesan rahasia pada akhir *file* citra. Dalam teknik EOF, data yang disisipkan di akhir diberi tanda khusus sebagai pengenal *start* dari data tersebut dan pengenal akhir dari data tersebut. Dengan metode ini, tidak ada batasan jumlah pesan yang dapat disisipkan. Metode EOF juga tidak mengubah isi *file* yang disisipkan. Ini adalah salah satu keunggulan metode EOF dibandingkan metode lain, karena disisipkan di akhir *file*, pesan yang disisipkan tidak bersinggungan dengan isi *file* yang ditumpangi. Tentu saja, jika ada keunggulan, ada juga kelemahannya, metode EOF ini mengubah ukuran *file* sesuai dengan ukuran pesan yang disisipkan.

2.3 Metode Waterfall

Metode *waterfall* merupakan model pengembangan sistem informasi yang sistematis dan berurutan [5]. Pengembangan aplikasi ini akan menggunakan metode *waterfall*. Metode *waterfall* memiliki beberapa tahapan yaitu analisis, desain, implementasi dan pengujian. Tahap analisis akan menentukan kapabilitas yang harus dimiliki oleh sistem untuk memenuhi apa yang dibutuhkan pengguna dalam menjalankan aplikasi ini.

1. Bahasa pemrograman yang digunakan dalam pengembangan system ini adalah Python.
2. Untuk dapat menjalankan sistem ini, diharuskan untuk menginstall beberapa *library* Python yang digunakan dalam sistem ini.
3. Untuk dapat menjalankan sistem ini, diharuskan untuk memiliki file *user interface* yang telah dibuat dan diletakkan pada direktori yang sama.

Pada tahap desain, kebutuhan yang dibutuhkan dalam pembangunan aplikasi ini adalah sebagai berikut.

a. Halaman *Home*

Halaman home ini memiliki 2 pilihan, penyisipan dan ekstraksi. Pengguna akan dibawa ke halaman penyisipan saat pengguna memilih menu penyisipan, dan sebaliknya pengguna akan dibawa ke halaman ekstraksi saat pengguna memilih menu ekstraksi.

b. Halaman Penyisipan

Pada halaman penyisipan ini, *user* akan melakukan proses penyisipan yang di mana *user* diminta untuk memilih file teks berekstensi .txt yang ingin diamankan dan *file* citra digital berekstensi .png yang akan digunakan sebagai media penyisipan. *User* juga diminta untuk memasukkan kunci yang akan digunakan pada proses enkripsi file teks sebelum disisipkan ke

dalam *file* citra digital. User juga diminta untuk memilih tempat penyimpanan untuk *file* citra digital yang telah disisipi pesan rahasia. Setelah semua dirasa sudah benar, *user* dapat mengklik tombol submit dan proses penyisipan akan dilakukan.

c. Halaman Ekstraksi

Pada halaman ekstraksi ini, *user* akan melakukan proses ekstraksi yang di mana *user* diminta untuk memilih *file* citra digital berekstensi .png yang telah disisipi pesan rahasia. *User* juga diminta untuk memasukkan kunci yang akan digunakan pada proses dekripsi setelah pesan di dalam *file* citra digital berhasil diekstraksi. *User* juga diminta untuk memilih tempat penyimpanan untuk *file* teks yang akan diekstraksi. Setelah semua dirasa sudah benar, *user* dapat mengklik tombol submit dan proses ekstraksi akan dilakukan.

3. Hasil dan Pembahasan

Implementasi pembangunan aplikasi dibuat berdasarkan tahap desain yang akan diterjemahkan ke dalam sebuah program.

3.1. Pengkodean

Implementasi sistem ini adalah aplikasi dapat mengenkripsi pesan, menyisipi pesan ke dalam citra digital, mengekstrak pesan dari dalam citra digital, dan mendekripsi pesan terenkripsi.

a. Tampilan Antarmuka Sistem



Gambar 3. Tampilan Halaman Home

Pada Gambar 3 menunjukkan tampilan halaman *home*. Halaman *home* merupakan tampilan awal dari aplikasi yang dikembangkan. Ada dua menu utama di halaman ini, menu penyisipan yang membawa pengguna ke halaman penyisipan saat ditekan, dan menu ekstraksi yang membawa pengguna ke halaman ekstraksi saat ditekan. Pada halaman ini juga terdapat tombol *exit* yang jika ditekan akan mengarahkan pengguna untuk keluar dari sistem.

Pada Gambar 4 menunjukkan tampilan halaman penyisipan. Halaman penyisipan merupakan tampilan yang akan dilihat pengguna ketika memilih menu penyisipan pada halaman *home*. Pada halaman ini pengguna diminta untuk memasukkan 4 buah masukan yang diperlukan dalam proses penyisipan yaitu *file* teks, citra digital, kunci, dan *path* untuk menyimpan *output* nanti. Pada halaman ini juga terdapat tombol *back* yang dapat digunakan untuk kembali ke halaman *home*.



Gambar 4. Tampilan Halaman Penyisipan



Gambar 5. Tampilan Halaman Ekstraksi

Pada Gambar 5 menunjukkan tampilan halaman ekstraksi. Halaman ekstraksi merupakan tampilan yang akan dilihat pengguna ketika memilih menu ekstraksi pada halaman *home*. Pada halaman ini pengguna diminta untuk memasukkan 3 buah masukan yang diperlukan dalam proses ekstraksi yaitu *stego-image*, kunci dan *path* untuk menyimpan *output* nanti. Pada halaman ini juga terdapat tombol *back* yang dapat digunakan untuk kembali ke halaman *home*.

3.2. Pengujian

Data penelitian yang akan digunakan untuk menguji sistem ini adalah 6 *file* citra digital dan 5 *file* teks. Data tersebut dapat dilihat pada Tabel 1 dan Tabel 2.







Pada Tabel 1, isi dari setiap file teks diambil secara random, yang membedakan adalah jumlah karakter pada setiap filenya, di mana tiap filenya akan ditambahkan karakter sebanyak 700. Dimulai dari file teks 1 memiliki jumlah karakter 700, file teks 2 memiliki jumlah karakter 1400, file teks 3 memiliki jumlah karakter 2100, file teks 4 memiliki jumlah karakter 2800, dan file teks 5 memiliki jumlah karakter 3500.

Pengujian kualitas hasil penyisipan dilakukan untuk mengetahui bagaimana kualitas citra setelah disisipkan pesan rahasia. Pengujian dilakukan dengan menggunakan metode *Peak Signal Noise Ratio* (PSNR) dengan membandingkan citra asli dan citra hasil penyisipan sesuai dengan data penelitian yang telah dijelaskan di atas. Semakin kecil nilai MSE maka semakin mirip citra dengan citra aslinya, di mana setiap piksel berada pada posisi yang sama. Tentunya semakin kecil nilai MSE maka semakin besar nilai PSNR, maka dapat dikatakan semakin besar nilai PSNR dan semakin kecil nilai MSE berarti penyisipan dapat dikatakan berhasil dilakukan dengan baik, dan begitu juga sebaliknya.

Tabel 1. Data Teks

No.	Nama File (.txt)	Panjang Karakter	Ukuran File
1	Teks1	700	699 bytes
2	Teks2	1400	1.36 KB
3	Teks3	2100	2.05 KB
4	Teks4	2800	2.73 KB
5	Teks5	3500	3.41 KB

Tabel 2. Data Citra Digital

No.	Nama File (.png)	Resolusi	Ukuran File	Preview
1	Citra1	1080x720	50 KB	
2	Citra2	760x458	72.6 KB	
3	Citra3	1600x1000	328 KB	
4	Citra4	1920x1080	347 KB	
5	Citra5	2560x1707	505 KB	
6	Citra6	2000x1335	828 KB	

Tabel 3. Hasil Pengujian Kualitas Citra

No	Media Penampung (.png)	Objek Penyisipan (.txt)	MSE	PSNR (dB)
1	Citra1	Teks1	0.001	76.943
		Teks2	0.002	73.924
		Teks3	0.003	72.161
		Teks4	0.005	70.918
		Teks5	0.007	69.917
2	Citra2	Teks1	0.003	73.445
		Teks2	0.006	70.328
		Teks3	0.009	68.589
		Teks4	0.012	67.337
		Teks5	0.015	66.378
3	Citra3	Teks1	0.001	80.173
		Teks2	0.001	77.114
		Teks3	0.002	75.306
		Teks4	0.003	74.071
		Teks5	0.003	73.103
4	Citra4	Teks1	0.001	81.070
		Teks2	0.001	78.025
		Teks3	0.002	76.275
		Teks4	0.002	75.029
		Teks5	0.003	74.057
5	Citra5	Teks1	0.001	84.375
		Teks2	0.001	81.367
		Teks3	0.001	79.579
		Teks4	0.001	78.324
		Teks5	0.001	77.340
6	Citra6	Teks1	0.001	82.414
		Teks2	0.001	79.370
		Teks3	0.001	77.613
		Teks4	0.002	76.369
		Teks5	0.002	75.390

Tabel 3 menunjukkan hasil kualitas penyisipan dengan metode PSNR. Tabel tersebut menampilkan informasi berupa nilai MSE dan PSNR dari perbandingan citra asli dan citra hasil penyisipan.

Nilai PSNR tertinggi dari pengujian ini adalah 84.375 pada Citra5 dan Teks1 sedangkan nilai PSNR terendah dari pengujian ini adalah 66.378 pada Citra2 dan Teks5. Dari pengujian ini didapatkan rata-rata nilai PSNR yaitu 75.543 dB, dan dapat dikatakan citra hasil penyisipan memiliki kualitas yang baik karena citra yang baik adalah citra yang memiliki nilai PSNR di atas 40 dB [6].

4. Kesimpulan

Kualitas hasil penyisipan dengan teknik steganografi metode *End Of File* dapat dikatakan baik. Hal ini dibuktikan dengan nilai PSNR pada kualitas citra hasil penyisipan, di mana nilai rata-rata dari keseluruhan pengujian adalah 75.543 dB. Nilai PSNR yang baik berada di atas 40 dB dan menunjukkan kualitas hasil penyisipan yang baik. Berdasarkan pernyataan sebelumnya, aplikasi ini telah berhasil mencapai tujuan dengan mengamankan pesan teks ke dalam sebuah citra digital dan mengekstraknya kembali tanpa kehilangan data sedikit pun.

Aplikasi ini dapat dikembangkan dengan menambah jenis ekstensi *file* yang dapat diproses agar pengguna tidak perlu mengganti ekstensi filenya ketika ingin menggunakan aplikasi ini. Aplikasi ini juga dapat dikembangkan agar dapat dijalankan pada berbagai sistem operasi, sehingga pengguna dapat menggunakan aplikasi ini dengan lebih mudah.

Daftar Pustaka

- [1] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. Florida: CRC Press, 1996.
- [2] D. Darwis, Wamiliana, and A. Junaidi, "Proses Pengamanan Data Menggunakan Kombinasi Metode Kriptografi Data Encryption Standard dan Steganografi End Of File," no. 978, pp. 228–240, 2017.
- [3] R. Munir, *Kriptografi*. Bandung: Informatika, 2006.
- [4] R. Munir, *Matematika Diskrit*. Bandung: Informatika, 2012.
- [5] G. Wiro Sasmito, "Penerapan Metode Waterfall Pada Desain Sistem Informasi Geografis Industri Kabupaten Tegal," *J. Inform. Pengemb. IT*, vol. 2, no. 1, pp. 6–12, 2017.
- [6] A. Solichin, "Mengukur Kualitas Citra Hasil Steganografi," *Mengukur Kualitas Citra Hasil Steganografi*, no. April, pp. 1–4, 2015.

Aplikasi Identifikasi Nada Darbuka Dengan Onset Detection, MFCC, Dan KNN

Hairul Lana^{a1}, Ngurah Agus Sanjaya^{a2}, I Dewa Made Bayu Atmaja^{a3}, Anak Agung Istri Ngurah Eka Karyawati^{a4}, I Made Widiartha^{a5}, I Gusti Agung Gede Arya Kadyanan^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
¹hairullana99@gmail.com
²agus_sanjaya@unud.ac.id
³dewabayu@unud.ac.id
⁴eka.karyawati@unud.ac.id
⁵madewidiartha@unud.ac.id
⁶gungde@unud.ac.id

Abstract

Darbuka is one of the hadrah musical instruments that acts as a marker when the vocals raise or lower the rhythm of the sound. In learning the Darbuka, the trainer needs to check whether the sound produced is correct or not. With the Darbuka tone recognition system, it will be easier for someone to learn hadrah without a coach. The system developed in this study uses onset detection to break the tone pattern. Then each note goes through a feature extraction process using MFCC with parameters of frame length, overlap length, and the number of coefficients. Then the results of feature extraction through a classification process using KNN. The results of the system test show that the best combination of parameters in the identification of Darbuka tones with a frame length of 30 ms, overlap length of 30%, the number of MFCC coefficients as much as 19 and a value of $K=7$ produces a basic tone identification accuracy of 100%, a tone pattern identification accuracy of 86,67%, and the accuracy of basic tone identification in the tone pattern is 72%.

Keywords: Hadrah, Tone, MFCC, KNN, Onset

1. Pendahuluan

Darbuka adalah salah satu alat musik dari kesenian hadrah yang memiliki bentuk seperti dandang dengan rata-rata diameter sepanjang 22 cm dan terbuat dari bahan aluminium [1]. Darbuka memiliki peranan cukup penting sebagai penanda vokal untuk menaikkan atau menurunkan ritme suara dan berperan sebagai melodi dalam instrumen musik umum. Kata "Darbuka" sendiri berasal dari kata "*daraba*" atau "*doroba*" yang berarti "pukul". Sesuai dengan asal katanya, Darbuka dimainkan dengan cara dipukul. Terdapat empat jenis pukulan Darbuka yaitu "dum", "tak", "ka", dan "slap" dimana pukulan "tak" dan "ka" menghasilkan suara nada yang sama, yang membedakan adalah pukulan "tak" menggunakan tangan kanan sedangkan pukulan "ka" menggunakan tangan kiri.

Pada umumnya dalam mempelajari Darbuka dibutuhkan pelatih untuk mengecek nada yang dihasilkan sudah benar atau tidak, namun tidak semua daerah memiliki kelompok hadrah sehingga ketersediaan pelatih Darbuka juga terbatas. Hal ini mengakibatkan sebagian orang kesulitan untuk berlatih Darbuka. Dengan adanya aplikasi identifikasi nada Darbuka diharapkan dapat mengatasi masalah yang dihadapi oleh sebagian orang yang ingin mempelajari Darbuka tanpa pelatih.

Aplikasi identifikasi nada Darbuka yang dikembangkan pada penelitian ini dapat mendeteksi *onset* sehingga sistem dapat melakukan identifikasi pada sebuah pola nada Darbuka. Sistem identifikasi nada Darbuka menggunakan algoritma *Mel Frequency Cepstral Coefficients* untuk ekstraksi fitur suara sehingga didapatkan suatu ciri dari sebuah file suara, kemudian akan dilakukan klasifikasi berdasarkan fitur suara dengan algoritma *K-Nearest Neighbor*.

2. Metode Penelitian

Metode yang digunakan pada penelitian ini menggunakan metode kuantitatif. Sedangkan metode untuk pembuatan sistem identifikasi nada Darbuka menggunakan algoritma *Mel Frequency Cepstral Coefficients* untuk ekstraksi fitur suara dan algoritma *K-Nearest Neighbor* untuk klasifikasi nada Darbuka.

2.1. Data Penelitian

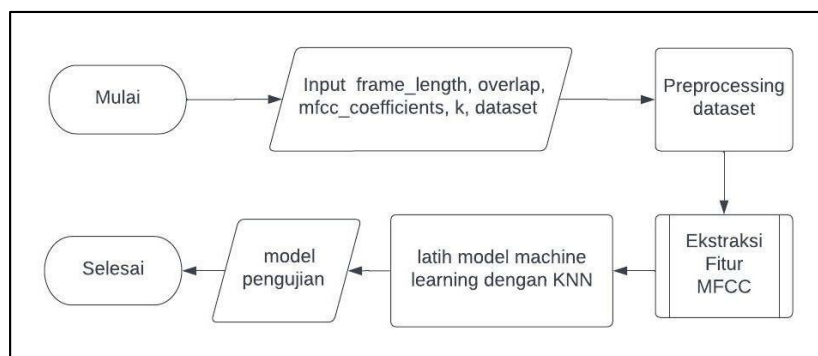
Jenis data yang digunakan pada penelitian ini merupakan data primer yang merupakan hasil perekaman nada Darbuka. Sedangkan dalam pengambilan data yang digunakan adalah metode observasi dimana data diambil secara langsung melalui proses perekaman nada Darbuka yang dilakukan bersama Kelompok Hadrah Masjid Al-Mahdi Kusamba, Kecamatan Dawan, Kabupaten Klungkung. Pengambilan data berupa tiga jenis nada dasar yaitu dum, tak dan slap dengan jumlah setiap nada sebanyak 70 nada. Sedangkan untuk data pola nada yang diambil adalah *baladi*, *maqsum*, dan *sayyidi* sebanyak 10 data tiap pola nada. Jumlah data tiap jenis nada dapat dilihat pada tabel 1.

Tabel 1 Data Penelitian

Jenis Pukulan	Jumlah Data Latih	Jumlah Data Uji
Dum	50	20
Tak	50	20
Slap	50	20
Baladi	-	10
Maqsum	-	10
Sayyidi	-	10

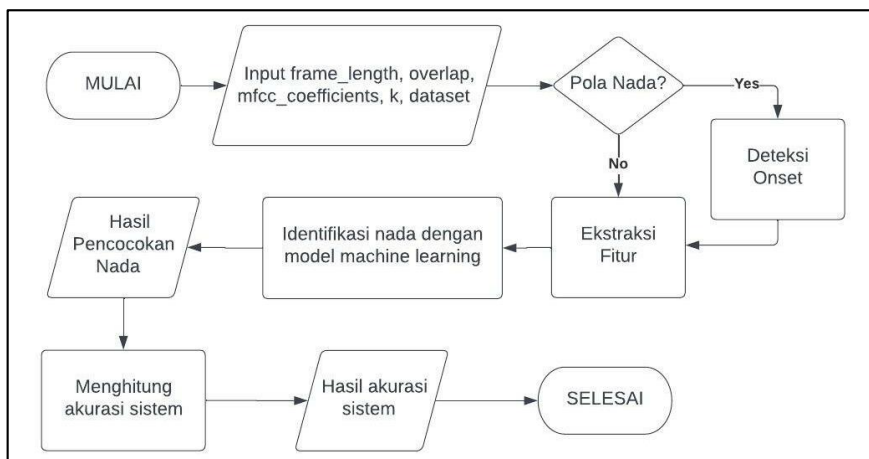
2.2. Desain Sistem

Aplikasi identifikasi nada Darbuka yang dikembangkan pada penelitian ini merupakan aplikasi berbasis website dengan bahasa pemrograman utama Python dengan *Framework* Django dengan tampilan antarmuka menggunakan HTML, CSS dengan *framework* Bootstrap dan Javascript dengan *library* JQuery. Alur sistem terbagi menjadi dua proses utama yaitu tahap pelatihan dan tahap pengujian. Tahap pelatihan adalah tahap untuk mengambil ciri data latih melalui proses ekstraksi fitur suara menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC) kemudian dilakukan proses pelatihan menggunakan *K-Nearest Neighbor* sehingga didapatkan suatu model *machine learning* untuk proses identifikasi nada Darbuka. *Flowchart* tahap pelatihan dapat dilihat pada gambar 1.



Gambar 1 Flowchart Tahap Pelatihan

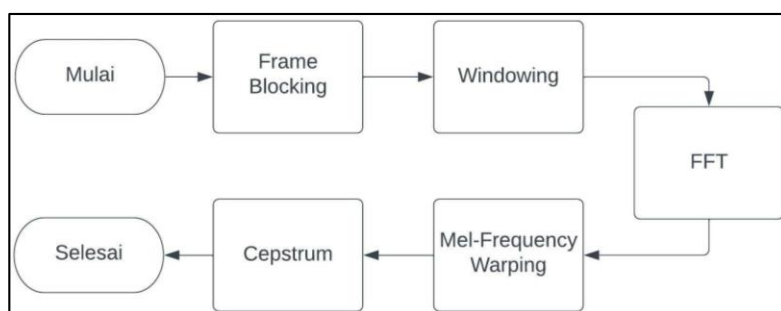
Tahap pengujian adalah tahap untuk menguji apakah data uji sudah benar atau tidak. Berbeda dengan tahap pelatihan, tahap pengujian akan melewati proses deteksi *onset* jika data uji berupa pola nada, proses deteksi *onset* berfungsi untuk mensegmentasikan tiap pukulan dan memecahnya menjadi beberapa pukulan atau nada dasar. Tiap nada dasar akan ekstraksi menggunakan algoritma MFCC, kemudian hasil ekstraksi fitur akan dilakukan identifikasi nada menggunakan algoritma KNN menggunakan model yang sudah dibuat. Kemudian sistem akan memberikan luaran apakah data uji benar atau tidak, selain itu sistem juga akan memberikan informasi akurasi yang dihasilkan sistem. *Flowchart* tahap pengujian dapat dilihat pada gambar 2.



Gambar 2 Flowchart Pengujian

Jika data uji berupa pola nada, maka sistem akan melalui proses deteksi onset untuk memecahnya menjadi nada dasar. Deteksi onset adalah kemampuan untuk melakukan deteksi awal dari suatu *note* dan durasi dari setiap *note* dapat dihitung [5]. Deteksi onset pada penelitian ini bertujuan mendeteksi awal suatu *note*, karena alat musik Darbuka adalah salah satu alat musik dari kesenian Hadrah dan cara memainkannya dengan cara dipukul. Secara umum, deteksi onset akan melalui proses *preprocessing*, *reduction*, dan *peak picking*. Tahap *preprocessing* berfungsi memecah sinyal kepada beberapa pita frekuensi dan memisahkan bagian *transient/steady state*. Tahap *reduction* merupakan kunci dalam proses deteksi *onset*. Metode *reduction* dipecah menjadi dua kelompok yaitu metode *reduction* berdasarkan penggunaan fitur sinyal suara yang sudah ditetapkan secara eksplisit (amplitude sinyal, perubahan frekuensi, dan fase) dan metode *reduction* berdasarkan model probabilistik sinyal [6]. Proses terakhir adalah *peak picking* yang menghasilkan waktu onset dari hadrah, kemudian akan dilakukan pemotongan nada sesuai dengan waktu onsetnya dengan cara memilih fitur yang lebih besar dari ambang sesuai dengan onset.

Ekstraksi fitur suara adalah proses yang bertujuan mendapatkan vektor ciri dari sebuah suara. Pada penelitian ini ekstraksi fitur suara dilakukan menggunakan algoritma MFCC yang didasarkan pada persepsi sistem pendengaran manusia. Secara umum MFCC mempunyai lima proses yaitu *frame blocking*, *windowing*, *FFT*, *mel-frequency wrapping*, dan *cepstrum* [2]. Flowchart ekstraksi fitur menggunakan algoritma MFCC dapat dilihat pada gambar 3.

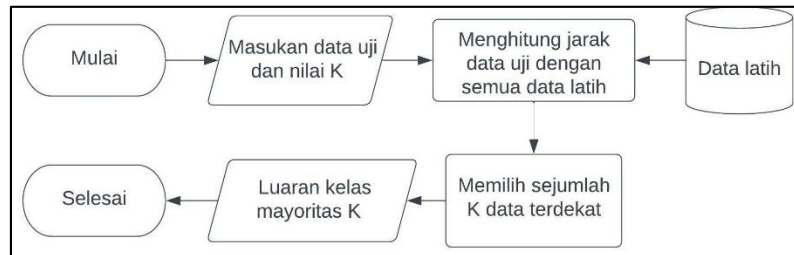


Gambar 3 Flowchart MFCC

Proses pertama MFCC adalah *frame blocking* di mana sinyal dipotong menjadi beberapa bagian yang disebut dengan *frame*. Proses *frame blocking*, sebuah sinyal suara akan dibagi menjadi beberapa potong *frame* yang saling bertumpuk atau *overlap*, proses ini berfungsi untuk meminimalisasi sinyal yang hilang (*deletion*) dimana proses terus berulang sampai semua sinyal masuk ke dalam *frame* [3]. Selanjutnya tiap *frame* akan melalui proses *windowing* untuk meminimalisasi kebocoran spektral yang terjadi karena rendahnya jumlah *sampling rate* atau karena proses *frame blocking* yang menyebabkan sinyal menjadi berhenti [4]. Proses selanjutnya adalah *fast fourier transform* (FFT) merupakan metode analisis *fourier*. Analisis *fourier* adalah metode yang berguna untuk melakukan analisa kepada sinyal yang dimasukkan berupa *spectrogram*. Proses selanjutnya adalah *mel-frequency warping* yang merupakan filter untuk mengetahui ukuran dari sebuah energi dari *frequency band* tertentu dalam sinyal

suara. Proses terakhir adalah *cepstrum* yaitu sebuah teknik untuk meningkatkan kualitas pengenalan sinyal suara.

K-Nearest Neighbor yaitu algoritma klasifikasi yang sederhana, karena KNN akan mencari label dari data latih dengan jarak terpendek sejumlah K. Konsep dasar KNN seperti pada algoritma *Nearest Neighbor*, yaitu dengan mencari jarak terdekat antara data latih dengan data uji [7]. *Flowchart* KNN dapat dilihat pada gambar 4.



Gambar 4 Flowchart KNN

Langkah pertama algoritma KNN adalah menentukan nilai K, penentuan nilai K sangat penting karena akan sangat mempengaruhi tingkat akurasi dari proses klasifikasi. Langkah selanjutnya menghitung jarak antara data latih ke data uji. Ada beberapa persamaan dalam menghitung jarak seperti *Manhattan Distance*, *Minkowsky Distance*, *Chebychev Distance*, dan *Euclidean Distance*. Perhitungan jarak pada penelitian ini menggunakan persamaan *Euclidean Distance*. Penerapan dari rumus *Euclidean Distance* adalah dengan mengakarkan nilai dari variabel data latih dikurangi dengan nilai variabel data uji yang sudah dipangkatkan dengan dua. Jika terdapat lebih dari satu variabel, maka akumulasikan pemangkatan dua yang sebelumnya sudah dilakukan pengurangan data latih dikurangi dengan data uji. Persamaan *Euclidean Distance* dapat dilihat pada persamaan 1.

$$D_{(x,y)} = \sqrt{\sum_{k=1}^n (y_k - x_k)^2} \quad (1)$$

(Prasetyo, Suta Wijaya and Yudo Husodo, 2019)

Keterangan :

$D(x,y)$ = hasil *eucledian distance*

x = data uji

y = data latih

k = variabel data

n = jumlah data latih

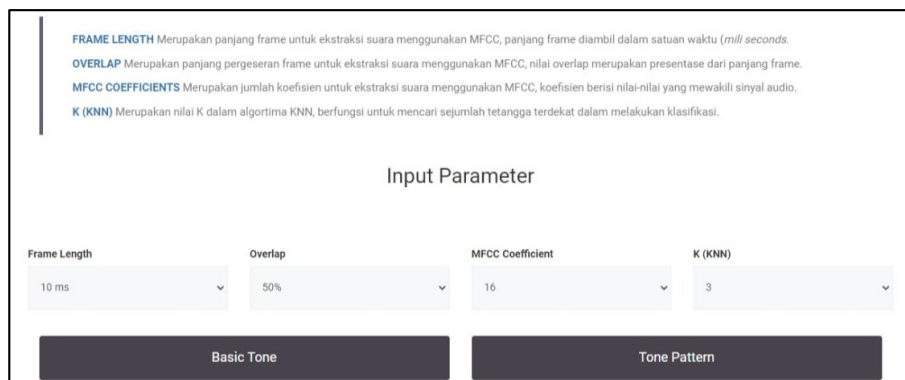
Setelah jarak ditentukan, selanjutnya adalah mencari jarak terdekat antara data uji dengan seluruh data latih dari yang jarak yang terdekat sampai dengan jarak yang terjauh. Kemudian diambil sejumlah K dengan nilai jarak terdekat dengan data uji. Dari nilai K tersebut akan dikelompokkan sesuai dengan mayoritas K sehingga sistem dapat menentukan apakah suatu nada alat musik Darbuka sudah benar atau tidak.

3. Hasil dan Diskusi

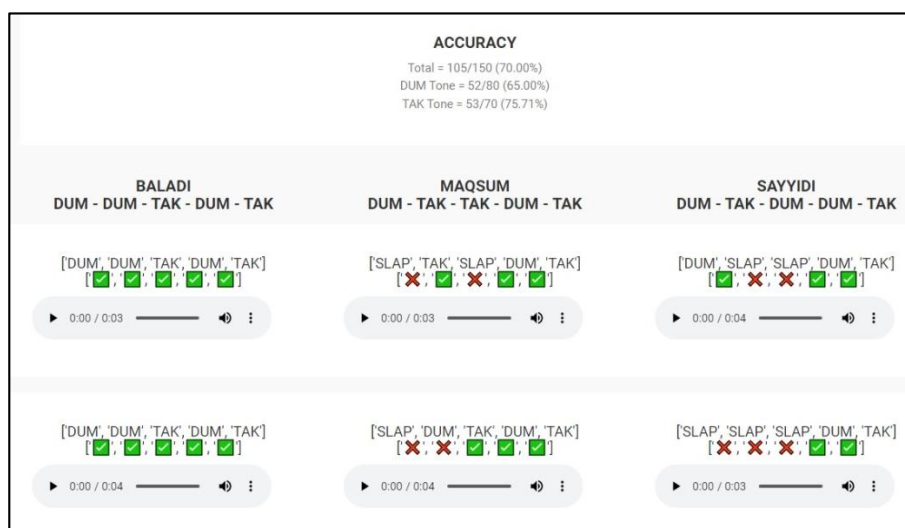
3.1. Tampilan Awal Aplikasi

Aplikasi identifikasi nada Darbuka memiliki tiga halaman yaitu halaman awal, halaman pengujian dan halaman identifikasi. Halaman awal berisi informasi mengenai kesenian Hadrah yang bertujuan untuk memberikan pemahaman kepada pengguna mengenai kesenian Hadrah terlebih alat musik Darbuka. Sedangkan halaman pengujian adalah halaman untuk melakukan identifikasi seluruh data uji yang ada pada sistem sehingga didapatkan akurasi sistem dalam mengidentifikasi nada Darbuka. Pada halaman pengujian, pengguna diminta memasukkan parameter yang digunakan dalam ekstraksi fitur menggunakan MFCC yaitu panjang *frame*, panjang *overlap*, dan jumlah koefisien MFCC, selain itu pengguna juga diminta untuk memasukkan nilai K untuk proses klasifikasi menggunakan KNN karena halaman pengujian dikhususkan untuk *developer* dalam mencari kombinasi parameter dengan akurasi terbaik. Hasil dari proses pengujian menampilkan hasil identifikasi dari semua data uji yang ada pada

sistem serta akurasi yang didapatkan. Dalam halaman pengujian terdapat dua tombol yaitu “*Basic Tone*” untuk pengujian pada nada dasar dan “*Tone Pattern*” untuk pengujian pada pola nada. Tampilan halaman pengujian dapat dilihat pada gambar 5 dan gambar 6.

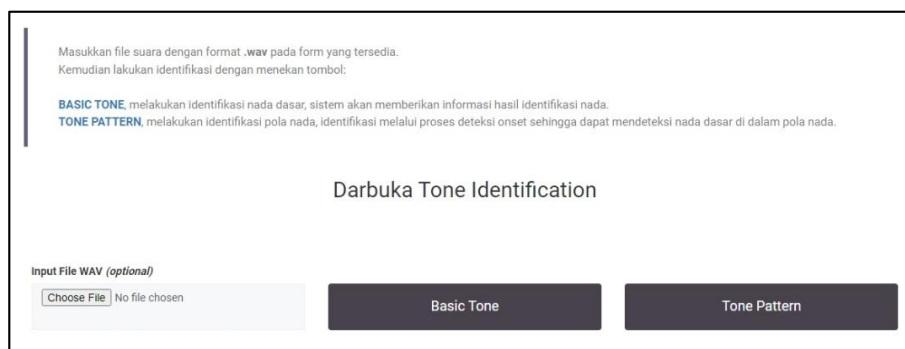


Gambar 5 Antarmuka Halaman Pengujian

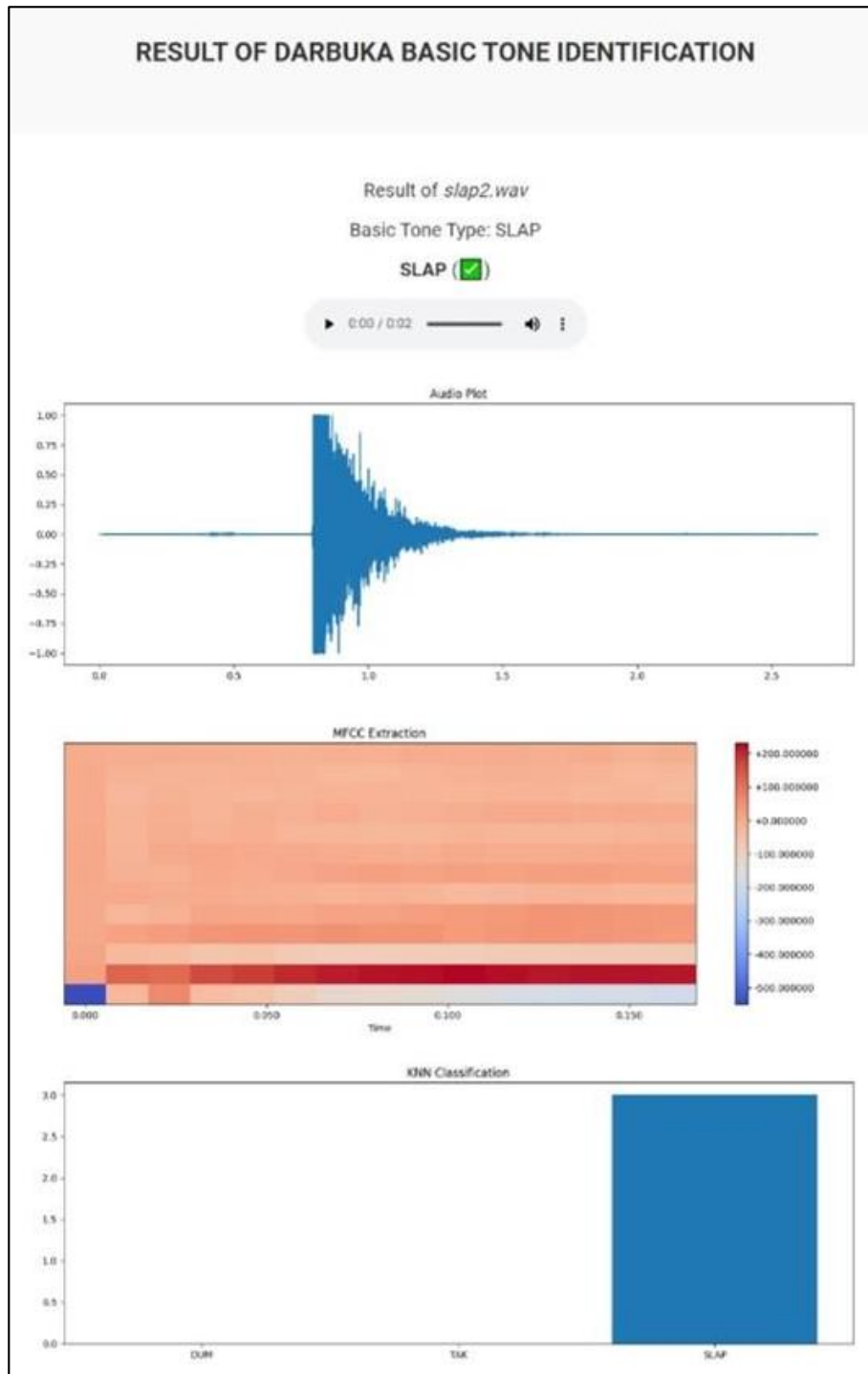


Gambar 6 Antarmuka Hasil Pengujian

Halaman ketiga adalah halaman identifikasi yang berfungsi untuk mengidentifikasi sebuah nada dari file yang dimasukkan oleh pengguna. Pada halaman identifikasi, pengguna tidak perlu memasukkan parameter seperti pada halaman pengujian karena pada identifikasi pada halaman ini menggunakan model terbaik dan dikhususkan untuk pengguna umum. Sama seperti halaman pengujian, halaman identifikasi akan terdapat dua tombol yaitu “*Basic Tone*” untuk pengujian pada nada dasar dan “*Tone Pattern*” untuk pengujian pada pola nada. Tampilan halaman pengujian dapat dilihat pada gambar 7 dan gambar 8.



Gambar 7 Antarmuka Halaman Identifikasi



Gambar 8 Antarmuka Hasil Identifikasi

3.2. Hasil Analisis Pengaruh Panjang *Frame* dan Panjang *Overlap* Terhadap Akurasi

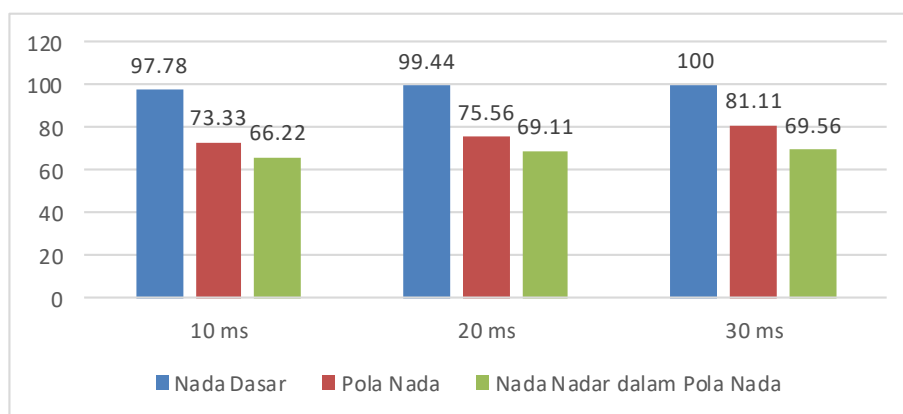
Panjang *frame* dan *overlap* merupakan variabel yang menentukan jumlah *frame* dalam proses ekstraksi fitur MFCC. Pembagian sampling menjadi beberapa *frame* ini terjadi pada proses *frame blocking*. Pengujian ini menggunakan kombinasi panjang *frame* 10, 20 dan 30 dengan *overlap* yang digunakan adalah sebesar 30%, 40%, dan 50%. Sedangkan untuk parameter lain menggunakan parameter terbaik

yaitu jumlah koefisien MFCC sebanyak 19 dan nilai $K=7$. Hasil pengujian pengaruh panjang frame dan panjang *overlap* dapat dilihat pada tabel 2.

Tabel 2 Analisis Pengaruh Panjang Frame dan Overlap

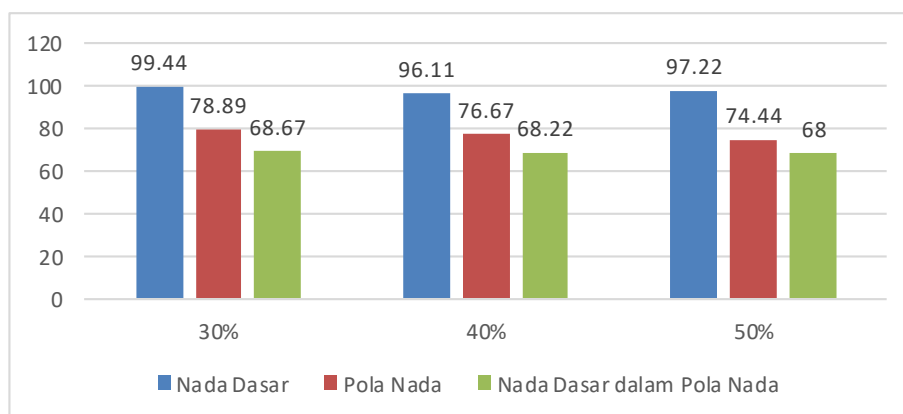
Panjang Frame	Panjang Overlap	Akurasi Identifikasi Nada Dasar	Akurasi Identifikasi Pola Nada	Akurasi Identifikasi Nada Dasar Dalam Pola Nada
10 ms	30%	93,33%	73,33%	66,00%
	40%	88,33%	73,33%	66,00%
	50%	93,33%	73,33%	66,67%
20 ms	30%	100,00%	76,67%	68,00%
	40%	100,00%	76,67%	70,00%
	50%	98,33%	73,33%	69,33%
30 ms	30%	100,00%	86,67%	72,00%
	40%	100,00%	80,00%	68,67%
	50%	100,00%	76,67%	68,00%

Gambar grafik pengaruh panjang *frame* dan dapat dilihat pada gambar 9.



Gambar 9 Grafik Pengaruh Panjang Frame

Melalui grafik pada gambar 9, penggunaan panjang *frame* 30 ms menghasilkan akurasi terbaik untuk identifikasi nada dasar dengan rata-rata akurasi 100%, identifikasi pola nada dengan rata-rata akurasi 81,11%, dan identifikasi nada dasar dalam pola nada terbaik dengan rata-rata akurasi 69,56%. Kemudian gambar grafik pengaruh panjang *overlap* dan dapat dilihat pada gambar 10.



Gambar 10 Grafik Pengaruh Panjang Overlap

Melalui grafik pada gambar 10, penggunaan panjang *overlap* 30% menghasilkan akurasi terbaik untuk identifikasi nada dasar dengan rata-rata akurasi 99,44%, identifikasi pola nada dengan rata-rata akurasi 78,89%, dan identifikasi nada dasar dalam pola nada dengan rata-rata akurasi 68,67%.

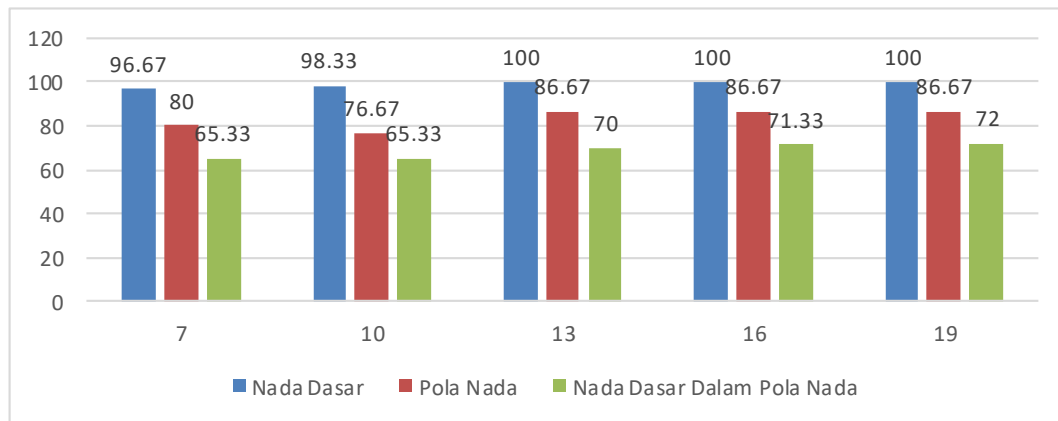
3.3. Hasil Analisis Pengaruh Jumlah Koefisien MFCC Terhadap Akurasi

Koefisien MFCC merupakan jumlah koefisien pada proses *discrete cosine transform* yang terdapat pada proses MFCC. Pengaruh dari koefisien MFCC dapat diketahui dengan menjadikan koefisien MFCC sebagai variabel bebas. Koefisien MFCC yang diuji adalah 7, 10, 13, 16, dan 19. Sedangkan untuk parameter lain menggunakan parameter terbaik yaitu panjang *frame* 30 ms, panjang *overlap* 30%, dan nilai $K=7$. Hasil pengujian pengaruh koefisien MFCC terhadap akurasi identifikasi nada dapat ditampilkan pada tabel 3.

Tabel 3 Analisis Pengaruh Jumlah Koefisien

Jumlah Koefisien MFCC	Akurasi Identifikasi Nada Dasar	Akurasi Identifikasi Pola Nada	Akurasi Identifikasi Nada Dasar Dalam Pola Nada
7	96,67%	80,00%	65,33%
10	98,33%	76,67%	65,33%
13	100,00%	86,67%	70,00%
16	100,00%	86,67%	71,33%
19	100,00%	86,67%	72,00%

Gambar grafik pengaruh panjang *overlap* dan dapat dilihat pada gambar 11.



Gambar 11 Grafik Pengaruh Panjang Overlap

Melalui grafik pada gambar 11, penggunaan jumlah koefisien MFCC sebanyak 13, 16 dan 19 menghasilkan akurasi terbaik untuk identifikasi nada dasar dengan akurasi 100% dan identifikasi pola nada dengan akurasi 86,67%. Sedangkan jumlah koefisien MFCC sebanyak 19 menghasilkan akurasi terbaik untuk identifikasi nada dasar dalam pola nada dengan akurasi 72%.

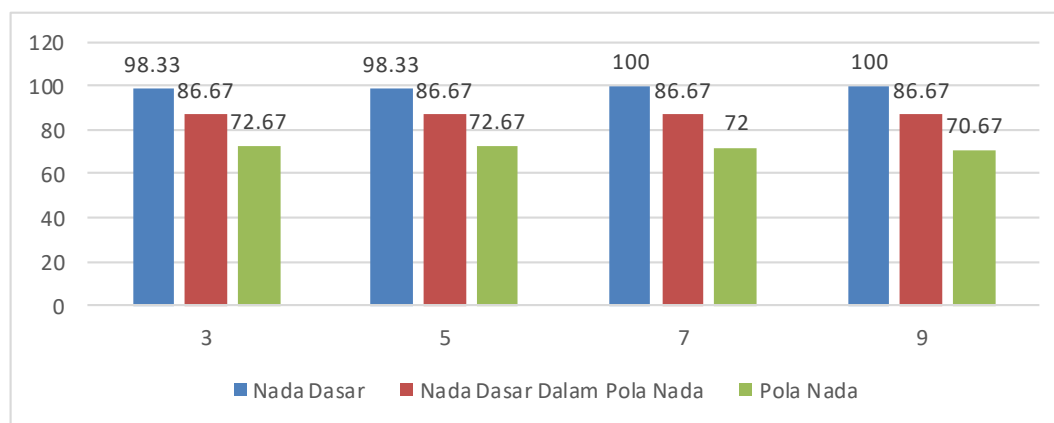
3.4. Analisis Jumlah K Terhadap Akurasi

Jumlah K pada algoritma *K-Nearest Neighbor* sangat mempengaruhi akurasi dalam identifikasi nada alat musik Darbuka, hal ini karena jumlah K mewakili jumlah tetangga terdekat yang merupakan data latih yang menjadi acuan dalam melakukan identifikasi sehingga jumlah K dalam penelitian ini menjadi variabel bebas. Jumlah K yang diuji adalah 1, 3, 5, 7, dan 9. Sedangkan untuk parameter lain menggunakan parameter terbaik yaitu panjang *frame* 30 ms, panjang *overlap* 30%, dan jumlah koefisien MFCC sebanyak 19. Hasil pengujian pengaruh jumlah K terhadap akurasi identifikasi nada dapat ditampilkan pada tabel 4.

Tabel 4 Analisis Pengaruh Jumlah K

K	Akurasi Identifikasi Nada Dasar	Akurasi Identifikasi Pola Nada	Akurasi Identifikasi Nada Dasar Dalam Pola Nada
3	98,33%	86,67%	72,67%
5	98,33%	86,67%	72,67%
7	100,00%	86,67%	72,00%
9	100,00%	86,67%	70,67%

Gambar grafik pengaruh jumlah K dan dapat dilihat pada gambar 12.



Gambar 12 Grafik Pengaruh Jumlah K

Melalui grafik pada gambar 12, penggunaan nilai K=7 dan K=9 menghasilkan akurasi terbaik untuk nada dasar sebesar 100%, sedangkan penggunaan nilai K menghasilkan akurasi yang sama untuk identifikasi pola nada dengan akurasi 86,67%, dan penggunaan nilai K=3 dan K=5 menghasilkan akurasi terbaik untuk identifikasi nada dasar dalam pola nada dengan akurasi 72,67%.

4. Kesimpulan

Berdasarkan implementasi penelitian yang telah dilakukan dan hasil yang telah didapat, maka dapat ditarik kesimpulan sebagai berikut:

1. Penggunaan panjang *frame* 30 ms menghasilkan akurasi terbaik untuk identifikasi nada dasar dengan akurasi 100%, identifikasi pola nada dengan akurasi 81,11%, dan identifikasi nada dasar dengan akurasi 69,56%. Sedangkan penggunaan panjang *overlap* 30% menghasilkan akurasi terbaik untuk identifikasi nada dasar dengan akurasi 99,44%, identifikasi pola nada dengan akurasi 78,89%, dan identifikasi nada dasar dalam pola nada dengan akurasi 68,67%. Sehingga dapat dikatakan semakin besar panjang *frame* dan semakin kecil panjang *overlap* maka semakin besar akurasi yang didapat.
2. Jumlah koefisien MFCC memiliki korelasi yang kuat terhadap tingkat akurasi identifikasi nada dasar alat musik Darbuka. Korelasi yang bersifat positif di mana semakin besar nilai koefisien MFCC yang digunakan akan semakin besar nilai akurasi yang didapatkan. Pada penelitian ini penggunaan panjang *frame* 13, 16, dan 19 menghasilkan akurasi tertinggi pada identifikasi nada dasar dengan akurasi 100% dan identifikasi pola nada dengan akurasi 86,67%, sedangkan pada identifikasi nada dasar dalam pola nada menghasilkan akurasi terbaik sebesar 72% dengan jumlah koefisien MFCC sebesar 19.
3. Jumlah K pada algoritma *K-Nearest Neighbor* sangat mempengaruhi tingkat akurasi dalam identifikasi nada alat musik Darbuka karena jumlah K mewakili jumlah tetangga terdekat yang merupakan data latih yang menjadi acuan dalam melakukan identifikasi. Nilai K=7 dan K=9 menghasilkan akurasi terbaik identifikasi nada dasar dengan akurasi 100%. Kemudian untuk identifikasi pola nada memiliki akurasi yang sama untuk semua nilai K yang diujikan dengan akurasi 86,67%. Sedangkan untuk identifikasi nada dasar dalam pola nada menghasilkan akurasi terbaik sebesar 72,67% dengan penggunaan K=3 dan K=5.

4. Aplikasi identifikasi nada alat music Darbuka menggunakan MFCC, dan KNN sudah berjalan dengan akurasi yang cukup tinggi sampai dengan 100% untuk identifikasi nada dasar Darbuka, namun masih perlu pengembangan lebih lanjut terlebih dalam identifikasi pola nada Darbuka dengan akurasi yang terbilang kecil, sehingga dapat membantu orang-orang dalam belajar alat musik Darbuka.

References

- [1] N. Marufah, "Komunikasi Seni Hadrah Majelis Ahbaabul Musthofa Yogyakarta," *Alhadharah J. Ilmu Dakwah*, vol. 19, no. 2, pp. 91–116, 2020.
- [2] N. Nurhamidah, E. C. Djamal, and R. Ilyas, "Perintah Menggunakan Sinyal Suara dengan Mel-Frequency Cepstrum Coefficients dan Learning Vector Quantization," *Semin. Nas. Apl. Teknol. Inf. 2017*, pp. 11–16, 2017.
- [3] R. Umar, I. Riadi, and A. Hanif, "Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequency Cepstral Coefficients (MFCC)," *CogITo Smart J.*, vol. 4, no. 2, p. 294, 2019, doi: 10.31154/cogito.v4i2.130.294-304.
- [4] F. Muhammad, Y. Indrawaty, and I. Amelia, "Identifikasi Nada Antara Suling Sunda Dan Suling Rekorder Dengan Menggunakan Metode Mel Frequency Cepstral Coefficients (Mfcc) Dan Dynamic Time Warping (Dtw)," *JTIK J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 145–154, 2020, doi: 10.25126/jtiik.202071649.
- [5] A. Suryarismi and R. Pulungan, "Penyusunan Notasi Musik dengan Menggunakan Onset Detection pada Sinyal Audio," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, p. 167, 2013, doi: 10.22146/ijccs.3357.
- [6] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A Tutorial on Onset Detection in Music Signals," *Speech Audio ...*, pp. 1–13, 2005, [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1495485.
- [7] A. F. Ryamizard, B. Hidayat, and S. Saidah, "Deteksi Nada Tunggal Alat Musik Kecapi Bugis Makassar Menggunakan Metode Mel Frequency Cepstral Coefficient (Mfcc) Dan Klasifikasi K-Nearest Neighbour (Knn)," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 4715–4721, 2018.

Classification of Typhus and Dengue Fever Using the Pseudo Nearest Neighbor Algorithm

I Dewa Ngurah Tri Hendrawan^{a1}, Ida Bagus Gede Dwidasmara^{a2}, I Gusti Agung Gede Arya Kadyanan^{a3}, I Putu Gede Hendra Suputra^{a4}, AAIN Eka Karyawati^{a5}, Ida Bagus Made Mahendra^{a6}

^{a1}Informatics Department, Faculty of Mathematics and Natural Sciences, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹dewahendrawan99@gmail.com

²dwidasmara@unud.ac.id

³gungde@unud.ac.id

⁴hendra.suputra@unud.ac.id

⁵eka.karyawati@unud.ac.id

⁶ibm.mahendra@unud.ac.id

Abstract

Typhus and dengue fever are diseases that often occur in Indonesia. The spread of these two diseases is relatively fast with similar symptoms. This could be a fatal thing if there is a misdiagnosis. Therefore, an application was developed to assist in the classification of typhus and dengue fever based on the patient's clinical symptoms using the PNN (Pseudo Nearest Neighbor) algorithm. This application receives input in the form of clinical symptoms experienced by the patient, then a preprocessing process is carried out to convert user input into discrete data, and the results are processed in classification using the PNN method. From the validation process with 5-fold cross validation obtained the best k value is k=6. Then, the accuracy testing process concluded that the accuracy of the classification process for typhus and dengue fever with the PNN method is 68.97%. Then, from 25 respondents in the user acceptance test obtained that 88.4% of respondents strongly agree with the application design, 87.6% respondents strongly agree with the ease of application, and 86.6% respondents strongly agree with the efficiency provided by the application.

Keywords: Classification, Pseudo Nearest Neighbor, Typhus, Dengue Fever, User Acceptance Test

1. Introduction

Typhus and dengue fever are public health problems that often occur in Indonesia. The spread of the virus from these two diseases is relatively fast, so the number of people with typhus or dengue fever tends to increase every year. In addition, the symptoms possessed by typhus or dengue fever also have similarities with each other. This can be fatal to the patient if an error occurs in the diagnosis. Therefore, a classification system was built on whether the patient had dengue fever or typhus based on the clinical symptoms experienced by the patient.

In this study, classification is done by applying Pseudo Nearest Neighbor in the process of diagnosing patients according to their clinical symptoms. Pseudo Nearest Neighbor (PNN) is a variant of the kNN algorithm that improves accuracy on small data. The PNN algorithm uses a combination of the weighted k-NN and the k-NN closest local average method [1]. PNN classifies by calculating the total distance between the test data and the k closest neighbors in each class by weighting proportionally based on the distance, then the class with the smallest total distance will be used as the class label [2]. In a study conducted by [1], it was discussed about the application of the PNN algorithm for the credit score classification of borrowers at banks. In this study, the smallest error range produced by the PNN algorithm was 20.75% with a value of k = 13.

Based on the research above, a classification system for typhus or dengue fever was built based on the clinical symptoms experienced by patients using the PNN (Pseudo Nearest Neighbor) algorithm. This research is hoped can help doctors and medical personnel to diagnose diseases based on clinical symptoms experienced by patients.

2. Research Methods

The research begins with the literature study to find related references for this study. After that, it continued with the process of collecting data in the form of disease medical record data. Then based on the literature study, an application was built to classify typhus and dengue fever. The system that has been built is tested using the specified test method. And then the results of the testing process will be used as a conclusion in this study.

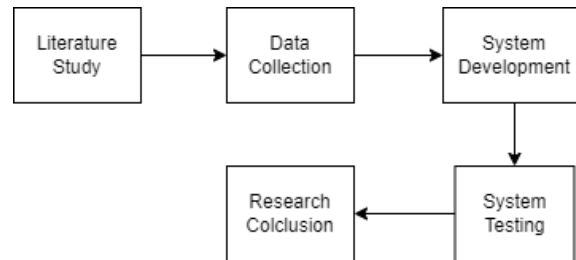


Figure 1. Research Flow

2.1. Research Data

This research used primary data in the classification process. The data collected for this research is clinical symptoms of patients with typhus, dengue fever, and several diseases that have similar symptoms to typhus and dengue fever. Then the data collected was divided into 80% of training data and 20% of test data.

2.2. Preprocessing Data

After all the data is collected, the research data was converted into discrete data using the ordinal encoding method. Ordinal encoding is a method used to convert categorical data into numeric data. In ordinal coding, each category is converted to an integer. This will have no effect on the number of categories, but only implies the order of the categories [3]. The purpose of this stage is to adjust the data so that it can be used in the classification process with the Pseudo Nearest Neighbor (PNN) algorithm.

2.3. K-Fold Cross Validation

K-Fold Cross Validation is one of the most used validation methods. It divides the data to be training data and testing data[4]. Training data is used to learn the data's pattern, and the testing data is used to validate the model. The data is divided into similar amounts and then trained and tested in the k iteration, as in Figure 2. Researchers commonly used k=5 or k=10 to get a not-biased model [5].

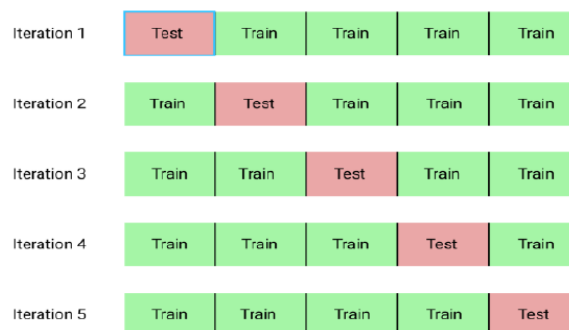


Figure 2. Illustration of 5-Fold Cross Validation

2.4. Pseudo Nearest Neighbor Algorithm

Pseudo Nearest Neighbor (PNN) is a variant of K-Nearest Neighbor which is developed from the Weighted k-Nearest Neighbor (WkNN) and local mean-based k-Nearest Neighbor (LMkNN) [1]. PNN calculates the total distance between k nearest neighbors in each class and test data with proportional weighting based on distance and concludes with the closest distance as the class for the test data [2]. PNN calculates the total distance from the test data with the number of k nearest

neighbors in each existing class with proportional weighting based on distance, then the class is determined based on the total closest distance obtained.

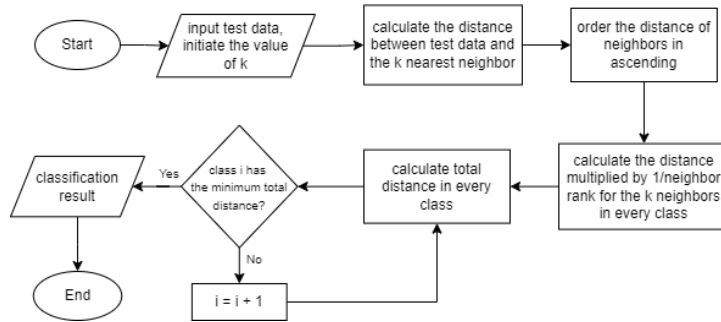


Figure 3. Classification with PNN

Based on the classification in Figure 3, a multi-label classification process was carried out to determine whether the patient had typhus, dengue fever, had both diseases, or other diseases. The multi-label classification flow can be seen in Figure 4.

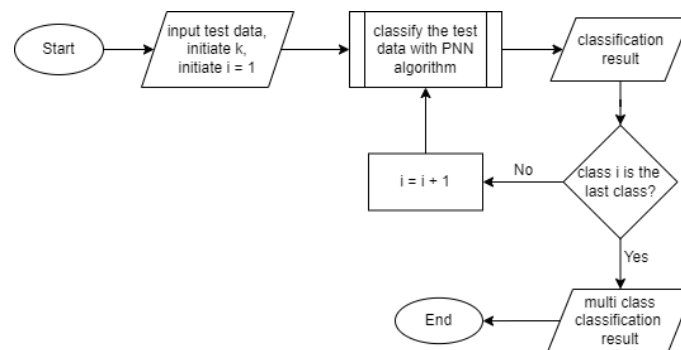


Figure 4. Multi Label Classification

In the multi-label classification, classification process is held twice. First classification to determine whether the test data was classified as typhus or not, and the second process is classification to determine whether or not the test data was included in dengue fever. From each processes the output produced is in the form of a value of 1 or 0 for each label being tested. The output in each label will be recorded, so, the conclusions obtained in the form [1,1] for patients with typhus fever and dengue, [1,0] or [0,1] for patients who have one of the two diseases, and [0,0] for patients who do not suffer from typhus or dengue fever.

2.5. System Development Method

The system was developed using the waterfall model. This model was used because the system requirements are clear, the possibility of changes in system requirements is small, and this model only requires small resources in its implementation [6]. Waterfall method include analysis, design, implementation, testing, and maintenance.

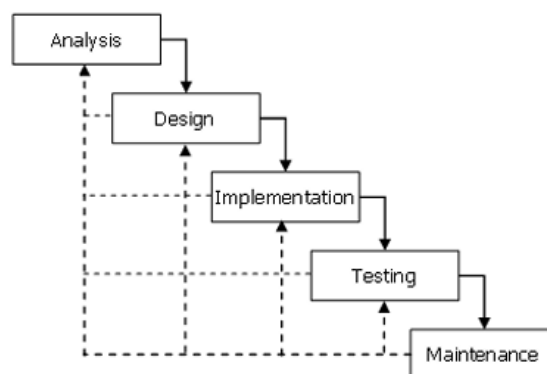


Figure 5. Waterfall Model

a. Analysis

The first step in waterfall model is analysis phase. Analysis phase is a description of the behavior of application to be developed. This step describes functional and non-functional requirements.

1. Functional requirements

Functional requirements are the requirements that will be implemented into the system. The functional requirements that will be made are as follows:

Table 1. Functional Requirements

No	Requirements	User
1	Login and logout	Admin & Petugas
2	Admin can add new petugas, delete petugas, and access list of petugas	Admin
3	Use disease prediction feature and access the result	Petugas
4	Access patient list and patient details	Petugas

2. Non-functional requirements

The non-functional requirements that will be made are as follows:

Table 2. Non-functional Requirements

No	Requirements
1	Only the registered users who can access the system
2	System can be accessed 24 hours
3	Web based system

b. System Design

The Unified Model Language (UML) was used to design the system. UML allows developers to easily make the systems blueprint, so it is easy to understand and it could be an effective way to communicate the system design to the others [7]. UML has several diagrams for the developers, some of them are use case diagram, activity diagram, and sequence diagram.

1. Use Case Diagram

Use case diagram is a representation of functional requirements, so the system will be easily developed. In another word, a use case diagram is an explanation of system function from the user's side [8]. So, use case diagram can describe the interaction that happens between the user (actor) and the system.

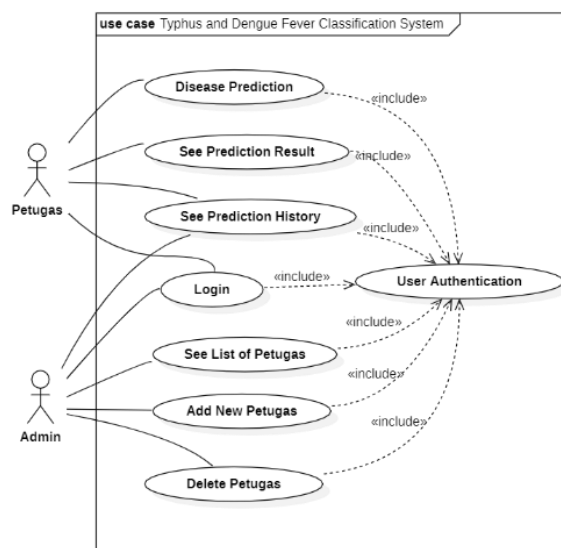


Figure 6. Use Case Diagram

2. Activity Diagram

An activity diagram is a description of the activity flows that happen in the system[7]. It is also describing how the activities started, the decision that might happen, and how it ended.

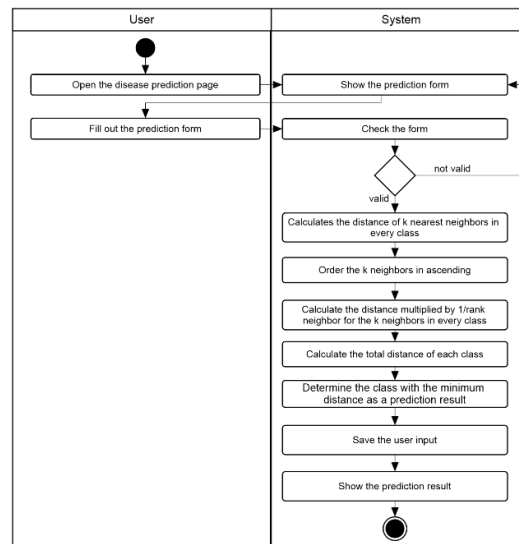


Figure 7. Activity Diagram

3. Sequence Diagram

Sequence diagram is an interaction diagram that describes the order of the system events. So, in the sequence diagram, the interaction between object/class has been described and it also indicates the communication that happens in the system [8].

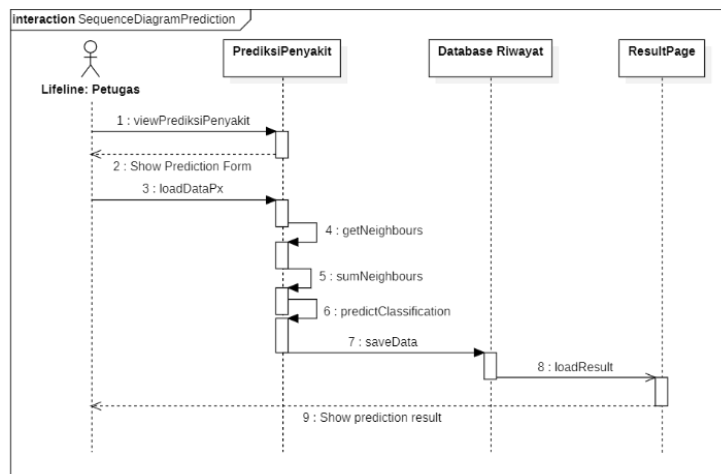


Figure 8. Sequence Diagram

2.6. Testing

a. Accuracy Testing

Accuracy testing is carried out to find out how much accuracy is produced by the classification system for typhus or dengue fever with the PNN algorithm [9]. The accuracy rate is obtained by dividing the correct predictions by the number of the test data, the result obtained would be in the form of a percentage. To calculate the level of system accuracy, the following formula is used:

$$P(PN) = \frac{Pn}{n} \times 100\% \quad (1)$$

Explanation:

- P(PN) = Accuracy rate
- Pn = The number of classification results by the system that deserves to be true
- n = Number of data tested

b. User Acceptable Test (UAT)

User Acceptance Test is carried out to find out what the system was doing, and to know the system assessment based on the end-user's point of view [10]. In this study, end-users are asked to use the system that had been built, then end-users are asked to answer a questionnaire regarding user ratings of the system with the linkert scale [11].

Table 3. Linkert Scale

Statement	Scale
Strongly Disagree	1
Disagree	2
Neutral	3
Agree	4
Strongly Agree	5

Then the questionnaire result is calculated as follows:

$$P = \frac{\sum f \times Ls}{n \times 5} \times 100\% \quad (2)$$

Explanation:

P = Percentage

Ls = Linkert Scale

f = frequent of the answer to be choose

n = total respondent

After the percentage calculation, the conclusion of UAT process obtained as Table 4 [11].

Table 4. UAT Result

Percentage	Conclusion
0% - 20%	Strongly Disagree
21% - 40%	Disagree
41% - 60%	Neutral
61% - 80%	Agree
81% - 100%	Strongly Agree

3. Result and Discussion

3.1. Data Collection

This study used primary data in the form of clinical symptom data of typhus patients, dengue fever patients, and several other diseases in 2020 and 2021 which were obtained from the patient's medical record book at Buleleng District Public Hospital. From the data collection process, 290 data were obtained in the form of categorical and numerical data. Then the clinical symptoms data of each patient will be processed to be used in predicting the patient's disease.

3.2. Preprocessing

After the data has been collected, the preprocessing stage then carried out. Attributes with categorical data type are converted into discrete data, using the LabelEncoder method found in the Scikit Learn library. From the results of preprocessing obtained datasets that are ready to be used in the classification process with the PNN method.

Lama Demam	Suhu	Kes	Nafsu Makan	Pusing	Mual	Muntah	Batuk	Sembelit	...	Perut kembung	Mimisan	Nyeri Kepala	Nyeri Otot	Nyeri Sendi	Manifestasi Pendarahan	Ruam Kulit	Lidah Kotor	TF	DHF
4	37.2	cm	baik	Tidak	Ya	Tidak	Ya	Tidak	...	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak
7	37.5	cm	turun	Tidak	Ya	Ya	Ya	Tidak	...	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak
23	39.7	cm	turun	Tidak	Tidak	Tidak	Tidak	Tidak	...	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak
7	36.7	cm	turun	Tidak	Ya	Ya	Tidak	Tidak	...	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Ya	Tidak
5	38.8	semi koma	turun	Tidak	Tidak	Tidak	Tidak	Tidak	...	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak

↓

Lama Demam	Suhu	Kes	Nafsu Makan	Pusing	Mual	Muntah	Batuk	Sembelit	...	Perut kembung	Mimisan	Nyeri Kepala	Nyeri Otot	Nyeri Sendi	Manifestasi Pendarahan	Ruam Kulit	Lidah Kotor	TF	DHF	
4	37.2	1	0	0	1	0	1	0	...	0	0	0	0	0	0	0	0	0	1	0
7	37.5	1	1	0	1	1	1	0	...	0	0	0	0	0	0	0	0	0	1	0
23	39.7	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1	0
7	36.7	1	1	0	1	1	0	0	...	0	0	0	1	0	0	0	0	0	1	0
5	38.8	2	1	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0	1	0

Figure 9. Preprocessing

Figure 9 shows the result of preprocessing process. All of the categorical data were converted into discrete data, so the data could be used in the calculation of PNN algorithm.

3.3. Model Validation with K-Fold Cross Validation

This research implemented 5-fold cross validation in the validation model process with 232 training data. To validate the model, 232 training data were divided into 5 similar sets and then were trained and tested in 5 iterations. To obtained the optimal accuracy, this research tested the value of k=1 until k=10.

Table 5.5-Fold Cross Validation Result

K	Accuracy in every iteration (%)					
	1 st	2 nd	3 rd	4 th	5 th	Average
1	19,57	34,78	17,39	23,91	23,91	23,91
2	21,74	23,91	19,57	17,39	21,74	20,87
3	23,91	19,56	26,09	21,74	17,39	21,74
4	21,74	28,26	17,39	26,09	17,39	22,17
5	34,78	21,74	32,6	21,74	28,26	27,83
6	28,26	30,43	19,56	32,6	30,43	28,26
7	34,78	26,08	26,08	32,6	13,04	26,52
8	17,39	17,39	23,91	17,39	30,43	21,3
9	17,39	17,39	28,26	32,6	23,91	23,91
10	15,21	17,39	19,56	30,43	28,26	22,17

From the Table 5, the optimal accuracy of the PNN is obtained by k=6 with an average accuracy of 28.26%. So that PNN algorithm with parameter k=6 is used as a model for the application in this research.

3.4. System Interface

Implementation of the system interface was done by doing the coding process using HTML, CSS, JavaScript, and the Bootstrap framework. The result of the interface system are petugas login, home page, prediction, prediction result, prediction history, admin login, admin homepage, list of petugas, and add petugas. There are some of the application's interface that has been deployed.

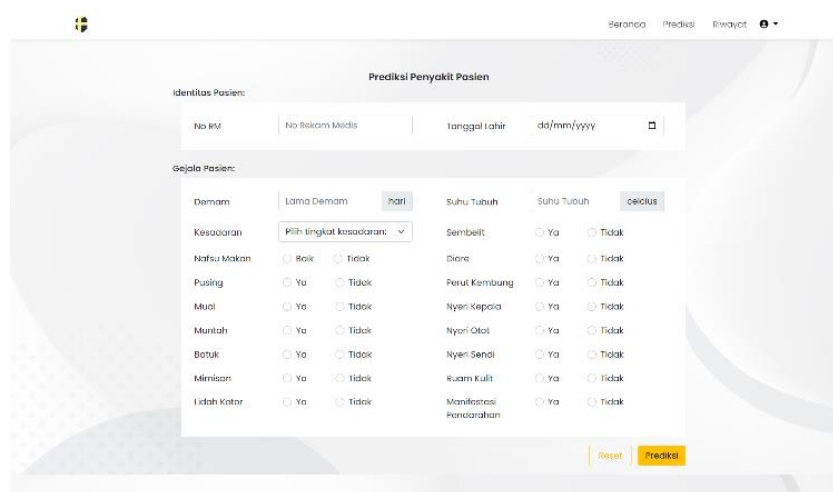


Figure 10. Prediction Interface

Figure 10 is where Petugas do the prediction. In this page, Petugas need to fill patient's identity and symptoms. After all of the columns was filled by Petugas, then Petugas need to click 'Prediksi' button to do the prediction.

Hendrawan, Dwidasmara, Kadyanan, Suputra, Karyawati, and Mahendra
 Development of Typhus and Dengue Fever Detection Applications Using the Pseudo Nearest Neighbor Algorithm

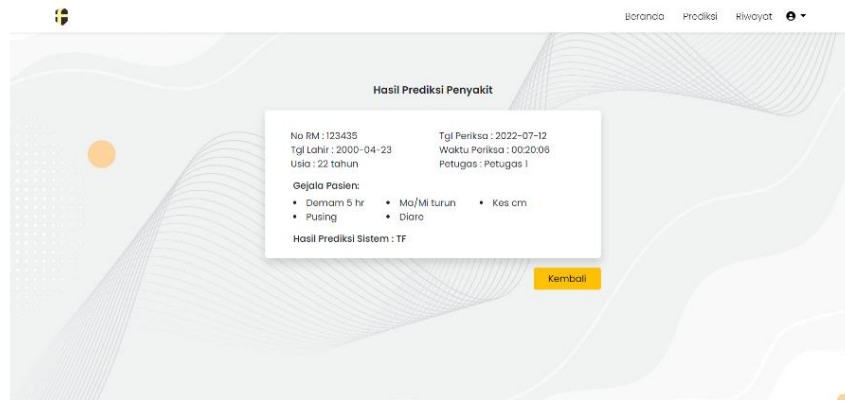


Figure 11. Prediction Result Interface

Prediction result page contains information about the prediction result done by Petugas. In this application there are four types of output, first 'TF' which means typhus, the second one is 'DHF' which means dengue fever, then there is 'Penyakit Lain' which is mean the patient neither has typhus or dengue fever, and the last one is 'TF dan DHF' which is mean that the patient suffers typhus and dengue fever.

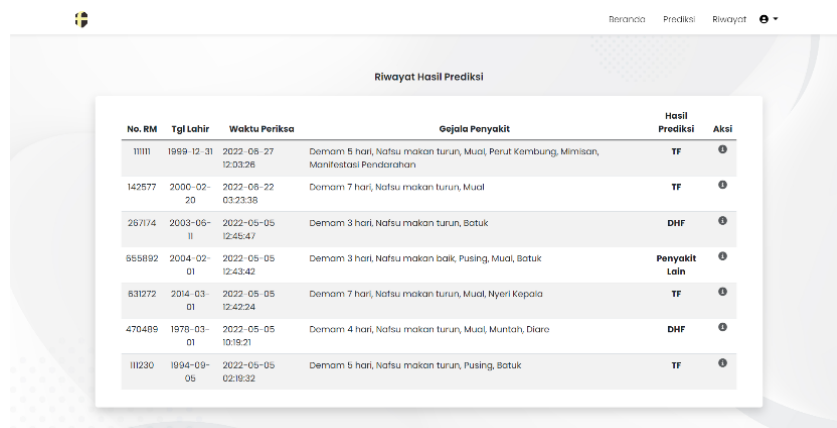


Figure 12. Prediction History Interface

Prediction history page contains all the prediction that has been done by Petugas. From this page, Petugas can access the patient's detail by clicking the button in the right columns.

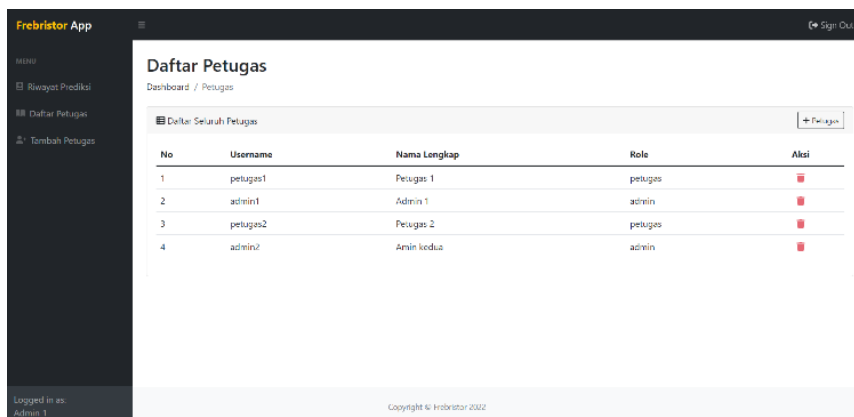


Figure 13. Petugas List Interface

Figure 13 is a page of list Petugas. From this page, Admin can see Petugas information, add new Petugas, and delete Petugas. Admin can click the '+Petugas' button to add new Petugas and has to click the trash icon to delete Petugas.

3.5. Testing Result

a. Accuracy Testing

After the PNN algorithm has been successfully implemented in the system, an accuracy test is carried out to determine how accurate the application can classify patient's diseases. In this process, all of the 290 data that have been collected were divided into 58 test data and 232 training data. This process calculates the result by dividing the total correct predictions of the application by the total test data used. From the testing process, the accuracy obtained from the implementation of PNN with k=6 as the model in the application is 68.97% with total 40 correct predictions by the application.

b. User Acceptance Test (UAT)

User Acceptance Test is carried out to know how users' perception of the application that was developed. In this research, there were 25 participants in the UAT process. The participants of this test are doctors and nurses. In this test, the participants managed to answer a questionnaire in Google Form about their perception of the developed system. There were some statements about the user perception in the questionnaire, which the participants must respond to whether they agree or disagree with the statements, using the rating scale in Table 3. From the UAT process obtained a result as follows:

Table 6. UAT Score

Code	Indicator	Score					Total Score	%
		SD (1)	D (2)	N (3)	A (4)	SA (5)		
P1	Design	0	1	1	10	13	115	92,0%
P2		0	0	6	7	12	106	84,8%
P3	Easiness	0	0	3	8	14	111	88,8%
P4		0	0	4	9	12	108	86,4%
P5	Efficiency	0	0	6	5	14	108	86,4%
P6		0	0	5	6	14	109	87,2%
P7		0	0	3	7	15	112	89,6%
P8		0	0	4	12	9	105	84,0%

Based on the questionnaire's result that has been carried out from 25 respondents, the average percentage value of the three indicators, namely design, easiness, and efficiency are as follows.

Table 7. Conclusion of UAT

Indicator	Percentage (%)	Description
Design	88,4%	Strongly Agree
Easiness	87,6%	Strongly Agree
Efficiency	86,8%	Strongly Agree

In Table 7, it is known that from 25 respondents in the user acceptance test obtained that 88.4% respondents strongly agree with the application design, then 87.6% respondents strongly agree that the application is easy to use, and 86.6% respondents strongly agree with the efficiency that given by the application.

4. Conclusion

This research succeeded in classifying typhus and dengue fever with the Pseudo Nearest Neighbor (PNN) algorithm. This research used a total of 290 data that divided into 232 training data and 58 test data. To get the most optimal result, this research used k-fold cross-validation to validate the best k value in the PNN algorithm. From the 5-fold cross-validation process, the best k in the range k=1 until k=10 is k=6, with an average accuracy of 28.26%. After the application was successfully built, several testing processes were carried out to find out how well the application could work. An accuracy test was carried out with 58 test data and the best accuracy rate obtained from the PNN model in this application was 68.97%. Then, from 25 respondents in user acceptance test obtained that 88.4% of them strongly agree with the application design, 87.6% respondents strongly agree with the ease of application, and 86.6% respondents strongly agree with the efficiency provided by the application.

References

- [1] H. Pratiwi, M. A. Mukid, A. Hoyyi, and T. Widiharah, "Credit scoring analysis using pseudo nearest neighbor," *Journal of Physics: Conference Series*, vol. 1217, p. 012100, 2019, doi: 10.1088/1742-6596/1217/1/012100.
- [2] Suyanto, *Data Mining untuk Klasifikasi dan Klusterisasi Data*, no. May. Bandung: Informatika, 2017.
- [3] K. Potdar, T. S. Pardawala, and C. D. Pai, "A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers," *International Journal of Computer Applications*, vol. 175, no. 4, pp. 7–9, Oct. 2017, doi: 10.5120/IJCA2017915495.
- [4] Y. Widyaningsih, G. P. Arum, and K. Prawira, "APLIKASI K-FOLD CROSS VALIDATION DALAM PENENTUAN MODEL REGRESI BINOMIAL NEGATIF TERBAIK," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, pp. 315–322, Jun. 2021, doi: 10.30598/BAREKENGVOL15ISS2PP315-322.
- [5] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation pada Data Pasien Pengidap Penyakit Jantung," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [6] A. K. A. Gumawang and A. Rakhmadi, "Pengembangan Sistem Perancangan Manajemen Usaha Kecil Menengah Bidang Kuliner dengan Metode Swot," in *The 7th University Research Colloquium 2018*, 2018, pp. 159–170. Accessed: May 20, 2022. [Online]. Available: <http://repository.urecol.org/index.php/proceeding/article/view/30/27>
- [7] M. T. Prihandoyo, "Unified Modeling Language (UML) Model Untuk Pengembangan Sistem Informasi Akademik Berbasis Web," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 1, pp. 126–129, Jan. 2018, doi: 10.30591/JPIT.V3I1.765.
- [8] Y. R. L. Kelen and B. J. Belalawe, "IMPLEMENTASI MODEL-VIEW-CONTROLLER (MVC) PADA UJIAN ONLINE MELALUI PENERAPAN FRAMEWORK CODEIGNITER," 2018.
- [9] R. Kohavi and F. Provost, "Glossary of terms. Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process.," *Machine Learning*, vol. 30, pp. 271–274, 1998, doi: 10.1023/A:1017181826899.
- [10] D. W. Utomo, D. Kurniawan, and Y. P. Astuti, "TEKNIK PENGUJIAN PERANGKAT LUNAK DALAM EVALUASI SISTEM LAYANAN MANDIRI PEMANTAUAN HAJI PADA KEMENTERIAN AGAMA PROVINSI JAWA TENGAH," *Simetris: Jurnal Teknik Industri, Mesin, Elektro dan Ilmu Komputer*, vol. 9, no. 2, pp. 731–746, Nov. 2018, doi: 10.24176/SIMET.V9I2.2289.
- [11] B. Priyatna, A. L. Hananto, and M. Nova, "Application of UAT (User Acceptance Test) Evaluation Model in Minggon E-Meeting Software Development," *Systematics Journal*, vol. 2, no. 3, pp. 110–117, Dec. 2020, Accessed: Jul. 09, 2022. [Online]. Available: <https://journal.unsika.ac.id/index.php/systematics/article/view/4947>

Implementasi Metode *Convolutional Neural Network* Untuk Pengenalan Pola Motif Kain Tenun Rote Ndao Berbasis Android

Derwi Rainord Nalle^{a1}, Luh Gede Astuti^{a2}, I Gede Santi Astawa^{a3}, Luh Arida Ayu Rahning Putri^{a4},
Anak Agung Istri Ngurah Eka Karyawati^{a5}, I Wayan Supriana^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan alam, Universitas Udayana
Badung, Bali, Indonesia

¹rainord.nalle@student.unud.ac.id

²g.astuti@unud.ac.id

³santi.astawa@unud.ac.id

⁴rahningputri@unud.ac.id

⁵eka.karyawati@unud.ac.id

⁶wayan.supriana@unud.ac.id

Abstract

Rote ndao Ikat Weaving has interesting characteristics in each fabric motif because it has different motifs which indicate the ethnic differences contained in each of the resulting motifs. Rote Ndao weaving has a variety of motifs that are still unknown to many people, so in this study a classification of motifs of rote ndao woven fabrics was carried out using the Convolutional Neural Network method. Weaving motif classification uses 3 motifs with a total of 1050 data including 70 data for Ai Bunak, Dula Kakaik and Lafa Langgak motifs each. Data for 3 fabric motifs is divided into 80% training data and 20% testing data. The classification of the Rote Ndao Weaving motif is carried out by building an architectural model of MobileNetV2 plus Dropout and using a Learning Rate of 0.0003 which is trained and evaluated using K-Fold Cross Validation with a value of K=5, obtaining an accuracy of 93%; The precision is 94% for the Ai Bunak motif, 88% for the Dula Kakaik motif, and 100% for the Lafa Langgak motif. Recall of 87% for the Ai Bunak motif, 94% for the Dula Kakaik motif and 100% for the Lafa Langgak motif. Then the F-Score value obtained is 90% for the Ai Bunak motif, 91% for the Dula Kakaik motif and 100% for the Lafa Langgak motif.

Keywords: Convolutional Neural network, MobileNetV2, Pattern Recognition, Weaving Rote Ndao

1. Pendahuluan

Tenun Ikat Rote Ndao merupakan karya budaya yang memiliki makna, nilai sejarah, dan teknik yang signifikan baik dari segi warna, motif, maupun bahan atau benang yang digunakan. Kain tenun rote ndao memiliki ciri yang menarik pada setiap motif kain karena memiliki perbedaan motif yang menunjukkan adanya perbedaan etnis yang terdapat pada masing-masing motif yang dihasilkan [1]. Seiring berkembangnya zaman, banyak kain tenun Rote Ndao yang mengalami modifikasi, sehingga memiliki motif yang lebih beragam dan membuat generasi muda bahkan masyarakat luas kesulitan untuk mengenali motif tenun rote ndao yang beredar luas. Untuk itu perlu adanya sebuah aplikasi android yang dapat dimanfaatkan untuk melakukan Pengenalan pola motif tenun Rote Ndao tanpa dibatasi ruang dan waktu yang dapat mempermudah generasi muda dan masyarakat luas untuk mendapatkan informasi tentang kain tenun Rote Ndao.

Pengenalan pola merupakan subbidang kecerdasan buatan yang berhubungan dengan deteksi pola dalam data. Pengelompokan otomatis data numerik dan simbolik, termasuk gambar, dicapai melalui pengenalan pola (dalam hal ini komputer) dengan menggunakan metode klasifikasi yang dapat mengelompokkan serta memudahkan dalam mengidentifikasi suatu objek pada suatu citra ke dalam satu kelas tertentu [2]. Objek yang dilihat manusia dapat dikenali karena otak manusia telah belajar mengklasifikasikan objek yang ditemuinya di alam agar dapat membedakan satu objek dengan objek lainnya. Mesin komputer mencoba meniru sistem visual manusia dengan cara Mesin Komputer

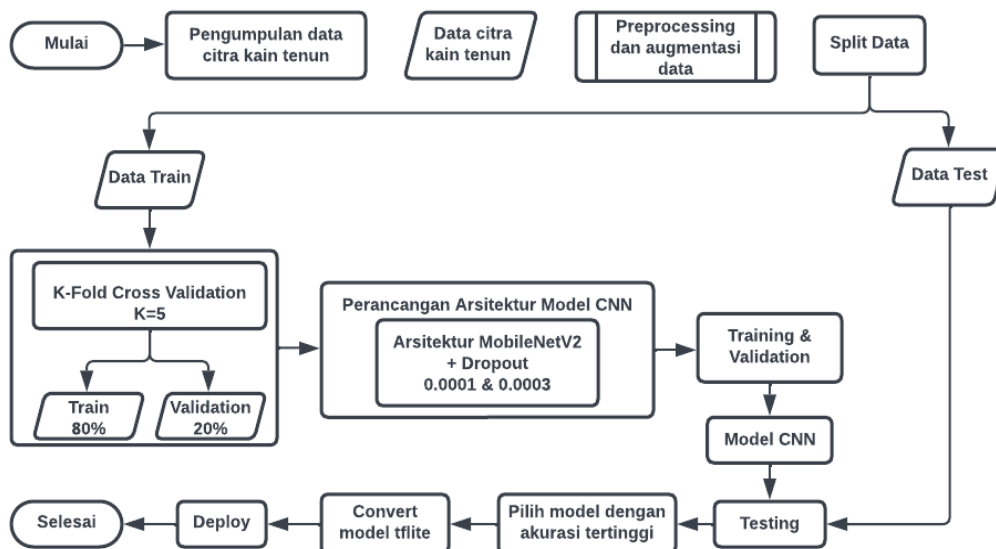
menerima input gambar objek untuk diidentifikasi, memprosesnya, dan mengeluarkan deskripsi objek dalam gambar [3].

Pada penelitian [4], dalam kasus pengenalan motif kain songket menggunakan metode *Convolutional Neural Network* menghasilkan tingkat akurasi yang mencapai 93% dengan penggunaan Dropout, memberikan efek regularisasi, yang meningkatkan akurasi data pengujian, dan momentum dengan nilai 0,9 mengurangi waktu pelatihan dua kali lipat. Berbeda dengan convolution layer CNN pretrain yang telah dilatih dengan dataset yang besar untuk menghasilkan akurasi 100% untuk klasifikasi songket. Menurut [5] Metode CNN adalah Metode *Deep Learning* yang memiliki hasil paling signifikan didalam pengenalan citra. Karena kemampuannya yang signifikan dalam memodelkan data kompleks seperti gambar dan suara, *Deep Learning* telah menjadi topik hangat diskusi di dunia *Machine Learning*. Hal ini disebabkan oleh upaya metode CNN untuk meniru pengenalan gambar di korteks visual yang sama dengan manusia agar mereka dapat memproses informasi yang sama.

Berdasarkan masalah diatas dan meninjau dari penelitian - penelitian yang telah dilakukan sebelumnya, maka pada penelitian ini dilakukan pengenalan motif kain tenun rote ndao menggunakan metode *Convolutional Neural Network* dan pada tahap pelatihan akan digunakan *K-Fold Cross Validation* untuk melatih dan memvalidasi model yang dibangun dengan $K=5$. Selanjutnya digunakan *Hyperparameter* berupa *Learning Rate* dengan nilai 0.0001 dan 0.0003, sehingga *output* yang dihasilkan untuk setiap *fold* ($2 \times 5\text{fold} = 10$ model CNN), setelah itu masing-masing model tersebut akan diuji terhadap data uji, dievaluasi, dipilih model dengan akurasi tinggi, di*convert* dan akhirnya di-*deploy* ke aplikasi android.

2. Metode Penelitian

Gambar 1 menunjukkan Alur Umum penelitian Proses klasifikasi motif kain tenun Ndao terdiri dari beberapa tahapan. yakni, pertama, dataset citra tenun dikumpulkan langsung dari lapangan. Kedua, preprocessing dan augmentasi kumpulan data untuk memastikan bahwa data gambar memenuhi kriteria standar untuk masuk ke arsitektur CNN. Augmentasi gambar digunakan untuk menambah jumlah gambar. Ketiga, kumpulan data akan dibagi menjadi data pelatihan dan pengujian, dengan data pelatihan digunakan untuk melatih model dan data pengujian digunakan untuk menguji model. Keempat, data train (pelatihan) akan dibagi menjadi 80% training data dan 20% validasi data. Tujuan dari pembagian data ini adalah untuk mengevaluasi model menggunakan K-fold cross-validation dengan $K=5$. Kelima, akan dilakukan perancangan arsitektur CNN. Pada penelitian ini akan digunakan MobileNetV2 dengan Dropout. Arsitektur akan dilatih dan divalidasi terhadap dataset. Selain itu, digunakan *hyperparameter* sebagai *learning rate* dengan nilai 0,0001 dan 0,0003, yang akan menghasilkan satu model untuk setiap fold (total $2 \times 5\text{fold} = 10$ model CNN), dan model tersebut akan diuji terhadap data testing. Setelah melakukan pengujian dan mendapatkan hasil akurasi model testing maka kemudian akan dipilih model dengan akurasi tertinggi untuk di *convert* ke model *tf lite* yang akan digunakan untuk *deploy* ke aplikasi android. Berikut adalah Alur penelitian dari sistem klasifikasi tenun yang akan dibangun.



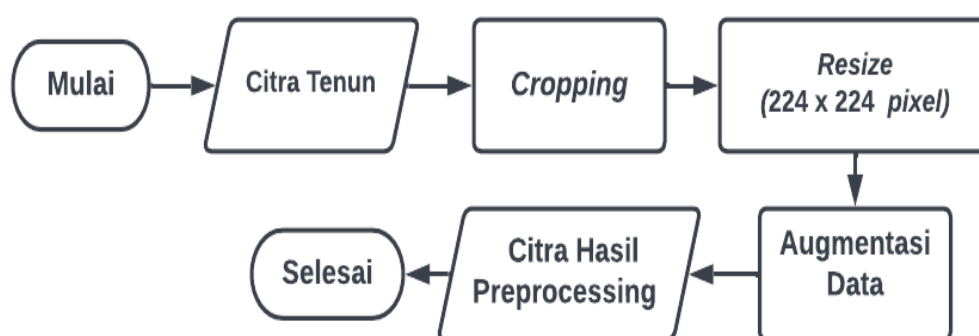
Gambar 1. Alur Umum penelitian

2.1 Dataset

Pada penelitian ini digunakan data berupa citra dari 3 motif Kain Tenun yang berbeda. Tiga motif kain tenun tersebut adalah Ai unak, dula kakaik, dan lafa langgak. Citra input berupa citra warna RGB (Red, Green, Blue) dengan ukuran 224 x 224 piksel. Total data yang digunakan sebanyak 1.050 data, dengan rincian 350 data untuk setiap motifnya. Data tersebut kemudian akan dipecah menjadi data pelatihan dan pengujian, dengan 80% pelatihan dan 20% pengujian. Data pelatihan kemudian dibagi menjadi 80% data pelatihan dan 20% validasi data. Tujuan dari pembagian data ini adalah untuk mengevaluasi model menggunakan K-Fold Cross Validation dengan K = 5.

2.2 Preprocessing

Gambar 2 menunjukkan alur tahapan *preprocessing*. Pada penelitian ini setiap data citra akan di crop sehingga hanya terlihat ciri khasnya saja. setelah melakukan proses cropping, citra kemudian di reziise kedalam dimensi yang berukuran 224 x 224 piksel dan dilakukan peningkatan (augmentasi) dataset untuk memastikan bahwa data citra memenuhi kriteria standar untuk masuk ke dalam arsitektur CNN.



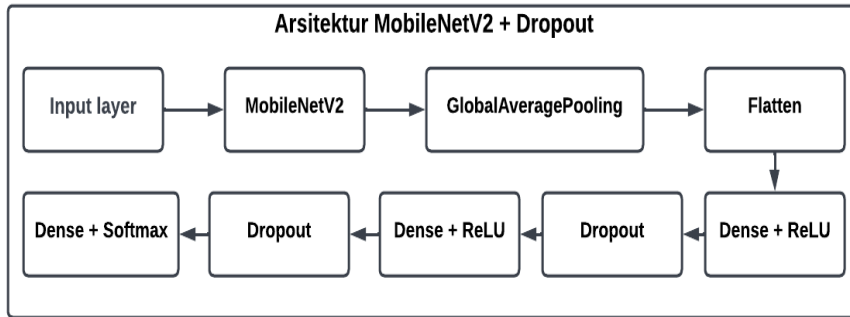
Gambar 2. Alur Tahapan *Preprocessing*

2.3 Convolutional Neural Network

Convolutional Neural Networks (CNN) adalah jenis multilayer perceptron (MLP) yang dirancang untuk memproses gambar. Pada Penelitian ini digunakan salah satu jenis arsitektur CNN yakni MobileNetV2 ditambah *Dropout*. Arsitektur MobileNetV2 adalah arsitektur jaringan tiruan yang dirancang dengan tujuan memberikan fungsionalitas yang memuaskan pada perangkat seluler (mobile). MobileNetV2 didasarkan pada struktur residual terbalik, yang berarti bahwa koneksi antara lapisan bottleneck adalah yang dianggap sebagai koneksi residual. Untuk menyaring fitur yang berpotensi menjadi sumber non-linier, Layer intermediate expansion menggunakan menggunakan lightweight depthwise convolutions. Arsitektur MobileNetV2, secara keseluruhan, terdiri dari lapisan awal yang sepenuhnya convolutional dan berisi 32 filter, diikuti oleh 19 lapisan yang merupakan *residual bottleneck* [6].

Gambar 3 menunjukkan rancangan arsitektur MobileNetV2 dengan *Dropout*. Arsitektur MobileNetV2 terdiri atas beberapa layer yakni, Input layer sebagai layer untuk citra masukan. *Convolutional layer* yang bertujuan untuk memproses citra dengan filter yang telah didefinisikan tanpa merusak struktur citra awal. *Pooling layer* sebagai *layer* yang akan melakukan ekstraksi fitur dari citra inputan untuk mengubah fitur map dengan *array* multidimensi yang kemudian digunakan pada *Dense layer*. Selanjutnya *Flatten layer* merupakan *layer* yang harus dilewati sebelum masuk ke *Dense Layer* karena fitur map yang dihasilkan sebelumnya masih berbentuk *array* multidimensi sehingga *flatten layer* bertujuan untuk membentuk ulang fitur map agar menjadi vector (array satu dimensi) sehingga dapat digunakan sebagai input ke *tahapan Dense Layer*. *Dense layer* merupakan *hidden layer* dari *fully connected layer* yang digunakan untuk melakukan proses klasifikasi citra sesuai dengan kelasnya. Kemudian digunakan fungsi aktivasi ReLU untuk membantu menyesuaikan input yang masuk ke *fully connected layer* agar citra input sesuai dengan kelas klasifikasinya. *Dropout layer* berfungsi untuk mengurangi *overfitting* dari model yang telah dibangun agar model tidak terlalu fokus pada data training hingga terjadi kesalahan prediksi jika digunakan data testing. *Overfitting* bertujuan untuk menangkap *noise* agar hasil akurasi menjadi lebih tinggi karena jika terjadi *overfitting* model maka akan memiliki akurasi yang rendah. Fungsi *Aktivasi Softmax* hanya akan digunakan pada

output layer untuk membantu menangani kasus klasifikasi lebih dari 2 kelas label karena pada penelitian ini digunakan 3 kelas label citra untuk klasifikasi.



Gambar 3. Rancangan Arsitektur MobileNetV2 dengan *Dropout*

2.4 K-Fold Cross Validation

K-Fold Cross Validation merupakan teknik statistik untuk mengevaluasi model dimana data *train* dibagi menjadi dua subset, yaitu data *train* dan data validasi. Model atau algoritma akan dilatih dan divalidasi menggunakan data *train* dan validasi. Teknik ini biasanya digunakan untuk membuat prediksi dari model pelatihan untuk menentukan akurasi model. *K-Folds Cross Validation* membagi subset menjadi sebanyak K bagian dan mengulangi proses tanpa batas, menghasilkan model K pada akhir tahap pelatihan dan validasi. Biasanya, pilihan K adalah 5 atau 10, tetapi tidak ada aturan formal [7]. Pada penelitian ini akan dilakukan evaluasi terhadap model CNN yang telah dirancang, menggunakan *K-Fold Cross Validation*. Evaluasi model *K-Fold Cross Validation* dengan data *train* dibagi menjadi dua subset, yaitu data *train* dan data validasi. Arsitektur Model CNN yang telah dirancang akan dilatih dan divalidasi menggunakan data *train* dan validasi. Teknik ini digunakan untuk membuat prediksi dari model pelatihan untuk menentukan akurasi model. Untuk pengujian pada penelitian ini digunakan nilai K = 5fold. Berikut adalah contoh tabel evaluasi dengan *K-Fold Cross Validation*.

Tabel 1. Contoh Tabel Evaluasi dengan *K-Fold Cross Validation*

Dataset					
Data Train					Data Test
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	
Validasi	Train	Train	Train	Train	Akurasi Model 1
Train	Validasi	Train	Train	Train	Akurasi Model 2
Train	Train	Validasi	Train	Train	Akurasi Model 3
Train	Train	Train	Validasi	Train	Akurasi Model 4
Train	Train	Train	Train	Validasi	Akurasi Model 5

2.5 Evaluasi Sistem

Untuk mengetahui performa hasil model klasifikasi CNN, diperlukan sebuah teknik untuk pengukuran evaluasi terhadap kelas aslinya. Pengujian dilakukan dengan *Confusion Matrix*. Satuan ukur evaluasi yang digunakan adalah *Accuracy*, *Precision*, *Recall*, dan *F-Score*. *Accuracy* merupakan hasil seberapa akurat model dalam mengklasifikasikan dengan benar. *Precision* merupakan perbandingan antara total hasil prediksi kelas positif yang benar dengan total keseluruhan data yang diprediksi sebagai kelas positif. *Recall* merupakan perbandingan antara total hasil prediksi kelas positif yang benar dengan total keseluruhan data yang benar-benar positif. Setelah mendapat nilai presisi dan recall, maka dilakukan perhitungan menggunakan *F-Score*. *F-Score* digunakan untuk mengukur kombinasi hasil *Precision* dan *Recall*, sehingga menjadi satu nilai pengukuran. Berikut adalah contoh tabel ilustrasi *Confusion Matrix* untuk klasifikasi.

Tabel 2. Contoh Tabel Evaluasi dengan *Confusion Matrix*

CONFUSION MATRIX		PREDICTED	
		Yes	No
ACTUAL	Yes	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
	No	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Keterangan:

TP = True Positive, Prediksi kelas positif, dan kelas sebenarnya positif.

FP = False Positive, Prediksi kelas positif, dan kelas sebenarnya negatif.

TN = True Negative, Prediksi kelas negatif, dan kelas sebenarnya negatif.

FN = False Negative, Prediksi kelas negatif, dan kelas sebenarnya positif.

Untuk mengukur kinerja dari *performance matrix* diatas, digunakan beberapa *matrix* yakni *Accuracy*, *Precision*, *Recall*, dan *F-Score*. Berikut adalah rumus dari evaluasi kinerja *Confusion Matrix*.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots (1)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

$$F - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \dots\dots\dots (4)$$

3. Hasil dan Pembahasan

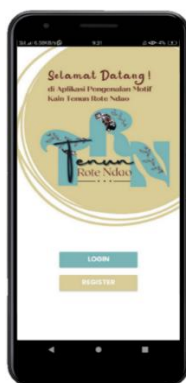
3.1. Tampilan Antarmuka Aplikasi Pengenalan Motif Kain Tenun Rote Ndao

Tampilan Antarmuka aplikasi terdiri dari halaman *Splash Screen*, menu *start*, menu *register*, menu *login*, menu utama, menu klasifikasi dan menu perpustakaan tenun. Halaman yang pertama muncul saat aplikasi dijalankan adalah *splash screen* yang dapat dilihat pada Gambar 4.



Gambar 4. Halaman *Splash Screen*

Menu *Start* yang dapat dilihat pada Gambar 5. Merupakan menu yang berisi pilihan untuk melakukan *login* atau registrasi. Jika pengguna sudah memiliki akun sebelumnya, maka memilih menu *login*, namun jika pengguna belum memiliki akun, maka pengguna akan melakukan registrasi terlebih dahulu.

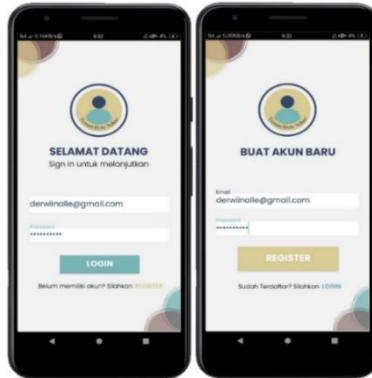


Gambar 5. Menu *Start*

Menu *Login* dan Menu *Register* yang dapat dilihat pada Gambar 6. Merupakan Menu yang dipilih pengguna untuk melakukan proses *login* apabila sudah memiliki akun dengan menginputkan *email* dan *password*. Jika pengguna belum memiliki akun, pengguna bisa mengklik pada tulisan *register*

Implementasi Metode *Convolutional Neural Network* Untuk Pengenalan Pola Motif Kain Tenun Rote Ndao Berbasis Android

yang berwarna kuning untuk masuk ke menu *register*. Menu *Register* merupakan menu yang dipilih pengguna untuk melakukan pendaftaran akun dengan menginputkan *email* dan *password*. Setelah melakukan proses *login* ataupun registrasi pada menu *login* dan registrasi, selanjutnya pengguna akan masuk ke menu utama yang berisi pilihan untuk ke menu klasifikasi, perpustakaan tenun dan keluar aplikasi.



Gambar 6. Menu *Login* dan Menu *Register*

Menu Utama yang dapat dilihat pada Gambar 7. Merupakan menu yang berisi pilihan untuk melakukan klasifikasi, melihat informasi tentang tenun di menu perpustakaan tenun dan keluar aplikasi.



Gambar 7. Menu Utama

Menu Klasifikasi yang dapat dilihat pada Gambar 8. Merupakan Menu yang berisi pilihan ambil gambar secara langsung menggunakan kamera atau memilih gambar dari galeri.



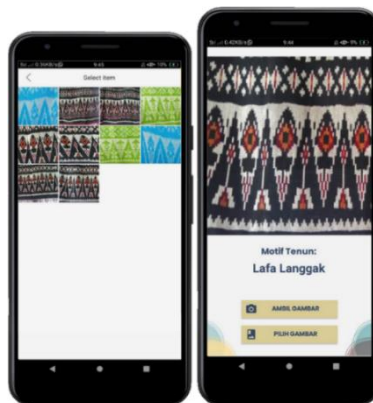
Gambar 8. Menu Klasifikasi

Ketika pengguna memilih ambil gambar, maka pengguna akan diarahkan ke kamera. Setelah mengambil gambar dari kamera maka pengguna akan mendapat keluaran hasil inputan dari gambar yang diambil berupa nama motif kain tenun. Hasil klasifikasi dengan citra yang diambil dengan kamera terdapat pada Gambar 9.



Gambar 9. Hasil Klasifikasi dengan citra yang diambil dengan kamera.

Selain menggunakan kamera secara langsung, pengguna juga dapat mengambil citra yang sudah ada digaleri dengan menggunakan pilih gambar maka pengguna akan diarahkan ke galeri. Setelah memilih gambar yang ingin diklasifikasikan dari galeri, maka pengguna akan mendapat keluaran hasil inputan dari gambar yang dipilih berupa nama motif kain tenun. Hasil klasifikasi dengan citra yang dipilih dari galeri dapat dilihat pada gambar terdapat pada Gambar 10.



Gambar 10. Hasil Klasifikasi dengan citra yang dipilih dari galeri

Selanjutnya Menu Perpustakaan Tenun yang dapat dilihat pada Gambar 11. Merupakan menu tambahan yang berisi informasi terkait, Sejarah tenun, fungsi tenun, ragam motif tenun, dan aksesoris pendamping tenun.



Gambar 11. Menu Perpustakaan Tenun

3.2. Hasil Pelatihan dan Pengujian Model Arsitektur MobileNetV2

Pada penelitian ini akan dilakukan evaluasi terhadap model CNN yang telah dirancang, menggunakan *K-Fold Cross Validation*. Evaluasi model dengan *K-Fold Cross Validation* dengan data *train* sejumlah 840 data kemudian dibagi menjadi dua subset, yaitu data *train* 80% (672 data) dan data validasi 20% (168 data). Arsitektur Model MobileNetV2 dengan *dropout* yang telah dirancang

akan dilatih dan divalidasi menggunakan data *train* dan *validasi* dengan *hyperparameter* berupa *learning rate* 0.0001 dan 0.0003 yang akan menghasilkan model untuk setiap Fold. Kemudian model tersebut akan diujikan dengan data testing. Teknik ini digunakan untuk membuat prediksi dari model pelatihan untuk menentukan akurasi model. Untuk pengujian pada penelitian ini digunakan nilai $K=5$ dengan skenario pada tabel 3.

Tabel 3. Skenario Pelatihan dan Pengujian Model Arsitektur MobileNetV2

Skenario	Keterangan
Data Training	672 Data Citra
Data Validasi	168 Data Citra
Batch Size	32 (21 Iterasi setiap Epoch)
Epoch	10
Learning Rate	- 0.0003 - 0.0001

Berdasarkan skenario evaluasi model menggunakan *K-Fold Cross Validation* pada tabel 3, didapatkan hasil evaluasi model dengan learning rate 0.0001 pada tabel 4, dan 0.0003 pada tabel 5.

Tabel 4. Hasil Pelatihan dan Pengujian Model dengan *Learning Rate* 0.0001

FOLD	Arsitektur Model MobileNetV2 dengan <i>Dropout</i> (<i>Adam Optimizer 0.0001</i>)		
	Akurasi Training	Akurasi Validasi	Akurasi Testing
1	84,08%	92,26%	87,61%
2	86,01%	88,69%	85,71%
3	84,38%	88,09%	88,57%
4	82,74%	89,28%	86,19%
5	84,52%	85,11%	88,09%

Pada Tabel 4, dapat dilihat bahwa hasil pelatihan menggunakan 840 data dengan learning rate 0.0001 mendapatkan akurasi tertinggi pada Fold-1 yaitu 92,26%. Hasil Pengujian Model dari setiap fold mendapatkan akurasi testing tertinggi pada Fold-3 yaitu 88,57%

Tabel 5. Hasil Pelatihan dan Pengujian Model dengan *Learning Rate* 0.0003

FOLD	Arsitektur Model MobileNetV2 dengan <i>Dropout</i> (<i>Adam Optimizer 0.0003</i>)		
	Akurasi Training	Akurasi Validasi	Akurasi Testing
1	95,54%	95,83%	93,81%
2	96,13%	91,07%	91,90%
3	93,60%	94,04%	93,81%
4	94,64%	95,23%	92,38%
5	94,94%	90,47%	91,90%

Sedangkan pada Tabel 5, dapat dilihat bahwa hasil pelatihan menggunakan 840 data dengan learning rate 0.0003 mendapatkan akurasi tertinggi pada Fold-1 yaitu 95,83%. Hasil Pengujian Model dari setiap fold mendapatkan akurasi testing tertinggi pada Fold-1 yaitu 93,81%. Hal ini berarti penggunaan *learning rate* berpengaruh untuk tingkat akurasi pelatihan terhadap model dan juga *learning rate* 0.0003 memiliki performa yang cukup baik dibandingkan learning rate 0.0001 dalam pelatihan model karena learning rate 0.0003 mendapatkan hasil akurasi pelatihan yang lebih tinggi daripada 0.0001.

3.3. Evaluasi Model Arsitektur MobileNetV2

Pada Tahap ini dilakukan Pada penelitian ini dilakukan evaluasi sistem, untuk mengetahui performa hasil model klasifikasi CNN. Model CNN yang telah didapatkan pada proses *training* model kemudian diuji dengan data testing untuk mengetahui performa model terhadap data *testing* menggunakan kombinasi *learning rate* 0.0001 dan 0.0003. Dari Hasil Testing didapat akurasi model tertinggi dengan Learning rate 0.0001 pada Fold-3 dan Hasil akurasi model tertinggi dengan Learning rate 0.0003 pada Fold-1. Proses Pengujian dilakukan dengan *Classification Report* dan *Confusion Matrix* untuk pengukuran evaluasi terhadap kelas aslinya.

Tabel 6. Hasil *Classification Report* dengan *Learning Rate* 0.0001 pada *Fold-3*

Classification Report	Precision	Recall	F-Score
Ai Bunak	81%	86%	83%
Dula Kakaik	85%	83%	84%
Lafa Laggak	100%	97%	99%
Accuracy			88%

Pada tabel 6, dapat dilihat bahwa hasil *Classification Report* mendapatkan *Accurary* model dengan learning rate 0.0001 pada *Fold-3* terhadap data testing sebesar 88% dengan penggunaan jumlah data testing 210 data yang terdiri dari 70 data untuk setiap motif yakni Ai Bunak, Dula Kakaik dan Lafa Laggak. *Precision* sebesar 81 % untuk motif ai bunak, 85% motif dula kakaik, dan 100% untuk motif lafa laggak. *Recall* sebesar 86% untuk motif ai bunak, 83% untuk motif dula kakaik dan 97% untuk motif lafa laggak. Kemudian nilai *F-Score* yang diperoleh yaitu 83% untuk motif ai bunak, 84% untuk motif dula kakaik dan 99% untuk motif lafa laggak.

Tabel 7. Hasil *Confusion Matrix* dengan *Learning Rate* 0.0001 pada *Fold-3*

CONFUSION MATRIX		PREDICTED		
		Ai Bunak	Dula Kakaik	Lafa Laggak
ACTUAL	Ai Bunak	60	10	0
	Dula Kakaik	12	58	0
	Lafa Laggak	2	0	68

Berdasarkan hasil *Confusion Matrix* dengan *Learning Rate* 0.0001 pada Tabel 7. Didapatkan hasil untuk pengukuran evaluasi motif kain tenun rote ndao terhadap kelas aslinya. Masing-masing pengukuran evaluasi yang didapatkan yakni, untuk motif Ai Bunak yang terdiri dari 70 data motif berhasil dikenali 60 data sesuai kelasnya, 10 data salah dikenali sebagai motif dula kakaik, kemudian Motif Dula Kakaik yang terdiri dari 70 data motif berhasil dikenali 58 data sesuai kelasnya dan 12 data salah dikenali sebagai motif Ai Bunak. Motif Lafa Laggak yang terdiri dari 70 data motif berhasil dikenali 68 data sesuai kelasnya dan 2 data salah dikenali sebagai ai bunak.

Tabel 8. Hasil *Classification Report* dengan *Learning Rate* 0.0003 pada *Fold-1*

Classification Report	Precision	Recall	F-Score
Ai Bunak	94%	87%	90%
Dula Kakaik	88%	94%	91%
Lafa Laggak	100%	100%	100%
Accuracy			93%

Pada tabel 8, dapat dilihat bahwa hasil *classification report* mendapatkan *Accurary* model dengan learning rate 0.0003 pada *Fold-1* terhadap data testing sebesar 93% dengan penggunaan jumlah data testing 210 data yang terdiri dari 70 data untuk setiap motif yakni Ai Bunak, Dula Kakaik dan Lafa Laggak. *Precision* sebesar 94% untuk motif ai bunak, 88% motif dula kakaik, dan 100% untuk motif lafa laggak. *Recall* sebesar 87% untuk motif ai bunak, 94% untuk motif dula kakaik dan 100% untuk motif lafa laggak. Kemudian nilai *f-score* yang diperoleh yaitu 90% untuk motif ai bunak, 91% untuk motif dula kakaik dan 100% untuk motif lafa laggak.

Tabel 9. Hasil *Confusion Matrix* dengan *Learning Rate* 0.0003 pada *Fold-1*

CONFUSION MATRIX		PREDICTED		
		Ai Bunak	Dula Kakaik	Lafa Laggak
ACTUAL	Ai Bunak	61	9	0
	Dula Kakaik	4	66	0
	Lafa Laggak	0	0	70

Berdasarkan hasil *Confusion Matrix* dengan *Learning Rate* 0.0003 pada tabel 9. Didapatkan hasil untuk pengukuran evaluasi motif kain tenun rote ndao terhadap kelas aslinya. Masing-masing

pengukuran evaluasi yang didapatkan yakni, untuk motif Ai Bunak yang terdiri dari 70 data motif berhasil dikenali 61 data sesuai kelasnya dan 9 data salah dikenali sebagai motif dula kakaik, kemudian Motif Dula Kakaik yang terdiri dari 70 data motif berhasil dikenali 66 data sesuai kelasnya dan 4 data salah dikenali sebagai motif Ai Bunak. Motif Lafa Langgak yang terdiri dari 70 data motif berhasil dikenali seluruhnya sesuai kelasnya.

Dilihat dari hasil pengujian model CNN terhadap data *testing* menggunakan kombinasi *learning rate* 0.0001 dan 0.0003. Didapatkan skor akurasi tertinggi pada Fold-3 sebesar 88% untuk learning rate 0.0001 dan 93% untuk learning rate 0.0003 pada *Fold-1* yang berarti:

- a. Arsitektur MobileNetV2 teruji dapat mengklasifikasikan dengan baik serta mendapatkan skor akurasi yang tinggi dengan penggunaan *learning rate* yang tepat (experimental).
- b. *Dropout* dapat digunakan untuk mengurangi overfitting pada model CNN karena mencapai nilai akurasi yang tinggi.
- c. *Learning Rate* 0.0003 mendapatkan akurasi lebih baik dari 0.0001 pada penelitian ini.

Setelah melalui proses pelatihan dan pengujian model arsitektur MobileNetV2 dengan *Dropout* dan didapatkan skor akurasi tertinggi pada model dengan *learning rate* 0.0003 pada *Fold-1*, maka model tersebut akan di *convert* ke *tf lite* yang kemudian akan di *deploy* pada aplikasi android sehingga aplikasi dapat melakukan pengenalan motif kain tenun rote ndao dengan baik. Untuk hasil implementasi Aplikasi pengenalan motif kain tenun rote ndao dapat dilihat pada sub bab 3.1.

4. Kesimpulan

Berdasarkan pada penelitian yang telah dilakukan dengan hasil yang diperoleh maka dapat disimpulkan bahwa model cnn mobilenetv2 teruji baik dalam melakukan klasifikasi terhadap 3 motif tenun Rote Ndao yakni Ai Bunak, Dula Kakaik, dan Lafa Langgak dengan penggunaan *Dropout* serta *Hyperparameter* berupa *Learning Rate* 0.0003 pada *Fold-1* mendapatkan akurasi sebesar 93%; *Precision* sebesar 94 % untuk motif Ai Bunak, 88% motif Dula Kakaik, dan 100% untuk motif Lafa Langgak. *Recall* sebesar 87% untuk motif Ai Bunak, 94% untuk motif Dula Kakaik dan 100% untuk motif Lafa Langgak. Kemudian nilai *F-Score* yang diperoleh yaitu 90% untuk motif Ai Bunak, 91% untuk motif Dula Kakaik dan 100% untuk motif Lafa Langgak. Dari penelitian yang telah dilakukan serta hasil yang diperoleh, maka saran-saran yang dapat disampaikan untuk dapat dipertimbangan dalam pengembangan penelitian selanjutnya yaitu, arsitektur MobileNetV2 dapat diganti dengan arsitektur CNN lainnya untuk mendapatkan hasil klasifikasi yang lebih optimal, selain itu diharapkan pada penelitian selanjutnya dapat menggunakan kombinasi parameter lainnya ketika melakukan pengujian. Kemudian memperbanyak data jenis motif kain tenun sehingga mendapatkan hasil akurasi yang lebih baik.

Daftar Pustaka

- [1] J. A. Therik, "Tenun Ikat Dari Timur (Keindahan Anggun Warisan Leluhur/ (Ikat in Eastern Archipelago (An Esoteric Beauty of Ancestral Entity))". Jakarta: Diandra Kreatif, 2017.
- [2] P. N. Andono, T. Sutojo, Muljono, Pengolahan citra digital. Yogyakarta: Penerbit Andi, 2017.
- [3] R. Munir, Pengolahan Citra Digital. Bandung: Informatika, 2004.
- [4] M. H. Nashr, M. Fachrurrozi, K. J. Miraswan, and E. Triningsih. "Pengenalan Motif Kain Songket Pada Citra Kamera Smartphone Dengan Beragam Sudut Pandang Menggunakan CNN.", vol 12. no 1. 2020.
- [5] R. F. Hakim, Implementasi Convolutional Neural Network Terhadap Transportasi Tradisional Menggunakan Keras, 2018.
- [6] M. Sandler, A. Howard, M. Ahu, A. Zhmoginov, And L. C. Chen, "Mobilenet2: Inverted residuals and linear bottlenecks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520, 2018.
- [7] J. Brownlee, A gentle introduction to k-fold cross-validation. Machine learning mastery, 2018

Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur

Farin Istighfarizky^{a1}, Ngurah Agus Sanjaya ER^{a2}, I Made Widiartha^{a3}, Luh Gede Astuti^{a4}, I Gusti Ngurah Anom Cahyadi Putra^{a5}, I Ketut Gede Suhartana^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia

¹farin79.istighfarizky@gmail.com

²agus_sanjaya@unud.ac.id

³madewidiartha@unud.ac.id

⁴lq.astuti@unud.ac.id

⁵anom.cp@unud.ac.id

⁶ikg.suhartana@unud.ac.id

Abstract

Classification is a process that automatically places text documents into a text based on the content of the text. Classification can help us classifying many text documents that have been published, with the classification, these text documents can be reached easily and quickly. Feature selection can be used to improve the performance of text classification in terms of learning speed and effectiveness. In the Chi-Square feature selection experiment, a 1% threshold combination with a parameter value of $k=6$ is the combination chosen to be the best model. In testing the new data, the K-Nearest Neighbor model by selecting the Chi-Square feature produces precision performance, recall, F1-Score, and accuracy respectively, namely 85%, 83.3%, 88.2%, and 92.3%. In the Gini Index feature selection experiment, 1% threshold combination with a parameter value of $k=4$ is the combination chosen to be the best model. This threshold selects about 31 features with the highest Gini Index value. In testing the new data, the K-Nearest Neighbor model by selecting the Gini Index feature produces precision performance, recall, F1-Score, and accuracy respectively, namely 81.2%, 80.3%, 81.6%, and 86.6%.

Keywords: Classification, Chi-Square, Gini Index, Features Selection, K-Nearest Neighbor

1. Pendahuluan

Penemuan dibidang informatika mengalami perkembangan yang sangat pesat. Ratusan penelitian dilakukan di berbagai bidang disetiap tahunnya yang mana dengan harapan hasil penelitian tersebut dapat digunakan untuk penemuan berikutnya. Tidak semua penemuan akan relevan terhadap penelitian yang dilakukan oleh seseorang, oleh karena itu diperlukannya pengelompokan penelitian agar lebih mudah dalam mencari penelitian yang kita inginkan atau butuhkan. Jumlah studi yang dipublikasikan ada ratusan bahkan ribuan penelitian setiap tahunnya, seseorang akan membutuhkan terlalu banyak usaha dan dana yang besar dalam mengelompokkan jurnal-jurnal penelitian. Masalah ini dapat diselesaikan dengan klasifikasi teks.

Klasifikasi adalah proses menempatkan dokumen teks secara otomatis ke dalam kategori berdasarkan teks [1]. Klasifikasi membantu mengklasifikasikan jumlah dokumen teks yang diterbitkan. Klasifikasi membuatnya cepat dan mudah untuk mengelompokkan dokumen tekstual. Dasar dari algoritma seleksi fitur adalah untuk menemukan semua kemungkinan kombinasi atribut dalam data. Ini digunakan untuk menemukan subset terbaik untuk prediksi. Pemilihan fitur dapat digunakan untuk meningkatkan kinerja klasifikasi teks dalam hal kecepatan dan efektivitas pembelajaran [2].

Studi klasifikasi telah dilakukan oleh beberapa peneliti, seperti tentang mengkategorikan soal ujian secara otomatis. Pada penelitian tersebut, penulis menggunakan metode KNN dengan seleksi fitur *Chi-Square*. Hasil pada penelitian tersebut menunjukkan bahwa metode seleksi fitur *Chi-Square* terbukti mampu meningkatkan performa dari metode KNN [3]. Penelitian selanjutnya adalah

membahas tentang kognitif soal pada taksonomi *bloom* dengan KNN. Hasil yang didapatkan dari algoritma KNN dengan seleksi fitur *Gini Index* pada penelitian tersebut adalah akurasi sebesar 68,37% dan kappa tertinggi sebesar 0,607. Berdasarkan hasil tersebut, *Gini Index* mampu mengurangi dimensi fitur yang tinggi [4].

Seleksi fitur *Chi Square* menggunakan teori statistik untuk menentukan independensi suatu *term* dari kategorinya. Dalam seleksi fitur *Chi Square* berdasarkan teori statistika, dua peristiwa di antaranya adalah kemunculan dari fitur dan kemunculan dari kategori yang kemudian nilai *term* diurutkan dari yang tertinggi. Sedangkan seleksi fitur *Gini Index* mampu mengurangi dimensi fitur yang tinggi pada klasifikasi teks.

Berdasarkan penelitian yang dilakukan sebelumnya, pada penelitian kali ini penulis melakukan klasifikasi jurnal menggunakan metode *K-Nearest Neighbor* dengan menggunakan dua seleksi fitur yaitu *Chi-Square* dan *Gini Index*. Penulis berharap bahwa dengan menggunakan kombinasi metode ini dapat menghasilkan performa *precision*, *recall*, *f1-score*, dan akurasi yang lebih baik dibandingkan penelitian sebelumnya.

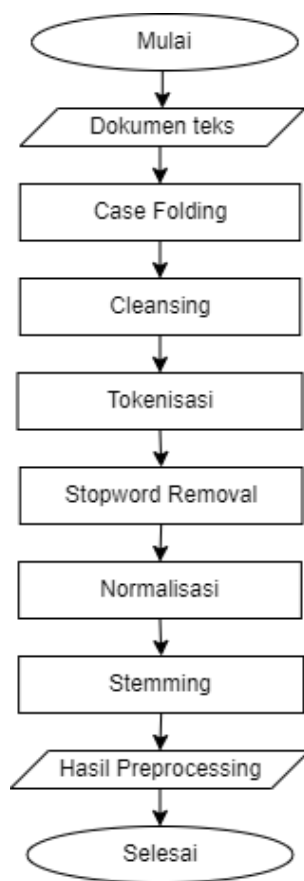
2. Metode Penelitian

2.2. Dataset

Data yang digunakan pada penelitian ini adalah jurnal yang dipublikasikan oleh SINTA (*Science and Technology Index*) dan *Google Scholar*, data dapat diperoleh dari *website* SINTA (*Science and Technology*) (<https://sinta.ristekbrin.go.id/>) dan *Google Scholar* (<https://scholar.google.com>) Kemudian data yang digunakan untuk proses klasifikasi adalah pada bagian teks abstrak jurnal yang berbahasa Indonesia yang disimpan dalam bentuk (.xlsx) untuk digunakan sebagai data latih dan data uji. Dokumen yang digunakan berjumlah 100 data per kelas disetiap artikel ilmiah yaitu: pendidikan, ekonomi, dan informatika. Data *testing* yang diuji sebanyak 60 data jurnal dan data *training* sebanyak 240 data jurnal. Data pada penelitian ini adalah data sekunder.

2.2. Preprocessing

Preprocessing proses pertama mempersiapkan dataset sebelum pembobotan, tujuannya adalah untuk menyederhanakan pemrosesan data dan juga untuk mendapatkan tingkat performa yang tinggi. Proses *preprocessing* dapat dilihat pada gambar 1.

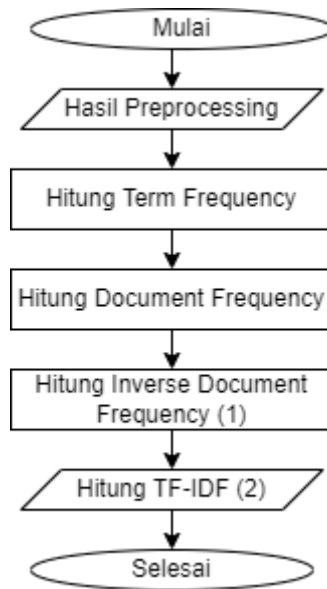


Gambar 1. Proses *preprocessing*

Pada tahap ini, terdapat beberapa proses yaitu *case folding*, *cleansing*, tokenisasi, *stopword*, normalisasi, dan *stemming*. *Case folding* yaitu proses mengubah semua huruf menjadi huruf kecil. *Cleansing* merupakan proses penghapusan karakter yang tidak relevan dengan klasifikasi jurnal. Tokenisasi yaitu pemisahan kata-kata paragraf atau kalimat menjadi token-token tertentu. *Stopword removal* yaitu penghapusan kata yang tidak mempengaruhi klasifikasi jurnal. Proses normalisasi yaitu mengubah dan mengembalikan bentuk penulisan tidak baku ke bentuk penulisan yang sesuai dengan KBBI. Proses terakhir adalah *stemming*, yaitu mengekstrak kata yang dilampirkan ke dalam kata dasar [5].

2.3. Term Frequency Inverse-Document Frequency(TF-IDF)

Metode *Term Frequency Invers Document Frequency* adalah metode pembobotan yang menggabungkan frekuensi istilah dalam satu set dokumen dan kelangkaannya. Metode ini menggabungkan dua konsep pembobotan, yaitu frekuensi kemunculan kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang berisi kata tersebut. Berapa kali sebuah kata muncul dalam dokumen tertentu menunjukkan pentingnya kata tersebut dalam dokumen itu. Frekuensi dokumen yang berisi kata-kata menunjukkan seberapa sering kata-kata itu muncul. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen. Proses TF-IDF dapat dilihat pada gambar 2.



Gambar 2. TF-IDF

- a. Hitung jumlah kemunculan *term* i dalam dokumen j ($tf_{i,j}$).
- b. Hitung jumlah dokumen yang mengandung *term* i (df)
- c. Menghitung nilai bobot *inverse document frequency* (idf) dengan menggunakan persamaan:

$$idf_i = \log \left(\frac{N}{df_i} \right) \quad (1)$$

Keterangan:

N = jumlah dokumen secara keseluruhan

- d. Menghitung nilai bobot TF-IDF dengan menggunakan persamaan:

$$w_{i,j} = tf_{i,j} \times idf_i \quad (2)$$

Keterangan:

$w_{i,j}$ = bobot *term* i terhadap dokumen j

$tf_{i,j}$ = frekuensi *term* i pada dokumen j

idf_i = nilai bobot IDF pada *term* i

2.4. *K-Nearest Neighbor* (KNN)

Algoritma *K-Nearest Neighbor* adalah metode untuk mengklasifikasi objek berdasarkan data latih yang paling dekat dengan objek tersebut. Metode *K-Nearest Neighbor* adalah algoritma pembelajaran terawasi, dan hasil dari *query instance* baru dikategorikan berdasarkan sebagian besar kategori algoritma *K-Nearest Neighbor*. Kelas yang paling sering ditampilkan adalah kelas yang diperoleh dari hasil klasifikasi. Kedekatan didefinisikan dalam jarak metrik, seperti jarak *Euclidean* [6].

$$D_{x,y} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Keterangan:

D = jarak kedekatan

x = data training

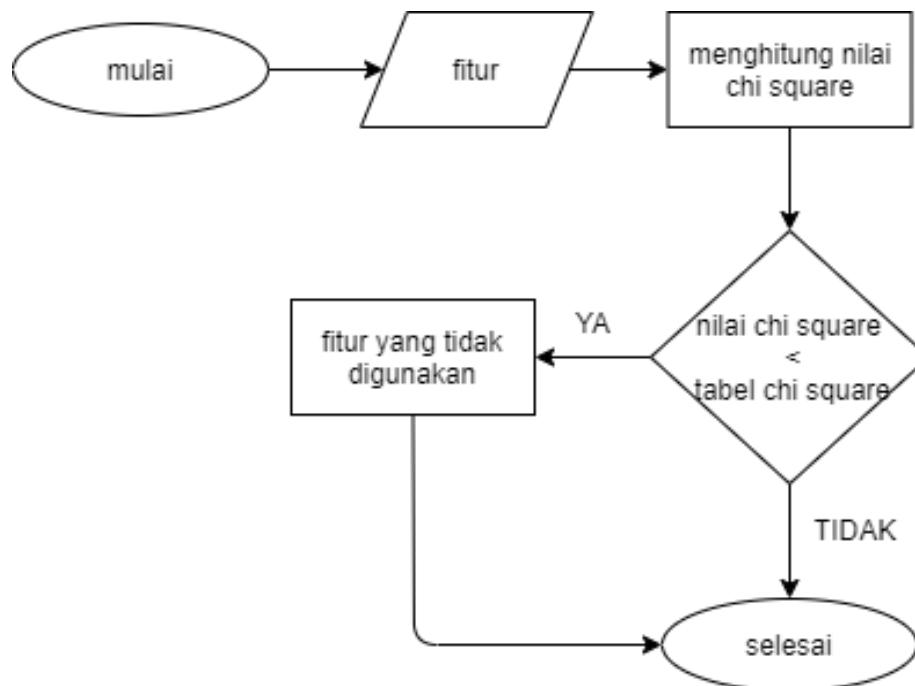
y = data testing

n = jumlah atribut individu antara 1 s.d n

i = atribut individu antara 1 s.d n

2.5. Chi-Square

Chi-Square adalah metode untuk menghitung ketergantungan fitur. Pada pemrosesan teks biasanya menggunakan dua kelas untuk mengukur ketergantungan antara dua label dan kata – antara kelas tertentu c. Tahapan *Chi-Square* dapat dilihat pada gambar 3.



Gambar 3. *Chi-Square*

Berikut ini adalah rumus yang digunakan untuk menerapkan metode seleksi fitur *Chi-Square* [7].

$$X^2(t, c) = \frac{N(AD - CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (4)$$

Keterangan:

t = kata

c = kelas/kategori

N = jumlah data latih

A = jumlah dokumen pada kelas c yang memuat t,

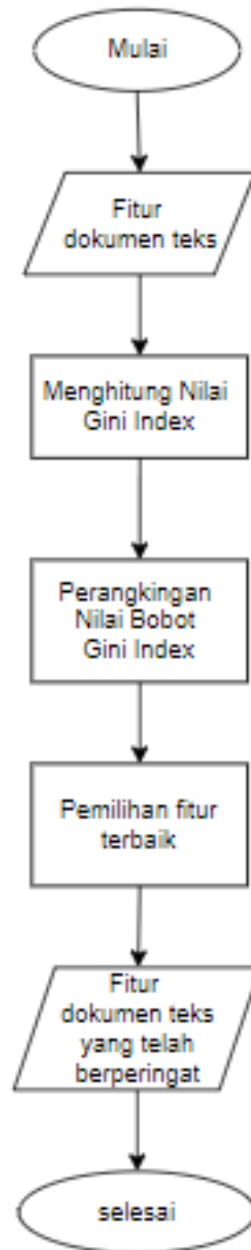
B = jumlah dokumen yang tidak ditemukan pada kelas c tapi memuat t,

C = jumlah dokumen pada kelas c yang tidak memuat t,

D = jumlah dokumen yang bukan merupakan dokumen kelas c dan tidak memuat term t

2.6. Gini Index

Gini Index adalah kriteria berbasis ketidakmurnian data yang mengukur perbedaan antara distribusi probabilitas dari nilai atribut. *Gini Index* umumnya dipakai dalam Algoritma *Classification and Regression Trees* yang merepresentasikan ukuran seberapa acak pilihan objek dari data latih. Ukuran ketidakmurnian mencapai 0 ketika hanya 1 kelas saja yang ada pada sebuah titik. Namun sebaliknya akan mencapai maksimum ketika ukuran kelas pada titik tersebut seimbang. *Gini Index* dapat dianggap sebagai probabilitas dari dua data yang dipilih secara acak dari kelas yang berbeda akan digunakan dalam penelitian ini untuk mengukur divergensi yang digunakan sebagai dasar bobot setiap. Cocok untuk pemilahan, sistem biner, nilai numerik terus menerus, dan lain-lain. Tahapan *Gini Index* dapat dilihat pada gambar 4.



Gambar 4. *Gini Index*

Berikut ini adalah rumus yang digunakan untuk menerapkan metode seleksi fitur *Gini Index* [8].

$$GI(t) = 1 - \sum_{i=1}^C [p(i|t)]^2 \quad (5)$$

Keterangan:

C = total kelas

t = term

p(i|t) = peluang kelas i terhadap term t

2.7. Evaluasi

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur performansi suatu metode klasifikasi. Pada dasarnya *confusion matrix* ini berisi informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi sebagaimana mestinya. Saat mengukur kinerja menggunakan *confusion matrix*, ada empat istilah yang menggambarkan hasil dari proses klasifikasi. Keempat istilah tersebut adalah *True Positive*, *True Negative*, *False Positive*,

dan *False Negative*. Nilai *True Negative* (TN) adalah jumlah data negatif yang terdeteksi dengan benar, dan *False Positive* (FP) adalah data negatif tetapi terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) di sisi lain adalah data positif yang dikenali dengan benar. *False Negative* (FN) adalah kebalikan dari *True Positive*, sehingga datanya positif tetapi dikenali sebagai data negatif. Tabel *Confusion matrix* dapat dilihat pada tabel 1 [9].

Tabel 1. *Confusion Matrix*

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
Positif	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Dimana:

- *True Positive*, jumlah data positif yang diklasifikasikan dengan benar oleh sistem.
- *True Negative*, jumlah data negatif yang diklasifikasikan dengan benar oleh sistem.
- *False Negative*, jumlah data negatif namun diklasifikasikan salah oleh sistem.
- *False Positive*, jumlah data positif namun diklasifikasikan salah oleh sistem.

Berdasarkan nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP) dapat diperoleh nilai akurasi, presisi dan *recall*. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan (6). Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan persamaan (7) sementara itu, *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai *recall* diperoleh dengan persamaan 8.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (6)$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\% \quad (7)$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\% \quad (8)$$

Setelah mendapat nilai *recall* dan *precision*, maka dilakukan perhitungan menggunakan *F1-score*. *F1-score* digunakan untuk mengukur kombinasi hasil *precision* dan *recall*, sehingga menjadi satu nilai pengukuran. *F1-Score* dapat dihitung menggunakan persamaan 9:

$$F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \quad (9)$$

3. Hasil dan Pembahasan

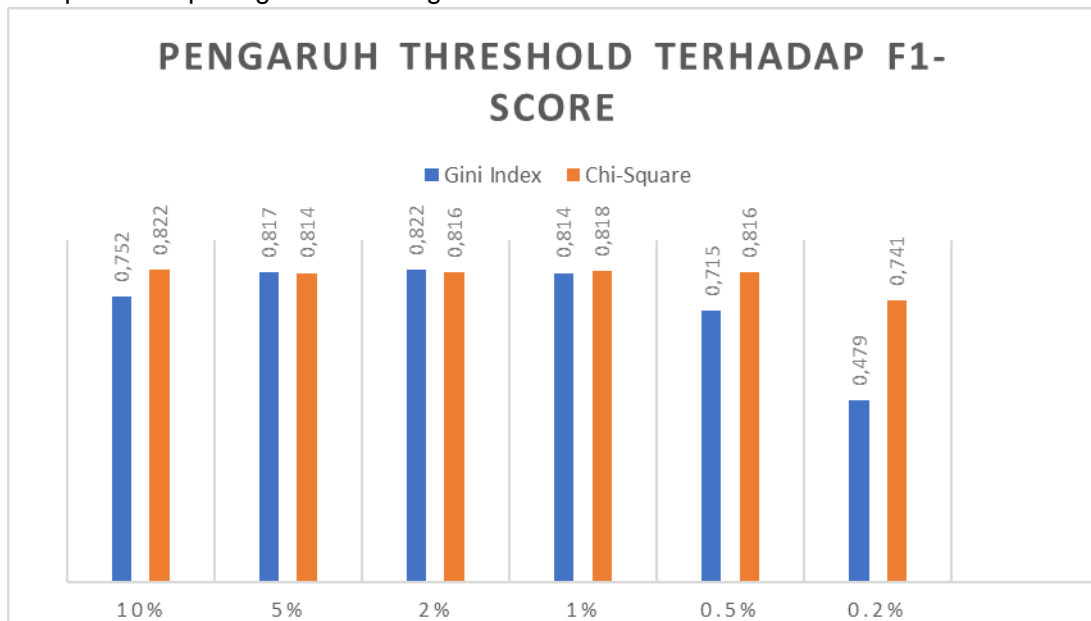
Pada penelitian yang dilakukan ada 80% total data digunakan selama tahap pelatihan dan juga validasi. Perubahan nilai k pada percobaan adalah k = 4, k = 6, k = 7, k = 9, dan k = 11. Uji *threshold* dilakukan dengan menggunakan seleksi fitur *Gini Index* dan seleksi fitur *Chi-Square*. *Threshold* adalah persentase jumlah fitur yang dipilih dari semua fitur yang diurutkan. *Threshold* yang digunakan adalah 10%, 5%, 2%, 1%, 0.5%, dan 0.2%. Pada setiap iterasi dari *10-Fold Cross Validation*, akan dihitung rata-rata performa *F1-Score* dan akurasi dengan menggunakan persamaan (9) dan (6). Nilai k dengan kinerja *F1-Score* tertinggi akan dipilih sebagai model yang terbaik. Nilai k dengan *F1-Score* tertinggi berarti hasil klasifikasi jurnal lebih akurat. Setelah melakukan proses pelatihan dan validasi pada model *K-Nearest Neighbor* menggunakan uji *10-Fold Cross Validation*,

didapatkan nilai k dengan performansi *F1-Score* terbaik. Uji kombinasi *threshold Chi-Square* dan *Gini Index* dan nilai k yang sudah dilakukan menghasilkan beragam performa yang berbeda. Dari hasil pengujian yang diperoleh, dapat dilihat pengaruh dari *threshold* yang digunakan untuk performansi dari metode *K-Nearest Neighbor*, seperti yang ditunjukkan pada tabel 5.

Tabel 5. Hasil Evaluasi Pengujian KNN dengan Seleksi Fitur Chi-Square dan Gini Index

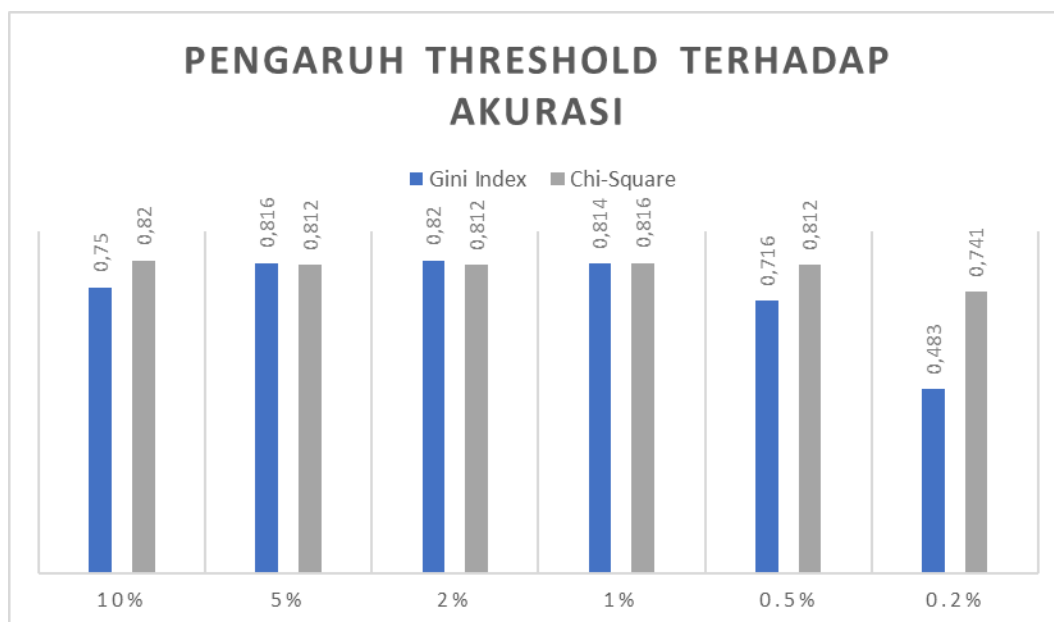
Threshold	Ukuran Evaluasi (Rata-Rata Fold)			
	F-1 Score Chi-Square	F1-Sscore Gini Index	Akurasi Chi-Square	Akurasi Gini Index
10%	82,2%	75,2%	82%	75%
5%	81,4%	81,7%	81,2%	81,6%
2%	81,6%	82,2%	81,2%	82%
1%	81,8%	81,4%	81,6%	80,8%
0.5%	81,6%	71,5%	81,2%	71,6%
0.2%	74,1%	47,9%	74,1%	48,3%

Pengujian kombinasi *threshold Chi-Square* dan *Gini Index* dan nilai k yang sudah dilakukan menghasilkan beragam performa yang berbeda. Dari hasil pengujian yang diperoleh dapat diketahui pengaruh *threshold* yang digunakan terhadap evaluasi kinerja metode *K-Nearest Neighbor*, hal tersebut dapat dilihat pada gambar 5 dan gambar 6.



Gambar 5. Pengaruh Threshold terhadap F1-Score

Gambar 5 menunjukkan pengaruh *threshold* terhadap performa *F1-Score* metode *K-Nearest Neighbor* dengan seleksi fitur *Chi-Square* dan seleksi fitur *Gini Index*. *F1-Score* yang tertera pada Gambar 5 adalah nilai *F1-Score* dari kombinasi nilai k dengan akurasi tertinggi. Nilai *threshold* diketahui memiliki pengaruh terhadap *F1-Score*.



Gambar 6. Pengaruh Threshold terhadap akurasi

Akurasi yang tertera pada Gambar 6 adalah akurasi tertinggi dari kombinasi nilai k yang sudah diuji. Nilai threshold diketahui memiliki pengaruh terhadap akurasi. Model dengan threshold yang menghasilkan akurasi di atas nilai rata-rata tersebut dapat dikatakan sebagai model dengan performa yang baik dalam mengklasifikasikan jurnal.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, ditarik kesimpulan bahwa implementasi seleksi fitur dapat meningkatkan performa *precision*, *recall*, *F1-Score*, dan akurasi dari metode *K-Nearest Neighbor* dalam mengklasifikasikan jurnal dan seleksi fitur *Chi-Square* lebih unggul daripada seleksi fitur *Gini Index*. Pada eksperimen seleksi fitur *Chi-Square*, kombinasi *threshold* 1% dengan parameter nilai $k=6$ adalah kombinasi yang dipilih menjadi model terbaik. *Threshold* ini menyeleksi sekitar 31 fitur dengan nilai *Chi-Square* tertinggi. Pada pengujian data baru, model *K-Nearest Neighbor* dengan seleksi fitur *Chi-Square* menghasilkan performa *precision*, *recall*, *F1-Score*, dan akurasi secara berturut-turut yaitu 85%, 83.3%, 88.2%, dan 92.3%. Pada eksperimen seleksi fitur *Gini Index*, kombinasi *threshold* 1% dengan parameter nilai $k=4$ adalah kombinasi yang dipilih menjadi model terbaik. *Threshold* ini menyeleksi sekitar 31 fitur dengan nilai *Gini Index* tertinggi. Pada pengujian data baru, model *K-Nearest Neighbor* dengan seleksi fitur *Gini Index* menghasilkan performa *precision*, *recall*, *F1-Score*, dan akurasi secara berturut-turut yaitu 81.2%, 80.3%, 81.6%, dan 86.6%

Daftar Pustaka

- [1] Z. XIONG, J. JIANG and Y. ZHANG, "New feature selection approach (CDF) for text categorization", *Journal of Computer Applications*, vol. 29, no. 7, pp. 1755-1757, 2009. Available: 10.3724/sp.j.1087.2009.01755.
- [2] H. Alshalabi, S. Tiun, N. Omar and M. Albared, "Experiments on the Use of Feature Selection and Machine Learning Methods in Automatic Malay Text Categorization", *Procedia Technology*, vol. 11, pp. 748-754, 2013. Available: 10.1016/j.protcy.2013.12.254.
- [3] I. Listiowarni and N. Puspa Dewi, "Pemanfaatan Klasifikasi Soal Biologi Cognitive Domain Bloom's Taxonomy Menggunakan KNN Chi-Square Sebagai Penyusunan Naskah Soal", *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 2, pp. 186-197, 2020. Available:

10.31849/digitalzone.v11i2.4798.

- [4] T. Setiyorini and R. Asmono, "PENERAPAN METODE K-NEAREST NEIGHBOR DAN GINI INDEX PADA KLASIFIKASI KINERJA SISWA", *Jurnal Techno Nusa Mandiri*, vol. 16, no. 2, pp. 121-126, 2019. Available: 10.33480/techno.v16i2.747.
- [5] K. Yonatha Wijaya and A. Karyawati, "The Effects of Different Kernels in SVM Sentiment Analysis on Mass Social Distancing", *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 9, no. 2, p. 161, 2020. Available: 10.24843/jlk.2020.v09.i02.p01.
- [6] H. Hadi and T. Sukamto, "Klasifikasi Jenis Laporan Masyarakat Dengan K-Nearest Neighbor Algorithm", *JOINS (Journal of Information System)*, vol. 5, no. 1, pp. 77-85, 2020. Available: 10.33633/joins.v5i1.3355.
- [7] C. Suharno, M. Fauzi and R. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square", *Systemic: Information System and Informatics Journal*, vol. 3, no. 1, pp. 25-32, 2017. Available: 10.29080/systemic.v3i1.191.
- [8] C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing Switzerland, 2015. Available: 10.1007/978-3-319-14142-8.
- [9] M. Imron and B. Prasetyo, "Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn", *Shmpublisher.com*, 2022. [Online]. Available: <https://shmpublisher.com/index.php/joscecx/article/view/7>.

Aplikasi Pencarian Informasi AI-Quran dengan Ontologi Web Semantik

Safira^{a1}, I Wayan Santiyasa^{a2}, I Gede Arta Wibawa^{a3}, Ngurah Agus Sanjaya ER^{a4}, Ida Bagus Gede Dwidasmara^{a5}, Luh Gede Astuti^{a6}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia

¹virasafira233@gmail.com

²santiyasa@unud.ac.id

³gede.arta@unud.ac.id

⁴agus_sanjaya@unud.ac.id

⁵dwidasmara@unud.ac.id

⁶lg.astuti@unud.ac.id

Abstract

AI-Quran is the holy book of Muslims which consists of 30 chapters, 114 surahs, and 6236 verses. Each juz in the Quran can contain one or more surah and each surah consists of a different number of verses. Seeing the many contents of the knowledge of the Quran, then this can help someone to find and determine the theme of the surah to be read manually. In this study, the authors build a system that is expected to facilitate one's work in searching for the contents of the AI-Quran by using ontology as the backbone of the system, Methontology as a method for building an ontology model, and Prototyping as a system development method.

Keywords: Knowledge Management System, AI-Quran, Ontology, Methontology, Prototyping

1. Pendahuluan

Indonesia merupakan negara dengan penduduk muslim terbanyak di dunia. Muslim adalah sebutan bagi seseorang yang menganut agama Islam. Menurut hasil dari Data Sensus Penduduk Indonesia pada tahun 2018, sebesar 86.7% dari total penduduk Indonesia adalah penganut agama Islam. AI-Quran adalah kitab suci agama Islam yang mengandung beberapa nilai seperti akhlak, syariah, keimanan, serta peraturan yang mengatur kehidupan manusia.

Menurut Kementerian Agama Republik Indonesia, AI-Quran terdiri dari 30 bagian (juz), 114 bab (surah), dan 6236 ayat. Konsep ilmu AI-Quran didefinisikan dengan menggunakan hirarki pengetahuan AI-Quran yang ditentukan oleh urutan bagian (juz), bab (surah) dan ayat [1]. Dalam hirarki pengetahuan AI-Quran, setiap bagian (juz) dapat berisikan satu atau lebih surah, kemudian beberapa surah dapat berada pada lebih dari satu bagian (juz), dan setiap surah memiliki jumlah ayat dan arti surah yang berbeda-beda. Jika melihat dari rumitnya hirarki pengetahuan AI-Quran, maka akan sulit bagi seseorang untuk melakukan pencarian terhadap isi AI-Quran secara manual dan cepat. Dalam hal ini, salah satu cara untuk mempermudah pekerjaan seseorang dalam melakukan pencarian isi AI-Qura adalah dengan memanfaatkan teknologi informasi.

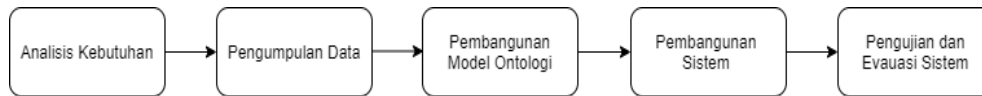
Dalam penelitian ini, teknologi web semantik diadopsi untuk mengatasi permasalahan dalam melakukan pencarian terhadap isi AI-Quran. Web Semantik merupakan perpanjangan dari web yang sudah ada, dimana informasi yang diberikan memiliki arti yang lebih jelas sehingga memungkinkan manusia dan komputer bekerja sama dalam mencari pemahaman yang sama antara konsep yang ada. [2]. Ontologi merupakan salah satu teknologi pendukung dari web semantik yang merupakan suatu rangkaian representasi konsep yang menjelaskan hubungan dengan konsep tersebut [3]. Dalam hal ini, semantik mengacu pada hubungan hirarki bagian (juz) dan bab (surah) pada AI-Quran.

Penelitian ini bertujuan untuk mengembangkan model ontologi pada domain AI-Quran. Dimana pada penelitian ini, dilakukan pengujian awal terhadap model ontologi AI-Quran dengan memberikan

serangkaian pertanyaan terkait informasi isi AI-Quran. Sehingga dari model ontologi AI-Quran yang telah dibangun diharapkan dapat menampilkan dan memberikan informasi mengenai AI-Quran secara sistematis, cepat, dan tepat.

2. Metode Penelitian

Methontologi digunakan sebagai metode untuk membangun model ontologi dan metode *Prototyping* sebagai metode untuk pembangunan sistem. Alur dari penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Analisis Kebutuhan

Tahap ini bertujuan untuk menganalisis kebutuhan sistem yang meliputi data yang digunakan, pembelajaran dari referensi yang sudah ada, dan perangkat yang digunakan. Tahap ini dibagi menjadi dua, yaitu tahap analisis kebutuhan fungsional yang meliputi kegunaan dari sistem dan kebutuhan non-fungsional yang meliputi komponen-komponen pendukung dalam implementasi sistem seperti kebutuhan perangkat keras yang berupa laptop atau komputer dan kebutuhan perangkat lunak yang berupa aplikasi Protégé, vscode, dan Apache Jena Fuseki.

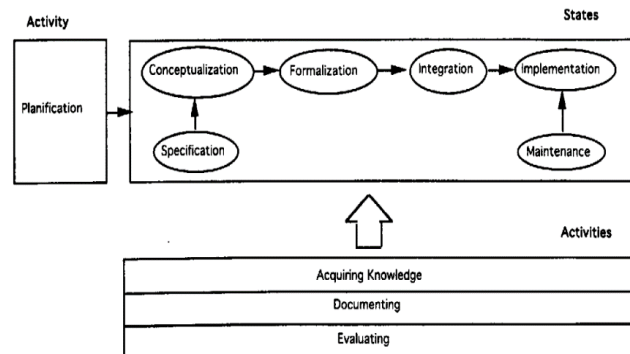
2.2. Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data yang berhubungan dengan AI-Quran yang diperoleh dan dikumpulkan melalui sumber internet dan jurnal. Kemudian data AI-Quran yang telah dikumpulkan diproses dengan menggunakan aplikasi Protégé untuk membangun model ontologi. Adapun data yang dikumpulkan adalah data mengenai si AI-Quran yang berupa juz, surah, ayat, arti surah, golongan surah, dan tema surah.

2.3. Pembangunan Model Ontologi

Metode yang digunakan untuk pembangunan model ontologi adalah Methontologi. Methontologi merupakan salah satu metode untuk pengembangan model ontologi yang memiliki keunggulan dalam mendeskripsikan setiap aktivitas [4]. Dalam hal ini, Methontologi memiliki kemampuan yaitu dapat menggunakan kembali ontologi yang sudah dibangun sebelumnya untuk pengembangan sistem selanjutnya [5]. Seperti yang ditunjukkan pada Gambar 2, tahapan dalam metode Methontologi terdiri atas:

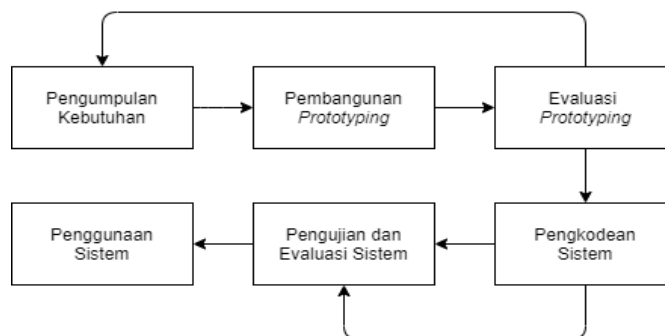
- Aktivitas (perencanaan, akuisisi pengetahuan, dokumentasi, dan evaluasi). Aktivitas perencanaan dilakukan di awal proyek pengembangan.
- States (spesifikasi, konseptualisasi, formalisasi, integrasi, implementasi, dan pemeliharaan)



Gambar 2. Tahapan dalam Metode Methontologi

2.4. Pembangunan Sistem

Metode yang digunakan untuk pembangunan sistem adalah metode *Prototyping*. Pada metode ini, dilakukan proses pembuatan model perangkat lunak sederhana yang memungkinkan pengguna memiliki gambaran dasar tentang program serta melakukan pengujian awal. Tahapan dalam metode *Prototyping* dapat dilihat pada Gambar 3.



Gambar 3. Tahapan dalam Metode *Prototyping*

2.5. Pengujian dan Evaluasi Sistem

Proses pengujian dan evaluasi sistem dilakukan dengan membuat beberapa pertanyaan yang berkaitan dengan fitur penjelajahan dan fitur pencarian, serta persepsi kegunaan dan kemudahan penggunaan yang dirasakan oleh pengguna sistem. Data jawaban yang didapatkan dari proses pengujian sistem, kemudian diolah dan dianalisis untuk mendapatkan nilai akurasi. Untuk pertanyaan yang berkaitan dengan pengujian fitur, data diolah dengan menandai masing-masing kiriman jawaban dan mengklasifikasikan skema penandaan menjadi 3 (tiga) kategori yaitu:

- Salah, dengan memberikan nilai skor 0 (nol). Nilai skor diberikan ketika peserta tidak memberikan jawaban benar terhadap tugas pertanyaan yang diberikan.
- Sebagian benar, dengan memberinilai skor 1 (satu). Nilai skor ini diberikan ketika jawaban yang diberikan oleh peserta sebagian benar.
- Sepenuhnya benar, dengan memberi nilai skor 2 (dua). Nilai skor ini diberikan ketika jawaban yang diberikan oleh peserta sepenuhnya benar.

Sedangkan untuk pertanyaan yang berkaitan dengan persepsi kegunaan dan kemudahan penggunaan sistem, diukur dengan menggunakan skala 5 poin yaitu sangat setuju = 5, setuju = 4, netral = 3, tidak setuju = 2, dan sangat tidak setuju = 1. Kemudian setelah data diolah, dilakukan analisis statistik terhadap data kiriman jawaban peserta pengujian dengan menerapkan hasil analisis statistik sebagai berikut:

- Rerata (*mean*). Dengan analisis ini, didapatkan rerata skor yang diberikan peserta pada masing-masing pertanyaan. Rerata ini akan menggambarkan seberapa berguna dan mudah digunakan sistem dalam persepsi peserta.
- Nilai tengah (*median*). Dengan analisis ini, didapatkan nilai tengah dari seluruh skor peserta pada masing-masing pertanyaan
- Nilai terendah (*minimum*). Dengan analisis ini, didapatkan nilai terendah dari seluruh skor peserta pada masing-masing pertanyaan.
- Nilai tertinggi (*maximum*). Dengan analisis ini, didapatkan nilai tertinggi dari seluruh skor peserta pada masing-masing pertanyaan.

3. Hasil dan Pembahasan

3.1. Implementasi Ontologi

Berikut ini merupakan implementasi dari tahapan pada metode pembangunan model ontologi dengan Methontologi.

a. Tahap Spesifikasi

Pada tahap ini, dihasilkan deskripsi dari ontologi AI-Quran sebagai berikut.

1. Domain : AI-Quran
2. Tujuan : Untuk membangun modelontologi sebagai representasi informasi dalam semantic ontologi pada domain AI-Quran
3. Tingkat Formalitas : Formal
4. Ruang Lingkup : Isi AI-Quran
5. Sumber Pengetahuan : Jurnal dan internet

b. Tahap Akuisisi Pengetahuan

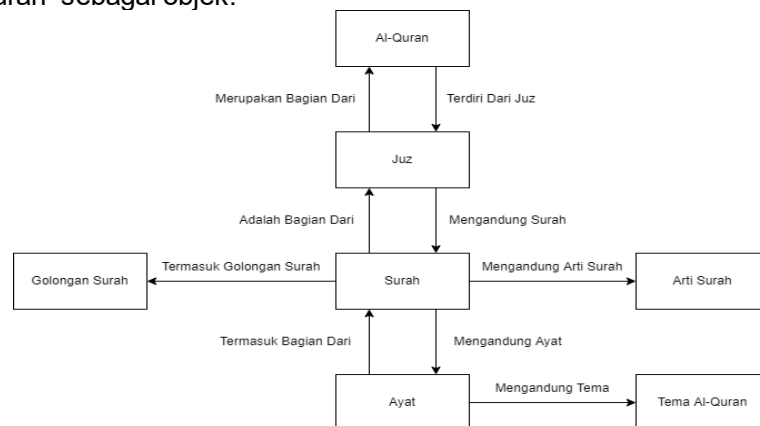
Pada tahap ini, teknik-teknik yang digunakan penulis untuk mengakuisisi pengetahuan dalam ontologi AI-Quran adalah sebagai berikut.

1. Berdiskusi dengan dosen pembimbing maupun mempelajari sumber terkait untuk membangun draf awal dokumen spesifikasi persyaratan.
2. Melakukan analisis teks informal, untuk mempelajari konsep-konsep utama yang diberikan dalam buku dan studi pegangan.
3. Melakukan analisis teks formal. Dalam hal ini, hal yang dilakukan adalah mengidentifikasi struktur yang akan dideteksi (definisi, penegasan, dan lain-lain) dan jenis pengetahuan yang dikontribusikan oleh masing-masing (konsep, atribut, nilai, dan hubungan).

c. Tahap Konseptualisasi

Pada tahap ini, pengetahuan yang didapatkan selama proses akuisisi pengetahuan dikelola dan diatur menjadi model konseptual. Setelah dibangun, model konseptual diubah menjadi model model formal, yang selanjutnya diimplementasikan menggunakan bahasa implementasi ontologi. Konseptual dari ontologi AI-Quran dapat dilihat pada Gambar 4.

Pada Gambar 4, terlihat gambaran dari model formal ontologi, dimana setiap relasinya dapat memiliki triplet (subjek-predikat-objek). Seperti contoh adalah pada triplet dari Surah, individual dari "Surah" menjadi subjek, "Mengandung Arti Surah" (*object properties*) sebagai predikat, dan individual "Arti Surah" sebagai objek.



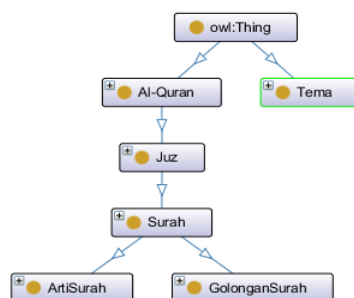
Gambar 4. Konseptual Ontologi AI-Quran

d. Tahap Integrasi

Pada tahap ini, diintegrasikan model ontologi dibangun ke dalam kerangka yang telah dikonsepsikan bersama dengan ahli ontologi.

e. Tahap Implementasi

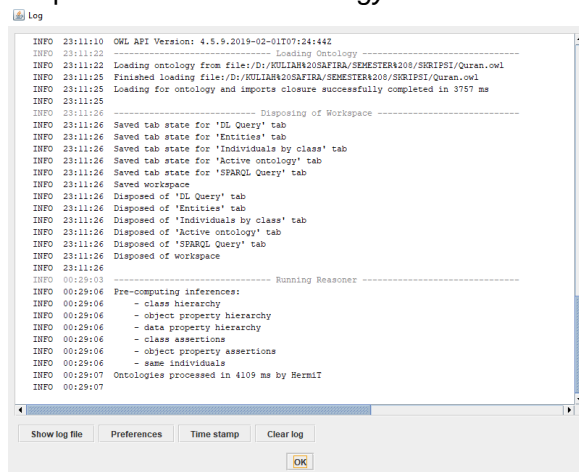
Pada tahap ini, perancangan konseptual ontologi dilakukan dengan menggunakan metode Methontologi yang kemudian diformalisasikan menggunakan perangkat lunak Protégé 5.5.0. Hasil perancangan ontologi AI-Quran yang merupakan ontologi yang dihasilkan dari perancangan sebelumnya dapat dilihat pada Gambar 5.



Gambar 5. Diagram Ontograf Sistem Manajemen Informasi AI-Quran

f. Tahap Evaluasi

Pada tahap ini, dilakukan proses *Reasoner* dengan Hermit untuk memetakan bahasa ontologi ke formalisme logis dengan melihat apakah bahasa ontologi yang dibangun sudah konsisten atau belum. Pada Gambar 6, proses *Reasoner* telah berhasil dijalankan dan tidak terjadi inkonsisten dalam memetakan Bahasa ontologi yang dibuat. Hal ini dapat diketahui karena tidak adanya pesan error yang muncul saat melakukan proses *Reasoners Ontology*.

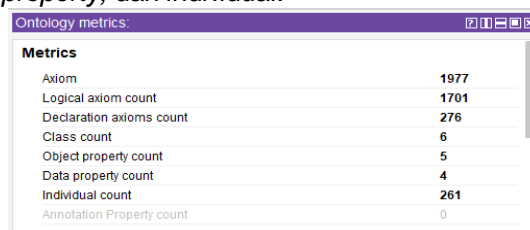


```
Log
INFO 23:11:10 OWL API Version: 4.5.9.2019-02-01T07:24:44Z
INFO 23:11:22 Loading Ontology -----
INFO 23:11:25 Finished loading file: /D:/MULIAH2020SAFIRA/SEMESTER4200/SRIPIS1/Quran.owl
INFO 23:11:25 Loading for ontology and imports closure successfully completed in 3757 ms
INFO 23:11:25 ----- Disposing of Workspace -----
INFO 23:11:26 Saved tab state for 'DL Query' tab
INFO 23:11:26 Saved tab state for 'Entities' tab
INFO 23:11:26 Saved tab state for 'Individuals by class' tab
INFO 23:11:26 Saved tab state for 'Active ontology' tab
INFO 23:11:26 Saved tab state for 'SPARQL Query' tab
INFO 23:11:26 Saved workspace
INFO 23:11:26 Disposed of 'DL Query' tab
INFO 23:11:26 Disposed of 'Entities' tab
INFO 23:11:26 Disposed of 'Individuals by class' tab
INFO 23:11:26 Disposed of 'Active ontology' tab
INFO 23:11:26 Disposed of 'SPARQL Query' tab
INFO 23:11:26 Disposed of workspace
INFO 00:29:03 ----- Running Reasoner -----
INFO 00:29:06 Pre-computing inferences:
INFO 00:29:06 - class hierarchy
INFO 00:29:06 - object property hierarchy
INFO 00:29:06 - data property hierarchy
INFO 00:29:06 - class assertions
INFO 00:29:06 - object property assertions
INFO 00:29:06 - same individuals
INFO 00:29:07 Ontologies processed in 4109 ms by Hermit
INFO 00:29:07
```

Gambar 6. Log Proses Reasoning Ontologi AI-Quran

g. Tahap Dokumentasi

Pada tahap ini, dilakukan dengan memproses dokumentasi ontologi AI-Quran baik dalam kode ontologi, teks bahasa alami yang dilampirkan pada definisi formal, maupun makalah yang diterbitkan dalam proses konferensi dan jurnal yang mengatur pertanyaan-pertanyaan penting dari ontologi yang sudah dibangun. Pada Gambar 7, tersusun *ontology metrics* yang memberikan gambaran secara sistematis mengenai komponen yang ada dalam rancangan ontologi AI-Quran seperti jumlah *class*, *object property*, *data property*, dan *individual*.



Ontology metrics:	
Metrics	
Axiom	1977
Logical axiom count	1701
Declaration axioms count	276
Class count	6
Object property count	5
Data property count	4
Individual count	261
Annotation Property count	0

Gambar 7. Metriks Ontologi AI-Quran

3.2. Implementasi Sistem

Pada bagian ini akan dijabarkan mengenai implementasi dari sistem sesuai dengan tahapan yang telah ditentukan.

a. Lingkup Implementasi

Terdapat beberapa perangkat lunak maupun *library* yang digunakan dalam pembuatan sistem manajemen informasi AI-Quran yang berupa Windows 10 64-bit, Protégé, PHP 8.0.15, Laravel 8, Visual Studio Code, Diagrams.net, Bootstrap 5, Apache Jena Fuseki, Google Chrome, dan Microsoft Office Excel 2016.

b. Implementasi Ontologi ke Dalam Sistem

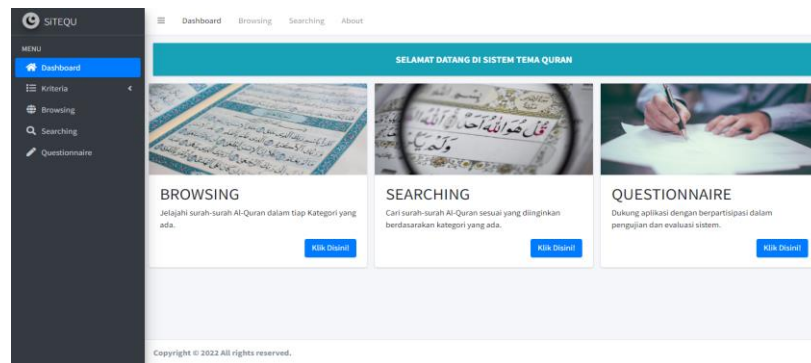
Pada tahap ini, terdiri dari proses mengunggah file ontologi ke dalam server Fuseki agar dapat digunakan oleh sistem. Kemudian dilakukan proses koneksi untuk menghubungkan antar ontologi dengan sistem yang telah dibangun.

c. Implementasi Antarmuka Sistem

Pada penelitian ini, antarmuka sistem diimplementasikan menggunakan bahasa pemrograman HTML dan CSS dengan *framework* Laravel. Berikut merupakan hasil implementasi dari sistem yang telah dibuat.

1. Antarmuka Halaman Utama (*Dashboard*)

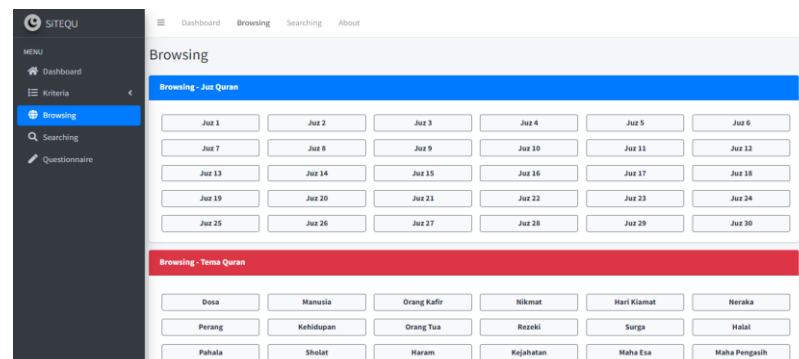
Pada halaman utama seperti yang dapat dilihat pada Gambar 8, terdapat daftar fitur-fitur yang dapat digunakan oleh *user* seperti fitur penjelajahan (*browsing*), fitur pencarian (*searching*), serta kuesioner. Pada halaman ini, *user* dapat memilih tautan yang diinginkan.



Gambar 8. Implementasi Antarmuka Halaman Utama (*Dashboard*)

2. Antarmuka Halaman Penjelajahan (*Browsing*)

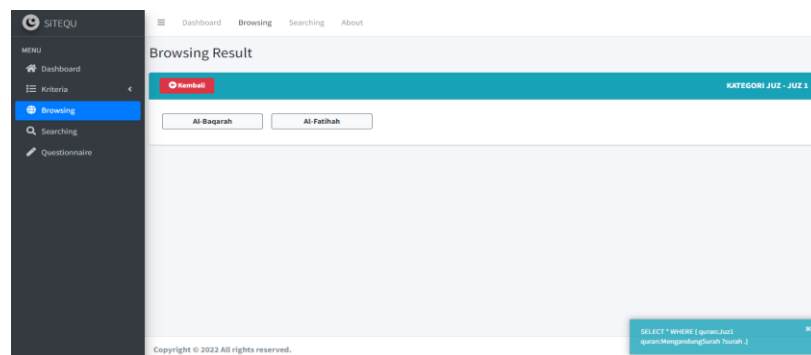
Pada halaman penjelajahan seperti yang dapat dilihat pada Gambar 9, ditampilkan semua data individual atau kriteria dari ontologi AI-Quran yang telah dibangun yaitu data juz, surah, golongan surah, dan tema yang berkaitan dengan AI-Quran.



Gambar 9. Implementasi Antarmuka Halaman Penjelajahan (*Browsing*)

3. Antarmuka Halaman Hasil Penjelajahan

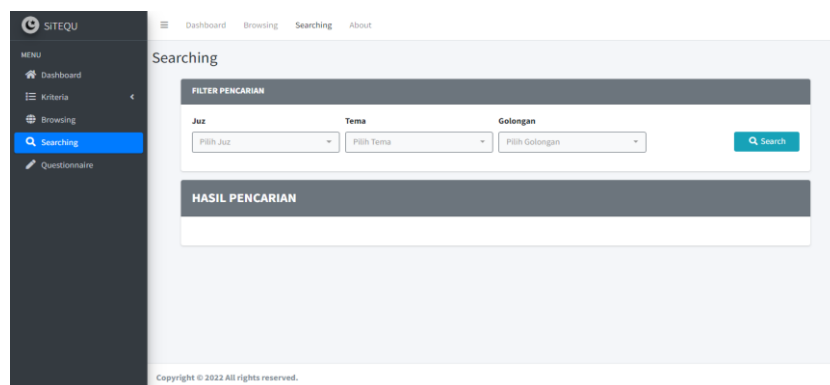
Pada halaman hasil penjelajahan seperti yang dapat dilihat pada Gambar 10, ditampilkan hasil dari bagian individual yang dipilih atau dibuka oleh *user*.



Gambar 10. Implementasi Antarmuka Halaman Hasil Penjelajahan (*Browsing*)

4. Antarmuka Halaman Pencarian (*Searching*)

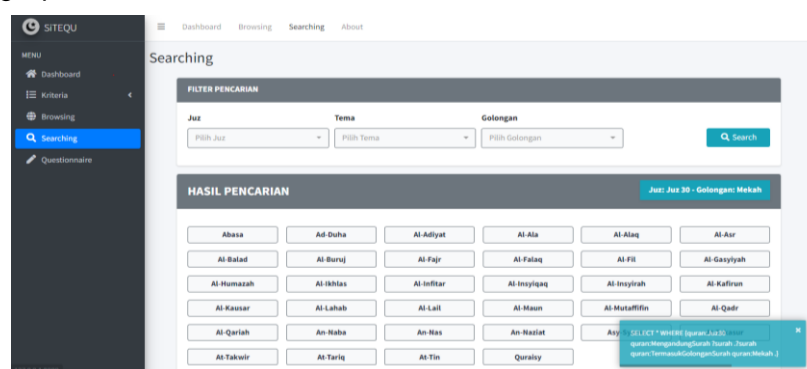
Pada halaman pencarian seperti yang dapat dilihat pada Gambar 11, *user* dapat memilih *input* kriteria yang diinginkan untuk mendapatkan *output* yang berupa surah AI-Quran. Pada halaman ini, *user* dapat melakukan pencarian dengan memasukkan kriteria yang ada pada sistem yaitu juz, tema, dan golongan surah.



Gambar 11. Implementasi Antarmuka Halaman Pencarian (*Searching*)

5. Antarmuka Halaman Hasil Pencarian

Pada halaman hasil pencarian seperti yang dapat dilihat pada Gambar 12, ditampilkan hasil dari pencarian yang dipilih oleh *user*.



Gambar 12. Implementasi Antarmuka Halaman Pencarian (*Searching*)

d. Pengolahan dan Analisis Data

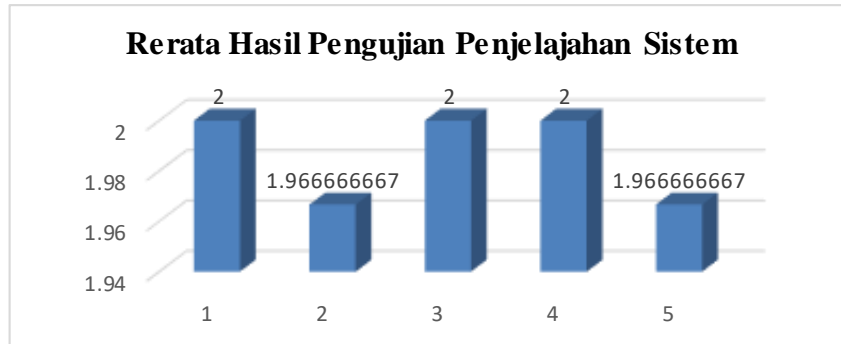
Pada penelitian ini, proses pengujian sistem dilakukan oleh 30 partisipan dengan masing-masing partisipan menjawab beberapa pertanyaan yang sama yang berkaitan dengan sistem. Kemudian data yang didapatkan dari hasil jawaban peserta pengujian diolah dan dianalisis sehingga didapatkan hasil sebagai berikut.

1. Pengolahan Data Pengujian Tugas Penjelajahan (*Browsing*)

Tabel 1. Hasil Klasifikasi Skim Penandaan Tugas Penjelajahan

Kategori	Skala	Jumlah	
		Kiriman Jawaban	Persen
Salah	0	0	0%
Sebagian Benar	1	2	1,33%
Sepenuhnya Benar	2	148	98,67%
Total		150	100%

Dari hasil klasifikasi skim penandaan tugas penjelajahan yang dapat dilihat pada Tabel 1, dapat disimpulkan bahwa kebanyakan peserta dapat memberikan jawaban yang sepenuhnya benar. Rerata hasil pengujian akurasi penjelajahan sistem dapat dilihat pada Gambar 13.



Gambar 13. Grafik Rerata Hasil Pengujian Akurasi Penjelajahan Sistem

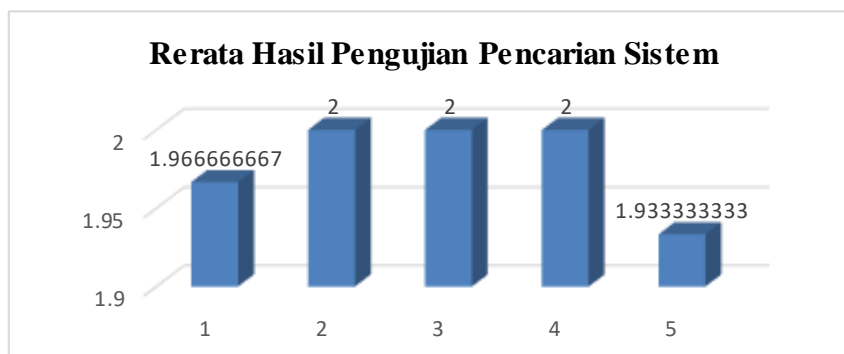
Jika dilihat dari grafik rerata hasil pengujian akurasi penjelajahan sistem, maka rerata skor tertinggi terdapat pada pertanyaan pertama (P1), pertanyaan ketiga (P3), dan pertanyaan keempat (P4) dengan nilai skor 2. Hal ini menandakan bahwa pertanyaan pertama, ketiga dan keempat merupakan pertanyaan yang dijawab oleh peserta pengujian dengan tingkat kebenaran tertinggi. Sedangkan rerata skor terendah terdapat pada pertanyaan kedua (P2) dan pertanyaan kelima (P5) dengan nilai skor 1,967. Hal ini menandakan bahwa pertanyaan kedua dan kelima merupakan pertanyaan yang dijawab oleh peserta dengan tingkat kebenaran terendah.

2. Pengolahan Data Pengujian Tugas Pencarian (*Searching*)

Tabel 2. Hasil Klasifikasi Skim Penandaan Tugas Pencarian

Kategori	Skala	Jumlah	
		Kiriman Jawaban	Persen
Salah	0	0	0%
Sebagian Benar	1	3	2,00%
Sepenuhnya Benar	2	147	98,00%
Total		150	100%

Dari hasil klasifikasi skim penandaan tugas pencarian yang dapat dilihat pada Tabel 2, dapat disimpulkan bahwa kebanyakan peserta dapat memberikan jawaban yang sepenuhnya benar. Rerata hasil pengujian akurasi penjelajahan sistem dapat dilihat pada Gambar 14.



Gambar 14. Grafik Rerata Hasil Pengujian Akurasi Pencarian Sistem

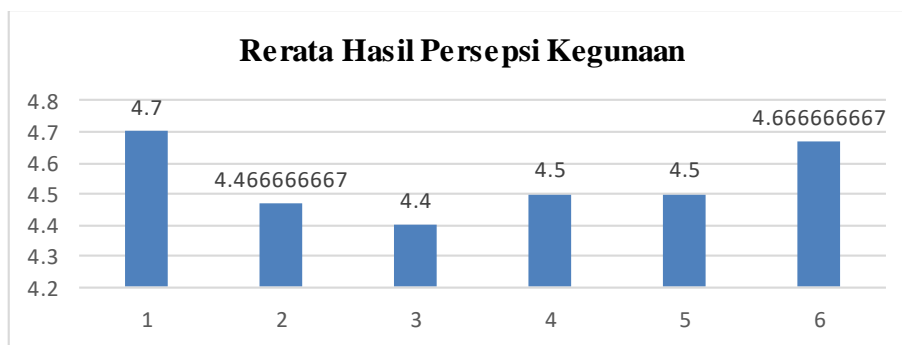
Jika dilihat dari grafik rerata hasil pengujian akurasi penjelajahan sistem, maka rerata skor tertinggi terdapat pada pertanyaan kedua (P2), pertanyaan ketiga (P3), dan pertanyaan keempat (P4) dengan nilai skor 2. Hal ini menandakan bahwa pertanyaan kedua, ketiga dan keempat merupakan pertanyaan yang dijawab oleh peserta pengujian dengan tingkat kebenaran tertinggi. Sedangkan rerata skor terendah terdapat pada pertanyaan kelima (P5) dengan nilai skor 1,933. Hal ini menandakan bahwa pertanyaan kelima merupakan pertanyaan yang dijawab oleh peserta dengan tingkat kebenaran terendah.

3. Pengolahan Data Evaluasi Persepsi Kegunaan yang Dirasakan

Tabel 3. Persentase Jawaban Data Evaluasi Persepsi Kegunaan

Kategori	Skala	Jumlah	
		Total Nilai	Persen
Sangat Tidak Setuju	1	0	0%
Tidak Setuju	2	0	0%
Netral	3	7	4%
Setuju	4	69	38%
Sangat Setuju	5	104	58%
Total		180	100%

Tabel 3 merupakan hasil analisis statistik dari data evaluasi persepsi kegunaan yang dirasakan yang telah dilakukan dengan menggunakan perangkat lunak Microsoft Excel. Rerata hasil pengujian akurasi persepsi kegunaan yang dirasakan dapat dilihat pada Gambar 15.



Gambar 15. Grafik Rerata Hasil Evaluasi Persepsi Kegunaan

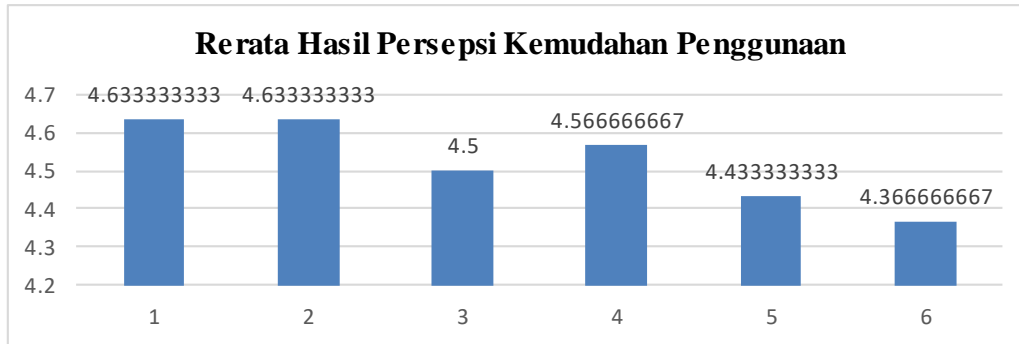
Jika dilihat dari grafik rerata hasil evaluasi persepsi kegunaan sistem, maka rerata skor tertinggi terdapat pada pertanyaan pertama (P1) dengan nilai skor 4,7. Hal ini menandakan bahwa persepsi kesetujuan peserta pengujian bahwa “sistem yang dibangun memungkinkan untuk menyelesaikan tugas lebih cepat” adalah yang tertinggi. Sedangkan rerata skor terendah terdapat pada pertanyaan ketiga (P3) dengan skor 4,4. Hal ini menandakan bahwa persepsi kesetujuan peserta pengujian bahwa “sistem yang dibangun meningkatkan produktivitas dalam pekerjaan” adalah yang terendah.

4. Pengolahan Data Evaluasi Persepsi Kemudahan Penggunaan

Tabel 4. Persentase Jawaban Data Evaluasi Persepsi Kemudahan Penggunaan

Kategori	Skala	Jumlah	
		Total Nilai	Persen
Sangat Tidak Setuju	1	0	0%
Tidak Setuju	2	0	0%
Netral	3	6	3%
Setuju	4	68	38%
Sangat Setuju	5	106	59%
Total		180	100%

Tabel 4 merupakan hasil analisis statistik dari data evaluasi persepsi kemudahan penggunaan yang telah dilakukan dengan menggunakan perangkat lunak Microsoft Excel. Rerata hasil pengujian akurasi persepsi mudahan penggunaan dapat dilihat pada Gambar 16.



Gambar 16. Grafik Rerata Hasil Evaluasi Persepsi Kemudahan Penggunaan

Jika dilihat dari grafik rerata hasil evaluasi persepsi kemudahan penggunaan sistem, maka rerata skor tertinggi terdapat pada pertanyaan pertama (P1) dan pertanyaan kedua (P2) dengan skor 4,633. Hal ini menandakan bahwa persepsi kesetujuan peserta pengujian bahwa “sistem yang dibangun mudah untuk dipelajari dan digunakan” adalah yang tertinggi. Sedangkan rerata skor terendah terdapat pada pertanyaan ketiga (P6) dengan skor 4,367. Hal ini menandakan bahwa persepsi kesetujuan peserta pengujian bahwa “sistem yang dibangun berguna dalam pekerjaan” adalah yang terendah.

4. Kesimpulan

Kesimpulan yang didapatkan dari penelitian ini yaitu telah diimplementasikan metode Methontology dalam mengembangkan ontologi pada domain AI-Quran, yang terdiri dari tahap spesifikasi, tahap akuisisi pengetahuan, tahap konseptualisasi, tahap integrasi, tahap implementasi, tahap evaluasi, dan tahap dokumentasi. Data yang didapatkan terdiri dari 6 *class*, 4 *data property*, 5 *object property*, dan 261 *individual*. Terdapat dua fitur utama pada sistem yaitu fitur penjelajahan (*browsing*) dan pencarian (*searching*). Kemudian dari pengujian yang dilakukan terhadap sistem dengan melibatkan 30 peserta pengujian, didapatkan hasil pengujian berupa nilai persentase akurasi dari fitur penjelajahan dan pencarian yaitu masing-masing sebesar 99,33% dan 99%. Berdasarkan hasil tersebut, maka dapat disimpulkan bahwa hasil dari penjelajahan dan pencarian sistem telah dapat dianggap akurat, dengan rata-rata peserta dapat menjawab tugas penjelajahan dan tugas pencarian sepenuhnya benar. Sedangkan berdasarkan hasil evaluasi yang diberikan peserta terhadap sistem mendapatkan hasil evaluasi berupa nilai persentase persepsi kegunaan yang dirasakan dan persepsi kemudahan penggunaan masing-masing sebesar 90,78% dan 90,44%. Berdasarkan hasil tersebut, maka dapat disimpulkan bahwa hasil rata-rata peserta sangat setuju bahwa sistem yang dibangun adalah sistem yang memiliki kegunaan dan mudah digunakan.

Daftar Pustaka

- [1] A. Ta'a, Q. A. Abed, and M. Ahmad, "AI-Quran ontology based on knowledge themes," *J. Fundam. Appl. Sci.*, vol. 9, no. 5S, p. 800, 2018, doi: 10.4314/jfas.v9i5s.57.
- [2] C. Pramatha, P. S. Informatika, U. Udayana, D. P. Kelod, and K. Denpasar, "PENGEMBANGAN ONTOLOGI TUJUAN WISATA BALI DENGAN," vol. 3, no. 2, pp. 77–89, 2020.
- [3] A. Ta'a, Q. A. Abed, B. M. Ali, and M. Ahmad, "Ontology-Based Approach for Knowledge Retrieval in AI-Quran Holy Book," *Int. J. Comput. Eng. Res. Ontol.*, vol. 06, no. 03, pp. 8–15, 2016.
- [4] M. Fernandez, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," *Proc. AAAI97 Spring Symp. Ser. Ontol. Eng.*, no. March, pp. 33–40, 1997, [Online]. Available: <http://speech.inesc.pt/~joana/prc/artigos/06c METHONTOLOGY from Ontological Art towards Ontological Engineering - Fernandez, Perez, Juristo - AAAI - 1997.pdf>.
- [5] C. Pramatha and J. G. Davis, "Digital preservation of cultural heritage: Balinese Kukul artefact and practices," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10058 LNCS, pp. 491–500, 2016, doi: 10.1007/978-3-319-48496-9_38.

Implementasi Metode *K-Nearest Neighbor* Dalam Mengklasifikasikan Jenis Suara Berdasarkan Jangkauan Vokal

Ni Made Putri Wahyuni^{a1}, Luh Arida Ayu Rahning Putri^{a2}, I Gusti Ngurah Anom Cahyadi Putra^{b3}, I Dewa Made Bayu Atmaja Darmawan^{b4}, Made Agung Raharja^{b5}, Agus Muliantara^{b6}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bali, Indonesia

¹nimadeputriwahyuni29@email.com (Corresponding author)

²rahningputri@unud.ac.id

³anom.cp@unud.ac.id

⁴dewabayu@unud.ac.id

⁵made.agung@unud.ac.id

⁶muliantara@unud.ac.id

Abstract

Humans have voice characteristics with different vocal ranges, namely the male voice consists of Tenor, Baritone, and Bass, while the female voice consists of Soprano, Mezzosoprano, and Alto. Determining the voice range, especially for a singer, requires a vocal trainer or musical instrument that is quite difficult to access. Therefore, a sound classification system created based on vocal range using the Harmonic Product Spectrum (HPS) feature extraction method and the K-Nearest Neighbors (KNN) classification method uses k parameters from 1 to 40. The test gets the highest accuracy on parameter k=8, which is 88.88%, so that from the resulting accuracy to prove the K-Nearest Neighbor (KNN) method gives good results in classifying the type of voice.

Keywords: Classification, Vocal range, Harmonic Product Spectrum, K-Nearest Neighbors

1. Pendahuluan

Bagi penyanyi pemula dalam meng-cover sebuah lagu, banyaknya pilihan lagu dengan jangkauan nada yang bervariasi seringkali membuat orang yang memiliki hobi bernyanyi salah dalam memilih lagu. Hal ini biasanya yang membuat antara suara dan instrumen lagu tidak sesuai. Untuk melakukan cover terhadap lagu, seseorang harus menyesuaikan suara yang dimiliki dengan nada instrumen pada lagu tersebut. Oleh karena itu untuk mengatasi masalah tersebut diperlukan pengolahan suara yang dapat mendeteksi nada vokal dan mengidentifikasi jenis suara berdasarkan jangkauan vokal, sehingga dapat membantu para penyanyi pemula dalam meng-cover lagu[1].

Penelitian identifikasi jenis suara juga dilakukan oleh Annisa [2] dalam penelitian ini suara yang diteliti berfokus pada suara anak-anak, peneliti menggunakan metode Fast Fourier Transform (FFT). FFT dapat mengubah sinyal dari domain waktu ke domain frekuensi, metode ini digunakan untuk melihat frekuensi yang dihasilkan oleh suara anak tersebut kemudian diklasifikasikan dengan tipe suara. Namun dalam penelitian [2] ini hanya mengetahui frekuensi yang dihasilkan dari suara yang telah direkam menggunakan metode Fast Fourier Transform (FFT). Proses pengolahan data suara akan dihasilkan beberapa macam frekuensi suara, yaitu frekuensi untuk pelafalan suara a untuk mengetahui frekuensi nada dasar, suara do re mi fa sol la si do' digunakan untuk mendeteksi range frekuensi rata-rata suara dan suara do' digunakan untuk mendeteksi suara tertinggi yang dapat dijangkau. Dari hasil data frekuensi tersebut akan dilihat jangkauan frekuensi yang dihasilkan, yang nantinya akan digunakan untuk menentukan tipe suara[3].

Berdasarkan paparan penelitian yang telah dilakukan sebelumnya, penulis tertarik dalam mengklasifikasikan jenis suara berdasarkan jangkauan vokal dengan menggunakan metode klasifikasi KNN. Pemilihan metode K-NN dilakukan karena metode KNN merupakan metode yang paling sederhana, sehingga akan berpengaruh pada efisiensi waktu. Algoritma Harmonic Product Spectrum (HPS) digunakan sebagai metode ekstraksi fitur untuk melihat frekuensi yang terdapat pada sinyal suara, sinyal suara tersebut sebelumnya telah diubah ke domain frekuensi menggunakan Fast Fourier

Transform (FFT). Hasil dari ekstraksi fitur menggunakan metode HPS ini kemudian akan klasifikasi menggunakan klasifikasi *K-Nearest Neighbor* (KNN) untuk menentukan jenis suara, yaitu sopran, mezzo-sopran, alto, tenor, baritone, atau bass.

2. Metode Penelitian

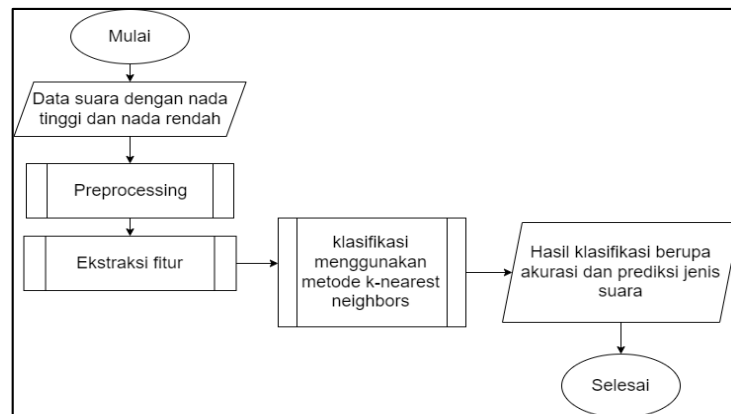
2.1. Pengumpulan Data

Dalam penelitian ini menggunakan data primer suara yang diperoleh langsung dari melakukan perekaman kepada 30 pria dan 30 wanita yang mewakili semua kelas, setiap partisipan memiliki rentang usia antara 15-23 tahun. Terdapat 10 orang dengan masing-masing jenis suara, setiap kelas telah divalidasi oleh guru vokal. Proses perekaman dilakukan dua kali kepada masing-masing partisipan, dengan melafalkan huruf vokal "a" dengan mengambil nada terendah dan nada tertinggi yang bisa dicapai oleh masing-masing partisipan. File suara yang direkam disimpan dalam format (.wav), *sample rate* yang digunakan yaitu dengan frekuensi sebesar 44100 Hz yang terdapat 44100 sampel per detik. Jumlah data yang didapatkan sebanyak 60, yang memiliki fitur dengan frekuensi fundamental nada tinggi dan frekuensi fundamental nada tinggi.

2.2. Perancangan Sistem

Tahapan pertama merupakan *input* data yang diambil dari hasil rekaman suara berupa melafalkan huruf vokal "a" dengan mengambil nada terendah dan nada tertinggi yang bisa dicapai setiap partisipan. Data rekaman menggunakan format "wav" dengan frekuensi *sampling* yang digunakan pada masing-masing suara adalah 44100 Hz. Kemudian amplitudonya dinormalisasikan, setelah itu sinyal dari stereo ke mono. Selanjutnya dilakukan tahap *windowing* dengan *hamming window* untuk mencegah kebocoran frekuensi ketika proses FFT. Keluaran dari proses FFT adalah *spectrum* frekuensi yang terdiri dari fundamental frekuensi dan harmoniknya. Spektrum ini yang dimasukkan ke dalam algoritma *Harmonic Product Spectrum* untuk dikompresi dan kemudian menentukan frekuensi dari nada yang terdeteksi pada sinyal suara.

Proses HPS akan mengalikan nilai spektrum frekuensi untuk menghilangkan sinyal harmonik pada frekuensi dan dapat mengenali frekuensi dari nada yang terdeteksi pada sinyal suara masukan. Klasifikasi KNN untuk melakukan pencocokan objek berdasarkan data yang paling dekat dengan objek tersebut. Secara umum, langkah-langkah perancangan sistem digambarkan dalam diagram alir seperti gambar 1.



Gambar 1. Diagram alir perancangan sistem.

2.3. Preprocessing

Dalam proses ini, masukan mengalami beberapa proses sehingga menghasilkan keluaran yang nantinya menjadi masukan untuk proses selanjutnya. Adapun tahap preprocessing terdiri dari tahap normalisasi yaitu menyamakan rentang amplitudo, selanjutnya mengubah sinyal audio menjadi monostereo.

2.3.1. Normalisasi Sinyal Suara

Proses ini bertujuan untuk menyamakan amplitudo dari setiap suara nyanyian yang direkam oleh sistem sehingga berada dalam rentang -1 dan +1. Proses normalisasi awal dilakukan dengan cara

membagi tiap nilai data masukan yaitu suara terekam dengan nilai absolut maksimal dari data masukan tersebut. Berikut rumus untuk proses normalisasi [4].

$$X_{norm} = \frac{x}{\max(|x|)} \quad (1)$$

Dimana :

X_{norm} = hasil data sinyal normalisasi

x = data input

2.3.2 Mengubah Stereo ke Mono

Pada tahap ini, data suara akan diubah dari stereo ke mono dengan mencari rata-rata dua saluran pada data stereo yang akan dikonversi menjadi hanya satu saluran (mono). Data yang semula adalah kanal stereo dan memiliki frekuensi sampling sebesar 44100 Hz dirubah menjadi kanal mono dengan cara mengubah data yang dua kanal menjadi satu kanal atau satu kolom matriks datanya saja. Tahap ini berguna untuk mempersingkat waktu kerja sistem karena mengolah data yang lebih sedikit. Hal ini tidak menghilangkan informasi karena matriks data kolom satu tidak jauh berbeda dengan matriks kolom dua[5].

2.4. Ekstraksi Fitur

Adapun tahap ekstraksi fitur terdiri dari tahap *windowing* menggunakan *hamming window*, kemudian dilanjutkan dengan tahap mengubah sinyal suara kedalam domain frekuensi dengan metode *Fast Fourier Transform* (FFT). Selanjutnya dilakukan tahap mencari frekuensi dari nada yang terdeteksi pada sinyal suara dengan menggunakan metode *Harmonic Product Spectrum* (HPS)[6].

2.4.1. Hamming Window

Hamming Window mempunyai *side lobe* yang paling kecil dan *main lobe* yang paling besar sehingga hasil *windowing* akan lebih halus dalam menghilangkan efek diskontinuitas. Ketika digunakan FFT untuk mengukur frekuensi dari data, maka harus ditentukan dasar analisa pada sebuah data yang pasti. FFT mengasumsikan bahwa data yang tetap adalah sebuah periode dari period signa [7]!. Secara matematis fungsi *hamming window* dituliskan dengan persamaan :

$$w(n) = \alpha - \beta \left(\frac{2\pi n}{N-1} \right) \quad (2)$$

Dimana:

$$\alpha = 0.54$$

$$\beta = 1 - \alpha$$

$$= 0.46$$

2.4.2. Fast Fourier Transform (FFT)

Proses selanjutnya adalah transformasi sinyal untuk menghasilkan spektrum dari sinyal hasil *windowing*, *spectrum* domain waktu diubah menjadi sinyal dalam domain frekuensi dengan menggunakan algoritma *Fast Fourier Transform* (FFT). Spektrum yang dibangkitkan inilah yang akan dianalisis untuk mengetahui frekuensi dari nada yang terdeteksi oleh sinyal suara. Proses FFT sinyal dibagi menjadi beberapa bagian yang lebih kecil yang bertujuan untuk memperoleh waktu proses yang lebih cepat. DFT adalah prosedur matematis yang digunakan untuk menentukan harmonik, atau frekuensi serta isi dari urutan sinyal diskrit [2]. Berikut rumus dasar *Discrete Fourier Transformation*:

$$F(k) = \sum_{n=0}^N f(n) \cos\left(\frac{j2\pi nk}{N}\right) - j f(n) \sin\left(\frac{j2\pi nk}{N}\right) \quad (3)$$

Untuk mendapatkan nilai j menggunakan persamaan seperti persamaan 3

$$|f(\mu)| = [R^2 + l^2]^{1/2} \quad (4)$$

Dimana :

N = jumlah sampel yang akan diproses.

x(n) = nilai sampel signal

k = variable frekuensi discrete

$\pi = 3,14$

n = indeks data nilai sampling

2.4.3. Harmonic product spectrum (HPS)

Harmonic Product Spectrum merupakan metode *pitch detection algorithm* yang paling simpel dan bekerja dengan baik dalam berbagai kondisi, metode HPS berfungsi untuk melihat frekuensi dasar yang terdapat pada sinyal input. HPS adalah suatu metode yang berfungsi untuk melihat frekuensi dasar yang terdapat pada sinyal input [8]. Secara matematis HPS dapat dirumuskan seperti persamaan 5

$$HPS(k) = (\prod_{n=1}^N Y(n, k)) \quad (5)$$

Dimana:

HPS = hasil spektrum harmonik

k = indeks frekuensi spektrum harmonik

Y = besar spektrum pada frekuensi positif

N = jumlah harmonik yang digunakan

2.5. Klasifikasi *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini banyak digunakan dalam bidang pengenalan pola, klasifikasi KNN didasarkan pada membandingkan sebuah data uji dengan sejumlah data latih. Data latih terdiri dari n atribut, tiap data merepresentasikan titik pada sebuah ruang berdimensi n, dengan begitu semua data latih disimpan di dalam ruang pola berdimensi n. Ketika diberikan sebuah data yang tidak diketahui kelasnya, *k-nearest neighbor* akan mencari pola ruang untuk data latih k yang terdekat. Data latih k ini merupakan k "*nearest neighbor*" dari data yang tidak diketahui tersebut [9]. Jarak antara data latih dan data uji dihitung menggunakan persamaan Euclidean Distance [10]. *Euclidean Distance* dapat didefinisikan pada persamaan 6.

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (6)$$

Dimana :

dist = jarak kedekatan

x_1 = data latih

x_2 = data uji

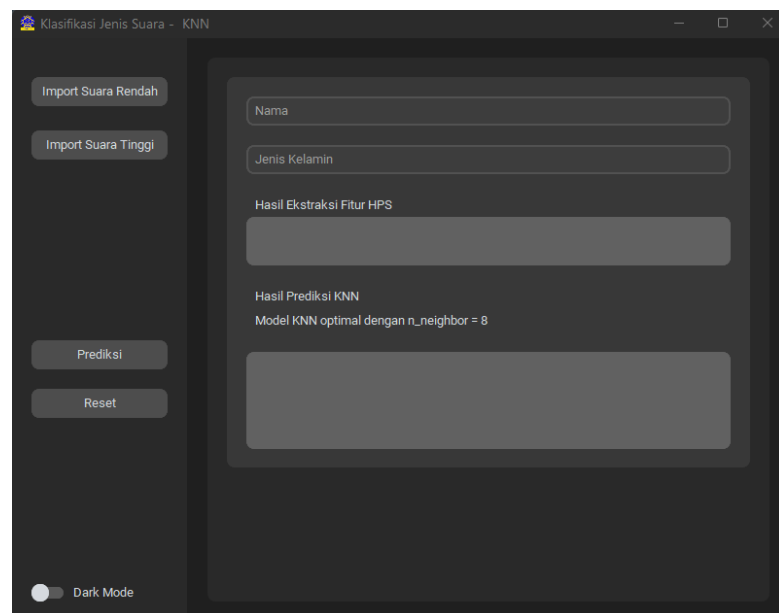
n = jumlah atribut antara 1 sampai n

i = atribut antara 1 sampai n

3. Hasil dan Pembahasan

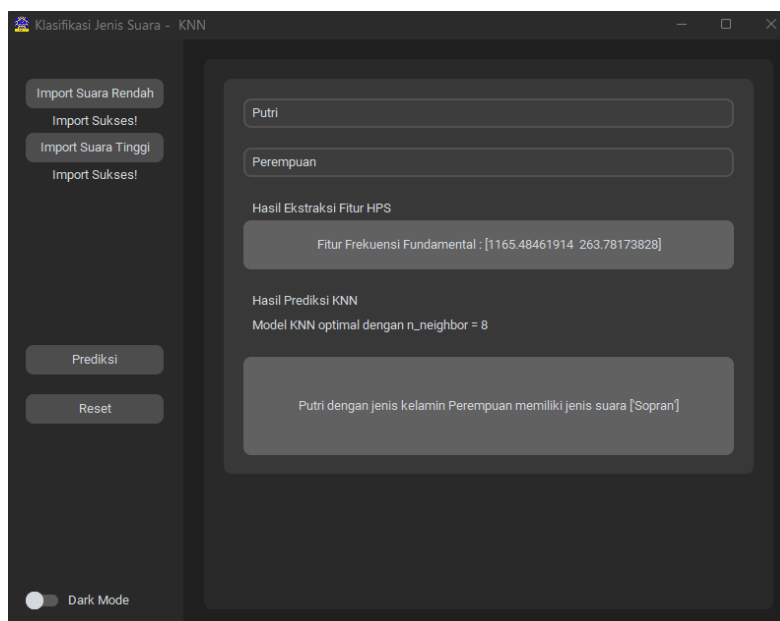
3.1. Tampilan Antar Muka Sistem

Antarmuka sistem terdiri dari satu halaman, Sistem menggunakan Bahasa pemrograman *python* berbasis GUI. Berikut pada gambar 2 merupakan tampilan antarmuka sebelum dijalankan.



Gambar 2. Implementasi antarmuka sistem

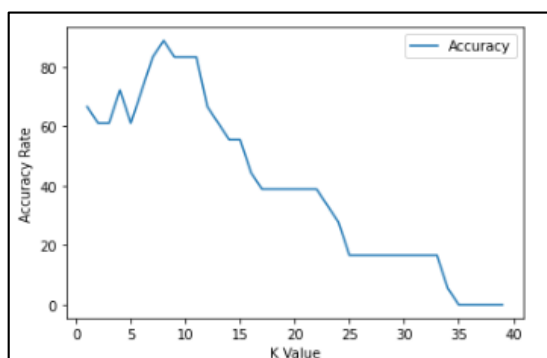
Pengguna dapat melakukan prediksi jenis suara dengan memasukkan data suara dengan format .wav. Pengguna memasukkan dua file suara yaitu suara terendah dan tertinggi yang bisa dicapai dengan menekan tombol "Import Suara Rendah" untuk data suara rendah dan tombol "Import Suara Tinggi" untuk data suara tinggi. Setelah data sukses dimasukkan, pengguna mengisikan data diri berupa nama dan jenis kelamin. Selanjutnya pengguna menekan tombol "Prediksi" untuk menjalankan proses ekstraksi fitur dan prediksi. Keluaran dari hasil prediksi akan menampilkan frekuensi fundamental dari hasil ekstraksi fitur dan menampilkan prediksi jenis suara dari pengguna. Tombol "Reset" digunakan untuk mengatur ulang sistem. Implementasi antar muka sistem setelah dijalankan ditunjukkan pada gambar 3.



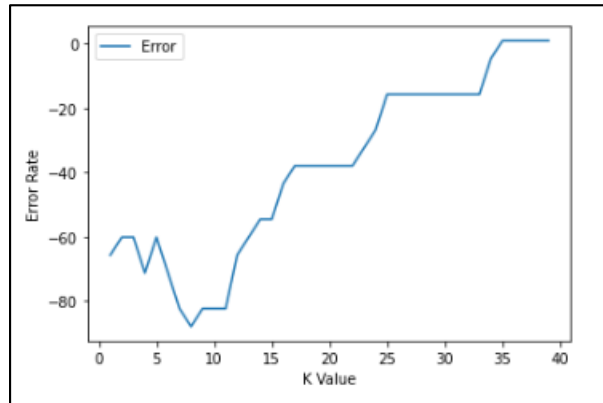
Gambar 2. Implementasi antarmuka sistem hasil

3.2. Pengujian Sistem

Hasil ekstraksi fitur terdiri dari dua fitur yaitu frekuensi tinggi dan frekuensi rendah, jumlah fitur HPS sebanyak 60 data. Pengujian jenis suara berdasarkan jangkauan vokal yaitu alto, mezzo, sopran, tenor, bariton, dan bass menggunakan ekstraksi fitur HPS dan klasifikasi KNN. Pembagian dataset antara data latih dan data uji yaitu 70:30. Jadi 70% merupakan data untuk pelatihan model sedangkan 30% merupakan data untuk pengujian model, dimana 42 data sebagai data uji dan 18 sebagai data latih. Penelitian ini menggunakan nilai k tetangga dari 1 sampai 40, berdasarkan pengujian pada setiap nilai k dilihat gambar 4, akurasi tertinggi diperoleh k=8 dengan akurasi sebesar 88.88%. Sedangkan dari k=14 sampai dengan k=40 memperoleh akurasi dibawah 60%.



Gambar 3. Grafik akurasi setiap parameter k

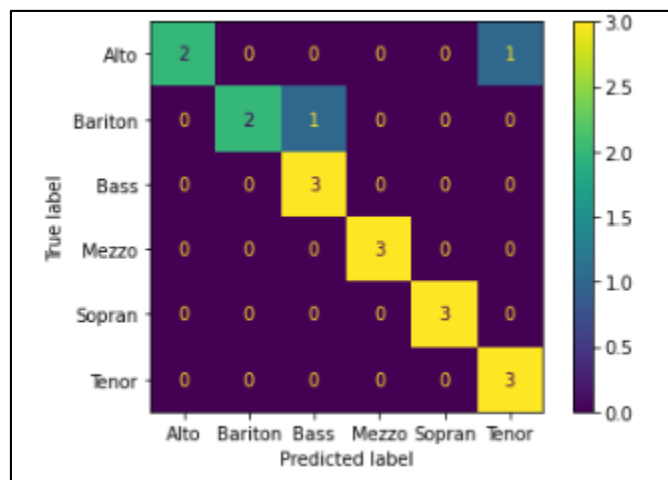


Gambar 4. Grafik Galat KNN

Berdasarkan hasil keseluruhan akurasi terlihat membentuk pola menurun dan terjadi pengulangan nilai akurasi pada beberapa pengujian nilai k, seperti pada pengujian k 17 – 22 dimana mendapatkan akurasi yang sama yaitu 38.8 % dan pengujian k 25 – 33 mendapatkan akurasi yang sama yaitu 16.6 %. Pengujian ini membuktikan bahwa model klasifikasi menggunakan metode KNN pada klasifikasi jenis suara berdasarkan jangkauan vokal memiliki pola menurun terlihat pada gambar 5, dimana semakin besar nilai k maka semakin kecil akurasi yang didapatkan.

Pengujian dilakukan dengan membagi dataset yang berjumlah 60 data menjadi dua bagian yang terdiri atas 48 data sebagai data training dan 18 data sebagai data uji. Proporsi jumlah data untuk masing-masing kelas data pada data training dan data uji dibuat 70% dan 30%. Proses training dilakukan dengan menggunakan metode *K-nearest neighbor* dengan nilai parameter k = 1 sampai dengan k = 40. Berdasar hasil pengujian model, dapat dilihat bahwa model yang terbaik adalah model dengan nilai parameter k=8. Pertimbangannya adalah model ini selain memiliki tingkat akurasi yang baik yaitu sebesar 88.88% .

Sehingga berdasarkan pengujian model dengan akurasi terbaik sebesar 88.88% pada parameter k = 8, hal ini dapat dibuktikan dari confusion matriks ditampilkan pada gambar 6, dapat dilihat gambar 6 pada setiap pengujian terdapat 3 data yang menjadi data uji setiap kelas. Untuk kelas dengan jenis suara alto terdeteksi benar sebanyak 2 dan 1 data terdeteksi dengan jenis suara tenor. Untuk kelas bariton terdeteksi benar sebanyak 2 dan 1 data terdeteksi dengan jenis suara bass. Untuk kelas bass, mezzo, sopran dan tenor semua terdeteksi dengan benar.



Gambar 5. Hasil prediksi k=8

4. Kesimpulan

Pada penelitian ini mengklafikasikan jenis suara berdasarkan jangkauan vokal menggunakan metode ekstraksi fitur *Harmonic Product Spectrum* (HPS) dan metode klasifikasi *K-Nearest Neighbor*

(KNN) dengan pengujian parameter nilai k, parameter k yang digunakan dari 1 sampai 40. Berdasarkan hasil penelitian dapat disimpulkan hasil sebagai berikut.

1. Akurasi tertinggi yang dihasilkan dari pengujian klasifikasi jenis suara berdasarkan jangkauan vokal menggunakan metode klasifikasi KNN dengan nilai $k = 8$ yaitu sebesar 88.88%, sehingga dari akurasi yang dihasilkan membuktikan metode *K-Nearest Neighbor* (KNN) memberikan hasil yang baik dalam mengklasifikasikan jenis suara.
2. Akurasi terendah pada $k = 14$ sampai $k=40$ dengan akurasi dibawah 60%, sehingga nilai k yang tinggi kurang bagus digunakan dalam mengklasifikasikan jenis suara.
3. Pengujian ini juga membentuk pola menurun antara akurasi dengan parameter k, dimana semakin tinggi parameter k maka akurasi semakin rendah.

Daftar Pustaka

- [1] F. Rifqi, *Aplikasi Rekomendasi Transpose Lagu Berdasarkan Vocal Range Menggunakan Algoritma Genetika Berbasis Android Aplikasi Rekomendasi Transpose Lagu Berdasarkan Vocal Range Menggunakan Algoritma Genetika Berbasis Android Skripsi*. 2019.
- [2] R. Nugra Annisa, D. Suprayogi, and H. Bethaningtyas, "Klasifikasi Suara Anak-Anak Dengan Menggunakan Metode Fast Fourier Transform Classification of Children'S Voices Using Fast Fourier Transform," vol. 6, no. 1, pp. 1141–1148, 2019.
- [3] P. D. Prasetyo, I. G. P. Suta Wijaya, and A. Yudo Husodo, "Klasifikasi Genre Musik Menggunakan Metode Mel-Frequency Cepstrum Coefficients dan K-Nearest Neighbors Classifier," *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 1, no. 2, pp. 189–197, 2019, doi: 10.29303/jtika.v1i2.41.
- [4] I. Wijayanto and R. Dwifebrianti, "Jenis Tipe Jangkauan Suara Pada Pria Dan Wanita Menggunakan Metoda Mel-Frequency Cepstral Coefficient," *Konfrensi Nas. Sist. dan Inform.*, no. October 2013, pp. 2–10, 2013.
- [5] G. Harsemadi, M. Sudarma, and N. Pramaita, "Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 1, pp. 14–20, 2017, doi: 10.24843/mite.1601.03.
- [6] W. S. M. Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferense System (ANFIS) sebagai Kontrol Lampu Otomatis," *AI-HAZEN J. Phys.*, vol. 1, no. 1, 2014, [Online]. Available: <http://journal.uinsgd.ac.id/index.php/ahjop/article/view/129>.
- [7] J. Adler, M. Azhar, and S. Supatmi, "Identifikasi Suara dengan MATLAB sebagai Aplikasi Jaringan Syaraf Tiruan Speech Recognition in MATLAB as Artificial Neural Network Application," vol. 1, no. 1, pp. 16–23, 2013.
- [8] V. A. Egisthi, D. Andreswari, and Y. Setiawan, "Aplikasi Latih Vokal Dengan Menggunakan Metode Harmonic Product Spectrum (Hps) Dan Boyer Moore Berbasis Android," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 2, p. 501, 2016, doi: 10.24176/simet.v7i2.761.
- [9] F. Darmadi, A. Rizal, and U. Sunarya, "Deteksi Sleep Apnea Melalui Analisis Suara Dengkuran Dengan Metode Mel Frekuensi Cepstrum," vol. 2, no. 2, pp. 2681–2686, 2015.
- [10] E. Susanti, S. Sasongko, and I. G. Suta W, "Klasifikasi Suara Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient (Mfcc)," vol. 4, no. 2, pp. 120–126, 2017.

This page is intentionally left blank

Basic Mathematics Learning Media System Based Semantic Web For Junior High School

Samson Cornelius Gele Yowe^{a1}, I Gede Santi Astawa^{a2}, I Ketut Gede Suhartana^{a3}, I Wayan Santiyasa^{a4}, Ida Bagus Gede Dwidasmara^{a5}, I Gede Arta Wibawa^{a6}

Informatics Department, Udayana University
Bali, Indonesia

¹Samcornelio16@gmail.com

²santi.astawa@unud.ac.id

³kg.suhartana@unud.ac.id

⁴santiyasa@unud.ac.id

⁵dwidasmara@unud.ac.id

⁶gede.arta@unud.ac.id

Abstract

Mathematics is a very important science because it is related to human life. Therefore, it is imperative to master mathematics and understand mathematical concepts correctly. In Indonesia, students' ability to understand mathematics is still relatively low. From this, understanding of mathematical concepts needs to be improved again so that students clearly understand mathematical concepts. In order to improve understanding of mathematical concepts, various efforts were made. One of the efforts carried out by the author is a semantic web-based learning media system using concept maps in basic mathematical ontologies. This system will use ontology to build a system that will represent a mathematical concept as a domain of knowledge explicitly about other concepts related to previous concepts by giving meaning, properties, and relations so as to form a knowledge base. In this research, the writer uses methodological method to build the basic mathematical ontology and the system development will be done by prototyping method. The system built has 2 features, namely exploration and semantic search, aiming to access the existing knowledge in the system systematically and according to user needs. To ensure that the system built can run as expected, system testing with black box testing has shown which features in the program run according to their functions and ontology evaluation using OntoQA on Schema Metrics carried out by measuring Relationship Richness (RR), Attribute Richness (AR), and Inheritance Richness (IR). In the evaluation of this ontology, it shows that the Relationship Richness (RR) value is 0.06; Attribute Richness (AR) value is 0.5; and the value of Inheritance Richness (IR) is 14.5.

Keywords: *semantic web, ontology, mathematics, metontology, black box, OntoQA, Schema Metric*

1 Introduction

Mathematics is a very important science because it is directly linked to human life. Even in this era of technology and digital, mathematics has a important role in human being life [1]. This role makes mathematics a tool for studying other sciences. Therefore, mastering mathematics is essential and a correct understanding of the concept of mathematics. This is because the concept of mathematics is a continuous series in which specific concepts are arranged based on the previous concept. A misunderstanding of one concept leads to a subsequent misunderstanding of the concept [2].

In Indonesia, the ability of students to understand mathematics is still relatively low. The poor math skills of Indonesian students are evident in the 2018 *Program for International Student Assessment (PISA)* and *Trends in International Mathematics and Science Study (TIMSS)* 2015 reports. In the math category, math skills scores reached 379 in 2018, making Indonesia ranked 79th countries participating in PISA [3]. Indonesia is ranked 44th out of 49 countries with a math

score of 397 in the 2015 TIMSS results [4]. The report suggests that Indonesian students misunderstand mathematical concepts. Based on these facts, in order for students to have a clear understanding of the concept of mathematics, they need to improve their understanding of the concept of mathematics again. To improve the understanding of mathematics the author proposes learning media systems using ontology based on concept maps. A concept map is a networked representation of a concept as a result of a structure that represents the relationship between the concepts and principles that govern mathematical structures and relationships in an easy-to-understand way [5].

Basic mathematics learning media system based semantic web is a system that help improve understanding of basic mathematics concepts through the use of concept maps. This system is mathematical as an area of knowledge that is explicitly related to other concepts related to previous concepts by using ontology to give meanings, properties, and relationships to form a knowledge base. Build a system that represents the concept. The ontology of this system can explicitly describe the domain of knowledge [6]. You can share your understanding of structured information and reuse your domain knowledge. This knowledge base is expected to help teachers determine the level of understanding and the direction of learning.

2 Research Method

The method used in this research is the methontology method and the prototyping method where the methontology method is used as a general ontology model development from scratch, This approach provides us with a set of guidelines on how to perform the activities identified during the development of the ontology and which technique is most appropriate in each case what activities and products each produce. By using this prototyping method, development and users can interact with each other during the system creation process for knowing more about the importance of the system and its functions. In Figure 1, there is a research flow that begins with data collection and then designs an ontology model which will be divided into the browsing process and the searching process. When the ontology model has been completed, a system performance evaluation will be carried out.

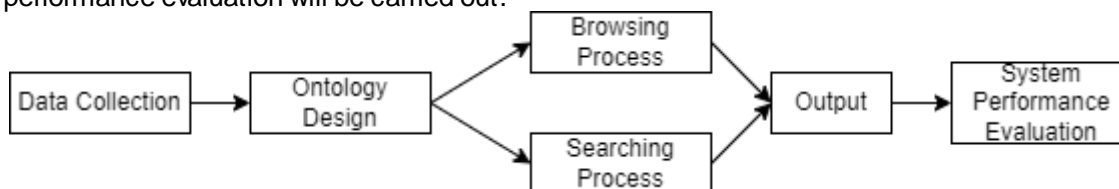


Figure 1. Flowchart System

2.1 Ontology Design Stage

In making ontology, the author uses a method of methontology that has the advantage of ontology being reusable for further system development. In addition, the advantage of the methodological description is that it explains each activity that must be carried out in detail. Some of the steps that must be carried out in the Methontology method are more clearly shown in Figure 2 which is adapted from the methodological flow [7].

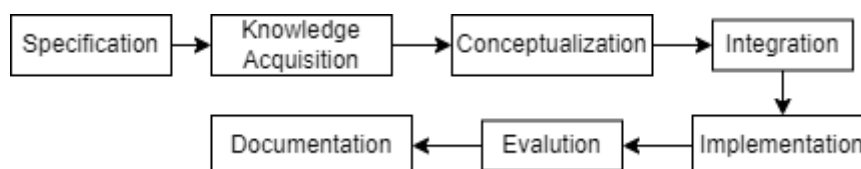


Figure 2. Methontology Method Activity Flow

In this method of methontology, the researcher wants to convert the concept map into an ontology model. While the definition of the ontology itself is an explicit specification of conceptualism. With the concept map which is a form of visualization of interrelated concepts, relationships between concepts, and meaningful relationships between these relevant concepts.

2.1.1 Specification

The goal of the ontology specification is to compile a document that includes the primary objectives of ontology, the objectives, and main scope of ontology. The document is written in natural language using a series of intermediate representations or using competency questions.

2.1.2 Knowledge Acquisition

Acquisition of knowledge is a separate activity in the development of ontology that coincides with other activities. Most of the knowledge is acquired at the beginning of the developing ontology and decreases with the progress of the process of developing ontology.

2.1.3 Conceptualization

This section will compile domain knowledge into a conceptual model, describing the problem and its solution related to the domain vocabulary defined in the ontology specification activity [8]. The first thing to do is to build a complete glossary. Thus, the glossary defines and brings together all useful and usable domain knowledge and what they mean.

2.1.4 Integration

In the integration stage, the definitions that have been made in an ontology are reconsidered to identify terms that may be present in other ontologies. In consideration of the definitions already in use, the author examines the terms that correspond to the concept of the domain. This happens so that the new definition or term is based on the same set of basic terms.

2.1.5 Implementation

This stage is the implementation process of the ontology design.

2.1.6 Evaluation

Evaluation includes conducting technical reviews of ontologies, software environments, and documentation against the terms of reference during each and between phases of their lifecycle. Review the Summary of Verification and Validation requirements.

2.1.7 Documentation

There are no uniform guidelines on how to create an ontological profile. In most cases, the only available literature is found in ontological codes, natural language texts with official definitions, and papers published in conferences and seminars, journals that address important ontological, ontological questions, and ontograph.

2.2 System Design Stage

The method used in the development of the system is prototyping. Prototyping is the process of creating a simple software model that allows the user to get a basic idea of the program and perform preliminary testing. Prototyping allows developers and users to interact with each other during production, so developers can easily model the software to be created.

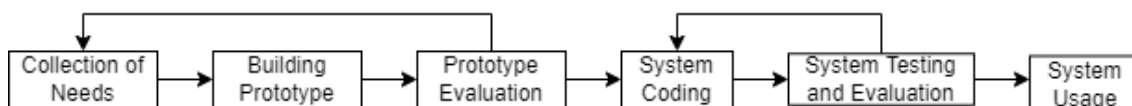


Figure 3. Stages of the Prototyping Method

2.2.1 Collection of Requirements

This stage aims to identify the format of the entire software, identify all requirements, and outline the system to be created.

2.2.2 Building Prototyping

Build prototyping by making temporary designs that focus on presentation, for example by making input and output formats. Forms of system design made include making system flowcharts, Use Case Diagrams, Activity Diagrams, and user interface designs. The following is a flowchart of the basic mathematics learning media system flowchart. In Figure 4, there is a flowchart system that will start with the home page then the user selects a menu, namely Browsing or Searching. If the user selects the browsing menu, it will display the Browsing Page, then the user selects content and sub contents. If the user selects the searching menu, it will display the Searching Page, then the user enters the output and input. when finished entering output and input, it will display the output according to the user's choice.

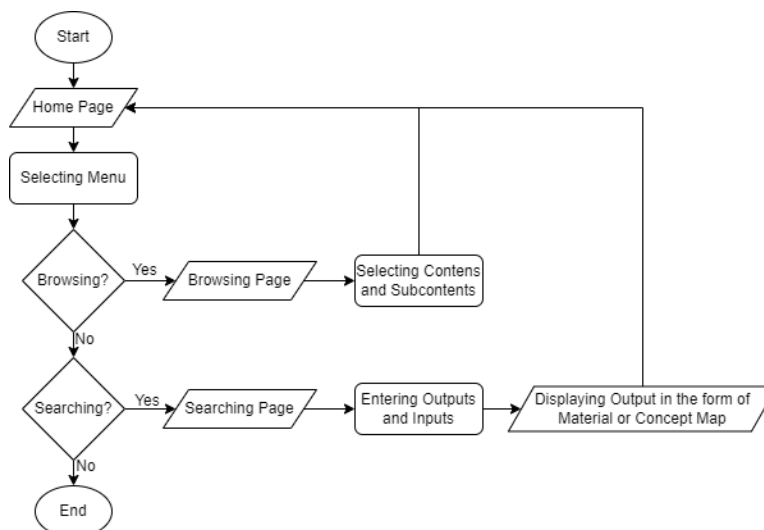


Figure 4. Flowchart System

2.2.3 Evaluation of Prototyping

At the evaluation stage, the prototype that has been made is evaluated to suit customer needs. Evaluation is also carried out on the developed system so as not to deviate from the concept of the initial system. If the prototype is correct, it will be continued in the fourth step. If the prototype does not match, it must be corrected by repeating steps 1, 2, and 3.

2.2.4 System Coding

In this stage the agreed prototyping is translated into the appropriate programming language.

2.2.5 System Testing and Evaluation

The system testing and evaluation stage is the stage for testing and evaluating the system. This stage aims to determine the extent to which the system that has been developed can run and solve predetermined problems [9].

At this stage a test will be held using black-box testing, this is done to determine the performance of the system made whether the functions in the system are appropriate or not, and to look for errors from the system display. At evaluation stage will discuss the evaluation of the ontology that has been built. Ontology evaluation will be carried out using *Schema Metrics* in *Ontology Quality Analysis (OntoQA)*. *Schema metrics* consists of measurement of *Relationship Richness (RR)*, *Inheritance Richness (IR)*, and *Attribute Richness (AR)*. These measurements can provide an evaluation of the characteristics and potential of the knowledge representation of the ontology design.

3 Result and Discussion

3.1 Implementation of Methontology in Ontology Design

3.1.1 Spesificaion

Specification phase is to compile a document that includes the primary objectives of ontology, the objectives, and main scope of ontology. The document is written in natural language using a series of intermediate representations or using competency questions. Below the authors include a table of the specifications of this research in Table 1.

Table 1. Spesification

Domain	Basic mathematic of SMP
Goal	To build an ontology model as the basis for learning media for basic mathematics of SMP and developing a semantic web-based system.
level of formality	Semi formal
Scope	Basic math knowledge of SMP
Source of knowledge	Book

3.1.2 Knowledge Acquisition

Knowledge acquisition is an independent activity within ontology development. However, it coincides with other activities. We acquired most knowledge at the start of ontology development. The level of knowledge acquisition fell as development progressed and we became more familiar with the application domain.

The ontology that is built will represent the knowledge of a set of concepts in a basic junior high school mathematics domain and the relationship between these concepts. The development of the ontology model is based on the concept map contained in the mathematics book so that at this stage the first glossary will be described with terms that are potentially relevant in the domain of basic mathematics of SMP.

3.1.3 Conceptualization

This section will compile domain knowledge in a conceptual model, that describes the problem and its solution concerning domain vocabulary identified in the ontology specification activity [8]. The first thing to do is build a complete glossary. Thus, the glossary identifies and collects all useful and potentially usable domain knowledge and their meanings. At this stage, the authors use the Protégé 5.5.0 software.

The process of forming an ontology class for junior high school mathematics knowledge is based on classes that represent every concept of junior high school mathematics knowledge.

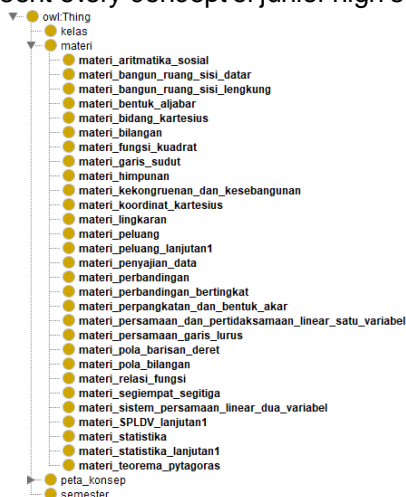


Figure 5. Class Ontology

After the class concept is designed, the next step is to input individual data from each class that has been created. In the next picture, several individuals or instances have been worked on.

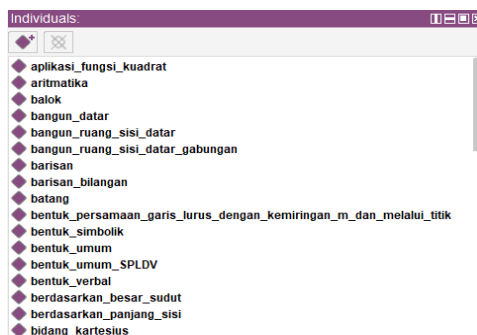


Figure 6. Individual Ontology

After the classes and individuals from the basic mathematics domain of SMP have been created, the next step is to define the class properties. A class if it stands alone will not provide sufficient information without the properties attached to it.

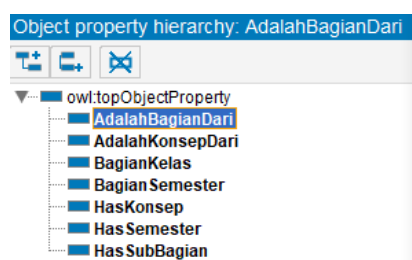


Figure 7. Object Property Ontology

3.1.4 Integration

At this stage, the existing definitions in the ontology will be reconsidered so that the use of such definitions corresponds to the concept of laptop domains. The purpose of this new definition check is to ensure that the new term set used is based on the same basis. Later, the author knows which ontology library gives the definition of semantic terms and how their implementation fits into the terms defined in the concept.

3.1.5 Implementation

In its implementation the author uses the software protege 5.5.0 Build Beta-9. In the implementation the author will execute the ontology that has been created using SPARQL Query. SPARQL is a query language to get information from Graph RDF, which provides facilities such as extracting information in the form of URI, graph RDF consists of triples formed from Subject, Predicate and Object, RDF can be defined in RDF Concepts and Abstract Syntax Concepts.

In this implementation process, a search experiment will be carried out by entering the characteristics or attributes of a concept. Then it will be known whether the concept exists or not. This is done in order to find out whether the ontology that has been built is connected and in accordance with the needs of the author's goals. Of the many individuals related to the concept, the author will try to find a concept which is a material concept from the seventh grade and first semester sections.

Figure 8. Query Process and Results

The results of the first trial query were 67 individuals. It turned out that in the first experiment it was still not in accordance with the needs of the author's goals because the query results contained individuals that were not intended, for example individual *materi_himpunan* which is an individual of material.

3.1.6 Evaluation

Evaluation means carrying out a technical assessment of the ontology, software environment and documentation in relation to the terms of reference. Evaluation summarizes the terms Verification and Validation. Verification refers to the technical process that ensures the correctness of the ontology, the related software environment. After the design is complete, an evaluation of the consistency and hierarchy is carried out using Reasoner and SPARQL queries. The reasoner used is HermiT 1.4.3.456 which by default is available on Protégé. The reasoner

will check the consistency of the ontology concept and class hierarchy in the basic mathematics domain of SMP as a whole, where the results of the check will appear in the reasoner log.

3.1.7 Documentation

There are no uniform guidelines on how to create an ontological profile. In most cases, the only available literature is found in ontological codes, natural language texts with official definitions, and papers published in conferences and seminars, journals that address important ontological, ontological questions, and ontograph. A simple schematic of the ontology of basic mathematics learning media systems for junior high schools designed with Protégé is shown in the following figure.

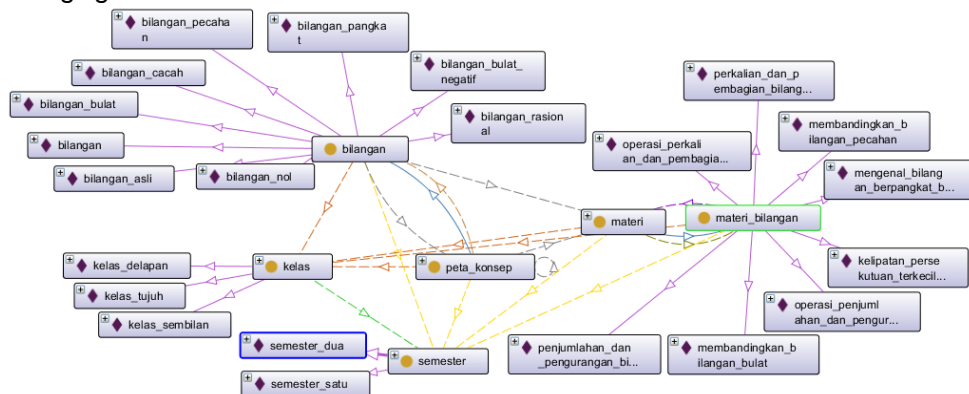


Figure 99. Simple Example of Basic Mathematical of SMP System Ontograph

3.2 Implementation of Prototyping Method

The system interface used in this study uses a PHP programming language. The author divides the services on the system into 2 main services, namely:

1. Browsing: allows system users to browse basic mathematics learning media that have been recorded in the system with available information.
2. Searching: allows system users to browse basic mathematics learning media that have been recorded in the system with available information.

There is 2 pages, namely : Browsing Page (Figure 10) and Searching Page (Figure 11).

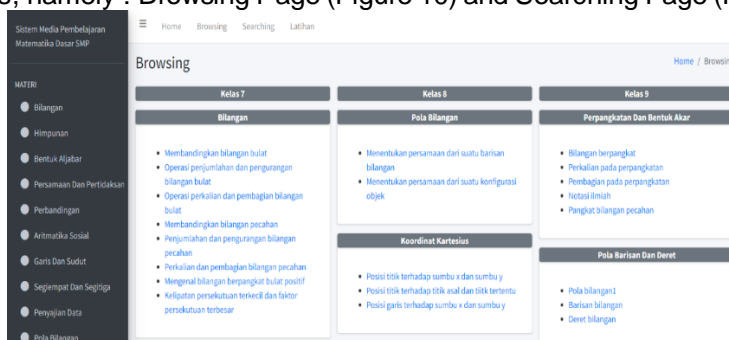


Figure 10. Interface Browsing Page

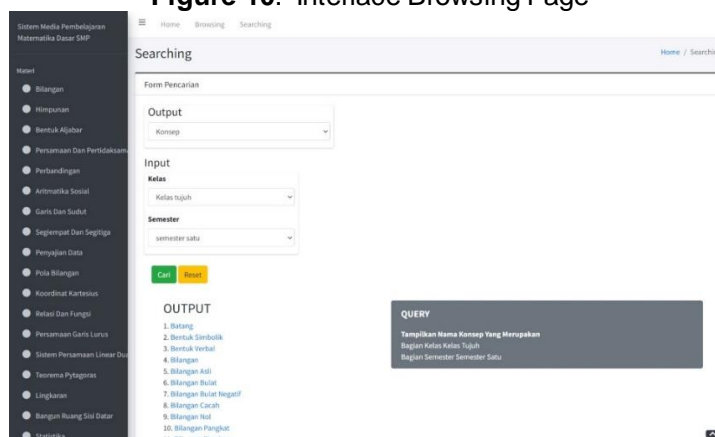


Figure 11. Interface Searching Page

3.3 System Testing And Evaluation

3.3.1 Testing With Black Box

The system is tested through a semantic search and search process by determining the outputs and inputs in the form of materials and concepts to be searched. The test will be carried out by black box testing, where the test aims to see the media system program for basic mathematics learning in junior high school is the same as the task of the system program without knowing the program code used. This test only checks the output values based on the respective input values. The black box testing process is done by trying the system that has been created in order to find out the semantic browsing and searching process according to what the author needs [10].

Table 2. Test result

No	Test Description	Expected results	Test result	Result
1	Go to the Browsing page, then select the "Memahami Korespondensi Satu-satu" sub-material on the "Relasi dan Fungsi" material tab or through the sidebar select "Relasi dan Fungsi" then select "Memahami Korespondensi Satu-satu"	The system will display the submaterial page "Memahami Korespondensi Satu-satu"	The system displays the submaterial page "Memahami Korespondensi Satu-satu"	Matching
2	Enter the output "materi" with the input "Kelas 8" and "Semester 1" click search and select the material "Memahami Korespondensi Satu-satu"	The system will display the output of any sub-materials that will appear and if you select the "Memahami Korespondensi Satu-satu" sub-material, the sub-material page will appear.	The system will display the output and display the appropriate material.	Matching

3.3.2 Ontology Evaluation

Ontology quality assessment is important for several reasons such as enabling ontology development to identify areas that need improvement, knowing which parts may have problems, and comparing several ontologies for consideration.

The metrics used are not a specific standard in ontology measurement. However, the metrics are intended to evaluate certain aspects of the basic mathematics ontology of SMP and its potential for further knowledge representation. Rather than describing an ontology as effective or not, metrics describe a particular aspect of an ontology because in most cases, the way an ontology is constructed largely depends on the domain in which it is designed [11].

In OntoQA there are two categories of metrics, namely schema metrics and instances metrics. In this paper, the metrics used are schema metrics because these metrics discuss the ontology design. Although it is not known whether the ontology design models knowledge correctly. Schema metrics represent the richness, width, depth, and inheritance of an ontology. There are 3 types of OntoQA schema metrics testing as follows.

3.3.2.1 Relationship Richness (RR)

Relationship Richness (RR) describes an ontology's diversity of relationships or relationships. An ontology that contains more class-subclass relations is richer than a taxonomy that only has class-subclass relations. RR is defined as the ratio of the number of classes (P) in

the ontology schema divided by the number of subclasses (SC) plus the number of classes as in equation (1).

$$RR = \frac{|P|}{|SC| + |P|} \dots\dots\dots (1)$$

The RR value that is closer to one indicates that the ontology contains more non-inheritance relations. This means that the ontology brings richer knowledge than the ontology which has an RR value close to 0. The calculation of the RR value for the basic mathematics ontology of SMP is as follows.

$$RR = \frac{|P|}{|SC| + |P|} = \frac{4}{58 + 4} = 0,06$$

In the basic mathematics ontology of SMP, non-inheritance relations or relations between classes are 4, while the number of inheritance or subclass relations is 58, resulting in an RR value of 0.06. This RR value that is close to zero means that it carries very little information or knowledge on the basic mathematics ontology of junior high school.

3.3.2.2 Attribute Richness (AR)

Attribute Richness can measure the amount of information available. AR indicates the quality of the ontology design and the amount of information associated with the data instances. In general, it is assumed that the more attributes defined, the more knowledge expressed in the ontology. AR in equation (2), is defined as the average number of attributes per class. AR is obtained by dividing the number of attributes for all classes (att) by the number of classes (C).

$$AR = \frac{|Att|}{|C|} \dots\dots\dots (2)$$

The higher the AR value, the more information that can be provided per class. The calculation of the AR value in the basic mathematics ontology of SMP is as follows.

$$AR = \frac{|Att|}{|C|} = \frac{2}{4} = 0,5$$

Based on the AR calculation value in the basic mathematics ontology of SMP, where there are 2 attributes from 4 classes, it is 0.5. This means that the developed ontology characteristics have 0.5 attributes per class. In other words, every 1 class has the opportunity to have attributes or not with a ratio value of 0.5.

3.3.2.3 Inheritance Richness (IR)

Inheritance Richness is a way to measure the distribution of information. In this measurement, it can be distinguished ontologies that have deep or shallow ontology characters. An ontology that has many levels of inheritance is a deep ontology. That is, the ontology has a representation of information that is more in-depth or specific to a domain. This can be seen with a small IR value. On the other hand, a large IR value indicates that the representation of the information held is less in-depth but covers a wider area of a domain or is shallow. The IR value shown in equation (3) is obtained from the average number of subclasses per class.

$$IR = \frac{|H(C)|}{|C|} \dots\dots\dots (3)$$

Ontology with a small IR value reflects a more detailed type of knowledge representation of a domain (specific domain), while a larger IR value indicates that the ontology represents general knowledge. The total subclass of 4 classes is 58. So, the calculation of the IR value in the basic mathematics ontology of SMP is as follows.

$$IR = \frac{|H(C)|}{|C|} = \frac{58}{4} = 14,5$$

From the results of the calculations above, it can be said that the basic mathematics ontology of SMP represents a more general knowledge or it can be said that the existing knowledge of the basic mathematics ontology of SMP is less deep and covers a wide area.

4 Conclusion

The semantic web-based learning media system with the development of elementary mathematics ontology for junior high schools presents two main features, namely browsing and searching features. The browsing process helps users search for information or knowledge related to basic junior high school mathematics material, and the searching process helps users find material or concepts according to the input that the user will choose in the system.

The design of a semantic web-based learning media system is carried out by developing a semantic ontology of basic mathematics for junior high school. One of the advantages of using a reasoner is that it can check whether all statements and definitions in the ontology are consistent

with each other and can help maintain the hierarchy correctly to minimize errors in the process of building a basic junior high mathematics ontology database.

The features in the system have been running well to the author's wishes and the system has succeeded in displaying the desired materials. System feature testing has been done with black box testing. The ontology of the elementary mathematics domain of SMP has also been well designed according to system requirements. The implementation of the ontology with SPARQL queries proves that the ontology has been successfully created according to system requirements. But unfortunately, it is true that the ontology is well designed, but not yet efficient. This is because the results of ontology evaluation with Schema Metrics OntoQA through three measurements (1) Relationship Richness (RR) has a value of 0.06. This RR value that is close to zero means that it carries very little information or knowledge on the elementary mathematics ontology of junior high school. (2) The Attribute Richness (AR) value, measuring the amount of information available, in the elementary mathematics ontology of SMP where 2 attributes from 4 classes are 0.5. This means that the developed ontology characteristics have 0.5 attributes per class. In other words, every 1 class has the opportunity to have attributes or not with a ratio value of 0.5. (3) Inheritance Richness (IR), is a way to measure the distribution of information. The IR value of the elementary mathematics ontology of junior high school is 14.5. This large IR value indicates that the representation of information owned by the system is not deep and feels broad or shallow.

References

- [1] N. R. Siregar, "Persepsi siswa pada pelajaran matematika: studi pendahuluan pada siswa yang menyenangkan game," *Pros. Temu Ilm. X Ikat. Psikol. Perkemb. Indones.*, pp. 224–232, 2017, [Online]. Available: <http://jurnal.unissula.ac.id/index.php/ippi/article/view/2193>.
- [2] M. Rahman, "PEMBELAJARAN DENGAN PETA KONSEP BIDANG STUDI MATEMATIKA DI KELOMPOK BELAJAR PAKET B," *Makalah*, vol. 2, 2014.
- [3] M. Tohir, "Hasil PISA Indonesia Tahun 2018 Turun Dibanding Tahun 2015," no. January, pp. 10–12, 2019, doi: 10.31219/osf.io/pcjvx.
- [4] S. Hadi and Novaliyosi, "TIMSS Indonesia (Trends in International Mathematics and Science Study)," *Pros. Semin. Nas. Call Pap. Progr. Stud. Magister Pendidik. Mat. Univ. Siliwangi*, pp. 562–569, 2019.
- [5] M. T. Rahman, "Peta Konsep (Concept Map)," pp. 1–12, 2016.
- [6] Y. F. Badron, F. Agus, and H. R. Hatta, "Studi Tentang Pemodelan Ontologi Web Semantik Artikel Jurnal Ilmiah," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, 2017.
- [7] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," 1997.
- [8] P. R. Ganeswara, C. Rai, and A. Pramatha, "Ontology-Based Approach for Klungkung Royal Family," vol. 8, no. 4, pp. 497–505, 2020.
- [9] O. B. Utami *et al.*, "Implementasi Methontology Untuk Pembangunan Model Ontology Program Studi Pada Perguruan Tinggi Di Bali," *Basic Educ.*, vol. 4, no. 1, pp. 2338–5081, 2019, doi: 10.1109/AXMEDIS.2007.19.
- [10] F. C. Ningrum, D. Suherman, S. Aryanti, H. A. Prasetya, and A. Saifudin, "Pengujian Black Box pada Aplikasi Sistem Seleksi Sales Terbaik Menggunakan Teknik Equivalence Partitions," *J. Inform. Univ. Pamulang*, vol. 4, no. 4, pp. 125–130, 2019.
- [11] S. Tartir and I. B. Arpinar, "Ontology evaluation and ranking using OntoQA," in *International conference on semantic computing (ICSC 2007)*, 2007, pp. 185–192.

Application Of Learning Vector Quantization Algorithm In Educational Games Learning To Write Aksara Bali

Agus Wisnawa^{a1}, I Gede Arta Wibawa^{a2}, Ida Bagus Gede Dwidasmara^{a3}, Made Agung Raharja^{a4}, I Komang Ari Mogi^{a5}, Luh Arida Ayu Rahning Putri^{a6}

^aInformatics Department, Faculty of Mathematics and Natural Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia

¹haguswisnawa03@gmail.com

²gede.arta@unud.ac.id

³dwidasmara@unud.ac.id

⁴made.agung@unud.ac.id

⁵arimogi@unud.ac.id

⁶rahningputri@unud.ac.id

Abstract

Almost every region in the nusantara has its own regional script, one of which is in Bali the people know the Balinese script as a script commonly used by their ancestors in writing. In order to keep this Balinese script sustainable and not extinct, learning about this script is applied in schools in Balinese language subjects. In addition to applying learning about script in schools, it would be better if learning about Balinese script could be done in more interesting ways and concepts. One of them is by making an educational game for the introduction of Balinese characters. Which in this application the user will provide input in the form of Balinese script characters written on the application, which then the input will go through a preprocessing process then proceed with diagonal feature extraction, then the results of the feature extraction will go through a classification process using the Learning Vector Quantization method. . The result is a web-based application that can recognize Balinese script writing using the Learning Vector Quantization classification method with an accuracy rate of 58.6% and get a good response from each respondent who has tested the application.

Keywords: Balinese Script, Game Application, Diagonal Feature Extraction, LVQ, Classification.

1. Introduction

Balinese script is a regional writing originating from Bali. Balinese script has been used since ancient times until now. Balinese script can be used in writing Balinese and Sanskrit in various fields such as in religious ceremonies, arts, writing Balinese literature and so on. Therefore, the Balinese script must be preserved so that it is not lost and forgotten by the younger generation. One way to preserve Balinese culture, especially in learning Balinese script, is to teach the younger generation from an early age to learn Balinese script [1].

Based on a digital report from Hootsuite We are Social Indonesian in January 2021, the percentage of Indonesian internet users aged 16 to 64 who play games on a desktop / computer reaches 53% [2]. This indicates that the game on the computer is quite popular with the people of Indonesia. Games can also be used as learning media. Based on this, the researcher aims to create an educational game with the topic of Balinese script.

In the manufacturing process, a classification stage is needed to recognize the written characters. The Learning Vector Quantization (LVQ) method was chosen as the classification method because it has high accuracy, it is based on research from Solichin and Rahman who used the LVQ method as vehicle number plate identification with an accuracy of 95.32% [3].

2. Research Method

2.1 System Analysis

Classification is a process that aims to determine an object into a pre-determined class or category [4]. Before doing the classification there is a process that must be done, the process consists of preprocessing, feature extraction, and classification. The feature extraction method used in this study is diagonal feature extraction. Classification of data in this study using the method of Learning Vector Quantization (LVQ). In this study, the process of introducing Balinese script starts from the input image obtained from application users. This input image will then undergo preprocessing. After getting the preprocessing results, then proceed with the feature extraction process to get the class weight of the image. After getting the weights for each class, the classification process is carried out using the LVQ method. This classification process is divided into two, namely training or also called training and testing or can be called testing. The classification process aims to determine whether the input from the user is in accordance with the given command. After the classification process is complete, this process will end. The system analysis flowchart can be seen in Figure 1.

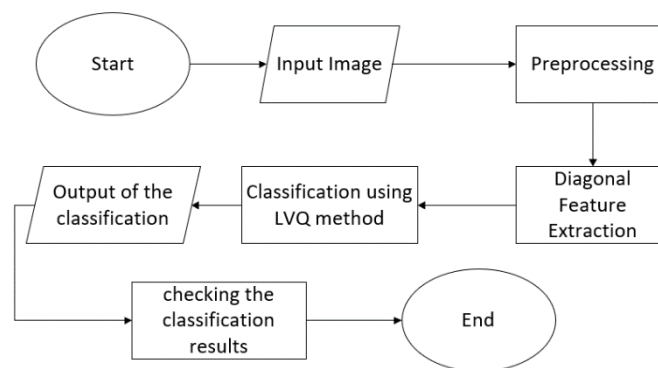


Figure 1 System Analysis Flowchart

2.2 Data

The wreastra script consists of 18 characters, namely ha, na, ca, ra, ka, da, ta, sa, wa, la, ma, ga, ba, nga, pa, ja, ya, nya [5]. The 18 characters are consonants, while the vowels are taken from the wisarga script, which is added by a pengangge aksara suara. The 6 pengangge aksara suara are ulu, suku, taling, tedong, taling tedong, and pepet [6]. The data used in this study is divided into two types, namely training data and test data. Training data is the data used to train the architecture of the LVQ method. The training data used in this study were primary. Students are asked to write each character on the provided paper 9 times, so that the total training data used is 90 pieces of writing on each Balinese script character. The details of the characters used are the characters in the Wreastra script, totaling 18 characters and the voice character cast, totaling 5 characters so that the total characters are 23 characters so that the total of all training data obtained is 2070 training data.

2.3 Preprocessing

In the preprocessing the initial data processing is divided into 2 processes, namely the process of processing test data and training data where both these processes have the same stages but use different data. This process consists of binaryzation, normalization, and thinning.

2.4 Feature Extraction

Diagonal feature extraction is one method that can be used to find the characteristics of handwritten characters. The first thing to do is to divide the input image into several zones that have the same size. Diagonal feature extraction is one method that can be used to find the characteristics of handwritten characters. The first thing to do is to divide the input image into several zones that have the same size. The results of the preprocessing are broken down into a size of 10x10 pixels so that

the image has 54 areas or features. In each of these features a diagonal line will be drawn according to the size of the feature, where with the size of each feature of 10x10 pixels it will produce 19 diagonal lines. the value of all foreground pixels in the feature. These values will be averaged to form a single value for the features they occupy. This process will be repeated until all features are scored.

2.5 Learning Vector Quantization (LVQ)

The Learning Vector Quantization (LVQ) algorithm has the result in the form of an appropriate weight value for grouping vectors into destination classes that have been initialized during the formation of the LVQ network. The testing algorithm of LVQ aims to calculate the output value (vector class) closest to the input vector, or can be equated with the classification process [7]. The LVQ training algorithm is as follows:

1. Initialization of the value of learning rate (α), learning rate deduction ($dec \alpha$), training data, maximum iteration and minimum α used.
2. For each input do steps a to d:
 - a. Calculate the euclidean distance of the input data with the weight vector for each class with Equation :

$$D(x, w) = \|X_i - W_j\| \dots \dots \dots (1)$$
 - b. Determine the minimum distance from the distance calculation results so that the output C_j is obtained.

$$C_j = \text{Min } D(x, w) \dots \dots \dots (2)$$
 - c. Fix the weights with the condition:
 If $T = C$

$$W_j(\text{new}) = W_j(\text{old}) + \alpha[X_i - W_j(\text{old})] \dots \dots \dots (3)$$

 If $T \neq C$

$$W_j(\text{new}) = W_j(\text{old}) - \alpha[X_i - W_j(\text{old})] \dots \dots \dots (4)$$
 - d. Make improvements to the value of learning rate (α)

$$\alpha(\text{new}) = \alpha(\text{old}) \times dec \alpha \dots \dots \dots (5)$$
3. Stops when the iteration is maximum or the learning rate is minimum.

Where:

- W : Vector Weight
- D : Distance data to i
- X : Data
- T : Target
- (α) : Learning rate

There are 3 layers of the LVQ architecture created, namely the input layer, the competitive layer, and the output layer. The input layer in this study is an input image consisting of 69 features. In the competitive and output layers, the architecture of this research consists of 23 classes used. The Learning Vector Quantization architecture in this study can be seen in Figure 2.

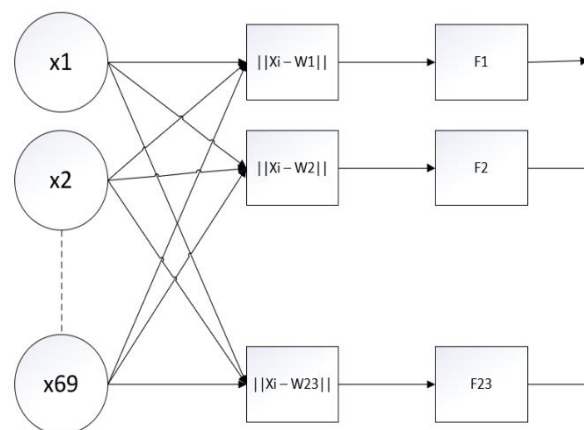


Figure 2. LVQ architecture

2.6 Testing

Testing process will be divided into 3 types, black box testing, accuracy test, and UAT (User Acceptance Test).

- a. Black Box Testing
Black box testing is a test that aims to show the function of the software about how it operates, whether the input and output have run as expected [8].
- b. Accuracy Testing
The level of character recognition accuracy is obtained by dividing the number of correct recognitions by all test results, the final result obtained will be in the form of a percentage [9]. As for this study, accuracy testing is carried out per character that appears and not per sentence. The results of this test are to determine the ability of the system in reading Balinese script for the Balinese script introduction game.
- c. User Acceptance Test (UAT)
User Acceptance Test or UAT is a test intended to determine user ratings of the applications made. The process of UAT testing is that prospective users are asked to use the application created and then the user is asked to fill out a questionnaire containing a perspective assessment of the application made [10].

3. Result and Discussion

3.1. Software Implementation

Interface Implementation Implementation of the interface on this system will be divided into 10 pages, namely the main menu page which can be seen in Figure 3, this page is the main page of the application. Next is the how to play page which can be seen in Figure 4, this page describes how to play. The next page about the application can be seen in Figure 5, this page contains information about the game being built such as the algorithm used, the technology to build the application, and also the creator of the application. The start page of the game which can be seen in Figure 6, this page becomes the second menu as an introduction for the user to the learning page or the practice page. On the learning page which can be seen in Figure 7, in this page users can see Balinese script and the latin writing. On the exercise page which can be seen in Figure 8, the user can practice the user's ability to write Balinese script which will be checked whether it is right or wrong. Furthermore, the correct answer page can be seen in Figure 9 or the wrong answer page which can be seen in Figure 10 is the page that is displayed after the user submits an answer on the exercise page.



Figure 3. Main Menu



Figure 4. How to Play



Figure 5. About Application



Figure 6. Game Start

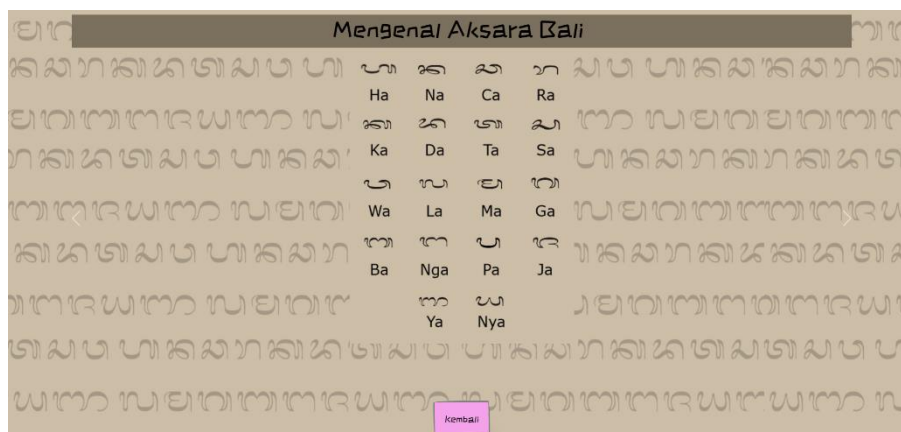


Figure 7. Study

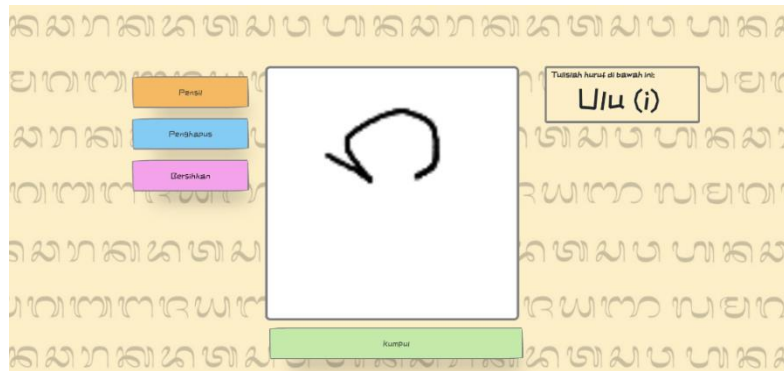


Figure 8. Practice



Figure 9. Correct Answer

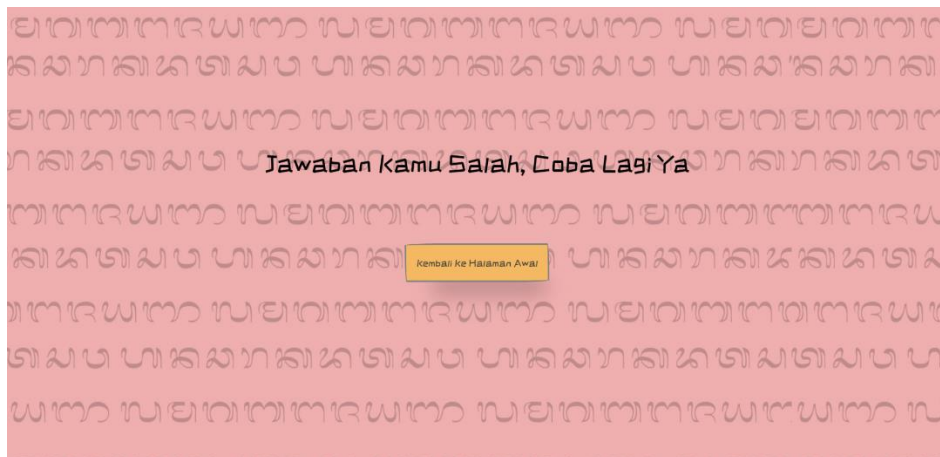


Figure 10. Wrong Answer

3.2. Black Box Testing

This test will test the functionality of the application whether it is in accordance with the expected expectations. The results of the black box testing are in the table 1.

Table 1 Black Box Testing

No	Testing	Result	Description
1.	Start the app	Success	The application can be opened and display the main page.
2.	Press the "Start Game" button	Success	The button can direct the application to the game mode select page.
3.	Pressing the "How to Play" button	Success	The button can direct the application to the how-to page.
4.	Pressing the "About" button	Success	The button can direct the app to the about page.
5.	Pressing the "Learn" button	Success	The button can direct the application to the study page.
6.	Pressing the "Practice" button	Success	The button can direct the app to the workout page.
7.	Writing on canvas	Success	Users can write on the provided canvas.
8.	Pressing the "Eraser" button	Success	The user enters erase mode and can delete the existing doodles in the canvas manually.
9.	Pressing the "Pencil" button	Success	Users can return to writing mode as usual after previously using erase mode.
10.	Pressing the "Clean" button	Success	Users can clean the entire canvas content automatically.
11.	Press the "Enter" button	Success	This button will stop writing activities and then send data to enter the classification process.
12.	Pressing the "Done" button	Success	This button will stop the learning process on the study menu and return to the previous page.
13.	Question	Success	Questions can appear in the practice menu.
14.	Press the "Back to Start Page" button	Success	This button will direct the application to the start page.

3.3. Accuracy Testing

From the tests carried out, where each character was tested 10 times so that the total tests carried out were 230 tests with a total of 135 trials which were true and 95 were false. It was found that the accuracy of the classification of Balinese characters using the Learning Vector Quantization method with diagonal feature extraction is 58.6%.

3.4. UAT Testing

curacy, next requires the calculation of metrics which is used to better present the performance of the model that has been created. Among them is Precision, which calculates how much data is really positive from all data that is predicted as positive, or in other words true positive, Recall calculates how much positive data is predicted as positive, then the last is the F-1 Score, which is a combination of Precision and Recall values to compare different models or commonly referred to as harmonics means of the two. The UAT test result can be seen in table 2.

Table 2 UAT Test Result

No.	Question	Answer					Score	Percentage
		A	B	C	D	E		
1.	Does the application look attractive?	0	7	3	0	0	37	74%
2.	Is the menu of the application easy to understand?	8	1	1	0	0	47	94%
3.	Is this app easy to use?	0	5	3	0	2	31	62%
4.	Can an example of Balinese script help you to know the shape of the Balinese script?	8	1	1	0	0	47	94%
5.	Can practice help you to know your ability to write Balinese script?	0	3	7	0	0	33	66
6.	Can this application help you to improve your writing skills in Balinese script?	0	1	8	1	0	30	60%
7.	Can this application make learning to write Balinese script fun?	7	2	1	0	0	46	92%
8.	Can this application be a good medium for learning Balinese script?	0	0	8	1	1	27	54%

Based on the UAT testing conducted, it can be concluded that the respondents agree that the application display is attractive. Furthermore, respondents strongly agree that the application menu is easy to understand. Respondents agree that the application is easy to use. Furthermore, respondents strongly agree that the existence of script samples can help to determine the shape of the Balinese script. Respondents agreed that the exercise can help respondents know the respondent's ability to write Balinese script. Furthermore, respondents agreed that this application can help respondents in improving their writing skills in Balinese script. Respondents strongly agree that this application can make the experience of learning to write Balinese script enjoyable. Finally, the respondents considered that this application could be a good medium for learning Balinese script. The total points

from all the questions obtained were 298 points or 74.5%. So it can be said that the application made has received a positive response from application users.

4. Conclusion

From the research that has been done, the results obtained are in the form of an educational game application that introduces Balinese script which is aimed at people, both students and the general public. The results of black box testing on this application show that all features and functions can run according to the desired expectations. Based on UAT testing, the application can be said to have received a positive response from the user.

This study also conducted an accuracy test. The purpose of this accuracy test is to measure the percentage of the ability of the LVQ (Learning Vector Quantization) algorithm with diagonal feature extraction in classifying Balinese script characters written by users. The results of this accuracy test are the percentage of success of the LVQ algorithm with diagonal feature extraction in classifying Balinese script writing, which is 58.6% in a total of 230 test trials.

References

- [1] P. P. G. Pertama, Suyoto And T. Suselo, "Pengembangan Aplikasi Mobile Pengenalan Aksara Bali Kendala Huruf Latin Dengan Augmented Reality," Seminar Nasional Teknologi Informasi Dan Komunikasi 2015(Sentika 2015), P. 1, 2015.
- [2] S. Kemp, "We Are Social Indonesian Digital Report," Hotsuite, 2021.
- [3] A. Solichin And Z. Rahman, "Identifikasi Plat Nomor Kendaraan Berbasis Mobile Dengan Metode Learning Vector Quantization," Jurnal Ticom Vol.3, 2015.
- [4] A. M. Puspitasari, D. E. Ratnawati And A. W. Widodo, "Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine," Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, Pp. 802-810, 2018.
- [5] I. Suasta, I. Mayun And W. Rupa, Modernisasi Dan Pelestarian Perkembangan Metode Dan Teknik Penulisan Aksara Bali, Jakarta: Proyek Pengkajian Dan Pembinaan Nilai-Nilai Budaya Direktorat Sejarah Dan Nilai Tradisional Direktorat Jenderal Kebudayaan Departemen Pendidikan Dan Kebudayaan , 1996.
- [6] M. I. K. Narwadha, Sareng Mapuruk Aksara Bali, Surabaya: Parāmita, 2006.
- [7] R. Meliawati, O. Soesanto And D. Kartini, "Penerapan Metode Learning Vector Qiantization (Lvq) Pada Prediksi Jurusan Di Sma Pgri 1 Banjarbaru," Kumpulan Jurnal Ilmu Komputer (Klik), 2016.
- [8] M. Adiwijaya, I. S And Y. Christyono, "Perancangan Game Edukasi Platform Belajar Matematika Berbasis Android Menggunakan Construct 2," Jurnal Teknik Elektro, Universitas Diponegoro Semarang, 2015.
- [9] R. Kohavi And F. Provost, Special Issue On Applications Of Machine Learning And The Knowledge Discovery Process, Kluwer Academic Publishers, 1998.
- [10] H. Supriyono, R. Rahmadzani, M. Adhantoro And A. Susilo, "Rancang Bangun Media Pembelajaran Dan Game Edukatif Pengenalan Aksara Jawa "Pandawa"," The 4th University Research Colloquium 2016, 2016.

This page is intentionally left blank



ISSN



E-ISSN
