

Klasifikasi Lirik Lagu Bertema Lingkungan dengan Metode Naive Bayes

Putu Ode Irfan Ardika¹, I Gusti Ngurah Anom Cahyadi Putra²

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹dirfanardika@gmail.com
²anom.cp@unud.ac.id

Abstract

Awareness of the importance of protecting the environment is becoming increasingly important in this modern era. Humans as inhabitants of the earth have a responsibility to protect and maintain the natural environment, they live in. Songs can be one of the important roles that can help awaken people to start protecting the environment they live in. This research makes it easy to find songs that have the theme of protecting the environment by classifying song lyrics. This research will create a system that can classify environment-themed song lyrics using the Naive Bayes method with a Multinomial model. The results of the Naive Bayes test with the Multinomial model get the best results on the composition of the training data and test data of 10:90 which produces a recall score of 38%, precision of 90.4%, F1 score of 53.5%, while for accuracy it gets the best score on the composition of 90:10 with a yield of 75%.

Keywords: Text Processing, TF-IDF, Naive Bayes, Lyrics

1. Pendahuluan

Kesadaran akan pentingnya menjaga lingkungan menjadi semakin penting di era modern ini. Manusia sebagai penghuni bumi memiliki tanggung jawab untuk melindungi dan memelihara lingkungan alam yang ditinggali. Lagu dapat menjadi salah satu peran penting yang bisa membantu menyadarkan untuk mulai menjaga lingkungan yang ditinggali. Penelitian ini dapat memudahkan untuk mencari lagu yang memiliki tema menjaga lingkungan dengan melakukan klasifikasi pada lirik lagu. Penelitian mengenai klasifikasi yang sudah dilakukan oleh peneliti sebelumnya dengan seperti Uji Akurasi Klasifikasi Emosi Pada Lirik Lagu Bahasa Indonesia dengan menggunakan metode SVM dengan hasil akurasi 92,13% [1]. Kemudian penelitian mengenai Klasifikasi Cerita Pendek Berbahasa Bali Berdasarkan Umur Pembaca dengan Metode Naive Bayes yang menghasilkan akurasi 72% dengan feature selection [2]. Lalu penelitian mengenai Klasifikasi Emosi Lagu Berdasarkan Lirik pada Teks Berbahasa Indonesia Menggunakan K-Nearest Neighbor dengan Pembobotan WIDF yang menghasilkan akurasi 66% [3]. Selanjutnya ada penelitian Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Support Vector Machine Dan Optimasi Particle Swarm Optimization menghasilkan akurasi paling besar 90% [4]. Terakhir ada penelitian Klasifikasi Emosi Lirik Lagu Menggunakan Improved K-Nearest Neighbor dengan Seleksi Fitur dan BM25 yang mendapatkan hasil rata-rata paling baik saat $k = 55$ dengan hasil f-measure 0,6693, recall 0,6582 dan precision 0,7427 [5]. Pada penelitian kali ini akan membangun sebuah sistem yang dapat mengklasifikasikan teks lirik lagu dengan metode Naive Bayes. Kategori yang digunakan adalah lingkungan dan non-lingkungan. Penelitian ini digunakan untuk membantu mencari lagu yang bertemakan menjaga lingkungan untuk membantu memberikan kesadaran terhadap pentingnya menjaga lingkungan.

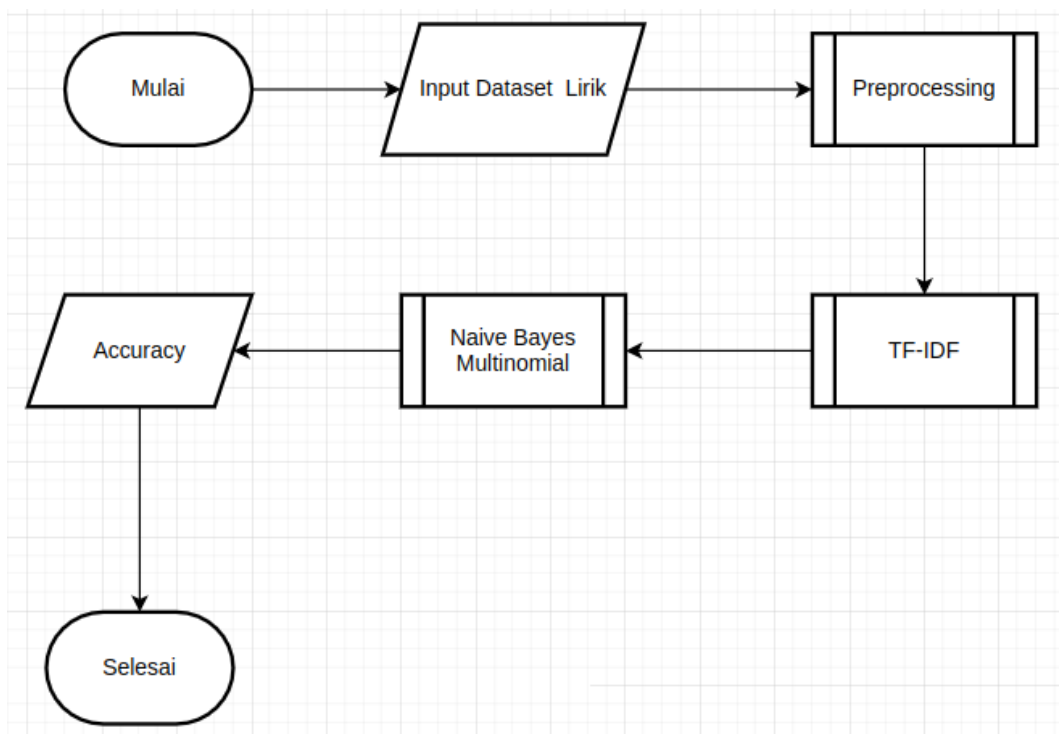
2. Metode Penelitian

2.1. Dataset

Data yang digunakan adalah 58 dokumen lirik lagu berbahasa Indonesia. Ada 2 kategori yang digunakan yaitu kategori lingkungan yang dilabel 1 dan kategori non-lingkungan dengan label 0. Data yang digunakan meliputi judul, isi, dan label.

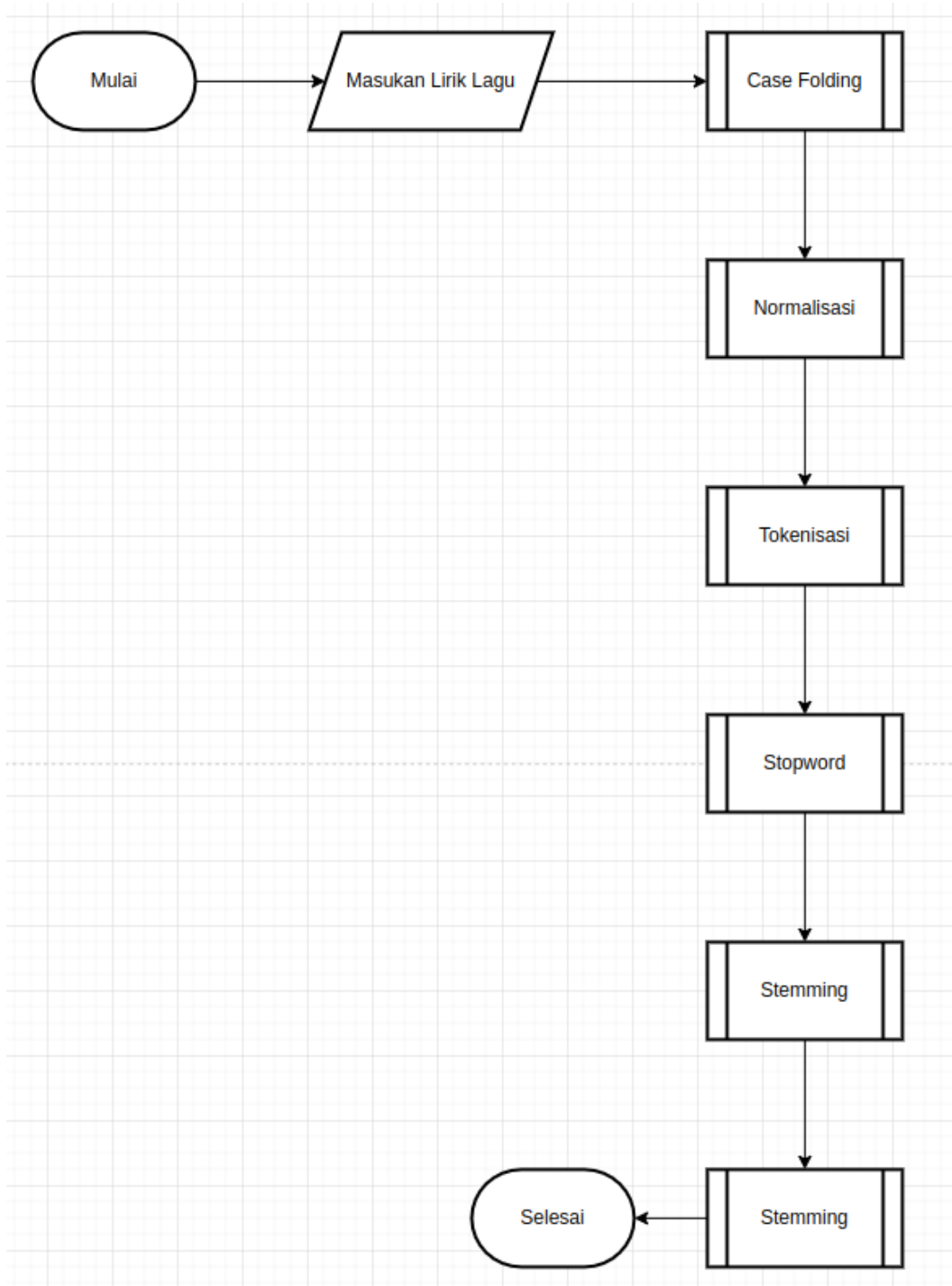
2.2. Alur Sistem

Tahap pertama yang dilakukan adalah pengumpulan data berupa lirik lagu berbahasa Indonesia dengan 2 kategori yaitu lingkungan dan non-lingkungan. Setelah terkumpul semua, data kemudian disimpan ke dalam Excel. Data kemudian melewati tahap preprocessing kemudian mengalami tahap pembobotan menggunakan TF-IDF. Kemudian hasil dari TF-IDF diproses menggunakan Naive Bayes Multinomial.



Gambar 1. Alur Penelitian

2.3 Preprocessing Data



Gambar 2. Alur Preprocessing

Dalam tahap ini dataset yang berupa lirik lagu diproses terlebih dahulu dimulai dengan tahap case folding untuk merubah menjadi lowercase kemudian menghilangkan kata dan karakter yang tidak diperlukan, seperti tanda baca, emoji, angka. Kemudian ditahap normalisasi menghilangkan dan memperbaiki kata dengan menerjemahkannya ke bahasa indonesia atau memperbaiki kata singkatan. Selanjutnya tahap tokenisasi untuk memisahkan lirik lagu menjadi token atau kata.

Tahap berikutnya stopwords untuk menghilangkan kata yang dianggap tidak penting. Terakhir tahap stemming untuk menjadikan kata ke bentuk dasar. Hasil dari preprocessing berupa matriks 2D.

2.4 TF-IDF

TF-IDF adalah salah satu metode pembobotan suatu kata dengan dokumen[1].

Keterangan:

$$\text{TF-IDF} = \text{TF} * \text{IDF}$$

TF = Term Frequency

TF = Berapa kali kata muncul dalam sebuah dokumen

IDF = Inverse Document Frequency

IDF = $\log(\text{total dokumen} / \text{jumlah dokumen yang mengandung kata})$

2.5 Naive Bayes

Naive Bayes adalah salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi, salah satunya klasifikasi teks. Model Naive Bayes yang digunakan untuk penelitian ini adalah Multinomial Naive Bayes. Data yang sudah diproses sebelumnya disiapkan untuk diproses menggunakan MNB, dengan X sebagai data hasil TF-IDF dan y sebagai label nya. Kemudian datanya dibagi menjadi data uji dan data latih dengan komposisi 10:90 sampai 90:10. Selanjutnya MNB melakukan prediksi dan akan menghasilkan label dari data uji, kemudian label dicocok untuk mendapatkan nilai akurasi, recall score, precision dan F1-score.

2.6 Evaluasi

Pada penelitian ini mendapatkan hasil akurasi, recall score, precision, dan F1-score yang dapat dilihat dari tabel dibawah.

Tabel 1. Hasil Pengujian

Komposisi	Akurasi	Recall Score	Precision	F1-score
10:90	68,5%	38%	90,4%	53,5%
20:80	56,9%	13,3%	85,7%	23,1%
30:70	59,7%	18,4%	77,7%	29,7%
40:60	70%	34,4%	83,3%	48,7%
50:50	67,4%	32%	80%	45,7%
60:40	68,1%	30%	85,7%	44,4%
70:30	65,7%	26,6%	80%	40%
80:20	66,6%	30%	75%	42,8%
90:10	75%	33,3%	50%	40%

3. Hasil dan Pembahasan

Pada pengujian metode Naive Bayes menggunakan model Multinomial mendapatkan hasil mendapatkan hasil akurasi yang paling tinggi saat menggunakan data latih dan data uji 90:10

yang mendapatkan hasil 75% sedangkan yang terendah adalah pada komposisi 20:80 yang hanya mendapatkan 56,9% dengan rata-rata hasilnya pengujian akurasinya adalah 66,4%. Kemudian untuk hasil recall score mendapatkan hasil yang tertinggi pada komposisi 10:90 yang mendapatkan hasil 38% sedangkan hasil yang terendah pada komposisi 20:80 yang hanya mendapatkan hasil 13,3% dengan rata-rata hasilnya adalah 28,4%. Selanjutnya untuk hasil precision hasil tertinggi didapatkan dengan komposisi 10:90 dengan hasil 90,4% sedangkan hasil terendah pada komposisi 90:10 yang hanya mendapatkan 50% dengan rata-rata hasilnya adalah 78,6%. Terakhir untuk F1-score mendapatkan hasil tertinggi pada komposisi 10:90 dengan 53,5% sedangkan hasil terendah terdapat pada komposisi 20:80 dengan hasil 23,1% dan rata-rata 40,8%.

4. Kesimpulan

Pada penelitian ini dilihat dari hasil pengujian. Hasil dari pengujian Naïve Bayes dengan model Multinomial mendapatkan hasil terbaik pada komposisi data latih dan data uji 10:90 yang menghasilkan recall score 38%, precision 90,4%, F1-score 53,5%, sedangkan untuk akurasi mendapatkan nilai terbaik pada komposisi 90:10 dengan hasil 75%. Sehingga bisa disimpulkan mengurangi data latihnya dapat meningkatkan hasil recall score, precision, dan F1-score, sedangkan untuk mendapatkan hasil akurasi yang tinggi bisa dengan mengurangi data uji nya.

Daftar Pustaka

- [1] Mega Noveanto, Helen Sastypratiwi, dan Hafiz Muhandi, "Uji Akurasi Klasifikasi Emosi Pada Lirik Lagu Bahasa Indonesia" *Jurnal Sistem dan Teknologi Informasi*, vol. 10, no. 3, Juli, 2022.
- [2] Luh Ristiari, AAIN Eka K., I P.G. Hendra S., Agus M., I D.M. Bayu A. D., dan I M. Widiartha "Klasifikasi Cerita Pendek Berbahasa Bali Berdasarkan Umur Pembaca dengan Metode Naive Bayes" *Jurnal Elektronik Ilmu Komputer Udayana.*, vol. 10, no. 4, May 2022.
- [3] Djajeng Ninda Armianti, Indriati dan Sigit Adinugroho "Klasifikasi Emosi Lagu Berdasarkan Lirik pada Teks Berbahasa Indonesia Menggunakan K-Nearest Neighbor dengan Pembobotan WIDF" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.*, vol. 3, no. 10, Oktober 2019.
- [4] Wahyudi Hermanto, Budhi Irawan dan Casi Setianingsih "Klasifikasi Emosi Pada Lirik Lagu Menggunakan Algoritma Support Vector Machine Dan Optimasi Particle Swarm Optimization" *e-Proceeding of Engineering*, vol. 8, no. 5, Oktober 2021.
- [5] Febrina Sarito Sinaga, Indriati dan Bayu Rahayudi "Klasifikasi Emosi Lirik Lagu Menggunakan Improved K-Nearest Neighbor dengan Seleksi Fitur dan BM25" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 6, Juni 2019.

Halaman ini sengaja dibiarkan kosong