

Literature Review Klasifikasi Data Menggunakan Metode Cosine Similarity dan Artificial Neural Network

Lely Meilina¹, I Nyoman Satya Kumara², Nyoman Setiawan³

[Submission: 21-06-2021, Accepted: 24-08-2021]

Abstract— One of the positive impacts arising from technological developments is the ease in conveying aspirations and in obtaining information very quickly. The benefits of this technological development can be felt by all sectors, including the government sector which must protect the people and the state. In improving the quality of public services, the government must implement a government based on digital information technology. Therefore, the central and regional governments have provided online-based public complaint services. To improve the quality of service, the online complaint system must run optimally. The methods that are widely used to find the similarity of the complaint text are the cosine similarity method and the Artificial Neural Network (ANN) method for classifying complaint data. This study reviews the application of the two methods to determine the level of accuracy before it can be implemented in the online complaint system. The results of the review state that the Cosine Similarity method has an accuracy rate of 71.5% and ANN has an accuracy rate of 77%. While other methods have an accuracy rate of 67%. From the percentage of these values, it can be concluded that the use of Cosine Similarity and ANN methods is feasible to use in classifying data in the Online Community Complaint System.

Keywords— Data Classification, Cosine Similarity, Artificial Neural Network

Intisari— Dampak positif yang ditimbulkan dari perkembangan teknologi salah satunya adalah kemudahan dalam menyampaikan aspirasi dan dalam mendapatkan informasi dengan sangat cepat. Manfaat dari perkembangan teknologi ini dapat dirasakan oleh semua sektor, termasuk sektor pemerintahan yang harus mengayomi masyarakat dan negara. Dalam meningkatkan kualitas pelayanan publik, pemerintah harus menerapkan pemerintahan yang berbasis teknologi informasi digital. Oleh karena itu, Pemerintah pusat maupun daerah telah menyediakan layanan pengaduan masyarakat yang berbasis online. Untuk meningkatkan kualitas pelayanan maka sistem pengaduan online harus berjalan dengan optimal. Metode yang banyak digunakan untuk mencari kemiripan teks pengaduan adalah metode cosine similarity dan metode Artificial Neural Network (ANN) untuk klasifikasi data pengaduan. Penelitian ini mereview penerapan kedua metode tersebut untuk mengetahui tingkat akurasinya sebelum dapat diimplementasikan pada sistem pengaduan online. Hasil dari review menyatakan bahwa metode Cosine Similarity memiliki tingkat akurasi sebesar 71,5% dan ANN memiliki tingkat akurasi-

sebesar 77%. Sedangkan metode lainnya memiliki tingkat akurasi sebesar 67%. Dari presentase nilai tersebut dapat disimpulkan bahwa penggunaan metode Cosine Similarity dan ANN layak untuk digunakan dalam mengklasifikasikan data pada Sistem Pengaduan Masyarakat Online.

Kata Kunci— Klasifikasi Data, Cosine Similarity, Artificial Neural Network

I. PENDAHULUAN

Teknologi komunikasi dan informasi masa kini yang semakin canggih sangat memudahkan berkomunikasi jarak jauh dan memiliki berbagai macam manfaat pada kehidupan sehari-hari. Dampak positif yang ditimbulkan dari perkembangan teknologi salah satunya adalah kemudahan dalam menyampaikan aspirasi dan dalam mendapatkan informasi dengan sangat cepat. Masyarakat yang ada pada dunia ini dapat dengan mudah dalam berinteraksi secara digital karena penyebaran informasi dan komunikasi pun tidak memiliki batas karena adanya perkembangan teknologi ini. Manfaat dari perkembangan teknologi ini dapat dirasakan oleh semua sektor, termasuk sektor pemerintahan yang harus mengayomi masyarakat dan negara. Memberikan pelayanan terhadap masyarakat termasuk salah satu tugas pemerintah. Dalam meningkatkan kualitas pelayanan terhadap masyarakat secara optimal, pemerintah harus menerapkan pemerintahan yang berbasis teknologi informasi digital dan online agar pelayanan publik yang diberikan dapat berjalan secara efektif dan efisien. Oleh sebab itu, Pemerintah pusat maupun daerah dapat menyediakan layanan pengaduan masyarakat yang berbasis online. Layanan tersebut digunakan agar masyarakat dapat berinteraksi dan berkomunikasi serta berpartisipasi secara langsung dengan pemerintah untuk melaporkan suatu kejadian yang menjadi tanggung jawab pemerintah ataupun dapat menyampaikan aspirasinya melalui platform atau aplikasi yang telah disediakan oleh pemerintah.

Demi meningkatkan kualitas pelayanan publik yang harus dilakukan oleh pemerintah terhadap masyarakatnya melalui sistem pengaduan online, maka dibutuhkan sebuah pengolahan data yang cepat dan akurat agar tidak menghambat proses penanganan pengaduan masyarakat tersebut. Text mining adalah metode pengolahan data yang memiliki arti sebagai penambang data berupa teks yang berasal dari dokumen, dan memiliki tujuan untuk mencari kata-kata atau kalimat berasal dari isi dokumen yang digunakan untuk menganalisa kethubungan antar dokumen satu dengan dokumen lainnya [1]. Adapun metode pengolahan data berikutnya yang dianggap efektif dalam meningkatkan kualitas kinerja sebuah sistem adalah *Machine Learning*. Pembelajaran Mesin atau yang biasa

¹Mahasiwa, Program Magister Teknik Elektro, Jln. Dalung Permai Blok A3 No 36, Badung, 80361 (telp: 085737601858; e-mail: lelymeilina452@gmail.com)

^{2, 3} Dosen, Program Pasca Sarjana, Magister Teknik Elektro Fakultas Teknik Universitas Udayana, Jln. P.B. Sudirman, Denpasar, Bali 80232 INDONESIA (telp: 0361-239599; email: ²satya.kumara@unud.ac.id)



disebut *Machine Learning* adalah sekumpulan algoritma pemrograman yang sering digunakan dalam mengoptimalkan kinerja komputer atau perangkat lunak berdasarkan data sampel yang sebelumnya sudah ada [2].

Metode *text mining* yang dapat diterapkan dalam pengolahan data pengaduan adalah Metode *Cosine Similarity* dimana metode tersebut digunakan untuk mencari kemiripan teks dokumen pengaduan dan metode *Artificial Neural Network* (ANN) untuk klasifikasi data pengaduan. Metode *Cosine Similarity* biasanya digunakan untuk dapat menghitung tingkat kesamaan dari lebih dari satu objek [3]. Sedangkan Metode ANN adalah sistem pengolahan data yang terinspirasi dari konfigurasi otak manusia dengan pemrosesan data yang saling berhubungan dan bertindak sama sekali hingga mencapai suatu permasalahan tertentu [4].

Pada artikel ini akan mencoba menelaah literatur yang membahas tentang penerapan Metode *Cosine Similarity* dan *Artificial Neural Network* sebagai acuan untuk meningkatkan kualitas sebuah sistem layanan pengaduan masyarakat. Hasil dari literatur review ini diharapkan dapat memberikan tingkat akurasi dalam penggunaan metode tersebut sebagai solusi untuk klasifikasi data dan pencarian kemiripan dokumen.

II. TINJAUAN PUSTAKA

A. Text Mining

Text Mining adalah proses pengumpulan dokumen menggunakan seperangkat alat analisis sebagai bentuk pengetahuan yang intensif kepada penggunanya [5]. Text mining juga memiliki tujuan untuk mencari kata-kata atau kalimat yang isinya terdapat dari dokumen untuk dapat dilakukan analisa keterhubungan antar dokumen satu dengan dokumen lainnya [1]. Adapun tahapan dari text mining secara umum adalah *case folding*, *tokenizing*, *filtering*, *stemming* dan *analyzing*.

B. Case Folding

Case Folding adalah sebuah proses untuk mengubah semua karakter teks yang ada menjadi huruf non kapital serta proses untuk menghilangkan angka dan tanda baca.

1). Tokenizing

Tokenizing adalah proses untuk menguraikan deskripsi yang awalnya berupa kalimat menjadi kata-kata berdasarkan pemisah kata dan menghilangkan delimiter seperti tanda koma (,), tanda titik (.), spasi dan karakter angka yang ada pada kalimat tersebut [6].

2). Filtering

Filtering digunakan untuk menghapus kata-kata yang terlalu sering muncul dan tidak dipakai dalam pemrosesan bahasa alami. Proses ini memiliki tujuan untuk mengurangi volume kata agar kata-kata yang didapatkan dalam dokumen hanya kata-kata penting saja. Algoritma dalam proses filtering terbagi menjadi dua yaitu *stop-list* dan *word-list*. Algoritma *stop-word* biasanya digunakan untuk mengeliminasi kata-kata yang tidak deskriptif, sedangkan algoritma *word-list* digunakan untuk menyimpan kata-kata yang memiliki nilai deskriptif.

3). Stemming

Stemming adalah proses yang digunakan sebagai penggabungan atau pemecahan suatu kata menjadi kata dasar. Proses ini dilakukan dengan cara menghilangkan awalan kata, akhiran, sisipan, serta kombinasi dari awalan dan akhiran.

4). Analyzing

Tahap *Analyzing* adalah tahap yang digunakan untuk menentukan seberapa jauh kemiripan dokumen teks satu dengan yang lain. Algoritma TF-IDF (*Term Frequency – Inversed Document Frequency*) yang dikombinasikan dengan metode *Cosine Similarity* dan *Artificial Neural Network* dapat digunakan sebagai tahap analisa pada tahapan ini.

C. Algoritma TF-IDF

Algoritma TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah suatu algoritma yang digunakan untuk memberi bobot hubungan suatu kata (*term*) terhadap dokumen. Terdapat dua konsep dalam algoritma ini dalam perhitungan bobot, yaitu *term frequency* atau frekuensi kemunculan sebuah term di dalam sebuah dokumen tertentu dan *inverse document frequency* atau inverse frekuensi dokumen yang mengandung term tersebut [7]. Perhitungan ini menggambarkan seberapa pentingnya kata dalam sebuah dokumen. Tujuan dari TF-IDF adalah menemukan jumlah kata yang diketahui (tf) setelah dikalikan dengan frekuensi aduan dimana suatu kata tersebut muncul (idf). Rumus yang digunakan adalah sebagai berikut [7]:

$$w_{dt} = tf_{dt} * IDF_t \quad (1)$$

d = Dokumen ke-d.

t = Kata ke-t dari kata kunci.

W = Bobot dokumen ke-d terhadap kata ke-t.

IDF = *Inverse Document Frequency* ($\text{Log}_2(D/df)$).

D = Total seluruh dokumen.

tf = Banyaknya dokumen yang mengandung kata terkait dengan kata yang dicari.

Setelah bobot (w) masing-masing dokumen diketahui, maka dilakukan proses sorting/pengurutan dimana semakin besar nilai w, semakin besar tingkat similaritas dokumen tersebut terhadap kata yang dicari, demikian sebaliknya.

D. Cosine Similarity

Cosine similarity adalah metode similaritas yang digunakan untuk menghitung similitas dua buah dokumen. Metode yang dipergunakan adalah melakukan perhitungan ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah vektor yang dibandingkan karena cosinus dari 0 adalah 1 dan kurang dari 1 untuk nilai sudut yang lain [8]. Kalau nilainya 0 maka dokumen tersebut dikatakan mirip jika hasilnya 1 maka nilai tersebut dikatakan tidak mirip. *Cosine similarity* tidak hanya digunakan untuk menghitung normalisasi panjang dokumen tapi juga menjadi salah satu ukuran kemiripan yang populer [9]. Rumus dapat dilihat sebagai berikut [3]:

$$\cos(d_j, q_k) = \frac{\sum_{i=1}^n (t_{d_{ij}} \times t_{q_{ik}})}{\sqrt{\sum_{i=1}^n t_{d_{ij}}^2 \times \sum_{i=1}^n t_{q_{ik}}^2}} \quad (2)$$

Keterangan:

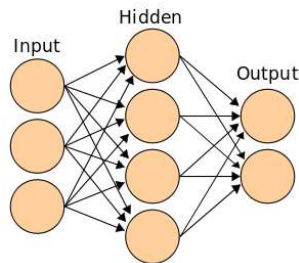
$\cos(dj, qk)$: Tingkat kesamaan dokumen dengan *query* tertentu
 td_{ij} : *Terms* ke- i dalam vektor untuk dokumen ke- j
 tq_{ik} : *Terms* ke- i dalam vektor untuk *query* ke- k
 n : Jumlah *term* unik dalam *dataset*.

Untuk mengukur kemiripan dokumen, dibutuhkan *query* input dan beberapa dokumen pembanding sebagai acuan dari *query* input tersebut. Kemudian dilakukan perhitungan menggunakan rumus diatas untuk mencari nilai similarity akhir.

E. Artificial Neural Network

Artificial Neural Networks (ANN) atau lebih dikenal dengan *Neural Networks* (NN) merupakan sebuah metode bagaimana komputer dapat mempelajari serta mengenali sesuatu tugas yang dihadapi komputer tersebut. Hal ini merupakan representasi dari jaringan jaringan biologis yang dimiliki oleh manusia. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola data Elemen yang paling mendasar dari jaringan syaraf adalah sel syaraf.

Secara matematis, NN seperti sebuah *graf*, NN memiliki neuron atau node (*vertex*) dan sinapsis (*edge*). Model ini meniru cara kerja jaringan neural biologis. Otak manusia terdiri atas sel-sel yang disebut neuron. Jenis sel ini memiliki keistimewaan dibanding jenis sel yang lain dimana sel lain selalu memproduksi dirinya sendiri kemudian mati akan tetap neuron dapat bertahan hidup. Hal ini yang menyebabkan informasi yang berada di dalamnya dapat bertahan. Neuron-neuron ini terbagi atas jaringan yang dibedakan atas fungsinya. Arsitektur *Artificial Neural Network* terdiri dari input [1], hidden layer dan output.



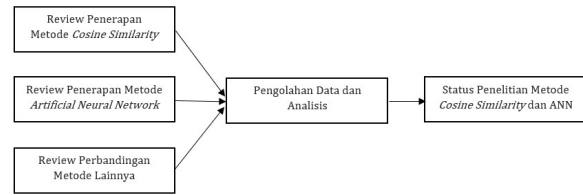
Gambar 1: Neural Network Concept

Input Layer merupakan layer tempat sebuah input dimasukkan (inisialisasi input) dan dari layer ini dilakukan proses-proses selanjutnya. Hidden Layer berfungsi untuk membantu proses, semakin banyak hidden layer digunakan maka semakin bagus dan semakin cepat pula didapat output yang diinginkan, tetapi waktu training akan semakin lama. Output Layer adalah layer yang menampung hasil proses suatu neural network.

III. METODOLOGI PENELITIAN

Sumber data yang digunakan pada penelitian ini berasal dari artikel jurnal, publikasi konferensi, mesin pencarian seperti Google Scholar, Garuda Ristekbrin, IEEE Xplore dan sumber lainnya. Penelitian ini memiliki tujuan untuk mengetahui tingkat akurasi penggunaan metode *Cosine Similarity* untuk pencarian

kemiripan dokumen dan *Artificial Neural Network* yang digunakan untuk klasifikasi data. Beberapa tahapan penelitian digambarkan dalam skematik yang dapat diperhatikan pada Gambar 1.



Gambar 2: Skematik Penelitian

Dapat dilihat pada Gambar 2, bahwa tahapan pertama dari penelitian ini adalah pengumpulan data. Proses pengumpulan data yang dilakukan pada penelitian ini adalah dengan cara mengumpulkan literature review yang berkaitan dengan penerapan metode *Cosine Similarity*, *Artificial Neural Network* (ANN) dan perbandingan metode lainnya sebagai pembanding keakuratan metode tersebut. Tahap kedua adalah melakukan pengolahan data dan Analisa dari literature review yang telah dilakukan. Tahap ketiga adalah melakukan Analisa terhadap status penelitian terkait dengan Metode *Cosine Similarity* dan *Artificial Neural Network* untuk mengetahui tingkat akurasi metode tersebut.

IV. HASIL DAN PEMBAHASAN

Berdasarkan dari tinjauan literatur yang sudah dilakukan pada penelitian ini, artikel ilmiah yang diperoleh sebanyak 40 buah dokumen. Namun yang relevan dengan penelitian yang ingin dilakukan adalah sebanyak 24 dokumen yang terdiri dari 10 artikel ilmiah yang membahas tentang penerapan metode *Cosine Similarity*, 7 artikel ilmiah yang membahas tentang penerapan metode *Artificial Neural Network* dan 7 artikel ilmiah yang membahas metode lain. Adapun penjelasan yang lebih detail dari tinjauan literatur yang telah dilakukan adalah sebagai berikut.

A. Penerapan Metode Cosine Similarity

Pada penelitian yang dilakukan oleh Melita, memiliki tujuan untuk membangun sebuah sistem Syarah Hadits dengan Bahasa Indonesia pada pembobotan kata yang menggunakan algoritma TFIDF dan metode *Cosine Similarity* untuk melakukan pencarian dokumen yang relevan. Tahapan pertama adalah melakukan Teks *Preprocessing* untuk mengubah teks menjadi data yang akan diolah. Kemudian tahap pembobotan kata digunakan untuk memberikan bobot suatu kata (*term*) pada dokumen dan tahap yang terakhir adalah perhitungan metode cosine similarity dalam mencari kemiripan teks antar dokumen. Presentase dari hasil pengujian stemming memiliki nilai akurasi sebesar 90,93% yang menunjukkan bahwa metode ini memiliki tingkat akurasi yang cukup tinggi. Setelah itu menggunakan metode pengujian confusion matrix yang mendapatkan nilai precision sebesar 100%, nilai recall sebesar 88,7%, nilai akurasi sebesar 88,73%, dan memiliki nilai error rate sebesar 11,27%. Sehingga dapat dikatakan bahwa sistem yang dibuat dalam penelitian ini berjalan sesuai harapan dengan baik [10].



Penelitian yang dilakukan oleh Firdaus, memiliki tujuan untuk dapat mengimplementasikan *Cosine Similarity* dalam meningkatkan akurasi pengukuran kesamaan dokumen dan klasifikasi dokumen berita menggunakan metode *K-Nearest Neighbors* (KNN). Nilai akurasi tertinggi yang didapat pada metode cosine similarity adalah sebesar 99,06%. Sedangkan nilai akurasi tertinggi pada euclidean distance adalah sebesar 60,05%. Hasil akhir dari perhitungan nilai rata-rata dengan klasifikasi KNN yang menggunakan *cosine similarity* adalah 98,12% dan *euclidean distance* 56,51% [11].

Pada penelitian yang dilakukan oleh Syahroni, bertujuan untuk mengukur dan mengidentifikasi kemiripan teks dokumen laporan antara query dengan dokumen pada sistem Layanan Pengaduan dan Aspirasi Online Rakyat (LAPOR!). Penanganan laporan pada sistem "LAPOR!" bergantung pada administrator. Penelitian ini menggunakan metode *cosine similarity* yang mendapatkan hasil akurasi secara optimal dengan proses pengujian *preprocessing Stemming* dan 75% rasio data latih memiliki hasil terbaik dari semua fitur dan rasio data uji sebesar 25%. Pada fitur *Cosine Similarity* berbasis terms weighting TF-IDF memiliki presentase sebesar 84%. Kemudian fitur CoSim-TFICF memiliki 40% rasio data latih dan data uji sebesar 60% [12].

Pada penelitian yang dilakukan oleh Arief, membuat sistem rekomendasi musik, di mana sistem itu sendiri dapat merekomendasikan lagu berdasarkan kesamaan artis yang disukai atau pernah didengar pengguna. Penelitian ini menggunakan metode *Colaborativ Filtering* dengan algoritma *Cosine Similarity* dan *K-Nearest Neighbors* (KNN). Algoritma KNN mengukur jarak dalam menentukan "kedekatan" dari instance artis dan menemukan 372 tetangga terdekat dari instance tersebut. Setelah didapatkan nilai kemiripan (jarak) antar seniman maka akan diprediksi rekomendasi artis dengan *k-neighbor* terdekat, dimana nilai terdekat akan diurutkan dari nilai yang tertinggi hingga nilai terendah dan rekomendasi teratas dipilih dari nilai tersebut [13].

Penelitian yang dilakukan oleh Thada adalah melakukan analisis komparatif untuk menemukan dokumen yang paling relevan untuk kumpulan kata kunci tertentu dengan menggunakan tiga koefisien kesamaan yaitu koefisien Jaccard, Dice dan Cosine. Penulis telah memilih fungsi *roulette* atau pemilihan kromosom yang paling cocok setelah setiap generasi. Pengkodean lengkap telah dilakukan di perangkat lunak Matlab versi R2009b [14].

Penelitian yang dilakukan oleh Abdullah, berkonsentrasi pada pengembangan algoritma *Multi Scanning Filter* (MSF), yang bekerja pada dokumen penelitian yang ditemukan di berbagai database ilmiah, seperti ISI, SCOPUS atau EBSCO, dll. Ide penelitian ini bergantung pada Google Search Engine, di mana algoritma yang diusulkan terdiri dari tiga bagian. Hasil yang didapat menurut filter yang diusulkan jauh lebih baik daripada yang dihasilkan menurut Mesin Pencari Google tradisional [15].

Pada Penelitian yang dilakukan oleh Lin bertujuan untuk mengusulkan algoritma pengambilan kesamaan kosinus baru untuk penalaran kasus, di mana skala pengukuran campuran dipertimbangkan. Hasil percobaan menunjukkan bahwa efisiensi pencocokan kesamaan dari algoritma pencocokan yang diusulkan meningkat secara signifikan dan menunjukkan kemampuannya dengan 73% presisi selama pengambilan. Hasil

eksperimen menunjukkan bahwa algoritma pencocokan kesamaan yang diusulkan pada sistem penalaran berbasis kasus efektif dan cocok untuk menangani pengambilan gambar desain interior berbasis kasus [16].

Penelitian yang dilakukan oleh Purba, dkk ini memiliki tujuan dalam menguji kesamaan Laporan Akhir Mahasiswa Politeknik Unggul LP3M Medan. Metode yang digunakan adalah Metode *Cosine Similarity*. Uji coba yang dilakukan terhadap 30 dokumen tugas akhir siswa yang berbeda diperoleh nilai kesamaan tertinggi sebesar 41%. Berdasarkan hasil pengujian dapat disimpulkan bahwa Metode *Cosine Similarity* memiliki normalisasi panjang vektor dengan cara melakukan perbandingan tingkat sejajar antar dokumen [17].

Penelitian yang dilakukan oleh Bagus Sujasman adalah mengukur kemiripan suatu produk dengan produk yang lain dengan menerapkan metode *Cosine Similarity*. Penelitian ini memiliki tujuan untuk mendapatkan kepastian akan produksi perangkat lunak yang berkualitas tinggi dan memenuhi kebutuhan pengguna dengan jadwal dan anggaran yang dapat diprediksi. Hasil dari penelitian ini adalah penerapan metode *Cosine Similarity* pada aplikasi penjualan berbasis mobile untuk mencari persentase kemiripan produk tertinggi [18].

Tujuan penelitian yang dilakukan oleh Zahrotun adalah untuk mencari kesamaan nilai yang paling optimal. Metode yang digunakan adalah gabungan kesamaan Jaccard dan *cosine similarity*. Hasil *cosine similarity* memiliki nilai tertinggi dibandingkan dengan Jaccard similarity dan joint antara Cosine dan Jaccard similarity. Dari pengelompokan hasil penggunaan parameter eps SNN sangat mempengaruhi pembentukan *cluster*. Semakin besar nilainya maka eps yang terbentuk *cluster* akan semakin sedikit [19].

B. Penerapan Metode Artificial Neural Network (ANN)

Penelitian yang dilakukan oleh Pardomuan, memiliki tujuan untuk klasifikasi kinerja PDAM berdasarkan indikator terpilih menggunakan algoritma *Artificial Neural Network* (ANN) dan Supporti *Vector Machine* (SVM). ANN digunakan untuk memprediksi kinerja perusahaan pada tahun berjalan dengan menggunakan 3 atribut yaitu Rasio Operasi, Jam Operasi Layanan per-hari, dan Rasio Jumlah Pegawai per 1000 pelanggan dengan hasil nilai rata-rata akurasi 83.93% untuk kinerja Sehat dan prediksi untuk kinerja Tidak Sehat sebesar 86.36%. Hal ini sedikit lebih baik apabila dibandingkan dengan algoritma SVM model terpilih yaitu memiliki rata-rata akurasi 82.14% dan untuk kinerja Tidak Sehat sebesar 80% [4].

Penelitian yang dilakukan oleh Yuliana ini membahas tentang penerapan *Artificial Neural Network* (ANN) dalam klasifikasi teks pengaduan masyarakat. Hasil uji dengan 10 label menunjukkan bahwa tingkat akurasi yang dihasilkan oleh metode ANN sebesar 43%. Dari presentase yang dihasilkan dapat dikatakan bahwa penggunaan metode ANN masih tergolong rendah tersebut masih tergolong rendah dan memiliki waktu yang relatif lama [20].

Pada penelitian yang dilakukan oleh Hadju menggunakan pembelajaran mesin yaitu *Artificial Neural Networks* untuk menentukan kemungkinan permintaan pertemanan Facebook itu asli atau tidak. Setiap persamaan pada setiap neuron (node) dimasukkan melalui fungsi *sigmoid* untuk menjaga hasil antara interval 0.0 dan 1.0. Hasil dari metode evaluasi ditugaskan

untuk skor. Skor tersebut digunakan untuk menentukan apakah profil tersebut asli atau palsu. Setiap persamaan di setiap neuron (node) dimasukkan melalui fungsi *sigmoid*. Penulis menggunakan kumpulan data pelatihan oleh Facebook atau jejaring sosial lainnya. Ini akan memungkinkan algoritma pembelajaran mendalam yang disajikan untuk mempelajari pola perilaku bot dengan propagasi mundur, meminimalkan fungsi biaya akhir dan menyesuaikan bobot [21].

Penelitian yang dilakukan oleh Farel, memiliki tujuan untuk melestarikan bahasa daerah agar pada perkembangan zaman ini bahasa tersebut tidak punah. Bahasa daerah dapat dilestarikan dengan cara mengenali karakter dari citra teks dokumen Bahasa Indonesia yang memiliki fokus pada hasil terjemahan menjadi bahasa daerah. Hasil dari klasifikasi citra teks dokumen, menunjukkan bahwa klasifikasi citra yang berhasil diklasifikasikan adalah sebanyak 1.591 dari total 5000 citra [22].

Penelitian yang dilakukan oleh Putra, bertujuan bertujuan untuk merancang arsitektur *Artificial Neural Network* (ANN) untuk memprediksi produksi padi yang menggunakan metode *Artificial Neural Network* dan Algoritma *Backpropagation*. Nilai MAPE yang didapatkan yaitu 11,86 % dan nilai akurasi atau keakuratan prediksi sebesar 88,14% dengan akurasi tertinggi mencapai 98,89%. Hasil dari nilai Mape didapatkan sesuai dengan prediksi data aktual, yaitu apabila nilai MAPE semakin kecil maka kinerja prediksi akan semakin baik. Dari implementasi sistem yang dirancang, diperoleh hasil prediksi dengan akurasi mencapai 88,14% [23].

Pada penelitian yang dilakukan oleh Yafitra, memprediksi suatu pergerakan saham dalam suatu perusahaan dilakukan dengan memanfaatkan ARIMA dan *Artificial Neural Network* (ANN). Skenario terbaik dalam memprediksi harga saham dengan menggunakan ANN adalah dengan arsitektur 20 – 16 – 1, yaitu 20 adalah data masukan, 16 adalah jumlah *hidden layer*, dan 1 adalah nilai keluarannya. Dari perhitungan error RMSE, model ARIMA (1,0,0) memiliki eror sebesar 1,3738 dan model ARIMA (2,0,0) memiliki eror sebesar 1,5514. Sedangkan metode *Artificial Neural Network* memiliki error sebesar 4.6814. Hasil akhir yang diperoleh dari penelitian ini adalah model ARIMA lebih tinggi akurasinya untuk memprediksikan harga saham pada PT. Bumi Citra Permai Tbk [24].

Penelitian yang dilakukan oleh Fitriana, membahas peramalan jumlah kasus sembuh, kasus meninggal, dan kasus sembuh terkait pandemic Covid-19 dengan menggunakan metode *Artificial Neural Network* (ANN) dan *Double Exponential Smothing*, yang kemudian nantinya akan dibandingkan hasil antara kedua metode tersebut. Dengan menggunakan metode DES diperoleh alpha optimum sebesar 0.199 dan beta optimum sebesar 0.055 pada kasus positif. Untuk kasus meninggal diperoleh Alpha optimum sebesar 0.173 dan beta optimum sebesar 0.017. Sedangkan kasus sembuh diperoleh nilai alpha optimum sebesar 0.349 dan beta optimum sebesar 0.030. Hal ini membuktikan bahwa pendekatan metode ANN lebih baik dibandingkan dengan DES, yang ditunjukkan oleh nilai RMSE yang lebih kecil pada metode ANN [25].

C. Perbandingan Metode Lainnya

Penelitian yang dilakukan oleh Gabriella, membahas tentang pengelompokan dokumen hasil pencarian berdasarkan kategori yang dapat menggambarkan isi dari suatu dokumen menggunakan metode *clustering*. Nilai akurasi saat menerapkan metode *clustering* berbeda jauh dengan tidak menerapkan *clustering* dengan nilai tertinggi hanya mencapai 62%. Sedangkan saat tidak menerapkan *clustering* data yang dibandingkan secara keseluruhan sehingga mendapat nilai persentase akurasi tinggi, yaitu mencapai 100%. Waktu yang dibutuhkan aplikasi untuk proses mendapat hasil kemiripan lebih sedikit atau lebih cepat waktu komputasinya ketika menerapkan metode *clustering* dibandingkan saat tidak menerapkan metode tersebut [26].

Metode yang digunakan pada penelitian Supriadi adalah mengelola dokumen dengan klasifikasi atau *clustering* dokumen. Dari hasil pengujian didapat nilai akurasi tertinggi yaitu 92%. Pengujian nilai presisi tertinggi yaitu 42%. Sedangkan pengujian nilai *recall* sebesar 72%. Penelitian ini bertujuan untuk memecah kata dengan waktu yang singkat dan cepat menggunakan metode algoritma N Gram dan Tanimoto Cosine untuk pencarian artikel bahasa indonesia berbasis online [27].

Pada penelitian yang dilakukan oleh Rozi ini ditujukan untuk memperoleh klasifikasi pada aplikasi web yang dibangun oleh pemerintah yaitu Sistem SP4N LAPOR! Penelitian yang menggunakan *Long Short Term Memory Recurent Neural Network* ini digunakan untuk memproses klasifikasi pada setiap dokumen laporan masyarakat. Pengujian pada penelitian ini menggunakan metode uji *KFold Cross Validation* dengan 10 data. Hasil pengujian menunjukkan rata-rata presentase dari f-mesure sebesar 85,69%. Nilai evaluasi rata-rata tertinggi adalah sebesar 88,82%. Dari hasil akurasi tertinggi dapat dilihat bahwa klasifikasi teks dokumen laporan masyarakat dapat menggunakan metode *Long Short Term Memory Recurent Neural Network* [28].

Penelitian yang dilakukan oleh Rismanto adalah mengembangkan sistem berbasis web dan menerapkan TF-IDF pembobotan kata dan metode kesamaan kosinus untuk memberikan rekomendasi kepada mahasiswa mengenai pembimbing tugas akhir yang memiliki melakukan penelitian sesuai dengan topik tugas akhir mahasiswa yang ditulis dalam bahasa Indonesia. Dalam 20 pengujian, akurasi perbandingan hasil rekomendasi sistem dengan data aktual diperoleh rata-rata 75% dengan membandingkan sistem rekomendasi dengan supervisor yang sebenarnya ditugaskan [29].

Penelitian yang dilakukan oleh Ningsih adalah membuat Sistem pengaduan masyarakat yang diterapkan di Kecamatan Kajen Kabupaten Pekalongan dalam menangani pengaduan masyarakat, atau dilakukan melalui *Short Message Service* (SMS), surat, e-mail, telepon, atau kotak pengaduan. Dalam pembangunan aplikasi ini penulis menggunakan sebuah Framework IONIC dan Angular JS sebagai front-end sistem, kemudian untuk bagian backend menggunakan bahasa pemrograman PHP dan MYSQL untuk database, serta Android SDK untuk membangun aplikasi menjadi bentuk mobile apk. Kemudian hasil aplikasi dapat diakses dengan internet sehingga pengguna harus memiliki internet untuk dapat menggunakan aplikasi ini [30].



Penelitian yang dilakukan oleh Sofyan ditujukan untuk memperoleh model klasifikasi pengaduan pelaporan rakyat dengan metode C4.5 berbasis *forward selection*. Hasil dari penelitian ini dengan pengujian pada 1100 dataset, 550 dataset, dan 275 dataset, dimana menghasilkan perbandingan akurasi yaitu metode c4.5 82,61%. Sedangkan dengan metode c4.5 berbasis *Forward selection* 85,27%, penggunaan dataset juga tidak mempengaruhi perbandingan tersebut, dimana 81,09% untuk c4.5 dan 85,27% untuk c4.5 berbasis *Forward Selection* sehingga dapat disimpulkan algoritma c4.5 *Forward Selection* lebih baik daripada c4.5 [31].

Penelitian yang dilakukan oleh Pramono adalah penggunaan algoritma *K-Nearest Neighbor* (KNN) dengan metode *text mining* sebagai klasifikasi laporan masyarakat agar Pemerintah Kota Semarang dapat mengelola pengaduan dengan lebih baik. Penelitian ini menggunakan *Confusion Matrix* sebagai uji validasi dan evaluasi, dimana parameter $k = 1$ memiliki nilai akurasi tertinggi yaitu 82% pada setiap kategori laporan. Hal ini menunjukkan bahwa klasifikasi dengan metode ini menghasilkan hasil yang cukup baik [32].

D. Hasil Pengukuran Metode Cosine Similarity

Penggunaan metode *Cosine Similarity* bertujuan untuk mencari kemiripan teks dokumen. Hasil pengukuran dari kajian studi literatur yang telah dilakukan berupa tingkat akurasi dari penerapan metode *cosine similarity* untuk menyelesaikan suatu permasalahan yang terjadi. Adapun hasil pengukuran dengan metode *Cosine Similarity* dapat diperhatikan pada Table 1.

TABLE I.
HASIL PENGUKURAN COSINE SIMILAIRTY

| Penelitian | Metode | Metode Uji | Akurasi (%) |
|------------|--------------------------------------|---|-------------|
| [10] | Algoritma TF-IDF & Cosine Similarity | Stemming | 90,93 |
| [11] | KNN & Cosine Similarity | Confusion matrix | 98,12 |
| [12] | Cosine Similarity | Stemming | 84 |
| [13] | KNN & Cosine Similarity | - | - |
| [14] | Jaccard, Dice & Cosine Similarity | Perbandingan Manual | 55 |
| [15] | Cosine Similarity dan Algoritma MSF | - | 85 |
| [16] | Cosine Similarity | Precision | 73 |
| [17] | Cosine Similarity | Confussion Matrix | 43 |
| [18] | Cosine Similarity | Perbandingan hasil dan perhitungan manual | 64,5 |
| [19] | Jaccard & Cosine Similarity | Perhitugan manual | 50 |
| Rata-rata | | | 71,5 |

Dari hasil pengukuran yang di dapat dari kajian literatur review, presentase rata-rata akurasi metode *cosine similarity* adalah 71.5% dari 10 buah dokumen. Penggunaan metode *cosine similarity* dapat dikatakan baik dan dapat di implementasikan untuk mencari kemiripan teks dokumen.

E. Hasil Pengukuran Metode Artificial Neural Network

Hasil pengukuran dari metode *Artificial Neural Network* (ANN) digunakan agar dapat mengetahui tingkat akurasi penggunaan metode dalam melakukan klasifikasi data. Adapun hasil pengukuran yang dilakukan pada studi literatur ini dapat diperhatikan pada Table 2.

TABLE II.
HASIL PENGUKURAN ARTIFICIAL NEURAL NETWORK

| Penelitian | Metode | Metode Uji | Akurasi (%) |
|------------|---------------------|-------------------|-------------|
| [4] | Algoritma ANN & SVM | Confussion Matrix | 83,10 |
| [20] | ANN | Confussion Matrix | 43 |
| [21] | ANN | - | - |
| [22] | ANN | - | - |
| [23] | ANN | Confussion Matrix | 88,14 |
| [24] | ANN & ARIMA | Error RMSE | 76 |
| [25] | ANN | Error RMSE | 95 |
| Rata-rata | | | 77 |

Kajian literatur yang telah dilakukan terkait dengan penggunaan metode ANN dalam klasifikasi dokumen mendapat nilai akurasi rata-rata sebesar 77% dari 7 buah dokumen. Hasil pengukuran dengan nilai tersebut menunjukkan bahwa penerapan metode ANN layak untuk di implementasikan.

F. Hasil Pengukuran Metode Lainnya

Hasil pengukuran dari metode lainnya digunakan sebagai perbandingan metode *Cosine Similarity* dan ANN untuk mengetahui seberapa tinggi tingkat akurasi dari kedua metode yang akan digunakan dengan metode lainnya seperti metode *Clustering*, Algoritma N Gram & Tanimoto Cosine, LSTM, TF-IDF, KNN, Algoritma C4.5 dan *Short Message Service* pada sistem pengaduan masyarakat. Adapun hasil pengukuran dari berbagai macam metode yang digunakan dalam klasifikasi data dan pencarian kemiripan dokumen dapat diperhatikan pada table 3.

TABLE III.
HASIL PENGUKURAN METODE LAINNYA

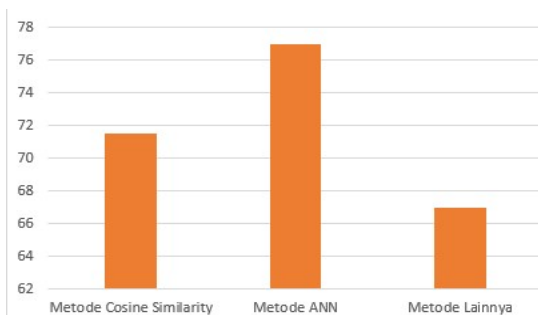
| Penelitian | Metode | Metode Uji | Akurasi (%) |
|------------|------------------------------------|-------------------------|-------------|
| [26] | Clustering | - | 62 |
| [27] | Algoritma N Gram & Tanimoto Cosine | Confussion Matrix | 92 |
| [28] | LSTM | K-Fold Cross Validation | 88,82 |
| [29] | TF-IDF | Comparation Data | 75 |

| | | | |
|-----------|-----------------------|-------------------|-------|
| [30] | Short Message Service | - | - |
| [31] | Algoritma C4.5 | Confussion Matrix | 71,27 |
| [32] | KNN | Confussion Matrix | 82 |
| Rata-rata | | | 67 |

Dari hasil pengukuran pada table 3, tingkat akurasi penggunaan metode lainnya memiliki nilai rata-rata sebesar 67% dari 7 buah dokumen. Hasil pengukuran tersebut menunjukkan bahwa penggunaan metode lainnya untuk mencari kemiripan teks dokumen dan klasifikasi data masih layak digunakan namun nilai akurasi yang dihasilkan lebih rendah.

G. Hasil Perbandingan Metode

Pada tahap sebelumnya telah dilakukan pengukuran nilai akurasi dari masing-masing metode untuk mengetahui seberapa layak metode yang akan diterapkan pada penelitian selanjutnya. Hasil pengukuran tersebut memiliki rata-rata nilai akurasi yang berbeda. Adapun perbandingan presentase nilai rata-rata dari metode *Cosine Similarity*, *Artificial Neural Network* dan Metode lainnya dapat diperhatikan pada Gambar 3.



Gambar 3. Akurasi Perbandingan Metode

Dapat dilihat bahwa grafik pada Gambar 3, nilai akurasi dari ketiga metode yang telah ditelaah tidak terlalu jauh presentase nilai perbandingannya. Metode *Artificial Neural Network* mendapat nilai presentase tertinggi yaitu 77% sedangkan metode *Cosine Similarity* mendapat nilai 71,5%. Penggunaan metode *cosine similarity* menentukan kesamaan dua buah objek dalam nilai sudut kosinus antara dua vektor dimana nilai terkecil adalah 0 dan nilai terbesar adalah 1. Hal ini menandakan bahwa apabila hasil yang diperoleh semakin mendekati nilai 1 maka tingkat kemiripan dokumennya semakin besar. Pada penelitian ini, nilai akurasi metode *cosine similarity* pada beberapa dokumen yang telah di review adalah 0,715 yang menandakan bahwa metode tersebut memiliki tingkat akurasi cukup tinggi karena berada diantara nilai 0 dan 1. Metode lainnya yang berkaitan dengan klasifikasi data dan pencarian persamaan teks dokumen seperti metode *Clustering*, Algoritma N Gram & Tanimoto Cosine, LSTM, TF-IDF, *Short Message Service*, Algoritma C4.5 dan KNN mendapatkan nilai presentase lebih rendah yaitu 67%. Setelah dilakukan Lely Meilina: Literature Review Klasifikasi Data...

perbandingan nilai presentase, dapat disimpulkan bahwa semua metode yang ditelaah layak digunakan dalam mencari kemiripan teks dokumen dan klasifikasi data pengaduan. Namun dalam memilih metode yang akan digunakan pada penelitian selanjutnya, dapat dipertimbangkan presentase nilai akurasi dari masing-masing metode sebagai acuan agar penerapan dan hasil akhir yang diperoleh lebih optimal dan akurat.

V. KESIMPULAN

Penelitian ini telah melakukan review penerapan metode *Cosine Similarity* dan *Artificial Neural Network* yang bertujuan untuk mengetahui tingkat akurasi penggunaan metode tersebut dibandingkan dengan metode lainnya dalam klasifikasi data yang akan digunakan pada sistem pengaduan masyarakat. Setelah melakukan literature review terdapat kesenjangan gap dari beberapa dokumen yang telah di review dalam mengklasifikasikan data sehingga dilakukan pengukuran tingkat akurasi dari masing-masing dokumen untuk mengetahui apakah metode yang diteliti masih layak digunakan. Dari hasil review menunjukkan bahwa rata-rata nilai hasil penerapan *Cosine Similarity* dan ANN memiliki tingkat akurasi yang lebih tinggi yaitu sebesar 71,5% dan 77% dibandingkan dengan metode lainnya yang memiliki presentase 67%. Penerapan menggunakan *Cosine Similarity* dan ANN dapat dikatakan layak digunakan karena memiliki presentase rata-rata tingkat akurasi lebih dari 50% sehingga metode tersebut lebih banyak digunakan untuk mencari kemiripan teks dokumen dan klasifikasi data dengan pemrosesan data yang lebih baik dan cepat. Sehingga kedua metode tersebut dapat digunakan untuk mencari kemiripan teks dokumen dan klasifikasi data pada sistem pengaduan masyarakat.

REFERENSI

- [1] H. C. Milkha, "Text Mining," 15 November 2020. [Online]. Available: <https://kesehatankerja.depkes.go.id>.
- [2] E. Alpaydin, "Introduction to Machine Learning Fourth Edition," *Adaptive Computation and Machine Learning series*, 2020.
- [3] Sugiyanto, B. Surarso och A. Sugiharto, "Analisa Performa Metode Cosine dan Jaccard Pada Pengujian Kesamaan Dokumen," *Jurnal Masyarakat Informatika*, p. Vol. 5 No 10, 2014.
- [4] P. R. Sihombing och O. P. Hendarsin, "Perbandingan Metode Artificial Neural Network (ANN) dan Support Vector Machine (SVM) untuk Klasifikasi Kinerja Perusahaan Daerah Air Minum (PDAM) di Indonesia," *Jurnal Ilmu Komputer*, p. Vol. XIII No 1, 2019.
- [5] R. Feldman och J. Sanger, *Text Mining Hand Book*, New York: Cambridge University Press, 2007.
- [6] S. M. Weiss, N. Indurkha, T. Zhang och F. J. Damerou, *Predictive Methods for Analyzing Unstructured Information*, New York: Springer, 2005.
- [7] A. Aziz och R. Saptono, "Implementasi Vector Space Model dalam Pembangkitan Frequently Asked Question Otomatis dan Solusi yang Relevan untuk Keluhan Pelanggan," *Scientific Journal of Informatics*, pp. Vol.2 Hal. 111-112, 2015.
- [8] K. Cios och L. Kurgan, *Data Mining: A Knowledge Discovery Approach*, Springer, 2007.
- [9] Tata, Sandeep, Patel och Jignesh, *Estimating The Selectivity Of TF-IDF Based Cosine Similarity Predicates*, Departement of Electrical Engineering and Computer Science University Of Michigan O'Brien, J. A., dan Marakas, G. M. *Management Information Systems*, 10 ed, 2007.



- [10] R. Melita, V. Amrizal, H. B. Suseno and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam)," *Jurnal Teknik Informatika*, vol. 12 No.2, 2018.
- [11] Firdaus, Pasnur and Wabdillah, "Implementasi Cosine Similarity Untuk Peningkatan Akurasi Pengukuran Kesamaan Dokumen Pada Klasifikasi Dokumen Berita Dengan K Nearest Neighbour," *Jurnal Teknologi Informasi dan Komunikasi*, Vol. 1, Nomor 1, 2019.
- [12] S. W. Iriananda, M. A. Muslim and H. S. Dachlan, "Identifikasi Kemiripan Teks Menggunakan Class Indexing Based dan Cosine Similarity," *Jurnal Ilmu Komputer dan Teknologi Informasi*, Vol. 1, No.2, pp. 30-38, 2018.
- [13] M. A. Budiman and G. A. V. Mastrika Giri, "Song Recommendations Based on Artists with Cosine Similarity Algorithms and K-Nearest Neighbor," *Jurnal Elektronik Ilmu Komputer Udayana*, 2019.
- [14] V. Thada and D. V. Jaglan, "Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm," *International Journal of Innovations in Engineering and Technology (IJET)*, 2018.
- [15] M. A. H. Al-Hagery, "Google Search Filter Using Cosine Similarity Measure to Find All Relevant Documents of a Specific Research Topic," *International Journal Of Education And Information Technologies*, 2016.
- [16] K.-S. Lin, "A case-based reasoning system for interior design using a new cosine similarity retrieval algorithm," *Journal of Information and Telecommunication*, 2019.
- [17] R. A. Purba, Suparno and Giatman, "The optimalization of cosine similarity method in detecting similarity degree of final project by the college students," *IOP Conf. Series: Materials Science and Engineering*, 2020.
- [18] M. Sujasman, Diana and A. Syazili, "IMPLEMENTASI METODE COSINE SIMILARITY UNTUK REKOMENDASI PRODUK PADA APLIKASI PENJUALAN BERBASIS MOBILE," *Bina Darma Conference on Computer Science*, 2020.
- [19] L. Zahrotun, "Comparison Jaccard similarity, Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method," *Computer Engineering and Applications*, 2016.
- [20] D. Yuliana, Purwanto and C. Supriyanto, "Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network," *UPI YPTK Jurnal KomTekInfo*, Vol. 1, No.3, pp. 92-116, 2019.
- [21] G. Hadju, Y. Minoso, R. Lopez, M. Acosta and A. Elleithy, "Use Of Artificial Neural Networks to Identify Fake Profiles," *IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, 2019.
- [22] F. Fathurrahman, M. M. Santoni and A. Muliawati, "Penerapan Artificial Neural Network Untuk Klasifikasi Citra Teks Dalam Penerjemahan Bahasa Daerah," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2020.
- [23] H. Putra and N. U. Walmi, "Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," *Jurnal Nasional Teknologi dan Sistem Informasi*, 2020.
- [24] B. Y. Pandji, Indwiarti and A. A. Rohmawati, "PERBANDINGAN PREDIKSI HARGA SAHAM DENGAN MODEL ARIMA DAN ARTIFICIAL NEURAL NETWORK," *Ind. Journal on Computing*, 2019.
- [25] N. Fitriana, R. Ghazian, S. Khoirina, T. Salsabila, V. Baby and Kariyam, "Perbandingan metode double exponential smoothing dan artificial neural network untuk meramalkan perkembangan covid-19 di Indonesia," *Seminar Nasional Matematika dan Pendidikan Matematika*, 2020.
- [26] G. E. Kambey, R. Sengkey and A. Jacobus, "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," *Jurnal Teknik Informatika*, 2020.
- [27] C. Supriadi, H. D. Purnomo and I. Sembiring, "Sensitivitas Sistem Pencarian Artikel Bahasa Indonesia Menggunakan Metode n-gram Dan Tanimoto Cosine," *TRANSFORMATIKA*, 2020.
- [28] I. F. Rozi, V. N. Wijayaningrum and N. Khozin, "Klasifikasi Teks Laporan Masyarakat Pada Situs Lapo! Menggunakan Recurrent Neural Network," *SISTEMASI : Jurnal Sistem Informasi*, Vol. 1, No.3, 2020.
- [29] R. Rismanto, A. R. Syulisto and B. P. C. Agusta, "Research Supervisor Recommendation System Based on Topic Conformity," *Modern Education and Computer Science*, pp. 26-34, 2020.
- [30] E. Y. Ningsih, I. Rosyadi and H. Handayani, "Sistem Informasi Pengaduan Online Pada Masyarakat Kecamatan Kajen Kabupaten Pekalongan Berbasis Web dan Android.," *Surya Informatika*, Vol. 1, No.1, 2020.
- [31] A. Sofyan and S. Santosa, "Text Mining Untuk Klasifikasi Pengaduan Pada Sistem Laporan Menggunakan Metode C4.5 Berbasis Forward Selection," *Jurnal Teknologi Informasi*, Vol. 212, No.1, 2016.
- [32] H. P. Hadi and T. S. Sukamto, "Klasifikasi Jenis Laporan Masyarakat dengan K-Nearest Neighbor Algorithm," *Journal of Information System*, Vol.25, No.1, 2020.