

Analisis Sentimen pada Ulasan Penyedia Layanan Menggunakan Algoritma C4.5

Miske Marcillia¹, Muhammad Fadlan², Lies Hartono³

¹Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
²Teknik Informatika, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
Email: miskemarcillia@gmail.com, fadlan@ppkia.ac.id, lies@ppkia.ac.id

Abstrak

Analisis sentimen adalah proses memahami dan mengolah data tekstual secara otomatis untuk memperoleh informasi. PT.XYZ merupakan salah satu perusahaan yang bergerak dalam pengoperasian sistem telekomunikasi Telepon Bergerak Seluler dengan sistem GSM (Global System for Mobile Communication). Perusahaan belum memiliki aplikasi khusus untuk menganalisis ulasan pengunjung terhadap pelayanan yang diberikan, sehingga sulit memperoleh pemahaman detail terkait ulasan konsumen guna meningkatkan kualitas pelayanan. Penelitian ini mengusulkan Aplikasi yang dapat menganalisis ulasan pada PT. XYZ guna menyelesaikan permasalahan. Data penelitian adalah ulasan konsumen PT. XYZ pada software Google Map sebanyak 140 ulasan. Metode penelitian ini yaitu Algoritma C4.5 untuk mengklasifikasi dan memprediksi ulasan dengan membentuk pohon keputusan untuk menemukan hubungan antarvariabel. Tahapan penelitian: Crawling data, Text Preprocessing, perhitungan TF (Term Frequency) kemudian analisis Algoritma C4.5 untuk menentukan kelas klasifikasi ulasan layanan. Penelitian ini telah berhasil memprediksi sesuai dengan tujuan penelitian. Penelitian dengan 126 data latih dan 14 data uji memperoleh nilai akurasi sebesar 78,57%, nilai Precision sebesar 83,33%, dan nilai Recall sebesar 90,91%. Selanjutnya, hasil pengujian mengungkapkan bahwa semakin banyak data latih maka semakin banyak pola yang diperoleh. Akurasi akan meningkat seiring bertambahnya jumlah data latih.

Kata Kunci: analisis, C4.5, sentimen, ulasan

Sentiment Analysis on Service Provider Reviews Using the C4.5 Algorithm

Abstract

Sentiment analysis is essential for understanding and processing textual data to derive meaningful insights. PT. XYZ, a company operating in the GSM cellular telecommunications sector, faces a challenge due to the lack of a specific application for analyzing visitor reviews on their services. This gap impedes their ability to gain detailed insights into consumer feedback, hindering efforts to improve service quality. This research addresses this issue by developing an application that utilizes the C4.5 algorithm to analyze PT. XYZ's reviews. The study uses 140 consumer reviews collected from Google Maps. The C4.5 algorithm, which creates decision trees to find relationships among variables, is employed for classifying and predicting service review sentiments. The research involves several stages: data crawling, text preprocessing, term frequency (TF) calculation, and applying the C4.5 algorithm for classification. The results demonstrate the effectiveness of this approach. With 126 training data samples and 14 test samples, the model achieved an accuracy of 78.57%, precision of 83.33%, and recall of 90.91%. These findings indicate that increasing the amount of training data enhances pattern recognition and accuracy. The study successfully meets its objectives, proving that sentiment analysis using the C4.5 algorithm can effectively predict service review sentiments and aid in improving service quality.

Keywords: analysis, C4.5, sentiments, reviews

I. PENDAHULUAN

PT. XYZ, sebuah perusahaan penyedia layanan telekomunikasi berbasis GSM di Indonesia, menghadapi tantangan dalam memahami secara mendalam umpan balik konsumen guna meningkatkan kualitas layanannya. Analisis sentimen adalah teknik otomatis dalam pemrosesan teks untuk

mengidentifikasi opini, yang lebih kompleks dibandingkan klasifikasi tradisional karena melibatkan ambiguitas bahasa, ketiadaan intonasi dalam teks, serta dinamika perubahan bahasa [1]. Tujuan dari analisis sentimen adalah mengidentifikasi dan mengelompokkan opini ke dalam

kategori sentimen seperti positif, negatif, atau netral, sehingga mempermudah interpretasi dan evaluasi data [2].

Text mining adalah teknik yang bertujuan mengekstraksi informasi dari dokumen teks melalui alat analisis yang berasal dari data mining, seperti pengkategorian. Karena data dalam text mining biasanya berbentuk teks tidak terstruktur, langkah awal yang diperlukan adalah menyusun teks tersebut agar lebih terstruktur sebelum menghitung frekuensi kemunculan kata [3]. Dokumen merupakan unit dasar dalam text mining, dan tujuan akhirnya adalah menemukan pola menarik dalam data teks tidak terstruktur yang dapat memberikan informasi bernilai. Tahap awal dalam text mining disebut Text Preprocessing, yang bertujuan untuk memilih data yang relevan dari setiap dokumen guna mempercepat keseluruhan proses analisis. Tahap ini meliputi Case Folding, Penghapusan Tanda Baca, Tokenisasi, Penghapusan Stop Word, dan Stemming [4].

Algoritma C4.5, yang dikembangkan oleh J. Ross Quinlan, adalah pengembangan dari algoritma ID3. Sementara ID3 hanya dapat menangani atribut diskrit dan tidak bisa mengatasi nilai yang hilang, Algoritma ini menentukan atribut terbaik untuk membagi data berdasarkan nilai informasi atau gain yang dihitung dari entropi. Entropi mengukur ketidakpastian dalam dataset, dan gain menunjukkan seberapa banyak informasi yang diperoleh melalui pemisahan dataset berdasarkan atribut tertentu [5]. Beberapa penelitian telah menggunakan C4.5 dalam melakukan analisis sentimen, diantaranya Penelitian oleh Ezza, dkk (2021) menggunakan C4.5 untuk menganalisis sentimen kepuasan mahasiswa dalam proses pembelajaran secara daring [6], selain itu penelitian oleh Rakhman (2019) menggunakan C4.5 untuk analisis sentimen terhadap review pengguna pada sebuah media massa [7].

Penelitian ini bertujuan untuk mengembangkan aplikasi analisis ulasan bagi PT. XYZ menggunakan metode Algoritma C4.5. Algoritma C4.5 merupakan metode klasifikasi yang membangun pohon keputusan guna mengidentifikasi hubungan antara variabel. Keunggulannya adalah kemampuan menangani data numerik maupun kategori, mengatasi nilai atribut yang hilang, dan menghasilkan aturan yang mudah dipahami.

Dalam penelitian ini, akurasi sistem diukur menggunakan metrik Precision dan Recall. Precision menilai seberapa tepat sistem dalam mengklasifikasikan data ke dalam kelas positif atau negatif, sementara Recall mengukur kemampuan sistem dalam mengenali dengan benar informasi yang berkaitan dengan data pada kelas-kelas tersebut.

II. METODOLOGI PENELITIAN

Penelitian ini mengembangkan sistem analisis sentimen menggunakan algoritma C4.5 pada ulasan berbahasa Indonesia mengenai pelayanan PT. XYZ. Data ulasan dikumpulkan dan diberi label secara manual ke dalam kategori positif dan negatif. Sebelum dianalisis menggunakan algoritma C4.5, data melalui beberapa tahap text preprocessing serta pembobotan kata menggunakan metode TF-IDF.

Tahap awal text preprocessing meliputi [8]:

- Case folding, yang mengubah semua huruf dalam dokumen menjadi huruf kecil.
- Remove punctuation, yaitu penghapusan semua tanda baca dari dokumen.
- Tokenization, yang memecah kalimat menjadi kata-kata dengan menggunakan fitur unigram.
- Stopword removal, untuk menghapus kata-kata yang tidak relevan agar hanya tersisa kata-kata bermakna.
- Stemming, yang menggunakan algoritma Nazief-Adriani untuk menghapus imbuhan seperti awalan, sisipan, atau akhiran dari kata-kata.

Setelah melalui tahapan tersebut, dokumen dilatih menggunakan pembobotan kata dengan metode TF-IDF, kemudian diterapkan algoritma C4.5 untuk memprediksi kelas sentimen berdasarkan nilai akurasi yang dihasilkan.

Dalam penelitian ini pengukuran akurasi menggunakan *Confusion Matrix* yang merupakan metode yang umum digunakan untuk mengukur akurasi dalam data mining dan sistem pendukung keputusan. Matrix ini mencakup empat elemen utama: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Dari *confusion matrix*, dapat dihitung metrik evaluasi seperti akurasi, precision, dan recall [9].

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data ulasan berbahasa Indonesia dari Google Maps yang terkait dengan pelayanan PT. XYZ. Ulasan tersebut diberi label manual menjadi dua kelas—Positif dan Negatif—oleh Ibu Anny Susilowati, S.Kom., M.H. Data dibagi menjadi data latih, yang telah diberi label, dan data uji, yang belum dilabeli. Data ulasan beserta kelasnya dapat dilihat pada Tabel 1.

Tabel 1. Data Ulasan

Data	Username	Ulasan	Label
D1	Okky Bima Saputra	Pelayanan yg cukup baik	Positif
D2	Slaymaster	Pelayanan super... Ramah...	Positif
D3	Nandia Karyajiwa	Pelayanan Kurang baik	Negatif
D4	Donna Rhamdan	Ramah	Positif
D5	Rahmad Fauzi	Tempat nyaman lahan parkir sempit	Negatif

Tahap preprocessing adalah langkah krusial dalam analisis sentimen. Pada fase ini, data disiapkan sebelum memasuki proses perhitungan. Proses ini mencakup transformasi data menjadi bentuk yang relevan untuk analisis nilai sentimen.

Langkah pertama dalam preprocessing adalah case folding, yaitu proses konversi seluruh huruf dalam kalimat ulasan menjadi huruf kecil. Huruf kapital dalam kalimat akan diubah menjadi huruf kecil, karena huruf kapital tidak mempengaruhi makna kata dalam konteks analisis sentimen dan dianggap

tidak berpengaruh terhadap nilai sentimen. Contoh penerapan case folding dapat dilihat pada Tabel 2.

Tabel 2. Hasil Case Folding

Data	Username	Ulasan
D1	Pelayanan yg cukup baik	pelayanan yg cukup baik
D2	Pelayanan super... Ramah...	pelayanan super... ramah...
D3	Pelayanan Kurang baik,	pelayanan kurang baik,
D4	Ramah	ramah
D5	Tempat nyaman lahan parkir sempit	tempat nyaman lahan parkir sempit

Langkah selanjutnya dalam proses ini adalah menghapus tanda baca atau Remove Punctuation. Pada tahap ini, data ulasan akan mengalami proses eliminasi angka (misalnya angka 1, 2, 3, 4, 5, 6, 7, 8, 9, 0) serta tanda baca seperti titik (.), koma (,), tanda tanya (?), tanda seru (!), dan simbol lainnya.

Angka dan tanda baca dianggap tidak memberikan kontribusi signifikan terhadap analisis nilai sentimen karena tidak memberi pengaruh pada data, sehingga komponen-komponen ini dihilangkan dari analisa. Contoh penerapan proses penghapusan tanda baca dan angka pada data ulasan dapat dilihat pada Tabel 3.

Tabel 3. Hasil Remove Punctuation

Dok	Hasil Case Folding	Hasil Remove Punctuation
D1	pelayanan yg cukup baik	pelayanan yg cukup baik
D2	pelayanan super... ramah...	pelayanan super ramah
D3	pelayanan kurang baik,	pelayanan kurang baik
D4	ramah	ramah
D5	tempat nyaman lahan parkir sempit	tempat nyaman lahan parkir sempit

Tahap berikutnya adalah tokenisasi. Pada fase ini, analisa dilakukan dengan menggunakan fitur unigram. Setiap kata akan dipisahkan menjadi unit kata individual. Contoh tokenisasi pada data ulasan dapat dilihat pada Tabel 4.

Tabel 4. Hasil Tokenize

Hasil Remove Punctuation	Hasil Tokenize			
pelayanan yg cukup baik	pelayanan	yg	cukup	baik
pelayanan super ramah	pelayanan	super	ramah	
pelayanan kurang baik	pelayanan	kurang	Baik	
ramah	ramah			

tempat nyaman lahan parkir sempit	tempat	nyaman	lahan	Parkir
-----------------------------------	--------	--------	-------	--------

Kumpulan kata dalam ulasan yang telah menjalani proses tokenisasi akan mengalami normalisasi kata. Proses ini mencakup perbaikan kata tidak baku menjadi bentuk kata baku serta normalisasi bahasa gaul dan singkatan. Normalisasi dilakukan dengan mencocokkan kata-kata tersebut dengan tabel normalisasi; jika kata ditemukan dalam tabel, maka kata tersebut akan diubah menjadi bentuk baku, sementara kata yang tidak ditemukan akan tetap dalam bentuk aslinya.

Setelah proses normalisasi, data akan melalui tahap penghapusan stopword, yang bertujuan untuk menghilangkan kata-kata yang tidak berkontribusi terhadap analisis sentimen. Pada tahap ini, kata-kata yang ada dibandingkan dengan daftar stoplist; kata-kata yang terdaftar dalam stoplist akan dihapus. Contoh penerapan penghapusan stopword pada data ulasan dapat dilihat pada Tabel 5.

Tabel 5. Hasil Tokenize

Data	Hasil Tokenize			
D1	pelayanan cukup	yg baik	pelayanan cukup	- baik
D2	pelayanan ramah	super	pelayanan ramah	super
D3	pelayanan baik	kurang	Pelayanan baik	kurang
D4	ramah		ramah	
D5	tempat lahan parkir sempit	nyaman parkir	-	nyaman parkir

Kumpulan kata dalam ulasan yang telah menjalani tahap penghapusan stopword selanjutnya akan mengalami proses stemming. Pada tahap ini, kata-kata diubah menjadi bentuk dasar dengan menghilangkan imbuhan seperti awalan, sisipan, akhiran, serta gabungan awalan dan akhiran. Jika sebuah kata memiliki imbuhan, maka kata tersebut akan dikembalikan ke bentuk dasarnya; jika tidak, kata tersebut akan dipertahankan dalam bentuk aslinya. Contoh penerapan stemming pada data ulasan dapat dilihat pada Tabel 6.

Tabel 6. Hasil Stemming

Data	Hasil Tokenize		
D1	layan	cukup	baik
D2	layan	super	ramah
D3	layan	kurang	baik
D4	ramah		
D5	nyaman	parkir	sempit

Tahap selanjutnya adalah metode TF-IDF (Term Frequency-Inverse Document Frequency). Tujuan dari tahap ini adalah untuk mengukur pentingnya sebuah kata dalam konteks dokumen. Data yang telah melalui proses text

preprocessing kemudian dianalisis untuk menentukan frekuensi kemunculan setiap kata dalam dokumen.

Proses TF-IDF menghitung bobot kata dengan menggabungkan frekuensi kata atau *term frequency* (TF) dan frekuensi dokumen terbalik atau *inverse document frequency* (IDF). TF mengukur bobot setiap istilah dalam ulasan, sementara *document frequency* (DF) menunjukkan jumlah dokumen yang mengandung istilah tertentu. Contoh data ulasan yang telah menjalani proses text preprocessing dapat dilihat pada Tabel 7.

Tabel 7. Hasil Text Preprocessing

Data	Hasil Preprocessing			Label
D1	layan	cukup	baik	Positif
D2	layan	super	ramah	Positif
D3	layan	kurang	baik	Negatif
D4	ramah			Positif
D5	nyaman	parkir	sempit	Negatif

Tahapan berikutnya adalah proses pembobotan, yang bertujuan untuk mengubah data teks menjadi representasi numerik dalam bentuk matriks. Proses ini melibatkan konversi istilah teks menjadi vektor numerik yang mencerminkan bobot masing-masing istilah berdasarkan frekuensi kemunculannya dan relevansi terhadap dokumen lainnya. Matriks numerik yang dihasilkan memungkinkan menampilkan data dalam model klasifikasi, di mana model tersebut dapat melakukan analisis dan prediksi. Contoh hasil pembobotan dapat dilihat pada Tabel 8.

Tabel 8. Hasil perhitungan TF, DF, IDF dan TF-IDF

Kata	TF					DF	D/DF	IDF
	U1	U2	U3	U4	U5			
layan	1	1	1	0	0	3	1.67	0.22
cukup	1	0	0	0	0	1	5	0.70
baik	1	0	1	0	0	2	2.50	0.40
super	0	1	0	0	0	1	5	0.70
ramah	0	1	0	1	0	2	2.50	0.40
nyaman	0	0	0	0	1	1	5	0.70
kurang	0	0	1	0	0	1	5	0.70
parkir	0	0	0	0	1	1	5	0.70
sempit	0	0	0	0	1	1	5	0.70

Selanjutnya, jumlah frekuensi dari hasil perhitungan TF-IDF berdasarkan kelasnya dapat dilihat pada Tabel 9.

Tabel 9. Jumlah Frekuensi

Jumlah Frekuensi		
Term	Positif	Negatif
layan	0.44	0.22
cukup	0.70	
baik	0.40	0.40
super	0.70	

ramah	0.80	
nyaman		0.70
kurang		0.70
parkir		0.70
sempit		0.70

Selanjutnya melakukan perhitungan entropy untuk seluruh training data menggunakan persamaan (1).

$$Entropy(S) = \sum_{j=1}^K - p_j * \log_2 p_j \quad (1)$$

Keterangan:

S = Himpunan partisi S

K = Banyaknya partisi S

Pj = Probabilitas dari sum (Ya dan Tidak)/Total Kasus

Rumus diatas juga digunakan untuk mencari entropi pada setiap kelas (positif dan negatif). Setelah mendapatkan nilai entropi, pemilihan atribut dilakukan dengan mencari nilai gain terbesar dengan rumus berikut.

$$Gain(A) = E(S) \sum_{i=1}^k (a_n \cos \frac{|s_i|}{|S|} * E(s_i)) \quad (2)$$

Keterangan:

A = Atribut

E = Entropi

S = Total nilai entropi

|Si| = Total jumlah (Ya dan Tidak)

|S| = Total jumlah pada kelas (Negatif dan Positif)

Contoh perhitungan entropy atribut "Layan" dengan nilai 0 dan 1 dapat dilihat pada perhitungan (3), (4), dan (5) berikut.

$$E(0) = \left(\frac{-41}{52}\right) * \log_2 \left(\frac{41}{52}\right) + \left(\frac{-11}{52}\right) * \log_2 \left(\frac{11}{52}\right) = 0.744 \quad (3)$$

$$E(1) = \left(\frac{-49}{60}\right) * \log_2 \left(\frac{49}{60}\right) + \left(\frac{-11}{60}\right) * \log_2 \left(\frac{11}{60}\right) = 0.687 \quad (4)$$

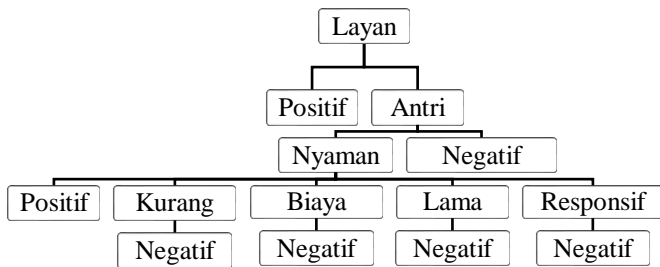
$$Gain(total, layan) = 11.664 - \left(\frac{52}{112}\right) * (0.744) + \left(\frac{60}{112}\right) * 0.687 = 11.686 \quad (5)$$

Proses perhitungan dilanjutkan hingga diperoleh simpul daun. Hasil perhitungan nilai gain yang digunakan untuk membentuk simpul akar pohon keputusan disajikan dalam Tabel 10.

Tabel 10. Perhitungan Gain untuk Simpul Akar

Kata	Pos (0)	Neg (0)	E (0)	P (1)	N (1)	E (1)	Gain
layan	41	11	0.744	49	11	0.687	11.686
Cukup	83	21	0.726	7	1	0.544	11.029
baik	77	21	0.750	13	1	0.371	11.054
super	89	22	0.718	1	0	0	10.952
ramah	67	20	0.778	23	2	0.402	11.149

Karena nilai entropy untuk cabang simpul “Layan” dengan nilai 0 masih tidak sama dengan 0, maka akan dilakukan perhitungan kembali untuk menentukan simpul selanjutnya. Dari hasil perhitungan nilai gain untuk simpul akar didapatkan rule tree yang dapat dilihat pada Gambar 1. Karena nilai entropy untuk cabang simpul “Kurang”, “Biaya”, “Lama”, dan “Responsif” telah mencapai hasil sama dengan 0 maka cabang atribut “Lumayan” berupa leaf node dan tidak dilakukan perhitungan gain untuk menentukan simpul selanjutnya.



Gambar 1. Rule Tree

Selanjutnya, untuk menghitung nilai akurasi, precision, dan recall dapat dilakukan dengan menggunakan hasil pengujian berdasarkan data training dan data testing yang telah diklasifikasikan dengan metode C4.5. Dengan menggunakan 126 data training dan 14 data uji sebagai berikut yang dapat dilihat pada contoh perhitungan berikut.

$$T_{pos} = 10 \mid T_{neg} = 1 \mid F_{pos} = 2 \mid F_{neg} = 1$$

Nilai akurasi, precision, dan recall dapat dihitung menggunakan rumus 2.7, 2.8 dan 2.9 sebagai berikut :

$$Akurasi = (10 + 1)/14 = 11/14 = 0,79 = 78,57\%$$

$$Precision = 10/(10 + 2) = 10/12 = 0,83 = 83,33\%$$

$$Recall = 10/(10 + 1) = 10/11 = 0,91 = 90,91\%$$

IV. KESIMPULAN

Dari hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa program yang digunakan untuk menganalisa sentimen terhadap data ulasan penyedia layanan dapat berjalan dengan yang diprediksikan, yaitu dapat menampilkan hasil klasifikasi data. Dengan program yang mengimplementasikan metode C4.5 ini, data ulasan dapat diklasifikasikan setelah melalui tahap text preprocessing,

pembobotan TF-IDF dan tahap klasifikasi dengan membagi data menjadi dua, yaitu data latih dan data uji. Data latih digunakan sebagai acuan pada tahap selanjutnya. Kemudian data uji akan melalui proses klasifikasi dengan metode C4.5 untuk mendapatkan hasil analisis sentimen. Dari hasil uji, didapatkan nilai akurasi sebesar 64,29% dengan menggunakan data latih sebanyak 98 data dan data uji sebanyak 42 data, dengan nilai precision sebesar 62,16% dan nilai recall sebesar 95,83%. Jika menggunakan data latih sebanyak 112 data dan data uji sebanyak 28 data, diperoleh nilai akurasi sebesar 71,43%, nilai precision 70,83% dan recall 94,44%. Sedangkan jika menggunakan data latih sebanyak 126 data dan data uji sebanyak 14 data, diperoleh nilai akurasi sebesar 78,57%, nilai precision 83,33% dan recall 90,91%. Dari hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa semakin banyak data latih maka akan semakin banyak pola yang diperoleh. Hasil Akurasi akan meningkat seiring bertambahnya data latih.

REFERENSI

- [1] J. A. A. Muhammad, A. Adiwijaya et al., *Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma Random Forest*, Jurnal e-Proceeding of Engineering, vol. 8, no. 5, 2021.
- [2] U. Fathia, *Penggunaan Google Review Sebagai Penilaian Kepuasan Pengunjung dalam Pariwisata*, Jurnal Ilmiah TORNARE, vol. 2, no. 1, 2020. doi:-10.24198/tornare.v2i1.25826
- [3] N. A. Dewi et al., *C4.5 Algorithm Implementation for Public Sentiment Analysis*, Jurnal Teknologi Informasi dan Komunikasi, vol. 13, no. 02, 2022.
- [4] L. Sri and S. Sudin, *Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Metode Naïve Bayes Classifier*, Jurnal Sismatik, vol. 1, no. 1, 2021. 10.24198/tornare.v2i1.25826.
- [5] Confusion Matrix, Binus University, 1 Nov. 2020. [Online]. Available: <https://socs.binus.ac.id/2020/11/01/confusion-matrix>.
- [6] A. Rakhman and M. R. Tsani, “Analisis Sentimen Review Media Massa Menggunakan Metode C4. 5 Berbasis Forward Selection,” *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 8, no. 2, pp. 78–82, 2019.
- [7] E. R. M. Razaq, D. W. Jacob, and F. Hamami, “Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi,” *eProceedings of Engineering*, vol. 8, no. 5, 2021.
- [8] R. Ade et al., *Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen*, Jurnal Ilmiah JLK, vol. 2, no. 1, 2019.
- [9] N. A. Dewiet al., *C4.5 Algorithm Implementation for Public Sentiment Analysis*, Jurnal Teknologi Informasi dan Komunikasi, vol. 13, no. 02, 2022.