
Desain Aplikasi Analisis Sentimen Ulasan Chatgpt pada Google Play Menggunakan Metode Naïve Bayes

Munirah¹, Aura Kristiani Pongamba², Suprianto*³, Lies Hartono⁴

^{1,2,3*}Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara

⁴ Program Studi Teknik Informatika, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara

Email: ¹2250009@student.ppkia.ac.id, ²2350108@student.ppkia.ac.id, ³supri@ppkia.ac.id, ⁴lies@ppkia.ac.id

Abstrak

Perkembangan teknologi kecerdasan buatan telah meningkatkan penggunaan aplikasi ChatGPT di Google Play Store sehingga menghasilkan berbagai ulasan yang mencerminkan pengalaman pengguna dalam bentuk sentimen positif, netral, maupun negatif. Ulasan tersebut dapat dimanfaatkan sebagai sumber informasi untuk mengetahui tingkat kepuasan pengguna serta sebagai bahan evaluasi terhadap kualitas layanan aplikasi. Namun, banyaknya jumlah ulasan yang tersedia menyebabkan proses analisis secara manual menjadi kurang efektif dan memerlukan waktu yang lebih lama. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna ChatGPT dan membangun sistem klasifikasi sentimen berbasis web menggunakan metode Naïve Bayes. Data penelitian berupa 300 ulasan berbahasa Indonesia yang dikumpulkan melalui Google Play Scraper pada periode Agustus–September 2025. Tahapan pengolahan data meliputi preprocessing yang terdiri atas case folding, tokenizing, normalisasi, stopword removal, dan stemming, kemudian dilanjutkan dengan pembobotan TF-IDF, pembagian data latih dan data uji, serta evaluasi menggunakan confusion matrix. Hasil pengujian menunjukkan bahwa dari 10 data uji, sebanyak 7 data berhasil diklasifikasikan dengan benar dan 3 data mengalami kesalahan klasifikasi sehingga diperoleh tingkat akurasi sebesar 70%. Selain itu, kelas sentimen negatif menunjukkan performa terbaik dengan nilai recall 100%, precision 83%, dan F1-Score 91%, sedangkan kelas netral dan positif masih memiliki nilai precision dan recall yang relatif rendah. Berdasarkan hasil penelitian, metode Naïve Bayes mampu memberikan kinerja yang cukup baik dalam mengklasifikasikan sentimen ulasan pengguna dan dapat digunakan sebagai salah satu pendekatan untuk mendukung evaluasi serta peningkatan kualitas layanan aplikasi.

Kata Kunci: Analisis Sentimen, ChatGPT, Google Play Store, Naïve Bayes.

Design of a Sentiment Analysis Application for ChatGPT Reviews on Google Play Using the Naïve Bayes Method

Abstract

The rapid development of artificial intelligence technology has increased the use of the ChatGPT application on Google Play Store, resulting in numerous user reviews that reflect positive, neutral, and negative sentiments. These reviews can be utilized as a valuable source of information to identify user satisfaction levels and evaluate the quality of application services. However, the large volume of available reviews makes manual analysis less effective and time-consuming. This study aims to analyze user sentiments toward ChatGPT reviews and develop a web-based sentiment classification system using the Naïve Bayes method. The dataset consisted of 300 Indonesian-language reviews collected through Google Play Scraper during the period of August–September 2025. The data processing stages included preprocessing techniques such as case folding, tokenizing, normalization, stopword removal, and stemming. Furthermore, TF-IDF weighting, training and testing data splitting, and model evaluation using a confusion matrix were performed. The testing results showed that 7 out of 10 testing data were successfully classified correctly, while 3 data instances were misclassified, resulting in an accuracy of 70%. In addition, the negative sentiment class achieved the best performance with a recall of 100%, precision of 83%, and F1-score of 91%, while the neutral and positive classes still showed relatively lower precision and recall values. Based on the findings, the Naïve Bayes method demonstrated satisfactory performance in classifying user review sentiments and can be utilized as an approach to support application evaluation and service quality improvement.

Keywords: *sentimen analysis, ChatGPT, Google Play Store, Naïve Bayes.*

I. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) telah menghasilkan berbagai inovasi yang membantu dan memudahkan aktivitas manusia dalam berbagai bidang kehidupan. Salah satu teknologi yang saat ini banyak digunakan adalah ChatGPT yang dikembangkan oleh OpenAI. ChatGPT tersedia dalam bentuk website maupun aplikasi yang dapat diunduh melalui Google Play Store dan telah dimanfaatkan oleh jutaan pengguna dari berbagai negara. Meningkatnya jumlah pengguna turut menyebabkan bertambahnya ulasan yang diberikan terhadap aplikasi, baik yang bersifat positif, netral, maupun negatif. Ulasan tersebut dapat digunakan sebagai sumber informasi untuk mengetahui tanggapan dan tingkat kepuasan pengguna, sekaligus menjadi bahan evaluasi dalam upaya meningkatkan kualitas layanan aplikasi.

Google Play Store merupakan salah satu platform distribusi aplikasi yang menyediakan berbagai ulasan pengguna secara terbuka. Ulasan tersebut berisi pengalaman pengguna dalam menggunakan aplikasi, mulai dari tingkat kepuasan, kritik, keluhan, hingga harapan terhadap layanan yang diberikan. Banyaknya jumlah ulasan yang terus meningkat dari waktu ke waktu menyebabkan proses analisis secara manual menjadi kurang optimal, membutuhkan waktu yang lebih lama, serta berpotensi menimbulkan subjektivitas dalam penilaian. Oleh karena itu, diperlukan metode analisis otomatis yang mampu mengelompokkan sentimen ulasan secara cepat, terstruktur, dan akurat sehingga informasi yang terkandung dalam ulasan dapat dimanfaatkan secara lebih efektif [1].

Analisis sentimen merupakan salah satu teknik dalam text mining yang digunakan untuk mengetahui kecenderungan opini atau tanggapan yang terdapat pada suatu teks. Untuk melakukan proses klasifikasi sentimen, salah satu metode yang dapat digunakan adalah Naïve Bayes. Metode ini termasuk algoritma klasifikasi berbasis probabilitas yang memiliki proses perhitungan sederhana, mampu bekerja dengan cepat, dan cukup efektif dalam mengolah data teks menjadi informasi yang lebih terstruktur [2].

Beberapa penelitian terdahulu menunjukkan efektivitas metode Naïve Bayes dalam analisis sentimen. Penelitian oleh Sari et al. menunjukkan bahwa metode ini mampu memberikan akurasi yang baik dalam klasifikasi ulasan aplikasi mobile [3]. Penelitian Pratama dan Wijaya juga menyatakan bahwa penggunaan preprocessing dan pembobotan TF-IDF dapat meningkatkan performa klasifikasi [4]. Sementara itu, Rahman et al. menemukan bahwa Naïve Bayes memiliki keunggulan dalam kecepatan dan kestabilan hasil dibandingkan metode lain pada analisis ulasan pengguna [5]. Penelitian lain juga menunjukkan bahwa metode klasifikasi berbasis probabilistik seperti Naïve Bayes masih efektif dalam analisis sentimen skala besar [6] serta bahwa penerapan preprocessing dan TF-IDF dapat meningkatkan akurasi klasifikasi teks [7].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengolah dan menganalisis sentimen ulasan pengguna aplikasi ChatGPT pada Google Play Store dengan jumlah data yang besar dan beragam. Data yang digunakan sebanyak 300 ulasan berbahasa Indonesia yang diambil pada periode Agustus 2024 hingga September 2025 menggunakan

Google Play Scraper. Proses analisis meliputi tahapan preprocessing teks seperti case folding, tokenizing, normalisasi, stopword removal, dan stemming menggunakan algoritma Nazief Adriani, kemudian dilakukan pembobotan TF-IDF serta klasifikasi menggunakan metode Naïve Bayes. Implementasi sistem dilakukan berbasis web menggunakan bahasa pemrograman PHP. Hasil penelitian diharapkan dapat memberikan gambaran mengenai kecenderungan sentimen pengguna serta menjadi bahan evaluasi dalam meningkatkan kualitas layanan aplikasi ChatGPT.

II. METODOLOGI PENELITIAN

A. Pengumpulan Data

Pada penelitian ini, data diperoleh dari ulasan pengguna aplikasi ChatGPT yang tersedia di Google Play Store. Proses pengumpulan data dilakukan menggunakan teknik web scraping dengan memanfaatkan Google Play Scraper untuk mengambil data ulasan secara otomatis. Data yang dikumpulkan berupa ulasan berbahasa Indonesia yang berisi pengalaman, tanggapan, serta penilaian pengguna terhadap aplikasi ChatGPT. Dari proses tersebut diperoleh sebanyak 300 ulasan yang kemudian digunakan sebagai dataset penelitian. Selanjutnya, data ulasan tersebut diproses pada tahap preprocessing untuk membersihkan dan mempersiapkan teks sebelum dilakukan klasifikasi sentimen menggunakan metode Naïve Bayes. Penggunaan teknik web scraping dinilai efektif karena mampu mengumpulkan data dari platform digital secara otomatis, cepat, dan efisien [8].

B. Text Preprocessing

Text preprocessing merupakan tahapan awal dalam pengolahan data teks yang dilakukan sebelum proses analisis atau klasifikasi. Pada tahap ini, data yang diperoleh akan diseleksi dan diproses untuk menghasilkan bentuk data yang lebih terstruktur dan mudah diolah. Umumnya, data teks masih mengandung berbagai karakter, simbol, atau informasi yang tidak diperlukan sehingga perlu dilakukan pembersihan terlebih dahulu. Tujuan dari preprocessing adalah mengurangi noise dan meningkatkan kualitas data agar lebih siap digunakan pada tahap analisis berikutnya. Beberapa proses yang dilakukan dalam tahapan ini meliputi cleaning, case folding, tokenizing, filtering, dan stemming [9].

C. Pembobotan Kata (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan salah satu metode pembobotan kata yang banyak digunakan dalam pengolahan teks dan Natural Language Processing (NLP). Metode ini digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen dengan mempertimbangkan frekuensi kemunculannya pada dokumen tersebut serta kemunculannya pada kumpulan dokumen secara keseluruhan. Dengan menggunakan TF-IDF, kata-kata yang memiliki informasi lebih relevan akan memperoleh bobot yang lebih tinggi dibandingkan kata yang sering muncul pada banyak dokumen [10].

Tahapan dalam metode TF-IDF terdiri dari beberapa langkah, salah satunya adalah:

a) TF (Term Frequency)

Term Frequency (TF) merupakan komponen yang digunakan untuk menghitung frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai TF menunjukkan seberapa sering suatu term muncul pada dokumen yang dianalisis. Semakin tinggi frekuensi kemunculan sebuah kata, maka semakin besar pula nilai TF yang diperoleh. Perhitungan nilai TF dapat dilakukan menggunakan rumus sebagai berikut:

$$TF(t,d) = f(t,d) \tag{2.1}$$

Keterangan:

t = kata tertentu (term)

d = dokumen tertentu

TF(t,d) = nilai Term Frequency dari kata t pada dokumen d

f(t,d) = jumlah kemunculan kata t dalam dokumen d

b) IDF (Inverse Document Frequency)

Inverse Document Frequency (IDF) merupakan komponen yang digunakan untuk mengetahui tingkat keunikan suatu kata dalam kumpulan dokumen. Nilai IDF akan semakin besar jika suatu kata jarang muncul pada dokumen lain, sehingga kata tersebut dianggap memiliki informasi yang lebih penting. Perhitungan nilai IDF dilakukan menggunakan Rumus 2.2.

$$IDF(t) = \log\left(\frac{N}{DFt}\right) \tag{2.2}$$

Keterangan :

T = kata tertentu (term).

N = jumlah total dokumen dalam korpus.

DFt = jumlah dokumen yang mengandung kata t.

c) TF-IDF

Tahap terakhir adalah mengalikan TF dan IDF dari setiap kata untuk memperoleh bobot akhir. TF-IDF dihitung menggunakan rumus 2.3.

$$TF-IDF = TF \times IDF \tag{2.3}$$

Keterangan :

TF = hasil perhitungan

IDF = hasil perhitungan IDF

D. Klasifikasi Naïve Bayes

Naïve Bayes merupakan salah satu algoritma klasifikasi dalam bidang data mining dan machine learning yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan pendekatan probabilitas. Metode ini melakukan proses klasifikasi dengan menghitung kemungkinan suatu data termasuk ke dalam suatu kelas berdasarkan pola kemunculan fitur yang diperoleh dari data pelatihan. Karena memiliki proses perhitungan yang relatif sederhana, Naïve Bayes banyak digunakan dalam berbagai penelitian klasifikasi, termasuk pada analisis sentimen [11].

Rumus ini digunakan untuk menghitung probabilitas awal (prior probability) dari suatu kelas. dengan rumus 2.4.

$$P(V_j) = \frac{|does\ j_i|}{|contoh|} \tag{2.4}$$

Keterangan :

P(V_j)= Probabilitas setiap dokumen dari dokumen berkategori j

|docs_j|= Jumlah dokumen berkategori j

|contoh|= Jumlah dokumen yang digunakan sebagai contoh

Untuk probabilitas kata a_i pada kategori V_j, P(a_i | V_j).

Perhitungan dapat dilihat pada Rumus 2.5.

$$P(a_i|V_j) = \frac{nk + 1}{n + |kosakata|} \tag{2.5}$$

Keterangan :

Nk= Jumlah kemunculan kata a_i pada dokumen berkategori

V_jn= Jumlah seluruh kata pada dokumen berkategori

V_j|kosakata|= Banyaknya kata dalam data training

Nilai P(V_j) dihitung pada saat data training, Perhitungan dapat dilihat pada rumus 2.6.

$$V_{map} = v_j \in v P(V_j) \parallel P(a_i | V_j) \tag{2.6}$$

E. Confusion Matrix

Confusion matrix merupakan salah satu metode evaluasi yang digunakan untuk menilai performa model klasifikasi setelah proses pengolahan data selesai dilakukan. Penggunaan confusion matrix dilakukan dengan membandingkan hasil prediksi model terhadap data uji dengan label sebenarnya. Dari hasil perbandingan tersebut dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun yang mengalami kesalahan klasifikasi. Selain itu, confusion matrix juga digunakan untuk menghitung berbagai metrik evaluasi, seperti accuracy, precision, recall, dan F1-score berdasarkan rumus yang digunakan dalam penelitian [12].

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.7}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.8}$$

$$Presicion = \frac{TP}{TP + FP} \tag{2.9}$$

$$F1\ Score = \frac{2 \times (Presicion \times recall)}{Presicion + Recall} \tag{2.10}$$

Keterangan:

TP = nilai True Positive

FP = nilai False Positive

FN = nilai False Negative

TN = nilai True Negaitve

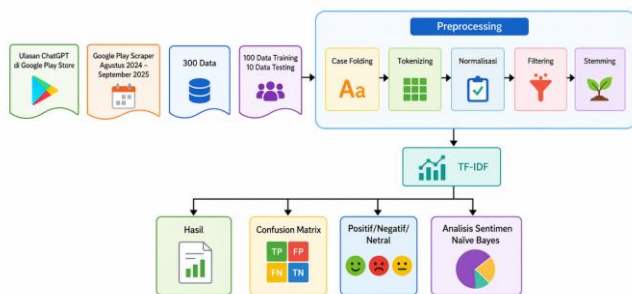
III. HASIL DAN PEMBAHASAN

Data penelitian diperoleh dari ulasan pengguna aplikasi ChatGPT di Google Play Store menggunakan Google Play Scraper. Sebanyak 300 ulasan berbahasa Indonesia berhasil dikumpulkan dan digunakan sebagai dataset penelitian. Data tersebut kemudian diproses melalui tahapan preprocessing yang meliputi case folding, tokenizing, normalisasi, stopword removal, dan stemming menggunakan

algoritma Nazief dan Adriani untuk menghasilkan data yang lebih bersih dan terstruktur.

Setelah proses preprocessing selesai dilakukan, data teks diubah menjadi representasi numerik menggunakan metode TF-IDF. Pembobotan TF-IDF digunakan untuk menentukan tingkat kepentingan setiap kata dalam dokumen. Nilai bobot yang dihasilkan selanjutnya digunakan pada proses training untuk membangun model klasifikasi Naïve Bayes. Model yang telah terbentuk kemudian diuji menggunakan data testing untuk mengelompokkan sentimen ulasan ke dalam tiga kategori, yaitu positif, netral, dan negatif.

Kinerja model dievaluasi menggunakan confusion matrix dengan menghitung nilai accuracy, precision, recall, dan F1-score. Hasil evaluasi tersebut digunakan untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen ulasan pengguna secara keseluruhan. Pada bagian ini juga disajikan hasil penelitian beserta pembahasannya secara komprehensif. Penyajian hasil dilakukan dalam bentuk gambar, grafik, maupun tabel agar informasi yang diperoleh dapat dipahami dengan lebih mudah oleh pembaca. Diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur

A. Data Ulasan

Pengumpulan data dalam penelitian ini dilakukan menggunakan metode web scraping terhadap ulasan pengguna aplikasi ChatGPT pada Google Play Store dengan bantuan Google Play Scraper. Data yang diperoleh disimpan dalam bentuk file .csv dan berjumlah 300 ulasan berbahasa Indonesia, yang kemudian digunakan sebagai dataset dalam penelitian analisis sentimen.

Pada tabel berikut ditampilkan beberapa sampel data ulasan pengguna yang digunakan dalam penelitian ini. Dataset yang diperoleh berjumlah 300 ulasan. Namun, pada tahap perhitungan manual menggunakan Microsoft Excel digunakan 100 data Training dan 10 data testing sebagai sampel. Tabel ini hanya menampilkan contoh ulasan mewakili masing-masing label sentimen yaitu positif, netral, dan negatif dapat dilihat pada Tabel 1.

Tabel 1. Data Ulasan

Dok	Ulasan	Label
D1	jelek banget!! Bagus sih tapi ngeselin banget aku uda nungguin satu jam lebih ga di	Negatif

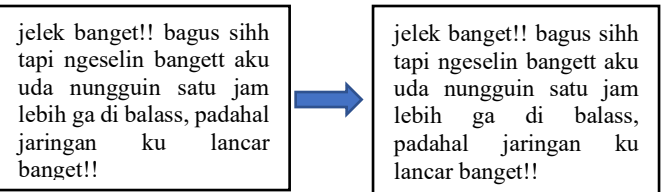
	balass, padahal jaringan ku lancar banget!!	
D2	menurut aku bagus tapi kadang suka lupa mesal nya di suruh ngetik kaya yg tadi tapi malah ngetik yg beda bikin kesel asli	Netral
D3	Ini sangat membantu sekali dalam menuangkan ide-ide di pikiran saya untuk menulis sebuah karya sastra.	Positif
-----	-----	-----
D100	sangat enak menggunakan aplikasi ini saat kesusahan mengerjakan pr tugas dan lain lain enak belajar di chat gpt	Positif
D101	Respon ChatGPT sering kali tidak sesuai dengan memori yang sudah di tanamkan oleh pengguna. Tidak berguna!	?

B. Text Preprocessing

Tahapan preprocessing ini meliputi beberapa tahap yaitu case folding, tokenizing, normalisasi, filtering dan stemming.

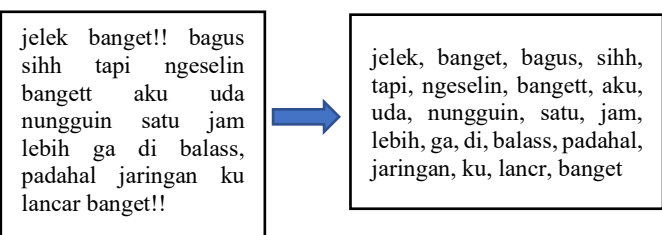
a) Case folding

Case Folding merupakan proses yang dilakukan untuk mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase). Pada tahap ini, karakter yang tidak diperlukan seperti tanda baca, simbol, dan angka juga dihapus sehingga hanya menyisakan teks yang akan digunakan dalam proses analisis. Contoh hasil proses case folding dapat dilihat sebagai berikut.



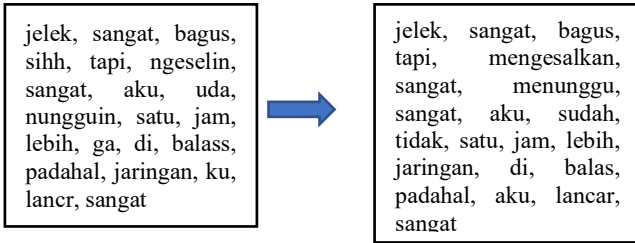
b) Tokenizing

Tokenizing merupakan proses pemisahan teks menjadi bagian-bagian yang lebih kecil berupa kata atau token. Setiap kata yang dihasilkan akan digunakan sebagai unit data untuk diproses pada tahapan berikutnya. Contoh hasil dari proses tokenizing dapat dilihat sebagai berikut.



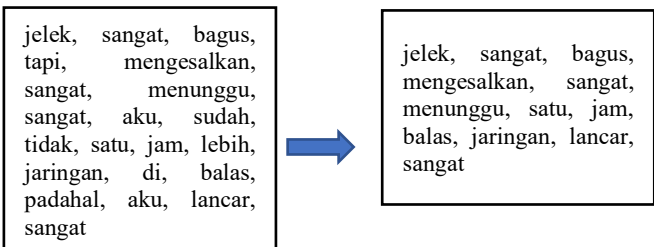
c) Normalisasi

Normalisasi merupakan proses perbaikan kata yang bertujuan untuk mengubah kata tidak baku atau kata yang mengalami kesalahan penulisan menjadi bentuk kata baku yang sesuai. Hasil dari proses normalisasi kemudian digunakan pada tahap pengolahan data berikutnya.



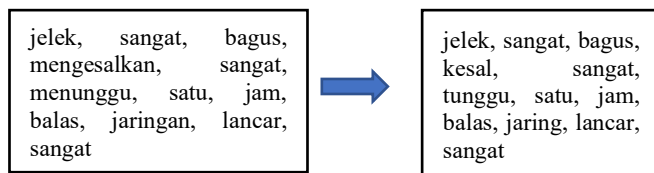
d) Filtering

Filtering merupakan proses penyaringan kata dengan mempertahankan kata-kata yang dianggap penting dan menghapus kata-kata yang kurang memiliki makna dalam proses analisis. Pada penelitian ini, proses filtering dilakukan menggunakan library NLTK. Contoh hasil dari proses filtering dapat dilihat sebagai berikut.



e) Stemming

Stemming merupakan proses mengubah kata yang memiliki imbuhan menjadi bentuk kata dasar. Pada tahap ini, token-token hasil proses sebelumnya juga digabungkan kembali menjadi sebuah kalimat. Proses stemming dilakukan menggunakan library Sastrawi yang menerapkan algoritma Nazief dan Adriani. Contoh hasil dari proses stemming dapat dilihat sebagai berikut.



C. Pembobotan Kata (TF-IDF)

Tabel 1 menampilkan sebagian contoh data ulasan yang digunakan dalam penelitian ini. Setelah melalui tahapan preprocessing, data ulasan selanjutnya diproses menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk menghasilkan bobot pada setiap kata. Pembobotan TF-IDF dilakukan dengan mempertimbangkan frekuensi kemunculan kata dalam dokumen serta tingkat relevansinya terhadap keseluruhan dokumen ulasan. Nilai bobot yang diperoleh kemudian digunakan sebagai fitur pada proses klasifikasi sentimen menggunakan metode Naïve Bayes.

Tabel 2. Hasil Perhitungan TF-IDF

kata	TF	DF	IDF	TF-IDF
jelek	1	2	1,69897	1,69897
sangat	3	19	0,721246	5,09691
bagus	1	35	0,455932	1,69897
kesal	1	6	1,221849	1,69897
----	----	----	----	----
hati	0	3	2,00000	0

Pada Tabel 2 dapat dilihat untuk bobot setiap kata. Untuk mengetahui bobot kata “jelek” pertama kita harus mengetahui nilai TF dan IDF terlebih dahulu. Kata “jelek” muncul sekali pada dokumen 1 maka nilai TF nya bernilai 1. Kata “mantap” muncul pada 2 dokumen (DF). Perhitungan nilai IDF untuk kata “jelek”: menggunakan rumus 2.2

$$IDF = \log (100 / 2) = 1,69897$$

Bobot kata “jelek” dengan menggunakan rumus 2.3

$$TF-IDF = 1 \times 1,6990 = 1,6990$$

Selanjutnya menghitung nilai kelas berdasarkan jumlah frekuensi dari perhitungan TF-IDF. Untuk jumlah frekuensi dari hasil perhitungan TF-IDF berdasarkan kelas dapat dilihat pada tabel 3.

Tabel 3. Jumlah Frekuensi

kata	Negatif	Netral	positif
jelek	3,39794	0	0
banget	23,78558	10,19382	5,09691
bagus	11,89279	30,5814601	22,08661
kesal	3,39794	6,79588002	0
----	----	----	----
hati	0	0	1,69897

D. Perhitungan Naïve Bayes

Setelah dilakukan pembobotan kata menggunakan metode TF-IDF, langkah selanjutnya adalah proses klasifikasi sentimen menggunakan algoritma Naïve Bayes pada data Testing. Metode ini digunakan untuk menghitung probabilitas suatu dokumen termasuk ke dalam kelas sentimen tertentu, yaitu positif, netral, dan negatif. Nilai probabilitas setiap kelas dihitung berdasarkan kemunculan kata pada dokumen yang telah melalui tahap preprocessing dan pembobotan TF-IDF. Hasil perhitungan data testing dapat dilihat pada tabel 4.

Tabel 4. Hasil Perhitungan Data Testing

DO K	Vmap			Label	
	Negatif	Netral	Positif	PREDIK SI	METO DE
D101	7,34589E-17	7,03077E-17	1,07114E-17	Negatif	Negatif
D102	3,70521E-16	2,2886E-17	9,29426E-17	Negatif	Negatif
D103	6,97692E-16	4,26981E-16	6,34894E-18	Negatif	Negatif

D104	4,0889E-31	3,53545E-32	8,90391E-36	Negatif	Negatif
D105	1,0675E-27	1,479E-25	7,05148E-26	Netral	Netral
D106	1,68634E-19	2,72471E-21	5,25986E-22	Negatif	Negatif
D107	3,59768E-25	2,6773E-24	3,26957E-24	Positif	Positif
D108	5,465E-14	5,66065E-13	8,72062E-14	Positif	Netral
D109	1,31036E-20	4,57538E-22	4,63218E-21	Positif	Negatif
D110	1,31449E-20	4,22415E-19	2,17223E-20	Positif	Netral

E. Confusion Matrix

Berdasarkan Tabel 3.5 terdapat 7 data yang berhasil diklasifikasikan dengan benar, yaitu 5 data negatif, dan 1 data positif. Sedangkan 3 data lainnya mengalami kesalahan klasifikasi, yang seluruhnya berasal dari kelas positif dapat dilihat pada Tabel 5.

Tabel 5. Confusion Matrix

Prediksi	TP	FN	FP	TN
Negatif	5	0	0	5
Netral	0	1	0	1
Positif	1	2	1	4
Total	6	3	1	10

Berdasarkan tabel evaluasi, kelas Negatif memiliki performa terbaik dengan recall 100%, precision 83%, dan F1-Score 91% sehingga model sangat baik dalam mendeteksi data negatif. Kelas Netral memiliki recall 100% namun precision rendah sebesar 33% dengan F1-Score 50%, yang menunjukkan masih terdapat kesalahan prediksi. Sementara itu, kelas positif memiliki precision 100% tetapi recall hanya 25% dengan F1-Score 40%, sehingga model masih kurang optimal dalam mendeteksi data positif dapat dilihat pada Tabel 6.

Tabel 6. Hasil Persentase Confusion Matrix

Kelas	Recall	Precision	F1-Score
Negatif	100%	83%	91%
Netral	100%	33%	50%
Positif	25%	100%	40%

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode Naïve Bayes dapat diterapkan dengan baik untuk mengklasifikasikan sentimen ulasan pengguna aplikasi ChatGPT ke dalam kategori positif, netral, dan negatif. Hasil penelitian menunjukkan bahwa metode ini mampu mengidentifikasi kecenderungan sentimen pengguna secara cukup efektif berdasarkan data ulasan yang dianalisis. Dari 300 ulasan berbahasa Indonesia yang berhasil dikumpulkan, sebanyak 100 data digunakan sebagai data pelatihan dan 10 data digunakan sebagai data pengujian yang

hitung secara manual menggunakan Microsoft Excel, sedangkan 190 data lainnya digunakan dalam proses pengolahan dan pengujian melalui sistem yang dikembangkan. Hasil pengujian menunjukkan bahwa model memperoleh tingkat akurasi sebesar 70% dengan nilai rata-rata AUC sebesar 0,8046 yang menunjukkan bahwa performa klasifikasi model berada pada kategori baik. Dengan hasil tersebut, metode Naïve Bayes dapat dimanfaatkan untuk membantu memahami persepsi serta tingkat kepuasan pengguna terhadap aplikasi ChatGPT secara lebih objektif. Selain itu, hasil analisis sentimen yang diperoleh juga dapat digunakan sebagai bahan evaluasi dalam upaya pengembangan dan peningkatan kualitas layanan aplikasi di masa mendatang.

REFERENSI

- [1] Jatmiko, S., & Dometian, C. (2025). Analisis Sentimen Ulasan Google Play Store: Studi Komparatif Algoritma SVM, Naïve Bayes, dan Logistic Regression. *Jurnal Fasilkom*, 15(3).
- [2] J. Brownlee, "Naive Bayes for Machine Learning," *Machine Learning Mastery*, 2023.
- [3] A. Sari, D. Putri, dan B. Nugroho, "Sentiment analysis of mobile application reviews using Naïve Bayes method," *Journal of Information Systems and Technology*, vol. 12, no. 2, pp. 101–110, 2024.
- [4] R. Pratama dan A. Wijaya, "Implementation of TF-IDF and Naïve Bayes for sentiment analysis on user reviews," *Indonesian Journal of Computer Science*, vol. 15, no. 1, pp. 45–53, 2024.
- [5] F. Rahman, T. Hidayat, dan M. Saputra, "Comparative analysis of Naïve Bayes and other classification methods in user review sentiment analysis," *International Journal of Data Science and Applications*, vol. 9, no. 1, pp. 25–34, 2025.
- [6] A. Kumar, S. Singh, dan R. Patel, "Sentiment analysis using machine learning techniques: A comparative study," *International Journal of Data Science*, vol. 8, no. 2, pp. 45–55, 2024.
- [7] M. Zhang, L. Wang, dan Y. Chen, "Improving text classification accuracy using TF-IDF and preprocessing techniques," *Journal of Artificial Intelligence Research*, vol. 12, no. 1, pp. 78–89, 2025.
- [8] S. K. Sharma dan R. Gupta, "Web scraping techniques for data extraction and analysis: A review," *International Journal of Computer Applications*, vol. 185, no. 12, pp. 15–22, 2023.
- [9] A. Kumar dan S. Sharma, "Text preprocessing techniques in natural language processing: A review," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 500–507, 2023.
- [10] Y. Li, X. Zhang, dan H. Liu, "TF-IDF based text representation for natural language processing: A review," *Journal of Information Processing Systems*, vol. 20, no. 1, pp. 45–55, 2024.
- [11] A. Tharwat, "Naïve Bayes classifier: A review," *Journal of Artificial Intelligence Research*, vol. 68, pp. 1–24, 2023.
- [12] Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 27(4S). <https://doi.org/10.53555/AJBR.v27i4S.434>