

## PENINGKATAN MODEL KLASIFIKASI SENTIMEN PENGGUNA APLIKASI TOMORO COFFEE MENGGUNAKAN ALGORITMA NAÏVE BAYES

\*Dina Audina<sup>1</sup>, Ade Irma Purnamasari<sup>2</sup>, Agus Bahtiar<sup>3</sup>, Edi Tohidi<sup>4</sup>

<sup>1,2</sup>Teknik Informatika, STMIK IKMI Cirebon

<sup>3</sup>Sistem Informasi, STMIK IKMI Cirebon, <sup>4</sup>Komputerisasi Akuntansi, STMIK IKMI Cirebon

Jl. Perjuangan No. 10B, Kota Cirebon, Provinsi Jawa Barat

<sup>1</sup>[audinaaa410@gmail.com](mailto:audinaaa410@gmail.com), <sup>2</sup>[irma2974@mail.com](mailto:irma2974@mail.com), <sup>3</sup>[agusbahtiar038@mail.com](mailto:agusbahtiar038@mail.com),

<sup>4</sup>[editohidi.ikmi@gmail.com](mailto:editohidi.ikmi@gmail.com)

### Abstract

Advances in information technology have revolutionized the way businesses interact with customers through mobile applications, including in the food and beverage sector. Tomoro Coffee application faces challenges in maintaining user satisfaction due to limited features and technical issues. This study aims to apply the Naïve Bayes algorithm to improve the sentiment classification model of user reviews, analyze the distribution of positive and negative sentiments and the main factors that influence them, and evaluate the performance of the model based on accuracy, precision, recall, and F1-score. Review data was collected from Google Play Store and processed using the Knowledge Discovery in Database (KDD) method, which includes data cleaning, tokenization, stopword removal, stemming, and feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). The results showed that the Naïve Bayes algorithm achieved 90% accuracy, with 91.3% precision, 87.3% recall, and 88.7% F1-score. The findings provide strategic insights for app developers in improving services and features based on user sentiment analysis. From the analysis results, 64.4% of reviews were positive, dominated by comments such as "the coffee is delicious", while 35.6% of negative reviews generally contained technical complaints, such as "not available".

Keywords : naïve bayes, sentiment analysis, TF-IDF, tomoro coffee, user reviews

### Abstrak

Kemajuan teknologi informasi telah merevolusi cara bisnis berinteraksi dengan pelanggan melalui aplikasi mobile, termasuk dalam sektor makanan dan minuman. Aplikasi Tomoro Coffee menghadapi tantangan dalam mempertahankan kepuasan pengguna akibat keterbatasan fitur dan masalah teknis. Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes guna meningkatkan model klasifikasi sentimen ulasan pengguna, menganalisis distribusi sentimen positif dan negatif beserta faktor utama yang memengaruhinya, serta mengevaluasi performa model berdasarkan akurasi, presisi, recall, dan F1-score. Data ulasan dikumpulkan dari Google Play Store dan diolah menggunakan metode Knowledge Discovery in Database (KDD), yang mencakup pembersihan data, tokenisasi, penghapusan stopword, stemming, serta ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi sebesar 90%, dengan presisi 91,3%, recall 87,3%, dan F1-score 88,7%. Temuan ini memberikan wawasan strategis bagi pengembang aplikasi dalam meningkatkan layanan dan fitur berdasarkan analisis sentimen pengguna. Dari hasil analisis, 64,4% ulasan tergolong positif, didominasi oleh komentar seperti "kopinya enak", sementara 35,6% ulasan negatif umumnya berisi keluhan teknis, seperti "tidak tersedia".

Kata kunci : Analisis sentiment, Naïve Bayes, TF-IDF, Tomoro Coffee, Ulasan pengguna

## 1. PENDAHULUAN

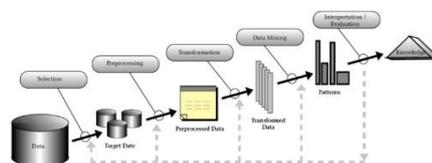
Perkembangan teknologi digital telah mengubah cara konsumen berinteraksi dengan bisnis, termasuk dalam industri *food and beverage* (F&B). Ulasan pengguna melalui aplikasi menjadi sumber informasi penting bagi pelaku bisnis untuk memahami kepuasan pelanggan dan meningkatkan kualitas layanan. *Tomoro Coffee* salah satu *coffee shop* dengan aplikasi pemesanan sendiri, menghadapi tantangan dalam menganalisis ulasan pelanggan secara efisien dan akurat. Sebagaimana dijelaskan oleh [1] analisis sentimen dapat membantu memahami ulasan pengguna untuk meningkatkan kualitas layanan melalui pengolahan data ulasan secara lebih terstruktur dan akurat.

Analisis sentimen dengan *machine learning* telah diterapkan secara luas untuk mengklasifikasikan opini publik. Algoritma *Naïve Bayes* menjadi pilihan yang populer karena kesederhanaannya dalam menangani klasifikasi teks. Penelitian sebelumnya menunjukkan bahwa algoritma ini mampu mencapai akurasi tinggi dalam menganalisis opini publik. Sebagai contoh, penelitian oleh Tavares dan Nurraharjo [2] membuktikan bahwa metode *Naïve Bayes Classifier* mampu mencapai akurasi sebesar 90% dalam mengklasifikasikan sentimen terkait peningkatan kasus *Omicron* berdasarkan data dari Twitter. Namun, penelitian khusus pada aplikasi F&B seperti *Tomoro Coffee* masih terbatas. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Naïve Bayes* guna meningkatkan model klasifikasi sentimen ulasan pengguna aplikasi *Tomoro Coffee*. Selain itu, penelitian ini mengevaluasi distribusi sentimen positif dan negatif serta kinerja model

menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Inovasi penelitian ini terletak pada integrasi algoritma *Naïve Bayes* dengan teknik *preprocessing* berbasis *Natural Language Toolkit* (NLTK) dan *Sastrawi* untuk memberikan wawasan yang lebih mendalam terhadap preferensi pengguna.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD) untuk menganalisis sentimen ulasan pengguna aplikasi *Tomoro Coffee*. Metode ini terdiri dari tahapan yang sistematis, mulai dari pengumpulan data hingga interpretasi hasil analisis, dengan implementasi algoritma *Naïve Bayes* untuk membangun model klasifikasi sentimen. Metode KDD dipilih karena mampu menangani pengolahan data dalam jumlah besar secara efisien, khususnya untuk aplikasi berbasis teks.



Gambar 1. *Knowledge Discovery in Database*

### 2.1. Pengumpulan Dataset

Penelitian ini menggunakan data sekunder berupa ulasan pengguna aplikasi *Tomoro Coffee* yang diambil dari *Google Play Store*. Data dikumpulkan dengan teknik *web scraping* menggunakan skrip *Python*. Periode data yang diambil adalah Januari 2023 hingga November 2024, dengan fokus pada satu tahun terakhir untuk relevansi analisis. Hasil jumlah data ulasan yang didapatkan adalah sebanyak 671.

## 2.2. Pre-Processing

Tahapan *preprocessing* dalam analisis sentimen memiliki peran esensial untuk memastikan bahwa data ulasan yang diolah bersih dan siap untuk dianalisis. Penelitian oleh [3] menunjukkan bahwa *preprocessing* dapat meningkatkan kualitas analisis sentimen dengan membersihkan teks ulasan dari elemen yang dapat mengganggu akurasi model. Hal serupa juga ditekankan oleh [4] yang menyoroti pentingnya proses ini dalam menyiapkan data yang akurat untuk klasifikasi sentimen menggunakan algoritma *Naïve Bayes*.

Berikut adalah tahapan-tahapan *preprocessing* yang dilakukan:

### a. Labelling Data

Setelah melalui proses pemfilteran data, setiap skor ulasan dikategorikan menjadi label sentimen "Positif" atau "Negatif." Proses pelabelan ini bertujuan untuk mempermudah klasifikasi sentimen dengan cara menetapkan label "Negatif" untuk skor di bawah 3 dan "Positif" untuk skor 4 dan 5. Sebagaimana disebutkan oleh [5] pelabelan adalah tahap penting dalam analisis sentimen yang menentukan akurasi model dalam mengenali pola sentimen.

### b. Data Cleaning

Proses data *cleaning* ini melibatkan penghapusan elemen-elemen yang tidak relevan, seperti simbol, tanda baca, angka, dan karakter khusus, yang dapat mengganggu hasil analisis. Penelitian oleh [6] menunjukkan bahwa pembersihan data secara menyeluruh dapat mengurangi *noise*, sehingga

meningkatkan akurasi model analisis sentimen. Hal ini diperkuat oleh [7] yang mengungkapkan bahwa data yang bersih berperan besar dalam meningkatkan kinerja algoritma *Naïve Bayes*, terutama dalam klasifikasi sentimen.

### c. Stopword

Setelah tahap pembersihan data, langkah selanjutnya adalah menghilangkan *stopword*, yaitu kata-kata umum seperti "dan," "di," atau "yang" yang tidak memiliki kontribusi signifikan dalam analisis sentimen. Penghapusan *stopword* bertujuan untuk menyederhanakan data serta memungkinkan algoritma lebih fokus pada kata-kata yang lebih relevan terhadap sentimen yang dianalisis. Proses penghapusan *stopword* dilakukan dengan memanfaatkan pustaka *Natural Language Toolkit* (NLTK) dalam *Python*, yang menyediakan daftar *stopword* bahasa Indonesia.

### d. Tokenizing

Setelah tahap penghapusan *stopword* selesai, langkah selanjutnya adalah proses *tokenizing*, yaitu memecah teks menjadi unit-unit kecil seperti kata, frasa, atau karakter sesuai dengan kebutuhan analisis. Menurut [8] *tokenizing* berfungsi untuk menyederhanakan data teks sehingga lebih mudah diproses, memungkinkan model untuk mendeteksi pola atau sentimen dengan lebih baik.

e. *Stemming*

Tahap akhir adalah proses *stemming*, yaitu transformasi kata menjadi bentuk dasarnya untuk mengurangi variasi kata dengan makna serupa. Memanfaatkan pustaka seperti *Sastrawi*, proses ini menyederhanakan data teks sehingga lebih relevan untuk dianalisis [9].

### 2.3. Transformation

*Transformation* adalah tahap penting dalam pemrosesan teks, di mana data teks mentah diubah menjadi representasi numerik yang dapat digunakan oleh algoritma pembelajaran mesin. Dua metode utama yang digunakan adalah *Bag of Words (BoW)* dan *TF-IDF (Term Frequency-Inverse Document Frequency)*. Dalam metode BoW, setiap kata dalam dokumen diberikan nilai berdasarkan frekuensi kemunculannya, yang diimplementasikan menggunakan *CountVectorizer* dari pustaka *scikit-learn*. Untuk memperhitungkan bobot kata berdasarkan kelangkaannya di seluruh dokumen, data BoW kemudian diubah menjadi bobot TF-IDF menggunakan *TfidfTransformer*. Transformasi ini memungkinkan analisis data teks menjadi lebih akurat, terutama dalam tugas seperti analisis sentimen dan klasifikasi teks [10]

### 2.4. Data Mining

Pada tahap *Data Mining*, dataset ulasan aplikasi *Tomoro Coffee* yang telah diproses melalui pembersihan teks, tokenisasi, penghapusan *stopword*, dan *stemming* dibagi menjadi dua bagian: data latih dan data uji. Pembagian ini dilakukan dengan rasio 80:20, 70:30, dan 60:40, untuk menemukan komposisi terbaik dalam melatih model. Model yang digunakan adalah algoritma Naïve Bayes, yang dikenal unggul dalam analisis teks karena

efisiensi dan akurasi tinggi. Penelitian oleh Zakasih dan Handoko [11] mengindikasikan bahwa metode Naïve Bayes Classifier mampu mencapai tingkat akurasi sebesar 92,33% dalam klasifikasi opini publik terkait NFT di Twitter dengan nilai precision sebesar 92,12%, recall sebesar 94%, dan F1-score sebesar 92,45%. Metode ini mengandalkan probabilitas dan frekuensi kemunculan kata dalam masing-masing kategori sentimen (positif atau negatif) untuk memprediksi sentimen dari ulasan baru [5]. Langkah ini selaras dengan penelitian lain yang memanfaatkan pembagian dataset sebagai bagian penting dalam melatih model dan mengukur performanya [12].

### 2.5. Evaluasi

Tahap evaluasi dilakukan dengan menguji model *Naïve Bayes* pada data uji dan mengukur kinerjanya menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score* [13]. Untuk memperjelas hasil evaluasi, digunakan visualisasi seperti *confusion matrix* yang menggambarkan jumlah prediksi benar dan salah untuk setiap kelas sentimen, serta diagram yang menunjukkan persentase sentimen positif dan negatif dari dataset [14]. Selain itu, visualisasi dalam bentuk wordcloud juga dibuat untuk menampilkan kata-kata yang paling sering muncul dalam setiap kategori sentimen, memberikan wawasan lebih lanjut tentang pola kata yang mendominasi ulasan. [15].

## 3. HASIL DAN PEMBAHASAN

Pengumpulan data dilakukan melalui teknik *scraping* menggunakan bahasa pemrograman *Python*. Dari proses *scraping* ini, diperoleh

sebanyak 671 baris data. Hasil pengumpulan data melalui *scraping* tersebut dapat dilihat pada tabel 1.

TABEL I. PENGUMPULAN DATA

Content	Score
Itu koffr latte nya,, tu kopi apa jamu pahit bos nggak ada efek manis nya	1
Mantap kode voucher nya!	5

### 3.1 Pelabelan Data

Setelah pengumpulan data, tahap awal dalam analisis data adalah pelabelan sentimen berdasarkan skor ulasan. Pelabelan dilakukan dengan memanfaatkan skrip *Python* yang menetapkan ulasan dengan skor lebih dari dengan 3 sebagai sentimen positif, sedangkan ulasan dengan skor di bawah 3 sebagai sentimen negatif. Hasil dari proses pelabelan dapat dilihat pada tabel 2.

TABEL II. PELABELAN DATA

Content	Score	Label
Itu koffr latte nya,, tu kopi apa jamu pahit bos nggak ada efek manis nya	1	Negatif

Mantap kode voucher nya! 5 Positif

### 3.2 Data Cleaning

#### a. Case Folding

Setelah dilakukan proses pelabelan langkah selanjutnya adalah pembersihan data. Hasil akhir yang telah melalui tahap *cleaning* dapat dilihat pada tabel 3.

TABEL III. CLEANING DATA

Content	Text Clean
Itu koffr latte nya,, tu kopi apa jamu pahit bos nggak ada efek manis nya	itu koffr latte nya tu kopi apa jamu pahit bos nggak ada efek manis nya
Mantap kode voucher nya!	lumayan enak nya!

#### b. Stopword Removal

Proses berikutnya adalah *stopword removal*, yaitu tahap menghilangkan kata-kata yang dianggap tidak berkontribusi dalam analisis, seperti "di," "dan," atau "yang". Hasil proses *stopword* dapat dilihat pada tabel 1.

TABEL IV. STOPWORD REMOVAL

Content	Text Stopword
Itu koffr latte nya,, tu kopi apa jamu pahit bos nggak ada efek manis nya	koffr latte nyatu kopi jamu pahit bos nggak efek manis nya

Mantap kode voucher    mantap    kode  
 nya!                            voucher nya

c. *Tokenizing*

Setelah proses penghapusan stopword, langkah berikutnya adalah tokenisasi, yaitu membagi kalimat menjadi unit-unit kecil seperti kata. Hasil dari proses *Tokenizing* dapat dilihat pada tabel 2.

TABELV. *TOKENIZING*

Content	Text Tokens
Itu koffr latte nya,, tu kopi apa jamu pahit nggak ada efek manis nya	[koffr, latte, nyatu, kopi, jamu, pahit, bos, nggak, efek, manis, nya]
Mantap kode voucher nya!	[mantap, kode, voucher, nya]

d. *Stemming*

Stemming merupakan langkah terakhir dalam proses prapemrosesan data, yang bertujuan mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan atau akhiran. Proses stemming menggunakan pustaka Sastrawi dalam penerapannya, yang dirancang khusus untuk mendukung pemrosesan teks dalam bahasa Indonesia. Hasil akhir dari seluruh tahapan prapemrosesan dapat dilihat pada tabel 3.

. TABEL VI. *STEMMING*

Content	Text Stemming
Itu koffr latte nya,, tu kopi apa jamu pahit bos nggak ada efek manis nya	koffr latte nyatu kopi jamu pahit bos nggak efek manis nya
Mantap kode voucher nya!	mantap kode voucher nya

3.3 *Data Transformation*

Tahap transformasi merupakan proses mengubah atau memanipulasi data ke dalam format yang lebih sesuai dan berguna untuk analisis atau pemodelan. Tujuan utama dari transformasi data adalah meningkatkan kualitas dan relevansi data, mengatasi masalah ketidakseimbangan, serta mempersiapkan data agar lebih siap untuk digunakan dalam analisis statistik atau pemodelan pembelajaran mesin.

3.4 *Split Data*

Pada Pembagian data ini dilakukan dengan berbagai rasio, seperti 90:10, 80:20, 70:30, dan 60:40, untuk menemukan komposisi terbaik dalam melatih model. Model yang digunakan adalah algoritma *Naïve Bayes*, yang dikenal unggul dalam analisis teks karena efisiensi dan akurasi tinggi

3.5 *Data Mining*

Penelitian ini menggunakan metodologi *data mining* dengan fokus pada *naïve bayes* untuk menganalisis sentimen. Algoritma *naïve bayes* mengidentifikasi sentimen berdasarkan berbagai elemen tekstual, termasuk pemilihan kata dan struktur kalimat positif dan negatif. Algoritma *naïve bayes* juga menghitung kemungkinan klasifikasi sentimen untuk setiap ulasan yang

dianalisis. Implementasi teknis dimulai dengan mengambil komponen *MultinomialNB* dari *scikit-learn*. Sebuah instance dari kelas ini dibuat untuk membangun model *Naïve Bayes* yang akan digunakan dalam analisis sentimen. Selanjutnya, model *Naïve Bayes* dikonfigurasi dengan membuat *objek* dari kelas tersebut, kemudian dilatih menggunakan fungsi *fit*, di mana input berupa data yang telah ditransformasi ke format TF-IDF (*X\_train\_tfidf*) beserta label sentimennya (*y\_train*). Proses pelatihan ini memungkinkan model memahami pola dan hubungan antara fitur-fitur dalam data dengan label yang diberikan. Setelah proses pelatihan selesai, model dapat digunakan untuk memprediksi sentimen pada data uji.

### 3.6 Evaluation

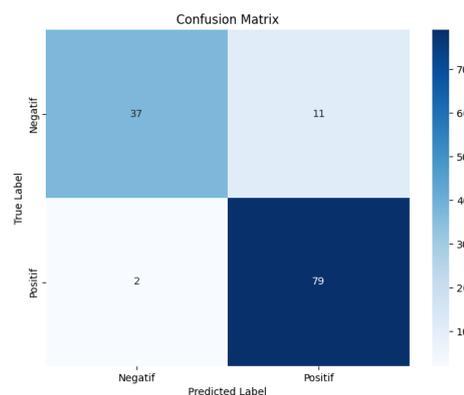
Evaluasi dalam analisis sentimen bertujuan untuk menilai seberapa baik model atau sistem mampu mengenali dan memahami sentimen dari teks. Evaluasi model umumnya dilakukan dengan metrik seperti akurasi, presisi, recall, dan F1-score sebagai indikator utama kinerjanya. Sebelum diterapkan pada data uji, model terlebih dahulu dilatih menggunakan data latih berlabel. Proses pelatihan ini mencakup ekstraksi fitur, pembelajaran pola melalui algoritma seperti *Naïve Bayes*, serta pengujian performa untuk memastikan model dapat mengenali pola sentimen dengan baik pada data latih. Setelah tahap ini selesai, model dapat digunakan untuk memprediksi data baru dengan tingkat akurasi yang lebih optimal. Hasil evaluasi menunjukkan bahwa proporsi terbaik antara data latih dan uji adalah 80:20, dengan akurasi akhir yang dapat dilihat pada Gambar 2.

```
MultinomialNB Accuracy: 0.8992248862015504
MultinomialNB Precision: 0.9487179487179487
MultinomialNB Recall: 0.7708333333333334
MultinomialNB f1_score: 0.8505747126436781
confusion_matrix:
[[37 11]
 [ 2 79]]
```

	precision	recall	f1-score	support
Negatif	0.95	0.77	0.85	48
Positif	0.88	0.98	0.92	81
accuracy			0.90	129
macro avg	0.91	0.87	0.89	129
weighted avg	0.90	0.90	0.90	129

Gambar 3. Hasil Akurasi

Selain itu, *confusion matrix* memberikan representasi visual yang menggambarkan hasil prediksi, baik yang benar maupun yang salah, untuk setiap kelas sentimen. Evaluasi ini menjadi panduan penting dalam pengembangan model, sehingga dapat meningkatkan akurasi prediksi pada teks yang dianalisis. Hasil visual *confusion matrix* dapat dilihat pada Gambar 4.

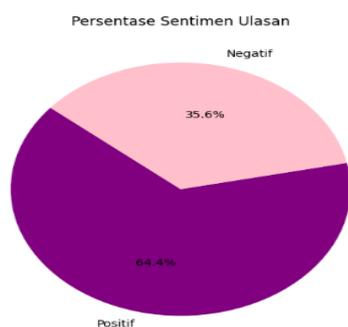


Gambar 5. Confusion Matrix

### 3.7 Visualisasi

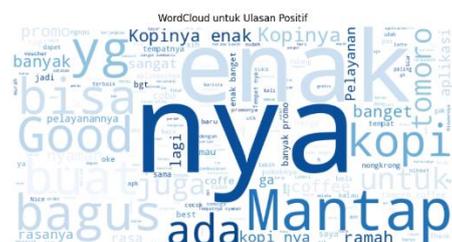
Diagram *pie chart* yang menggambarkan persentase sentimen menghasilkan 64,4% sentimen positif, menunjukkan bahwa sebagian besar pengguna merasa puas dengan aplikasi tersebut. Sementara itu, 35,6% ulasan mengandung sentimen negatif, yang meskipun jumlahnya lebih kecil, mencerminkan keluhan dan kritik pengguna. Diagram ini memberikan gambaran yang jelas tentang distribusi sentimen

positif dan negatif, yang dapat digunakan untuk lebih memahami respons pengguna terhadap aplikasi *Tomoro Coffee*. Melalui penerapan algoritma *Naïve Bayes*, data ini dapat dianalisis lebih lanjut untuk mengidentifikasi pola dan meningkatkan akurasi prediksi sentimen secara keseluruhan. Hasil visualisasi diagram sentimen dapat dilihat pada gambar 6



Gambar 7. Diagram *Pie Chart* Sentimen

Selain itu, visualisasi *WordCloud* mengungkap perbedaan yang jelas antara sentimen positif dan negatif dalam ulasan pengguna aplikasi *Tomoro Coffee*. *WordCloud* untuk sentimen positif menampilkan kata-kata yang sering muncul dalam ulasan yang memberikan apresiasi terhadap aplikasi, memberikan wawasan tentang aspek-aspek yang paling dihargai oleh pengguna. Ukuran setiap kata mencerminkan frekuensi kemunculannya, menyoroti elemen-elemen yang mendapat perhatian positif. Hasil visualisasi *wordcloud* sentimen positif dapat dilihat pada gambar 8.



Gambar 9. *Wordcloud* Sentimen Positif

Sebaliknya, sentimen negatif *wordcloud* menunjukkan kata-kata yang sering muncul dalam ulasan yang mengungkapkan ketidakpuasan, memberikan wawasan tentang aspek-aspek yang perlu ditingkatkan atau disempurnakan untuk meningkatkan pengalaman pengguna. Hasil visualisasi *wordcloud* sentimen positif dapat dilihat pada gambar 10.



Gambar 11. *Wordcloud* Sentimen Negatif

#### 4. Kesimpulan dan Saran

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan algoritma *Naive Bayes* untuk klasifikasi sentimen ulasan pengguna pada aplikasi *Tomoro Coffee* memberikan hasil yang sangat memuaskan, dengan tingkat akurasi terbaik mencapai 90% pada pembagian data latih dan data uji sebesar 80:20. Keberhasilan ini tidak terlepas dari pentingnya penerapan teknik *preprocessing* dan ekstraksi fitur menggunakan TF-IDF, yang secara signifikan membantu meningkatkan performa model. Hasil analisis sentimen juga mengungkapkan bahwa 64,4% ulasan pengguna bersifat positif, sedangkan 35,6% bersifat negatif. Hal ini mengindikasikan bahwa mayoritas pengguna merasa puas dengan pengalaman mereka menggunakan aplikasi *Tomoro Coffee*. Selain itu, evaluasi model melalui metrik seperti akurasi, presisi, *recall*, dan *F1-score* menunjukkan performa yang sangat baik.

Model berhasil mencapai rata-rata akurasi 90%, dengan presisi 91,3%, recall 87,3%, dan F1-score 88,7%. Secara keseluruhan, model *naïve bayes* menunjukkan kinerja yang lebih unggul dalam mendeteksi sentimen positif dibandingkan sentimen negatif.

Bagi peneliti selanjutnya disarankan untuk meningkatkan akurasi model dengan menambah jumlah data latih, sehingga model dapat mengenali lebih banyak pola, terutama dalam ulasan dengan sentimen negatif. Selain itu, eksplorasi metode ekstraksi fitur seperti *Word2Vec* atau *GloVe* dapat membantu model memahami konteks kata secara lebih mendalam. Jika dataset yang digunakan tidak seimbang, teknik seperti SMOTE dapat diterapkan untuk mengurangi bias terhadap sentimen mayoritas. Penelitian mendatang juga dapat membandingkan performa algoritma *Naïve Bayes* dengan metode lain, seperti *Super Vector Machine SVM* atau *Random Forest*, guna menentukan algoritma yang paling optimal.

#### Daftar Pustaka:

- [1] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1329, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [2] A. L. D. tavares Duarte and Eddy Nurraharjo, "Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 1–8, 2023, doi: 10.36595/jire.v6i1.779.
- [3] D. Abror, "Analisis Sentimen Review Aplikasi PeduliLindungi Menggunakan Seleksi Fitur Information Gain Berbasis SVM," *Indones. J. Softw. Eng.*, vol. 9, no. 1, pp. 1–8, 2023, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse>
- [4] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [5] S. N. Salsabila, B. N. Sari, and R. Mayasari, "Klasifikasi Ulasan Pengguna Aplikasi Discord Menggunakan Metode Information Gain Dan Naïve Bayes Classifier," *INFOTECH J.*, vol. 9, no. 2, pp. 383–392, 2023, doi: 10.31949/infotech.v9i2.6277.
- [6] M. K. Insan, U. Hayati, and O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI," vol. 7, no. 1, pp. 478–483, 2023.
- [7] Meliyawati and F. N. Hasan, "Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naïve Bayes," *Media Online*, vol. 4, no. 4, pp. 2272–2280, 2024, doi: 10.30865/klik.v4i4.1555.
- [8] E. H. Muktafin and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [9] S. A. Helmayanti, F. Hamami, and R. Y. Fa'rifah, "Penerapan Algoritma Tf-Idf Dan Naïve Bayes Untuk Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Flip Pada Google Play Store," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1822–1834, 2023, doi: 10.35870/jimik.v4i3.415.
- [10] A. Nurian, B. N. Sari, U. S. Karawang, and T. Timur, "Analisis sentimen ulasan pengguna aplikasi google play menggunakan naïve bayes," vol. 11, no. 3, pp. 829–835, 2023.
- [11] M. Iqbal Zakasih and W. Tri Handoko, "Analisis Sentimen Pengguna Twitter Tentang Nft (Non Fungible Token) Dengan Metode Naive Bayes Classifier," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 221–229, 2022, doi: 10.36595/jire.v5i2.694.
- [12] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi:

- 10.57152/malcom.v4i1.983.
- [13] S. A. Saputra, Didi Rosiyadi, Windu Gata, and Syepry Maulana Husain, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2021, doi: 10.29207/resti.v3i3.1118.
- [14] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, "Jurnal Computer Science and Information Technology ( CoSciTech ) samsat jawa tengah Performance analysis of the naïve bayes algorithm on the sentiment dataset of central java samsat NEWSAKPOLE applications," vol. 3, no. 3, 2022.
- [15] A. I. Tanggraeni, M. N.N, and Sitokdana, "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes," vol. 9, no. 2, pp. 785–795, 2022.