JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i1.8776

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI FLO MENGGUNAKAN ALGORITMA NAIVE BAYES

Eti Kurniawati[1], Ade Irma Purnamasari[2], Irfan Ali[3], Rudi Kurniawan[4] , Odi Nurdiawan[5]

[1,3,4]Rekasaya Perangkat Lunak, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135
[2]Teknik Informatika, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135
[5]Sistem Informasi, STMIK IKMI CIREBON, Jl. Perjuangan No.10B Cirebon, 45135

Keywords:
sentiment analysis; Flo application; Naive Bayes; text mining; mHealth.

Corespondent Email:
etikurniawati722@gmail.com

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Flo pada Google Play Store menggunakan algoritma Multinomial Naive Bayes. Flo merupakan aplikasi mobile health (mHealth) populer yang digunakan untuk memantau siklus menstruasi dan kesehatan reproduksi. Data dikumpulkan melalui web scraping dan menghasilkan 10.000 ulasan yang setelah pembersihan menjadi 6.908 data valid. Proses pra-pemrosesan meliputi case folding, cleaning, normalisasi, tokenisasi, stopword removal, dan stemming menggunakan Sastrawi. Pelabelan sentimen dilakukan secara semi-otomatis berbasis lexicon InSet dan rating. Ekstraksi fitur menggunakan CountVectorizer menghasilkan representasi Bag-of-Words sebagai input model. Hasil evaluasi menunjukkan bahwa algoritma Naive Bayes mencapai akurasi sebesar 73,6% dengan nilai precision, recall, dan F1-score yang seimbang pada tiga kelas sentimen. Temuan ini menunjukkan bahwa Naive Bayes efektif digunakan dalam mengolah ulasan teks pendek dan informal berbahasa Indonesia. Penelitian ini berkontribusi dalam pemanfaatan machine learning untuk analisis sentimen aplikasi mHealth serta menyediakan wawasan yang dapat digunakan pengembang untuk meningkatkan kualitas layanan aplikasi Flo.

Abstract. *This study aims to analyze user reviews of the Flo application on Google Play Store using the Multinomial Naive Bayes algorithm. Flo is a popular mobile health (mHealth) application for tracking menstrual cycles and reproductive health. Data were collected using web scraping, obtaining 10,000 initial reviews, with 6,908 valid reviews after cleaning. Preprocessing included case folding, cleaning, normalization, tokenization, stopword removal, and stemming using Sastrawi. Sentiment labeling was performed semi-automatically using the InSet lexicon and rating-based rules. Feature extraction used CountVectorizer with the Bag-of-Words approach. The evaluation shows that Naive Bayes achieved an accuracy of 73.6% with balanced precision, recall, and F1-score across sentiment classes. These results indicate that Naive Bayes is effective for processing short and informal Indonesian text* reviews. This research contributes to the application of machine learning in mHealth sentiment analysis and provides insights for developers to improve the quality of the Flo application.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong pertumbuhan pesat aplikasi berbasis mobile, khususnya pada sektor kesehatan atau mobile health (mHealth). Aplikasi mHealth digunakan untuk mendukung pemantauan kesehatan, edukasi medis, serta pengelolaan kondisi kesehatan secara mandiri oleh pengguna [1] ,[2]. Salah satu aplikasi mHealth yang populer di kalangan perempuan adalah Flo – Ovulation & Period Tracker, yang

berfungsi untuk memantau siklus menstruasi dan kesehatan reproduksi.

Seiring dengan meningkatnya jumlah pengguna, aplikasi Flo menerima ribuan ulasan di Google Play Store. Ulasan pengguna mencerminkan tingkat kepuasan, keluhan, dan kebutuhan pengguna terhadap aplikasi [3] Namun, jumlah ulasan yang besar menyebabkan kesulitan dalam menganalisis opini pengguna secara manual. Oleh karena itu, diperlukan pendekatan otomatis untuk mengolah dan mengklasifikasikan opini pengguna.

Analisis sentimen merupakan salah satu teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi polaritas opini dalam teks, seperti sentimen positif, negatif, dan netral [4],[5]. Metode ini telah banyak diterapkan pada ulasan aplikasi untuk mengevaluasi kualitas layanan dan pengalaman pengguna [6], [7].

Berbagai algoritma machine learning telah digunakan dalam analisis sentimen, salah satunya adalah Naive Bayes Classifier. Algoritma ini dikenal sederhana, efisien, dan efektif dalam klasifikasi teks [8]. Penelitian [9] menunjukkan bahwa Naive Bayes mampu memberikan performa yang baik dalam analisis sentimen berbahasa Indonesia pada media sosial Twitter.

Meskipun telah banyak penelitian terkait analisis sentimen ulasan aplikasi, kajian khusus terhadap aplikasi kesehatan reproduksi seperti Flo masih terbatas, terutama menggunakan data ulasan berbahasa Indonesia. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Flo menggunakan algoritma Naive Bayes. Hasil penelitian diharapkan dapat memberikan gambaran persepsi pengguna serta menjadi bahan.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Ulasan pengguna telah diakui sebagai sumber data penting dalam memahami persepsi terhadap aplikasi digital. Penelitian oleh Fadhil dan Schueller menunjukkan bahwa ulasan aplikasi mHealth mengandung ekspresi emosional yang beragam sehingga relevan untuk dianalisis menggunakan teknik analisis sentimen. Analisis skala besar terhadap ulasan aplikasi kesehatan juga menunjukkan bahwa umpan balik pengguna dapat digunakan untuk mengidentifikasi kebutuhan dan permasalahan aplikasi [10].

Dalam konteks bahasa Indonesia, preprocessing teks memegang peranan penting. [1] serta [11] menekankan bahwa normalisasi dan stemming mampu meningkatkan performa model analisis sentimen. Beberapa penelitian menunjukkan bahwa Multinomial Naive Bayes efektif dan efisien untuk analisis sentimen teks pendek [6], [2], [7].

Pendekatan lexicon-based juga masih relevan untuk pelabelan otomatis dataset, terutama ketika data berlabel manual terbatas. Selain itu, pemilihan representasi fitur seperti Bag-of-Words terbukti berpengaruh signifikan terhadap performa model klasifikasi teks [15].

Penelitian [12] yang dipublikasikan pada JITET menunjukkan bahwa Naive Bayes mampu memberikan hasil kompetitif pada analisis sentimen berbahasa Indonesia di media sosial, sehingga mendukung penggunaan algoritma ini pada penelitian serupa.

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks ke dalam kategori tertentu, seperti positif, negatif, dan netral [13] Teknik ini banyak digunakan untuk mengevaluasi persepsi publik terhadap produk, layanan, dan aplikasi digital.

2.2.2 Text Mining dan NLP

Text mining bertujuan mengubah teks tidak terstruktur menjadi informasi yang bermakna melalui tahapan preprocessing dan ekstraksi fitur. NLP digunakan untuk mengubah teks

menjadi representasi numerik agar dapat diproses oleh algoritma machine learning [14].

2.2.3 Multinomial Naive Bayes

Multinomial Naive Bayes merupakan algoritma probabilistik berbasis Teorema Bayes yang menghitung peluang suatu dokumen termasuk ke dalam kelas tertentu berdasarkan frekuensi kata [15] Algoritma ini cocok untuk teks pendek dan bekerja optimal dengan fitur Bag-of-Words.

2.2.4 Evaluasi Model

Evaluasi model klasifikasi dilakukan menggunakan accuracy, precision, recall, dan F1-score. Confusion matrix digunakan untuk menganalisis kesalahan klasifikasi pada tiap kelas sentimen.

2.3 Kesenjangan Penelitian

Penelitian terdahulu belum secara spesifik membahas analisis sentimen ulasan aplikasi Flo pada konteks pengguna Indonesia. Selain itu, evaluasi performa per kelas sentimen, khususnya sentimen netral, masih jarang dibahas secara mendalam. Penelitian ini mengisi kesenjangan tersebut dengan analisis komprehensif serta mengaitkan hasil dengan pengembangan aplikasi.

2.4 Kerangka Berpikir

Penelitian dimulai dari pengumpulan data ulasan aplikasi Flo, dilanjutkan dengan pra-pemrosesan teks, pelabelan sentimen, ekstraksi fitur, pelatihan model Multinomial Naive Bayes, dan evaluasi performa model. Seluruh tahapan dirancang untuk menghasilkan model analisis sentimen yang akurat dan aplikatif.

3. METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis machine learning, bertujuan mengklasifikasikan ulasan pengguna aplikasi Flo Period Tracker pada Google Play Store ke dalam tiga kategori sentimen: positif, netral, dan negatif. Pendekatan ini dipilih karena sesuai untuk menganalisis data teks berukuran besar dan memungkinkan proses perhitungan yang objektif serta terukur (Swaminathan & al., 2024)

Desain penelitian disusun agar *reproducible*, mengikuti tahapan terstruktur yang terdiri dari: (1) pengumpulan data ulasan, (2) pra-pemrosesan teks, (3) pelabelan sentimen, (4) ekstraksi fitur, (5) pelatihan model Multinomial Naive Bayes, dan (6) evaluasi performa menggunakan metrik standar machine learning.

3.2 Arsitektur dan Tahapan Penelitian

Tahapan penelitian disusun untuk menjamin hasil yang sistematis dan dapat diulang kembali. Arsitektur proses penelitian dapat dijelaskan sebagai berikut:

(1) Pengumpulan Data → (2) Pra-pemrosesan Teks → (3) Pelabelan Sentimen → (4) Ekstraksi Fitur → (5) Pembagian Data → (6) Pelatihan Model Naive Bayes → (7) Evaluasi Model

Secara umum, alur penelitian mengadopsi pipeline analisis teks modern yang digunakan pada penelitian machine learning kontemporer.

3.2.1 Pengumpulan Data

Data penelitian berasal dari ulasan pengguna aplikasi Flo Period Tracker pada Google Play Store. Pengumpulan data dilakukan dengan metode web scraping menggunakan library *google-play-scraper* pada Python.

Data yang dikumpulkan meliputi:

1. ID Ulasan
2. Nama Pengguna
3. Rating
4. Teks Ulasan
5. Tanggal Ulasan

Seluruh data diekstraksi dalam format CSV agar dapat diolah pada tahap selanjutnya. Metode pengumpulan ini memastikan data bersumber dari umpan balik pengguna aktual (*user-generated content*) sehingga representatif untuk dianalisis.

3.2.2 Pra-Pemrosesan Data (Preprocessing)

Pra-pemrosesan dilakukan untuk membersihkan dan menormalkan teks sebelum diterapkan ke model. Tahapan meliputi:

1. Case Folding: seluruh teks diubah menjadi huruf kecil.
2. Cleaning: menghapus angka, emotikon, URL, tanda baca, dan simbol non-alfabet.
3. Normalisasi Teks: mengganti kata tidak baku menggunakan kamus normalisasi.
4. Tokenisasi: memecah teks menjadi satuan kata (token).
5. Stopword Removal: menghapus kata umum yang tidak membawa makna penting.
6. Stemming (Sastrawi)**:** mengubah kata menjadi bentuk dasar.

Tahapan ini diperlukan untuk mengurangi noise sehingga fitur yang terbentuk lebih konsisten dan relevan.

3.2.3 Metode Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan pendekatan semi-otomatis yang memadukan:

1. Lexicon-based Sentiment Analysis**,**
2. Rating-based Labeling, dan
3. Validasi manual pada sebagian data.

Label yang ditetapkan adalah:

Label sentimen ditentukan dengan kriteria sebagai berikut:

1. Positif: ulasan dengan dominasi kata bermakna positif dan rating tinggi,
2. Netral: ulasan bersifat deskriptif tanpa ekspresi emosional yang jelas,
3. Negatif: ulasan dengan dominasi kata bermakna keluhan dan rating rendah.

Pendekatan hybrid ini dipilih untuk meminimalkan bias dan meningkatkan konsistensi label.

3.2.4 Ekstraksi Fitur (Feature Extraction)

Ekstraksi fitur dilakukan menggunakan teknik Bag-of-Words (BoW) dengan *CountVectorizer*. Tahapannya meliputi :

1. Membentuk *vocabulary* dari seluruh kata unik.
2. Menghitung frekuensi kemunculan kata pada setiap dokumen.
3. Membentuk *document-term matrix* sebagai input model klasifikasi.

Pendekatan ini dinilai tepat untuk teks pendek seperti ulasan aplikasi mobile.

3.2.5 Pembagian Dataset

Dataset dibagi menjadi dua bagian:

1. 80% untuk pelatihan (training),
2. 20% untuk pengujian (testing).

Pembagian dilakukan secara acak menggunakan fungsi *train_test_split* agar memenuhi kaidah *generalization* dalam machine learning.

3.2.6 Model Multinomial Naive Bayes

Algoritma yang digunakan adalah Multinomial Naive Bayes, dengan tahapan:

1. Melatih model menggunakan data training.
2. Melakukan prediksi sentimen pada data testing.

3. Menghasilkan probabilitas setiap kelas sentimen.
4. Membandingkan prediksi dengan label aktual.

Algoritma ini dipilih karena efektif untuk teks pendek, cepat, dan tidak memerlukan komputasi tinggi.

Model Naive Bayes menggunakan rumus dasar Teorema Bayes sebagai berikut:

$$P(C|X) = \frac{P(X|C)\ X\ P(C)}{P(X)}$$

Keterangan:

1. *P(C|X)* = probabilitas suatu kelas *C* diberikan fitur X
2. *P(X|C)* = probabilitas fitur *X* muncul pada kelas *C*
3. *P(C)P* = probabilitas prior dari kelas
4. *P(X)* = probabilitas keseluruhan fitur

3.2.7 Evaluasi Model

Evaluasi performa model dilakukan menggunakan metrik:

1. Accuracy – tingkat ketepatan prediksi keseluruhan,
2. Precision – ketepatan prediksi per kelas,
3. Recall – kemampuan model menemukan data aktual per kelas,
4. F1-Score – keseimbangan precision dan recall.

Selain itu digunakan confusion matrix untuk memvisualisasikan distribusi prediksi benar dan salah pada setiap kelas sentimen. Metrik ini dipilih karena memberikan gambaran menyeluruh tentang kinerja model, khususnya pada dataset yang berpotensi tidak seimbang.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data ulasan pengguna aplikasi Flo – Ovulation & Period Tracker diperoleh melalui proses web scraping pada platform Google Play Store menggunakan library *google-play-scraper*. Proses pengambilan data dilakukan dengan filter bahasa Indonesia dan wilayah Indonesia sehingga data yang diperoleh relevan dengan konteks pengguna lokal.

Sebanyak 10.000 ulasan awal berhasil dikumpulkan. Selanjutnya dilakukan proses seleksi data berupa penghapusan ulasan duplikat, ulasan kosong, serta ulasan non-bahasa Indonesia. Setelah proses pembersihan tersebut, diperoleh 6.908 ulasan valid yang digunakan sebagai dataset penelitian.

Dataset yang digunakan memiliki atribut utama berupa Review ID, Username, Rating, Review Text, dan Date. Struktur data yang konsisten ini memudahkan proses pra-pemrosesan dan analisis sentimen berbasis machine learning, serta memastikan bahwa data yang digunakan benar-benar merepresentasikan pengalaman pengguna aplikasi flo

4.2 Pra-Pemrosesan Data

Sebelum digunakan sebagai input pemodelan, teks ulasan dibersihkan melalui tahapan *preprocessing* yang meliputi:

1. Cleaning – menghapus URL, HTML tag, emoji, angka, tanda baca, dan simbol yang berpotensi menjadi noise.
2. Case Folding – mengubah seluruh karakter menjadi huruf kecil guna menyeragamkan format penulisan.
3. Normalisasi – mengganti kata tidak baku menjadi kata baku menggunakan kamus baku–tidak baku (misal: “bgt” → “banget”).
4. Tokenisasi – memecah teks menjadi token/kata.
5. Stopword Removal – menghapus kata umum seperti “yang”, “dan”, “di”, “atau”.
6. Stemming – mengubah kata menjadi bentuk dasarnya menggunakan Sastrawi (misal: “membantu” → “bantu”).

No	Review Text (Asli)	Cleaning	Case Folding	Normalisasi	Tokenize	Stopword Removal	Steming Data
1	*sangat bagus*	sangat bagus	sangat bagus	sangat bagus	['sangat','bagus']	['bagus']	bagus
2	*sangat membantu*	sangat membantu	sangat membantu	sangat membantu	['sangat','membantu']	['membantu']	bantu
3	*Sangat membantu*	Sangat membantu	sangat membantu	sangat membantu	['sangat','membantu']	['membantu']	bantu
4	*suka bngt samaa flo, selalu deket sama hari ha…*	suka bngt samaa flo selalu deket sama hari hai	suka bngt samaa flo selalu deket sama hari hai	suka banget sama flo selalu dekat sama hari haid	['suka','banget','sama','flo','selalu','dekat','sama','haid']	['suka','banget','flo','dekat','haid']	suka banget flo dekat haid

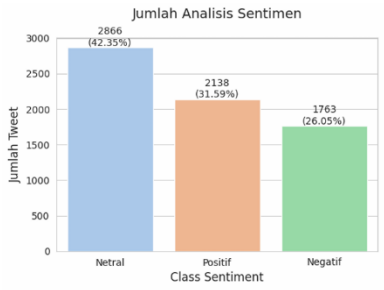
Tabel 4. 1 Hasil Preprocessing

Pada tabel 4.1 Hasil *preprocessing* menunjukkan bahwa teks ulasan menjadi lebih bersih, terstruktur, dan siap diproses pada tahap ekstraksi fitur. Word cloud yang dihasilkan memperlihatkan penurunan drastis terhadap kata-kata tidak baku dan noise, serta menonjolkan kata bermakna seperti “bagus”, “bantu”, dan “akurat”.

4.3 Labeling Data

Pelabelan dilakukan menggunakan Indonesian Sentiment Lexicon (InSet) dengan pendekatan *lexicon-based*. Penentuan sentimen dilakukan berdasarkan dominasi skor kata positif dan negatif pada setiap ulasan:

1. Positif → skor positif > negatif
2. Negatif → skor negatif > positif
3. Netral → skor seimbang atau tidak mengandung kata berpolaritas.



Gambar 4. 1 Distribusi Sentimen

Distribusi sentimen menunjukkan bahwa sentimen netral mendominasi, disusul sentimen positif dan negatif. Hal ini sejalan dengan karakter ulasan aplikasi yang sering berupa deskripsi penggunaan dan informasi umum tanpa ekspresi emosional.

4.4 Ekstraksi Fitur

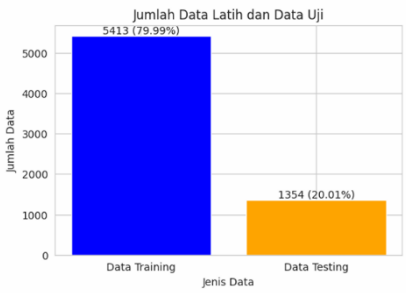
Fitur teks diubah menjadi representasi numerik menggunakan CountVectorizer (Bag-of-Words). Proses ini menghasilkan :

1. 3.891 fitur unik dari seluruh dataset,
2. Matriks dokumen kata yang menjadi input bagi model Naive Bayes.

Pembagian dataset dilakukan menggunakan metode stratified train-test split (80%–20%), sehingga distribusi sentimen tetap terjaga pada data latih dan data uji.

4.5 Pembagian Dataset

Tahap pembagian dataset dilakukan untuk memisahkan data menjadi dua bagian, yaitu data latih (training set) dan data uji (testing set) yang digunakan dalam proses pelatihan dan evaluasi model. Pembagian ini bertujuan memastikan bahwa model tidak hanya belajar dari data yang tersedia, tetapi juga diuji menggunakan data yang belum pernah dilihat sebelumnya sehingga performanya dapat dinilai secara objektif.



Gambar 4. 2 Pembagian Dataset

Gambar 4.1 pembagian dataset pada penelitian ini digunakan rasio pembagian 80% data latih dan 20% data uji. Rasio ini merupakan praktik umum pada penelitian machine learning karena dianggap mampu memberikan keseimbangan antara kebutuhan pembelajaran model dan evaluasi generalisasi. Selain itu, pembagian

dilakukan menggunakan pendekatan stratified sampling agar proporsi sentimen positif, netral, dan negatif tetap seimbang pada kedua subset. Teknik ini penting untuk mencegah bias model terhadap kelas tertentu, terutama karena distribusi ulasan cenderung tidak merata antar kategori sentimen.

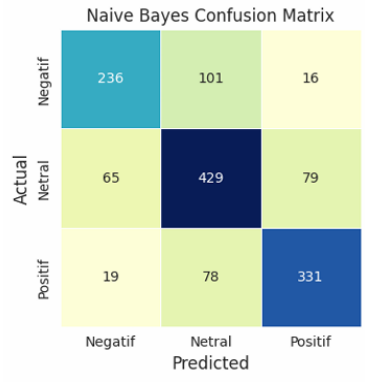
Proses pembagian dataset dilakukan menggunakan fungsi *train_test_split* dari *Scikit-learn*. Hasil pembagian menunjukkan bahwa data latih terdiri dari sejumlah besar ulasan yang memungkinkan model untuk mempelajari pola kata secara efektif, sementara data uji menyediakan sampel yang representatif untuk mengukur kemampuan prediksi model terhadap data baru.

Pembagian dataset selanjutnya divisualisasikan dalam bentuk diagram batang untuk memperlihatkan komposisi data secara lebih jelas. Visualisasi ini menunjukkan bahwa proporsi data telah sesuai dengan rasio yang ditetapkan, serta memberikan gambaran mengenai volume data yang digunakan pada setiap tahap pemodelan.

Secara keseluruhan, proses pembagian dataset memberikan dasar penting bagi keberhasilan proses pelatihan model Multinomial Naive Bayes, karena memastikan model memperoleh data yang cukup untuk belajar sekaligus tetap memberikan evaluasi yang akurat dan tidak bias.

4.6 Penerapan Algoritma Multinomial Naive Bayes

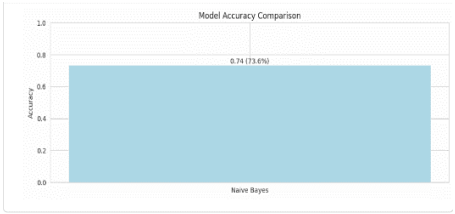
Model Multinomial Naive Bayes dilatih menggunakan data latih dan diuji menggunakan data uji. Evaluasi dilakukan menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score.



Gambar 4. 3 confusion Matrix

Hasil *confusion matrix* menunjukkan:

1. Kelas positif sangat mudah dikenali oleh model.
2. Kelas negatif cukup stabil karena mengandung kata keluhan yang jelas.
3. Kelas netral paling sering mengalami misklasifikasi, karena cenderung ambigu.



Gambar 4. 4 Hasil akurasi model

Akurasi model tercatat sebesar 73,6%**,** dengan nilai precision, recall, dan F1-score yang relatif seimbang pada tiap kelas.

4.7 Evaluasi Model

Hasil evaluasi menunjukkan:

1. Accuracy : 73,6%
2. F1-score tertinggi diperoleh pada sentimen negatif (0.775)
3. Sentimen positif memiliki F1-score 0.701
4. Sentimen netral memiliki recall tertinggi namun precision relatif lebih rendah.

Tabel 4. 2 Hasil Evaluasi

Kelas	Precision	Recall	F1-score	Support
Positif	0.738	0.669	0.701	353.000
Netral	0.706	0.749	0.727	573.000
Negatif	0.777	0.773	0.775	428.000
Accuracy	0.736	0.736	0.736	0.736
Macro Avg	0.740	0.730	0.734	1354.000
Weighted Avg	0.736	0.736	0.735	1354.000

Model menunjukkan performa yang stabil pada tiga kelas sentimen, namun tantangan terbesar ada pada sentimen netral yang secara linguistik lebih sulit diprediksi.

5. KESIMPULAN

Penelitian ini menghasilkan beberapa kesimpulan utama sebagai berikut:

1. Pengumpulan dan Pra-pemrosesan Data
   Teknik web scraping berhasil mengumpulkan 10.000 ulasan, yang kemudian diproses menjadi 6.908 ulasan valid melalui *preprocessing* komprehensif. Tahapan pembersihan, normalisasi, tokenisasi, stopword removal, dan stemming terbukti meningkatkan kualitas teks secara signifikan.
2. Efektivitas Multinomial Naive Bayes
   Algoritma Naive Bayes berhasil mengklasifikasikan sentimen ulasan pengguna dengan akurasi 73,6%, dengan nilai precision, recall, dan F1-score yang seimbang. Kelas positif dan negatif dapat dikenali dengan baik, sementara kelas netral masih menjadi tantangan.
3. Temuan Sentimen Pengguna
   Ulasan pengguna aplikasi Flo didominasi sentimen positif dan netral, mencerminkan apresiasi terhadap fitur aplikasi sekaligus adanya keluhan teknis. Word cloud dan distribusi sentimen memperlihatkan pola kata yang kuat terkait persepsi dan pengalaman pengguna.
4. Kelebihan dan Kekurangan Model
   Kelebihan model adalah efisiensi, interpretabilitas, dan performa tinggi pada teks pendek. Kekurangan utama adalah kesulitan membaca sentimen netral yang bersifat ambigu.
5. Arah Pengembangan Selanjutnya
   Penelitian lanjutan dapat menggunakan TF-IDF, bigram, model LSTM, atau BERT untuk meningkatkan akurasi terutama pada sentimen netral serta memperkaya analisis ke arah topic modeling dan dashboard analitik.

DAFTAR PUSTAKA

[1] Agastya, I. M. (2022). The Influence of the Indonesian Sastrawi Stemmer on Sentiment Analysis. *Indonesian NLP Conference Proceedings*.

[2] Ashbaugh, L., & Zhang, Y. (2024). A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning. *Computers*. https://doi.org/10.3390/computers13120340

[3] Ataei, P., Regula, S., Staegemann, D., & Malgaonkar, S. (2024). Filtering Useful App Reviews Using Naïve Bayes—Which Naïve Bayes? *AI (Switzerland)*, *5*(4), 2237–2259. https://doi.org/10.3390/ai5040110

[4] Ayash, L., Khan, H., & Mustafa, R. (2025). Advancements in Feature Selection and Extraction Methods in Text Mining: A Unified Conceptual Framework. *SN Applied Sciences*. https://doi.org/10.1007/s42452-025-07587-w

[5] Çaylak, P. Ç., Kayakuş, M., Eksili, N., Yiğit Açikgöz, F., Coşkun, A. E., Ichimov, M. A. M., & Moiceanu, G. (2024). Analysing Online Reviews Consumers' Experiences of Mobile Travel Applications with Sentiment Analysis and Topic Modelling: The Example of Booking and Expedia. *Applied Sciences (Switzerland)*, *14*(24). https://doi.org/10.3390/app142411800

[6] Dewi, C., Chen, R. C., Christanto, H. J., & Cauteruccio, F. (2023). Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database. *Vietnam Journal of Computer Science*, *10*(4), 485–498. https://doi.org/10.1142/S2196888823500100

[7] Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, *10*(1). https://doi.org/10.23960/jitet.v10i1.2262

[8] El Majid, A. U., & Nuari, R. (2025). Performance Comparison of BERT Metrics and Classical Machine Learning Models for Sentiment Analysis. *INOVTEK Polbeng – Seri Informatika*. https://doi.org/10.35314/wmh3rg23

[9] Fadhil, M., & Schueller, S. M. (2022). mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2022.857304

[10] Ghatora, P. S., Hosseini, S. E., Pervez, S., Iqbal, M. J., & Shaukat, N. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. *Big Data and Cognitive Computing*, *8*(12). https://doi.org/10.3390/bdcc8120199

[11] Haggag, O., Grundy, J., Abdelrazek, M., & Haggag, S. (2022). A large scale analysis of mHealth app user reviews. *Empirical Software Engineering*, *27*(7). https://doi.org/10.1007/s10664-022-10222-6

[12] Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

[13] Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review. *Journal of King Saud University – Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2024.102048

[14] Mercha, E. M., & Benbrahim, H. (2023). Machine Learning and Deep Learning for Sentiment Analysis Across Languages: A Survey. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2023.02.015

[15] Mihany, F. A., Galal-Edeen, G. H., Hassanein, E. E., & Moussa, H. (2025). Data-Driven Requirements Elicitation from App Reviews Framework Based on BERT. *Applied Sciences (Switzerland)*, *15*(17). https://doi.org/10.3390/app15179709

[16] Swaminathan, S., & al., et. (2024). Confusion Matrix–Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, *27*(4S). https://doi.org/10.53555/AJBR.v27i4S.4345