**JDMSI**, Vol. 7, No. 1, 2026, 11-19, ISSN: 2748-8485

# Analisis Sentimen Ulasan Pengguna Tiktok Menggunakan Ekstraksi Fitur Tf-Idf Dan Naive Bayes

**Rohur Rizqi Renata[1], Muhammad Rizqi Maulana[2], Heru Saputro[3]**

[1,2,3]Sistem Informasi, Universitas Islam Nahdlatul Ulama Jepara
*[1,2,3]Jl. Taman Siswa (Pekeng), Tahunan, Jepara, Jawa Tengah*
E-Mail: [1]renopakis@gmail.com, [2]mrizqim506@gmail.com, [3]herusaputro@unisnu.ac.id

**Abstract**

*This study classifies sentiment in TikTok user reviews on the Google Play Store using TF-IDF feature extraction combined with the Naive Bayes algorithm, implemented in Orange Data Mining version 3.38.1. The dataset comprises 100,000 reviews collected between August and December 2025, labeled into three sentiment classes: Positive (ratings 4–5), Neutral (rating 3), and Negative (ratings 1–2). To address class imbalance, balanced sampling was applied, selecting 3,334 samples per class for a total of 10,002 instances. Preprocessing included lowercasing, word tokenization, and Indonesian stopword removal, followed by Bag-of-Words representation weighted using sublinear TF. Model evaluation via 10-fold stratified cross-validation yielded an accuracy of 0.530, AUC of 0.760, F1-score of 0.509, precision of 0.537, recall of 0.530, and MCC of 0.310. Class-wise analysis showed the Positive class achieved the highest recall (84.43%), while the Negative class obtained the highest precision (58.53%). The Neutral class remained the most difficult to classify due to semantic ambiguity. The AUC of 0.760 confirms satisfactory discriminative capability despite moderate overall accuracy, reflecting the inherent challenges of three-class sentiment classification in informal Indonesian text.*

**Keyword:** *Naive Bayes, Orange Data Mining, Google Play Store, Sentiment, TF-IDF, TikTok)*

**Abstrak**

Penelitian ini mengklasifikasikan sentimen ulasan pengguna TikTok di Google Play Store menggunakan kombinasi TF-IDF dan algoritma Naive Bayes melalui Orange Data Mining versi 3.38.1. Dataset terdiri dari 100.000 ulasan yang dikumpulkan antara Agustus–Desember 2025, dengan tiga kelas sentimen: Positif (rating 4–5), Netral (rating 3), dan Negatif (rating 1–2). Untuk mengatasi ketidakseimbangan kelas, diterapkan balanced sampling dengan 3.334 sampel per kelas, menghasilkan total 10.002 instance. Tahap prapemrosesan meliputi lowercasing, tokenisasi kata, dan penghapusan stopword Bahasa Indonesia, dilanjutkan dengan representasi Bag-of-Words yang diberi bobot sublinear TF. Evaluasi model menggunakan 10-fold stratified cross-validation menghasilkan akurasi 0,530, AUC 0,760, F1-score 0,509, presisi 0,537, recall 0,530, dan MCC 0,310. Analisis per kelas menunjukkan bahwa kelas Positif memiliki recall tertinggi (84,43%), sedangkan kelas Negatif memperoleh presisi tertinggi (58,53%). Kelas Netral terbukti paling sulit diklasifikasikan akibat ambiguitas semantik. Nilai AUC sebesar 0,760 menegaskan bahwa model memiliki kemampuan diskriminatif yang cukup baik meskipun akurasi keseluruhan tergolong moderat, mencerminkan tantangan inheren klasifikasi sentimen tiga kelas pada teks Bahasa Indonesia yang informal.

**Kata Kunci:** naive bayes, orange data mining, play store, sentimen, TF-IDF, TikTok

# 1. PENDAHULUAN

Perkembangan teknologi informasi dan meluasnya penggunaan internet telah mengubah pola konsumsi informasi dan interaksi sosial masyarakat secara fundamental. Media sosial berbasis video pendek, khususnya TikTok, telah mengalami pertumbuhan pengguna yang sangat pesat. Di Indonesia, TikTok menjadi salah satu aplikasi paling banyak diunduh dan digunakan, dengan puluhan juta pengguna aktif yang secara konsisten memberikan ulasan melalui platform distribusi aplikasi Google Play Store. Akumulasi data ulasan dalam jumlah sangat besar ini menghadirkan peluang sekaligus tantangan dalam analisis opini publik secara otomatis.

Analisis sentimen merupakan cabang pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan mengidentifikasi dan mengekstrak opini, perasaan, serta sikap yang terkandung dalam teks [1]. Dalam konteks ulasan aplikasi mobile, analisis sentimen memungkinkan pengembang memahami umpan balik pengguna secara efisien tanpa harus membaca jutaan ulasan secara manual. Hal ini berdampak langsung pada pengambilan keputusan terkait pengembangan fitur, perbaikan bug, dan peningkatan kualitas layanan.

Salah satu pendekatan yang paling umum digunakan dalam analisis sentimen teks berbahasa Indonesia adalah algoritma Naive Bayes. Naive Bayes merupakan metode klasifikasi berbasis probabilistik yang dikenal karena kesederhanaan, efisiensi komputasi, dan performa yang kompetitif pada tugas klasifikasi teks [2]. Beberapa penelitian telah membuktikan efektivitasnya pada ulasan aplikasi berbahasa Indonesia. Hilmi dan Irwiensyah [3] menganalisis sentimen ulasan TikTok di Google Play Store menggunakan Naive Bayes dan memperoleh akurasi 83,66%, presisi 82,97%, dan recall 91,97%. Maulana et al. [4] juga menggunakan Naive Bayes dengan TF-IDF pada ulasan TikTok dan mencapai akurasi 83%. Hidayah et al. [5] membandingkan Naive Bayes dan SVM pada 5.000 ulasan TikTok dan menemukan bahwa SVM (84%) mengungguli Naive Bayes (79%), namun Naive Bayes tetap menunjukkan performa yang memadai dengan waktu komputasi yang lebih singkat.

Penelitian terkait lainnya adalah studi Pratama dan Pamungkas [6] yang menganalisis sentimen komentar video viral (FYP) TikTok menggunakan Naive Bayes dengan pembobotan TF-IDF dan berhasil mencapai akurasi terbaik 90% dengan dua kelas sentimen. Setiawan [1] dalam tinjauan sistematis analisis sentimen berbahasa Indonesia tahun 2023–2024 menegaskan bahwa kombinasi preprocessing yang komprehensif dan pembobotan TF-IDF secara konsisten meningkatkan performa model Naive Bayes. Ardinata et al. [7] menekankan pentingnya normalisasi teks slang pada teks informal berbahasa Indonesia, yang secara langsung memengaruhi kualitas fitur yang diekstrak.

Sebagian besar penelitian sebelumnya menggunakan dataset yang terbatas (di bawah 12.500 data) dan hanya menangani dua kelas sentimen. Penelitian ini mengisi celah tersebut dengan menggunakan dataset yang jauh lebih besar (100.000 ulasan), menangani tiga kelas sentimen secara seimbang melalui balanced sampling, serta mengimplementasikan seluruh alur kerja menggunakan platform Orange Data Mining yang memungkinkan reprodusibilitas eksperimen yang lebih baik. Tujuan penelitian ini adalah: (1) membangun model klasifikasi sentimen ulasan pengguna TikTok menggunakan Orange Data Mining; (2) mengevaluasi efektivitas kombinasi TF-IDF (Sublinear) dan Naive Bayes pada tiga kelas sentimen yang seimbang; dan (3) menganalisis performa model secara komprehensif menggunakan AUC, akurasi, F1, precision, recall, dan MCC.

# 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental berbasis pembelajaran mesin. Seluruh alur kerja diimplementasikan menggunakan Orange Data Mining 3.38.1, sebuah platform visual machine learning berbasis Python yang memungkinkan pembuatan pipeline analisis teks secara intuitif. Alur penelitian terdiri dari: pengumpulan dan pelabelan data, balanced sampling, preprocessing teks, representasi Bag of Words berbobot TF-IDF Sublinear, pelatihan model Naive Bayes, dan evaluasi menggunakan 10-fold stratified cross-validation.

## 2.1 Dataset dan Pelabelan

Dataset yang digunakan adalah ulasan pengguna aplikasi TikTok yang diperoleh dari Google Play Store melalui metode web scraping. Data terdiri dari 100.000 ulasan yang dikumpulkan antara Agustus hingga Desember 2025. Setiap rekaman data mencakup enam atribut: nama pengguna, teks ulasan (Ulasan), skor rating bintang (1–5), tanggal ulasan (Tanggal), jumlah likes, dan versi aplikasi. Pendekatan scraping serupa digunakan oleh Hidayah et al. [5] dan Apriani et al. [8] dalam mengumpulkan data ulasan TikTok dari Play Store.

Pelabelan sentimen dilakukan secara otomatis berdasarkan skor rating, sebuah pendekatan yang telah lazim digunakan dalam penelitian analisis sentimen ulasan aplikasi [3], [4], [5]: rating 4–5 = Positif, rating 3 = Netral, dan rating 1–2 = Negatif. Distribusi kelas sentimen dari keseluruhan dataset awal disajikan pada Tabel 1.

**Tabel 1.** Distribusi Kelas Sentimen Dataset Awal (100.000 Ulasan)

| Kelas Sentimen | Jumlah Data | Persentase (%) | Rating |
|---|---|---|---|
| Positif | 56.938 | 56,94% | 4–5 |
| Netral | 7.260 | 7,26% | 3 |
| Negatif | 35.802 | 35,80% | 1–2 |
| **Total** | **100.000** | **100,00%** | — |

Tabel 1 menunjukkan distribusi kelas yang sangat tidak seimbang, dengan kelas Positif mendominasi (56,94%) dan Netral menjadi kelas sangat minoritas (7,26%). Untuk memastikan model tidak bias terhadap kelas dominan dan agar evaluasi lebih representatif, dilakukan balanced sampling menggunakan Python Script pada Orange: setiap kelas diambil sebanyak 3.334 sampel secara acak, sehingga total dataset setelah sampling adalah 10.002 dokumen. Distribusi setelah balanced sampling ditampilkan pada Tabel 2.

**Tabel 2.** Distribusi Kelas Sentimen Setelah Balanced Sampling

| Kelas Sentimen | Jumlah Data | Persentase (%) | Rating |
|---|---|---|---|
| Positif | 3.334 | 33,33% | 4–5 |
| Netral | 3.334 | 33,33% | 3 |
| Negatif | 3.334 | 33,33% | 1–2 |
| **Total** | **10.002** | **100,00%** | — |

### 2.2 Preprocessing Teks

Preprocessing dilakukan menggunakan widget Preprocess Text pada Orange Data Mining. Teks ulasan bahasa Indonesia informal mengandung banyak noise seperti emoji, angka, singkatan, kata slang, dan ejaan tidak baku [7]. Tahapan preprocessing yang diterapkan dalam penelitian ini mengadaptasi praktik terbaik dari literatur analisis ardsentimen bahasa Indonesia [1], [3], [6] dan disajikan pada Tabel 3.

**Tabel 3.** Tahapan Preprocessing Teks dalam Orange

| Tahap | Proses | Contoh Input | Contoh Output |
|---|---|---|---|
| 1. Lowercase | Konversi ke huruf kecil | Aplikasi TikTok BAGUS | aplikasi tiktok bagus |
| 2. Cleaning | Hapus angka, simbol, emoji | tiktok 2025 🔥 luar biasa! | tiktok luar biasa |
| 3. Tokenisasi | Pisah kalimat menjadi token kata | tiktok sangat bagus | ['tiktok','sangat','bagus'] |
| 4. Stopword Removal | Hapus stopword bahasa Indonesia | aplikasi ini sangat bagus | aplikasi bagus |
| 5. Normalisasi Slang | Koreksi kata tidak baku [7] | apk oke bgt | aplikasi oke banget |

Pada widget Preprocess Text, transformasi Lowercase diaktifkan, tokenisasi menggunakan mode Word Punctuation (pola \w+), dan stopword dihapus menggunakan daftar stopword bahasa Indonesia bawaan Orange. Pratama dan Pamungkas [6] membuktikan bahwa kelengkapan tahapan preprocessing secara langsung memengaruhi performa Naive Bayes pada teks komentar TikTok.

### 2.3 Representasi Fitur: Bag of Words dengan TF-IDF Sublinear

Representasi fitur dilakukan menggunakan widget Bag of Words pada Orange dengan konfigurasi Term Frequency: Sublinear dan Document Frequency: IDF. Konfigurasi ini secara efektif mengimplementasikan TF-IDF dengan modifikasi sublinear scaling pada TF untuk mengurangi dampak dominasi kata yang muncul sangat sering [9].

$$TF\_sub(t,d) = 1 + log(f(t,d))\ jika\ f(t,d) > 0, else\ 0 \quad (1)$$

$$IDF(t,D) = log(|D| / |\{d \in D : t \in d\}|) + 1 \quad (2)$$

$$TF-IDF\_sub(t,d,D) = TF\_sub(t,d) \times IDF(t,D) \quad (3)$$

Penggunaan Sublinear TF bertujuan mencegah kata yang muncul sangat sering mendominasi representasi vektor, sehingga kata-kata yang memiliki frekuensi moderat namun informatif tetap mendapatkan bobot yang proporsional. Achmad dan Budiman [9] menunjukkan bahwa konfigurasi pembobotan yang tepat secara signifikan memengaruhi performa Naive Bayes pada analisis sentimen teks e-commerce Indonesia.

### 2.4 Algoritma Naive Bayes

Algoritma Naive Bayes yang diimplementasikan pada Orange menggunakan prinsip Teorema Bayes dengan asumsi kondisional independensi antar fitur [2]. Meskipun asumsi ini jarang terpenuhi sempurna pada data teks, Naive Bayes tetap menunjukkan performa yang baik dan efisiensi komputasi yang tinggi [1], [3]. Teorema Bayes diformulasikan sebagai:

$$P(C|X) = P(X|C) \times P(C) / P(X) \quad (4)$$

$$\hat{C} = argmax\, P(C) \times \Pi\, P(x_i|C) \quad (5)$$

Apriani et al. [8] dan Maulana et al. [4] juga menggunakan varian Multinomial Naive Bayes untuk klasifikasi sentimen ulasan TikTok, membuktikan relevansi algoritma ini untuk domain yang sama.

### 2.5 Konfigurasi Eksperimen

Seluruh eksperimen dijalankan menggunakan alur kerja Orange Data Mining yang terdiri dari widget: CSV File Import → Python Script (balanced sampling) → Preprocess Text → Bag of Words → Naive Bayes → Test and Score → Confusion Matrix → ROC Analysis. Konfigurasi lengkap eksperimen disajikan pada Tabel 4.

**Tabel 4.** Konfigurasi Eksperimen Orange Data Mining

| Parameter | Nilai/Konfigurasi |
|---|---|
| Tool Eksperimen | Orange Data Mining 3.38.1 |
| Alur Kerja | CSV File Import → Python Script → Preprocess Text → Bag of Words → Naive Bayes → Test and Score → Confusion Matrix → ROC Analysis |
| Preprocessing | Lowercase, Tokenisasi (Word+Punctuation), Stopwords Bahasa Indonesia |
| Term Frequency | Sublinear |
| Document Frequency | IDF |
| Regularization | None |
| Teknik Sampling | Balanced sampling (3.334 per kelas) |
| Total Data Setelah Sampling | 10.002 dokumen |
| Metode Evaluasi | 10-Fold Cross Validation (Stratified) |
| Algoritma | Naive Bayes |

### 2.6 Evaluasi Model

Evaluasi dilakukan menggunakan metrik yang tersedia pada widget Test and Score Orange, yaitu: AUC (Area Under ROC Curve), CA (Classification Accuracy), F1-score, Precision, Recall, dan MCC (Matthews Correlation Coefficient). Evaluasi menggunakan 10-fold stratified cross-validation untuk memperoleh estimasi performa yang robust dan mengurangi varians akibat pembagian data [10].

$$CA = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

$$Precision = TP / (TP + FP) \quad (7)$$

$$Recall = TP / (TP + FN) \quad (8)$$

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (9)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Evaluasi Model secara Keseluruhan

Evaluasi menggunakan 10-fold stratified cross-validation pada 10.002 dokumen seimbang menghasilkan metrik yang dilaporkan oleh widget Test and Score Orange. Hasil lengkap disajikan pada Tabel 5.

**Tabel 5.** Hasil Evaluasi Model Naive Bayes (10-Fold Cross-Validation)

| Model | AUC | CA (Akurasi) | F1 | Precision | Recall | MCC |
|---|---|---|---|---|---|---|
| Naive Bayes | 0,760 | 0,530 (53,00%) | 0,509 | 0,537 | 0,530 | 0,310 |

Tabel 5 menunjukkan bahwa model Naive Bayes mencapai CA (akurasi) sebesar 0,530 (53,00%) dan AUC sebesar 0,760. Nilai AUC 0,760 mengindikasikan kemampuan diskriminatif model yang memuaskan; menurut standar interpretasi AUC, nilai di atas 0,7 dianggap sebagai klasifikasi yang cukup baik [10]. Nilai MCC sebesar 0,310 mengonfirmasi bahwa model memberikan prediksi yang lebih baik dibandingkan random chance (MCC = 0) meskipun masih ada ruang peningkatan yang signifikan. Akurasi 53,00% pada tiga kelas seimbang secara baseline jauh di atas random chance (33,33%), yang menunjukkan bahwa model mampu menangkap pola sentimen dalam teks.

### 3.2 Analisis Performa Per Kelas

Analisis performa per kelas dihitung berdasarkan confusion matrix yang dihasilkan Orange. Hasil lengkap per kelas disajikan pada Tabel 6.

**Tabel 6.** Hasil Evaluasi Per Kelas Sentimen

| Kelas | Precision (%) | Recall (%) | F1-Score (%) | Support | Keterangan |
|---|---|---|---|---|---|
| Negatif | 58,53 | 35,81 | 44,44 | 3.334 | Salah diklasif. ke Positif |
| Netral | 50,77 | 38,66 | 43,90 | 3.334 | Paling ambigu |
| Positif | 51,91 | 84,43 | 64,29 | 3.334 | Recall tertinggi |
| **Weighted Avg** | **53,74** | **52,97** | **50,87** | **10.002** | — |

Tabel 6 mengungkapkan pola performa yang bervariasi antar kelas. Kelas Positif memperoleh recall tertinggi (84,43%), yang berarti model sangat sensitif dalam mengenali ulasan positif. Namun, presisi kelas Positif relatif rendah (51,91%), mengindikasikan bahwa banyak ulasan dari kelas lain yang salah diprediksi sebagai Positif. Hal ini konsisten dengan kecenderungan bias model Naive Bayes terhadap kelas yang fitur-fiturnya lebih mudah dipelajari dari data pelatihan.

Kelas Negatif memiliki presisi tertinggi (58,53%) namun recall terendah (35,81%), menunjukkan bahwa model lebih 'konservatif' dalam memprediksi ulasan negatif. Kelas Netral menunjukkan performa paling moderat dengan presisi 50,77% dan recall 38,66%, yang mencerminkan ambiguitas semantik ulasan rating 3. Shafira dan Nugraha [11] juga melaporkan tantangan serupa pada klasifikasi tiga kelas sentimen dengan Naive Bayes.

### 3.3 Analisis Confusion Matrix

Confusion matrix dari 10-fold cross-validation pada 10.002 dokumen disajikan pada Tabel 7.

**Tabel 7.** Confusion Matrix Hasil Klasifikasi Naive Bayes

| | **Pred. Negatif** | **Pred. Netral** | **Pred. Positif** |
|---|---|---|---|
| Aktual Negatif | 1.194 | 891 | 1.249 |
| Aktual Netral | 686 | 1.289 | 1.359 |
| Aktual Positif | 160 | 359 | 2.815 |

Tabel 7 mengungkapkan pola kesalahan yang sistematis dan bermakna. Kesalahan terbesar terjadi pada kelas Negatif: dari 3.334 ulasan aktual Negatif, sebanyak 1.249 (37,46%) salah diklasifikasikan sebagai Positif dan 891 (26,72%) sebagai Netral — hanya 1.194 (35,81%) yang diprediksi benar. Hal ini menunjukkan bahwa teks ulasan negatif dalam bahasa Indonesia informal seringkali mengandung kata-kata yang secara leksikal mirip dengan ulasan positif, misalnya penggunaan sarkasme.

Pada kelas Netral, 686 ulasan (20,58%) salah diklasifikasikan sebagai Negatif dan 1.359 (40,77%) sebagai Positif, dengan hanya 1.289 (38,66%) yang benar. Kelas Positif adalah yang paling berhasil diklasifikasikan, dengan 2.815 dari 3.334 (84,43%) diprediksi benar, yang konsisten dengan karakteristik ulasan positif yang cenderung menggunakan kata-kata seperti 'bagus', 'mantap', dan 'suka' yang mudah dikenali secara leksikal.

### 3.4 Analisis Kurva ROC

Kurva ROC yang dihasilkan oleh widget ROC Analysis Orange (target kelas Negatif) menunjukkan kurva yang terletak secara signifikan di atas garis diagonal baseline (random classifier). Nilai AUC = 0,760 mengindikasikan bahwa model memiliki kemampuan membedakan kelas Negatif dari kelas lainnya yang memuaskan. Meskipun nilai akurasi keseluruhan moderat (53,00%), AUC yang lebih tinggi menunjukkan bahwa model sebenarnya mampu melakukan ranking prediksi dengan baik. Ayomi et al. [10] juga menggunakan AUC sebagai metrik evaluasi utama dan mencatat bahwa AUC memberikan gambaran yang lebih lengkap tentang kemampuan model dibandingkan akurasi saja.

### 3.5 Perbandingan dengan Penelitian Terdahulu

Tabel 8 menyajikan perbandingan hasil penelitian ini dengan penelitian-penelitian terdahulu yang relevan.

**Tabel 8.** Perbandingan Hasil dengan Penelitian Terdahulu

| **Penelitian** | **Metode & Dataset** | **Akurasi** | **Jumlah Kelas** |
|---|---|---|---|
| Hilmi & Irwiensyah [3] | Naive Bayes, ulasan TikTok Play Store | 83,66% | 2 kelas |
| Maulana et al. [4] | Naive Bayes + TF-IDF, ulasan TikTok | 83,00% | 3 kelas |
| Hidayah et al. [5] | Naive Bayes vs SVM, 5.000 ulasan TikTok | 79,00% (NB) | 2 kelas |
| Pratama & | Naive Bayes + TF-IDF, komentar FYP | 90,00% | 2 kelas |

| Penelitian | Metode & Dataset | Akurasi | Jumlah Kelas |
|---|---|---|---|
| Pamungkas [6] | TikTok | | |
| Apriani et al. [8] | Naive Bayes, TikTok media pembelajaran | 52,27% | 3 kelas |
| Shafira & Nugraha [9] | Naive Bayes + TF-IDF, Netflix Play Store | 77,96% | 3 kelas |
| Budaya & Suniantara [10] | NB + SMOTE + TF-IDF, Google Reviews Puskesmas | 89,00% | 2 kelas |
| **Penelitian Ini** | **Naive Bayes + BoW (TF-IDF), Orange, 10.002 data balanced** | **53,00%** | **3 kelas seimbang** |

Tabel 8 menjelaskan mengapa akurasi penelitian ini (53,00%) lebih rendah dibandingkan sebagian besar penelitian terdahulu. Perbedaan ini disebabkan oleh beberapa faktor metodologis yang fundamental: (1) penelitian ini menggunakan tiga kelas seimbang, sementara sebagian besar penelitian terdahulu hanya menggunakan dua kelas; (2) balanced sampling dengan tiga kelas seimbang secara inheren lebih menantang dibandingkan klasifikasi biner; dan (3) ulasan kelas Netral (rating 3) dimasukkan secara eksplisit. Apriani et al. [8] yang juga menggunakan tiga kelas hanya memperoleh akurasi 52,27%, sangat dekat dengan hasil penelitian ini dan mengonfirmasi bahwa tingkat kesulitan klasifikasi tiga kelas memang lebih tinggi.

Pateman et al. [12] menunjukkan bahwa penerapan SMOTE meningkatkan akurasi SVM dari 61% menjadi 76% pada data TikTok yang tidak seimbang. Halim et al. [13] membuktikan bahwa model hybrid deep learning CNN-LSTM dapat melampaui performa Naive Bayes secara signifikan, namun dengan biaya komputasi dan kompleksitas implementasi yang jauh lebih tinggi. Naive Bayes tetap relevan sebagai baseline yang efisien, transparan, dan mudah diinterpretasikan.

## 4. KESIMPULAN

Penelitian ini telah berhasil membangun dan mengevaluasi model analisis sentimen ulasan pengguna aplikasi TikTok menggunakan kombinasi TF-IDF (Sublinear) dan algoritma Naive Bayes yang diimplementasikan pada Orange Data Mining 3.38.1. Dataset awal berjumlah 100.000 ulasan dari Google Play Store (Agustus–Desember 2025), dengan distribusi sentimen 56,94% Positif, 35,80% Negatif, dan 7,26% Netral. Setelah balanced sampling sebanyak 3.334 dokumen per kelas, total 10.002 dokumen seimbang digunakan dalam pemodelan.

Evaluasi menggunakan 10-fold stratified cross-validation menghasilkan: CA = 0,530, AUC = 0,760, F1 = 0,509, Precision = 0,537, Recall = 0,530, dan MCC = 0,310. Analisis per kelas menunjukkan kelas Positif memperoleh recall tertinggi (84,43%), kelas Negatif memiliki presisi tertinggi (58,53%), sementara kelas Netral paling sulit diklasifikasikan (F1 = 43,90%) akibat ambiguitas semantiknya. Nilai AUC 0,760 mengkonfirmasi kemampuan diskriminatif model yang memuaskan meskipun akurasi keseluruhan masih moderat.

Untuk penelitian selanjutnya, disarankan: (1) penerapan teknik penyeimbangan data yang lebih canggih seperti SMOTE [12], [14] atau ADASYN; (2) eksplorasi model deep learning seperti CNN-LSTM [13] atau IndoBERT [1] yang telah terbukti melampaui Naive Bayes pada dataset yang lebih kompleks; (3) penggunaan emotion-based preprocessing [15] untuk menangkap nuansa emosional yang lebih halus; serta (4) pengembangan kamus normalisasi slang bahasa Indonesia

yang lebih komprehensif [7]. Penelitian ini diharapkan menjadi kontribusi dan referensi dalam pengembangan sistem analisis sentimen berbahasa Indonesia untuk domain ulasan aplikasi mobile.

## UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

## REFERENCES

[1] B. Setiawan, "A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024," vol. 08, pp. 71–83, 2024.

[2] A. H. Ramadhani, H. Q. Djauhari, V. Lius, and A. Nugroho, "Hybrid Naive Bayes TF-IDF Algorithm and Lexicon Approach for Sentiment Analysis of Reviews," vol. 12, no. 7, 2024.

[3] N. F. Hilmi, "Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naïve Bayes," vol. 14, no. 1, pp. 146–156, 2024.

[4] M. T. F. Maulana, B. I. Nugroho, E. Unggul, and S. Utami, "Analisis Sentimen Ulasan Aplikasi TikTok di Google Playstore Menggunakan Algoritma Naive Bayes," vol. 4, no. 3, pp. 6627–6635, 2025.

[5] I. A. Hidayah, R. Kusumawati, Z. Abidin, and M. Imamuddin, "Journal of Computer Networks , Architecture and High Performance Computing Analysis of Public Sentiment Towards The Tiktok Application Using The Naive Bayes Algorithm and Support Vector Machine Journal of Computer Networks , Architecture and High Performance Computing," vol. 6, no. 2, pp. 881–891, 2024.

[6] P. W. Pratama, E. W. Pamungkas, T. Informatika, and U. M. Surakarta, "Analisis Sentimen Terhadap Komentar Pada Video Viral ( Fyp ) Tiktok Menggunakan Metode Naïve Bayes," vol. 9, pp. 35–46, 2026.

[7] P. M. S. Ardinata, A. A. J. Permana, and I. N. S. W. Wijaya, "IDENTIFIKASI DAN NORMALISASI TEKS SLANG DENGAN," vol. 21, no. 1, 2024.

[8] E. Apriani, I. F. Hanif, and F. Oktavianalisti, "Sentiment Analysis of Using TikTok as a Learning Media Using the Naïve Bayes Classifiers Algorithm Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier," vol. 4, no. July, pp. 1160–1168, 2024.

[9] F. Budiman, "Optimizing Feature Extraction for Naïve Bayes Sentiment Analysis," vol. 10, no. 1, pp. 619–629, 2026.

[10] J. M. Ayomi, A. V. Vitianingsih, Y. Kristyawan, A. Lidya, and T. Widiartin, "Sentiment Analysis of User Reviews for the PLN Mobile Application Using Naïve Bayes and Long Short-Term Memory," vol. 7, no. 4, pp. 3849–3873, 2025, doi: 10.63158/journalisi.v7i4.1342.

[11] F. Shafira and A. H. Nugraha, "Sentiment Analysis of Netflix App Reviews on Google Play Store using the Naive Bayes," vol. 1, no. November, pp. 1–17, 2025.

[12] D. Pateman, T. F. Prasetyo, H. Sujadi, and U. Majalengka, "SENTIMENT ANALYSIS OF GOVERNMENT ON TIKTOK AND X," vol. 10, no. 4, pp. 900–908, 2025, doi: 10.33480/jitk.v10i4.6645.(SVM).

[13] A. Halim *et al.*, "Fine-Tuning Hybrid Deep Learning for Sentiment Analysis of Indonesian Product Reviews," vol. 20, no. 1, pp. 127–137, 2026.

[14] I. G. Bintang, A. Budaya, and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," vol. 4, no. July, pp. 1077–1086, 2024.

[15] B. R. Ermawan, M. B. Prayoga, A. R. Fadhillah, and E. Utami, "Optimizing Sentiment Classification Models for TikTok Comments using Emotion-Based Preprocessing and Grid Search," vol. 10, no. 1, pp. 535–547, 2026.