

## PERBANDINGAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBORS DALAM ANALISIS SENTIMEN ULASAN APLIKASI GOJEK

Rousyati<sup>✉</sup>, Dany Pratmanto, Fandhilah

Universitas Bina Sarana Informatika, Kampus Kota Tegal, Tegal, Indonesia

Email: [rousyati.rou@bsi.ac.id](mailto:rousyati.rou@bsi.ac.id)

DOI: <https://doi.org/10.46880/jmika.Vol10No1.pp186-191>

### ABSTRACT

*Sentiment analysis of mobile application reviews is important for monitoring perceived service quality in digital platforms. This study compares Multinomial Naive Bayes and K-Nearest Neighbors for classifying Indonesian Gojek reviews collected from Google Play Store. The dataset consisted of 5,000 extreme-rating reviews collected from 13 January to 31 March 2026; after preprocessing, 4,685 valid reviews remained, comprising 2,482 negative and 2,203 positive reviews. The method followed the Knowledge Discovery in Databases workflow, including text cleaning, Indonesian stopword removal, PySastrawi stemming, TF-IDF feature extraction with unigram and bigram features, an 80:20 stratified train-test split, 5-fold cross-validation, and McNemar testing for paired model comparison. Naive Bayes achieved 89.43% accuracy, 90.09% weighted precision, 89.43% recall, 89.34% F1-score, and 91.22% +/- 0.71% cross-validation accuracy, outperforming KNN, which achieved 86.77% accuracy, 86.78% precision, 86.77% recall, 86.75% F1-score, and 87.83% +/- 1.61% cross-validation accuracy. The McNemar test indicated that the difference between both classifiers was statistically meaningful ( $p = 0.0016$ ). The contribution of this study is a reproducible empirical pipeline for Indonesian app-review sentiment analysis and evidence that Naive Bayes is a more efficient baseline than KNN for sparse TF-IDF review data.*

**Keyword:** *Sentiment Analysis, Naive Bayes, K-Nearest Neighbors, TF-IDF, PySastrawi, Gojek, Google Play Store.*

### ABSTRAK

*Analisis sentimen terhadap ulasan aplikasi mobile penting untuk memantau persepsi pengguna terhadap kualitas layanan digital. Penelitian ini membandingkan Multinomial Naive Bayes dan K-Nearest Neighbors dalam klasifikasi sentimen ulasan aplikasi Gojek berbahasa Indonesia yang diperoleh dari Google Play Store. Dataset terdiri dari 5.000 ulasan berlabel ekstrem yang dikumpulkan pada 13 Januari sampai 31 Maret 2026; setelah preprocessing, diperoleh 4.685 ulasan valid, yaitu 2.482 ulasan negatif dan 2.203 ulasan positif. Metode penelitian mengikuti alur Knowledge Discovery in Databases, meliputi pembersihan teks, penghapusan stopwords bahasa Indonesia, stemming PySastrawi, ekstraksi fitur TF-IDF unigram dan bigram, pembagian data 80:20 secara stratifikasi, validasi silang 5-fold, serta uji McNemar untuk membandingkan prediksi kedua model secara berpasangan. Naive Bayes menghasilkan akurasi 89,43%, presisi tertimbang 90,09%, recall 89,43%, F1-score 89,34%, dan akurasi validasi silang 91,22% +/- 0,71%. KNN menghasilkan akurasi 86,77%, presisi tertimbang 86,78%, recall 86,77%, F1-score 86,75%, dan akurasi validasi silang 87,83% +/- 1,61%. Uji McNemar menunjukkan perbedaan performa yang signifikan secara statistik ( $p = 0,0016$ ). Kontribusi penelitian ini adalah penyediaan pipeline empiris yang dapat direplikasi untuk analisis sentimen ulasan aplikasi Indonesia serta bukti bahwa Naive Bayes lebih efisien daripada KNN pada data ulasan TF-IDF yang jarang dan berdimensi tinggi.*

**Kata Kunci:** *Analisis Sentimen, Naive Bayes, K-Nearest Neighbors, TF-IDF, PySastrawi, Gojek, Google Play Store.*

### PENDAHULUAN

Pertumbuhan layanan digital di Indonesia menjadikan ulasan aplikasi sebagai sumber data yang semakin penting bagi pengembang platform. Laporan Digital 2026 menunjukkan bahwa Indonesia memiliki sekitar 230 juta pengguna internet, dengan tingkat penetrasi di atas 80% dan lebih dari 331 juta koneksi seluler aktif (DataReportal, 2025). Skala penggunaan

tersebut memperbesar volume interaksi pengguna dengan layanan berbasis aplikasi, termasuk transportasi, pesan-antar makanan, pembayaran digital, dan layanan on-demand lain yang berada dalam ekosistem Gojek.

Urgensi penelitian ini juga diperkuat oleh besarnya aktivitas ekonomi pada ekosistem GoTo. Pada laporan kinerja tahun buku 2025, GoTo



melaporkan pertumbuhan pendapatan bersih 24% secara tahunan menjadi Rp18,3 triliun dan Annual Transacting Users meningkat 24% menjadi 66 juta (Group, 2026). Dalam konteks tersebut, ulasan di Google Play Store tidak hanya berisi komentar spontan, tetapi juga sinyal kualitas layanan, kestabilan aplikasi, pengalaman pengguna, dan persepsi terhadap mitra layanan.

Masalah utama yang muncul adalah volume dan keragaman bahasa pada ulasan pengguna. Ulasan aplikasi cenderung pendek, informal, mengandung singkatan, repetisi, kata tidak baku, serta campuran istilah Indonesia-Inggris. Pembacaan manual terhadap ribuan ulasan menjadi tidak efisien, sedangkan pendekatan analisis sentimen dapat membantu mengubah data teks tidak terstruktur menjadi informasi yang dapat ditindaklanjuti. Dengan klasifikasi otomatis, pengembang dapat memantau proporsi keluhan dan apresiasi secara lebih cepat, meskipun hasilnya tetap perlu ditafsirkan dengan hati-hati.

Sejumlah penelitian telah membandingkan algoritma pembelajaran mesin untuk analisis sentimen bahasa Indonesia. Namun, research gap masih terlihat pada tiga aspek. Pertama, banyak studi menggunakan data media sosial seperti Twitter/X, sementara karakteristik ulasan aplikasi berbeda karena lebih terkait langsung dengan pengalaman penggunaan produk. Kedua, sebagian studi hanya melaporkan akurasi tanpa menguji apakah selisih performa antarmodel bermakna secara statistik. Ketiga, pembahasan tentang sumber bias dataset, validasi data, dan konsekuensi penggunaan label rating bintang masih terbatas. Penelitian ini mengisi celah tersebut dengan membandingkan Naive Bayes dan KNN pada ulasan terbaru Gojek, menyertakan validasi silang, uji McNemar, serta diskusi keterbatasan dataset.

Kontribusi utama penelitian ini adalah evaluasi empiris yang dapat direplikasi mengenai efektivitas Naive Bayes dan KNN untuk klasifikasi sentimen ulasan aplikasi Gojek berbahasa Indonesia dengan preprocessing PySastrawi dan representasi TF-IDF.

Rumusan masalah dalam penelitian ini berfokus pada bagaimana performa algoritma Naive Bayes dan K-Nearest Neighbors (KNN) dalam mengklasifikasikan sentimen ulasan aplikasi Gojek berbahasa Indonesia dengan menggunakan fitur TF-IDF serta tahapan preprocessing menggunakan PySastrawi. Selain itu, penelitian ini juga mengkaji apakah terdapat perbedaan performa antara kedua model yang signifikan secara statistik, serta mengidentifikasi keterbatasan dataset dan metode yang digunakan yang perlu diperhatikan dalam interpretasi hasil analisis. Adapun tujuan dari penelitian ini adalah

untuk membangun sebuah pipeline analisis sentimen terhadap ulasan aplikasi Gojek yang diambil dari Google Play Store, mengevaluasi dan membandingkan performa algoritma Naive Bayes dan KNN pada data ulasan berbahasa Indonesia, menjelaskan secara teoretis dan praktis penyebab perbedaan performa kedua model, serta menyajikan keterbatasan penelitian beserta rekomendasi yang lebih spesifik untuk pengembangan penelitian selanjutnya.

## TINJAUAN PUSTAKA

Analisis sentimen merupakan bagian dari pemrosesan bahasa alami yang bertujuan mengidentifikasi polaritas opini dalam teks (Liu & Cardie, 2014). Dalam penelitian klasik, pendekatan berbasis pembelajaran mesin telah terbukti efektif untuk klasifikasi opini, terutama saat fitur teks direpresentasikan sebagai bag-of-words atau TF-IDF (Turney, 2002), (Pang et al., 2002). Namun, keberhasilan model tidak hanya ditentukan oleh algoritma, tetapi juga oleh karakteristik bahasa, domain data, strategi pelabelan, dan validasi eksperimen.

Pada teks bahasa Indonesia, preprocessing memiliki peran penting karena variasi morfologi, bahasa informal, dan campuran istilah asing sering muncul dalam ulasan pengguna. Stemming dan stopword removal dapat meningkatkan kinerja model pada kondisi tertentu, tetapi pengaruhnya tidak selalu seragam pada semua algoritma (Pradana & Hayaty, 2019). Oleh karena itu, konfigurasi preprocessing perlu dijelaskan secara eksplisit agar penelitian dapat direplikasi.

Ulasan aplikasi mobile memiliki karakteristik yang berbeda dari tweet atau komentar media sosial umum. Satu ulasan dapat memuat beberapa aspek layanan, misalnya aplikasi lambat, pesanan batal, driver ramah, atau promo menarik. Perbedaan konteks aspek-aspek tersebut akhirnya memengaruhi distribusi kata, panjang teks, sampai kemungkinan munculnya opini yang lebih kaya daripada sekadar polaritas dokumen. Meski demikian, klasifikasi biner tetap relevan untuk tahap awal pemantauan kualitas layanan karena dapat memberi gambaran cepat mengenai kecenderungan umum persepsi pengguna.

Naive Bayes sering menjadi baseline kuat untuk klasifikasi teks karena bekerja baik pada fitur sparse dan berdimensi tinggi. KNN, sebaliknya, sensitif terhadap pemilihan metrik jarak dan jumlah tetangga. Pada ruang fitur TF-IDF yang sangat jarang, jarak antardokumen yang pendek dan mirip dapat menjadi kurang stabil, sehingga keputusan kelas sangat dipengaruhi oleh tetangga terdekat yang belum tentu mewakili seluruh pola kelas. Pertimbangan ini menjadi

alasan teoretis mengapa Naive Bayes berpotensi unggul pada dataset ulasan pendek.

**Tabel 1.** Perbandingan dengan Penelitian Terdahulu

Studi	Fokus Data	Temuan Utama	Celah Relevan
Turney (Turney, 2002); Pang et al. (Pang et al., 2002)	Ulasan dan klasifikasi opini umum	Pendekatan supervised efektif untuk klasifikasi sentimen.	Belum spesifik pada ulasan aplikasi Indonesia.
Pradana dan Hayaty (Pradana & Hayaty, 2019)	Teks bahasa Indonesia	Stemming dan stopwords removal dapat memengaruhi akurasi.	Efek preprocessing perlu divalidasi per dataset.
Azhar dan Masruroh (Azhar et al., 2023)	Analisis sentimen Indonesia	Membandingkan Naive Bayes dan KNN.	Belum menekankan uji statistik dan bias label rating.
Pratama dan Suhartono (Andayani et al., 2024)	Twitter/X politik Indonesia	Naive Bayes kompetitif dibanding KNN.	Domain media sosial berbeda dari ulasan aplikasi.
Penelitian ini	Ulasan Gojek 2026	NB dibanding KNN dengan TF-IDF, validasi silang, McNemar, dan model tambahan	Memberi benchmark domain aplikasi on-demand Indonesia.

**METODE PENELITIAN**

Penelitian menggunakan kerangka Knowledge Discovery in Databases (KDD), yang terdiri atas seleksi data, preprocessing, transformasi fitur, data mining, serta interpretasi dan evaluasi. Kerangka ini dipilih karena cocok untuk penelitian yang berangkat dari data operasional tidak terstruktur dan memerlukan tahapan pembersihan sebelum pemodelan.

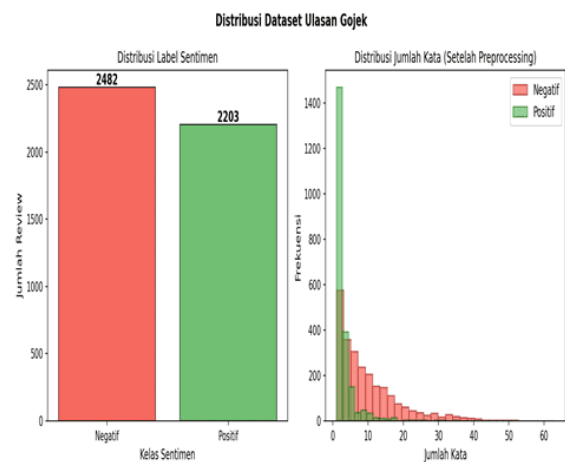
Dataset diperoleh dari halaman resmi aplikasi Gojek di Google Play Store dengan API tidak resmi melalui library google-play-scraper. Data mencakup ulasan pada 13 Januari sampai 31 Maret 2026. Penelitian ini menggunakan rating ekstrem untuk pelabelan awal: bintang 1 sebagai sentimen negatif dan bintang 5 sebagai sentimen positif. Ulasan bintang 2, 3, dan 4 tidak digunakan agar polaritas label lebih jelas.

Setelah penyaringan, dikumpulkan 5.000 ulasan awal, yaitu 2.500 ulasan negatif dan 2.500 ulasan positif.

Validasi dataset dilakukan melalui pemeriksaan kelengkapan teks, rentang rating, konsistensi label, tanggal ulasan, dan hasil preprocessing kosong. Setelah preprocessing, 315 ulasan menjadi kosong dan dikeluarkan, sehingga tersisa 4.685 ulasan valid. Dataset akhir mencakup 2.482 ulasan negatif (52,97%) dan 2.203 ulasan positif (47,03%). Potensi bias tetap ada karena model hanya belajar dari ulasan rating ekstrem, tidak mencakup opini netral, dan label rating dapat berbeda dari nuansa teks. Dengan demikian, hasil penelitian sebaiknya dipahami sebagai klasifikasi polaritas eksplisit, bukan pengukuran sentimen penuh pada seluruh populasi ulasan.

**Tabel 2.** Ringkasan Data Penelitian

Aspek Dataset	Keterangan
Sumber	Google Play Store aplikasi Gojek
Rentang waktu	13 Januari 2026 sampai 31 Maret 2026
Data awal	5.000 ulasan ekstrem: 2.500 bintang 1 dan 2.500 bintang 5
Data valid setelah preprocessing	4.685 ulasan
Distribusi kelas valid	2.482 negatif (52,97%) dan 2.203 positif (47,03%)
Split data	80:20 stratified split, 3.748 data latih dan 937 data uji
Fitur	TF-IDF unigram dan bigram, 3.322 fitur



**Gambar 1.** Distribusi Kelas Sentimen Setelah Preprocessing

**Preprocessing Teks**

Tahap preprocessing dilakukan untuk mengurangi noise dan menyatukan variasi bentuk kata. Proses meliputi case folding, penghapusan URL, tag HTML, angka, tanda baca, karakter non-alfanumerik,

emoji, dan spasi berlebih. Selanjutnya, teks dipecah menjadi token, stopword bahasa Indonesia dihapus dengan daftar Sastrawi yang diperkaya stopword domain aplikasi, dan stemming dilakukan menggunakan PySastrawi. Konfigurasi ini dipilih karena ulasan aplikasi sering mengandung bentuk kata tidak baku, repetisi, serta ekspresi pendek yang perlu dinormalisasi sebelum dihitung sebagai fitur.

**Ekstraksi Fitur dan Pemilihan Model**

Teks hasil preprocessing ditransformasikan menjadi representasi numerik menggunakan TF-IDF. Parameter yang digunakan adalah n-gram range (1,2), min\_df = 2, max\_df = 0,95, sublinear\_tf = True, dan max\_features = 5.000 fitur. Kombinasi unigram dan bigram digunakan agar model dapat menangkap ekspresi sentimen seperti 'tidak bisa', 'sangat kecewa', atau 'bagus cepat', yang sering kehilangan makna jika hanya diproses sebagai kata tunggal.

Naive Bayes dipilih karena merupakan baseline probabilistik yang kuat untuk klasifikasi teks sparse, sedangkan KNN dipilih sebagai pembanding berbasis kedekatan dokumen. Naive Bayes menggunakan MultinomialNB dengan alpha = 1,0 dan KNN menggunakan k = 7 dengan metrik cosine. Pada analisis tambahan, Logistic Regression dan Linear SVM digunakan sebagai model pembanding karena keduanya umum dipakai pada klasifikasi teks berbasis TF-IDF.

**Model dan Evaluasi**

Evaluasi dilakukan pada data uji 937 sampel dengan metrik akurasi, precision, recall, dan F1-score, baik secara tertimbang maupun per kelas. Validasi silang stratifikasi 5-fold digunakan untuk melihat stabilitas performa. Selain itu, uji McNemar digunakan untuk menguji apakah perbedaan prediksi dua model pada sampel yang sama signifikan secara statistik. Pendekatan ini memperkuat interpretasi karena selisih akurasi saja belum cukup untuk menyimpulkan keunggulan model.

**HASIL DAN PEMBAHASAN**

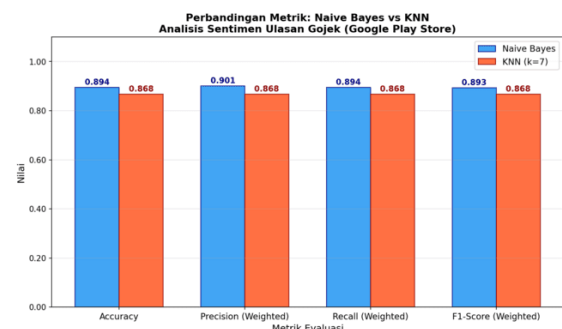
**Distribusi Dataset**

Distribusi kelas setelah preprocessing relatif seimbang, dengan kelas negatif sedikit lebih dominan. Kondisi ini mengurangi kebutuhan terhadap teknik penyeimbangan kelas seperti oversampling atau undersampling. Namun, proporsi negatif yang sedikit lebih tinggi tetap perlu diperhatikan karena pola teks pendek dapat memengaruhi sensitivitas model terhadap keluhan teknis.

**Perbandingan Metrik Evaluasi**

**Tabel 3.** Perbandingan Metrik Evaluasi NB dan KNN

Metrik	Naive Bayes	KNN	Selisih	Interpretasi
Akurasi	89,43%	86,77%	+2,66 %	NB lebih tepat secara keseluruhan.
Presisi tertimbang	90,09%	86,78%	+3,31 %	Prediksi NB lebih bersih dari false positive.
Recall tertimbang	89,43%	86,77%	+2,66 %	NB menangkap lebih banyak sampel benar.
F1-score tertimbang	89,34%	86,75%	+2,59 %	Keseimbangan presisi-recall NB lebih baik.
CV 5-fold	91,22% ± 0,71%	87,83% ± 1,61%	+3,39 %	NB lebih stabil lintas fold.



**Gambar 2.** Visualisasi Perbandingan Metrik Naive Bayes dan KNN

Tabel berikut menunjukkan bahwa Naive Bayes mengungguli KNN pada seluruh metrik utama. Selisih akurasi 2,66 poin persentase tampak moderat, tetapi menjadi penting karena disertai standar deviasi validasi silang yang lebih kecil. Naive Bayes yang berbasis probabilitas lebih sesuai dengan representasi TF-IDF yang sparse; setiap fitur kata diperlakukan sebagai indikator probabilistik terhadap kelas. Sebaliknya, KNN harus membandingkan kedekatan dokumen pada ruang berdimensi tinggi, sehingga jarak antardokumen dapat menjadi kurang informatif ketika banyak fitur bernilai nol.

Secara teoretis, keunggulan Naive Bayes dapat dijelaskan oleh kecocokannya dengan data teks sparse. Meskipun asumsi independensi antarfitur tidak sepenuhnya terpenuhi, model ini memanfaatkan distribusi kata yang sangat kontras antara sentimen negatif dan positif. KNN lebih rentan terhadap variasi kata pendek, slang, dan repetisi. Pada ulasan negatif, istilah seperti 'error', 'lambat', 'gagal', 'susah', atau 'kecewa' dapat memberi sinyal kelas yang kuat; TF-IDF

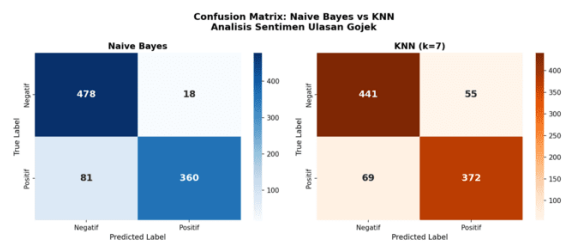
membantu mengangkat kata-kata tersebut menjadi fitur yang lebih informatif.

**Analisis Per Kelas**

Tabel 4. Evaluasi Per Kelas pada Data Uji

Model	Kelas	Presisi	Recall	F1-score	Dukungan
Naive Bayes	Negatif	0,8551	0,9637	0,9062	496
Naive Bayes	Positif	0,9524	0,8163	0,8791	441
KNN	Negatif	0,8647	0,8891	0,8767	496
KNN	Positif	0,8712	0,8435	0,8571	441

Naive Bayes memiliki recall sangat tinggi pada kelas negatif (96,37%), sehingga lebih jarang melewatkan ulasan yang benar-benar negatif. Hal ini penting secara praktis karena keluhan pengguna biasanya membutuhkan respons lebih cepat daripada apresiasi. Namun, recall kelas positif NB lebih rendah (81,63%), yang menunjukkan bahwa sebagian ulasan positif masih diprediksi sebagai negatif. Pola ini dapat terjadi ketika ulasan positif mengandung kritik ringan, misalnya pujian yang disertai keluhan fitur tertentu.



Gambar 3. Confusion Matrix Naive Bayes dan KNN

**Uji Statistik dan Model Pemodelan**

Tabel 5. Uji McNemar dan Pemodelan Eksploratif

Analisis	Sel A benar/B salah	Sel B salah/B benar	p-value	Interpretasi
NB vs KNN	48	21	0,0016	Perbedaan signifikan; NB lebih andal dari KNN.
Linear SVM vs NB	43	28	0,0959	SVM lebih tinggi, tetapi belum signifikan pada alpha 0,05.
Linear SVM vs KNN	73	31	<0,001	Linear SVM signifikan lebih baik dari KNN.

Uji McNemar menunjukkan bahwa keunggulan Naive Bayes terhadap KNN bukan sekadar variasi acak pada data uji. Hasil  $p = 0,0016$  memperkuat bahwa keputusan prediksi kedua model berbeda secara bermakna. Analisis tambahan juga menunjukkan bahwa Linear SVM memperoleh akurasi lebih tinggi

daripada Naive Bayes, tetapi selisih tersebut belum signifikan pada taraf 5%. Karena fokus penelitian adalah perbandingan Naive Bayes dan KNN, hasil tersebut diposisikan sebagai indikasi awal untuk penelitian lanjutan, bukan sebagai kesimpulan utama.

Tabel 6. Ringkasan Model Utama dan Pemodelan Tambahan

Model	Akurasi	Presisi	Recall	F1-score	CV 5-fold	Catatan
Naive Bayes	89,43%	90,09%	89,43%	89,34%	91,22% +/- 0,71%	Baseline efisien dan stabil.
KNN	86,77%	86,78%	86,77%	86,75%	87,83% +/- 1,61%	Lebih sensitif terhadap ruang fitur sparse.
Logistic Regression	88,79%	89,30%	88,79%	88,71%	90,32% +/- 0,74%	Kompetitif, tetapi tidak melampaui NB pada data uji.
Linear SVM	91,04%	91,05%	91,04%	91,04%	91,84% +/- 0,71%	Tertinggi pada analisis tambahan.

**KESIMPULAN**

Penelitian ini membandingkan Naive Bayes dan KNN pada 4.685 ulasan valid aplikasi Gojek dari Google Play Store. Dataset dibagi menjadi 3.748 data latih dan 937 data uji, dengan 3.322 fitur TF-IDF. Naive Bayes menghasilkan akurasi 89,43%, presisi 90,09%, recall 89,43%, F1-score 89,34%, serta validasi silang 91,22% +/- 0,71%. KNN memperoleh akurasi 86,77%, presisi 86,78%, recall 86,77%, F1-score 86,75%, serta validasi silang 87,83% +/- 1,61%. Uji McNemar menunjukkan perbedaan signifikan antara kedua model ( $p = 0,0016$ ).

Secara ilmiah, penelitian ini memperkuat bukti bahwa Naive Bayes merupakan baseline yang efektif untuk klasifikasi teks bahasa Indonesia berbasis TF-IDF, khususnya pada ulasan pendek yang sparse. Secara praktis, pipeline ini dapat membantu pengembang aplikasi memantau keluhan dan apresiasi pengguna secara lebih cepat. Hasil analisis tambahan menunjukkan Linear SVM sebagai kandidat model lanjutan karena memperoleh akurasi 91,04%, meskipun peningkatannya terhadap Naive Bayes belum signifikan pada taraf 5%.

Penelitian selanjutnya disarankan melakukan analisis berbasis aspek untuk memisahkan sentimen terhadap harga, driver, pembayaran, promosi, dan kestabilan aplikasi. Selain itu, penelitian dapat memperluas label menjadi negatif, netral, dan positif; membandingkan data Google Play Store dengan App Store; menambahkan model transformer bahasa Indonesia seperti IndoBERT; serta menguji performa



model secara temporal agar dapat melihat perubahan sentimen setelah pembaruan aplikasi atau perubahan kebijakan layanan.

#### DAFTAR PUSTAKA

- Andayani, M., Marisa, F., & Putra, R. P. (2024). Sentiment Analysis of Indonesia 2024 Election with a Comparison of Naive Bayes and KNN Algorithms on Twitter. *SAR Journal*, 7(3), 204–212. <https://doi.org/10.18421/SAR73>
- Azhar, Masruroh, S. U., Wardhani, L. K., & Okfalisa. (2023). Performance comparison of the Naive Bayes algorithm and the k-NN lexicon approach on Twitter media sentiment analysis. *Science, Technology, and Communication Journal*, 3(2), 35–40. <https://doi.org/10.59190/stc.v3i2.229>
- DataReportal. (2025). *Digital 2026: Indonesia*. <https://datareportal.com/>
- Fields, J., Chovanec, K., & Madiraju, P. (2024). A survey of text classification with transformers: How wide? how large? how long? how accurate? how expensive? how safe?. *IEEE Access*, 12, 6518-6531.
- GoTo Group. (2026). *GoTo beats guidance, achieving record results as it reports 2025 fourth quarter and full year earnings*. <https://www.gotocompany.com/>
- Liu, B., & Cardie, C. (2014). *Book Reviews Sentiment Analysis and Opinion Mining*. <https://doi.org/10.1162/COLI>
- Manning, C. D. (2008). *Introduction to information retrieval*. Syngress Publishing.
- Pang, B., Lee, L., Rd, H., & Jose, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2002)* (pp. 79-86).
- Pradana, A. W., & Hayaty, M. (2019). The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on indonesian-language texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(3).
- Turney, P. D. (2002). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. (July), (pp. 417-424).