

Perbandingan Kinerja Multinomial Naïve Bayes dan Complement Naïve Bayes pada Analisis Sentimen Ulasan Aplikasi Gojek Berbahasa Indonesia dengan Penanganan Class Imbalance Menggunakan SMOTE

Dimas Permana^{1*}, Riska Septiani²

¹⁻² Program Studi Sistem Informasi, Universitas Pamulang Kampus Serang

*Corresponding email: perdimas98@gmail.com^{1*} dosen02912@unpam.ac.id²

Abstrak: Penelitian ini bertujuan membandingkan kinerja algoritma *Multinomial Naïve Bayes* (MNB) dan *Complement Naïve Bayes* (CNB) dalam analisis sentimen ulasan aplikasi Gojek berbahasa Indonesia yang mengalami *class imbalance*, serta mengukur kontribusi *Synthetic Minority Over-sampling Technique* (SMOTE) terhadap CNB. Nilai tambah penelitian ini meliputi rancangan *pipeline preprocessing* lima tahap yang adaptif terhadap karakteristik bahasa ulasan informal Indonesia, serta aplikasi pengujian interaktif berbasis *Streamlit* untuk membandingkan hasil prediksi ketiga model. Dataset yang digunakan berjumlah 500 ulasan dari Kaggle (406 negatif, 94 positif), diekstraksi menggunakan TF-IDF, dan dibagi 80:20 secara *stratified*. Hasil pengujian menunjukkan MNB memperoleh akurasi 81,00% (Macro F1-Score 44,75%), CNB memperoleh akurasi 75,00% (Macro F1-Score 60,18%), sedangkan CNB+SMOTE memperoleh akurasi 80,00% dengan Macro F1-Score tertinggi 67,51% dan peningkatan F1 Positif +11,47 poin dibanding CNB tanpa SMOTE, dikuatkan validasi *5-Fold Stratified Cross Validation*. Penelitian ini menyimpulkan bahwa kombinasi CNB dan SMOTE merupakan konfigurasi terbaik untuk menangani *class imbalance* pada data ulasan Gojek.

Kata kunci: Analisis Sentimen, *Class Imbalance*, *Complement Naïve Bayes*, *Multinomial Naïve Bayes*, SMOTE.

Abstract: This study aims to compare the performance of *Multinomial Naïve Bayes* (MNB) and *Complement Naïve Bayes* (CNB) in sentiment analysis of Indonesian-language Gojek reviews affected by *class imbalance*, and to measure the contribution of the *Synthetic Minority Over-sampling Technique* (SMOTE) to CNB. Its added value lies in a five-stage preprocessing pipeline adaptive to informal Indonesian review text, and an interactive *Streamlit* application for comparing predictions from the three models. The dataset comprises 500 Kaggle reviews (406 negative, 94 positive), extracted using TF-IDF, and split 80:20 using a stratified scheme. Results show MNB achieved 81.00% accuracy (Macro F1-Score 44.75%), CNB achieved 75.00% accuracy (Macro F1-Score 60.18%), while CNB+SMOTE achieved 80.00% accuracy with the highest Macro F1-Score of 67.51% and an 11.47-point increase in Positive-class F1 over CNB without SMOTE, confirmed by *5-Fold Stratified Cross Validation*. This study concludes that CNB combined with SMOTE is the best configuration for handling class imbalance in Gojek review data.

Keywords: *Class Imbalance*, *Complement Naïve Bayes*, *Multinomial Naïve Bayes*, *Sentiment Analysis*, SMOTE.

PENDAHULUAN

Perkembangan teknologi informasi telah mendorong transformasi digital di berbagai sektor kehidupan, termasuk sektor transportasi. Transportasi online kini menjadi kebutuhan pokok masyarakat modern karena menyediakan akses kendaraan yang lebih mudah, cepat, dan fleksibel dibandingkan moda transportasi konvensional. Salah satu platform transportasi online yang paling banyak digunakan di Indonesia adalah Gojek, yang dipelopori oleh Nadiem Makarim dan kini menjadi bagian dari PT GoTo. Sebagai super-app terbesar di Indonesia, Gojek menghimpun beragam layanan dalam satu aplikasi, mulai dari *GoRide*, *GoCar*, *GoSend*, *GoMart*, hingga *GoFoods* (Nugraha & Gustian, 2023). Seluruh layanan tersebut dapat diakses melalui aplikasi di *Google Play Store*, yang juga menyediakan fitur ulasan (*review*) dan penilaian (*rating*) sebagai wadah bagi pengguna untuk menyampaikan pendapat dan pengalaman mereka secara terbuka (Pradipta et al., 2026).

Tingginya adopsi Gojek tercermin dari data unduhan aplikasi yang telah mencapai 142 juta pada tahun 2019 (Nugraha & Gustian, 2023). Volume ulasan pengguna di *Google Play Store* juga terus bertambah dari waktu ke waktu: tercatat sekitar 4,8 juta ulasan pada tahun 2022 dengan rerata rating 4,6, meningkat menjadi sekitar 5,9 juta ulasan pada tahun

2024. Pertambahan sekitar satu juta ulasan hanya dalam kurun waktu dua tahun tersebut mengindikasikan bahwa ribuan ulasan baru masuk ke platform setiap harinya, menjadikan ulasan pengguna sebagai sumber informasi tekstual yang besar dan sangat dinamis (Wahyudi & Kusumawardhana, 2021).

Berbagai penelitian terdahulu telah menerapkan algoritma *Naïve Bayes* pada tugas klasifikasi teks dan analisis sentimen karena kesederhanaan implementasinya yang tidak mengorbankan kualitas hasil, serta kecepatan komputasinya dibandingkan algoritma pembelajaran mesin lain yang lebih kompleks (Mubarokah et al., 2025). *Naïve Bayes* juga terbukti memberikan akurasi yang baik dan stabil pada data bertipe teks, khususnya ketika dipadukan dengan representasi fitur TF-IDF (Hendrawan & Kusniyati, 2024). Untuk menangani *class imbalance* secara khusus, *Complement Naïve Bayes* (CNB) dikembangkan dengan mengestimasi probabilitas dari komplemen kelas, sehingga secara teoritis lebih *robust* dibandingkan *Multinomial Naïve Bayes* (MNB) yang sensitif terhadap dominasi kelas mayoritas (Alenazi et al., 2023). Belum ditemukan kajian empiris yang secara khusus membandingkan kinerja MNB dan CNB pada dataset ulasan aplikasi transportasi online berbahasa Indonesia dengan kondisi *class imbalance* nyata, sekaligus memvalidasi hasilnya melalui *5-Fold Cross Validation*. Celah penelitian (*research gap*) inilah yang menjadi dasar dilakukannya penelitian ini.

Penelitian ini menggunakan dataset 500 ulasan pengguna aplikasi Gojek berbahasa Indonesia yang diperoleh dari platform *Kaggle*, diproses menggunakan *Google Colaboratory* dengan bahasa pemrograman *Python* dan *Library PySastrawi* untuk penanganan morfologi bahasa Indonesia. Sebagai solusi atas permasalahan yang telah diuraikan, penelitian ini merancang *pipeline preprocessing* lima tahap yang adaptif terhadap karakteristik bahasa ulasan informal, kemudian membandingkan kinerja *Multinomial Naïve Bayes* (MNB) sebagai baseline dengan *Complement Naïve Bayes* (CNB) sebagai model utama, serta menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE) pada skema CNB+SMOTE untuk mengukur sejauh mana teknik ini dapat meningkatkan kemampuan model dalam mendeteksi kelas minoritas. Sebagai luaran tambahan, penelitian ini juga mengembangkan aplikasi pengujian sederhana berbasis *Streamlit* yang mendemonstrasikan dan membandingkan hasil prediksi ketiga model secara interaktif.

Meskipun penelitian analisis sentimen ulasan aplikasi berkembang pesat, sebagian besar fokus pada algoritma standar tanpa menangani *class imbalance* atau tanpa menyediakan implementasi praktis. Celah ini dijawab melalui kebaruan penelitian berupa integrasi *Complement Naïve Bayes* (CNB) dan *SMOTE* pada ulasan transportasi online berbahasa Indonesia, dipadukan dengan *preprocessing* lima tahap dan *validasi 5-Fold Stratified Cross Validation*. Penelitian ini juga menghadirkan luaran praktis berupa aplikasi web *Streamlit* untuk menguji dan membandingkan prediksi model secara *real-time*, sehingga memberikan kontribusi akademik dalam mengatasi bias kelas mayoritas sekaligus menyediakan alat evaluasi layanan yang praktis.

LANDASAN TEORI

Analisis sentimen atau opinion mining merupakan cabang dari *Natural Language Processing* (NLP) dan *text mining* yang bertujuan untuk mengidentifikasi, mengekstrak, serta mengklasifikasikan pendapat, evaluasi, atau emosi yang terkandung dalam teks terhadap suatu entitas seperti produk, layanan, atau topik tertentu (Lee, 2023). Pada analisis ulasan aplikasi, sentimen umumnya diklasifikasikan ke dalam kategori *biner* (Positif dan Negatif) atau multi-kelas (Positif, Netral, Negatif) untuk memahami tingkat kepuasan pengguna secara otomatis (Firda et al., 2025).

Streamlit adalah pustaka (*library*) *Python* bersifat *open-source* yang digunakan untuk membangun aplikasi web interaktif berbasis data *science* dan *machine learning* dengan cepat, tanpa memerlukan keahlian pengembangan *front-end* yang mendalam. *Streamlit* memungkinkan model *machine learning* yang telah dilatih untuk diintegrasikan langsung ke dalam antarmuka web yang interaktif, sehingga sangat sesuai digunakan untuk membangun purwarupa aplikasi pengujian dan pembandingan kinerja model, sebagaimana diterapkan pada penelitian ini (Admira et al., 2025).

Gojek adalah perusahaan teknologi dan transportasi berbasis online yang berdiri pada tahun 2010, dipelopori oleh Nadiem Makarim, dan kini merupakan bagian dari PT. GoTo Gojek Tokopedia Tbk salah satu *decacorn* pertama di Asia Tenggara. Platform ini menawarkan ekosistem layanan terintegrasi yang mencakup *GoRide*, *GoCar*, *GoSend*, *GoMart*, dan *GoFood*. Volume ulasan yang sangat besar di *Google Play Store* dan sifatnya yang tidak terstruktur menjadikan teknik analisis sentimen otomatis sangat diperlukan untuk mengekstrak wawasan bermakna dari data tersebut (Tukino & Fifi, 2024).

Multinomial Naïve Bayes (MNB) merupakan varian algoritma *Naïve Bayes* yang dirancang untuk data berbentuk frekuensi sehingga sangat cocok untuk klasifikasi teks berbasis TF-IDF. MNB mengestimasi probabilitas fitur berdasarkan frekuensi kemunculan kata pada kelas target. Namun, MNB memiliki kelemahan fundamental saat diterapkan pada dataset dengan distribusi kelas yang tidak seimbang (*class imbalance*). Probabilitas *prior* yang proporsional dengan ukuran kelas menyebabkan bias yang kuat terhadap kelas mayoritas, yang dirumuskan sebagai berikut:

$$P(C) = N_C / N$$

dengan $P(C)$ adalah probabilitas *prior* dari kelas C , N_C adalah jumlah sampel dokumen dalam kelas C , dan N adalah total keseluruhan sampel dokumen dalam dataset (Valero-Carreras et al., 2023).

Complement Naïve Bayes (CNB) adalah varian dari algoritma *Naïve Bayes* yang dirancang secara spesifik untuk menangani kelemahan *Multinomial Naïve Bayes* (MNB) pada kondisi ketidakseimbangan kelas (*class imbalance*) (Alenazi et al., 2023).

Algoritma *Synthetic Minority Over-sampling Technique* (SMOTE) mengatasi permasalahan *class imbalance* tersebut dengan menghasilkan sampel sintesis baru melalui teknik interpolasi *K-Nearest Neighbors* (KNN) pada kelas minoritas, bukan sekadar menduplikasi data yang ada (Sapriadi & Nasution, 2024).

METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental berbasis *supervised learning* dengan mengadopsi kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri atas pemahaman permasalahan bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, hingga implementasi hasil. Pendekatan kuantitatif dipilih karena penelitian berorientasi pada pengukuran performa model secara numerik menggunakan indikator statistik seperti akurasi, presisi, *recall*, dan *F1-score*.

Dataset yang digunakan adalah data sekunder "*Gojek Reviews Relevant Sentiment*" dari platform *Kaggle* berformat CSV, berisi 500 ulasan pengguna aplikasi Gojek berbahasa Indonesia dengan tiga atribut *user*, *review*, dan *rating* (1-5 bintang). Pelabelan sentimen dilakukan secara otomatis berbasis *rating* dengan skema, *rating* 4-5 bintang dikategorikan Positif dan *rating* 1-3 bintang dikategorikan Negatif, menghasilkan distribusi kelas sebagaimana disajikan pada Tabel 1.

Tabel 1. Distribusi Rating dan Pelabelan Sentimen Dataset

Rating	Kategori	Jumlah	Persentase	Label Sentimen
1	Sangat Tidak Puas	287	57,40%	Negatif
2	Tidak Puas	62	12,40%	Negatif
3	Cukup	52	11,40%	Negatif
4	Puas	41	8,20%	Positif
5	Sangat Puas	53	10,60%	Positif
Total		500	100%	406 Negatif / 94 Positif

Rasio kelas Negatif dibanding Positif sebesar 4,3:1 mengindikasikan kondisi *class imbalance* yang signifikan, sekaligus menjadi justifikasi utama pemilihan *Complement Naïve Bayes* sebagai model utama dan penerapan *SMOTE* sebagai eksperimen tambahan.

Preprocessing Teks Berbahasa Indonesia

Teks ulasan yang diunggah pengguna pada platform digital umumnya bersifat informal, tidak terstruktur, serta mengandung banyak noise, singkatan, dan ejaan tidak baku. Untuk mengatasi permasalahan tersebut, dirancang sebuah *pipeline preprocessing* lima tahap yang adaptif terhadap karakteristik bahasa ulasan informal Indonesia.

1. Case Folding

Tahap pertama dilakukan dengan mengonversi seluruh huruf dalam teks menjadi format kecil (*lowercase*). Langkah ini bertujuan untuk menyelaraskan representasi kata sehingga kata yang sama tidak dianggap berbeda hanya karena perbedaan penggunaan huruf kapital.

2. Cleaning

Tahap kedua memfokuskan pada pembersihan teks dari elemen-elemen *non-tekstual* yang tidak memberikan kontribusi terhadap penentuan sentimen. Proses ini memanfaatkan ekspresi reguler (*regular expression*) untuk menghapus tautan URL, sebutan akun (*mention @*), tagar (*hashtag #*), karakter angka, serta seluruh simbol dan tanda baca.

3. Normalisasi Slang

Tahap ketiga bertujuan menstandarkan kosa kata tidak baku, singkatan, maupun istilah slang menjadi kata baku sesuai kaidah Bahasa Indonesia. Proses ini menggunakan kamus pemetaan khusus yang disusun berisi 127 entri entitas kata informal (sebagai contoh, kata "gak", "ga", atau "gda" dipetakan menjadi "tidak", dan kata "apk" dipetakan menjadi "aplikasi").

4. Stopword Removal

Tahap keempat adalah penghapusan kata-kata umum yang memiliki frekuensi kemunculan tinggi namun tidak membawa informasi kontekstual yang signifikan (*stop words*). Proses ini mengintegrasikan pustaka *PySastrawi*, namun dengan modifikasi krusial berupa pengecualian terhadap 12 kata negasi (seperti "tidak", "bukan", "belum", "jangan", dan sejenisnya). Kata negasi sengaja dipertahankan karena penghapusannya dapat mengubah atau menghilangkan arti sentimen sebuah frasa secara drastis.

5. Stemming

Tahap terakhir dalam *preprocessing* adalah mengubah kata-kata berimbuhan menjadi bentuk kata dasarnya. Algoritma yang digunakan adalah *Enhanced Confix Stripping* (ECS) yang dijalankan melalui pustaka *PySastrawi*. Proses ini

membantu mereduksi variabilitas kata sehingga fitur yang diekstrak menjadi lebih padat.

Ekstraksi Fitur TF-IDF

Teks bersih hasil *preprocessing* dikonversi menjadi representasi numerik menggunakan *TfidfVectorizer* dari *library Scikit-learn* dengan parameter: *max_features*=3.000 untuk membatasi jumlah fitur pada kata paling informatif, *ngram_range*=(1,2) untuk menangkap *unigram* dan *bigram* sehingga frasa seperti "\"tidak bagus\" dapat direpresentasikan sebagai satu fitur, *min_df*=2 untuk mengabaikan kata yang hanya muncul pada satu dokumen, dan *sublinear_tf*=True untuk menerapkan transformasi logaritmik guna mereduksi dominasi kata yang muncul sangat sering. Hasil ekstraksi menghasilkan matriks fitur berukuran 500 x 1.715, karena *vocabulary* aktual setelah *preprocessing* hanya terdiri dari 1.715 kata unik yang muncul minimal di dua dokumen.

Pembagian Data dan Pelatihan Model

Matriks fitur TF-IDF dibagi menjadi data *training* (80%) dan *testing* (20%) menggunakan *Stratified Train-Test Split* dengan *random_state*=42, sehingga proporsi kelas Positif:Negatif yang tidak seimbang (4,3:1) tetap terjaga pada kedua subset, menghasilkan 400 data *training* (75 Positif, 325 Negatif) dan 100 data *testing* (19 Positif, 81 Negatif). Tiga konfigurasi model dilatih dan dibandingkan: (1) *Multinomial Naïve Bayes* (MNB) sebagai *baseline*, (2) *Complement Naïve Bayes* (CNB) sebagai model utama, dan (3) CNB+SMOTE, yaitu CNB yang dilatih pada data *training* hasil penyeimbangan *SMOTE* dengan *k-neighbors*=3. Seluruh model menggunakan parameter *alpha*=0,5 yang dipilih berdasarkan pengujian beberapa kandidat nilai *alpha* (0,1; 0,3; 0,5; 0,7; 1,0) menggunakan *Macro F1-Score* pada *5-Fold Stratified Cross Validation*. Penerapan *SMOTE* dibatasi hanya pada data *training* untuk mencegah data *leakage*, jika *SMOTE* diterapkan sebelum pembagian data, sampel sintetis dari data *testing* berpotensi ikut masuk ke *training* set sehingga evaluasi menjadi terlalu optimis. Validasi tambahan dilakukan menggunakan *5-Fold Stratified Cross Validation*, dengan *SMOTE* untuk skema CNB+SMOTE diterapkan melalui *imblearn*. Pipeline agar *oversampling* hanya berlaku pada data *training* setiap *fold*.

Evaluasi Model

Evaluasi performa model dilakukan menggunakan data *testing* dengan metrik akurasi, presisi, *recall*, *F1-Score*, dan *confusion matrix*, sebagaimana dirumuskan pada Persamaan (1) hingga (4).

$$Akurasi = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (3)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

Penelitian ini dilaksanakan pada lingkungan pengembangan berbasis *cloud*, yaitu *Google Colaboratory*, dengan bahasa pemrograman *Python* serta pustaka *Scikit-learn* (TF-IDF, MNB, CNB, evaluasi), *PySastrawi* (*stopword* dan *stemming* bahasa Indonesia), *Imbalanced-learn* (*SMOTE*), *Pandas*, *Matplotlib*, dan *Seaborn*. Model terbaik selanjutnya diintegrasikan ke dalam aplikasi web menggunakan framework *Streamlit* yang memungkinkan pengguna membandingkan hasil prediksi ketiga model secara interaktif terhadap ulasan baru maupun dataset dalam jumlah banyak.

HASIL DAN PEMBAHASAN

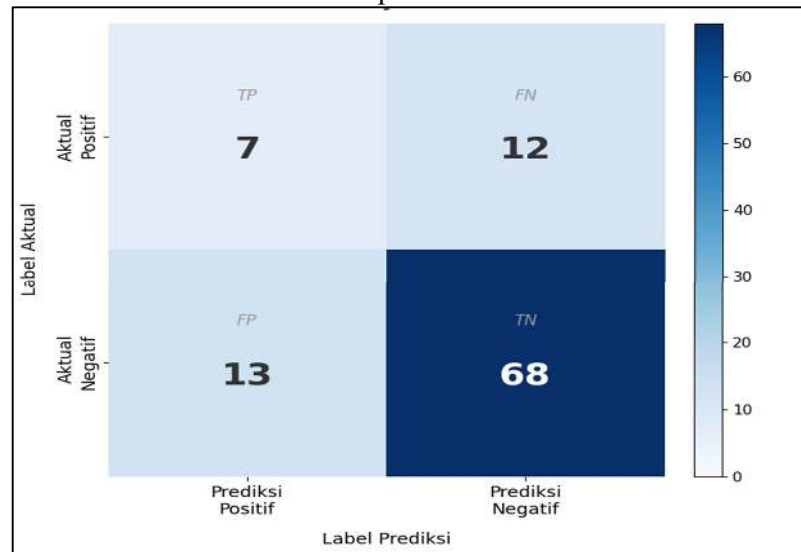
Hasil *Confusion Matrix*

Pengujian dilakukan terhadap 100 data testing (19 ulasan Positif, 81 ulasan Negatif) untuk membandingkan hasil prediksi ketiga konfigurasi model terhadap label ground truth.



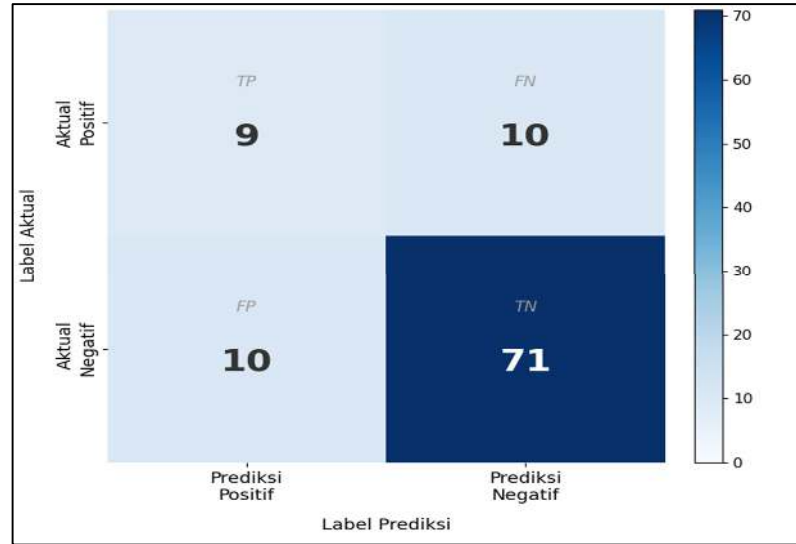
Gambar 1. *Confusion Matrix* Model MNB (Baseline)

Confusion matrix MNB menunjukkan model memprediksi seluruh 100 data testing sebagai kelas Negatif (TN=81, FP=0) tanpa satu pun prediksi Positif yang benar (TP=0, FN=19). *Accuracy* 81,00% yang dihasilkan bersifat menyesatkan (*misleading*) karena diperoleh dengan mengorbankan seluruh kemampuan mendeteksi kelas Positif, sehingga membuktikan bahwa MNB sangat rentan mengalami bias mayoritas dan tidak layak digunakan pada kondisi *class imbalance* seperti dataset ini.



Gambar 2. *Confusion Matrix* CNB (Model Utama)

Confusion matrix CNB menunjukkan model mulai mampu mengenali kelas Positif dengan benar mendeteksi 7 dari 19 ulasan Positif (TP=7, FN=12), meski harus mengorbankan sebagian akurasi pada kelas Negatif (FP=13, TN=68) sehingga *Accuracy* turun menjadi 75,00% dibandingkan MNB. *Trade-off* ini menunjukkan CNB mulai mempelajari pola kelas minoritas, bukan sekadar menebak kelas mayoritas seperti MNB.



Gambar 3. Confusion Matrix CNB+SMOTE

Confusion matrix CNB+SMOTE menunjukkan perbaikan pada seluruh sel dibandingkan CNB tanpa SMOTE: TP meningkat dari 7 menjadi 9, FN menurun dari 12 menjadi 10, FP menurun dari 13 menjadi 10, dan TN meningkat dari 68 menjadi 71. Peningkatan konsisten pada keempat sel ini membuktikan bahwa penerapan SMOTE pada data training tidak hanya meningkatkan deteksi kelas Positif, tetapi juga tidak mengorbankan performa pada kelas Negatif, sehingga CNB+SMOTE merupakan konfigurasi model terbaik secara holistik.

Metrik Evaluasi per Model

Tabel 2. Metrik Evaluasi Model MNB (Baseline)

Metrik Evaluasi	Kelas Positif	Kelas Negatif	Macro Average
Precision	0,00%	81,00%	40,50%
Recall	0,00%	100,00%	50,00%
F1-Score	0,00% Δ	89,50%	44,75%
Accuracy	—	—	81,00%

Tabel 3. Metrik Evaluasi Model CNB (Model Utama)

Metrik	Kelas Positif	Kelas Negatif	Macro Average
Precision	35,00%	85,00%	60,00%
Recall	36,84%	83,95%	60,40%
F1-Score	35,90%	84,47%	60,18%

<i>Accuracy</i>	—	—	75,00%
-----------------	---	---	--------

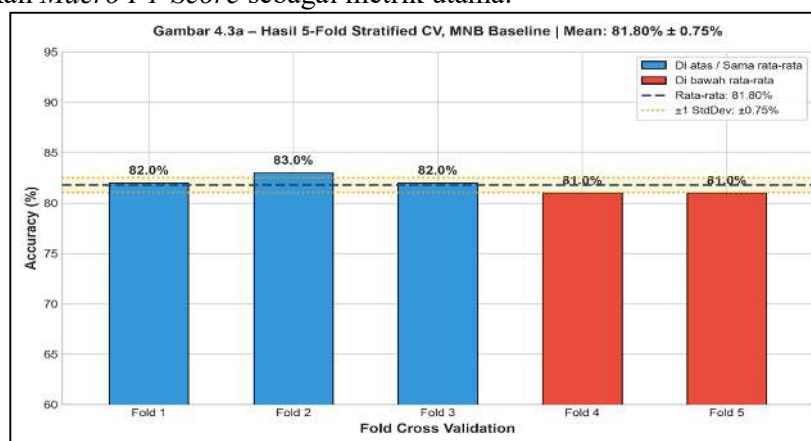
Tabel 4. Metrik Evaluasi Model CNB+SMOTE

Metrik Evaluasi	Kelas Positif	Kelas Negatif	Macro Average
<i>Precision</i>	47,37%	87,65%	67,51%
<i>Recall</i>	47,37%	87,65%	67,51%
<i>F1-Score</i>	47,37%	87,65%	67,51%
<i>Accuracy</i>	—	—	80,00%

Nilai *Accuracy* MNB sebesar 81,00% bersifat sangat menyesatkan karena diperoleh secara pasif dengan memprediksi seluruh data sebagai Negatif, sehingga *F1-Score* kelas Positif tercatat 0,00% (gagal total). CNB memberikan performa yang jauh lebih seimbang dengan *F1-Score* Positif 35,90% dan *Macro F1-Score* 60,18%, membuktikan keunggulan estimasi probabilitas komplemen pada kondisi *imbalance*. CNB+SMOTE mencapai *F1-Score* Positif 47,37% (meningkat +11,47 poin *persentase* dibanding CNB) serta *Macro F1-Score* tertinggi 67,51% dengan *Accuracy* 80,00%, menunjukkan manfaat nyata *oversampling* sintetis SMOTE dalam memperkuat representasi kelas minoritas.

Hasil 5-Fold Stratified Cross Validation

Untuk memastikan stabilitas dan generalisabilitas model pada berbagai subset data, dilakukan validasi *5-Fold Stratified Cross Validation* terhadap ketiga model menggunakan *Macro F1-Score* sebagai metrik utama.



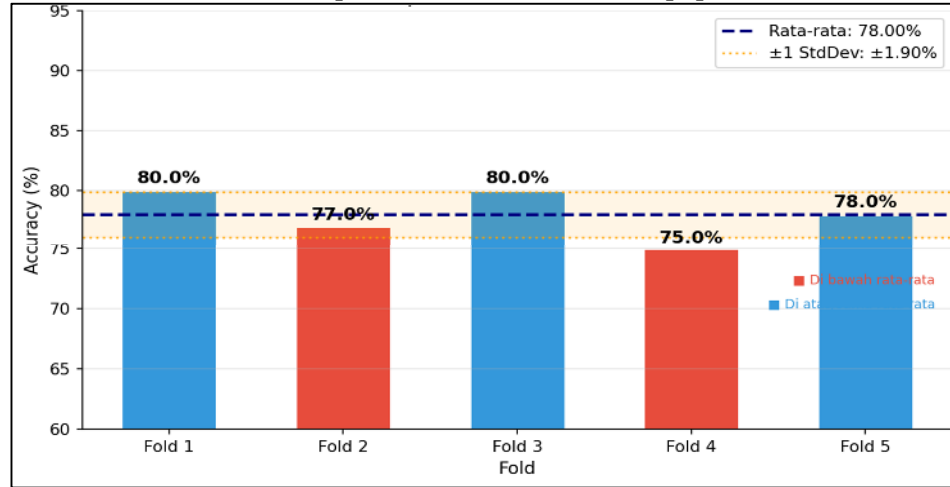
Gambar 4. Hasil 5-Fold Stratified Cross Validation Model MNB

Tabel 5. Rekapitulasi 5-Fold Cross Validation Model MNB

Fold ke-	Accuracy (%)	Macro F1 (%)
1	82,00	44,50
2	83,00	41,30

3	82,00	43,10
4	81,00	39,80
5	81,00	34,84
Rata-rata	81,80 ± 0,75	40,70 ± 3,53

Rata-rata *Accuracy* MNB pada validasi silang adalah 81,80% ± 0,75%, namun rata-rata *Macro F1* sangat rendah (40,70% ± 3,53%) dengan standar deviasi yang relatif tinggi, mengindikasikan ketidakstabilan performa MNB akibat bias populasi data.



Gambar 5. Hasil 5-Fold Stratified Cross Validation Model CNB

Tabel 6. Rekapitulasi 5-Fold Cross Validation Model CNB

Fold Ke-	Accuracy (%)	Macro F1 (%)
1	77,00%	61,50%
2	76,00%	60,80%
3	77,00%	61,20%
4	74,00%	58,90%
5	71,00%	58,60%
Rata-rata	75,00% ± 2,28%	60,20% ± 1,24%

Rata-rata *accuracy* CNB pada 5-Fold Stratified CV sebesar 75,00% ± 2,28% dengan standar deviasi kecil, mengindikasikan model tidak mengalami *overfitting* dan konsisten pada berbagai *subset* data. Nilai ini selaras dengan *accuracy* pada data *testing* (75,00%), mengonfirmasi validitas evaluasi.

Identifikasi Faktor yang Memengaruhi Sentimen

Berdasarkan analisis kata dengan bobot TF-IDF tertinggi per kelas (kata '*driver*' memiliki bobot tertinggi sebesar 0,049, diikuti '*aplikasi*' sebesar 0,044 dan '*gojek*' sebesar 0,034) serta penelaahan sampel ulasan, teridentifikasi lima faktor utama yang

memengaruhi sentimen pengguna Gojek: (a) kinerja driver merupakan faktor dominan, dengan keluhan cancel order dan driver tidak merespons mendominasi sentimen negatif sedangkan driver yang ramah dan tepat waktu mendominasi sentimen positif, (b) kualitas layanan *GoFood*, terutama lamanya waktu tunggu dan keterlambatan pengiriman makanan, (c) performa aplikasi, dengan keluhan *error*, *bug*, *crash*, dan sulit login yang menunjukkan perlunya peningkatan kualitas pengujian software, (d) sistem pembayaran *GoPay*, mencakup keluhan biaya admin yang tidak transparan dan kegagalan transaksi, serta (e) harga dan promo, di mana sentimen positif muncul saat harga terjangkau atau promosi aktif sedangkan sentimen negatif meningkat saat harga dianggap naik tiba-tiba.

Implementasi Aplikasi Streamlit

Ketiga model terbaik yang diperoleh dari proses pelatihan diintegrasikan ke dalam aplikasi pengujian interaktif berbasis *Streamlit* dengan delapan fungsi utama: input ulasan tunggal, pemilihan skenario ulasan *preset*, analisis sentimen menggunakan tiga model sekaligus (MNB, CNB, CNB+SMOTE), tampilan rincian *preprocessing* lima tahap, pemilihan preset dataset *benchmark*, unggah dataset CSV mandiri, visualisasi doughnut chart beserta tabel prediksi *batch*, dan tabel evaluasi kinerja. Aplikasi ini memungkinkan pengguna membandingkan secara langsung hasil prediksi ketiga model terhadap ulasan baru maupun dataset ulasan dalam jumlah banyak, sehingga hasil penelitian tidak hanya bersifat teoretis tetapi juga dapat diuji dan diaplikasikan secara nyata.

SIMPULAN

Model analisis sentimen otomatis berhasil dirancang dan diterapkan, dilengkapi pipeline preprocessing lima tahap (*case folding*, *cleansing*, *normalisasi slang*, *stopword removal* dengan retensi kata negasi, dan *stemming* menggunakan *Sastrawi*) yang berhasil mereduksi dimensi kosakata menjadi 407 kata bersih, serta representasi fitur TF-IDF. Model MNB sebagai baseline menghasilkan *Accuracy* 81,00% namun *Macro F1-Score* hanya 44,75% (F1 Positif 0,00%), sedangkan CNB sebagai model utama menghasilkan *Accuracy* 75,00% dengan *Macro F1-Score* 60,18% (F1 Positif 35,90%). Hasil ini menunjukkan bahwa model berhasil diterapkan untuk mengotomatisasi klasifikasi sentimen ulasan Gojek, meskipun tingkat keberhasilannya bergantung pada varian algoritma yang digunakan.

Perbandingan empiris menunjukkan bahwa CNB lebih unggul daripada MNB dalam mendeteksi kelas minoritas (positif). Penerapan SMOTE pada CNB (CNB+SMOTE) selanjutnya meningkatkan kinerja model, menghasilkan *Accuracy* 80,00% dengan *Macro F1-Score* tertinggi sebesar 67,51% (F1 Positif 47,37%, meningkat +11,47 poin persentase dibandingkan CNB tanpa SMOTE). Hierarki kinerja CNB+SMOTE > CNB > MNB ini konsisten pada evaluasi *hold-out* maupun *5-Fold Stratified Cross Validation*, sehingga kombinasi CNB dan SMOTE terbukti menjadi solusi terbaik untuk menangani *class imbalance* pada dataset ulasan Gojek.

Aplikasi pengujian interaktif berbasis *Streamlit* berhasil dirancang dan dibangun dengan delapan fungsi utama, meliputi input ulasan tunggal, pemilihan skenario ulasan *preset*, analisis sentimen menggunakan tiga model sekaligus, tampilan rincian *preprocessing*, pemilihan *preset* dataset *benchmark*, unggah dataset CSV mandiri, visualisasi *doughnut chart* beserta tabel prediksi *batch*, dan tabel evaluasi kinerja. Aplikasi ini memungkinkan pengguna/penguji membandingkan hasil prediksi ketiga model secara langsung dan interaktif.

DAFTAR PUSTAKA

- Admira, Tb. M. A., Sutarno, S., & Saefudin, M. (2025). Model Sentiment Analysis Berbasis Machine Learning Untuk Data Genz-Career Aspiration Menggunakan Flask Dan Naive Bayes. *Journal of Information System, Informatics and Computing*, 9(1), 25–39. <https://doi.org/10.52362/jisicom.v9i1.1880>
- Alenazi, F. S., Hindi, K. El, & AsSadhan, B. (2023). Complement-Class Harmonized Naïve Bayes Classifier. *Applied Science*, 13(4852), 1–18. <https://doi.org/10.3390/app13084852>
- Firda, H., Putra, P., Oktadini, N. R., Sevtiyuni, P. E., & Sevtiyuni, P. E. (2025). Perbandingan Pelabelan Rating-based dan Inset Lexicon-based dalam Analisis Sentimen Menggunakan SVM (Studi Kasus: Ulasan Aplikasi GoBiz di Google Play Store). *Sistemasi: Jurnal Sistem Informasi*, 14(2), 516–528. <https://doi.org/10.32520/stmsi.v14i2.4795>
- Hendrawan, G. N., & Kusniyati, H. (2024). Evaluasi Performa Naive Bayes dan SVM dalam Analisis Sentimen Kendaraan Listrik di Media Sosial Twitter. *CESS (Journal of Computer Engineering, System and Science)*, 9(1), 299–313. <https://doi.org/10.24114/cess.v9i1.54086>
- Lee, R. (2023). Natural Language Processing A Textbook With Python Implementations. In *Springer*. Springer. <https://doi.org/10.1007/978-981-99-1999-4>
- Mubarokah, A. S., Irawan, B., & Prihartono, W. (2025). PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI ADAKAMI DI GOOGLE PLAY STORE. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3), 2295–2305. <https://doi.org/10.23960/jitet.v13i3.6191>
- Nugraha, D., & Gustian, D. (2023). Analisis Sentimen Pengguna Aplikasi Transportasi Online pada Ulasan Google Play Store Menggunakan Algoritma SVM. *Seminar Nasional Sistem Informasi Dan Manajemen Informatika (SISMATIK)*, 3, 99–106. <https://sismatik.nusaputra.ac.id/index.php/sismatik/id/article/view/209>
- Pradipta, T. P., Anwariningsih, S. H., & Susilo, D. (2026). Perbandingan Kinerja Klasifikasi Sentimen Ulasan Aplikasi Over the Top di Google Play Store. *Journal of Computer Engineering, System and Science*, 11(1), 51–63. <https://doi.org/10.24114/cess.v11i1.70832>
- Sapriadi, & Nasution, M. (2024). Oversampling Menggunakan Pendekatan Latin Hypercube Sampling Dan K-Nearest Neighbor Untuk Meningkatkan Kinerja Klasifikasi. *Jurnal Komputer Terapan*, (November), 98–110.
- Tukino, & Fifi. (2024). Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online. *Jurnal Desain Dan Analisis Teknologi*, 3(2), 104–113. <https://doi.org/10.58520/jddat.v3i2.59>
- Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers and Operations Research*, 152(December 2022), 106131. <https://doi.org/10.1016/j.cor.2022.106131>
- Wahyudi, R., & Kusumawardhana, G. (2021). Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), 200–207. <https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/9681>