Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 4 No. 3 (2026) pp: 7476-7489
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Peningkatan Kinerja Multinomial Naïve Bayes untuk Analisis Sentimen Mobile JKN dengan SMOTE

Ridho, Abdullah
Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri
ridhomdta@gmail.com, abdialam@gmail.com

Abstrak

Aplikasi Mobile JKN merupakan salah satu layanan digital yang digunakan untuk mendukung akses peserta terhadap layanan Jaminan Kesehatan Nasional. Ulasan pengguna pada Google Play Store menyediakan informasi mengenai pengalaman dan tanggapan pengguna, tetapi distribusi sentimen yang tidak seimbang dapat menyebabkan model klasifikasi lebih dominan mengenali kelas mayoritas. Penelitian ini bertujuan menganalisis pengaruh Synthetic Minority Over-sampling Technique (SMOTE) terhadap kinerja Multinomial Naïve Bayes (MNB) dalam klasifikasi sentimen ulasan Mobile JKN. Data penelitian diperoleh melalui web scraping sebanyak 10.000 ulasan, kemudian melalui proses seleksi dan preprocessing sehingga diperoleh 7.726 ulasan untuk pemodelan dengan tiga kelas sentimen, yaitu negatif, netral, dan positif. Representasi teks menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), sedangkan evaluasi dilakukan menggunakan Stratified 5-Fold Cross-Validation dengan membandingkan MNB tanpa SMOTE dan MNB dengan SMOTE. Hasil menunjukkan bahwa MNB baseline memperoleh accuracy 86,77%, macro precision 59,48%, macro recall 59,50%, dan macro F1-score 58,97%. Penerapan SMOTE menghasilkan accuracy 74,37%, tetapi meningkatkan macro precision menjadi 63,25%, macro recall menjadi 64,10%, dan macro F1-score menjadi 60,13%. Recall kelas netral meningkat dari 0% menjadi 39,22%. Hasil tersebut menunjukkan bahwa SMOTE tidak meningkatkan accuracy keseluruhan, tetapi mampu meningkatkan kemampuan model dalam mengenali kelas minoritas dan menghasilkan kinerja yang lebih seimbang antar kelas pada data ulasan Mobile JKN tersebut.

Kata kunci: *Mobile JKN, Analisis Sentimen, Multinomial Naïve Bayes, TF-IDF, SMOTE*

Abstract

The Mobile JKN application is one of the digital services used to support participants' access to National Health Insurance services. User reviews on the Google Play Store provide information about users' experiences and responses; however, an imbalanced sentiment distribution may cause the classification model to be more dominant in recognizing the majority class. This study aims to analyze the effect of the Synthetic Minority Over-sampling Technique (SMOTE) on the performance of Multinomial Naïve Bayes (MNB) in sentiment classification of Mobile JKN reviews. The research data were obtained through web scraping of 10,000 reviews, followed by selection and preprocessing, resulting in 7,726 reviews for modeling with three sentiment classes, namely negative, neutral, and positive. Text representation used Term Frequency–Inverse Document Frequency (TF-IDF), while evaluation was conducted using Stratified 5-Fold Cross-Validation by comparing MNB without SMOTE and MNB with SMOTE. The results show that the baseline MNB achieved an accuracy of 86.77%, macro precision of 59.48%, macro recall of 59.50%, and macro F1-score of 58.97%. The application of SMOTE resulted in an accuracy of 74.37%, but increased macro precision to 63.25%, macro recall to 64.10%, and macro F1-score to 60.13%. The recall of the neutral class increased from 0% to 39.22%. These results indicate that SMOTE does not improve overall accuracy, but it can improve the model's ability to recognize minority classes and produce more balanced performance across classes in the Mobile JKN review data.

Keywords: Mobile JKN, Sentiment Analysis, Multinomial Naïve Bayes, TF-IDF, SMOTE

1. Pendahuluan

Aplikasi Mobile JKN merupakan salah satu layanan digital yang dikembangkan BPJS Kesehatan untuk memudahkan peserta dalam mengakses berbagai layanan Jaminan Kesehatan Nasional. Penggunaan aplikasi tersebut menghasilkan banyak ulasan pada Google Play Store yang mencerminkan pengalaman, kepuasan, maupun keluhan pengguna. Ali dan Kadafi (2025) menunjukkan bahwa pengalaman pengguna Mobile JKN dapat dilihat melalui aspek kemudahan, efisiensi, keandalan, dan pengalaman penggunaan pada fitur REHAB.

Ulasan pengguna juga dapat dimanfaatkan sebagai sumber data untuk mengevaluasi layanan Mobile JKN secara lebih luas. Prasetyo et al. (2026) menganalisis 100.000 ulasan Mobile JKN dan menemukan bahwa sentimen positif masih dominan, tetapi proporsi ulasan negatif juga cukup besar. Keluhan yang ditemukan terutama berkaitan dengan login dan autentikasi, kinerja aplikasi, antrean layanan, data kepesertaan, serta pembayaran iuran. Kondisi tersebut menunjukkan bahwa pengolahan ulasan secara otomatis diperlukan agar informasi yang terkandung di dalamnya dapat dianalisis secara lebih sistematis.

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen dengan metode pembelajaran mesin. Amaa et al. (2025) menerapkan Multinomial Naïve Bayes untuk mengklasifikasikan sentimen ulasan Mobile JKN menggunakan TF-IDF. Hasil penelitian tersebut menunjukkan bahwa kelas negatif dan positif dapat dikenali dengan baik, sedangkan kelas netral masih menjadi tantangan bagi model. Penggunaan metode Naïve Bayes untuk analisis ulasan aplikasi juga dikembangkan dalam penelitian Kurniawan et al. (2025) melalui penerapan SMOTE untuk menangani ketidakseimbangan kelas.

Ketidakseimbangan kelas menjadi permasalahan penting karena kelas dengan jumlah data yang lebih besar dapat lebih dominan dalam proses pembelajaran model. Tamami et al. (2025) menemukan adanya ketidakseimbangan kelas pada ulasan Mobile JKN dan menerapkan SMOTE bersama LSTM untuk mengatasinya. Hasil penelitian menunjukkan bahwa SMOTE mampu meningkatkan kemampuan model pada kelas minoritas meskipun terdapat sedikit penurunan accuracy. Safitri et al. (2026) juga menggunakan SMOTE pada klasifikasi sentimen tiga kelas untuk mengatasi permasalahan distribusi data yang tidak seimbang.

Penelitian analisis sentimen pada ulasan aplikasi terus berkembang dengan penggunaan metode dan objek yang beragam. Shafira dan Nugraha (2025) menggunakan Naïve Bayes untuk menganalisis ulasan aplikasi Netflix, sedangkan Suhaimi dan Lestari (2024) melakukan analisis sentimen ulasan TikTok dengan beberapa metode pembelajaran mesin. Apriliyanti et al. (2026) juga menerapkan Naïve Bayes pada ulasan aplikasi GetContact. Perkembangan tersebut menunjukkan bahwa ulasan pada Google Play Store dapat menjadi sumber data yang relevan untuk klasifikasi sentimen, tetapi kemampuan model dalam menangani distribusi kelas yang tidak seimbang masih perlu diperhatikan.

Berdasarkan penelitian terdahulu, masih diperlukan evaluasi yang secara khusus membandingkan Multinomial Naïve Bayes tanpa SMOTE dan dengan SMOTE pada ulasan Mobile JKN dengan tiga kelas sentimen, yaitu negatif, netral, dan positif. Penelitian ini menggunakan TF-IDF dan Stratified 5-Fold Cross-Validation untuk menjaga proporsi kelas pada proses evaluasi. Kedua skenario dibandingkan menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Kebaruan penelitian ini terletak pada evaluasi komparatif penerapan SMOTE pada Multinomial Naïve Bayes dengan perhatian khusus terhadap kemampuan model dalam mengenali kelas minoritas, khususnya kelas netral, pada ulasan Mobile JKN.

2. Metode Penelitian

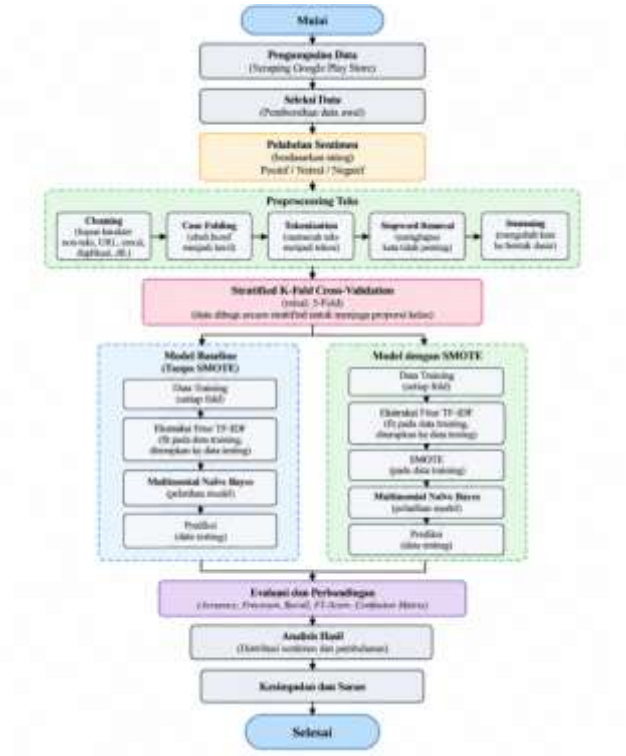
Penelitian ini menggunakan pendekatan klasifikasi teks untuk menganalisis sentimen pada ulasan pengguna aplikasi Mobile JKN yang diperoleh dari Google Play Store. Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan data hingga analisis hasil klasifikasi. Alur penelitian meliputi pengumpulan data, seleksi data, pelabelan sentimen, preprocessing teks, Stratified K-Fold Cross-Validation, pembentukan fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), penerapan Multinomial Naïve Bayes, penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE), evaluasi, serta perbandingan hasil kedua skenario model.

Penggunaan Multinomial Naïve Bayes untuk menganalisis sentimen ulasan aplikasi Mobile JKN telah dilakukan oleh Amaa et al. (2025) dengan memanfaatkan data ulasan Google Play Store dan representasi fitur TF-IDF. Penelitian ini mengembangkan pendekatan tersebut dengan memberikan perhatian pada permasalahan ketidakseimbangan kelas melalui penerapan SMOTE. Penerapan SMOTE pada klasifikasi sentimen menggunakan Multinomial Naïve Bayes juga telah digunakan pada penelitian ulasan aplikasi oleh Java et al. (2024).

Untuk memperoleh evaluasi yang lebih representatif, penelitian ini menggunakan 5-Fold Stratified Cross-Validation dengan mempertahankan proporsi kelas sentimen pada setiap fold. Situngkir et al. (2026) menunjukkan penggunaan Cross-Validation dalam evaluasi klasifikasi sentimen ulasan aplikasi, sedangkan

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

penerapan strategi stratifikasi dalam penelitian ini digunakan untuk menjaga distribusi kelas selama proses validasi. Selanjutnya, kinerja Multinomial Naïve Bayes tanpa SMOTE dibandingkan dengan Multinomial Naïve Bayes yang menggunakan SMOTE berdasarkan accuracy, precision, recall, dan F1-score.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data penelitian berupa ulasan pengguna aplikasi Mobile JKN yang tersedia pada Google Play Store. Ulasan dipilih karena memuat pengalaman, penilaian, dan tanggapan pengguna terhadap penggunaan aplikasi. Penggunaan ulasan Google Play Store untuk analisis sentimen Mobile JKN juga dilakukan oleh Amaa et al. (2025).

Pengumpulan data dilakukan menggunakan teknik web scraping untuk memperoleh ulasan secara terstruktur. Data yang dikumpulkan mencakup teks ulasan dan atribut pendukung, terutama rating yang digunakan dalam proses pelabelan. Pendekatan scraping ulasan Google Play Store juga digunakan dalam penelitian Zulfiqri et al. (2024).

Data hasil scraping kemudian disimpan sebagai dataset untuk tahap seleksi, pelabelan, dan preprocessing. Seluruh data diproses dengan prosedur yang sama agar eksperimen MNB baseline dan MNB dengan SMOTE dapat dibandingkan secara konsisten.

2.2 Seleksi Data

Data hasil pengumpulan selanjutnya melalui tahap seleksi untuk memastikan kesesuaiannya dengan kebutuhan penelitian. Seleksi dilakukan dengan memeriksa kelengkapan data serta mengidentifikasi ulasan yang tidak dapat digunakan, seperti data kosong, duplikat, dan teks yang tidak memadai. Tahap penyaringan diperlukan untuk memperoleh dataset yang lebih terstruktur sebelum memasuki proses pelabelan dan preprocessing. Pendekatan penyaringan data sebelum analisis sentimen juga diterapkan oleh Situngkir et al. (2026).

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Pada tahap ini, atribut yang tidak diperlukan dalam proses klasifikasi tidak digunakan sebagai fitur model. Data yang dipertahankan terutama berupa teks ulasan dan rating pengguna. Ulasan yang memenuhi kriteria diteruskan ke tahap pelabelan sentimen, sedangkan data yang tidak memenuhi kriteria dikeluarkan dari dataset.

2.3 Pelabelan Sentimen

Pelabelan sentimen dilakukan terhadap data hasil seleksi berdasarkan rating pengguna pada Google Play Store. Penelitian ini menggunakan tiga kelas sentimen, yaitu Negatif, Netral, dan Positif. Rating 1–2 dikategorikan sebagai Negatif, rating 3 sebagai Netral, sedangkan rating 4–5 sebagai Positif. Pendekatan pelabelan berdasarkan rating juga digunakan oleh Amaa et al. (2025) dalam analisis sentimen ulasan aplikasi Mobile JKN.

Hasil pelabelan selanjutnya digunakan sebagai target klasifikasi pada tahap preprocessing dan pemodelan. Pelabelan berbasis rating digunakan secara konsisten pada seluruh dataset agar kedua skenario model, yaitu MNB baseline dan MNB dengan SMOTE, menggunakan target kelas yang sama.

2.4 Preprocessing Teks

Tahap preprocessing dilakukan untuk membersihkan dan menormalisasi teks ulasan sebelum digunakan dalam proses klasifikasi. Tahapan yang digunakan meliputi cleaning, case folding, tokenization, stopword removal, dan stemming. Proses serupa digunakan dalam penelitian analisis sentimen ulasan aplikasi oleh Ramdani et al. (2024) dan Apriliyanti et al. (2026).

Pada tahap cleaning, karakter yang tidak diperlukan seperti URL, mention, angka, simbol, dan spasi berlebih dihapus. Selanjutnya, teks diubah menjadi huruf kecil melalui case folding dan dipecah menjadi token. Tahap stopword removal dilakukan dengan mempertahankan kata negasi seperti tidak, tak, bukan, jangan, dan belum agar makna sentimen tetap terjaga. Kata ok dan boleh juga dipertahankan karena memiliki makna dalam konteks ulasan. Tahap terakhir adalah stemming menggunakan Sastrawi untuk mengubah kata menjadi bentuk dasarnya.

Hasil preprocessing kemudian digunakan sebagai input pada tahap pembentukan fitur menggunakan TF-IDF sebelum proses klasifikasi Multinomial Naïve Bayes.

2.5 Stratified K-Fold Cross-Validation

Stratified K-Fold Cross-Validation digunakan untuk membagi dataset menjadi beberapa subset (fold) dengan tetap mempertahankan proporsi masing-masing kelas sentimen. Penelitian ini menggunakan 5-Fold Cross-Validation dengan proses stratified agar distribusi kelas Negatif, Netral, dan Positif tetap relatif seimbang pada data latih dan data uji. Penggunaan 5-Fold Cross-Validation juga diterapkan dalam penelitian Situngkir et al. (2026) untuk mengevaluasi model analisis sentimen.

Pada setiap fold, empat bagian data digunakan sebagai data latih dan satu bagian sebagai data uji. Proses pembagian dilakukan secara bergantian hingga seluruh data memperoleh kesempatan sebagai data uji. TF-IDF dan SMOTE diterapkan hanya pada data latih di setiap fold untuk mencegah data leakage, sedangkan data uji digunakan dalam kondisi asli untuk mengevaluasi kinerja model.

2.6 Pembobotan Fitur TF-IDF

Ekstraksi fitur dilakukan menggunakan Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah teks hasil preprocessing menjadi representasi numerik yang dapat digunakan oleh algoritma Multinomial Naïve Bayes. TF-IDF memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kemunculannya pada keseluruhan dokumen. Penggunaan TF-IDF dalam analisis sentimen dengan Naïve Bayes juga diterapkan oleh Helmayanti et al. (2023) dan Amaa et al. (2025).

Pada penelitian ini, TF-IDF diterapkan pada setiap fold secara terpisah. Vectorizer hanya melakukan fit pada data latih dan selanjutnya digunakan untuk melakukan transform pada data uji. Proses tersebut dilakukan untuk mencegah data leakage dan memastikan fitur yang digunakan model berasal dari data pelatihan.

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Hasil ekstraksi fitur TF-IDF kemudian digunakan sebagai input pada dua skenario klasifikasi, yaitu MNB baseline dan MNB dengan SMOTE.

2.7 Multinomial Naïve Bayes Tanpa SMOTE

Multinomial Naïve Bayes (MNB) digunakan sebagai algoritma klasifikasi untuk menentukan kelas sentimen berdasarkan fitur TF-IDF. MNB dipilih karena sesuai untuk klasifikasi data teks dan telah digunakan dalam analisis sentimen ulasan aplikasi, termasuk Mobile JKN. Amaa et al. (2025) menerapkan TF-IDF dan MNB untuk mengklasifikasikan sentimen ulasan Mobile JKN, sedangkan Situngkir et al. (2026) juga menggunakan MNB pada data ulasan aplikasi.

Pada penelitian ini, MNB digunakan dalam dua skenario, yaitu MNB baseline tanpa SMOTE dan MNB dengan SMOTE. Model dilatih menggunakan data hasil ekstraksi TF-IDF pada masing-masing fold, kemudian diuji menggunakan data uji untuk memperoleh nilai Accuracy, Precision, Recall, dan F1-Score.

2.8 Penerapan SMOTE pada Multinomial Naïve Bayes

Synthetic Minority Over-sampling Technique (SMOTE) digunakan untuk menangani ketidakseimbangan jumlah data antarkelas dengan menghasilkan sampel sintetis pada kelas yang memiliki jumlah data lebih sedikit (Chawla et al., 2002). Penerapan SMOTE pada analisis sentimen ulasan aplikasi telah digunakan oleh Java et al. (2024), sedangkan Situngkir et al. (2026) menunjukkan bahwa ketidakseimbangan kelas dapat memengaruhi kemampuan model dalam mengenali kelas tertentu.

Dalam penelitian ini, SMOTE diterapkan pada data latih setelah proses TF-IDF pada setiap fold. Data uji tidak dilakukan oversampling dan tetap digunakan dalam kondisi aslinya. Sampel sintetis yang dihasilkan SMOTE kemudian digunakan untuk melatih MNB sehingga distribusi kelas pada data latih menjadi lebih seimbang.

Penerapan tersebut dilakukan untuk membandingkan kinerja MNB baseline dengan MNB yang menggunakan SMOTE, khususnya dalam kemampuan model mengenali kelas Netral yang memiliki jumlah data paling sedikit.

2.9 Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja Multinomial Naïve Bayes pada kedua skenario penelitian. Pengukuran dilakukan berdasarkan hasil prediksi terhadap data testing pada setiap fold. Metrik yang digunakan terdiri atas accuracy, precision, recall, dan F1-score. Penggunaan metrik tersebut juga diterapkan dalam penelitian Java et al. (2024) untuk mengevaluasi kinerja model klasifikasi sentimen.

Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap keseluruhan data. Precision menunjukkan tingkat ketepatan model dalam memberikan prediksi pada suatu kelas, sedangkan recall menunjukkan kemampuan model dalam menemukan data yang sebenarnya termasuk dalam kelas tersebut. F1-score digunakan untuk memberikan ukuran gabungan antara precision dan recall.

Selain metrik tersebut, penelitian ini menggunakan confusion matrix untuk melihat distribusi hasil klasifikasi pada masing-masing kelas sentimen. Confusion matrix dapat menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar serta kesalahan prediksi yang terjadi antara kelas positif, netral, dan negatif. Penggunaan confusion matrix dalam evaluasi analisis sentimen ulasan aplikasi juga diterapkan oleh Zulfiqri et al. (2024). Nilai evaluasi dari masing-masing fold kemudian dirangkum untuk memperoleh gambaran kinerja model secara keseluruhan. Hasil evaluasi dari model tanpa SMOTE dan model dengan SMOTE selanjutnya digunakan pada tahap analisis dan perbandingan.

2.10 Analisis dan Perbandingan Hasil

Tahap analisis dilakukan setelah seluruh proses klasifikasi dan evaluasi pada lima fold selesai. Analisis diarahkan untuk membandingkan kinerja Multinomial Naïve Bayes tanpa SMOTE dengan Multinomial Naïve Bayes menggunakan SMOTE.

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Perbandingan dilakukan berdasarkan nilai accuracy, precision, recall, dan F1-score. Selain melihat kinerja secara keseluruhan, analisis juga memperhatikan hasil pada masing-masing kelas sentimen untuk mengetahui kemampuan model dalam mengenali kelas mayoritas maupun kelas dengan jumlah data yang lebih sedikit.

Confusion matrix digunakan sebagai informasi tambahan untuk mengidentifikasi pola kesalahan klasifikasi pada masing-masing skenario. Perubahan hasil klasifikasi setelah penerapan SMOTE kemudian dianalisis untuk mengetahui apakah teknik tersebut memberikan peningkatan terhadap kemampuan model dalam menangani ketidakseimbangan kelas.

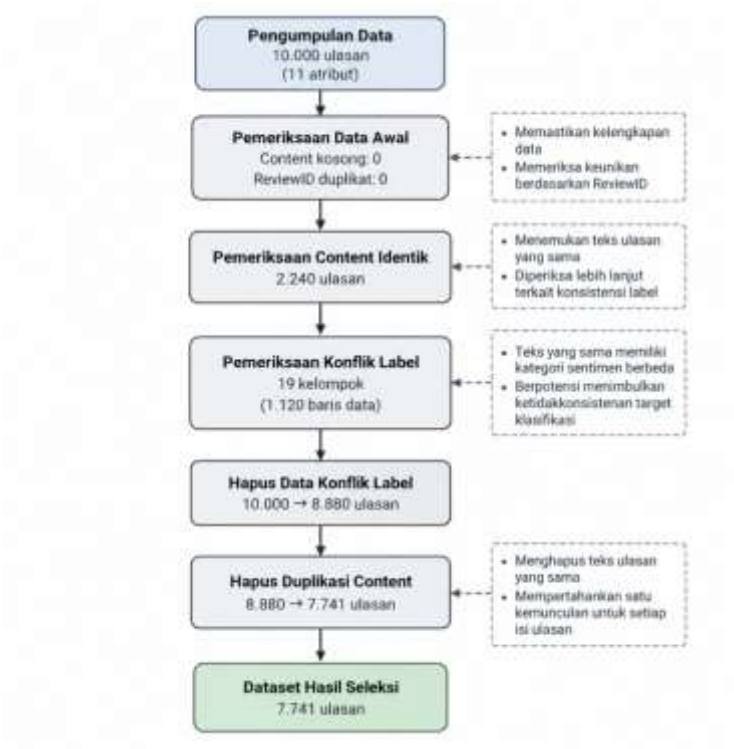
Hasil perbandingan digunakan untuk menjawab tujuan penelitian mengenai peningkatan kinerja Multinomial Naïve Bayes dalam analisis sentimen ulasan pengguna aplikasi Mobile JKN melalui penerapan SMOTE. Berdasarkan hasil tersebut, dapat ditentukan skenario model yang memberikan kinerja klasifikasi paling baik sesuai dengan hasil eksperimen.

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan dan Seleksi Data

Data penelitian diperoleh dari ulasan pengguna aplikasi Mobile JKN pada Google Play Store melalui proses web scraping. Pengumpulan data dilakukan menggunakan bahasa pemrograman Python pada Google Colab dengan library google_play_scraper. Proses tersebut menghasilkan 10.000 ulasan dengan 11 atribut. Atribut yang digunakan dalam penelitian adalah reviewId, content, score, dan at yang masing-masing digunakan untuk identitas ulasan, teks, rating, dan waktu ulasan.

Pemeriksaan awal menunjukkan seluruh data memiliki nilai pada content dan tidak terdapat duplikasi berdasarkan reviewId. Namun, ditemukan 2.240 ulasan dengan content yang identik. Pemeriksaan lebih lanjut menggunakan aturan pemetaan rating secara sementara menemukan 19 kelompok content dengan kategori sentimen berbeda yang melibatkan 1.120 baris data. Proses pemeriksaan dan seleksi data dilakukan menggunakan Python dengan library Pandas. Data dengan konflik label tersebut dihapus untuk menjaga konsistensi label pada proses klasifikasi.



Gambar 2. Alur Seleksi Dataset Ulasan Mobile JKN

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Setelah penghapusan 1.120 data konflik, diperoleh 8.880 ulasan. Selanjutnya, dilakukan penghapusan duplikasi berdasarkan content dengan mempertahankan satu kemunculan untuk setiap isi ulasan. Sebanyak 1.139 data duplikat dihapus sehingga diperoleh 7.741 ulasan sebagai dataset hasil seleksi.

Tabel 1. Tahapan Seleksi Dataset Ulasan Mobile JKN

Tahapan Seleksi	Data Awal	Data Dihapus	Data Tersisa
Pengumpulan data	–	–	10.000
Penghapusan konflik label	10.000	1.120	8.880
Penghapusan duplikasi *content*	8.880	1.139	7.741

3.2. Hasil Pelabelan Sentimen

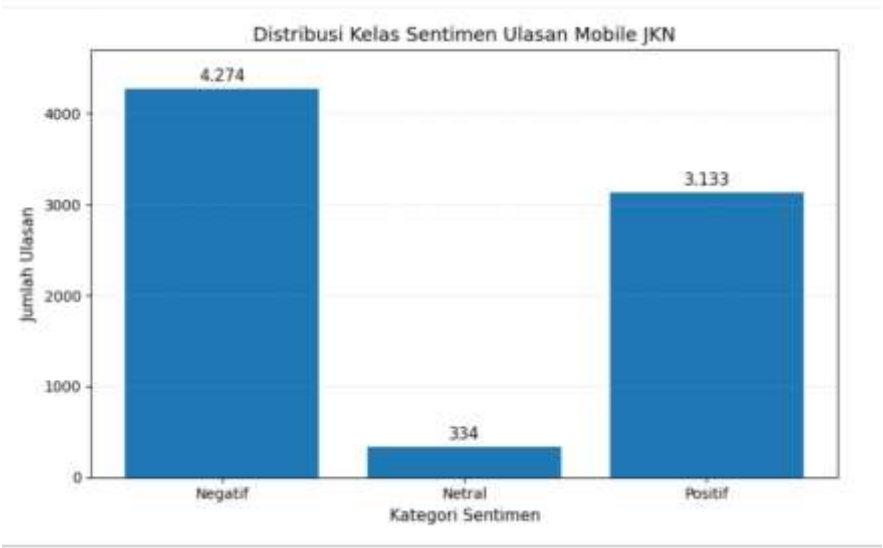
Sebanyak 7.741 ulasan hasil seleksi selanjutnya diberi label sentimen berdasarkan rating pengguna pada Google Play Store. Rating 1–2 dikategorikan sebagai Negatif, rating 3 sebagai Netral, sedangkan rating 4–5 sebagai Positif. Pelabelan ini digunakan untuk membentuk tiga kelas sentimen yang menjadi target dalam proses klasifikasi.

Tabel 2. Distribusi Kelas Sentimen Ulasan Mobile JKN

Kategori Sentimen	Jumlah Ulasan	Persentase
Negatif	4.274	55,21%
Netral	334	4,31%
Positif	3.133	40,47%
Total	7.741	100,00%

Berdasarkan Tabel 2, kelas Negatif merupakan kelas terbesar dengan 4.274 ulasan (55,21%), diikuti kelas Positif sebanyak 3.133 ulasan (40,47%). Sementara itu, kelas Netral hanya berjumlah 334 ulasan (4,31%). Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas yang cukup besar, terutama pada kelas Netral. Kondisi ini menjadi pertimbangan dalam tahap pemodelan karena model berpotensi lebih banyak mempelajari pola dari kelas dengan jumlah data yang lebih besar.

Untuk memperjelas perbedaan jumlah pada setiap kelas, distribusi sentimen dihitung menggunakan Python dengan library Pandas dan divisualisasikan menggunakan Matplotlib. Hasil visualisasi distribusi kelas sentimen ditampilkan pada Gambar 3.



Gambar 3. Distribusi Kelas Sentimen Ulasan Mobile JKN

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

3.3. Hasil Preprocessing Data

Preprocessing dilakukan terhadap 7.741 ulasan hasil seleksi untuk mengubah teks menjadi bentuk yang lebih terstruktur sebelum digunakan dalam proses klasifikasi. Seluruh proses preprocessing dilakukan menggunakan bahasa pemrograman Python pada Google Colab. Tahapan preprocessing meliputi cleaning, case folding, tokenization, stopword removal, dan stemming. Pada tahap cleaning, dilakukan penghapusan URL, mention, angka, simbol, karakter khusus, dan spasi berlebih serta normalisasi Unicode. Selanjutnya, case folding mengubah seluruh teks menjadi huruf kecil, sedangkan tokenization memecah teks menjadi token atau kata. Tahap stopword removal dilakukan dengan tetap mempertahankan kata negasi seperti “tidak”, “tak”, “bukan”, “jangan”, dan “belum” serta kata “ok” dan “boleh” karena dianggap relevan terhadap isi ulasan. Tahap terakhir adalah stemming menggunakan library Sastrawi untuk mengubah kata berimbuhan menjadi bentuk dasar.

Tabel 3. Tahapan Processing Data

Tahap	Proses
Cleaning	Menghapus URL, mention, angka, simbol, karakter khusus, dan spasi berlebih serta melakukan normalisasi Unicode
Case folding	Mengubah seluruh huruf menjadi huruf kecil
Tokenization	Memecah teks menjadi token atau kata
Stopword removal	Menghapus kata yang kurang informatif dengan mempertahankan kata negasi dan kata tertentu yang relevan
Stemming	Mengubah kata berimbuhan menjadi bentuk dasar menggunakan Sastrawi

Hasil preprocessing menunjukkan adanya perubahan bentuk teks menjadi lebih seragam. Sebagai contoh, ulasan “Sangat membantu banget” berubah menjadi “sangat bantu banget”, sedangkan “mudah digunakan dan cepat” berubah menjadi “mudah guna cepat”. Hasil perubahan teks tersebut diperoleh dari proses preprocessing menggunakan Python dan digunakan sebagai data terproses untuk tahap berikutnya.

Tabel 4. Contoh Hasil Preprocessing Ulasan

No.	**Teks Asli**	**Teks Setelah Preprocessing**	**Sentimen**
1	Sangat membantu banget	sangat bantu banget	Positif
2	apk jelek.. dikit dikit keluar nanti masuk lagi.. ribet banget... tolong lah diperbaiki	apk jelek dikit dikit keluar masuk ribet banget lah baik	Negatif
3	mudah digunakan dan cepat	mudah guna cepat	Positif
4	Ok	ok	Positif
5	masih ok saat ini	masih ok saat ini	Positif

Data yang hanya menghasilkan emoji atau simbol tidak dapat digunakan dalam proses klasifikasi karena tidak menghasilkan token berupa kata. Dari 7.741 ulasan hasil seleksi, terdapat 15 ulasan dengan kondisi tersebut sehingga dikeluarkan dari dataset pemodelan. Proses pemeriksaan dan penyaringan data dilakukan menggunakan Python dengan library Pandas. Dengan demikian, diperoleh 7.726 ulasan yang digunakan pada tahap pemodelan.

Tabel 5. Distribusi Dataset Setelah Preprocessing

Kategori Sentimen	Jumlah	Persentase
Negatif	4.273	55,31%
Netral	334	4,32%
Positif	3.119	40,37%
Total	7.726	100,00%

Berdasarkan Tabel 5, distribusi kelas masih tidak seimbang setelah preprocessing. Kelas Negatif merupakan kelas terbesar dengan 4.273 ulasan (55,31%), sedangkan kelas Netral hanya berjumlah 334 ulasan (4,32%). Kondisi tersebut menunjukkan bahwa masalah ketidakseimbangan kelas masih terdapat pada dataset yang akan digunakan untuk pemodelan.

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Selain distribusi kelas, panjang teks dianalisis untuk mengetahui karakteristik ulasan yang digunakan dalam pemodelan. Statistik panjang teks dihitung menggunakan Python dengan library Pandas berdasarkan jumlah token pada setiap ulasan. Hasil analisis menunjukkan rata-rata panjang teks sebesar 10,97 token, median 7 token, dan maksimum 87 token. Sebanyak 75% data memiliki panjang teks tidak lebih dari 15 token, sehingga sebagian besar ulasan memiliki panjang yang relatif pendek.

Tabel 6. Statistik Panjang Teks Setelah Preprocessing

Statistik	Nilai
Jumlah data	7.726
Rata-rata	10,97 token
Median	7 token
Minimum	1 token
Kuartil 1 (25%)	3 token
Kuartil 3 (75%)	15 token
Maksimum	87 token

Berdasarkan hasil preprocessing tersebut, diperoleh dataset akhir sebanyak 7.726 ulasan yang telah memiliki teks terproses dan siap digunakan pada tahap ekstraksi fitur. Data selanjutnya direpresentasikan dalam bentuk bobot Term Frequency–Inverse Document Frequency (TF-IDF) pada tahap berikutnya.

3.4. Hasil Ekstraksi Fitur TF-IDF

Sebanyak 7.726 ulasan hasil preprocessing selanjutnya diubah menjadi bentuk numerik menggunakan Term Frequency–Inverse Document Frequency (TF-IDF). Proses ekstraksi fitur dilakukan menggunakan bahasa pemrograman Python pada Google Colab dengan fungsi TfidfVectorizer dari library Scikit-learn. TF-IDF memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu ulasan dan tingkat kemunculannya pada keseluruhan dokumen. Hasil ekstraksi digunakan sebagai representasi numerik yang menjadi masukan bagi algoritma Multinomial Naïve Bayes.

Ekstraksi fitur dilakukan menggunakan Stratified 5-Fold Cross-Validation dengan mempertahankan proporsi kelas pada setiap fold. Pada setiap fold, TfidfVectorizer hanya dilakukan fit terhadap data training, sedangkan data testing ditransformasikan menggunakan vocabulary yang diperoleh dari data training. Prosedur tersebut dilakukan untuk mencegah data leakage sehingga informasi dari data pengujian tidak digunakan dalam pembentukan fitur.

Hasil ekstraksi fitur menunjukkan jumlah fitur yang berbeda pada setiap fold, yaitu 4.799 fitur pada Fold 1, 4.840 pada Fold 2, 4.790 pada Fold 3, 4.755 pada Fold 4, dan 4.781 pada Fold 5, dengan rata-rata 4.793 fitur. Perbedaan jumlah fitur terjadi karena vocabulary dibentuk berdasarkan data training pada masing-masing fold, sehingga kosakata yang terbentuk dapat berbeda.

```
================================================================================
RINGKASAN TF-IDF
================================================================================
 Fold  Train  Test  Fitur
    1   6180  1546   4799
    2   6181  1545   4840
    3   6181  1545   4790
    4   6181  1545   4755
    5   6181  1545   4781

Rata-rata jumlah fitur:
4793.0
```

Gambar 4. Hasil Ekstraksi Fitur TF-IDF pada Stratified 5-Fold Cross-Validation

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Fitur TF-IDF yang dihasilkan selanjutnya digunakan sebagai input untuk dua skenario klasifikasi, yaitu MNB baseline tanpa SMOTE dan MNB dengan SMOTE.

3.5. Hasil Klasifikasi Multinomial Naïve Bayes (MNB Baseline)

Sebanyak 7.726 data hasil preprocessing digunakan untuk pengujian menggunakan algoritma Multinomial Naïve Bayes (MNB) tanpa penerapan SMOTE sebagai skenario baseline. Seluruh eksperimen dilakukan menggunakan bahasa pemrograman Python pada Google Colab dengan library Scikit-learn. Pembagian data dilakukan menggunakan StratifiedKFold sebanyak 5 fold, sedangkan proses klasifikasi dilakukan menggunakan MultinomialNB. Pada setiap fold, data training digunakan untuk membentuk model, sedangkan data testing digunakan untuk memperoleh hasil prediksi.

Nilai accuracy, precision, recall, dan F1-score diperoleh dari hasil prediksi menggunakan fungsi classification_report pada modul sklearn.metrics. Hasil evaluasi dari lima fold kemudian dirangkum untuk memperoleh nilai rata-rata kinerja MNB baseline.

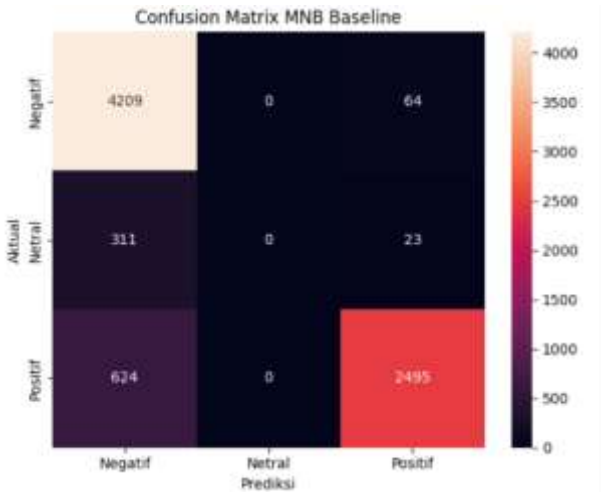
```
================================================================================
RINGKASAN MNB BASELINE PER FOLD
================================================================================
 Fold Accuracy Precision_Macro Recall_Macro F1_Macro
    1   0.8616          0.5927       0.5899   0.5852
    2   0.8777          0.6025       0.6021   0.5972
    3   0.8686          0.5935       0.5963   0.5904
    4   0.8654          0.5931       0.5935   0.5882
    5   0.8654          0.5925       0.5931   0.5877

================================================================================
RATA-RATA MNB BASELINE
================================================================================
Accuracy rata-rata  : 0.8677
Precision rata-rata : 0.5949
Recall rata-rata    : 0.5950
F1-Score rata-rata  : 0.5897
```

Gambar 5. Hasil Klasifikasi MNB Baseline pada Stratified 5-Fold Cross-Validation

Berdasarkan hasil pengujian, MNB baseline memperoleh rata-rata accuracy sebesar 86,77%, macro precision sebesar 59,49%, macro recall sebesar 59,50%, dan macro F1-score sebesar 58,97%. Nilai tersebut merupakan hasil perhitungan evaluasi dari lima fold menggunakan fungsi classification_report pada Scikit-learn.

Untuk melihat kinerja model pada masing-masing kelas secara lebih rinci, hasil prediksi dari seluruh data testing pada lima fold digabungkan menjadi Out-of-Fold (OOF) prediction. OOF prediction kemudian dibandingkan dengan label aktual menggunakan fungsi confusion_matrix dari modul sklearn.metrics. Hasil perhitungan tersebut divisualisasikan dalam bentuk heatmap menggunakan Python.



Gambar 6. Confusion Matrix MNB Baseline

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Berdasarkan Gambar 6, MNB baseline mampu mengklasifikasikan sebagian besar data pada kelas Negatif dan Positif, tetapi belum mampu mengidentifikasi kelas Netral. Dari 334 ulasan Netral, tidak terdapat data yang diklasifikasikan secara tepat sebagai Netral; sebanyak 311 data diklasifikasikan sebagai Negatif dan 23 data sebagai Positif. Kondisi tersebut menghasilkan recall kelas Netral sebesar 0%.

Temuan tersebut menunjukkan bahwa accuracy sebesar 86,77% belum sepenuhnya menggambarkan kemampuan model dalam menangani seluruh kelas secara seimbang. Kegagalan mengenali kelas Netral menjadi dasar untuk menguji penerapan SMOTE pada tahap berikutnya.

3.6 Hasil Klasifikasi Multinomial Naïve Bayes dengan SMOTE

Berdasarkan hasil MNB baseline, kelas Netral belum dapat diidentifikasi dengan baik, sehingga dilakukan eksperimen menggunakan Synthetic Minority Over-sampling Technique (SMOTE) pada data training. Seluruh eksperimen dilakukan menggunakan bahasa pemrograman Python pada Google Colab. Penerapan SMOTE menggunakan fungsi SMOTE dari library imbalanced-learn, sedangkan proses klasifikasi dan evaluasi menggunakan library Scikit-learn. SMOTE diterapkan secara terpisah pada data training di setiap fold, sedangkan data testing tetap menggunakan distribusi aslinya.

Pada Fold 1, data training sebelum SMOTE terdiri atas 3.418 data Negatif, 267 data Netral, dan 2.495 data Positif. Setelah penerapan SMOTE, jumlah setiap kelas menjadi 3.418 data, sehingga total data training meningkat dari 6.180 menjadi 10.254 data. Proses yang sama diterapkan pada fold lainnya dengan jumlah kelas mayoritas sebagai acuan oversampling.

Data training hasil SMOTE selanjutnya digunakan untuk melatih model MultinomialNB. Hasil prediksi terhadap data testing pada setiap fold dievaluasi menggunakan fungsi classification_report dari modul sklearn.metrics. Hasil evaluasi MNB baseline dan MNB dengan SMOTE kemudian dirangkum dan disajikan pada Gambar 7.

```
================================================================================
CLASSIFICATION REPORT — MNB BASELINE
================================================================================
              precision    recall  f1-score   support

     Negatif     0.8182    0.9850    0.8939      4273
      Netral     0.0000    0.0000    0.0000       334
     Positif     0.9663    0.7999    0.8753      3119

    accuracy                         0.8677      7726
   macro avg     0.5948    0.5950    0.5897      7726
weighted avg     0.8426    0.8677    0.8477      7726

================================================================================
CLASSIFICATION REPORT — MNB + SMOTE
================================================================================
              precision    recall  f1-score   support

     Negatif     0.8529    0.7285    0.7858      4273
      Netral     0.0900    0.3922    0.1465       334
     Positif     0.9546    0.8022    0.8718      3119

    accuracy                         0.7437      7726
   macro avg     0.6325    0.6410    0.6013      7726
weighted avg     0.8610    0.7437    0.7929      7726

================================================================================
PERBANDINGAN KINERJA PER KELAS
================================================================================
Sentimen Baseline_Precision Baseline_Recall Baseline_F1 SMOTE_Precision SMOTE_Recall SMOTE_F1
 Negatif            0.8182          0.9850      0.8939          0.8529       0.7285   0.7858
  Netral            0.0000          0.0000      0.0000          0.0900       0.3922   0.1465
 Positif            0.9663          0.7999      0.8753          0.9546       0.8022   0.8718
```

Gambar 7. Hasil Classification Report MNB Baseline dan MNB dengan SMOTE

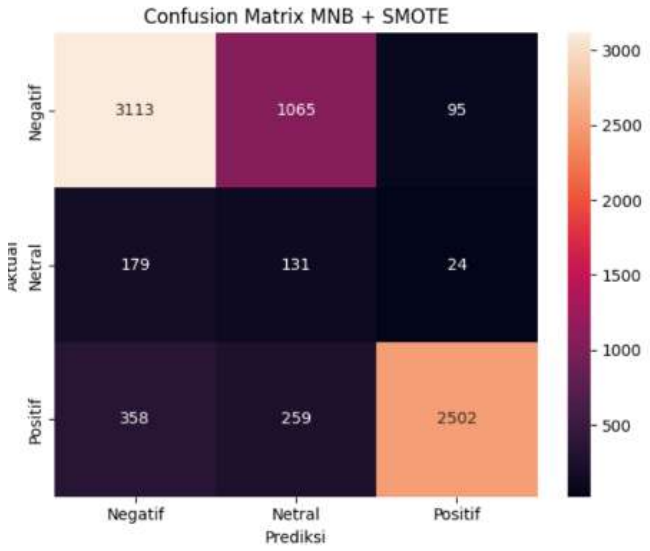
Hasil pengujian menunjukkan bahwa MNB dengan SMOTE memperoleh accuracy 74,37%, macro precision 63,25%, macro recall 64,10%, dan macro F1-score 60,13%. Dibandingkan MNB baseline, accuracy mengalami penurunan sebesar 12,40 poin persentase, sedangkan macro precision, macro recall, dan macro F1-score masing-masing meningkat sebesar 3,77, 4,60, dan 1,16 poin persentase.

Pada tingkat kelas, perubahan paling signifikan terjadi pada kelas Netral. MNB baseline menghasilkan recall 0%, sedangkan setelah penerapan SMOTE recall meningkat menjadi 39,22%, dengan precision 9,00% dan F1-score

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

14,65%. Sebaliknya, recall kelas Negatif menurun dari 98,50% menjadi 72,85%, sedangkan kinerja kelas Positif relatif stabil.

Untuk melihat perubahan distribusi hasil prediksi pada masing-masing kelas, hasil prediksi testing dari lima fold digabungkan menjadi OOF prediction. Selanjutnya, OOF prediction dibandingkan dengan label aktual menggunakan fungsi confusion_matrix dari modul sklearn.metrics. Hasil perhitungan kemudian divisualisasikan dalam bentuk heatmap menggunakan Python dan disajikan pada Gambar 8.



Gambar 8. Confusion Matrix MNB dengan SMOTE

Berdasarkan Gambar 8, sebanyak 131 dari 334 data Netral berhasil diklasifikasikan secara tepat sebagai Netral. Peningkatan tersebut menunjukkan bahwa SMOTE memberikan pengaruh terhadap kemampuan model dalam mengenali kelas minoritas. Namun, sebanyak 1.065 data Negatif diklasifikasikan sebagai Netral, yang berkontribusi terhadap penurunan recall kelas Negatif.

Secara keseluruhan, penerapan SMOTE menghasilkan trade-off antara accuracy dan kemampuan model dalam menangani kelas minoritas. SMOTE tidak meningkatkan accuracy, tetapi meningkatkan macro precision, macro recall, macro F1-score, serta recall kelas Netral. Dengan demikian, penerapan SMOTE menghasilkan kinerja yang lebih seimbang antar kelas, khususnya dalam mengenali kelas Netral.

3.7 Perbandingan dan Pembahasan Hasil

Penelitian ini membandingkan dua skenario, yaitu Multinomial Naïve Bayes (MNB) sebagai baseline dan MNB dengan penerapan Synthetic Minority Over-sampling Technique (SMOTE). Kedua model diuji menggunakan Stratified 5-Fold Cross-Validation dengan dataset dan pembagian fold yang sama. Rekapitulasi hasil eksperimen ditampilkan pada Gambar 9.

```
================================================================================
REKAP HASIL EKSPERIMEN FINAL
================================================================================
       Model  Accuracy  Macro_Precision  Macro_Recall  Macro_F1
MNB Baseline     86.77            59.48          59.5     58.97
 MNB + SMOTE     74.37            63.25          64.1     60.13
```

Gambar 9. Rekap Hasil Eksperimen MNB Baseline dan MNB dengan SMOTE

Berdasarkan Gambar 8, MNB baseline memperoleh accuracy sebesar 86,77%, macro precision 59,48%, macro recall 59,50%, dan macro F1-score 58,97%. Sementara itu, MNB dengan SMOTE memperoleh accuracy 74,37%, macro precision 63,25%, macro recall 64,10%, dan macro F1-score 60,13%.

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Penerapan SMOTE menyebabkan accuracy menurun sebesar 12,40 poin persentase, tetapi meningkatkan macro precision, macro recall, dan macro F1-score. Peningkatan tersebut terutama terlihat pada kelas Netral, dengan recall meningkat dari 0,00% menjadi 39,22%. Artinya, SMOTE membuat model mulai mampu mengenali kelas minoritas yang sebelumnya tidak berhasil dikenali oleh MNB baseline.

Namun, peningkatan tersebut disertai penurunan recall kelas Negatif dari 98,50% menjadi 72,85% karena sebagian data Negatif diprediksi sebagai Netral. Dengan demikian, SMOTE dalam penelitian ini tidak meningkatkan accuracy, tetapi memberikan kinerja yang lebih seimbang antar kelas, terutama dalam mengenali kelas Netral.

Secara keseluruhan, hasil eksperimen menunjukkan bahwa efektivitas SMOTE bergantung pada metrik yang digunakan. MNB baseline lebih unggul dalam accuracy, sedangkan MNB dengan SMOTE lebih baik berdasarkan metrik macro dan kemampuan mengenali kelas minoritas.

4. Kesimpulan

Berdasarkan hasil penelitian, penerapan Multinomial Naïve Bayes pada analisis sentimen ulasan aplikasi Mobile JKN menunjukkan bahwa ketidakseimbangan kelas berpengaruh terhadap kemampuan model dalam mengenali setiap kategori sentimen. Dataset pemodelan terdiri atas 7.726 ulasan dengan distribusi 4.273 ulasan negatif, 334 ulasan netral, dan 3.119 ulasan positif. Kondisi tersebut menyebabkan model Multinomial Naïve Bayes tanpa SMOTE memiliki accuracy yang relatif tinggi sebesar 86,77%, tetapi belum mampu mengenali kelas Netral dengan recall sebesar 0%. Penerapan SMOTE menghasilkan accuracy sebesar 74,37%, sehingga mengalami penurunan sebesar 12,40 poin persentase dibandingkan model baseline. Namun, penerapan SMOTE meningkatkan macro precision dari 59,48% menjadi 63,25%, macro recall dari 59,50% menjadi 64,10%, dan macro F1-score dari 58,97% menjadi 60,13%. Peningkatan paling signifikan terjadi pada kelas Netral, dengan recall meningkat dari 0% menjadi 39,22%. Hasil tersebut menunjukkan bahwa SMOTE tidak memberikan peningkatan pada accuracy keseluruhan, tetapi mampu meningkatkan kemampuan model dalam mengenali kelas minoritas dan menghasilkan kinerja yang lebih seimbang antar kelas. Dengan demikian, efektivitas SMOTE pada penelitian ini tidak tepat dinilai hanya berdasarkan accuracy. Pemilihan skenario model perlu mempertimbangkan tujuan klasifikasi dan kebutuhan terhadap kemampuan model dalam mengenali seluruh kelas sentimen. Untuk penelitian selanjutnya, pengembangan dapat diarahkan pada penggunaan strategi penyeimbangan data lainnya, optimasi parameter Multinomial Naïve Bayes, serta pendekatan representasi teks yang mampu menangkap konteks bahasa Indonesia secara lebih baik.

Referensi

1. Ali, A., & Kadafi, M. (2025). Analisis kepuasan pengguna terhadap fitur REHAB pada aplikasi Mobile JKN dengan menggunakan metode UEQ. TIN: Terapan Informatika Nusantara, 5(9), 569–577. https://doi.org/10.47065/tin.v5i9.7024
2. Amaa, R. P., Ali, I., Rahaningsih, N., & Prihartono, W. (2025). Kinerja Multinomial Naïve Bayes pada analisis sentimen ulasan pengguna aplikasi Mobile JKN. J-ENSITEC: Journal of Engineering and Sustainable Technology, 12(1), 10338–10345. https://doi.org/10.31949/j-ensitec.v12i01.16656
3. Apriliyanti, P. N., Dasuki, M., & Rahman, M. (2026). Klasifikasi sentimen positif dan negatif ulasan aplikasi GetContact dengan algoritma Naïve Bayes. JUSTIFY: Jurnal Sistem Informasi Ibrahimy, 4(2), 123–129. https://doi.org/10.35316/justify.v4i2.9133
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
5. Helmayanti, S. A., Hamami, F., & Fa'rifah, R. Y. (2023). Penerapan algoritma TF-IDF dan Naïve Bayes untuk analisis sentimen berbasis aspek ulasan aplikasi Flip pada Google Play Store. Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 4(3), 1822–1834. https://doi.org/10.35870/jimik.v4i3.415
6. Java, M. A., Syafrullah, M., Windarto, W., & Painem, P. (2024). Analisis sentimen ulasan pengguna aplikasi Threads pada Google Play Store menggunakan Multinomial Naive Bayes dan Support Vector Machine. Jurnal Ticom: Technology of Information and Communication, 12(2), 75–80. https://doi.org/10.70309/ticom.v12i2.112
7. Kurniawan, Y. I., Sidiq, R. H., Widadi, A. D. U., Yubiharto, A. R. A., Gibran, A. K., Aditama, M. R., & Laksono, F. A. T. (2025). Naive Bayes classifier with SMOTE for sentiment analysis of Blibli app reviews on the Google Play Store. Jurnal Penelitian Inovatif, 5(3), 2675–2688. https://doi.org/10.54082/jupin.1842
8. Prasetyo, T. A., Sulistiyowati, N., & Voutama, A. (2026). Evaluasi layanan Mobile JKN berbasis 100.000 ulasan Google Play Store dan COBIT 5. Jurnal Informatika dan Teknik Elektro Terapan, 14(2), 662–668. https://doi.org/10.23960/jitet.v14i2.9384
9. Ramdani, B., Saputra, A. D., & Zahir, M. R. A. (2024). Analisis sentimen terhadap ulasan aplikasi pinjaman online (PINJOL) di Google Play Store menggunakan Naive Bayes Classifier. Jurnal Riset Informatika dan Teknologi Informasi, 1(2), 61–64. https://doi.org/10.58776/jriti.v1i2.39
10. Safitri, A., Risaldi, M., Alwi, M. N. R., Surianto, D. F., Fadilah, N., & Parenreng, J. M. (2026). Improving sentiment classification of Kredit Pintar reviews using IndoBERT, SMOTE, and stacking ensemble. Jurnal Teknik Informatika (JUTIF), 7(3), 2411–2424. https://doi.org/10.52436/1.jutif.2026.7.3.5342
11. Shafira, F., & Nugraha, A. H. (2025). Sentiment analysis of Netflix app reviews on Google Play Store using the Naive Bayes. JITAR:

DOI: https://doi.org/10.69693/ijmst.v4i3.13994

Journal of Information Technology and Applications Research, 1(2), 1–17. https://doi.org/10.63956/jitar.v1i2.31
12. Situngkir, D. D. T., Anita, & Purba, D. L. (2026). Klasifikasi sentimen ulasan GoFood di Google Play Store dengan metode Naive Bayes. Jurnal Algoritma, 23(1), 1589–1600. https://doi.org/10.33364/algoritma/v.23-1.3479
13. Suhaimi, N. B. A., & Lestari, M. (2024). Sentiment analysis of TikTok app reviews on Google Play using several machine learning methods. International Journal of Global Operations Research, 5(4), 275–287. https://doi.org/10.47194/ijgor.v5i4.343
14. Tamami, G., Triyanto, W. A., & Muzid, S. (2025). Sentiment analysis Mobile JKN reviews using SMOTE based LSTM. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 19(1), 13–24. https://doi.org/10.22146/ijccs.101910
15. Zulfiqri, R., Sari, B. N., & Padilah, T. N. (2024). Analisis sentimen ulasan pengguna aplikasi media sosial Instagram pada situs Google Play Store menggunakan Naïve Bayes Classifier. Jurnal Informatika dan Teknik Elektro Terapan, 12(3), 2965–2973. https://doi.org/10.23960/jitet.v12i3.4995

DOI: https://doi.org/10.69693/ijmst.v4i3.13994