Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 15737- 15746
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Sentimen Wattpad di Play Store Menggunakan Naïve Bayes Berbasis TF-IDF dan SMOTE

Farah Diba Azkia[1], Agus Junaidi[2], Arif Ismail Husin[3]
[1-3]Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
farohafalasi12@gmail.com, agus.asj@bsi.ac.id, arif.aal@bsi.ac.id

Abstrak

Kajian ini menguji performa klasifikasi Multinomial Naive Bayes yang diintegrasikan dengan skema pembobotan TF-IDF untuk memilah ulasan aplikasi Wattpad ke dalam kategori Positif, Negatif, dan Netral, sembari menakar kontribusi Synthetic Minority Oversampling Technique (SMOTE) dalam memitigasi ketimpangan jumlah data. Sebanyak 2.829 ulasan valid hasil web scraping melalui library google-play-srcaper diproses melalui tahapan krusial, meliputi pembersihan teks, case folding, konversi kata tidak baku, tokenisasi, eliminasi 773 stopword, hingga stemming dengan PySastrawi. Transformasi data menjadi vektor TF-IDF dengan batasan max_features=5.000 serta ngram_range=(1,2) menghasilkan matriks berdimensi 2.829 × 4.499. Melalui skema pembagian stratified split 80:20, teknik SMOTE diaplikasikan pada sektor data latih untuk mendongkrak jumlah sampel dari 2.263 menjadi 4.056 agar komposisi antar kelas lebih proporsional. Hasil observasi memperlihatkan pergeseran metrik performa yang cukup signifikan pasca-intervensi SMOTE. Sebelum dilakukan penyeimbangan data, model memang mencatat akurasi 68,37% dengan presisi 62,83% dan F1-score 60,24%, namun sistem menunjukkan kelemahan fatal yakni kegagalan total dalam mengidentifikasi kelas Netral. Setelah SMOTE diterapkan, kendati terjadi penurunan akurasi ke angka 58,83%, model justru menunjukkan ketangguhan lebih baik melalui peningkatan presisi menjadi 65,25% dan F1-score ke level 61,22%, serta lonjakan recall pada kelas Netral dari titik nol menjadi 0,43. Secara keseluruhan, pemetaan sentimen didominasi oleh opini Negatif sebesar 59,51%, disusul opini Positif 26,01%, dan Netral 14,48%, di mana mayoritas keluhan pengguna berhulu pada persoalan teknis aplikasi seperti lonjakan iklan, hambatan login, serta rendahnya stabilitas sistem.

Kata Kunci: *Analisis Sentimen, Naive Bayes, TF-IDF, SMOTE, Wattpad, Google Play Store, Imbalance Class*

1. Latar Belakang

Google Play Store menjadi salah satu ekosistem distribusi perangkat lunak digital yang paling dominan, di mana para penggunanya dapat menyampaikan pendapat dalam bentuk ulasan teks maupun penilaian bintang terhadap aplikasi yang sudah pernah mereka coba [1]. Seiring dengan tren membaca digital yang terus meningkat, aplikasi yang bergerak di segmen literasi digital, khususnya yang berbasis cerita dan novel daring, mengalami peningkatan perhatian yang signifikan di platform tersebut. Yang paling menonjol di antaranya adalah Wattpad, sebuah komunitas daring yang memungkinkan penggunanya untuk menjelajahi, menulis, dan berbagi cerita dalam berbagai genre, baik fiksi maupun nonfiksi [1]. Berdasarkan data Similarweb, per September 2023 Wattpad menduduki peringkat kedua pada kategori Buku dan Literatur di tingkat global, dan Indonesia tercatat sebagai negara ketiga dengan pengunjung terbanyak sebesar 6,45% [1]. Tingginya jumlah pengguna aktif Wattpad dari Indonesia menjadikan koleksi ulasan yang terdapat di Google Play Store sebagai sumber data opini yang bernilai tinggi, kaya informasi, dan sangat layak untuk dijadikan objek kajian analisis sentimen. Dalam penerapannya pada ulasan teks, analisis sentimen bertugas mengelompokkan polaritas dokumen agar diketahui apakah opini pengguna bermakna positif, negatif, dan netral [2].

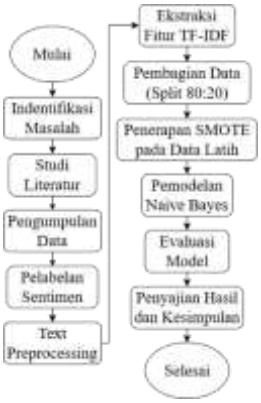
Analisis sentimen pada dasarnya merupakan serangkaian proses berbasis komputasi yang bertujuan menggali, mengolah, dan menginterpretasikan data berupa teks secara otomatis, sehingga dapat diketahui pandangan maupun sikap pengguna terhadap suatu produk atau layanan tertentu. Sejumlah penelitian terdahulu telah membuktikan keefektifan *Naive Bayes* dalam mengklasifikasikan sentimen ulasan aplikasi. Algoritma *Naive Bayes* dipilih lantaran termasuk dalam kategori metode klasifikasi teks yang sudah terbukti efisien, dengan proses implementasi yang relatif sederhana dan waktu komputasi yang singkat [3]. Setiap tahapan komputasi ini saling berkaitan dan sangat memengaruhi kualitas hasil akhir dari model sentimen yang dibangun [4]. Guna memperkuat kualitas representasi fitur teks, *Naive Bayes* kerap dipadukan dengan teknik pembobotan TF-IDF; dalam studi ulasan aplikasi misalnya, penggabungan kedua metode ini menghasilkan nilai akurasi rata-rata 0,84 untuk klasifikasi ulasan berdasarkan dimensi kecepatan layanan, keamanan transaksi, dan besaran biaya [5]. Selain tantangan distribusi data, kompleksitas linguistik ulasan berbahasa Indonesia di Google Play Store menjadi hambatan

signifikan yang memerlukan ketelitian tinggi. Pengguna bahasa yang tidak baku, singkatan kasual, hingga istilah slang menuntut prosedur *preprocessing* yang sangat spesifik. Untuk menjamin akurasi representasi fitur numerik, rangkaian prosedur sistematis dijalankan mulai dari pembersihan *noise*, standardisasi *case folding*, translasi 100 entri slang, segmentasi token, eliminasi 733 *stopword* yang dianggap redundan, hingga proses *stemming* melalui PySastrawi. Karena ulasan yang berjumlah sangat besar dan tidak teratur, diperlukan metodologi yang efektif untuk mendeteksi dan menganalisis ulasan pengguna terhadap suatu aplikasi secara otomatis [6].

Perbandingan riset ini dengan studi terdahulu yang digagas oleh Adhan dkk. [1], tampak pada perbedaan metode serta cakupan klasifikasi. Jika sebelumnya pengujian Random Forest hanya membagi sentimen ke dalam dua polaritas dengan kecenderungan dominasi kelas Positif sebesar 64,2%, investigasi ini justru mengeksplorasi tiga klasifikasi yaitu positif, negatif, dan netral, dengan temuan awal bahwa sentimen negatif justru menguasai dataset sebesar 59,51%. Ketimpangan proporsi antar kelas merupakan kendala klasik dalam machine learning yang menyebabkan model sering kali bias terhadap kelas mayoritas dan mengabaikan kelas minoritas [7]. Sebagai solusi penyeimbangan, teknik SMOTE diintegrasikan guna mendiversifikasi data minoritas melalui penciptaan sampel sintetis berbasis interpolasi antar tetangga terdekat, sehingga model mendapatkan representasi kelas yang jauh lebih berimbang sebelum memasuki fase pembelajaran [7]. Informasi kepemilikan korporasi digital [8] serta keaslian data resmi di halaman unduhan [9] menjadi pijakan kontekstual yang kuat dalam melihat dinamika platform Wattpad. Sinergi antara *Naive Bayes*, TF-IDF, dan SMOTE dalam konteks klasifikasi tiga sentimen ulasan Wattpad ini menjadi celah kebaruan yang belum tersentuh oleh studi-studi sebelumnya.

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini dibagi ke dalam beberapa tahapan sistematis, mulai dari pengumpulan data, pembersihan data (*preprocessing*), ekstraksi fitur, hingga tahap klasifikasi dan evaluasi model. Alur urutan tahapan tersebut digambarkan secara jelas pada kerangka penelitian di Gambar 1.



Gambar 1. Kerangka Penelitian

2.1. Data Penelitian dan Pelabelan Otomatis

Penelitian ini menerapkan metode kuantitatif eksperimental untuk menganalisis sentimen ulasan pengguna aplikasi Wattpad dengan identitas paket wp.wattpad di platform Google Play Store. Pengumpulan data sekunder dilakukan melalui teknik *web scraping* menggunakan pustaka google-play-scraper berbasis bahasa pemrograman Python pada lingkungan kerja Google Colaboratory [10]. Data ditarik berdasarkan ulasan terbaru berbahasa Indonesia dalam rentang waktu antara 1 Oktober 2025 hingga 31 Maret 2026 untuk menjaga aktualitas materi. Proses ekstraksi awal mengunduh 5.000 entri ulasan mentah yang kemudian disaring berdasarkan ketersediaan teks pada atribut content dan nilai peringkat pada atribut score, sehingga diperoleh sebanyak 2.845 entri ulasan mentah yang memiliki struktur atribut lengkap. Struktur atribut hasil pemindaian data tersebut disajikan pada Tabel 1.

Tabel 1. Perangkat Lunak dan Perangkat Keras Pendukung

Nama Atribut	Keterangan
reviewId	ID unik untuk setiap ulasan
userName	Nama pengguna yang memberikan ulasan
userImage	Tautan gambar profil pengguna
content	Isi teks ulasan yang ditulis pengguna

DOI: https://doi.org/10.31004/riggs.v5i2.12268

Nama Atribut	Keterangan
score	Nilai *rating* (1-5) yang diberikan pengguna
thumbsUpCount	Jumlah tanda suku (like) pada ulasan
reviewCreatedVersion	Versi aplikasi saat ulasan dibuat
at	Tanggal dan waktu ulasan dibuat
replyContent	Isi balasan resmi dari pihak developer (jika ada
repliedAt	Tanggal dan waktu balasan developer diberikan
appVersion	Versi aplikasi yang terpasang saat data ditarik

Pelabelan kelas sentimen ditentukan secara otomatis berdasarkan nilai *score* rating bintang pengguna [7]. Skor rating 1 dan 2 dikategorikan sebagai sentimen Negatif, skor 3 sebagai Netral, sedangkan skor 4 dan 5 dikelompokkan sebagai sentimen Positif. Aturan pengelompokan ini menghasilkan distribusi awal dataset yang menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang signifikan sebelum memasuki fase pembersihan teks, sebagaimana dirinci pada Tabel 2.

Tabel 2. Perangkat Lunak dan Perangkat Keras Pendukung

Nama Atribut	Keterangan	Jumlah Data	Persentase (%)
Negatif	1-2	1.693	59,51%
Netral	3	412	14,48%
Positif	4-5	740	26,01%
Total		**2.845**	**100%**

2.2. Tahapan Pra-Pemrosesan Teks

Data tekstual yang telah dikumpulkan diproses melalui enam tahapan berurutan untuk membersihkan data tidak terstruktur agar siap diolah dalam pembuatan model analisis sentimen [11]. Tahap pertama diawali dengan proses cleaning yang bertujuan untuk menyaring teks dari keberadaan elemen noise yang dianggap redundan atau tidak relevan, seperti tautan URL, angka, simbol khusus, tagar, serta spasi berlebih demi menghasilkan korpus yang terstandarisasi dan optimal [12]. Setelah teks dibersihkan, data diarahkan menuju tahap case folding untuk mengonversi seluruh aksara menjadi huruf kecil (*lowercase*) guna menyelaraskan representasi teks sehingga variasi penulisan kapitalisasi tetap diakui sebagai satu entitas kata yang identik oleh sistem [12]. Langkah berikutnya adalah normalisasi kosakata informal yang mencakup konversi istilah tidak baku, singkatan kasual, maupun jargon slang menjadi format leksikal standar dengan merujuk pada kamus mapping khusus berisi 100 entri kata slang berbahasa Indonesia, seperti melakukan translasi kata "gak" menjadi "tidak" atau "apk" menjadi "aplikasi" [13]. Setelah normalisasi, teks memasuki tahap tokenization untuk memecah struktur kalimat menjadi unit-unit kata tunggal atau token individual menggunakan bantuan fungsi word_tokenize dari pustaka NLTK [11]. Proses kemudian diteruskan ke tahap *stopword removal* untuk mengeliminasi kata-kata tugas yang kurang memiliki bobot semantik seperti kata depan atau kata hubung melalui daftar bawaan NLTK yang dikombinasikan dengan kata tugas spesifik penelitian hingga mencapai total 773 entri kata yang dibuang [11]. Sebagai pamungkas, prosedur stemming dieksekusi dengan memanfaatkan pustaka PySastrawi untuk mendekonstruksi kata berimbuhan dan mengembalikannya ke bentuk dasar leksikal bahasa Indonesia secara presisi [14]. Melalui rangkaian pra-pemrosesan ini, ulasan yang menghasilkan teks kosong (seperti ulasan yang hanya berisi emoji atau angka) dihapus dari dataset sehingga menyisakan sebanyak 2.829 ulasan valid untuk fase ekstraksi fitur..

2.3. Ekstraksi Fitur TF-IDF

Teks hasil pra-pemrosesan diubah menjadi representasi numerik menggunakan instrumen pembobotan TF-IDF untuk menetapkan bobot signifikansi sebuah kata di dalam dokumen dengan mempertimbangkan keberadaannya di dalam keseluruhan koleksi data korpus [3]. Pendekatan ini merupakan sintesis dari dua parameter utama, yakni

DOI: https://doi.org/10.31004/riggs.v5i2.12268

Term Frequency (TF) dan *Inverse Document Frequency* (IDF). Formula matematis untuk menentukan nilai TF dinyatakan sebagai berikut:

$$TF(t,d) = \frac{f(t,d)}{\max\{f(t\prime,d):t\prime\in d\}} \quad (1)$$

Di mana $TF(t,d)$ adalah nilai frekuensi kata t pada dokumen d, $f(t,d)$ melambangkan jumlah kemunculan kata t pada dokumen d, dan $\max\{f(t',d): t' \in d\}$ menyatakan total seluruh kata pada dokumen d. Sementara itu, perhitungan untuk aspek *Inverse Document Frequency* (IDF) dirumuskan melalui persamaan berikut:

$$IDF(t,D) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

Di mana $IDF(t,D)$ adalah nilai inverse frekuensi kata t dalam seluruh dokumen, N merepresentasikan jumlah seluruh dokumen dalam korpus, dan $df(t)$ menunjukkan jumlah dokumen yang mengandung kata t. Intergrasi bobot akhir TF-IDF diperoleh dari perkalian nilai Tf dan IDF melalui formula berikut:

$$\mathrm{W(t,d)} = \mathrm{TF(t,d)} \times \mathrm{IDF(t,D)} \quad (3)$$

Di mana $\mathrm{W(t,d)}$ melambangkan bobot akhir kata t pada dokumen d, $TF(t,d)$ adalah nilai frekuensi term kata t. Pada penelitian ini, metode tersebut diimplementasikan menggunakan *TfidVectorizer* dari pustaka *scikit-learn* dengan konfigurasi parameter pembatasan fitur max_features = 5.000, pertimbangan kombinasi kata ngram_range = (1,2) untuk unigram dan bigram, syarat kemunculan minimal min_df = 2, serta penerapan normalisasi logaritmik sublinear_tf = True [3], [5]. Konfigurasi tersebut menghasilkan matriks berdimensi 2.829 × 4.499 yang siap diumpankan ke model klasifikasi.

2.4. Ekstraksi Fitur TF-IDF

Algoritma *Naive Bayes* merupakan model klasifikasi berbasis probabilitas yang bersandar pada *Teorema Bayes* dengan premis dasar bahwa setiap fitur bersifat independen atau saling lepas satu sama lain [15]. Persamaan dasar Teorema Bayes diformulasikan sebagai berikut:

$$\mathrm{P(C|d)} = \frac{\mathrm{P(d|C)P(C)}}{\mathrm{P(d)}} \quad (4)$$

Di mana $P(C|d)$ melambangkan probabilitas posterior kelas C diberikan dokumen d, $P(d|C)$ menyatakan probabilitas kemunculan dokumen d pada kelas C atau *likelihood*, $P(C)$ merupakan probabilitas prior dari kelas C, dan $P(d)$ adalah probabilitas dokumen d yang berfungsi sebagai faktor normalisasi. Varian yang diterapkan dalam kajian teks ini adalah *Multinomial Naive Bayes* (MNB) yang bekerja berdasarkan bobot kemunculan kata dengan formulasi sebagai berikut [7]:

$$P(p|n) \propto P(p) \prod_{k=1}^{n_d} P\,(t_k|p) \quad (5)$$

Di mana $P(p|n)$ adalah probabilitas posterior kelas p terhadap dokumen n, $P(p)$ merepresentasikan prior probabilitas dokumen berada di kelas p, $P(t_k|p)$ menyatakan probabilitas kemunculan kata t_k pada kelas p, dan simbol perkalian $\prod$ melambangkan akumulasi perkalian *likelihood* seluruh kata dalam dokumen n. Guna menanggulangi kendala probabilitas bernilai nol bagi terminologi kata yang tidak terekam dalam fase pelatihan, studi ini mengadopsi prosedur *Laplace Smoothing* dengan menetapkan nilai parameter *smoothing* α sebesar 1,0 [3].

2.5. Referensi

Guna memitigasi problematika ketimpangan distribusi data (*imbalance class*) yang berisiko memicu bias pada model klasifikasi terhadap kategori mayoritas, teknik *Synthetic Minority Oversampling Technique* (SMOTE) diintegrasikan sebagai instrumen penyeimbang data [7]. Berbeda dengan metode konvensional yang mendplikasi data secara acak, SMOTE bekerja dengan cara memproduksi sampel baru sintetis berbasis interpolasi ruang di

DOI: https://doi.org/10.31004/riggs.v5i2.12268

antara pasangan data kelompok minoritas menggunakan pendekatan tetangga terdekat [7]. Rumus matematis untuk membangkitkan data sintetis tersebut dinyatakan sebagai berikut [7]:

$$X_{syn} = X_i + (X_{knn} - X_i) \times \lambda \quad (6)$$

Di mana X_{syn} melambangkan data sintetis baru yang dibentuk, X_i menyatakan sampel dari kelas minoritas yang dipilih, X_{knn} menunjukkan posisi salah satu tetangga terdekat (*k-nearest neighbor*) dari sampel X_i, dan λ merepresentasikan nilai acak acak antara rentang 0 dan 1. Dalam penelitian ini, SMOTE diaplikasikan eksklusif pada porsi data latih menggunakan konfigurasi parameter k_neighbors = 5 dan sampling_strategy = 'auto' untuk menghindari penyerapan data uji (*data leakage*) [7], [13]. Intervensi ini berhasil mendongkrak total sampel data latih dari 2.263 menjadi 4.056 unit dengan komposisi seimbang yakni 1.352 data di setiap kelas sentimen.

2.6. Kerangka Evaluasi Model

Dataset dipartisi secara sistematis menggunakan rasio pembagian data 80:20, di mana 80% dialokasikan sebagai data latih untuk proses pembelajaran model dan 20% difungsikan sebagai data uji untuk pengujian performa. Pembagian data mengeksekusi parameter stratified split guna memastikan proporsi sebaran pada ketiga kelas sentimen di bagian data latih dan data uji tetap ekuivalen dengan karakteristik distribusi data asli. Kinerja dari hasil klasifikasi model divalidasi memanfaatkan instrumen *confusion matrix* multikelas untuk memetakan kombinasi nilai prediksi dan nilai aktual [1]. Struktur pemetaan tersebut direpresentasikan pada Tabel 3.

Tabel 3. Confusion matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

Berdasarkan parameter fundamental dalam *confusion matrix* tersebut, dirumuskan empat metrik evaluasi utama untuk menakar akurasi, presisi, kepekaan, serta titik harmonisasi model sebagai berikut [1], [7]:

$$\text{Akurasi(Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\text{Presisi(Precision)} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

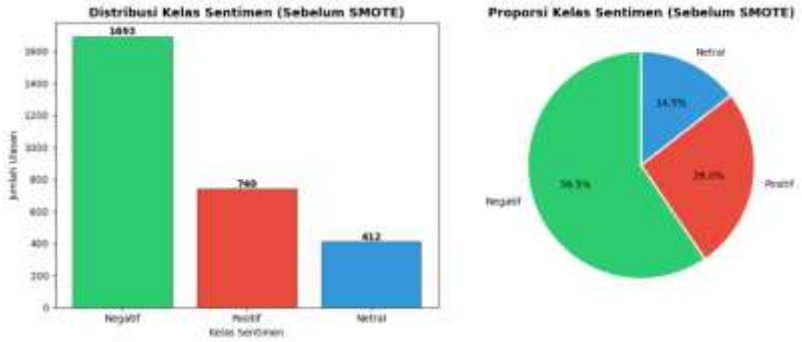
Mengingat penelitian ini menangani klasifikasi multikelas yang memiliki tiga kelas sentimen (Positif, Negatif, dan Netral), kalkulasi metrik presisi, *recall*, dan *F1-score* dihitung menggunakan metode *weighted average*. Pendekatan rata-rata terbobot ini mengintegrasikan proporsi jumlah sampel dari masing-masing kategori data uji secara adil sehingga mampu menyajikan performa evaluasi yang valid dan objektif meskipun terdapat perbedaan kuantitas data antar kelas [1].

DOI: https://doi.org/10.31004/riggs.v5i2.12268

3. Hasil dan Diskusi

3.1. Hasil Klasifikasi Model *Naive Bayes* Tanpa SMOTE

Eksperimen pertama dilakukan dengan menguji performa algoritma Multinomial *Naive Bayes* yang dikombinasikan dengan pembobotan TF-IDF tanpa melibatkan teknik penyeimbangan data. Pengujian ini menggunakan pembagian rasio data latih dan data uji sebesar 80:20 dari total 2.829 ulasan valid. Kondisi awal distribusi data sebelum penyeimbangan dapat dilihat pada Gambar 2.



Gambar 2. Grafik Distribusi Kelas Sentimen Sebelum SMOTE

Setelah model dilatih menggunakan data latih yang timpang, performa model diukur menggunakan data uji. Hasil evaluasi *classification report* untuk skenario tanpa SMOTE ini dirangkum dalam Tabel 4.

Tabel 4. Classification Report Model Naive Bayes Tanpa SMOTE

Kelas	Precision	Recall	F1-score
Negatif	0,66	1,00	0,79
Netral	0,00	0,00	0,00
Positif	0,91	0,34	0,50
Weighted Avg	**0,63**	**0,68**	**0,60**

Berdasarkan paparan hasil data pada Tabel 2, dapat diamati dengan jelas bahwa meskipun nilai akurasi global yang diraih model tanpa intervensi SMOTE tampak cukup tinggi yakni mencapai 68,37%, capaian tersebut sejatinya bersifat bias dan manipulatif. Model menunjukkan kegagalan total dalam mengidentifikasi kategori Netral (dengan nilai presisi, recall, maupun F1-score sebesar 0,00). Sistem hanya sanggup mendeteksi 34% dari total ulasan Positif yang ada pada data pengujian (recall 0,34), meskipun tingkat ketepatan prediksinya tergolong tinggi (precision 0,91). Fenomena ini berakar pada kecenderungan model untuk memusatkan keputusan klasifikasi ke dalam kelas mayoritas (Negatif) yang memegang porsi dominan pada training set. Akibatnya, nilai recall untuk kelas Negatif tercatat sempurna (1.00). namun kondisi tersebut harus dibayar mahal dengan terabaikannya kemamanpuan generalisasi model terhadap kategori minoritas.

3.2. Hasil Klasifikasi Model *Naive Bayes* Dengan SMOTE

Pengujian skenario kedua melibatkan model *Multinomial Naive Bayes* yang dikembangkan setelah porsi data latih diaugmentasi menggunakan teknik SMOTE hingga mencapai total 4.056 sampel dengan komposisi kelas yang seimbang. Melalui pengujian pada perangkat data uji yang sama, model ini menghasilkan nilai akurasi keseluruhan sebesar 58,83%, dengan presisi rata-rata tertimbang sebesar 65,25%, recall sebesar 58,83%, dan tingkat F1-score sebesar 61,22%. Meskipun terjadi penurunan pada nilai akurasi global, model ini menunjukkan ketangguhan dan generalisasi yang jauh lebih baik dalam mengenali pola sentimen secara menyeluruh. Peningkatan performa yang paling signifikan terlihat pada kategori Netral, di mana nilai recall melonjak drastis dari titik nol menjadi 0,43 dengan tingkat presisi sebesar 0,23 dan F1-score mencapai 0,30. Perbaikan performa juga terjadi pada sentimen Positif yang mencatatkan nilai recall sebesar 0,55, presisi sebesar 0,66, dan F1-score yang meningkat menjadi 0,60.

Tabel 5. Classification Report Model Naive Bayes Dengan SMOTE

Kelas	Precision	Recall	F1-score
Negatif	0,75	0,64	0,69

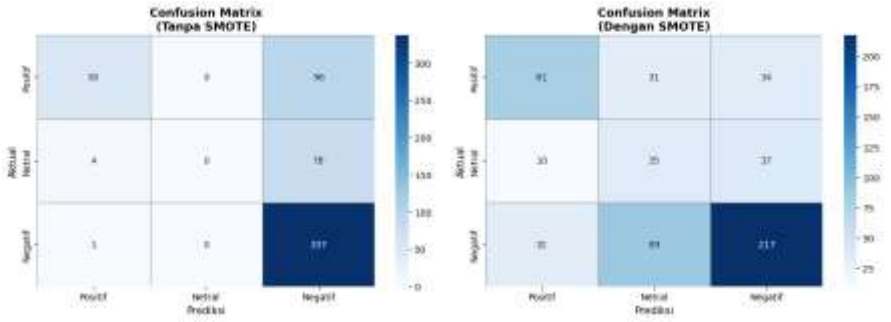
DOI: https://doi.org/10.31004/riggs.v5i2.12268

Netral	0,23	0,43	0,30
Positif	0,66	0,55	0,60
Weighted Avg	**0,65**	**0,59**	**0,61**

Berdasarkan analisis *confusion matrix* pasca-penerapan SMOTE, terlihat bahwa distribusi prediksi model telah bergeser secara proporsional dan tidak lagi memusat pada kelas mayoritas semata. Dari 146 data aktual kelas Positif, jumlah prediksi benar meningkat menjadi 81 data, sementara sisanya terbagi menjadi 31 prediksi Netral dan 34 prediksi Negatif. Pada kelas Netral yang berjumlah 82 data aktual, sistem kini berhasil menebak secara tepat sebanyak 35 data, sedangkan kesalahan prediksi terdistribusi menjadi 10 data ke kelas Positif dan 37 data ke kelas Negatif. Konsekuensi logis dari perubahan orientasi model ini menyebabkan penurunan performa pada kelas Negatif, di mana dari 338 data aktual, prediksi benar mengalami reduksi menjadi 217 data, dengan kesalahan klasifikasi berupa 32 data diprediksi sebagai Positif dan 89 data sebagai Netral. Penurunan recall kelas Negatif menjadi 0,64 ini menjadi bukti nyata bahwa model telah berhasil melepaskan ketergantungan atau bias terhadap kelas mayoritas, sehingga mampu memberikan ruang interpretasi fitur yang lebih adil bagi kategori Netral dan Positif.

3.3. Perbandingan Performa Komprehensif

Evaluasi komparatif antara model *Naive Bayes* tanpa SMOTE dan model dengan SMOTE memperlihatkan dinamika metrik performa yang bervariasi secara komprehensif. Berdasarkan akumulasi total nilai *weighted average*, implementasi teknik SMOTE berdampak pada penurunan nilai akurasi sebesar 9,54%, yakni merosot dari angka 68,37% menjadi 58,83%. Tren penurunan yang identik juga terjadi pada nilai recall keseluruhan yang terkontraksi dari 68,37% menjadi 58,83%. Namun, situasi sebaliknya ditunjukkan oleh parameter presisi total yang mengalami eskalasi sebesar 2,42% dari titik 62,83% menuju level 65,25%. Dinamika positif ini diperkuat oleh nilai F1-score global yang sukses merangkak naik sebesar 0,97%, berpindah dari level 60,24% menjadi 61,22%.



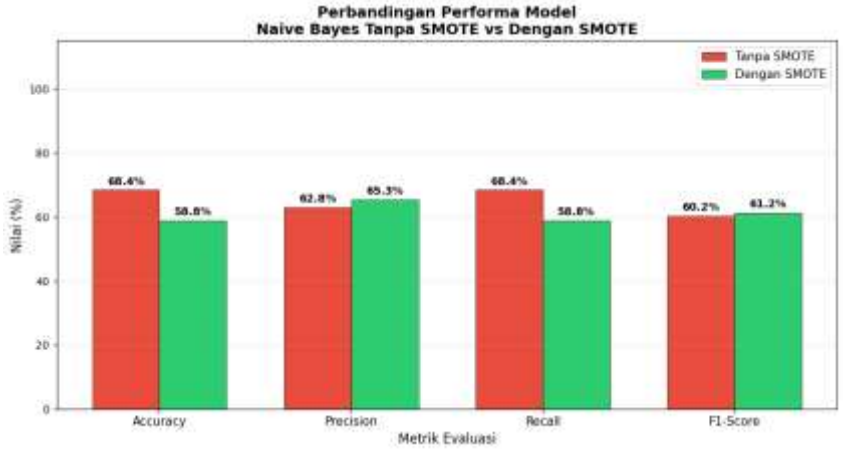
Gambar 3. Confusion Matrix Tanpa dan Dengan SMOTE

Tabel 6. Perbandingan Metrik Evaluasi Model Tanpa SMOTE dan Dengan SMOTE

Metrik	Tanpa SMOTE	Dengan SMOTE	Selisih
Accuracy	68,37%	58,83%	-9,54%
Precision	62,83%	65,25%	+2,42%
Recall	68,37%	58,83%	-9,54%
F1-score	60,24%	61,22%	+0,97%

DOI: https://doi.org/10.31004/riggs.v5i2.12268

Secara akademis, fenomena penurunan akurasi ini tidak dapat diinterpretasikan sebagai kegagalan dari metode penyeimbangan data. Pada dataset yang mengalami ketimpangan kelas yang parah, nilai akurasi dan recall global sering kali bersifat semu dan manipulatif karena nilainya didongkrak secara artifisial oleh performa tinggi pada kelas mayoritas. Saat model tanpa SMOTE menebak hampir seluruh data sebagai Negatif, nilai akurasinya terdengar tinggi karena kelompok Negatif memegang porsi dominan pada data uji, padahal sistem tersebut sejatinya lumpuh total dalam mengidentifikasi kelas Netral. Oleh sebab itu, dalam konteks penanganan *imbalanced class*, metrik F1-score yang mengombinasikan nilai presisi dan recall lewat rata-rata harmonis menjadi tolok ukur yang jauh lebih valid, objektif, dan tepercaya untuk mengukur keberhasilan generalisasi model. Keunggulan model berbasis SMOTE dalam meraih F1-score tertinggi menjadi bukti bahwa interpolasi data sintetis berhasil melatih model untuk mengenali batas-batas keputusan klasifikasi secara lebih proporsional di seluruh spektrum kelas.



Gambar 4. Perbandingan Performa Model Naive Bayes Tanpa dan Dengan SMOTE

3.4. Penguji Uji Prediksi Manual

Guna menguji validitas dan kemampuan generalisasi model final yang telah dioptimalkan dengan SMOTE terhadap data baru, dilakukan eksperimen pengujian mandiri menggunakan lima sampel teks ulasan yang dibuat secara artifisial. Sampel kalimat pertama yang berbunyi, "Aplikasinya bagus banget, ceritanya lengkap dan mudah digunakan!", berhasil diprediksi secara akurat ke dalam kategori Positif dengan tingkat probabilitas sebesar 61,54%. Keberhasilan yang sama juga ditunjukkan pada sampel kalimat kedua, "Sering error dan crash, sangat mengecewakan", di mana model mampu menetapkan label Negatif secara tepat dengan tingkat keyakinan sebesar 51,02%. Kedua hasil ini mengonfirmasi bahwa sistem memiliki ketajaman yang mumpuni dalam mengektraksi dan mengklasifikasikan dokumen teks yang mengandung muatan polaritas sentimen secara lugas, tegas, dan eksplisit.

Kendati demikian, hasil kurang memuaskan muncul saat model dihadapkan pada tiga sampel kalimat yang mengandung struktur bahasa informasi, kosakata ambigu, atau muatan sentimen campuran. Sampel ketiga yang menyatakan, "Lumayan sih aplikasinya, biasa saja", keliru diprediksi sebagai sentimen Positif dengan probabilitas 58,64%, padahal secara semantik bermakna Netral. Selanjutnya, sampel ulasan keempat, "Gabisa login dari tadi, tolong perbaiki segera", secara tidak tepat diklasifikasikan ke dalam label Netral dengan tingkat keyakinan 59,41%, meskipun esensi kalimat tersebut merupakan keluhan yang bernilai Negatif. Kesalahan terakhir tampak pada sampel kelima, "Update terbaru makin keren dan lebih cepat loadingnya", yang salah dipetakan menjadi label Netral dengan probabilitas 57,25% dari yang seharusnya bernilai Positif. Keterbatasan ini berakar pada karakteristik mendasar dari algoritma *Naive Bayes* yang memegang prinsip asumsi independensi fitur secara kaku. Karena algoritma ini menganggap setiap kemunculan kata saling lepas dan berdiri sendiri tanpa memedulikan urutan atau konteks antarkata, sistem menjadi kurang sensitif terhadap pergeseran makna semantik yang kompleks, frasa sarkasme, atau penekanan kalimat keluhan yang sering digunakan dalam gaya penulisan ulasan masyarakat Indonesia.

3.5. Jawaban Masalah & Diskusi Akademis

Melalui seluruh rangkaian eksperimen komputasi yang telah dijalankan, studi ini berhasil merumuskan jawaban ilmiah yang kritis untuk menyelesaikan tiga rumusan masalah yang diajukan pada bab pendahuluan. Terkait rumusan masalah pertama mengenai efektivitas kombinasi Multinomial *Naive Bayes* dan pembobotan TF-IDF, hasil pengujian membuktikan bahwa sinergi kedua metode ini memiliki kapabilitas yang andal dalam memetakan

DOI: https://doi.org/10.31004/riggs.v5i2.12268

dokumen teks ulasan ke dalam tiga spektrum polaritas sentimen. Model ini mampu menembus ambang batas performa dengan mencatatkan nilai akurasi di atas 60% pada skenario data orisinal. Capaian ini sejalan dengan landasan teori dan penelitian terdahulu oleh Nurian dan Sari (2023) serta Wati et al. (2023) yang menyatakan bahwa *Naive Bayes* merupakan metode klasifikasi teks yang efisien, berkecepatan komputasi tinggi, dan sangat kompatibel jika dipadukan dengan pembobotan fitur berbasis nilai frekuensi dokumen TF-IDF.

Untuk menjawab rumusan masalah kedua mengenai kontribusi penerapan SMOTE, hasil analisis memperlihatkan adanya transformasi performa model yang signifikan. Penerapan SMOTE pada sektor data latih terbukti secara empiris berhasil mendiversifikasi data minoritas melalui penciptaan sampel sintetis berbasis interpolasi tetangga terdekat, sehingga mampu mendongkrak nilai presisi total sebesar 2,42% dan menaikkan F1-score global menjadi 61,22%. Kontribusi paling esensial dari intervensi ini adalah pemulihan kemampuan model dalam mengenali kelas minoritas, yang ditunjukkan oleh lonjakan nilai recall kelas Netral dari angka nol menjadi 0,43. Fenomena penurunan akurasi global sebesar 9,54% sebagai kompensasi dari pemerataan kemampuan prediksi ini sesuai dengan teori dari Pulungan et al. (2024) serta Dharmendra et al. (2024) yang menegaskan bahwa penyeimbangan dataset lewat SMOTE memang berpotensi mengoreksi nilai akurasi total demi menghilangkan bias model terhadap kelas mayoritas, sekaligus meningkatkan akurasi pengenalan pada kelas-kelas minoritas yang sebelumnya terabaikan.

Terakhir, untuk menjawab rumusan masalah ketiga mengenai potret distribusi sentimen pengguna, akumulasi data valid menunjukkan secara mutlak bahwa persepsi publik terhadap aplikasi Wattpad di Google Play Store didominasi oleh opini bernuansa Negatif dengan persentase mencapai 59,51%. Angka ini berbanding jauh dengan sebaran opini Positif yang hanya menyerap porsi 26,01% dan sentimen Netral sebesar 14,48%. Temuan ini memberikan kontribusi praktis yang berharga bagi pihak manajemen dan tim pengembang Wattpad Corp karena menyingkapkan fakta bahwa mayoritas pengguna aktif di Indonesia merasa kurang puas terhadap performa sistem. Berdasarkan analisis teks korpus, berhulu tingginya angka sentimen negatif ini dipicu oleh akumulasi keluhan teknis yang berulang, seperti gangguan stabilitas aplikasi yang sering mengalami *error* atau *crash*, hambatan sistem saat memproses autentikasi akun atau *login*, serta lonjakan durasi tayangan iklan yang dirasa sangat mengganggu kenyamanan pengguna dalam membaca. Fakta sebaran ini sekaligus memperbarui studi literatur terdahulu dari Adhan et al. (2024) yang sebelumnya mencatat dominasi kelas positif pada aplikasi Wattpad menggunakan metode Random Forest, sehingga temuan dalam riset ini dapat dijadikan acuan ilmiah terkini yang lebih kontekstual dalam menangani dinamika ulasan digital berbahasa Indonesia.

4. Kesimpulan

Berdasarkan hasil eksperimen yang telah dilaksanakan, dapat disimpulkan bahwa integrasi Multinomial *Naive Bayes* dan pembobotan TF-IDF yang dikombinasikan dengan teknik SMOTE memiliki kontribusi signifikan dalam menyelesaikan kendala ketimpangan distribusi kelas sentimen pada ulasan aplikasi Wattpad di Google Play Store. Penerapan SMOTE secara empiris terbukti mampu memulihkan kelumpuhan model orisinal dalam mendeteksi kelas minoritas Netral, yang ditunjukkan oleh peningkatan recall kelas Netral dari 0,00 menjadi 0,43 serta kenaikan nilai F1-score global rata-rata berbobot dari 60,24% menjadi 61,22%. Penurunan akurasi keseluruhan sebesar 9,54% merupakan kompensasi logis yang valid dari proses redistribusi bobot keputusan klasifikasi agar tidak bias terhadap kelas mayoritas. Dari perspektif analisis data ulasan pengguna, persepsi publik didominasi secara mutlak oleh sentimen Negatif sebesar 59,51%, yang merefleksikan adanya ketidakpuasan mendasar dari para pengguna aktif di Indonesia terkait persoalan teknis kestabilan aplikasi, hambatan akses masuk akun, serta gangguan intensitas penayangan iklan komersial. Implikasi praktis penelitian ini memberikan acuan strategis yang tepercaya bagi pihak pengembang untuk memprioritaskan optimalisasi arsitektur teknis aplikasi demi menaikkan retensi dan kepuasan pengguna di masa mendatang. Untuk penelitian lanjutan, disarankan melakukan pengayaan ekstraksi fitur dengan pendekatan leksikon sentimen formal atau arsitektur *deep learning* kontekstual guna mengatasi kelemahan *Naive Bayes* dalam mengenali diksi campuran dan kosakata *slang* yang berkarakteristik ambigu.

Referensi

[1] S. N. Adhan *et al.*, “Analisis sentimen ulasan aplikasi wattpad di google play store dengan metode random forest,” vol. 2, no. 1, pp. 6–15, 2024.

[2] M. K. Insan, U. Hayati, and O. Nurdiawan, “ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI,” vol. 7, no. 1, pp. 478–483, 2023.

[3] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes Pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH TF-IDF Weighting Using Naïve Bayes on Public Sentiment on The Issue of Rising BIPIH,” vol. 13, no. April, pp. 84–93, 2023.

[4] N. C. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee

DOI: https://doi.org/10.31004/riggs.v5i2.12268

pada Google Play Store," vol. 2, no. April, pp. 47–54, 2022.
[5] S. A. Helmayanti, F. Hamami, and R. Y. Fa'rifah, "Jurnal Indonesia : Manajemen Informatika dan Komunikasi PENERAPAN ALGORITMA TF-IDF DAN NAÏVE BAYES Jurnal Indonesia : Manajemen Informatika dan Komunikasi," vol. 4, no. 3, pp. 1822–1834, 2023.
[6] M. R. Hanafi and R. K. R, "Sentiment Analysis on Sirekap App Reviews on Google Play Using Naive Bayes Algorithm Analisis Sentimen pada Ulasan Aplikasi Sirekap di Google Play Menggunakan Algoritma Naive Bayes," vol. 4, no. October, pp. 1578–1586, 2024.
[7] M. P. Pulungan, A. Purnomo, A. Kurniasih, P. Korespondensi, I. Class, and S. M. O. Technique, "PENERAPAN SMOTE UNTUK MENGATASI IMBALANCE CLASS DALAM KLASIFIKASI KEPRIBADIAN MBTI MENGGUNAKAN NAIVE BAYES APPLICATION OF SMOTE TO OVERCOME CLASS IMBALANCE IN THE MBTI PERSONALITY CLASSIFICATION USING THE NAÏVE BAYES CLASSIFIER," vol. 11, no. 5, pp. 1033–1042, 2024, doi: 10.25126/jtiik.2024117989.
[8] U. of Toronto, "'A Match Made in Heaven': Allen Lau on Naver's US$600-Million Acquisition of Wattpad," University of Toronto News. Accessed: Jul. 04, 2026. [Online]. Available: https://www.utoronto.ca/news/match-made-heaven-allen-lau-naver-s-us600-million-acquisition-wattpad
[9] Google, "Wattpad - Read & Write Stories," Google Play Store. Accessed: Jul. 04, 2026. [Online]. Available: https://play.google.com/store/apps/details?id=wp.wattpad
[10] R. N. Rahman, A. Rahim, and W. J. Pranoto, "Analisis Sentimen Ulasan Game eFootball 2024 Pada Playstore menggunakan Algoritma Naïve Bayes," no. 15, 2025.
[11] D. Rifaldi, A. Fadlil, and Herman, "DECODE : Jurnal Pendidikan Teknologi Informasi," vol. 3, no. 2, pp. 161–171, 2023.
[12] M. A. Java, M. Syafrullah, Windarto, and Painem, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine," vol. 12, pp. 75–80, 2024.
[13] I. K. Dharmendra, I. M. Agus, W. Putra, and Y. P. Atmojo, "Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet," vol. 9, no. 2, pp. 192–193, 2024.
[14] M. U. Albab, Y. K. P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," vol. 20, no. 2, pp. 1–10, 2023.
[15] A. Nurian and B. N. Sari, "Analisis sentimen ulasan pengguna aplikasi google play menggunakan naïve bayes," vol. 11, no. 3, pp. 829–835, 2023.

DOI: https://doi.org/10.31004/riggs.v5i2.12268