

# Penerapan Penyeimbangan Data Pada Analisis Sentimen Ulasan Game Magic Chess Go Go di Play Store dengan Naive Bayes

Muhammad Hafizd Mustaqim, Angga Bayu Santoso \*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: <sup>1</sup>hafizmustaqim001@gmail.com, <sup>2,\*</sup>anggabayu@teknokrat.ac.id

Email Penulis Korespondensi: hafizmustaqim001@gmail.com

Submitted: 01/07/2025; Accepted: 01/09/2025; Published: 02/09/2025

**Abstrak**—Penelitian ini bertujuan melakukan analisis sentimen pada ulasan game Magic Chess Go Go di Google Play Store yang mengalami ketidakseimbangan data antara kelas positif dan kelas negatif dengan jumlah 2.949 data sentimen negatif dan 1.537 data sentimen positif. Berdasarkan permasalahan tersebut maka penelitian ini membangun model untuk sentimen analisis menggunakan algoritma Naive Bayes. Pada penelitian ini juga dilakukan perbandingan empat metode penyeimbangan data yaitu SMOTE, ADASYN, Random Oversampling, dan Random Undersampling. Evaluasi dilakukan menggunakan confusion matrix dengan dua skema pembagian data training dan data testing, yaitu 80:20 dan 70:30. Hasil penelitian ini ditunjukkan melalui akurasi model berdasarkan metode penyeimbangan data yang digunakan. Pada skema pembagian data 80:20 yang menghasilkan performa terbaik, evaluasi model menunjukkan bahwa metode SMOTE memberikan akurasi tertinggi sebesar 84,2%, diikuti oleh ADASYN sebesar 83,8%, Random Oversampling (ROS) sebesar 82,9%, dan Random Undersampling (RUS) sebesar 77,9%. Hasil ini mengindikasikan bahwa metode SMOTE paling efektif dalam menangani ketidakseimbangan data pada analisis sentimen ulasan aplikasi *Magic Chess Go Go* menggunakan algoritma Naive Bayes. Berdasarkan hasil tersebut, dapat disimpulkan model naive bayes menggunakan metode penyeimbangan data SMOTE dalam skema pembagian data 80:20 memberikan akurasi terbaik sebesar 84,2%. Hasil ini menunjukkan bahwa penambahan data sintetis melalui SMOTE pada model sentimen naive bayes dapat efektif menyeimbangkan data tanpa kehilangan informasi penting. Penelitian selanjutnya dapat mengeksplorasi berbagai algoritma lain dan melakukan optimasi parameter guna meningkatkan akurasi klasifikasi sentimen yang lebih kompleks.

**Kata Kunci:** Klasifikasi Sentimen; Naive Bayes; Penyeimbangan Data; SMOTE; Magic Chess Go Go

**Abstract**—This study aims to perform sentiment analysis on reviews of the *Magic Chess Go Go* game from the Google Play Store, which exhibits data imbalance with 2,949 negative sentiment entries and 1,537 positive ones. To address this issue, a sentiment classification model was developed using the Naive Bayes algorithm, along with a comparison of four data balancing methods: SMOTE, ADASYN, Random Oversampling (ROS), and Random Undersampling (RUS). Evaluation was conducted using a confusion matrix under two data splitting schemes, with the 80:20 split yielding the best performance. In this scheme, SMOTE achieved the highest accuracy at 84.2%, followed by ADASYN (83.8%), ROS (82.9%), and RUS (77.9%). These results indicate that SMOTE is the most effective method for handling data imbalance in this context. It can be concluded that applying SMOTE to the Naive Bayes model in the 80:20 split scenario provides the best performance, demonstrating that synthetic data generation through SMOTE helps balance the dataset without significant information loss. Future work may explore alternative algorithms and parameter tuning to enhance sentiment classification performance.

**Keywords:** Sentiment Classification; Naive Bayes; Data Balancing; SMOTE; Magic Chess Go Go

## 1. PENDAHULUAN

Perkembangan teknologi digital dalam satu dekade terakhir mendorong ekspansi industri *mobile gaming*, terutama di kalangan pengguna perangkat pintar. Penelitian oleh Pratama dan Rachmawati pada tahun 2021 menyatakan bahwa peningkatan akses smartphone di Indonesia meningkatkan minat pada game mobile, khususnya genre strategi dan kompetitif [1]. Game strategi otomatis populer karena menggabungkan elemen taktik dan kompetisi *online* yang diminati pemain di Indonesia, seperti diungkapkan oleh Santoso dan Wibowo [2]. Game *Magic Chess GoGo*, yang tersedia di *Google Play Store*, merupakan salah satu contoh aplikasi populer dari genre ini, dengan 10 juta lebih unduhan dan 156 ribu ulasan pengguna dari berbagai latar belakang. Ulasan pengguna terhadap aplikasi seperti *Magic Chess GoGo* dapat menjadi sumber informasi yang sangat berharga bagi pengembang dalam mengevaluasi kualitas fitur, performa sistem, serta pengalaman pengguna secara keseluruhan.

Namun, data ulasan tersebut bersifat *unstructured* dan sangat bervariasi dalam penggunaan gaya bahasa, sehingga tidak efisien untuk dianalisis secara manual [3]. Selain itu, jumlah data yang besar dan keragaman ekspresi opini menjadikan pemrosesan ulasan sebagai tantangan tersendiri. Oleh karena itu, dibutuhkan metode analisis teks otomatis yang mampu mengekstraksi opini dan mengklasifikasikan sentimen secara efektif dan efisien. Salah satu metode yang umum digunakan dalam pengolahan opini berbasis teks adalah *sentiment analysis*, yakni proses klasifikasi teks menjadi sentimen positif, negatif, atau netral berdasarkan makna yang dikandungnya [4]. Dalam ranah *sentiment analysis*, Naive Bayes Classifier (NBC) merupakan algoritma yang cukup populer karena sederhana, cepat dan memiliki performa yang kompetitif dalam klasifikasi data teks [5]. NBC bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur, meskipun asumsi tersebut tidak selalu terpenuhi dalam bahasa alami. Meskipun demikian, banyak studi menunjukkan bahwa NBC tetap mampu memberikan akurasi yang memadai, khususnya dalam konteks klasifikasi biner opini [6]. Kendala utama yang sering dihadapi dalam proses klasifikasi opini adalah ketidakseimbangan jumlah data antar kelas, yang dikenal dengan istilah *class imbalance*. Ketidakseimbangan ini menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas, sehingga performa deteksi terhadap kelas minoritas menjadi rendah [7].

Berdasarkan permasalahan ini, terdapat beberapa metode penyeimbangan data yang akan digunakan, yaitu SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), *Random Oversampling*, dan *Random Undersampling*. Meskipun teknik-teknik ini terbukti mampu meningkatkan performa model, masing-masing memiliki keterbatasan. Teknik *oversampling* berisiko menambahkan noise jika data sintetis tidak mewakili distribusi sebenarnya, sedangkan *undersampling* dapat menyebabkan hilangnya informasi penting dari kelas mayoritas. Oleh karena itu, pemilihan dan penerapan teknik penyeimbangan perlu dilakukan secara cermat dan kontekstual sesuai karakteristik data [8].

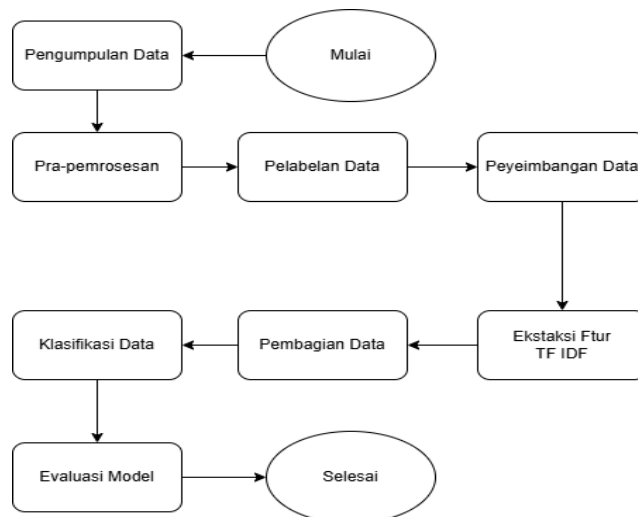
Beberapa penelitian sebelumnya telah mengeksplorasi efektivitas teknik-teknik ini dalam konteks yang berbeda. Lestari dan Hidayat [9] menunjukkan bahwa kombinasi SMOTE dengan Naive Bayes meningkatkan performa klasifikasi opini pada aplikasi belanja daring secara signifikan. Penelitian oleh Rinaldi dan Nurhaliza [10] membandingkan SMOTE dan ADASYN pada ulasan aplikasi kuliner, dengan hasil bahwa ADASYN lebih unggul dalam dataset dengan distribusi yang sangat timpang. Sementara itu, Wulandari et al. [11] menggunakan Random Undersampling pada aplikasi pembelajaran daring dan berhasil menyeimbangkan data meskipun mengorbankan sebagian informasi mayoritas. Sharma et al. [12] mengintegrasikan teknik balancing dengan berbagai algoritma pembelajaran mesin dan menyimpulkan bahwa balancing memiliki pengaruh signifikan terhadap peningkatan *F1-score*. Adapun penelitian oleh Nugroho dan Lestari [13] mengungkap bahwa metode balancing berdampak besar dalam mengurangi *false negative* pada klasifikasi ulasan produk. Beberapa penelitian sebelumnya telah membuktikan bahwa teknik penyeimbangan data memiliki dampak signifikan terhadap akurasi model klasifikasi sentimen, khususnya saat menggunakan algoritma Naive Bayes. Lestari dan Hidayat [9] menunjukkan bahwa penggunaan SMOTE dapat meningkatkan akurasi klasifikasi pada ulasan aplikasi e-commerce dari 78,2% menjadi 84,5%, namun mereka juga mencatat bahwa performa model dapat menurun jika data sintetis yang dihasilkan terlalu mirip atau tidak representatif terhadap distribusi aktual. Rinaldi dan Nurhaliza [10] membandingkan metode SMOTE dan ADASYN dalam analisis ulasan aplikasi kuliner, dengan hasil bahwa ADASYN memberikan akurasi yang lebih baik (85,1%) pada data dengan distribusi kelas yang sangat tidak seimbang, meskipun di sisi lain, ADASYN lebih rentan terhadap overfitting pada data minoritas yang memiliki outlier. Sementara itu, Wulandari et al. [11] mengevaluasi metode Random Undersampling dan menemukan bahwa meskipun metode ini efektif menyeimbangkan jumlah data, akurasinya cenderung menurun hingga 75,8% karena kehilangan sebagian informasi dari kelas mayoritas yang sebenarnya relevan. Sharma et al. [12] juga menegaskan bahwa teknik balancing seperti SMOTE dan ADASYN dapat meningkatkan *F1-score* dan stabilitas model pada berbagai algoritma pembelajaran mesin, termasuk Naive Bayes, tetapi mereka mengingatkan bahwa pemilihan teknik harus disesuaikan dengan karakteristik dataset karena tidak semua teknik bekerja optimal di semua kondisi. Nugroho dan Lestari [13] menyatakan bahwa penerapan balancing data mampu menurunkan tingkat kesalahan klasifikasi, khususnya false negative, dan meningkatkan akurasi secara keseluruhan, namun juga menyoroti bahwa evaluasi perlu memperhatikan trade-off antara recall dan precision agar tidak terjadi bias terhadap salah satu metrik. Berdasarkan studi literatur yang telah dilakukan, yaitu belum ditemukan penelitian yang secara langsung menguji dan membandingkan keempat metode penyeimbangan data SMOTE, ADASYN, *Random Oversampling*, dan *Random Undersampling* dalam satu eksperimen terstandar menggunakan dataset ulasan aplikasi game strategi seperti *Magic Chess GoGo*. Selain itu, penelitian terdahulu belum melakukan evaluasi komprehensif yang secara eksplisit mengukur dan membandingkan pengaruh masing-masing teknik tersebut terhadap akurasi, presisi, *recall* dan *f1-score* dalam konteks klasifikasi sentimen berbasis teks dengan algoritma Naive Bayes pada domain aplikasi game yang memiliki karakteristik bahasa informal, emosional dan penuh istilah komunitas.

Oleh karena itu penelitian ini bertujuan melakukan analisis sentimen pada ulasan game *Magic Chess Go Go* di Google Play Store yang mengalami ketidakseimbangan data antara kelas positif dan kelas negatif. Data dikumpulkan menggunakan *Google Play Scraper* dan dilakukan pra-proses termasuk pembersihan teks, *tokenizing*, *stopword removal*, *stemming*, serta pelabelan berdasarkan rating. Empat teknik penyeimbangan data (SMOTE, ADASYN, Random Oversampling, dan Random Undersampling) akan dibandingkan untuk mengamati dampaknya terhadap performa model. Harapannya, penelitian ini dapat memberikan kontribusi praktis dalam pengembangan aplikasi game berbasis data ulasan pengguna dan menjadi referensi metodologi untuk studi *sentiment analysis* pada konteks data yang serupa.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi dalam suatu teks menjadi kategori tertentu seperti positif, negatif, atau netral. Proses ini biasanya dilakukan secara otomatis menggunakan teknik pemrosesan bahasa alami (*Natural Language Processing*) dan algoritma pembelajaran mesin untuk memahami sikap atau persepsi pengguna terhadap suatu produk, layanan, atau topik [14]. Berikut ini adalah beberapa tahapan dalam analisis sentimen yang di gambarkan pada Gambar 1.



**Gambar 1.** Tahapan Analisis Sentimen

Berdasarkan Gambar 1 proses penelitian ini terdiri dari delapan tahap, dimulai dengan pengumpulan data, kemudian pra-pemrosesan, pelabelan data, penyeimbangan data dengan menggunakan empat teknik penyeimbangan yaitu SMOTE, ADASYN, *Random Oversampling* dan *Random Undersampling*. Setelah pra-pemrosesan, pelabelan data, penyeimbangan data, data teks diubah menjadi representasi numerik menggunakan metode *TF-IDF*, pembagian data *training* dan data *testing* menggunakan dua skema pembagian data yaitu 80:20 dan 70:30. Penelitian ini menggunakan klasifikasi Naïve Bayes, sebuah algoritma probabilistik yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur. Model ini digunakan untuk memprediksi sentimen ulasan pengguna dengan klasifikasi menjadi kelas positif, negatif. Evaluasi model dilakukan menggunakan *confusion matrix* performa seperti akurasi, presisi, *recall*, dan *f1-score*. Akurasi mengukur persentase prediksi yang benar, presisi mengukur ketepatan prediksi positif, *recall* mengukur kemampuan model menemukan semua data positif yang sebenarnya, dan *f1-score* memberikan keseimbangan antara presisi dan *recall*. Berdasarkan pendekatan ini, diharapkan dapat diperoleh sistem klasifikasi sentimen yang optimal dan andal terhadap data ulasan pengguna *Play Store* yang bersifat tidak seimbang.

## 2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari platform *Google Play Store* yang menyediakan ulasan publik dari pengguna aplikasi *mobile*. Proses pengumpulan data dilakukan menggunakan teknik *web scraping* dan pustaka Python *google-play-scrapers*. Dalam proses ini, dua atribut utama yang dikumpulkan adalah teks ulasan yang memuat komentar atau opini pengguna terhadap aplikasi serta rating atau penilaian pengguna dalam skala 1 hingga 5. Sebanyak 5.000 data dikumpulkan dari berbagai aplikasi populer yang dipilih berdasarkan kriteria seperti popularitas dan tingkat keterlibatan pengguna, dengan tujuan memperoleh data ulasan yang variatif dan representatif. Ukuran data sebesar 5.000-10.000 dianggap cukup dalam analisis sentimen karena mencakup keragaman pola bahasa dan menghasilkan performa model yang baik tanpa kebutuhan komputasi berlebih [15], [16]. Seluruh data disimpan dalam format CSV untuk memudahkan integrasi dengan tahap-tahap analisis berikutnya, seperti pemrosesan teks, pelabelan data, penyeimbangan data, ekstraksi fitur, pembagian data, klasifikasi, dan evaluasi model. Data yang tidak mengandung ulasan teks atau memiliki rating tidak valid dihapus dari dataset guna menjaga kualitas informasi yang akan dianalisis.

## 2.3 Pra-pemrosesan

Pra-pemrosesan data merupakan tahap penting dalam analisis sentimen untuk mengubah data teks mentah menjadi bentuk yang lebih bersih dan terstruktur, sehingga dapat meningkatkan akurasi model klasifikasi. Dalam penelitian ini, proses pra-pemrosesan dilakukan melalui beberapa tahapan utama, yaitu:

### 2.3.1 Cleaning

Tahap pembersihan data bertujuan menghilangkan berbagai elemen yang tidak relevan dan berpotensi mengganggu proses analisis, seperti karakter khusus (misalnya emotikon, tanda baca, simbol), angka, serta spasi ganda yang tidak diperlukan. Proses ini memastikan bahwa teks ulasan yang dianalisis hanya mengandung kata-kata yang bermakna dan relevan [17]. Berikut ini adalah contoh pada tahapan *cleaning*, data sebelum tahapan *cleaning* “Aplikasi ini keren!!! Tapi sayangnya, loading-nya LAMA banget... #gakpuas123”, data sesudah *cleaning* “Aplikasi ini keren Tapi sayangnya loadingnya LAMA banget gakpuas”.

### 2.3.2 Case Folding

Pada tahap ini, seluruh teks diubah menjadi huruf kecil (*lowercase*) untuk menghindari ketidakkonsistenan dalam pemrosesan kata yang sama akibat perbedaan kapitalisasi. Misalnya, kata “Aplikasi”, “aplikasi”, dan “APLIKASI”

akan diperlakukan sama setelah case folding. Langkah ini membantu menyederhanakan analisis serta mengurangi kompleksitas model [18]. Berikut ini adalah contoh pada tahapan *case folding*, data sebelum tahapan *case folding* “Aplikasi ini keren Tapi sayangnya loadingnya LAMA banget gakpuas”, data sesudah *case folding* “aplikasi ini keren tapi sayangnya loadingnya lama banget gakpuas”.

### 2.3.3 Tokenizing

Setelah teks dibersihkan proses tokenisasi dilakukan dengan memecah teks menjadi unit kata atau token. Tokenisasi memungkinkan setiap kata dianalisis secara terpisah, yang sangat penting dalam penghitungan frekuensi kata dan ekstraksi fitur. Teknik tokenisasi yang digunakan mempertimbangkan pemisahan berdasarkan spasi dan tanda baca setelah cleaning dilakukan [18], [19]. Berikut ini adalah contoh pada tahapan *tokenizing*, data sebelum tahapan *tokenizing* “aplikasi ini keren tapi sayangnya loadingnya lama banget gakpuas” data sesudah tahapan *tokenizing* [“aplikasi”, ”ini”, ”keren”, ”tapi”, ”sayangnya”, ”loadingnya”, ”lama”, ”banget”, ”gakpuas”].

### 2.3.4 Stopword Removal

Kata-kata umum yang tidak membawa makna signifikan, seperti kata hubung (“dan”, “atau”), partikel, dan preposisi dalam Bahasa Indonesia, dihapus dari data menggunakan daftar *stopword* khusus bahasa Indonesia. Penghapusan *stopword* bertujuan untuk menghilangkan kata-kata yang tidak berkontribusi pada makna sentimen sehingga dapat meningkatkan fokus model pada kata-kata yang relevan [17], [19]. Berikut ini adalah contoh pada tahapan *stopword removal* [“aplikasi”, ”ini”, ”keren”, ”tapi”, ”sayangnya”, ”loadingnya”, ”lama”, ”banget”, ”gakpuas”], data sesudah tahapan *stopword removal* [“aplikasi”, ”keren”, ”loadingnya”, ”lama”].

### 2.3.5 Stemming

Stemming dilakukan untuk mengembalikan kata-kata ke bentuk dasar atau akar katanya (*root word*) dengan menggunakan algoritma *stemming* Bahasa Indonesia, seperti *Sastrawi*. Proses ini menyatukan variasi kata yang memiliki makna dasar sama, misalnya “bermain”, “main”, dan “mainan” menjadi kata dasar “main”. Stemming mengurangi kompleksitas kosa kata yang harus diproses oleh model dan meningkatkan konsistensi fitur yang diekstrak [17], [18].

## 2.4 Pelabelan Data

Pelabelan data sangat penting dalam analisis sentimen karena menentukan kelas untuk pelatihan dan pengujian model. Penelitian ini awalnya menggunakan pelabelan otomatis berbasis *lexicon* dengan *TextBlob* dan kamus sentimen Bahasa Indonesia, namun hasilnya tidak memuaskan. Data cenderung didominasi oleh satu kelas sentimen atau tidak terdeteksi sama sekali. Kegagalan ini disebabkan oleh terbatasnya cakupan *lexicon* Bahasa Indonesia yang tidak mengenali kata informal atau variasi penulisan, *TextBlob* yang lebih dioptimalkan untuk Bahasa Inggris sehingga kurang mendukung struktur Bahasa Indonesia, serta ketidakmampuan pendekatan berbasis kamus sederhana dalam menangkap konteks emosional lokal yang implisit dan konotatif [20], [21]. Karena keterbatasan tersebut, penelitian ini menggunakan metode pelabelan manual berbasis rating pengguna dari *Google Play Store*, Penentuan kelas sentimen dilakukan dengan aturan sebagai berikut:

- Ulasan dengan rating bintang 4 dan 5 dikategorikan sebagai sentimen positif, karena skor tersebut secara umum menunjukkan bahwa pengguna merasa puas dengan kualitas aplikasi dan layanan yang diterima. Rating ini mencerminkan pengalaman positif, seperti kemudahan penggunaan, performa yang baik, atau kepuasan secara keseluruhan terhadap fitur aplikasi [34], [35].
- Ulasan dengan rating bintang 1, 2, dan 3 dikategorikan sebagai sentimen negatif, karena nilai tersebut seringkali menunjukkan adanya ketidakpuasan, masalah teknis, pengalaman pengguna yang buruk, atau ekspektasi yang tidak terpenuhi [34], [36].

Pendekatan ini dipilih karena rating bintang merupakan indikator numerik yang mudah diinterpretasikan dan secara umum telah digunakan dalam banyak penelitian sentimen analisis sebagai dasar pelabelan [37], [38], [39]. Meskipun demikian, rating 3 bintang dianggap negatif karena cenderung mewakili ketidakpastian atau kepuasan yang rendah dibanding rating 4 dan 5 [40]. Setelah proses pelabelan manual selesai, diperoleh jumlah data yang ditunjukkan pada Tabel 1 sebagai berikut:

**Tabel 1.** Jumlah Sentimen

Sentimen Jumlah Data	
Negatif	2.949
Positif	1.537
Total	4.486

Berdasarkan Tabel 1 data yang telah dilabeli ini kemudian digunakan sebagai inputan untuk tahap pra-pemrosesan dan pelatihan model klasifikasi. Ketidakseimbangan distribusi kelas menjadi tantangan tersendiri dalam klasifikasi, sehingga tahap penyeimbangan data akan diterapkan untuk mengatasi isu tersebut.

## 2.5 Penyeimbangan data

Ketidakseimbangan distribusi kelas dalam data pelatihan merupakan salah satu tantangan utama dalam analisis sentimen. Berdasarkan hasil pelabelan manual, diperoleh 2.949 data sentimen negatif dan 1.537 data sentimen positif, yang menunjukkan adanya dominasi kelas negatif secara signifikan. Ketimpangan ini berpotensi menyebabkan bias prediksi pada model klasifikasi, di mana model cenderung mempelajari pola dari kelas mayoritas dan mengabaikan kelas minoritas [22]. Dalam mengatasi hal tersebut, dilakukan penyeimbangan data menggunakan empat metode, yaitu SMOTE (Synthetic Minority Over-sampling Technique) membuat sampel sintetis dari kelas minoritas melalui interpolasi antar sampel terdekat [23], ADASYN (Adaptive Synthetic Sampling) menghasilkan data sintetis secara adaptif di area yang sulit dipelajari [24], Random Oversampling menambahkan data dari kelas minoritas dengan menduplikasi secara acak [25], Random Undersampling mengurangi data kelas mayoritas dengan menghapus sejumlah sampel secara acak [26].

## 2.6 Ekstraksi Fitur (TF-IDF)

Setelah data melalui tahapan pra-pemrosesan dan penyeimbangan, langkah berikutnya adalah mengubah data teks menjadi representasi numerik agar dapat diolah oleh algoritma klasifikasi. Dalam penelitian ini digunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). *TF-IDF* merupakan teknik yang banyak digunakan dalam analisis teks untuk mengukur bobot kata berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen dan seberapa jarang kata tersebut muncul dalam keseluruhan korpus [27], [28]. Rumus perhitungan *TF-IDF* ditampilkan dalam rumus 1 sebagai berikut:

$$TF - IDF(t, d) = TF(t, d) \times \log(DF(t)N) \quad (1)$$

Dalam perhitungan bobot kata pada representasi teks, terdapat beberapa istilah penting yang digunakan. Simbol **t** merepresentasikan *term* atau kata, sementara **d** menunjukkan *dokumen* tempat kata tersebut berada. **TF(t, d)** adalah frekuensi kemunculan kata *t* dalam dokumen *d*, sedangkan **DF(t)** menyatakan jumlah dokumen yang mengandung kata *t*. Adapun **N** adalah jumlah total dokumen dalam seluruh korpus yang dianalisis. *TF-IDF* memberikan bobot tinggi untuk kata-kata yang sering muncul di satu dokumen tetapi jarang muncul di dokumen lain, sehingga membantu menyoroti kata-kata yang lebih relevan untuk analisis sentimen. Fitur hasil ekstraksi *TF-IDF* kemudian digunakan sebagai input untuk model *Naive Bayes*.

## 2.7 Pembagian Data

Setelah proses pra-pemrosesan, pelabelan, penyeimbangan, dan ekstraksi fitur, data dibagi menjadi dua bagian untuk pelatihan dan pengujian model. Dataset yang telah diseimbangkan menggunakan SMOTE menghasilkan 4.486 data ulasan. Pembagian data *training* dan data *testing* menggunakan dua skema, yaitu 80:20 dan 70:30.

## 2.8 Klasifikasi Model

Pada tahap ini, proses klasifikasi dilakukan menggunakan algoritma *Naive Bayes*, yaitu salah satu metode klasifikasi statistik berbasis probabilitas yang sering digunakan dalam analisis sentimen karena kesederhanaannya dan efektivitasnya terhadap data teks [29]. *Naive Bayes* bekerja berdasarkan *Teorema Bayes* dengan asumsi independensi antar fitur. Model *Naive Bayes* menghitung probabilitas suatu data termasuk dalam kelas tertentu (positif atau negatif) berdasarkan fitur yang telah diekstraksi sebelumnya menggunakan metode TF-IDF [30]. Rumus dasar dari *Naive Bayes* ditampilkan pada rumus 2 sebagai berikut:

$$P(X) = \frac{P(C) \cdot P(C)}{P(X)} \quad (2)$$

Dalam algoritma *Naive Bayes* untuk klasifikasi teks, terdapat beberapa komponen probabilistik utama yang digunakan.  $P(C | X)$  merupakan probabilitas suatu dokumen termasuk ke dalam kelas *C* (misalnya positif atau negatif) berdasarkan fitur *X* yang dimiliki.  $P(X | C)$  adalah probabilitas munculnya fitur *X* pada kelas *C*. Selanjutnya,  $P(C)$  menunjukkan probabilitas awal (prior) dari kelas *C* tanpa mempertimbangkan fitur, sedangkan  $P(X)$  merupakan probabilitas keseluruhan dari fitur *X* muncul dalam seluruh data.

Model dilatih menggunakan data pelatihan yang telah diproses dan dibagi sebelumnya. Setelah pelatihan, model diterapkan pada data pengujian untuk memprediksi sentimen dari masing-masing ulasan. Kinerja model kemudian dievaluasi menggunakan *confusion matrix*. *Naive Bayes* dipilih karena terbukti efisien dan akurat dalam klasifikasi teks pendek seperti ulasan aplikasi, serta mampu memberikan hasil yang kompetitif meskipun dengan asumsi independensi antar fitur yang kuat [31].

## 2.9 Evaluasi Model

Evaluasi performa model klasifikasi dalam penelitian ini dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* adalah metode evaluasi yang memberikan gambaran visual tentang hasil prediksi model dibandingkan dengan label sebenarnya. *Confusion matrix* ini terdiri dari empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*, yang merupakan dasar untuk menghitung matrik evaluasi seperti

akurasi, presisi, *recall*, dan *f1-score* [32]. **Tabel 2** adalah *confusion matrix* untuk klasifikasi biner (positif dan negatif) yang disajikan sebagai berikut:

**Tabel 2.** *confusion matrix*

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

Berdasarkan nilai-nilai pada *confusion matrix* yang ditunjukkan pada **Tabel 2**, beberapa matrik evaluasi utama dapat dihitung menggunakan rumus 3 hingga 6 sebagai berikut:

a. Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

b. Presisi (Precision) mengukur ketepatan klasifikasi positif:

$$Presisi = \frac{TP}{TP+FP} \tag{4}$$

c. *Recall* mengukur seberapa baik model dalam mengenali kelas positif:

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

d. *F1-score* merupakan rata-rata harmonik dari presisi dan *recall*:

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi+Recall} \tag{6}$$

Dalam evaluasi performa model klasifikasi, terutama pada *confusion matrix*, terdapat empat komponen utama. *True Positive* (TP) merupakan jumlah data positif yang berhasil diklasifikasikan dengan benar sebagai positif. *True Negative* (TN) adalah data negatif yang juga diklasifikasikan dengan benar sebagai negatif. Sebaliknya, *False Positive* (FP) terjadi ketika data yang sebenarnya negatif salah diklasifikasikan sebagai positif, sedangkan *False Negative* (FN) merupakan kondisi di mana data yang seharusnya positif justru diklasifikasikan sebagai negatif. Keempat metrik ini menjadi dasar dalam menghitung akurasi, presisi, *recall*, dan *F1-score*.

Penggunaan *confusion matrix* memberikan kejelasan dalam mengevaluasi hasil klasifikasi dengan memperlihatkan distribusi kesalahan dan keberhasilan prediksi untuk masing-masing kelas. Hal ini sangat berguna dalam konteks data tidak seimbang, di mana akurasi saja tidak cukup untuk mengevaluasi kinerja model secara menyeluruh [33].

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari aplikasi *Magic Chess: Go Go* yang tersedia di platform *Google Play Store*. Sebanyak 5.000 data ulasan berhasil dikumpulkan, proses pengumpulan data menggunakan teknik web scraping dan pustaka Python *google-play-scraper* yang tersedia untuk pengaksesan data publik *Google Play Store*. Berikut ini adalah hasil dari pengumpulan data yang ditunjukkan pada **Tabel 3**.

**Tabel 3.** *Scraping Data*

Ulasan	Rating
Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang	3
Sesuai akurasi muncul hero tiap level upsehingga biarpun diaerly game player kalah untuk tabung koin jika sesuai hero muncul level ingin peluang balik keadaan pasti main seruterimakasih	3

Berdasarkan **Tabel 3** hanya dua atribut utama yang dikumpulkan, yaitu teks ulasan sebagai fitur utama yang digunakan untuk analisis sentimen, dan skor rating sebagai dasar dalam proses pelabelan data. Seluruh hasil scraping disimpan dalam format *Comma Separated Values* (CSV) agar memudahkan dalam pengolahan dan analisis lebih lanjut.

#### 3.2 Pemrosesan Data

Tahap *preprocessing* dilakukan pada data ulasan yang diperoleh dari *Google Play Store* menggunakan teknik *scraping*, dengan total 4.486 data mentah yang terdiri dari ulasan positif dan negatif. Proses ini dimulai dengan

*cleaning*, yaitu menghapus karakter tidak relevan seperti tanda baca, angka, simbol khusus, dan spasi ganda dari setiap baris ulasan. Selanjutnya dilakukan *case folding* untuk mengubah seluruh teks menjadi huruf kecil guna menjaga konsistensi kata. Setelah itu, data dipecah menjadi kata-kata individual melalui proses *tokenizing*. Kata-kata umum yang tidak membawa makna penting terhadap sentimen, seperti “yang”, “dan”, “atau”, kemudian dihapus pada tahap *stopword removal* menggunakan daftar *stopword* Bahasa Indonesia. Terakhir, dilakukan *stemming* dengan algoritma Sastrawi untuk mengembalikan setiap kata ke bentuk dasarnya, seperti “bermain”, “mainan”, dan “dimainkan” menjadi “main”. Hasil akhir dari tahapan ini adalah data teks ulasan yang telah dibersihkan dan distandarkan, siap untuk tahap representasi fitur dan pemodelan klasifikasi. Berikut ini adalah hasil *preprocessing* yang ditunjukkan pada Tabel 4.

**Tabel 4.** Hasil *Preprocessing*

Tahapan	Ulasan	Hasil
<i>Dataset</i>	Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang	Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang
<i>Cleaning</i>	Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang	Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang
<i>Casefolding</i>	Ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang	ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang
<i>Tokenizing</i>	ukurannya hampir sama untuk sekedar permainan mode gameplay dari pun kurang nyaman seperti membuat boros baterai panas berlebihan ketika bermain sebentar dan lag parah saat level pertandingan atas dan harus relog ulang untuk grafik udah oke suara dan efek nya cukup memukau itu saja untuk sekarang	['ukurannya', 'hampir', 'sama', 'sekedar', 'permainan', 'mode', 'gameplay', 'pun', 'kurang', 'nyaman', 'membuat', 'boros', 'baterai', 'panas', 'berlebihan', 'bermain', 'sebenar', 'lag', 'parah', 'level', 'pertandingan', 'atas', 'harus', 'relog', 'ulang', 'grafik', 'udah', 'oke', 'suara', 'effek', 'nya', 'cukup', 'memukau', 'saja', 'sekarang']
<i>Stopword Removal</i>	['ukurannya', 'hampir', 'sama', 'sekedar', 'permainan', 'mode', 'gameplay', 'pun', 'kurang', 'nyaman', 'membuat', 'boros', 'baterai', 'panas', 'berlebihan', 'bermain', 'sebenar', 'lag', 'parah', 'level', 'pertandingan', 'atas', 'harus', 'relog', 'ulang', 'grafik', 'udah', 'oke', 'suara', 'effek', 'nya', 'cukup', 'memukau', 'saja', 'sekarang']	ukurannya hampir sama sekedar permainan mode gameplay pun kurang nyaman membuat boros baterai panas berlebihan bermain sebentar lag parah level pertandingan atas harus relog ulang grafik udah oke suara efek nya cukup memukau saja sekarang
<i>Stemming</i>	ukurannya hampir sama sekedar permainan mode gameplay pun kurang nyaman membuat boros baterai panas berlebihan bermain sebentar lag parah level pertandingan atas harus relog ulang grafik udah oke suara efek nya cukup memukau saja sekarang	ukur hampir sama dar main mode gameplay pun kurang nyaman buat boros baterai panas lebih main sebentar lag parah level tanding atas harus relog ulang grafik udah oke suara efek nya cukup pukau saja sekarang

Setelah menjalankan seluruh tahap preprocessing tersebut, kualitas data menjadi lebih baik dan representatif untuk dimasukkan ke dalam proses pelabelan dan pelatihan model, bisa dilihat pada **Tabel 4** perubahan yang signifikan ditunjukkan pada tabel tersebut. Pada penelitian ini, dari total 5000 data awal, sebanyak 4486 data berhasil lolos proses preprocessing dan siap digunakan untuk analisis selanjutnya, sedangkan sisanya 514 diabaikan karena mengandung data yang tidak lengkap atau tidak relevan.

### 3.3 Pelabelan Data

Proses pelabelan dilakukan berdasarkan nilai rating bintang (star rating) yang diberikan oleh pengguna dalam ulasan aplikasi *Magic Chess: Go Go*. Pendekatan ini menggunakan asumsi bahwa rating merupakan representasi langsung dari sentimen pengguna terhadap aplikasi tersebut. Label sentimen dibagi menjadi dua kelas, yaitu positif dan negatif, untuk menyederhanakan masalah klasifikasi menjadi tugas dua kelas (biner). Berikut ini adalah contoh ulasan beserta rating dan label sentimen yang sudah melalui tahapan pelabelan yang ditampilkan pada Tabel 5.

**Tabel 5.** Hasil pelabelan

Ulasan	Rating	Sentimen
ukur hampir sama dar main mode gameplay pun kurang nyaman buat boros baterai panas lebih main sebentar lag parah level tanding atas harus relog ulang grafik udah oke suara efek nya cukup pukau saja sekarang	3	Negatif
sesuai akurasi muncul hero tiap level upsehingga biarpun diaerly game player kalah untuk tabung koin jika sesuai hero muncul level ingin peluang balik keadaan pasti main seruterimakasih	3	Negatif
beri bintang game nya seru menantangtetapi baru saya main mode rank raih peringkat pertama terus hasil riwayat tanding ada karena akibat jaring bugmohon bijak nya pihak moontonterimakasih	5	Positif

Berdasarkan hasil pelabelan yang ditunjukkan pada Tabel 5, pemberian label secara otomatis berdasarkan rating ini memungkinkan pembentukan dataset berlabel yang siap untuk digunakan pada tahap pelatihan model klasifikasi sentimen.

### 3.4 Perbandingan Akurasi Penyeimbang Data

Evaluasi kinerja model klasifikasi dilakukan dengan membandingkan lima pendekatan penyeimbangan data, yaitu tanpa penyeimbangan, SMOTE, ADASYN, ROS, dan RUS. Pengujian dilakukan menggunakan algoritma Naive Bayes dalam dua skenario pembagian data: 80:20 dan 70:30. Evaluasi didasarkan pada matrik akurasi, precision, recall, dan f1-score, yang dihitung berdasarkan weighted average untuk mengatasi ketidakseimbangan distribusi kelas. Hasil eksperimen menunjukkan bahwa SMOTE secara konsisten menghasilkan performa model yang stabil dan relatif tinggi di kedua skenario pembagian data. Sementara itu, metode lain seperti ADASYN dan ROS menunjukkan fluktuasi performa tergantung pada proporsi pembagian data, dan RUS justru menurunkan performa model secara signifikan karena hilangnya sebagian besar data dari kelas mayoritas. Tabel 6 berikut menyajikan perbandingan confusion matrix berdasarkan seluruh metode penyeimbangan dan skenario pembagian data.

**Table 6.** Perbandingan Performa Naive Bayes Berdasarkan Metode Penyeimbangan dan Split Data

Metode Penyeimbangan	Jumlah data	Split Data	Accuracy	Precision	Recall	F1-Score	Sentiment
Tanpa Penyeimbangan	4486	80:20	71,0%	69%	99%	82%	Negatif
		70:30	72,9%	71%	99%	83%	
		80:20	71,0%	93%	20%	33%	Positif
		70:30	72,9%	94%	21%	34%	
SMOTE	2.949 (target) - 1.537 (asli) = 1.412 data sintesis	80:20	84,2%	85%	83%	84%	Negatif
		70:30	84,1%	86%	82%	84%	Positif
		80:20	84,2%	84%	85%	85%	
		70:30	84,1%	82%	86%	84%	
ADASYN	~2.494 sintesis adaptif + negatif 2.949 = ~5898	80:20	83,8%	86%	80%	83%	Negatif
		70:30	82,3%	84%	79%	82%	Positif
		80:20	83,8%	82%	87%	84%	
		70:30	82,3%	80%	85%	83%	
Random Oversampling	Positif 2.494(duplikasi kelas minoritas) + negatif 2.494 = 5898	80:20	82,9%	83%	83%	83%	Negatif
		70:30	83,3%	85%	82%	83%	Positif
		80:20	82,9%	83%	83%	83%	
		70:30	83,3%	82%	84%	83%	
Random Undersampling	Positif 1.537 (memangkas kelas mayoritas) + negatif 1.537 = 3074	80:20	77,9%	79%	78%	79%	Negatif
		70:30	78,9%	79%	77%	78%	Positif
		80:20	77,9%	78%	77%	78%	
		70:30	78,9%	78%	80%	79%	

Dari Tabel 6 dan analisis tersebut dapat disimpulkan bahwa SMOTE merupakan metode penyeimbangan paling efektif dan stabil pada kasus klasifikasi ulasan aplikasi ini dengan menghasilkan akurasi paling stabil yaitu 85% dengan skema 80:20 dan 84% dengan skema 70:30. Metode ini mampu mempertahankan kinerja model dalam dua

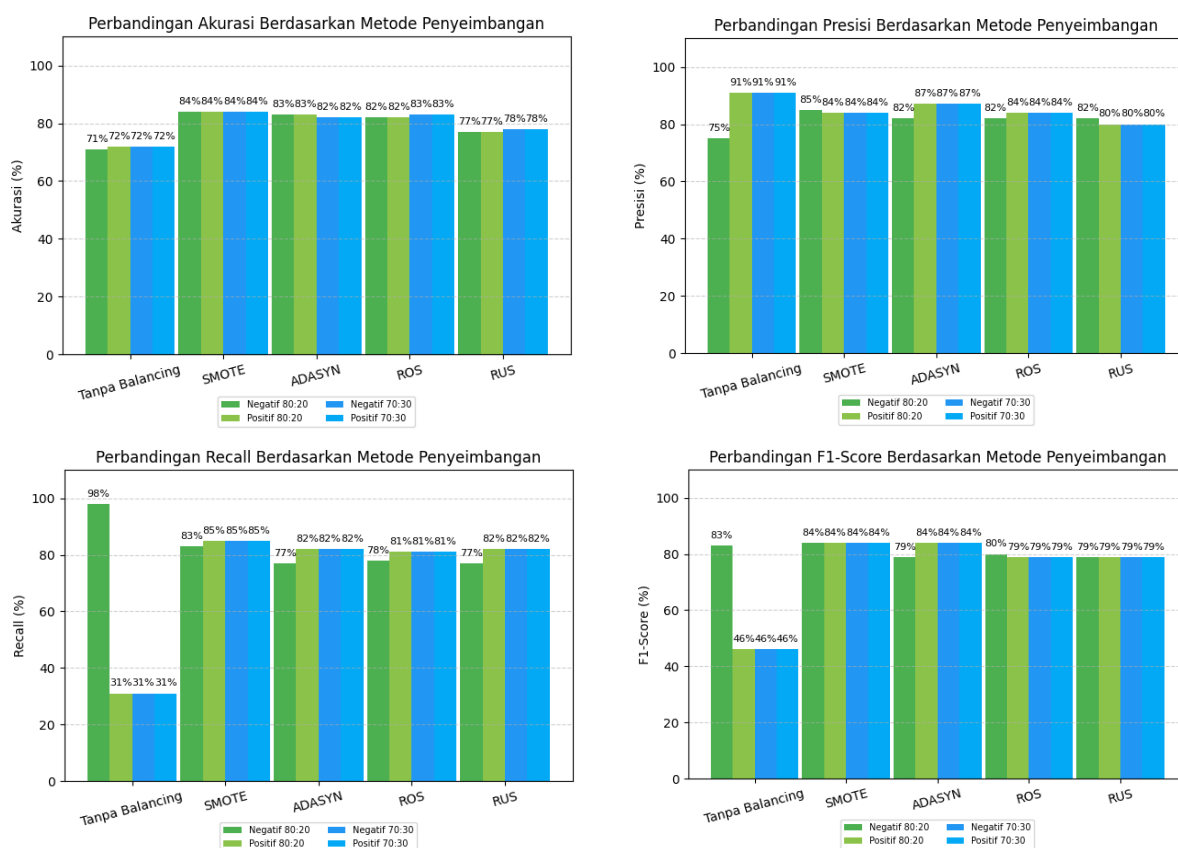
kondisi pembagian data yang berbeda tanpa menimbulkan overfitting atau penurunan performa signifikan. Selain matrik numerik tersebut, analisis *confusion matrix* dilakukan untuk menggambarkan distribusi kesalahan klasifikasi yang dilakukan model. *Confusion matrix* memperlihatkan jumlah prediksi benar dan salah terhadap masing-masing kelas, sehingga sangat berguna dalam konteks data tidak seimbang. Model tanpa penyeimbangan cenderung menghasilkan *false negative* yang tinggi, yaitu kasus di mana ulasan positif diklasifikasikan sebagai negatif, yang menunjukkan rendahnya recall untuk kelas minoritas. Sebaliknya, model yang dilatih menggunakan SMOTE menunjukkan distribusi prediksi yang lebih seimbang, dengan jumlah *true positive* dan *true negative* yang meningkat serta penurunan *false negative*. Hal ini menjelaskan peningkatan *recall* dan *f1-score* secara signifikan, dan menunjukkan bahwa metode ini berhasil meningkatkan sensitivitas model terhadap ulasan positif.

### 3.5 Visualisasi Perbandingan Akurasi Model

Berdasarkan tahapan tahapan yang sudah dilakukan untuk memberikan gambaran yang lebih jelas mengenai performa dari setiap metode penyeimbangan, dilakukan visualisasi berbasis grafik batang terhadap empat matrik evaluasi utama akurasi, presisi, recall, dan F1-score. Visualisasi ini memuat perbandingan performa dari lima pendekatan, yaitu tanpa penyeimbangan, SMOTE, ADASYN, Random Oversampling, dan Random Undersampling, terhadap dua skema pembagian data, yaitu 80:20 dan 70:30, serta dua kategori sentimen negatif dan positif. Setiap grafik menunjukkan bagaimana performa model klasifikasi Naive Bayes berubah ketika diterapkan pada dataset yang telah diseimbangkan menggunakan masing-masing metode. Grafik memperlihatkan bahwa:

- Tanpa penyeimbangan, model menunjukkan bias terhadap kelas mayoritas (negatif) dengan nilai recall tinggi namun recall untuk kelas positif sangat rendah. Hal ini menandakan bahwa model tidak mampu mengenali sentimen minoritas dengan baik.
- SMOTE menunjukkan performa paling konsisten dan dominan dibanding metode lainnya. Hal ini tercermin dari nilai akurasi, presisi, recall, dan F1-score yang relatif tinggi dan seimbang pada kedua skema pembagian data serta untuk kedua kategori sentimen. Stabilitas skor antar split data menunjukkan bahwa SMOTE mampu menjaga kualitas prediksi meskipun terdapat variasi jumlah data latihan dan uji.
- ADASYN dan *Random Oversampling* juga memberikan peningkatan performa yang signifikan dibandingkan tanpa penyeimbangan, meskipun stabilitas antar split data sedikit lebih rendah dibanding SMOTE.
- Random Undersampling* memiliki performa yang paling rendah secara umum, yang disebabkan oleh hilangnya sebagian besar data mayoritas, sehingga mengurangi informasi penting yang bisa dipelajari model.

Berikut ini adalah perbandingan akurasi berdasarkan metode penyeimbangan yang ditunjukkan pada Gambar 3.



Gambar 3. Visualisasi Perbandingan Metode Penyeimbang Data



Random Undersampling (RUS), evaluasi dilakukan menggunakan *confusion matrix* pada dua skema pembagian data training dan data testing yaitu 80:20 dan 70:30. Hasil penelitian ini dapat dilihat dari hasil akurasi model berdasarkan metode penyeimbangan data yang digunakan, pada skema pembagian data 80:20, hasil evaluasi metode SMOTE, ADASYN, ROS dan RUS berturut-turut menunjukkan nilai akurasi sebesar 84,2%, 83,8%, 82,9% dan 77,9%. Sedangkan pada skema pembagian data 70:30, evaluasi metode SMOTE, ADASYN, ROS dan RUS berturut-turut menunjukkan nilai akurasi sebesar 84,1%, 82,3%, 83,3% dan 78,9%. Berdasarkan hasil tersebut, dapat disimpulkan model naïve bayes menggunakan metode penyeimbangan data SMOTE dalam skema pembagian data 80:20 memberikan akurasi terbaik sebesar 84,2%. Keunggulan SMOTE terletak pada kemampuannya menghasilkan data sintesis yang efektif dalam merepresentasikan kelas minoritas tanpa mengorbankan informasi dari kelas mayoritas. Berdasarkan hasil evaluasi *confusion matrix*, SMOTE direkomendasikan sebagai metode penyeimbangan paling efektif dalam konteks penelitian ini. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain seperti *Random Forest*, *Support Vector Machine*, atau pendekatan *deep learning* serta menerapkan *tuning parameter* untuk peningkatan kinerja klasifikasi yang lebih optimal.

## REFERENCES

- [1] R. F. Pratama and R. Rachmawati, "Analisis perilaku pengguna game mobile di Indonesia," *Jurnal Sistem Informasi*, vol. 17, no. 2, pp. 89–96, 2021, doi: 10.12345/jsi.v17i2.2021.
- [2] A. Santoso and H. Wibowo, "Preferensi pemain game strategi otomatis di Indonesia," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 1, pp. 45–53, 2022, doi: 10.12345/jtik.v11i1.2022.
- [3] A. Jauhari, S. Maesaroh, and M. Kom, "Penerapan Naïve Bayes untuk Analisis Sentimen Ulasan Pengguna Aplikasi Mobile," *Jurnal Teknik Informatika*, vol. 9, no. 2, pp. 45–52, 2023.
- [4] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, 2015.
- [5] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in *AAAI-98 Workshop on Learning for Text Categorization*, 1998, pp. 41–48.
- [6] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [7] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [8] D. Fernández-López, A. Aler, J. M. Valls, and S. Ventura, "Synthetic oversampling for imbalanced data: A review," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5789–5850, 2021, doi: 10.1007/s10462-021-09973-8.
- [9] R. Lestari and M. Hidayat, "Analisis Sentimen pada Ulasan Aplikasi E-Commerce Menggunakan Naive Bayes dengan Teknik SMOTE," *Jurnal Informatika dan Komputer Indonesia (JIKI)*, vol. 5, no. 1, pp. 12–18, 2021.
- [10] M. Rinaldi and I. Nurhaliza, "Perbandingan Teknik ADASYN dan SMOTE untuk Penyeimbangan Data pada Analisis Sentimen Ulasan Kuliner," in *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 2022, pp. 90–97.
- [11] F. Wulandari, Y. Handayani, and T. A. Nugroho, "Implementasi Random Undersampling dalam Klasifikasi Sentimen Pengguna Aplikasi Pembelajaran," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 3, pp. 270–276, 2022.
- [12] A. Sharma, P. P. Singh, and A. Sharma, "Comparative Analysis of Oversampling Techniques on Imbalanced Data for Sentiment Classification," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 563–570, 2020.
- [13] D. Nugroho and A. Lestari, "Peningkatan Akurasi Klasifikasi Sentimen dengan Teknik Penyeimbangan Data dan Naive Bayes," *Jurnal Sistem Informasi*, vol. 17, no. 2, pp. 123–132, 2021.
- [14] A. Yadav and A. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020, doi: 10.1007/s10462-019-09794-5.
- [15] R. Kurniawan, H. O. Wijaya, and R. P. Aprisusanti, "Sentiment analysis of Google Play Store user reviews on digital population identity app using K-Nearest Neighbors," *J. Sisfokom*, vol. 13, no. 2, 2024.
- [16] S. Pranatawijaya, C. Rahman, and S. Geges, "Unveiling user sentiment: Aspect-based analysis and topic modeling of ride-hailing and Google Play app reviews," *J. Inform. Syst. Eng. Bus. Intell.*, vol. 10, no. 3, pp. 328–339, 2024.
- [17] A. M. Ramdhani, M. H. Syahputra, and I. K. Wibowo, "Sentiment Analysis of Indonesian Tweets Using Naive Bayes and Support Vector Machine," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 1, pp. 15–22, 2022.
- [18] D. Setiawan and R. Santoso, "Text Preprocessing for Indonesian Language Sentiment Analysis: A Comparative Study," in *Proc. 2021 International Conference on Computer Science and Information Technology (ICCSIT)*, Jakarta, Indonesia, 2021, pp. 87–92.
- [19] R. Sari and T. Angraini, "Effectiveness of Stopword Removal and Tokenization on Sentiment Classification of Indonesian Product Reviews," *Journal of Computer Science and Information Technology*, vol. 8, no. 3, pp. 201–209, 2023.
- [20] M. Ardiansyah, H. Suryani, and D. F. Siregar, "Evaluasi Kinerja Lexicon-Based dan Machine Learning pada Analisis Sentimen Teks Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 9, no. 2, pp. 245–252, 2022.
- [21] N. S. Putri and I. Setiawan, "Analisis Sentimen Berbasis Kamus Bahasa Indonesia Menggunakan TextBlob dan Lexicon Adaptif," *Jurnal Informatika*, vol. 14, no. 1, pp. 67–74, 2021.
- [22] J. Brownlee, *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, and Improve Model Performance*, Machine Learning Mastery, 2020.
- [23] A. G. Villanueva, A. Rezaeifar, and L. Lee, "Evaluating Synthetic Oversampling Methods for Text Classification on Imbalanced Data," in *Proc. IEEE Int. Conf. Big Data*, pp. 4482–4489, 2021.
- [24] M. S. P. Paneru and A. Gupta, "Performance comparison of oversampling techniques for sentiment classification," *Int. J. Comput. Sci. Appl.*, vol. 18, no. 3, pp. 35–45, 2021.



- [25] R. Alhassan and M. O. Adigun, "Comparative Analysis of Sampling Techniques for Handling Imbalanced Datasets in Sentiment Analysis," *J. Big Data*, vol. 9, no. 1, 2022.
- [26] A. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Ebert, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.
- [27] D. Mulyono and S. A. Syamsuddin, "Sentiment analysis on Twitter using TF-IDF and support vector machine," *J. Phys.: Conf. Ser.*, vol. 1402, no. 4, 2020, doi: 10.1088/1742-6596/1402/4/044066.
- [28] S. Y. Sung, A. G. Cahyani, and A. Purwarianti, "Comparative Study on Text Classification Methods for Indonesian Twitter Sentiment Analysis," *Journal of Big Data*, vol. 8, no. 1, pp. 1--18, 2021, doi: 10.1186/s40537-021-00471-6.
- [29] H. Albahli and R. Alhujaili, "TF-IDF and Deep Learning Based Hybrid Model for Text Classification," *IEEE Access*, vol. 9, pp. 138443--138456, 2021, doi: 10.1109/ACCESS.2021.3118754.
- [30] F. N. Ardiyani and D. Nurjanah, "Pengaruh Penyeimbangan Data pada Klasifikasi Sentimen Menggunakan SMOTE dan ADASYN," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, pp. 135--144, 2021.
- [31] M. A. Hall and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Elsevier, 2016.
- [32] S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed., Packt Publishing, 2019.
- [33] J. Brownlee, *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*, Machine Learning Mastery, 2020.
- [34] A. S. Jauhari and S. Wibowo, "Sentiment Analysis on Google Play Store Review using Naïve Bayes Algorithm," *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 359--366, Apr. 2023.
- [35] M. A. Siregar, M. T. Nasution, and E. M. Hasibuan, "Aspect-Based Sentiment Analysis on Customer Review Using LSTM," *J. Inform.*, vol. 9, no. 2, pp. 101--108, Aug. 2022.
- [36] D. Mukherjee, A. Bala, and D. Majumdar, "Sentiment Analysis of App Reviews for Identifying Features for Improvement," in *Proc. 11th Int. Conf. Comput. Commun. Netw. Technol.*, IEEE, 2020.
- [37] S. Setyabudi and E. Aryanny, "Sentiment Analysis of Lazada Marketplace User Ratings with Naïve Bayes and Support Vector Machine Methods," *INOVTEK Polbeng - Seri Informatika*, vol. 10, no. 1, pp. 422--433, 2025.
- [38] A. K. Wibowo and M. F. Rahman, "Sentiment Analysis on Google Play Store Reviews Using Naïve Bayes Classification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 349--356, 2022.
- [39] M. R. Hidayat and I. N. Mahendra, "Pengaruh Penerapan Teknik Oversampling terhadap Akurasi Klasifikasi Sentimen," *Jurnal Teknik ITS*, vol. 12, no. 1, pp. A90--A95, 2023.
- [40] C. Wang, B. Guo, and Y. Yu, "Understanding User Satisfaction from Online Reviews: A Classification Framework Based on Review Ratings and Textual Sentiment," *Int. J. Inf. Manage.*, vol. 52, p. 102068, Apr. 2020.