

# Perbandingan Algoritma Naïve Bayes, Random Forest, dan SVM Untuk Analisis Sentiment Aplikasi PLN Mobile Pada Google Play Store

Cici Nurita Kumala Sari<sup>1</sup>, Ryan Randy Suryono<sup>2,\*</sup>

<sup>1</sup>Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

<sup>2</sup>Fakultas Teknik dan Ilmu Komputer, Program Studi Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: <sup>1</sup>cici\_nurita@teknokrat.ac.id, <sup>2,\*</sup>ryan@teknokrat.ac.id

Email Penulis Korespondensi: ryan@teknokrat.ac.id

Submitted: 29/03/2025; Accepted: 03/05/2025; Published: 01/06/2025

**Abstrak**—PLN Mobile merupakan inovasi digital yang dikembangkan oleh PT PLN (Persero) untuk memberikan layanan ketenagalistrikan melalui perangkat seluler. Banyak pengguna menyampaikan keluhan dan ulasan mereka di Google Play Store. Studi ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi PLN Mobile dengan menggunakan tiga metode klasifikasi: Naïve Bayes, Random Forest, dan Support Vector Machine (SVM). Sebanyak 19.870 ulasan yang telah melalui tahap preprocessing dianalisis dalam penelitian ini. Performa model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Teknik SMOTE diterapkan untuk mengatasi ketidakseimbangan kelas sentimen. Hasil penelitian menunjukkan bahwa sebelum penerapan SMOTE, SVM memiliki performa terbaik dengan akurasi 92%, diikuti oleh Random Forest 79%, dan Naïve Bayes 62%. Setelah penerapan SMOTE, performa SVM meningkat menjadi 95%, Random Forest menjadi 85%, sedangkan Naïve Bayes tetap di angka 62%. Metrik evaluasi lainnya seperti recall dan F1-score juga menunjukkan peningkatan signifikan setelah penerapan SMOTE, terutama pada sentimen negatif dan netral. Hasil ini menunjukkan bahwa SMOTE mampu meningkatkan akurasi dan keseimbangan performa model, serta memberikan wawasan penting terhadap persepsi publik terhadap aplikasi PLN Mobile.

**Kata Kunci:** Analisis Sentimen; PLN Mobile; Playstore; SMOTE; Naive Bayes; Random Forest; SVM.

**Abstract**—PLN Mobile is a digital innovation developed by PT PLN (Persero) to provide electricity services through mobile devices. Many users submit their complaints and reviews on the Google Play Store. This study aims to analyze user sentiment towards the PLN Mobile application using three classification methods: Naïve Bayes, Random Forest, and Support Vector Machine (SVM). A total of 19,870 reviews that have gone through the preprocessing stage were analyzed in this study. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The SMOTE technique was applied to address the imbalance of sentiment classes. The results showed that before the implementation of SMOTE, SVM had the best performance with an accuracy of 92%, followed by Random Forest 79%, and Naïve Bayes 62%. After the implementation of SMOTE, SVM performance increased to 95%, Random Forest to 85%, while Naïve Bayes remained at 62%. Other evaluation metrics such as recall and F1-score also showed significant improvements after the implementation of SMOTE, especially for negative and neutral sentiments. These results show that SMOTE is able to improve the accuracy and balance of model performance, as well as provide important insights into public perception of the PLN Mobile application.

**Keywords:** Sentiment Analysis; PLN Mobile; Playstore; SMOTE; Naive Bayes; Random Forest; SVM.

## 1. PENDAHULUAN

PT PLN (Persero), yang biasa dikenal sebagai Perusahaan Listrik Negara, adalah BUMN di Indonesia yang bergerak di sektor kelistrikan. Sebagai penyedia listrik utama di negara ini, PLN berkomitmen untuk terus meningkatkan standar layanannya. Komitmen ini didorong oleh kebutuhan untuk memahami harapan pelanggan, meningkatkan kepuasan, dan mengembangkan strategi yang selaras dengan perilaku pelanggan yang terus berkembang dan kemajuan internal di dalam PLN. Sejalan dengan tujuan ini, PLN memperkenalkan aplikasi PLN Mobile sebagai inovasi digital yang dirancang untuk menawarkan kemudahan dan kenyamanan yang lebih besar bagi pelanggan dalam mengakses berbagai layanan terkait kelistrikan[1].

PLN Mobile merupakan platform digital yang dikembangkan oleh PT PLN (Persero) untuk menyediakan layanan listrik melewati perangkat seluler. Aplikasi ini menawarkan berbagai layanan daring, termasuk pembayaran tagihan listrik, pembelian token, pengajuan pembacaan meter, peningkatan daya, pelaporan pengaduan, penanganan gangguan, pemantauan penggunaan listrik pascabayar, pemberitahuan tagihan dan pemadaman listrik, serta pembaruan tentang pemeliharaan jaringan dan penyelesaian masalah[2].

Seiring waktu, jumlah pelanggan melaporkan keluhan dan gangguan layanan melalui aplikasi PLN Mobile terus meningkat. Namun, beberapa pengguna tetap merasa tidak puas atas layanan yang diberikan, sehingga mereka pun menyampaikan keluhannya di ulasan Google Play Store[3]. Penilaian di platform tersebut tidak setiap saat mencerminkan kualitas aplikasi secara akurat, karena terkadang pengguna memberikan nilai yang tidak sesuai dengan umpan balik tertulis. Selain itu, meninjau semua komentar pelanggan secara manual memakan waktu dan membutuhkan upaya yang signifikan. Untuk mengatasi hal ini, sistem klasifikasi diterapkan untuk menganalisis dan menilai sentimen publik terhadap aplikasi PLN Mobile secara lebih efisien[4].

NBC adalah algoritma klasifikasi langsung berdasarkan Teorema Bayes, yang mengasumsikan independensi antar variabel. Algoritma ini banyak digunakan di berbagai bidang, seperti analisis sentimen, klasifikasi teks, dan deteksi spam[5]. Sementara itu, algoritma SVM mengidentifikasi hyperplane optimal untuk memisahkan kelas-kelas berbeda dalam satu set data[6]. Algoritma ini sangat efektif untuk data berdimensi tinggi dan sering diterapkan dalam

tugas klasifikasi teks serta identifikasi pola. Sebaliknya, Random Forest yaitu algoritma pembelajaran ensemble yang meningkatkan akurasi klasifikasi dengan mengombinasikan beberapa pohon keputusan, membuatnya efektif dalam menangani kumpulan data dengan banyak fitur[7]. Ketiga algoritma ini telah diterapkan secara luas dan memiliki kelebihan serta keterbatasannya sendiri. Oleh karena itu, pengujian pada set data tertentu sangat penting untuk menentukan metode yang relevan untuk analisis.

Dalam beberapa tahun terakhir, analisis sentimen berbasis machine learning telah menjadi pendekatan yang semakin populer dalam mengevaluasi persepsi publik terhadap layanan digital, termasuk aplikasi seluler. Penelitian terdahulu telah membuktikan efektivitas berbagai algoritma klasifikasi dalam menganalisis opini pengguna, seperti Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Masing-masing metode memiliki keunggulan tersendiri dalam menangani jenis dan volume data yang berbeda. Naïve Bayes dikenal dengan efisiensinya dalam mengolah data teks dalam skala besar, sementara SVM menunjukkan performa tinggi dalam ruang fitur berdimensi besar. Random Forest sendiri unggul dalam kestabilan dan akurasi prediksi melalui pendekatan ensemble learning. Selain itu, penggunaan teknik Synthetic Minority Over-sampling Technique (SMOTE) juga mulai diterapkan secara luas dalam berbagai studi untuk mengatasi ketidakseimbangan kelas dalam data, sehingga dapat meningkatkan akurasi model secara keseluruhan. Integrasi teknik SMOTE bersama algoritma-algoritma tersebut telah menunjukkan potensi besar dalam meningkatkan performa analisis sentimen di berbagai domain, termasuk aplikasi layanan publik.

Dalam Studi sebelumnya telah menganalisis sentimen komentar aplikasi PLN Mobile dengan metode NBC dan KNN melibatkan 3000 data uji, Hasilnya, NBC menunjukkan performa lebih efisien dibandingkan KNN, dengan akurasi 77%, precision 59%, recall 53%, serta F1-Score 54%[8]. Sebuah studi lain yang membandingkan performa Naïve Bayes, Random Forest, serta SVM dalam memprediksi tidak hadir di tempat kerja menyatakan bahwa Random Forest memiliki kinerja unggul, mencapai akurasi 99%, recall 99%, precision 99% dan [9].

Dalam sebuah penelitian sebelumnya, teknik text mining diterapkan dalam analisis sentimen pengguna aplikasi PLN Mobile dengan metode NBC dan pembobotan TF-IDF. Data diperoleh dari Google Play Store, dan hasil studi menunjukkan bahwa model yang dikembangkan mencapai recall 100%, precision 76%, akurasi 76%, dan F1-Score 86%[10]. Studi lain yang membandingkan performa algoritma Naïve Bayes, SVM, dan Random Forest untuk prediksi penyakit ginjal kronis, dilakukan evaluasi terhadap ketiga algoritma tersebut. Hasilnya, Naïve Bayes mencapai akurasi 97,14%, SVM mencapai 92,50%, dan algoritma Random Forest memperoleh akurasi tertinggi, yaitu 99,64%[11]. mengindikasikan keunggulan algoritma Random Forest dalam memprediksi penyakit ginjal kronis.

Sebuah penelitian sebelumnya menganalisis sentimen terkait Metaverse di Twitter menggunakan model algoritma SVM dan Random Forest. Penelitian ini melibatkan 6348 tweet sebagai data. Hasilnya dengan penerapan metode SMOTE, Random Forest menunjukkan kinerja yang unggul, mencapai akurasi 91%, sementara SVM mencapai akurasi 90%.[12].

Meskipun beberapa penelitian telah menganalisis sentimen aplikasi PLN Mobile dan membandingkan algoritma klasifikasi, belum ada studi yang secara komprehensif membandingkan performa Naïve Bayes, Random Forest, dan SVM dengan mempertimbangkan teknik SMOTE dalam mengatasi ketidakseimbangan kelas pada data sentimen. Kesenjangan inilah yang menjadi fokus utama dalam penelitian ini.

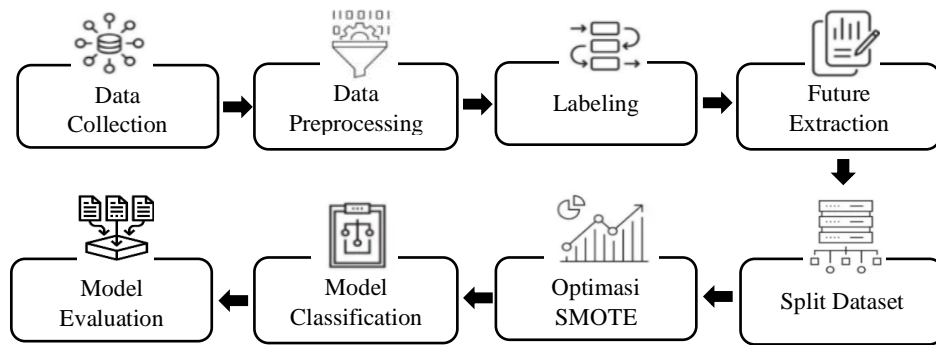
Studi ini bertujuan untuk menganalisis sentimen dalam komentar mengenai aplikasi PLN Mobile dengan metode Naïve Bayes, Random Forest, dan SVM. Dengan melakukan studi ini, PT PLN (Persero) dapat memperoleh wawasan yang lebih baik tentang kepuasan pengguna terhadap aplikasi PLN Mobile. Selain itu, studi ini menggabungkan metode optimasi SMOTE untuk meningkatkan kinerja ketiga algoritma ini, memastikan hasil yang lebih optimal. Tujuan akhirnya adalah mengidentifikasi algoritma klasifikasi yang paling efisien terhadap analisis sentimen. Temuan ini diharapkan dapat memberikan wawasan lebih dalam terhadap umpan balik pengguna aplikasi PLN Mobile.

## 2. METODOLOGI PENELITIAN

Tujuan dari studi ini untuk menganalisis sentimen pada komentar Play Store terhadap aplikasi PLN Mobile serta membandingkan performa tiga metode klasifikasi: Naïve Bayes, Random Forest, dan SVM. Proses studi mencakup beberapa tahap, meliputi pengumpulan data, *preprocessing*, *labelling*, penerapan teknik SMOTE, serta evaluasi model. Masing-masing tahap ini dijelaskan secara rinci di bawah ini.

### 2.1 Tahapan Penelitian

Alur penelitian ini terdiri dari delapan tahapan utama. Tahap pertama adalah pengumpulan data (*Data Collection*), kemudian *preprocessing* untuk membersihkan dan mempersiapkan data. Selanjutnya, dilakukan pelabelan data (*Labeling*) agar setiap sampel memiliki kategori yang sesuai. Selanjutnya, dilakukan ekstraksi fitur dengan metode TF-IDF[13]. Dalam mengatasi data tidak seimbang (*Handle Imbalance*), diterapkan teknik SMOTE sebelum data dibagi menjadi data latih dan data uji. Kemudian, bagian klasifikasi model dengan tiga metode, yaitu *Naïve Bayes*, Random Forest, dan SVM. Tahap terakhir, model dipertimbangkan dengan metrik akurasi, precision, recall, dan F1-score untuk membandingkan kinerja ketiga algoritma dalam analisis sentimen pada komentar aplikasi PLN Mobile. Gambar 1 merupakan ilustrasi dari alur penelitian ini.

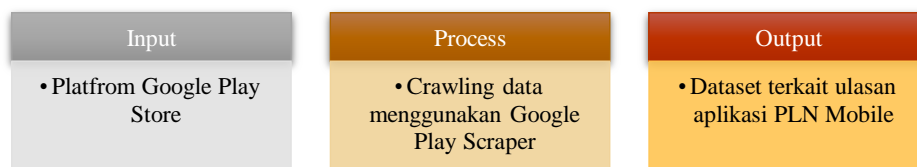


**Gambar 1.** Tahap Penelitian

Berikut adalah penjelasan dari Gambar 1 tentang alur tahapan penelitian yang dilakukan dalam proses analisis sentimen terhadap komentar pengguna aplikasi PLN Mobile.

**2.2.1 Data Collection (Pengumpulan Data)**

Dalam studi ini, data dikumpulkan dengan teknik web scraping. Dataset berisi komentar pengguna terhadap aplikasi PLN Mobile yang diperoleh dari Play Store. Total ulasan yang dikumpulkan adalah 20000 ulasan dalam kurun waktu 18 September 2020 hingga 16 Februari 2025. Data yang dikumpulkan berbahasa Indonesia dan disimpan dalam bentuk CSV untuk pemrosesan lebih lanjut. Gambar 2 merupakan ilustrasi dari proses pengumpulan data.



**Gambar 2.** Proses Pengumpulan Data

Gambar 2 menjelaskan proses pengumpulan data untuk analisis sentimen terhadap ulasan pengguna aplikasi PLN Mobile, yang dilakukan melalui tiga tahapan utama: Input, Process, dan Output.

a. Input

Sumber data berasal dari platform Google Play Store. Di sinilah pengguna memberikan ulasan atau komentar terkait pengalaman mereka menggunakan aplikasi PLN Mobile. Data ini mencakup teks komentar, tanggal ulasan, rating bintang, dan informasi terkait lainnya.

b. Process

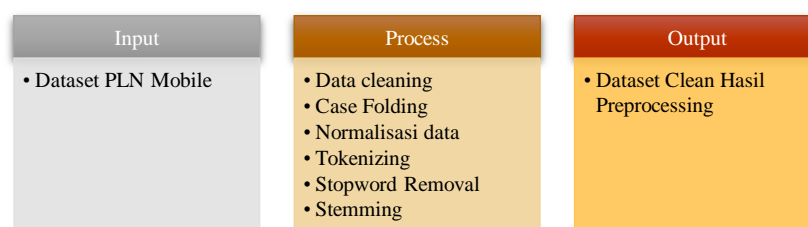
1. Tahap ini melibatkan proses web scraping, yaitu pengambilan data secara otomatis dari halaman aplikasi PLN Mobile di Google Play Store.
2. Alat yang digunakan adalah Google Play Scraper, yang memungkinkan peneliti untuk mengumpulkan ribuan data ulasan secara efisien dan terstruktur.
3. Dalam penelitian ini, data yang berhasil dikumpulkan berjumlah 20.000 ulasan, dikumpulkan selama periode 18 September 2020 hingga 16 Februari 2025.

c. Output

1. Hasil dari proses scraping adalah sebuah dataset yang berisi ulasan-ulasan pengguna aplikasi PLN Mobile.
2. Dataset ini nantinya akan digunakan untuk tahap preprocessing dan analisis sentimen lebih lanjut, seperti pelabelan, pembersihan data, dan penerapan algoritma klasifikasi.

**2.2.2 Preprocessing Data**

*Preprocessing* merupakan tahap mengonversi data tidak terstruktur ke dalam format terstruktur agar memastikan data mudah dibaca dan sesuai untuk analisis. Proses ini bertujuan untuk meningkatkan kualitas serta konsistensi data[14]. Gambar 3 mengilustrasikan alur kerja *Preprocessing*.



**Gambar 3.** Proses *Preprocessing*

Berikut adalah tahapan *preprocessing* data yang digunakan dalam penelitian ini untuk mempersiapkan data teks sebelum dilakukan analisis lebih lanjut:

a. *Data Cleaning*

Pembersihan data yaitu proses menghilangkan tanda baca seperti tanda seru, koma, titik, tanda kurung, dan tanda tanya, serta menghilangkan entri yang sama dan atribut seperti nama pengguna dan karakter khusus[15]. Setelah menerapkan proses pembersihan pada kumpulan data berisi 20000 ulasan, jumlah total ulasan berkurang menjadi 19870.

b. *Case Folding*

Case folding adalah proses mengubah semua huruf dalam dokumen teks menjadi huruf kecil untuk memastikan keseragaman data. Selain itu, karakter apa pun selain huruf dan angka dihilangkan[16].

c. *Normalisasi Data*

Normalisasi adalah proses mengubah kata-kata tidak baku atau slang menjadi bentuk baku yang sesuai dengan kaidah bahasa Indonesia. Dalam penelitian ini, proses normalisasi dilakukan dengan mengacu pada Kamus Besar Bahasa Indonesia (KBBI) untuk memastikan keseragaman dan keakuratan data teks sebelum dianalisis lebih lanjut[17].

d. *Tokenizing*

*Tokenizing* merupakan proses segmentasi dan pemisahan kalimat menjadi kata atau frasa individual menggunakan pemisah seperti koma (,), titik (.), dan pemisah lainnya. Metode *tokenizing* membagi teks menjadi bagian-bagian lebih kecil[18].

e. *Stopword Removal*

Penghapusan *stopword* melibatkan penghapusan kata tidak relevan, seperti "dan," "atau," dan "yang," karena kata-kata tersebut tidak memberikan makna yang signifikan terhadap analisis[19].

f. *Stemming*

*Stemming* adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya atau akar katanya. Tujuannya adalah untuk menyederhanakan variasi kata yang memiliki makna serupa, sehingga analisis teks menjadi lebih efisien dan konsisten[19].

### 2.2.3 Pelabelan Data

Setelah tahap *preprocessing* selesai, dataset berkurang menjadi 19870 ulasan. Kemudian dilakukan pelabelan sentimen untuk mengkategorikan setiap ulasan yang telah diproses. Proses pelabelan ini dilakukan dengan menggunakan pendekatan berbasis kamus, yaitu mengklasifikasikan ulasan di bagian sentimen positif, negatif, atau netral. Kamus sentimen yang dipakai studi ini merupakan kamus berbasis *lexicon based* yang banyak diaplikasikan dalam analisis sentimen di Indonesia.

Pendekatan berbasis *lexicon based* mengutamakan frasa, ekspresi, atau konten teks yang umum ditemukan dalam obrolan, dialog, kiriman, dan ulasan. Metode ini memanfaatkan kamus sentimen yang berisi kata-kata positif dan negatif, yang kemudian dicocokkan dengan kata-kata dalam ulasan untuk menentukan polaritas sentimennya. Klasifikasi sentimen menggunakan pendekatan berbasis *lexicon* dilakukan dengan menganalisis keberadaan kata positif dan negatif dalam komentar yang telah diproses. Jika jumlah kata positif lebih banyak dibandingkan kata negatif, maka ulasan tersebut dikategorikan sebagai ulasan positif. Sebaliknya, jika kata-kata negatif mendominasi, maka ulasan tersebut tergolong negatif. Jika kata-kata positif dan negatif muncul dalam jumlah yang sama, ulasan dikategorikan netral[20].

### 2.2.4 Feature Extraction

Ekstraksi fitur dalam penelitian ini dilakukan menggunakan TF-IDF. TF-IDF diterapkan untuk memproses data pelatihan dengan mengidentifikasi istilah yang sering muncul dan menetapkan nilai berdasarkan frekuensinya dalam dokumen dan kepentingannya relatif terhadap keseluruhan kumpulan data. Teknik TF-IDF dipakai untuk menilai signifikansi kata untuk analisis sentimen. Bobot suatu kata meningkat jika kata tersebut muncul lebih sering dalam dokumen sambil mempertimbangkan kepentingannya di seluruh dataset[14].

### 2.2.5 Split Dataset

Pada bagian split dataset, mempartisi data ke dalam data latih dan data uji untuk model pembelajaran mesin bekerja secara efektif. Data dibagi menggunakan rasio 70% untuk pelatihan dan 30% untuk pengujian[21].

### 2.2.6 Optimasi SMOTE

SMOTE digunakan dalam mengatasi ketidakseimbangan kelas yang menghasilkan sampel sintesis pada kelas minoritas, untuk membantu menyeimbangkan kumpulan data. Dalam pendekatan ini, contoh kelas minoritas yang baru dibuat dimasukkan ke dalam data pelatihan, yang memungkinkan pengklasifikasi untuk belajar dari distribusi yang lebih seimbang. Teknik ini sangat berguna untuk kumpulan data dengan distribusi sentimen yang tidak seimbang, seperti diskusi tentang perjudian daring, yang mungkin berisi lebih banyak komentar kritis atau peringatan[22].

### 2.2.7 Model Klasifikasi

Pada proses ini, model klasifikasi diterapkan dalam menganalisis data telah melewati serangkaian tahap sebelumnya, termasuk pembersihan data, ekstraksi fitur, penerapan SMOTE, serta pemisahan dataset. Berikut ini adalah penjelasan algoritma yang digunakan dalam studi ini:

a. *Naïve Bayes Classifier*

Naïve Bayes adalah algoritma klasifikasi langsung berdasarkan Teorema Bayes, dengan asumsi independensi antar fitur. Algoritma ini dapat diterapkan pada data numerik dan kategorikal. Selain itu, Naïve Bayes mampu menangani kumpulan data dominan serta data yang tidak stabil secara efisien. Algoritma ini umum diterapkan di berbagai bidang, seperti analisis sentimen, klasifikasi teks, dan penyaringan spam[9].

b. Random Forest

Random Forest yaitu metode berbasis *ensemble* yang terdiri dari banyak *decision trees*. Algoritma ini bekerja dengan menyatukan prediksi dengan beberapa pohon untuk meningkatkan akurasi serta mengurangi overfitting. Random Forest juga sering digunakan untuk tugas klasifikasi dan regresi.[7].

c. SVM

SVM beroperasi dengan mengidentifikasi *hyperplane* terbaik untuk memisahkan berbagai kelas dalam suatu data. Algoritma ini sangat handal dalam menangani data berdimensi tinggi dan unggul dalam mengklasifikasikan kumpulan data dengan batasan yang terdefinisi dengan baik[11].

### 2.2.8 Evaluasi Model

Pada tahap evaluasi, kinerja algoritma klasifikasi yang digunakan dalam studi ini dinilai untuk menentukan efektivitasnya. Evaluasi ini menggunakan metrik seperti akurasi, presisi, recall, dan skor F1. Hasil evaluasi ini memberikan wawasan tentang kekuatan serta kelambatan masing-masing metode, membantu mengidentifikasi model yang optimal dalam klasifikasi sentimen[23].

## 3. HASIL DAN PEMBAHASAN

Analisis sentimen pada aplikasi PLN Mobile dilakukan secara sistematis dengan menerapkan teknik yang diimplementasikan di Google Colab. Selain itu, bagian ini menjelaskan bagaimana penerapan teknik SMOTE secara efektif meningkatkan kinerja model, khususnya dalam mengatasi ketidakseimbangan data. Setiap tahap, termasuk pengumpulan data, *preprocessing*, pelabelan, serta penerapan SMOTE hingga evaluasi model, akan dibahas secara terperinci.

### 3.1 Pengumpulan Data

Dataset ini dikumpulkan dari komentar pengguna di Google Play Store melalui tahap *crawling*, yang menghasilkan total 20000 ulasan dalam kurun waktu 18 September 2020 sampai dengan 16 Februari 2025, data yang dikumpulkan berbahasa Indonesia dan disimpan dalam bentuk CSV untuk dianalisis dan diproses lebih lanjut.

### 3.2 Preprocessing Data

Data yang diperoleh melalui *crawling* masih dalam format mentah dan belum diolah. Untuk menyempurnakan dan menyusun data agar siap dianalisis, diperlukan serangkaian langkah *preprocessing*. Berikut adalah tahapan-tahapan dalam proses *preprocessing* data.

#### 3.1.1 Cleansing

Pada tahap *cleansing*, data teks dibersihkan dari karakter-karakter yang tidak relevan seperti simbol, emoji, dan tanda baca yang tidak diperlukan untuk analisis. Sebagaimana ditunjukkan pada **Tabel 1**, sebelum dilakukan *cleansing*, teks ulasan masih mengandung simbol yang tidak terbaca di awal kalimat (kemungkinan berupa emoji) serta penggunaan tanda baca koma yang menempel langsung pada kata sebelumnya, seperti pada frasa “Mobile,lebih”. Setelah melalui proses *cleansing*, simbol-simbol tersebut dihapus dan kalimat menjadi lebih bersih serta terstruktur. Selain itu, tanda baca yang mengganggu juga dihilangkan sehingga teks menjadi lebih siap untuk diproses pada tahap tokenisasi dan analisis selanjutnya.

**Tabel 1.** Hasil *Cleansing*

<i>Content</i>	<i>Cleansing</i>
👍👍 beli pulsa di PLN Mobile,lebih murah dan dapat undian setiap bulannya	beli pulsa di PLN Mobile lebih murah dan dapat undian setiap bulannya

#### 3.1.2 Case folding

Pada tahap *case folding*, seluruh huruf dalam data teks diubah menjadi huruf kecil (lowercase) untuk menyeragamkan format penulisan. Hal ini bertujuan untuk menghindari perbedaan makna atau pengelompokan yang tidak akurat akibat perbedaan penggunaan huruf kapital dan huruf kecil dalam teks. Sebagai contoh, seperti yang ditampilkan pada Tabel

di bawah, kata “App Koneksi” diubah menjadi “app koneksi”, “Cek Booking” menjadi “cek booking”, dan seluruh kata lainnya juga dikonversi ke huruf kecil. Proses ini penting untuk menjaga konsistensi data teks sebelum memasuki tahapan analisis lanjutan, seperti tokenisasi atau klasifikasi, sehingga setiap kata dapat dikenali secara seragam tanpa dipengaruhi oleh perbedaan kapitalisasi. Tabel 2 di bawah ini mengilustrasikan perbedaan data teks sebelum dan sesudah tahap *Case Folding*.

**Tabel 2.** Hasil *Casefolding*

<i>Cleansing</i>	<i>Case folding</i>
App Koneksi di HP sy tidak ada mslah tp tiap kali pilih fitur Cek Booking selalu muncul Kesalah Sambungan Tolong perbaiki bug itu jika tidak mau app ini dikategorikan sampah	app koneksi di hp sy tidak ada mslah tp tiap kali pilih fitur cek booking selalu muncul kesalahan sambungan tolong perbaiki bug itu jika tidak mau app ini dikategorikan sampah

### 3.1.3 Normalisasi Data

Pada tahap normalisasi data, kata-kata tidak baku atau bentuk singkatan diubah menjadi bentuk baku sesuai dengan standar bahasa Indonesia, agar tercapai konsistensi dalam teks. Dalam contoh pada **Tabel 3**, kata “aja” yang merupakan bentuk tidak baku disesuaikan menjadi “saja” untuk mengikuti bentuk baku menurut Kamus Besar Bahasa Indonesia (KBBI). Meskipun pada contoh ini hanya terdapat satu perubahan, proses normalisasi secara keseluruhan penting dilakukan untuk menyesuaikan seluruh kata yang tidak baku dalam data, seperti singkatan informal atau bahasa gaul, agar data yang digunakan dalam analisis lebih konsisten dan mudah diproses secara otomatis oleh algoritma analisis teks.

**Tabel 3.** Hasil Normalisasi

<i>Case folding</i>	Normalisasi
lumayan aja bisa cek pemakaian listrik prepaid token pln walaupun waktunya kurang akurat	lumayan saja bisa cek pemakaian listrik prepaid token pln walaupun waktunya kurang akurat

### 3.1.4 Tokenizing

Pada tahap tokenizing, kalimat hasil normalisasi dipecah menjadi potongan-potongan kata atau token agar dapat diproses lebih lanjut oleh algoritma analisis teks. Setiap kata dan tanda baca yang sebelumnya tergabung dalam satu kalimat dipisahkan menjadi elemen-elemen individual. Misalnya, kalimat “Kenapa aplikasi ini tiba-tiba tidak jelas? Saya dulu sudah daftar, tapi sekarang mau masuk, aktivasi gagal terus.” diubah menjadi daftar token seperti ["Kenapa", "aplikasi", "ini", "tiba-tiba", "tidak", "jelas", ...]. Tahapan ini sangat penting untuk mempermudah analisis sentimen karena model memerlukan input berupa token agar dapat mengenali struktur, frekuensi, dan makna kata dalam teks secara lebih akurat. Dengan memisahkan kata-kata ini, proses penghitungan kata, analisis konteks, dan penerapan teknik pembobotan menjadi lebih efektif. Tabel 4 di bawah ini menyajikan contoh perbedaan data ulasan sebelum dan sesudah tahap tokenizing.

**Tabel 4.** Hasil Tokenizing

Normalisasi	Tokenizing
Kenapa aplikasi ini tiba-tiba tidak jelas? Saya dulu sudah daftar, tapi sekarang mau masuk, aktivasi gagal terus.	["Kenapa", "aplikasi", "ini", "tiba-tiba", "tidak", "jelas", "?", "Saya", "dulu", "sudah", "daftar", ",", "tapi", "sekarang", "mau", "masuk", ",", "aktivasi", "gagal", "terus", "."]

### 3.1.5 Stopword Removal

Stopword removal yaitu proses dalam pemrosesan teks yang menghilangkan kata-kata umum yang tidak memberikan makna penting dalam analisis, seperti "untuk," "bisa," serta "karena," dihilangkan. Proses ini membantu mengurangi kompleksitas komputasi dan meningkatkan relevansi kata-kata dalam analisis sentimen[19]. Tabel 5 di bawah ini menyajikan contoh perbedaan data ulasan sebelum dan sesudah tahap penghapusan kata umum.

**Tabel 5.** Hasil penghapusan kata umum

Tokenizing	Stopword Removal
['semoga', 'untuk', 'tagihan', 'bisa', 'update', 'terus', 'karena', 'saya', 'setiap', 'buka', 'aplikasi', 'tagihan', 'tidak', 'berubah', 'supaya', 'berubah', 'harus', 'log', 'out', 'dulu', 'terus', 'log', 'in', 'lagi', 'mungkin', 'bisa', 'ditambahkan', 'fitur', 'tarik', 'untuk', 'refresh', 'jadi', 'untuk', 'update', 'tagihan', 'tidak', 'harus', 'log', 'out', 'dulu']	['semoga', ',', 'tagihan', ',', 'update', 'terus', ',', ',', ',', 'buka', 'aplikasi', 'tagihan', ',', 'berubah', ',', 'berubah', ',', 'log', 'out', 'dulu', 'terus', 'log', 'in', ',', 'mungkin', ',', 'ditambahkan', 'fitur', 'tarik', ',', 'refresh', 'jadi', ',', 'update', 'tagihan', ',', ',', 'log', 'out', 'dulu']



Hasil worcloud gambar 5 menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna terhadap aplikasi PLN Mobile. Semakin besar ukuran kata, semakin sering kata tersebut muncul dalam data. Dari visualisasi ini, terlihat bahwa kata "pln", "mobile", "aplikasi", "bantu", dan "terima kasih" menjadi kata yang paling dominan. Hal ini menunjukkan bahwa banyak pengguna memberikan ulasan positif terkait kemudahan penggunaan dan bantuan yang diberikan melalui aplikasi PLN Mobile. Kata-kata seperti "mudah", "cepat", "token", dan "lapor" juga sering muncul, yang mengindikasikan bahwa fitur pembelian token dan pelaporan gangguan cukup sering disebut dalam komentar pengguna.

### 3.5 Feature Extraction

Dalam tahap *feature extraction*, penelitian ini menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)* untuk mengonversi teks ke dalam bentuk numerik yang dapat diproses oleh model machine learning. Teknik ini memberikan bobot lebih rendah pada kata-kata yang sering muncul tetapi kurang bermakna, sementara kata-kata yang jarang muncul namun memiliki signifikansi tinggi diberi bobot lebih besar. Dengan demikian, fitur yang diekstraksi lebih menekankan kata-kata yang memiliki peran penting dalam menentukan sentimen dalam teks. Hasil ekstraksi fitur TF-IDF disajikan dalam Tabel 7.

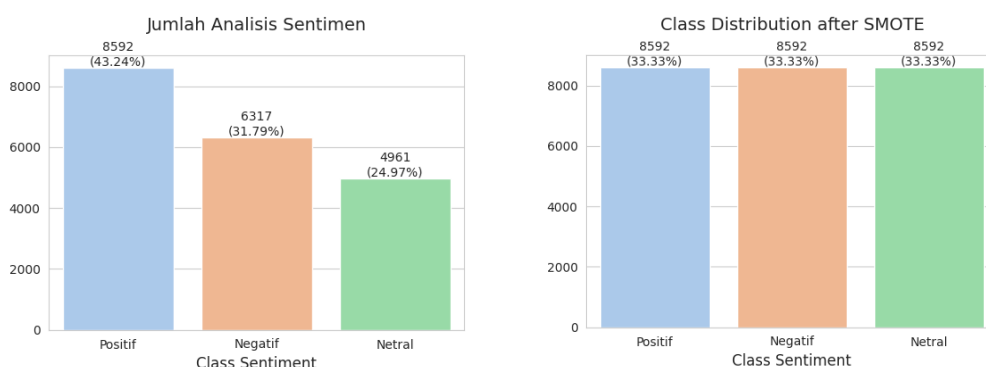
**Tabel 7.** Hasil TF-IDF

	Term	TFIDF_SUM
1000	pln	1670.870773
504	aplikasi	1565.659614
8208	mudah	1530.337908
7964	Mobile	1288.774188
1009	bantu	1283.328390

Tabel 7 menyajikan lima kata dengan nilai TF-IDF tertinggi yang ditemukan dalam ulasan Play Store terkait aplikasi PLN Mobile. Skor TF-IDF mencerminkan signifikansi kata dalam dokumen relatif terhadap seluruh kumpulan data. Kata-kata dengan nilai tinggi, seperti "pln", "aplikasi," dan "mudah," menyoroti topik yang dominan dalam opini publik. Kata-kata ini penting dalam membedakan sentimen selama analisis.

### 3.6 Optimasi SMOTE

SMOTE berguna untuk menangani ketidakseimbangan kelas dalam dataset dengan menghasilkan sampel sintesis pada kelas yang jumlahnya lebih sedikit, sehingga membantu menciptakan distribusi data yang lebih proporsional dan meningkatkan performa model dalam mengenali pola dari semua kelas. Gambar 6 menunjukkan ilustrasi distribusi dataset sebelum dan setelah penerapan SMOTE.



**Gambar 6.** Visualisasi sebelum dan sesudah di SMOTE

Gambar 6 mengilustrasikan distribusi kumpulan data sebelum dan setelah penerapan SMOTE. Pada grafik sebelah kiri, data asli menunjukkan jumlah data positif sebanyak 8592 (43,24%), negatif 6317 (31,79%), dan netral 4961 (24,97%). Ketimpangan ini membuat model cenderung mengklasifikasikan data ke kelas positif. Setelah dilakukan SMOTE, seperti yang terlihat pada grafik sebelah kanan, distribusi jumlah data menjadi seimbang untuk ketiga kelas, yaitu masing-masing sebanyak 8592 data (33,33%).

Penyeimbangan ini penting karena akan membantu model dalam mengenali karakteristik dari masing-masing kelas secara lebih adil. Hasilnya, model menjadi lebih responsif terhadap sentimen minoritas (negatif dan netral), serta mampu meningkatkan performa dan akurasi klasifikasi secara keseluruhan.

### 3.7 Tahap Pengujian

Setelah melalui serangkaian proses *preprocessing data*, tahap selanjutnya adalah melakukan pengujian model untuk mengukur performa algoritma yang digunakan. Dalam penelitian ini, algoritma yang diuji meliputi Naïve Bayes, Random Forest, dan Support Vector Machine (SVM). Untuk proses pengujian, data dibagi menjadi dua bagian, yaitu

70% sebagai data latih (training data) dan 30% sebagai data uji (testing data). Data latih digunakan untuk melatih model dalam mengenali pola, sedangkan data uji digunakan untuk mengukur kemampuan generalisasi model terhadap data baru yang belum pernah dilihat sebelumnya. Tabel 8 menunjukkan ilustrasi *data collection*.

**Tabel 8.** Data Collection

Matriks	Naive Bayes (NB)	Naïve Bayes (NB) + SMOTE	Random Forest (RF)	Random Forest (RF) + SMOTE	SVM	SVM + SMOTE
Accuracy Sentimen Negatif	0,62	0,67	0,79	0,85	0,92	0,95
Precision Sentimen Negatif	0,79	0,69	0,79	0,84	0,93	0,97
Recall Sentimen Negatif	0,61	0,72	0,87	0,91	0,94	0,95
F1-Score Sentimen Negatif	0,61	0,7	0,83	0,88	0,93	0,96
Precision Sentimen Positif	0,56	0,76	0,79	0,86	0,96	0,98
Recall Sentimen Positif	0,96	0,71	0,92	0,85	0,95	0,93
F1-Score Sentimen Positif	0,71	0,73	0,85	0,85	0,95	0,95
Precision Sentimen Netral	0,49	0,59	0,79	0,85	0,85	0,9
Recall Sentimen Netral	0,03	0,6	0,46	0,78	0,86	0,96
F1-Score Sentimen Netral	0,05	0,59	0,58	0,82	0,86	0,93

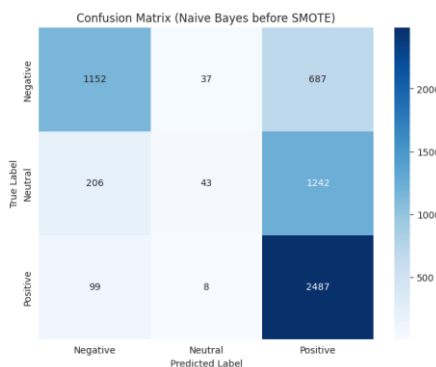
Berdasarkan dari tabel 8, semua algoritma menunjukkan tingkat akurasi yang bervariasi, berkisar antara 0,62 hingga 0,95. Penerapan teknik SMOTE meningkatkan akurasi semua model, khususnya Naïve Bayes (NB) mengalami peningkatan dari 0,62 menjadi 0,67, Random Forest meningkat dari 0,79 menjadi 0,85, dan Support Vector Machine (SVM) meningkat dari 0,92 menjadi 0,95.

Untuk sentimen negatif, SMOTE secara signifikan meningkatkan recall di semua algoritma. Misalnya, recall dalam NB meningkat dari 0,61 menjadi 0,72, sedangkan dalam RF, naik dari 0,72 menjadi 0,87. Tetapi, presisi sedikit menurun dalam NB dan RF, dengan NB turun dari 0,79 menjadi 0,69 serta RF dari 0,79 menjadi 0,78. Di sisi lain, SVM menunjukkan peningkatan presisi yang nyata, meningkat dari 0,84 menjadi 0,97 sesudah di SMOTE. Dalam hal Skor F1, terdapat sedikit peningkatan atau stabilitas, dengan RF meningkat dari 0,83 menjadi 0,88, sementara SVM mempertahankan nilai tinggi sebesar 0,93.

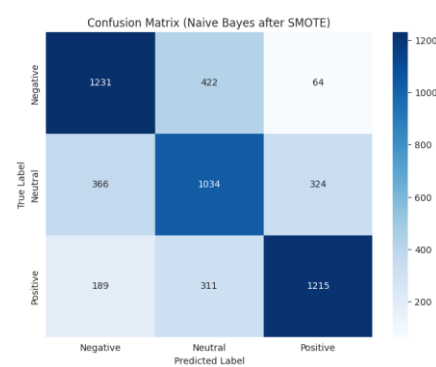
Untuk sentimen positif, SMOTE menyebabkan perolehan kembali yang lebih tinggi di seluruh algoritme. Penarikan kembali NB meningkat dari 0,56 menjadi 0,76, RF dari 0,79 menjadi 0,92, serta SVM dari 0,96 menjadi 0,98. Presisi tetap stabil atau sedikit membaik, seperti pada RF, yang tetap pada 0,85, sementara SVM meningkat sedikit dari 0,96 menjadi 0,98. Skor F1 juga menunjukkan peningkatan pada model tertentu, dengan NB meningkat dari 0,71 menjadi 0,73, dan RF mempertahankan stabilitas pada 0,85.

Dalam sentimen netral, SMOTE memiliki dampak signifikan pada peningkatan daya ingat, khususnya dalam NB, yang naik dari 0,03 menjadi 0,06, serta RF, yang melonjak dari 0,46 menjadi 0,78. Presisi mengalami peningkatan dalam SVM, naik dari 0,85 menjadi 0,90. Namun, NB mengalami sedikit penurunan presisi dari 0,49 menjadi 0,59. Skor F1 meningkat di semua model, terutama dalam SVM, yang naik dari 0,86 menjadi 0,93.

Secara keseluruhan, teknik SMOTE secara efektif meningkatkan kinerja model, khususnya dalam daya ingat, yang berarti model menjadi lebih baik dalam mengenali kelas minoritas. Namun, beberapa trade-off dalam presisi diamati dalam model tertentu. Algoritme SVM dengan SMOTE menghasilkan hasil terbaik, mencapai akurasi 0,95 dan Skor F1 tertinggi di semua kategori sentimen. gambar di bawah ini adalah hasil matriks konfusi.



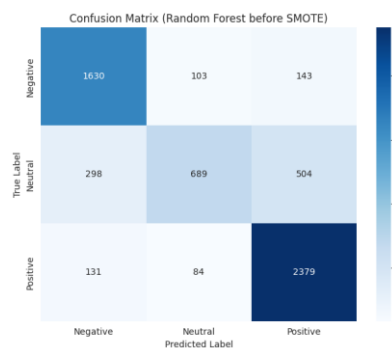
**Gambar 7.** Matriks Konfusi *Naive Bayes* Sebelum SMOTE



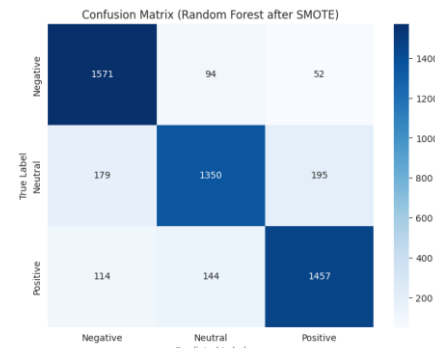
**Gambar 8.** Matriks Konfusi *Naive Bayes* Sesudah SMOTE

Pada Gambar 7 dan Gambar 8 penerapan SMOTE pada Naïve Bayes meningkatkan keseimbangan prediksi antar kelas, terutama dalam mengenali kelas negatif, di mana negatif sebenarnya meningkat dari 1152 menjadi 1231. Namun, kesalahan klasifikasi meningkat, terutama kelas Negatif yang lebih sering salah diklasifikasikan sebagai Netral (37 menjadi 422).

Pada kelas Netral, jumlah prediksi benar menurun (1242 menjadi 1034), dengan lebih banyak kesalahan ke Negatif dan Positif. Sementara itu, kelas Positif mengalami penurunan True Positive dari 2487 menjadi 1215, dengan peningkatan kesalahan klasifikasi ke Netral (8 menjadi 311). Secara keseluruhan, SMOTE meningkatkan recall tetapi menurunkan precision, terutama dalam membedakan kelas Netral dan lainnya.



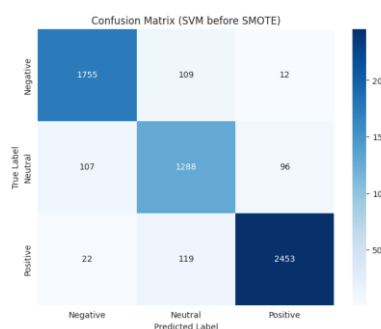
**Gambar 9.** Matriks Konfusi *Random Forest* Sebelum SMOTE



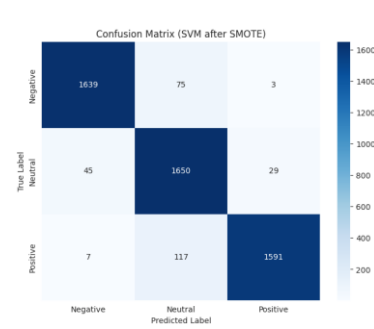
**Gambar 10.** Matriks Konfusi *Random Forest* Sesudah SMOTE

Untuk mengetahui dampak penerapan SMOTE terhadap performa klasifikasi, dilakukan analisis terhadap confusion matrix model Random Forest sebelum dan sesudah SMOTE diterapkan **Gambar 9** dan **Gambar 10** di atas ini menunjukkan perbandingan matriks hasil prediksi model. Penerapan SMOTE pada Random Forest meningkatkan prediksi untuk kelas Netral, dengan True Neutral naik dari 689 menjadi 1350. Namun, kelas Negatif sedikit menurun (1630 menjadi 1571), meskipun kesalahan klasifikasinya ke Netral berkurang (298 menjadi 179). Pada kelas Positif, jumlah True Positive menurun dari 2379 menjadi 1457, sementara kesalahan ke Netral meningkat dari 94 menjadi 144.

Secara keseluruhan, SMOTE berhasil meningkatkan sensitivitas model terhadap kelas minoritas (Netral), yang sebelumnya kurang terwakili. Namun, hal ini juga menyebabkan sedikit penurunan presisi pada kelas mayoritas (Positif), yang mengindikasikan adanya trade-off antara recall dan precision.



**Gambar 11.** Matriks Konfusi SVM Sebelum SMOTE

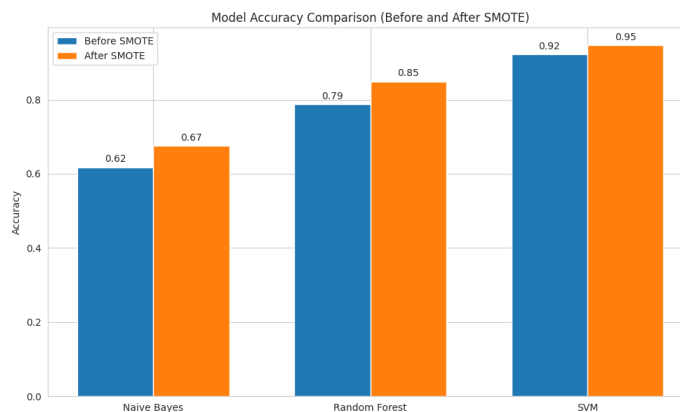


**Gambar 12.** Matriks Konfusi SVM Sesudah SMOTE

Untuk mengetahui dampak penerapan SMOTE terhadap performa klasifikasi, dilakukan analisis terhadap confusion matrix model SVM sebelum dan sesudah SMOTE diterapkan **Gambar 11** dan **Gambar 12** di atas ini menunjukkan perbandingan matriks hasil prediksi model. Penerapan SMOTE pada SVM meningkatkan True Neutral secara signifikan (1288 menjadi 1650), menunjukkan model lebih baik dalam mengenali kelas Netral. True Negative juga meningkat (1755 menjadi 1639), sementara kesalahan klasifikasi ke kelas Netral berkurang (109 menjadi 73). Pada kelas Positif, jumlah True Positive menurun dari 2463 menjadi 1591, dengan peningkatan kesalahan ke Netral (119 menjadi 117).

Secara keseluruhan, penerapan SMOTE pada model SVM membantu menyeimbangkan performa antar kelas, khususnya dengan meningkatkan recall pada kelas minoritas seperti Netral. Namun, hal ini juga mengorbankan presisi dan akurasi pada kelas mayoritas (Positif), yang terlihat dari menurunnya jumlah prediksi benar dan meningkatnya kesalahan ke kelas Netral.

### 3.8 Visualisasi Perbandingan Algoritma



**Gambar 13.** Visualisasi perbandingan akurasi algoritma

Gambar 10 di atas menyajikan grafik batang yang membandingkan akurasi model Naïve Bayes, Random Forest, dan SVM sebelum dan setelah penerapan metode SMOTE. Grafik tersebut menggambarkan ketiga model mengalami peningkatan akurasi setelah SMOTE diterapkan. Peningkatan yang paling menonjol terlihat pada model SVM, di mana akurasi meningkat dari 92% menjadi 95%. Model Random Forest juga menunjukkan pertumbuhan yang signifikan, dari 79% menjadi 85%. Sementara itu, model Naïve Bayes menunjukkan peningkatan sedang dari 62% menjadi 67%. Visualisasi ini menegaskan bahwa SMOTE secara efisien menangani masalah tidak keseimbangan kelas serta menambah kinerja ketiga algoritma dalam klasifikasi sentimen.

## 4. KESIMPULAN

Penelitian ini mengkaji sentimen publik terkait komentar aplikasi PLN Mobile pada playstore dengan metode Naïve Bayes, Random Forest, dan SVM. Selain itu, penelitian ini mengevaluasi efektivitas teknik SMOTE dalam mengatasi ketidakseimbangan data dalam kumpulan data. Temuan studi menunjukkan bahwa SMOTE sangat meningkatkan performa semua model, khususnya dalam mengklasifikasikan sentimen negatif dan netral yang sebelumnya kurang terwakili. Setelah penerapan SMOTE, SVM mencapai akurasi tertinggi sebesar 95%, kemudian Random Forest sebesar 85%, dan Naïve Bayes sebesar 67%. Teknik ini berhasil menyeimbangkan distribusi kelas sentimen, sehingga meningkatkan akurasi model secara keseluruhan. Temuan penelitian ini menyoroti pentingnya SMOTE dalam analisis sentimen yang melibatkan kumpulan data yang tidak seimbang, terutama dalam konteks media sosial, dan menunjukkan potensinya untuk meningkatkan prediksi dalam model klasifikasi sentimen di masa mendatang. Lebih jauh, hasil penelitian menunjukkan bahwa persepsi publik terhadap PLN Mobile bervariasi, dengan sentimen netral dan negatif yang cukup menonjol. Studi ini menawarkan wawasan berharga untuk pengembang PLN Mobile dalam meningkatkan kualitas layanan berdasarkan analisis sentimen pengguna. Namun, keterbatasan utamanya adalah data hanya bersumber dari satu platform. Penelitian di masa mendatang dapat menggabungkan data dari berbagai sumber dan menerapkan teknik analisis tingkat lanjut, seperti pembelajaran mendalam, untuk memperoleh pemahaman yang lebih komprehensif tentang opini pengguna. Pendekatan ini akan memungkinkan pengembangan PLN Mobile yang berkelanjutan agar lebih sesuai dengan harapan dan kebutuhan pelanggan.

## REFERENCES

- [1] H. Faisal, A. Febriandirza, and F. N. Hasan, "Analisis Sentimen Terkait Ulasan Pada Aplikasi PLN Mobile Menggunakan Metode Support Vector Machine," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 5, no. 1, pp. 303–312, 2024. doi: <https://doi.org/10.30645/kesatria.v5i1.339>
- [2] Ihsan Zulfahmi, "Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decision Tree," *J. Penelit. Rumpun Ilmu Tek.*, vol. 3, no. 1, pp. 11–21, 2023, doi: [10.55606/juprit.v3i1.3096](https://doi.org/10.55606/juprit.v3i1.3096).
- [3] U. Brawijaya, K. R. Hilal, N. Y. Setiawan, and D. E. Ratnawati, "Fakultas Ilmu Komputer Analisis Sentimen Berbasis Aspek Untuk Pengguna PLN Mobile Pada Google Playstore Menggunakan Metode Support Vector Machine (SVM)," *JPTIHK*, vol. 1, no. 1, pp. 2548–964, 2017
- [4] H. B. Tambunan and T. W. D. Hapsari, "Analisis Opini Pengguna Aplikasi New PLN Mobile Menggunakan Text Mining," *Petir*, vol. 15, no. 1, pp. 121–134, 2021, doi: [10.33322/petir.v15i1.1352](https://doi.org/10.33322/petir.v15i1.1352).
- [5] A. Komarudin and A. M. Hilda, "Analisis Sentimen Ulasan Aplikasi Identitas Kependudukan Digital Pada Play Store Menggunakan Metode Naïve Bayes," *Comput. Sci.*, vol. 4, no. 1, pp. 28–36, 2024, doi: [10.31294/coscience.v4i1.2955](https://doi.org/10.31294/coscience.v4i1.2955).
- [6] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: [10.30591/jpit.v3i3.977](https://doi.org/10.30591/jpit.v3i3.977).
- [7] C. Mario, R. R. Suryono, "Public Sentiment Analysis on Dirty Vote Movie on YouTube using Random Forest and Naïve Bayes," *Jurnal INOVTEX Polbeng*, vol. 10, no. 1, pp. 111–122, 2025. doi: <https://doi.org/10.35314/ev9j2g33>
- [8] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve



- Bayes Classifier dan K-Nearest Neighbor,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [9] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, “Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, 2021, doi: 10.32493/informatika.v5i4.7575.
- [10] A. Carla and H. Soetanto, “Implementasi Text Mining untuk Analisis Sentimen pada Pengguna PLN Mobile Menggunakan Metode Naive Bayes,” *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 21, no. 1, p. 72, 2024, doi: 10.36080/bit.v21i1.2974.
- [11] Imaniar Ikko Mulya Rizky, Suhendro Yusuf Irianto, and Sriyanto, “Perbandingan Kinerja Algoritma Naive Bayes, Support Vector Machine dan Random forest untuk Prediksi Penyakit Ginjal Kronis,” *Semin. Nas. Has. Penelit. dan Pengabd. Masy.*, vol. 18, pp. 139–151, 2023.
- [12] Sari, Putri Kumala, and Ryan Randy Suryono. “Komparasi algoritma Support Vector Machine dan Random Forest untuk analisis sentimen metaverse.” *Jurnal Mnemonic* 7.1 (2024): 31-39. <https://doi.org/10.36040/mnemonic.v7i1.8977>
- [13] R. R. Pratama, R. R. Suryono, “Performance Comparison Of Naive Bayes, Support Vector Machine And Random Forest Algorithm For Apple Vision Pro Sentiment Analysis,” *JUTIF*, vol. 6, no. 1, pp. 31–39, 2025. doi: <https://doi.org/10.52436/1.jutif.2025.6.1.4035>
- [14] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 16–22, 2021, doi: 10.30813/jbase.v4i2.3000.
- [15] I. N. Rizki, D. Prayoga, M. L. Puspita, and M. Q. Huda, “Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3856.
- [16] F. A. Artanto, “Implementasi Algoritma Random Forest dan Model Bag of Words Dalam Analisis Sentimen Mengenai E-Materai,” *Jurnal Satesi*, vol. 4, no. 2, pp. 139–145, 2024, doi: 10.54259/satesi.v4i2.3240.
- [17] Z. Efendy, “Normalisasi dalam Desain Database,” *J. CoreIT*, vol. 4, no. 1, pp. 34–43, 2018.
- [18] U. Pujianto, I. A. E. Zaeni, and K. I. Rasyida, “Comparison of Naive Bayes and Random Forests Classifier in the Classification of News Article Popularity as Learning Material,” *Proc. Ist UMGESHIC Int. Semin. Heal. Soc. Sci. Humanit. (UMGESHIC-ISHSSH 2020)*, vol. 585, pp. 229–242, 2021, doi: 10.2991/assehr.k.211020.036.
- [19] A. Santosa, I. Purnamasari, and Mayasari Rini, “Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM,” *J. Sains Komput. Inform.*, vol. 6, pp. 81–93, 2022. doi: <http://dx.doi.org/10.30645/j-sakti.v6i1.427>
- [20] Aminullah, Zulkifli, Deni Arifianto, and Habibatul Azizah Al Faruq, “Lexicon-Based Approach Pada Analisis Sentimen Ulasan Airbnb Menggunakan Vader Sentiment Lexicon-Based Approach to Sentiment Analysis of Airbnb Reviews Using Vader Sentiment,” *Jurnal Smart Teknologi*, vol. 5, no. 4, pp. 558–566, 2024.
- [21] Y. A. Prasetyo, E. Utami, and A. Yaqin, “Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naive Bayes dan SVM,” *Journal of Electrical Engineering and Computer*, vol. 6, no. 2, pp. 382–390, 2024, doi: 10.33650/jeeecom.v4i2.
- [22] S. Maula Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, “Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan kNN,” *Syntax J. Inform.*, vol. 11, no. 02, pp. 56–67, 2022, doi: 10.35706/syji.v11i02.6940.
- [23] D. N. Agustia, R. R. Suryono, “Comparison of Naive Bayes, Random Forest, and Logistic Regression Algorithms for Sentiment Analysis Online Gambling” *Jurnal INOVTEK Polbeng*, vol. 10, no. 1, pp. 284–295, 2025: <https://doi.org/10.35314/prk93630>
- [24] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, “Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile,” *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [25] B. F. S. Supriyanto and S. Rosalin, “Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency,” *J. Minfo Polgan*, vol. 12, no. 1, pp. 25–32, 2023, doi: 10.33395/jmp.v12i1.12312.