Jurnal TRANSFORMATIKA
Vol.24, No.1, July 2026, pp. 32 - 44
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Analisis Sentimen Ulasan Aplikasi *PLN Mobile* Menggunakan *Naïve Bayes, SVM*, dan *Random Forest* dengan SMOTE

Helmy Agta Al Fatah [1*] , Rara Sriartati Redjeki [2]
[1.] Universitas Stikubank / Sistem Informasi
Jl.Tri Lomba Juang No 1, Kota Semarang, e-mail: helmyagta3001@mhs.unisbank.ac.id
[2.] Universitas Stikubank / Sistem Informasi
Jl.Tri Lomba Juang No 1, Kota Semarang, e-mail: rara_artati@edu.unisbank.ac.id

ARTICLE INFO

History of the article:
Received 22 Juni 2026
Received in revised form 20 Juli 2026
Accepted 21 Juli 2026
Available online 25 Juli 2026

Keywords:

Analisis Sentimen; *PLN Mobile*; *Machine Learning*; SMOTE; TF-IDF

*** Correspondence:**

Telepon :
-

E-mail:
helmyagta3001@mhs.unisbank.ac.id

ABSTRACT

Aplikasi PLN Mobile sebagai saluran layanan digital PT PLN (Persero) menerima ribuan ulasan pengguna di *Google Play Store* yang sulit dianalisis secara manual. Penelitian ini bertujuan membandingkan kinerja tiga algoritma *machine learning Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* untuk mengklasifikasikan sentimen ulasan PLN Mobile ke dalam dua kategori (positif dan negatif). Ketiga algoritma dipilih karena mewakili tiga paradigma berbeda (*probabilistik*, berbasis *margin*, dan *ensemble*) yang terbukti efektif untuk klasifikasi teks. Kebaruan penelitian terletak pada kombinasi strategi pengumpulan data berimbang per kategori rating, penyeimbangan kelas dengan *Synthetic Minority Over-sampling Technique* (SMOTE), validasi silang 5-lipat, dan penyetelan *hiperparameter* SVM. Sebanyak 3.702 ulasan dikumpulkan melalui *web scraping*; setelah *prapemrosesan* teks dan pembobotan TF-IDF diperoleh 2.917 data (1.586 negatif dan 1.331 positif) dengan 4.160 fitur. Hasil *cross-validation* menunjukkan *Naïve Bayes* memperoleh rata-rata skor F1 tertinggi (0,846) dengan standar deviasi rendah (0,011), menandakan kinerja yang stabil. Pada pengujian terhadap 584 data uji, *Naïve Bayes* kembali menjadi yang terbaik dengan akurasi, presisi, *recall*, dan F1 sebesar 84%, diikuti SVM (83%) dan *Random Forest* (81%). Analisis kata dominan mengungkap apresiasi pengguna terhadap kemudahan dan kecepatan layanan, serta keluhan utama mengenai pemadaman listrik yang lama dan kendala proses pembayaran serta pengaduan. Temuan ini menegaskan keunggulan *Naïve Bayes* pada teks ulasan berbahasa Indonesia yang relatif sederhana sekaligus memberikan masukan praktis untuk meningkatkan kualitas layanan PLN Mobile.

PENDAHULUAN

Kemajuan di bidang teknologi informasi telah mendorong berbagai penyedia layanan publik untuk melakukan transformasi digital, termasuk di sektor ketenagalistrikan Indonesia. PT PLN (Persero) menghadirkan aplikasi PLN Mobile sebagai kanal layanan terpadu yang

memungkinkan pelanggan melakukan pembelian *token* listrik, pembayaran tagihan, pencatatan meter mandiri, pengaduan gangguan, serta memperoleh informasi pemadaman secara langsung melalui perangkat bergerak [1], [2]. Kehadiran aplikasi ini bertujuan meningkatkan kemudahan dan kualitas layanan kepada jutaan pelanggan PLN di seluruh Indonesia.

Sebagai aplikasi yang tersedia di *Google Play Store*, PLN Mobile menerima ribuan ulasan dari pengguna yang memuat beragam ekspresi kepuasan maupun keluhan, seperti kendala pemuatan lokasi saat pengaduan serta saldo terpotong namun kode *token* tidak muncul [3]. Ulasan tersebut menjadi umpan balik penting bagi pengembang untuk mengevaluasi dan memperbaiki kualitas layanan. Namun, jumlah ulasan yang sangat besar menyulitkan analisis secara manual, sehingga diperlukan pendekatan otomatis berbasis analisis sentimen [4]. Analisis sentimen adalah cabang dari pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam suatu teks, seperti ke dalam kategori positif dan negatif [5]. Pada tahap klasifikasi, bobot kata biasanya ditentukan oleh metode *Term Frequency Inverse Document Frequency* (TF-IDF), yang digambarkan sebagai vektor digital berdasarkan frekuensi dan signifikansi kata tersebut [6]. Berbagai algoritma *machine learning* telah terbukti efektif untuk tugas ini, di antaranya *Naïve Bayes* yang berbasis probabilistik, *Support Vector Machine* (SVM) yang mencari *hyperplane* pemisah optimal, dan *Random Forest* yang mengagregasi banyak *decision tree* [7], [8]. Sejumlah penelitian terdahulu telah menerapkan ketiga algoritma tersebut pada ulasan aplikasi di *Google Play Store*. Penelitian pada aplikasi Ajaib menunjukkan *Random Forest* unggul pada *recall* sementara SVM mencatat akurasi tertinggi [7], sedangkan studi pada aplikasi Halo BCA [8] dan Gojek [9] memberikan kesimpulan yang bervariasi mengenai algoritma terbaik.

Dalam konteks PLN Mobile itu sendiri, penelitian sebelumnya umumnya hanya membandingkan dua algoritma, seperti *Naïve Bayes* dan *K-Nearest Neighbour* [3], atau berfokus pada satu metode tertentu seperti *Random Forest* [1] dan SVM berbasis aspek [2]. Selain itu, banyak penelitian belum menangani persoalan ketidakseimbangan kelas yang lazim terjadi pada data ulasan, padahal distribusi kelas yang timpang dapat menghasilkan akurasi tinggi yang menyesatkan. Teknik *Synthetic Minority Over-sampling* (SMOTE) telah terbukti mampu meningkatkan kinerja klasifikasi pada data yang tidak seimbang [10], [11], [12].

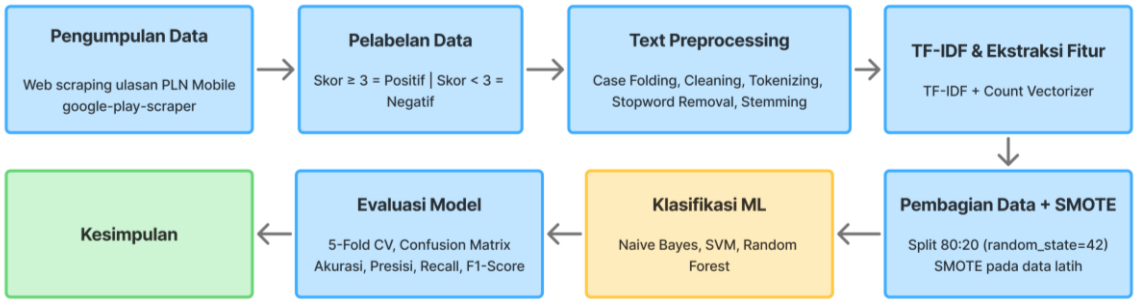
Berdasarkan pembahasan di atas, penelitian ini mengevaluasi opini aplikasi *mobile* PLN dengan algoritma *Naive Bayes*, SVM, dan *Random Forest* menggunakan skema peringkat dua kelas (positif dan negatif) serta membandingkan kinerja dengan skema pemeringkatan dua kelas (positif dan negatif) aplikasi *mobile* PLN. Kontribusi penelitian ini meliputi: (1) penerapan strategi pengumpulan data berimbang berdasarkan kategori *rating*; (2) penanganan ketidakseimbangan kelas menggunakan SMOTE; serta (3) validasi yang lebih kuat melalui 5-*fold cross validation* dan *tuning hyperparameter*, untuk menghasilkan evaluasi yang lebih objektif dan *akuntabel*.

Pemilihan ketiga algoritma tidak dilakukan secara acak, melainkan karena ketiganya sama-sama terbukti efektif untuk klasifikasi teks sekaligus mewakili tiga paradigma pembelajaran mesin yang berbeda: *Naïve Bayes* (*probabilistik*), *Support Vector Machine* (berbasis *margin*/geometris), dan *Random Forest* (*ensemble* berbasis pohon). Perbandingan tiga paradigma yang berbeda pada domain dan kondisi data yang sama memungkinkan identifikasi karakteristik algoritma yang paling sesuai untuk teks ulasan berbahasa Indonesia yang relatif singkat dan didominasi kata kunci spesifik. Kesenjangan dari studi terdahulu adalah: (1) penelitian sentimen PLN Mobile umumnya hanya membandingkan dua algoritma atau berfokus pada satu metode [1], [2], [3]; (2) sebagian besar belum menangani ketidakseimbangan kelas secara eksplisit sehingga akurasi berpotensi menyesatkan; dan (3) evaluasi jarang diperkuat dengan validasi silang maupun penyetelan *hiperparameter*. Dengan demikian, kebaruan penelitian ini terletak pada kerangka evaluasi terpadu yang lebih *akuntabel*, yakni kombinasi pengumpulan data berimbang per kategori *rating*, penyeimbangan kelas dengan SMOTE hanya pada data latih, validasi 5-*fold cross validation*, dan

optimasi *hiperparameter* SVM dengan *GridSearchCV*, yang belum diterapkan secara utuh pada objek ulasan PLN Mobile.

METODE PENELITIAN

Penelitian ini secara kuantitatif membandingkan kinerja tiga algoritma *machine learning* dengan *mengkategorikan* sentimen dalam ulasan pengguna dalam aplikasi mobile PLN. Tahapan penelitian meliputi pengumpulan data, penandaan label, *pra-pemrosesan* teks, pembobotan dan ekstraksi fitur, pemisahan data, penyeimbangan kelas, pelatihan dan evaluasi model, disusun secara sistematis dan berurutan. Seluruh proses tersebut dilaksanakan menggunakan bahasa pemrograman *Python* di lingkungan *Google Colaboratory*. Alur kerja penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1. Alur/diagram metode penelitian

1. Pengumpulan Data

Data dari penelitian ini terdiri dari umpan balik dari pengguna aplikasi *PLN Mobile* yang dikumpulkan dengan *web scraping* di *Google Play Store* menggunakan pustaka *google-play-scraper* dalam bahasa pemrograman *Python*. Objek penelitian adalah aplikasi *PLN Mobile* dengan *package* com.icon.pln123, yaitu aplikasi layanan kelistrikan resmi PT PLN (Persero) yang memungkinkan pelanggan membeli *token* listrik, membayar tagihan, dan melaporkan gangguan. Untuk menghindari dominasi ulasan berbintang tinggi yang dapat memicu ketidakseimbangan kelas, pengambilan data dilakukan secara berimbang dengan memfilter ulasan berdasarkan masing-masing kategori *rating* bintang 1 hingga 5 (maksimal 800 ulasan per kategori). Setelah penghapusan data duplikat, diperoleh 3.702 ulasan unik.

2. Pelabelan Data

Penandaan dilakukan secara otomatis berdasarkan skor peringkat yang diberikan oleh pengguna, mengacu pada pendekatan yang biasa digunakan dalam penelitian analisis sentimen dalam ulasan aplikasi. Ulasan dengan skor di atas 3 dikategorikan sebagai sentimen positif, dan skor di bawah 3 sebagai sentimen negatif. Ulasan dengan skor tepat 3 (netral) dikeluarkan dari analisis agar penelitian berfokus pada klasifikasi dua kelas (biner). Hasil pelabelan memperoleh 1.589 ulasan negatif dan 1.341 ulasan positif.

3. *Text Preprocessing*

Prapemrosesan teks merupakan langkah awal dalam membersihkan dan mempersiapkan data teks agar terstruktur dan bebas dari unsur-unsur yang tidak relevan [13], [14]. Seluruh proses dilakukan secara otomatis menggunakan *Python* melalui tahapan berikut:

a. Tahap *case folding* menyetarakan format teks dengan mengonversi setiap karakter ke huruf kecil.
b. *Text Cleaning* menghapus tanda baca, angka, URL, simbol, *emoji*, dan huruf berulang, diikuti dengan menormalkan kata-kata non-standar (bahasa gaul) ke dalam bentuk standar.
c. *Tokenizing* Bagi teks menjadi unit teks (*token*) untuk menyederhanakan pemrosesan.
d. *Stopword Removal* menghapus kata-kata umum tanpa arti analitis (seperti "dan", "yang", "di") menggunakan kamus kata kunci Indonesia dari pustaka Sastrawi.
e. *Stemming* Ubah kata sifat ke bentuk dasarnya menggunakan *stemmer* Sastrawi yang menerapkan algoritma Nazief Andriani dalam bahasa Indonesia[15].

Setelah *praproses* dan penghapusan teks kosong, diperoleh 2.917 data bersih (1.586 negatif dan 1.331 positif).

Tabel 1. Contoh hasil tiap tahap *preprocessing*

Tahap	Contoh Ulasan Positif	Contoh Ulasan Negatif
1. Ulasan Asli	Aplikasi sangat MUDAH digunakan, beli token listrik cepat sekali!! Mantappp	petugas suram tidak ada tindakan, sdh dilapor berkali-kali ga ada respon
2. *Case Folding*	aplikasi sangat mudah digunakan, beli token listrik cepat sekali!! Mantappp	petugas suram tidak ada tindakan, sdh dilapor berkali-kali ga ada respon
3. *Text Cleaning*	aplikasi sangat mudah digunakan beli token listrik cepat sekali mantap	petugas suram tidak ada tindakan sudah dilapor berkalikali tidak ada respon
4. *Tokenizing*	['aplikasi', 'sangat', 'mudah', 'digunakan', 'beli', 'token', 'listrik', 'cepat', 'sekali', 'mantap']	['petugas', 'suram', 'tidak', 'ada', 'tindakan', 'sudah', 'dilapor', 'berkalikali', 'tidak', 'ada', 'respon']
5. *Stopword Removal*	aplikasi sangat mudah digunakan beli token listrik cepat sekali mantap	petugas suram tindakan dilapor berkalikali respon
6. *Stemming*	aplikasi sangat mudah guna beli token listrik cepat sekali mantap	tugas suram tindak lapor berka likali respon

4. Pembobotan TF-IDF dan Ekstraksi Fitur

Bobot kata ditentukan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), yang menetapkan bobot kata-kata berdasarkan seberapa sering muncul dalam dokumen dan tingkat keunikan di seluruh rangkaian dokumen [15]. TF-IDF dirumuskan sebagai:

$$W_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (1)$$

dengan, $W_{t,d}$ dalah bobot kata t pada dokumen d, $tf_{t,d}$ frekuensi kata t pada dokumen d, N jumlah total dokumen, dan df_t jumlah dokumen yang memuat kata t.

Selanjutnya, ekstraksi fitur dilakukan dengan menggunakan *Count Vectoriser*, yang merepresentasikan teks sebagai vektor numerik berdasarkan frekuensi kata, sehingga menghasilkan 4.160 fitur (kata unik).

5. Pembagian Data dan Penyeimbangan Kelas (SMOTE)

Himpunan data dibagi 80% menjadi data pelatihan dan 20% menjadi data pengujian menggunakan metode partisi bertingkat *random_state* = 42 untuk memungkinkan *reproduktifitas* hasil. Rasio 80:20 dipilih karena menyeimbangkan kecukupan data latih untuk membentuk model yang andal dengan porsi data uji yang memadai untuk estimasi kinerja yang valid; secara teoretis rasio ini optimal karena memaksimalkan keandalan estimasi galat [16], dan telah lazim diterapkan pada penelitian sentimen ulasan sejenis [1], [4]. Karena ketidakseimbangan kelas masih tetap signifikan, teknik *synthetic minority oversampling* (SMOTE) diterapkan pada data pelatihan. SMOTE mengatasi ketidakseimbangan ini dengan menghasilkan sampel sintetis dari kelas minoritas, menginterpolasi berdasarkan *k-nearest neighbor* dari setiap titik data di kelas itu, daripada hanya membuat salinan data yang ada [17]. Teknik ini terbukti meningkatkan kinerja klasifikasi pada data ulasan yang tidak seimbang [11], [12]. Penyeimbangan hanya diterapkan pada data latih (menjadi 1.268 positif dan 1.268 negatif), sedangkan data uji dibiarkan dalam kondisi asli agar evaluasi mencerminkan kondisi nyata.

6. Algoritma Klasifikasi

Ketiga algoritma yang dibandingkan merupakan metode pembelajaran mesin yang paling umum digunakan dan terbukti efektif untuk analisis sentimen ulasan aplikasi berbahasa Indonesia [7], [8], [9]. Ketiganya juga mewakili tiga paradigma pembelajaran yang berbeda, sebagaimana diuraikan berikut:

6.1 *Naïve Bayes*

Algoritma *probabilistik* yang didasarkan pada *teorema Bayes*, dengan asumsi kemandirian antar fitur. Algoritma ini efisien dan sesuai untuk data teks berdimensi tinggi dan jarang (*sparse*) seperti representasi TF-IDF, serta terbukti menjadi *baseline* yang kuat pada klasifikasi sentimen ulasan berbahasa Indonesia [3], [13]. Model *Naïve Bayes* dinyatakan sebagai:

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

dengan $P(C|X)$ adalah probabilitas posterior kelas C jika diberikan fitur X.

6.2 *Support Vector Machine (SVM)*

Algoritma yang memungkinkan penentuan *hiperplane* optimal yang memberikan *margin* terbesar untuk memisahkan dua kelas. SVM sangat sesuai untuk analisis sentimen karena efektif pada ruang fitur berdimensi tinggi seperti teks dan memiliki kemampuan generalisasi yang baik melalui *margin* maksimum sehingga kerap berkinerja terbaik [2], [6]. Fungsi keputusan:

$$f(x) = \sum_{i=1}^{n} \alpha_i \, y_i K(x_i, x) + b \quad (3)$$

dengan α_i bobot, y_i label kelas, $K(x_i, x)$ fungsi kernel, dan b bias.

DOI : https://doi.org/10.26623/transformatika.v24i1.14934

6.3 *Random Forest*

Metode kombinasi yang membangun sejumlah pohon keputusan dan menentukan kategori akhir berdasarkan suara terbanyak (*majority voting*) yang, sebagai metode *ensemble*, sesuai untuk data ulasan karena tahan terhadap *overfitting* dan mampu menangkap interaksi fitur non-linier sehingga memberikan prediksi yang stabil [1], [9]:

$$f(x) = \text{mode} f_1(x), f_2(x), \dots, f_n(x) \quad (4)$$

7. Evaluasi Model

Stabilitas model diverifikasi menggunakan validasi silang 5-lipat, sedangkan optimisasi parameter SVM dilakukan menggunakan *GridSearchCV* pada ruang pencarian C = {0,1; 1; 10; 100}, gamma = {1; 0,1; 0,01; 0,001}, dan *kernel* = {RBF, linier}. Kinerja model tersebut dievaluasi menggunakan *Confusion Matrix* dengan empat indikator utama akurasi, presisi, *recall*, dan skor F1 untuk mendapatkan penilaian yang komprehensif yang tidak hanya bergantung pada akurasi [16], [18] :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

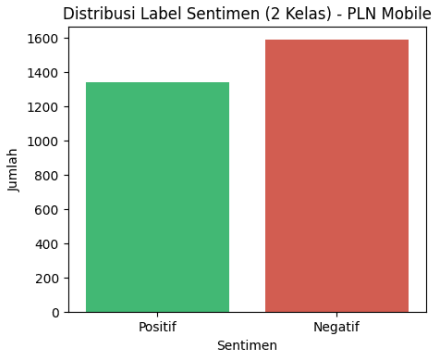
$$F1-score = 2 \times \frac{Presisi \times Recall}{Presisi+Recall} \quad (8)$$

Dengan TP (*True* Positif), TN (*True* Negatif), FP (*False* Positif), dan FN (*False* Negatif) merupakan komponen dari *Confusion Matrix*.

HASIL

1. Hasil Pengumpulan dan Distribusi Data

Pengumpulan data melalui *web scraping* berimbang per kategori *rating* menghasilkan 3.702 ulasan unik. Setelah melalui tahap *text preprocessing* dan penghapusan ulasan netral serta teks kosong, diperoleh 2.917 data yang terdiri atas 1.586 ulasan negatif (54,4%) dan 1.331 ulasan positif (45,6%). Komposisi ini menunjukkan distribusi kelas yang relatif seimbang, sehingga proses klasifikasi tidak didominasi oleh satu kelas tertentu.



Gambar 2. Grafik distribusi label sentimen

2. Hasil Pembobotan TF-IDF

Pembobotan TF-IDF mengidentifikasi kata-kata terpenting dalam korpus ulasan. 10 kata dengan skor rata-rata TF-IDF tertinggi ditunjukkan pada Tabel 2. Kata "aplikasi" (0,054), "PLN" (0,048), dan "sangat" (0,045) menempati peringkat atas, menunjukkan bahwa tema utama ulasan terkait dengan pengalaman menggunakan aplikasi dan layanan PLN.

Tabel 2. Hasil pembobotan TF-IDF (10 kata teratas)

	Kata	TF-IDF
0	aplikasi	0.054128
1	pln	0.047748
2	sangat	0.044993
3	mudah	0.033906
4	mobile	0.031258
5	bantu	0.031244
6	listrik	0.030265
7	layan	0.029178

Proses ekstraksi fitur menggunakan *Count Vectorizer* menghasilkan matriks fitur 2.917 × 4.160, artinya 4.160 kata unik (kosakata) digunakan sebagai fitur klasifikasi.

3. Hasil Validasi 5-*Fold Cross Validation*

Untuk menguji kestabilan model, dilakukan validasi silang 5 lipatan (5-*fold cross validation*) terhadap ketiga algoritma. Hasilnya ditunjukkan pada Tabel 3. *Naïve Bayes* memperoleh rata-rata *F1-score* tertinggi sebesar 0,846 dengan standar deviasi 0,011, diikuti *Random Forest* (0,834) dan SVM (0,823). Nilai standar deviasi yang kecil (≤0,011) pada seluruh model mengindikasikan bahwa performa bersifat stabil dan konsisten di semua lipatan data.

Tabel 3. Hasil 5-*fold cross validation* (*F1-Score*)

Model	Rata-rata F1	Standar Deviasi	F1 per-*fold*
Naïve Bayes	0,846	0,011	0,853; 0,853; 0,834; 0,834; 0,859
Random Forest	0,834	0,010	0,835; 0,849; 0,823; 0,822; 0,839
SVM	0,823	0,008	0,831; 0,821; 0,808; 0,829; 0,825

4. Hasil *Tuning Hyperparameter* SVM

Optimasi parameter SVM menggunakan *GridSearchCV* terhadap 32 kombinasi parameter (dengan 5 lipatan, total 160 pelatihan) menghasilkan konfigurasi terbaik: C = 1, gamma = 0,1, dan *kernel* = RBF, dengan skor *cross validation F1-score* sebesar 0,846. Konfigurasi inilah yang digunakan pada pengujian akhir terhadap data uji.

5. Hasil Klasifikasi dan *Confusion Matrix*

Pengujian dilakukan pada 584 data uji (318 negatif dan 266 positif). Hasil klasifikasi masing-masing model dievaluasi menggunakan *Confusion Matrix* dan laporan klasifikasi.

DOI : https://doi.org/10.26623/transformatika.v24i1.14934

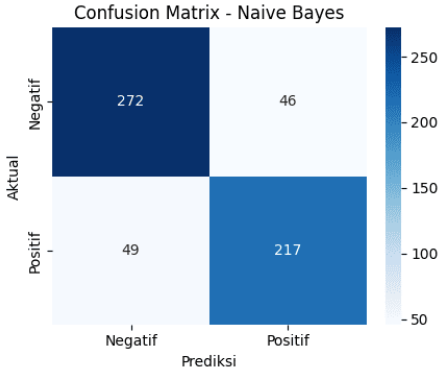
5.1 *Naïve Bayes*

Model ini mencapai akurasi 84%, dengan kategori negatif mencapai akurasi 85% dan *recall* 86%, dan kategori *anoda* mencatat akurasi 83% dan *recall* 82%. Model ini menunjukkan keseimbangan optimal antara kedua kelas.

```
======================================================
HASIL KLASIFIKASI - Naive Bayes
======================================================
              precision    recall  f1-score   support

     Negatif       0.85      0.86      0.85       318
     Positif       0.83      0.82      0.82       266

    accuracy                           0.84       584
   macro avg       0.84      0.84      0.84       584
weighted avg       0.84      0.84      0.84       584
```



Gambar 3. *Confusion matrix naïve bayes*

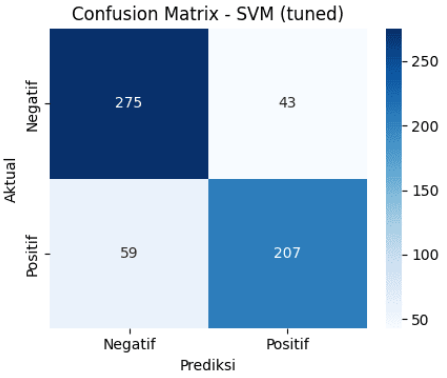
5.2 *Support Vector Machine (SVM)*

Model ini mencapai akurasi 83%, dengan nilai *recall* yang tinggi di kelas negatif (86%) tetapi nilai *recall* yang lebih rendah di kelas positif (78%).

```
======================================================
HASIL KLASIFIKASI - SVM (tuned)
======================================================
              precision    recall  f1-score   support

     Negatif       0.82      0.86      0.84       318
     Positif       0.83      0.78      0.80       266

    accuracy                           0.83       584
   macro avg       0.83      0.82      0.82       584
weighted avg       0.83      0.83      0.82       584
```



Gambar 4. *Confusion matrix* SVM

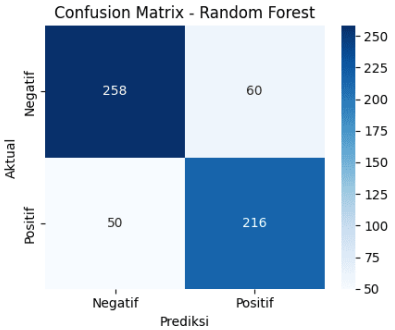
5.3 *Random Forest*

Memperoleh akurasi 81%, dengan *precision* 84% pada kelas negatif dan *recall* 81% pada kedua kelas.

```
======================================================
HASIL KLASIFIKASI - Random Forest
======================================================
              precision    recall  f1-score   support

     Negatif       0.84      0.81      0.82       318
     Positif       0.78      0.81      0.80       266

    accuracy                           0.81       584
   macro avg       0.81      0.81      0.81       584
weighted avg       0.81      0.81      0.81       584
```



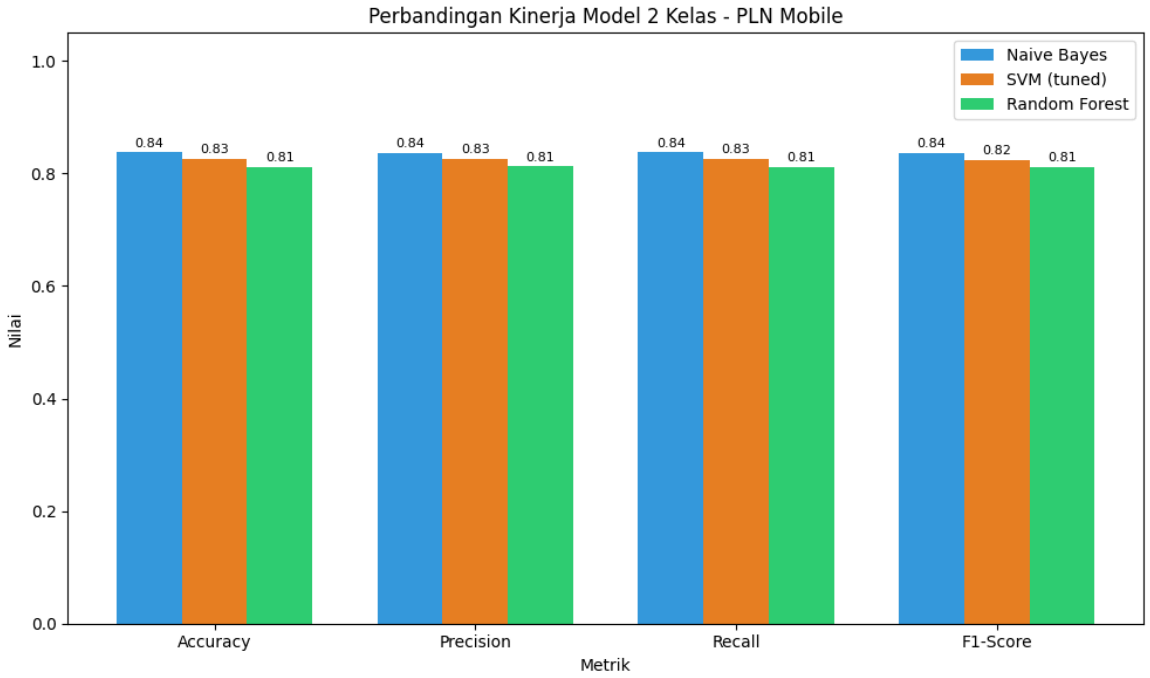
Gambar 5. *Confusion matrix random forest*

6. Perbandingan Kinerja Model

Perbandingan empat metrik utama dari ketiga model disajikan pada Tabel 4 dan ditampilkan pada Gambar 6. *Naïve Bayes* adalah model berkinerja terbaik pada semua metrik (akurasi, presisi, *recall*, dan skor F1 dengan 84%), diikuti oleh SVM (83%) dan *Random Forest* (81%).

Tabel 4. Perbandingan kinerja model pada data uji

	Model	Accuracy	Precision
0	*Naive Bayes*	84%	84%
1	SVM (tuned)	83%	83%
2	*Random Forest*	81%	81%



Gambar 6. Grafik batang perbandingan kinerja model

7. Analisis Kata Dominan dan *Word Cloud*

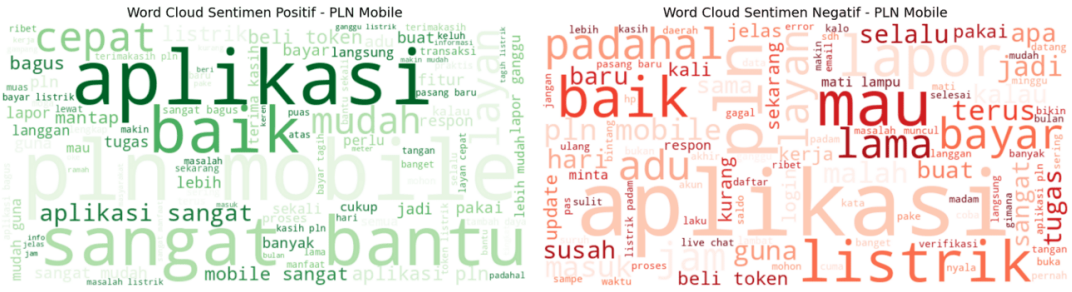
Visualisasi *Word Cloud* dan analisis frekuensi kata memberikan wawasan terhadap aspek yang diapresiasi dan dikeluhkan pengguna. Pada sentimen positif, kata dominan meliputi "mudah" (299), "bantu" (257), "cepat" (206), "layan" (185), "baik" (174), "bagus" (132), "bayar" (123), dan "*token*" (109). Hal ini menunjukkan pengguna mengapresiasi kemudahan, kecepatan, dan manfaat aplikasi dalam pembayaran tagihan serta pembelian *token* listrik.

Pada sentimen negatif, kata dominan meliputi "listrik" (355), "bayar" (239), "adu/pengaduan" (206), "jam" (192), "mati" (189), "lapor" (176), "lama" (163), dan "terus" (159). Pola ini mengindikasikan keluhan utama berkaitan dengan pemadaman listrik yang lama dan berulang serta kendala teknis pada proses pembayaran dan pengaduan.

Tabel 5. Pengelompokan kata dominan berdasarkan aspek layanan

Aspek Layanan	Kata Dominan Positif	Kata Dominan Negatif	Implikasi
Kemudahan & kegunaan (*usability*)	mudah (299), bantu (257), baik (174), bagus (132)	–	Antarmuka dinilai mudah & membantu; pertahankan
Kecepatan & *responsivitas*	cepat (206), layan (185)	lama (163), terus (159), jam (192)	Sebagian puas, namun banyak keluhan proses lambat/berulang
Transaksi & pembayaran	bayar (123), *token* (109)	bayar (239)	Beli *token*/bayar tagihan mudah, tetapi ada kendala transaksi
Keandalan pasokan listrik	–	listrik (355), mati (189), jam (192)	Keluhan terbesar: pemadaman listrik lama
Pengaduan & penanganan keluhan	–	adu/pengaduan (206), lapor (176)	Respons pengaduan dinilai lambat

Pengelompokan pada Tabel 5 memperlihatkan bahwa apresiasi pengguna terkonsentrasi pada aspek kemudahan penggunaan dan kecepatan layanan aplikasi, sedangkan keluhan didominasi aspek di luar aplikasi itu sendiri, yaitu keandalan pasokan listrik (pemadaman lama dan berulang) serta penanganan pengaduan. Kata "bayar" muncul pada kedua sentimen, menandakan fitur pembayaran dihargai sebagian pengguna namun juga menjadi sumber kendala bagi sebagian lainnya. Temuan ini memberi arahan prioritas pengembangan bagi PT PLN (Persero): (1) meningkatkan keandalan dan transparansi informasi pemadaman, (2) memperbaiki stabilitas proses pembayaran dan pembelian *token*, serta (3) mempercepat tindak lanjut pengaduan melalui aplikasi.



Gambar 7. Sentimen positif dan negatif *Word Cloud*

PEMBAHASAN

Temuan penelitian ini memberikan perspektif yang menarik bila dibandingkan dengan penelitian terdahulu. Pada domain aplikasi investasi seperti Ajaib [7], SVM dan *Random Forest* cenderung unggul. Namun, pada kasus ulasan aplikasi layanan publik *PLN Mobile* dengan skema klasifikasi dua kelas ini, *Naïve Bayes* justru memberikan kinerja terbaik dan paling stabil. Hal ini sejalan dengan temuan pada penelitian *PLN Mobile* sebelumnya yang menunjukkan *Naïve Bayes* sebagai *baseline* yang kuat [3], [4], dan diduga disebabkan oleh karakteristik teks ulasan *PLN*

Mobile yang relatif lugas serta didominasi kata kunci spesifik (seperti "mati", "lama", "mudah", "bantu"), sehingga pendekatan *probabilistik Naïve Bayes* mampu memodelkan distribusi kata secara efektif. Penerapan strategi pengumpulan data berimbang dan teknik SMOTE terbukti penting dalam penelitian ini. Tanpa penyeimbangan, distribusi ulasan *PLN Mobile* sangat timpang ke arah positif, sehingga model cenderung mengabaikan kelas minoritas dan menghasilkan akurasi tinggi yang menyesatkan [10]. Dengan data yang seimbang, validasi silang, dan *tuning* parameter, hasil evaluasi pada kisaran 81–84% jauh lebih representatif dan kredibel dalam menggambarkan kinerja model yang sebenarnya. Hasil analisis kata dominan juga memberikan kontribusi praktis bagi pengembang *PLN Mobile*, yaitu perlunya peningkatan keandalan informasi pemadaman serta perbaikan proses pembayaran dan pengaduan.

KESIMPULAN DAN REKOMENDASI

Penelitian ini mampu membandingkan kinerja antara algoritma *Naïve Bayes*, *Support Vector Machine* (SVM) dan *Random Forest* dalam analisis sentimen pada ulasan aplikasi PLN Mobile menggunakan skema klasifikasi dua kelas (positif dan negatif). Berdasarkan hasil dan pembahasan, *Naïve Bayes* terbukti menjadi model yang paling unggul dan stabil dengan akurasi, presisi, *recall*, dan skor F1 84% pada data tes, serta skor F1 rata-rata 0,846 dalam validasi silang 5 kali lipat. Model ini mengungguli SVM (83%) dan *Random Forest* (81%). Keunggulan ini menunjukkan bahwa pendekatan *probabilistik Naïve Bayes* sangat cocok dengan karakteristik teks *proofreading* Indonesia yang relatif sederhana dan didominasi oleh kata kunci tertentu. Penerapan strategi pengumpulan data yang seimbang, SMOTE, validasi silang, dan optimasi *hiperparameter* terbukti efektif dalam menghasilkan penilaian yang lebih objektif dan kredibel, menjadikannya kontribusi metodologis utama dari penelitian ini. Secara praktis, analisis kata-kata dominan mengungkapkan bahwa pengguna menghargai kemudahan dan kecepatan layanan, sedangkan keluhan utama adalah tentang pemadaman panjang dan hambatan dalam proses pembayaran dan pengaduan; Temuan ini dapat menjadi dasar bagi PT PLN (Persero) untuk meningkatkan keandalan informasi pemadaman dan meningkatkan fungsionalitas transaksi. Untuk meningkatkan akurasi dan pemahaman konteks dari tinjauan, penelitian di masa depan disarankan menggunakan kumpulan data lintas platform yang lebih besar, mengeksplorasi klasifikasi multi-kelas termasuk sentimen netral, serta menerapkan pendekatan pembelajaran mendalam dan model berbasis transformator seperti *IndoBERT*.

REFERENSI

[1] M. D. Rizkiyanto, M. D. Purbolaksono, and W. Astuti, “Sentiment Analysis Classification on PLN Mobile Application Reviews using *Random Forest* Method and TF-IDF Feature Extraction,” *INTEK: Jurnal Penelitian*, vol. 11, no. 1, pp. 37–43, Apr. 2024, doi: 10.31963/intek.v11i1.4774.

[2] 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),” 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[3] S. Syafrizal, M. Afdal, and R. Novita, “Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma *Naïve Bayes* Classifier dan *K-Nearest Neighbor*,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.

[4] B. A. Prabowo, A. Hindasyah, and A. Khalid Rivai, "Sentiment Analysis of PLN Mobile Application Services Using Naive Bayes, *Support Vector Machine* (SVM) and *Decision Tree* Methods," *Jurnal Riset Informatika*, vol. 7, no. 3, pp. 236–243, Jun. 2025, doi: 10.34288/jri.v7i3.378.

[5] J. M. Ayomi, A. V. Vitianingsih, Y. Kristyawan, A. L. Maukar, and T. Widiartin, "Sentiment Analysis of User Reviews for the PLN Mobile Application Using *Naïve Bayes* and *Long Short-Term Memory*," *Journal of Information Systems and Informatics*, vol. 7, no. 4, pp. 3849–3873, Dec. 2025, doi: 10.63158/journalisi.v7i4.1342.

[6] M. Arya Java, M. Syafrullah, and F. Teknologi, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan *Support Vector Machine*," *Jurnal TICOM: Technology of Information and Communication*, vol. 12, no. 2, 2024, [Online]. Available: https://github.com/nasalsabila/kamus-alay

[7] N. D. Kurniawan, P. R. Ferdian, and N. Hidayati, "Analisis Sentimen Algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Random Forest* Pada Ulasan Aplikasi Ajaib," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 11, no. 1, pp. 87–97, May 2025, doi: 10.25077/teknosi.v11i1.2025.87-97.

[8] S. A. S. Mola, D. L. B. Baun, I. O. Nunes, and M. M. A. R. Sani, "Analisis Sentimen Aplikasi Halo BCA di Google Play Store Menggunakan Metode Naive Bayes, *Support Vector Machine* dan *Random Forest*," *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, vol. 15, no. 2, pp. 69–79, Dec. 2024, doi: 10.52972/hoaq.vol15no2.p69-79.

[9] M. Winda, P. 1 Ai, I. Warnilah, B. K. Simpony, V. Christian, and M. Dhafa Alfareza, "Implementasi Algoritma Model *Random Forest*, SVM Dan Naive Bayes Untuk Sentimen Analisis Aplikasi Gojek Di Playstore," *Jurnal Sains dan Manajemen*, vol. 13, no. 2, 2025.

[10] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Suryady (Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class)," 2024.

[11] I. G. B. A. Budaya and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE *Oversampling* and TF-IDF Implementation on Google Reviews for Public Health Centers," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1077–1086, Jul. 2024, doi: 10.57152/malcom.v4i3.1459.

[12] D. Andriyani, A. Faqih, and S. E. Permana, "Journal of Artificial Intelligence and Engineering Applications The Effect of SMOTE Application on *Support Vector Machine* Performance in Sentiment Classification on Imbalanced Datasets," 2025. [Online]. Available: https://ioinformatic.org/

[13] E. A. Lisangan *et al.*, "Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia," 2022.

[14] T. Kemendikbud *et al.*, "J-INTECH (Journal of Information and Technology) Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna".

[15] T. Kemendikbud *et al.*, "J-INTECH (Journal of Information and Technology) Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna".

[16] A. Gholamy, V. Kreinovich, O. Kosheleva, A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation," *Departmental Technical Reports (CS)*, Feb. 2018, Accessed: Jul. 20, 2026. [Online]. Available: https://scholarworks.utep.edu/cs_techrep/1209

[17] R. Marta Dinata, E. Rayhana, V. Hadi, and U. Alkaf, "Analisis Komprehensif Kinerja Model Klasifikasi Sentimen: Evaluasi Lintas Metrik pada Dataset Tweet Film Bahasa Indonesia," *Jurnal Rekayasa Informasi*, vol. 14, no. 1, p. 2025.

[18] N. Hidayah and D. Dodiman, "Implementasi Algoritma Multinomial *Naïve Bayes*, TF-IDF dan *Confusion Matrix* dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," *Jurnal Akademik Pendidikan Matematika*, pp. 8–15, May 2024, doi: 10.55340/japm.v10i1.1491.