Jurnal Teknologi Sistem Informasi dan Aplikasi **ISSN: 2654-3788**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2654-4229**
Vol. 9, No. 1, January 2026 (23-33) **DOI: https://doi.org/10.32493/jtsi.v9i1.54847**

Perbandingan Kinerja Antara Algoritma Naive Bayes dan Long Short-Term Memory dalam Menganalisis Sentimen Ulasan Aplikasi PLN Mobile

Ario Kusumo[1], Rinna Rachmatika[2*]

[1,2*] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten, Indonesia, 15310
e-mail: [1]arioorder2002@gmail.com. [2*]rinnarachmatika@unpam.ac.id.

Submitted Date: November 19th, 2025 Reviewed Date: December 15th, 2025
Revised Date: December 30th, 2025 Accepted Date: January 20th, 2026

Abstract

The rapid advancement of digital technology has increased the use of public service applications such as PLN Mobile. However, numerous user reviews highlight technical issues that may reduce customer satisfaction. This study aims to compare the performance of Naïve Bayes and Long Short-Term Memory (LSTM) algorithms in classifying user review sentiments on the PLN Mobile application. The dataset was collected from Google Play Store using Google Play Scraper. The research involved text preprocessing (case folding, tokenizing, stopword removal, stemming, TF-IDF), oversampling with SMOTE, and evaluation using accuracy, precision, recall, and F1-score metrics. The results show that Naïve Bayes achieved 96% accuracy, while LSTM reached 98%. LSTM also outperformed Naïve Bayes in all evaluation metrics, indicating superior capability in capturing contextual and sequential patterns in textual data. Therefore, LSTM is more recommended for sentiment analysis of complex text data.

Keywords: PLN Mobile; Sentiment Analysis; Naïve Bayes; LSTM; Machine Learning

Abstrak

Perkembangan teknologi digital mendorong meningkatnya penggunaan aplikasi layanan publik seperti PLN Mobile. Namun, banyak ulasan pengguna menunjukkan kendala teknis yang dapat menurunkan kepuasan pelanggan. Penelitian ini bertujuan membandingkan kinerja algoritma Naïve Bayes dan Long Short-Term Memory (LSTM) dalam mengklasifikasikan sentimen ulasan pengguna aplikasi PLN Mobile. Data diperoleh dari Google Play Store menggunakan Google Play Scraper. Tahapan penelitian meliputi text preprocessing (case folding, tokenizing, stopword removal, stemming, TF-IDF), oversampling dengan SMOTE, serta evaluasi menggunakan accuracy, precision, recall, dan F1-score. Hasil menunjukkan bahwa Naïve Bayes mencapai akurasi 96%, sedangkan LSTM mencapai 98%. LSTM juga unggul pada seluruh metrik evaluasi, membuktikan kemampuannya menangkap konteks dan pola sekuensial teks lebih baik dibandingkan Naïve Bayes. Dengan demikian, LSTM lebih direkomendasikan untuk analisis sentimen teks kompleks.

Kata Kunci: PLN Mobile; Analisis Sentimen; Naïve Bayes; LSTM; Machine Learning

1. Pendahuluan

Transformasi digital telah menjadi faktor utama dalam memodernisasi layanan publik di Indonesia. Salah satu bentuk implementasinya adalah pengembangan aplikasi mobile yang memungkinkan masyarakat mengakses layanan secara mandiri dan efisien dalam berbagai sektor, termasuk pendidikan, kesehatan, dan layanan publik (Putra & Ratnawati, 2025). PLN Mobile, yang diluncurkan oleh PT. PLN (Persero) pada tahun 2016, merupakan inovasi digital yang menyediakan berbagai fitur seperti pembayaran tagihan listrik, pelaporan gangguan, pencatatan meter mandiri, dan informasi penggunaan energi. Aplikasi

ini dirancang untuk meningkatkan kualitas pelayanan dan memperkuat interaksi antara penyedia layanan dan pelanggan.

Namun, seiring meningkatnya jumlah pengguna, berbagai ulasan di *Google Play Store* menunjukkan bahwa aplikasi ini masih menghadapi sejumlah kendala teknis. Beberapa keluhan yang sering muncul meliputi kesalahan informasi, notifikasi yang tidak akurat, serta hambatan dalam proses login dan transaksi. Ulasan tersebut mencerminkan persepsi dan tingkat kepuasan pengguna yang penting untuk dianalisis secara sistematis. Dalam konteks ini, analisis sentimen menjadi pendekatan yang relevan untuk mengevaluasi opini pengguna secara otomatis melalui teks ulasan, sehingga dapat memberikan gambaran terhadap kualitas layanan dan area yang perlu diperbaiki (Informatika, 2025)

Dalam perkembangan penelitian analisis sentimen terhadap aplikasi layanan publik, berbagai algoritma telah diuji untuk memperoleh performa terbaik. Menurut Luthfi, analisis terhadap 10.000 data ulasan pengguna aplikasi Mobile JKN menunjukkan bahwa *Naïve Bayes* lebih unggul dengan akurasi sebesar 91,3% (Naufal et al., 2025). Selanjutnya, penelitian yang dilakukan oleh Yuliana dan Fahmi memperlihatkan bahwa metode *Long Short-Term Memory* (LSTM) mampu menghasilkan performa yang lebih baik dengan capaian 77% akurasi, 84% presisi, 75% recall, dan 80% F1-*Score*, sedangkan *Naïve Bayes* hanya memperoleh 76% akurasi serta masing-masing 75% untuk presisi, recall, dan F1-*Score* (Romadhoni & Holle, 2022). Temuan-temuan tersebut menegaskan bahwa pemilihan algoritma yang tepat sangat berpengaruh terhadap kualitas hasil analisis sentimen, sehingga mendorong perlunya eksplorasi lebih lanjut terhadap model-model pembelajaran mesin yang adaptif dan akurat dalam mengolah data ulasan pengguna aplikasi layanan publik.

Berbagai penelitian lainnya menunjukkan bahwa metode *Long Short-Term Memory* (LSTM) memiliki efektivitas tinggi dalam analisis sentimen di media sosial maupun ulasan produk, dengan penerapan yang beragam mulai dari opini publik terhadap Tokocrypto (Sandag & Waworundeng, 2022), fenomena *childfree* dengan kombinasi LSTM dan BERT yang mencapai akurasi 95,85% (Farida & Rochmawati, 2024), hingga *review skincare* dengan integrasi Word2Vec yang meningkatkan akurasi dari 74% menjadi 91% (Purnasiwi et al., 2023). Selain itu, perbandingan LSTM dengan *Naïve Bayes* pada isu penundaan Pemilu 2024 menunjukkan keunggulan LSTM dengan akurasi 92% dibandingkan 80% (Cholis et al., 2023), sementara analisis sentimen terhadap kenaikan harga BBM juga menegaskan performa LSTM dengan akurasi hingga 90% (Ikhsan, 2023). Secara keseluruhan, hasil-hasil tersebut menegaskan bahwa LSTM, baik digunakan sendiri maupun dikombinasikan dengan metode lain seperti BERT atau *Word2Vec*, mampu memberikan klasifikasi sentimen yang lebih akurat dan relevan untuk memahami opini publik terhadap isu teknologi, sosial, politik, maupun produk komersial.

Untuk mendukung analisis tersebut, penelitian ini membandingkan dua algoritma populer dalam klasifikasi teks, yaitu *Naïve Bayes* dan *Long Short-Term Memory* (LSTM). *Naïve Bayes* merupakan metode probabilistik yang sederhana dan efisien, sedangkan LSTM adalah varian dari *Recurrent Neural Network* (RNN) yang mampu menangkap pola dan dependensi jangka panjang dalam data teks. Dengan membandingkan performa kedua algoritma dalam mengklasifikasikan sentimen ulasan pengguna terhadap PLN Mobile, penelitian ini bertujuan untuk mengidentifikasi model yang paling akurat dan efektif dalam memahami persepsi publik terhadap layanan digital pemerintah.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen terhadap ulasan pengguna aplikasi PLN Mobile yang diperoleh melalui teknik web scraping. Data dikumpulkan menggunakan *Google Play Scraper* dari halaman aplikasi PLN Mobile di *Google Play Store*. Jumlah data yang berhasil dihimpun sebanyak 18.724 ulasan terbaru dalam rentang waktu 9 Januari 2025 hingga 14 April 2025. Pemanfaatan scraping sebagai teknik pengumpulan data digital telah terbukti efektif dalam penelitian berbasis ulasan aplikasi (Roiqoh & Zaman, 2023).

Tahap pra-pengolahan data dilakukan untuk memastikan kualitas teks sebelum dianalisis. Proses ini meliputi sebagai berikut:

a. *Cleaning*: Menghapus karakter non-alfabet, simbol, dan tautan.
b. *Case Folding*: Mengubah semua huruf menjadi huruf kecil
c. *Tokenizing:* Memecah kalimat menjadi kata-kata.

d. *Stopword Removal*: Menghapus kata-kata umum yang tidak memiliki makna signifikan (misalnya "yang", "dan", "di")
e. *Stemming*: Mengubah kata ke bentuk dasar menggunakan algoritma Nazief-Adrian
f. *Representasi* TF-IDF: Mengubah teks menjadi vektor numerik berdasarkan frekuensi kata dan kekhususan dalam dokumen.

Tahapan ini penting agar model dapat mengenali pola linguistik secara konsisten. Pra-pengolahan teks merupakan faktor krusial dalam meningkatkan akurasi model klasifikasi sentimen.

Data yang telah diproses kemudian dibagi menjadi data latih dan data uji menggunakan teknik 10-*fold cross validation* dengan rasio 90%:10% untuk memastikan generalisasi model dan menghindari overfitting (Akbar et al., 2024). Setelah melalui proses pra-processing terhadap setiap tweet, data tersebut dianalisis dengan menggunakan variabel prediktor berupa kata dasar dari setiap tweet dan variabel respon berupa klasifikasi sentimen tweet, yaitu positif dan negatif, yang dapat dilihat pada Tabel 1 berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala Data
y	Sentimen 0 = sentimen negatif 1 = sentimen positif	Nominal
xk	Bobot kata dasar dari sentimen tweet yang diperoleh dari hasil TF-IDF	Rasio

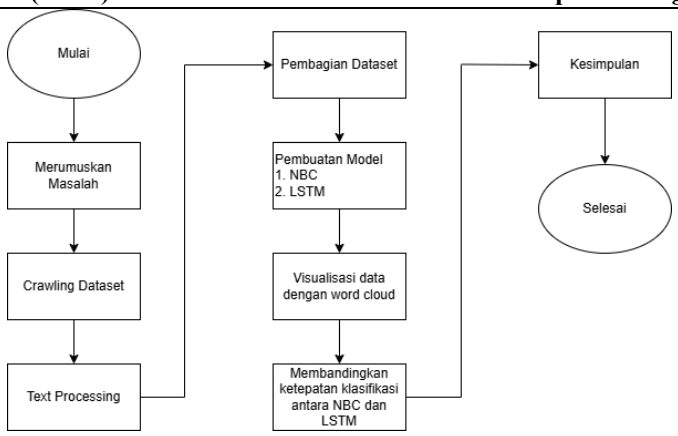
Dua algoritma digunakan dalam klasifikasi sentimen, yaitu *Naïve Bayes Classifier* (NBC) dan *Long Short-Term Memory* (LSTM). NBC dipilih karena kesederhanaannya dalam menghitung probabilitas kata terhadap kategori sentimen, sedangkan LSTM digunakan untuk menangkap konteks urutan kata dalam ulasan. Kombinasi kedua pendekatan ini memungkinkan perbandingan kinerja antara model berbasis probabilistik dan model berbasis jaringan saraf tiruan berulang (RNN) (Hadiprakoso et al., n.d.).

Evaluasi model dilakukan dengan menggunakan metrik sebagai berikut:
a. *Accuracy*: Proporsi prediksi yang benar.
b. *Precision*: Ketepatan prediksi positif.
c. *Recall*: Kemampuan model menangkap semua data positif.
d. F1-*Score*: Harmoni antara *precision* dan *recall*.

Selain itu, visualisasi data berupa *word cloud* dibuat untuk menampilkan kata-kata yang paling sering muncul dalam ulasan pengguna. Visualisasi ini membantu memberikan gambaran umum mengenai persepsi masyarakat terhadap aplikasi PLN Mobile. Menurut (Limbong et al., 2022), word cloud dapat digunakan sebagai alat eksplorasi awal untuk memahami distribusi kata dalam analisis sentimen.

Hasil akhir penelitian berupa perbandingan performa NBC dan LSTM dalam mengklasifikasikan sentimen ulasan pengguna. Interpretasi dilakukan untuk menentukan model yang lebih unggul serta memberikan rekomendasi pengembangan aplikasi PLN Mobile berdasarkan temuan analisis. Dengan metodologi ini, penelitian diharapkan mampu memberikan kontribusi terhadap pengembangan sistem analisis sentimen berbasis ulasan aplikasi digital di Indonesia. Diagram alir penelitian disajikan pada gambar 1. sebagai berikut.

ISSN: 2654-3788
e-ISSN: 2654-4229
DOI: https://doi.org/10.32493/jtsi.v9i1.54847



Gambar 1. Diagram Alir Analisis Data

3. Hasil dan Pembahasan

a. Frekuensi Dataset

Dataset penelitian diperoleh melalui *web scraping* dengan *Google Play Scraper*, berupa ulasan pengguna aplikasi PLN Mobile di *Google Play Store*, lengkap dengan atribut pendukung seperti tanggal, rating, dan nama pengguna.

b. Frekuensi Kemunculan Data

Analisis frekuensi kata digunakan untuk menemukan kata yang paling sering muncul dalam ulasan PLN Mobile, sehingga dapat diidentifikasi kata kunci dominan yang mewakili sentimen atau topik utama. Kata yang sering muncul ditunjukkan pada tabel 2 dibawah ini.

Tabel 2. Frekuensi Kemunculan Kata

Kata	Frekuensi
Pln	8657
Sangat	6607
Mobile	5384
Aplikasi	5377
Mudah	4974
Bantu	4687
Dan	4350
Listrik	3056
Cepat	2646
Baik	2618

c. Preprocessing Text

Tahap *praprosesing teks* merupakan komponen krusial dalam analisis sentimen terhadap ulasan pengguna aplikasi PLN Mobile yang bersifat tidak terstruktur dan mengandung elemen-elemen tidak relevan seperti tanda baca, angka, huruf kapital, dan *stopwords*. Untuk meningkatkan efektivitas klasifikasi oleh algoritma machine learning, dilakukan serangkaian tahapan praprosesing meliputi *case folding, tokenizing, stopword removal, stemming*, dan transformasi teks akhir. *Case folding* menyamakan format huruf menjadi kecil, tokenizing memecah kalimat menjadi kata-kata, *stopword removal* menghapus kata umum yang tidak bermakna analitis, dan stemming mengembalikan kata ke bentuk dasar untuk mengurangi keragaman makna. Hasil akhir berupa *final_text* digunakan sebagai input dalam pelatihan dan pengujian model *Naïve Bayes* dan LSTM (Waluyo, 2023).

ISSN: 2654-3788
e-ISSN: 2654-4229
DOI: https://doi.org/10.32493/jtsi.v9i1.54847

d. Cleaning

Cleaning adalah tahap awal praproses teks untuk menghapus elemen tidak relevan seperti emotikon, simbol, dan karakter khusus, sehingga menghasilkan data ulasan yang bersih dan siap dianalisis secara optimal. Hasil output cleaning ditunjukkan pada tabel 3 dibawah ini.

Tabel 3. Hasil Cleaning

Sebelum	Sesudah
Aplikasi Yang Sangat Bagus Dan Sangat Membantu Ðÿ‘□Ðÿ□»	Aplikasi Yang Sangat Bagus Dan Sangat Membantu
Kerenðÿ˜ž Ðÿ‘□	Keren
PLN Terbaikðÿ”¥	Pln Terbaik
Semoga PLN Sukses	Semoga Pln Sukses
Selaluðÿ‘□Ðÿ‘□Ðÿ‘□Ðÿ‘□	Selalu
Terimakasih PLN Ðÿ‘□	Terimakasih Pln

e. Case Folding

Case folding adalah tahap awal praproses teks yang mengubah huruf kapital menjadi huruf kecil untuk menyeragamkan penulisan dan mencegah kesalahan identifikasi kata akibat perbedaan huruf (Wijaya et al., 2026). Hasil output case folding ditunjukkan pada tabel 4 dibawah ini.

Tabel 4. Hasil Case Folding

Sebelum	Sesudah
Bagus Dan Bermanfaat	Bagus Dan Bermanfaat
PLN Mobile Sangat Memudahkan Sekali Dari Segi Apapun	Pln Mobile Sangat Memudahkan Sekali Dari Segi Apapun
Aplikasi Yang Sangat Bagus Dan Sangat Membantu	Aplikasi Yang Sangat Bagus Dan Sangat Membantu
Lapor Gangguan Jadi Lebih Mudah	Lapor Gangguan Jadi Lebih Mudah
Layanannya Cepat Dan Memuaskan	Layanannya Cepat Dan Memuaskan

f. Tokenzing

Tokenizing adalah tahap praproses teks yang memecah kalimat menjadi token (kata-kata terpisah), sehingga algoritma machine learning dapat menganalisis teks secara lebih akurat (Polgan et al., 2025). Hasil output tokenizing ditunjukkan pada tabel 5 dibawah ini.

Tabel 5. Hasil Tokenizing

Sebelum	Sesudah
Pln Mobile Aplikasi Terbaik	['Pln', 'Mobile', 'Aplikasi', 'Terbaik']
Bagus Dan Bermanfaat	['Bagus', 'Dan', 'Bermanfaat']
Lapor Gangguan Jadi Lebih Mudah...	['Lapor', 'Gangguan', 'Jadi', 'Lebih', 'Mudah...']
Mau Catat Meter Mandiri, Bayar Listrik Pakai Aplikasi	['Mau', 'Catat', 'Meter', 'Mandiri,', 'Bayar', ‘Listrik’, ‘Pakai’, ‘Aplikasi’]
Aplikasinya Bagus Banget Sangat Memudahkan	['Aplikasinya', 'Bagus', 'Banget', 'Sangat', 'Memudahkan']

ISSN: 2654-3788
e-ISSN: 2654-4229
DOI: https://doi.org/10.32493/jtsi.v9i1.54847

g. Stopword Removal

Stopword removal adalah tahap praproses teks untuk menghapus kata-kata umum seperti "dan", "yang", atau "di" yang tidak memiliki nilai informasi penting, sehingga analisis sentimen menjadi lebih fokus dan akurat (Chandra et al., 2025). Hasil output stopword removal ditunjukkan pada tabel 6 dibawah ini.

Tabel 6. Hasil Stopword Removal

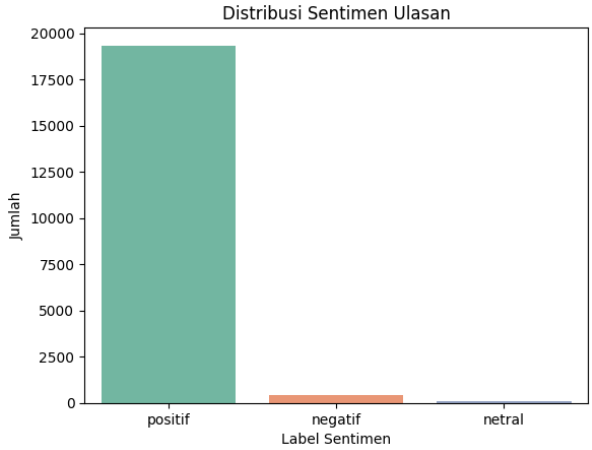
Sebelum	Sesudah
['bagus', 'dan', 'bermanfaat']	['bagus', 'bermanfaat']
['aplikasi', 'yang', 'sangat', 'bagus']	['aplikasi', 'sangat', 'bagus']
lapor gangguan jadi lebih mudah...	['lapor', 'gangguan', 'jadi', 'lebih', 'mudah...']
mau catat meter mandiri, bayar listrik pakai aplikasi	['mau', 'catat', 'meter', 'mandiri,', 'bayar', 'listrik', 'pakai', 'aplikasi']
aplikasinya bagus banget sangat memudahkan	['aplikasinya', 'bagus', 'banget', 'sangat', 'memudahkan']

h. Labelling

Labeling adalah tahap akhir praproses teks yang dilakukan dengan pendekatan *lexicon-based* menggunakan kamus sentimen bahasa Indonesia. Ulasan diklasifikasikan sebagai positif, negatif, atau netral berdasarkan dominasi kata bermuatan sentimen (Ardana et al., 2025). Metode ini dipilih karena praktis, efisien, dan cukup akurat untuk pelatihan model *supervised* learning seperti *Naïve Bayes* dan LSTM. Hasil labelling dan output distribusi sentimen ulasan pada tabel 7 dan gambar 2 dibawah ini.

Tabel 7. Hasil Labelling

Final text	Label Sentimen
aplikasi sangat bantu pln makin maju	positif
pln mobile mantap	positif
aplikasi tidak konsisten tidak semua adu di layan	negatif
tidak bantu	negatif
cukup	netral
kurang efektif apl	netral



Gambar 2. Distribusi Sentimen Ulasan

i. TF-IDF

TF-IDF adalah teknik representasi teks yang menilai pentingnya kata dalam dokumen relatif terhadap korpus. Dalam penelitian ini, TF-IDF diterapkan pada ulasan PLN Mobile setelah tahap praproses (*cleaning, stopword removal, stemming*) untuk menghasilkan bobot kata yang representatif. Nilai rata-rata TF-IDF tiap kategori sentimen digunakan untuk mengidentifikasi kata kunci dominan, sehingga memberikan wawasan tentang pola bahasa dan fokus pengguna (Syah et al., 2025). Hasil output TF-IDF dari positif, negatif, hingga netral ditunjukkan pada tabel 8,9, dan 10 dibawah ini.

Jurnal Teknologi Sistem Informasi dan Aplikasi **ISSN: 2654-3788**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **e-ISSN: 2654-4229**
Vol. 9, No. 1, January 2026 (23-33) **DOI: https://doi.org/10.32493/jtsi.v9i1.54847**

Tabel 8. Hasil TF-IDF Sentimen Positif

Kata	Rata-Rata TF-IDF
pln	0.097700
sangat	0.094681
bantu	0.075223
mobile	0.068527
mudah	0.063554
baik	0.062031
aplikasi	0.061470
mantap	0.059085
dan	0.049014
cepat	0.041835
layan	0.040317
bagus	0.039678
listrik	0.033641
makin	0.032143
untuk	0.020990

Tabel 9. Hasil TF-IDF Sentimen Negatif

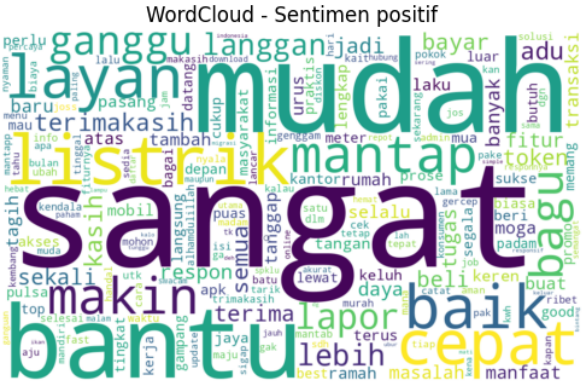
Kata	Rata-Rata TF-IDF
parah	0.173486
tidak	0.063867
di	0.042253
login	0.034448
ada	0.034354
bisa	0.031212
aplikasi	0.028013
lambat	0.025969
sudah	0.025941
gak	0.025581
mau	0.024670
saya	0.024155
jam	0.023097
masuk	0.021952
sama	0.020591

Tabel 10. Hasil TF-IDF Sentimen Netral

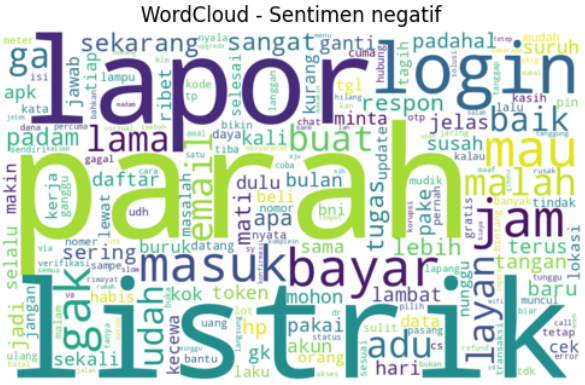
Kata	Rata-Rata TF-IDF
cukup	0.148724
di	0.047744
bisa	0.042056
nya	0.040610
saya	0.040496
bayar	0.037463
gak	0.034516
tidak	0.031680
saldo	0.031423
kok	0.030664
ada	0.030255
mau	0.029938
email	0.029823
kenapa	0.029645
masih	0.027832

j. Word Cloud

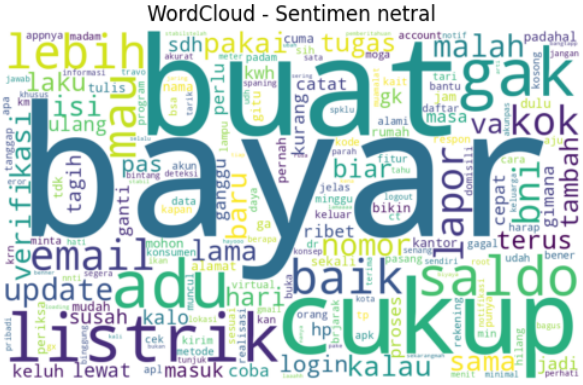
Word cloud adalah teknik visualisasi teks yang menampilkan kata-kata dominan dalam ulasan PLN Mobile berdasarkan frekuensi kemunculannya, sehingga memudahkan peneliti mengenali topik, keluhan, dan kata penting yang menjadi sorotan pengguna. Hasil *output visualisasi teks word cloud* dari sentimen positif, negatif, hingga netral ditunjukkan pada pada gambar 3,4, dan 5 dibawah ini.



Gambar 3. Word Cloud - Sentimen Positif



Gambar 4. Word Cloud - Sentimen Negatif

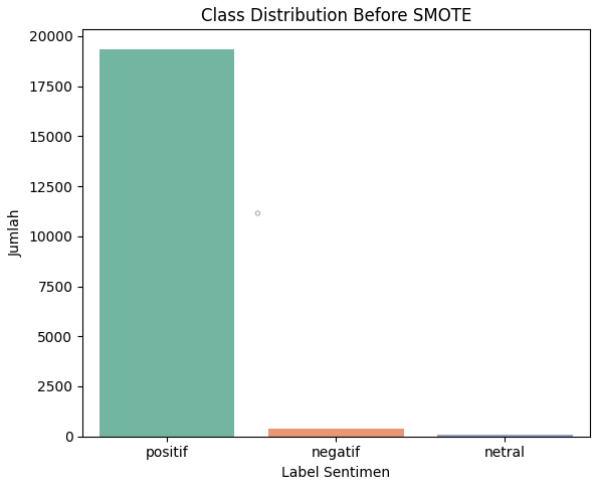


Gambar 5. Word Cloud - Sentimen Netral

ISSN: 2654-3788
e-ISSN: 2654-4229
DOI: https://doi.org/10.32493/jtsi.v9i1.54847

k. Oversampling Dengan SMOTE

Penelitian ini menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi ketidakseimbangan kelas sentimen dalam ulasan PLN Mobile, yang didominasi oleh kelas positif. SMOTE menghasilkan data sintetis bagi kelas minoritas (negatif dan netral) melalui interpolasi, sehingga distribusi menjadi lebih seimbang (Hadiprakoso et al., n.d.). Implementasi dilakukan setelah tahap praproses dan transformasi data ke bentuk numerik: TF-IDF untuk Naïve Bayes dan urutan integer untuk LSTM. Dengan parameter sampling_strategy=auto dan k_neighbors=5, oversampling dilakukan menggunakan library imblearn. Hasilnya, dataset yang lebih proporsional diharapkan meningkatkan akurasi dan objektivitas model dalam mengenali semua kategori sentimen. Hasil output distribusi kelas sentimen sebelum dan sesudah dilakukan proses oversampling menggunakan metode SMOTE pada gambar 6 dan 7 dibawah ini.



Gambar 6. Hasil Oversampling Dengan SMOTE - Sebelum

Gambar 7. Hasil Oversampling Dengan SMOTE - Sesudah

Gambar 6 menunjukkan ketidakseimbangan kelas sentimen dengan dominasi lebih dari 19.000 ulasan positif, sementara kelas negatif dan netral sangat sedikit. Setelah diterapkan SMOTE (Gambar 7), jumlah data tiap kelas berhasil diseimbangkan sekitar 19.000 entri, sehingga model dapat belajar lebih proporsional dan menghasilkan prediksi yang lebih adil serta akurat. Proses oversampling ini menjadi langkah penting dalam pra-pelatihan model analisis sentimen.

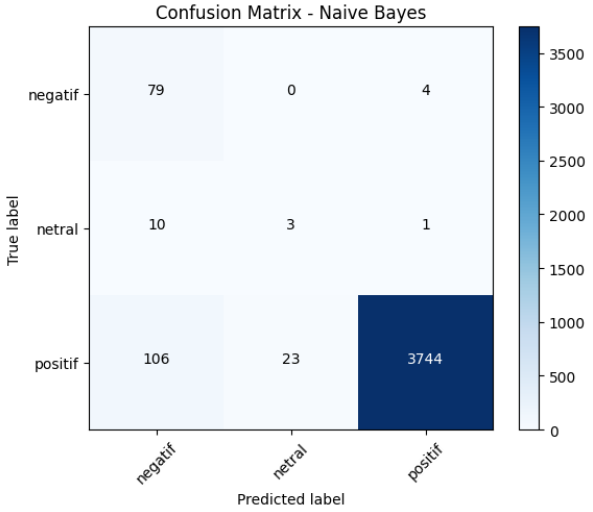
l. Implementasi Model Klasifikasi Naïve Bayes

Proses klasifikasi dilakukan dengan algoritma *Multinomial Naïve Bayes* yang berbasis probabilistik dan sesuai untuk representasi *bag-of-words* atau TF-IDF (Pasaribu et al., 2025). Data ulasan PLN Mobile yang telah diseimbangkan dengan SMOTE digunakan sebagai input, sehingga model diharapkan lebih akurat dalam membedakan sentimen positif, negatif, dan netral. Hasil output dari tabel 11 dan gambar 8 dibawah ini.

Tabel 11. Laporan Hasil Klasifikasi Naïve Bayes

	Precision	Recall	F-1 Score	Support
Negatif	0.41	0.95	0.57	83
Netral	0.12	0.21	0.15	14
Positif	1.00	0.97	0.98	3873
Accuracy			0.96	3970
Macro avg	0.51	0.71	0.57	3970
Weighted avg	0.98	0.96	0.97	3970

ISSN: 2654-3788
e-ISSN: 2654-4229
DOI: https://doi.org/10.32493/jtsi.v9i1.54847



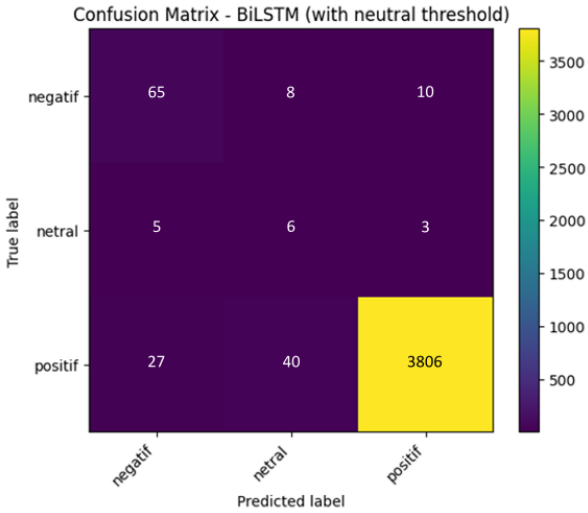
Gambar 8. Visualisasi Confusion Matrix – Naïve Bayes

m. Implementasi Model Klasifikasi Long Short Term Memory (LSTM)

Tahap selanjutnya setelah Naïve Bayes adalah klasifikasi dengan LSTM, arsitektur RNN yang mampu mengingat informasi jangka panjang dan mengatasi vanishing gradient. Data ulasan yang telah dipraproses diubah menjadi token berurutan dan dipadankan sebelum dimasukkan ke model. Dengan penerapan class weighting, LSTM dilatih agar lebih adil dalam mengenali semua kategori sentimen dan diharapkan menghasilkan prediksi yang lebih akurat dibanding metode statistik seperti *Naïve Bayes*. Hasil output dari tabel 12 dan gambar 9 dibawah ini.

Tabel 12. Laporan Hasil Klasifikasi Naïve Bayes

	Precision	Recall	F-1 Score	Support
Negatif	0.67	0.78	0.72	83
Netral	0.11	0.43	0.18	14
Positif	1.00	0.98	0.99	3873
Accuracy			0.98	3970
Macro avg	0.59	0.73	0.63	3970
Weighted avg	0.99	0.98	0.98	3970



Gambar 9. Visualisasi Confusion Matrix - LSTM

n. Perbandingan Kinerja Naïve Bayes Dan LSTM

Setelah klasifikasi dengan *Naïve Bayes* dan LSTM, dilakukan evaluasi komparatif menggunakan metrik akurasi, *precision, recall*, F1-*score*, serta *confusion matrix*. Tujuannya adalah menilai kekuatan dan kelemahan masing-masing model dalam mengklasifikasikan sentimen negatif, netral, dan positif setelah penyeimbangan data dengan SMOTE, sehingga dapat ditentukan algoritma terbaik untuk analisis sentimen ulasan PLN Mobile. Berikut hasil perbandingan kedua.

Tabel 13. Perbandingan Kinerja Naïve Bayes dan LSTM

	Accuracy	Precision	Recall	F-1 Score	Support
Naïve Bayes	0.96	0.12 – 1.00	0.21 – 0.97	0.15 – 0.98	3970
LSTM	0.98	0.11 – 1.00	0.43 – 0.98	0.18 – 0.99	3970

4. Kesimpulan

Berdasarkan hasil penelitian terhadap analisis sentimen ulasan pengguna aplikasi PLN Mobile menggunakan algoritma *Naïve Bayes* dan LSTM, dapat disimpulkan bahwa kedua algoritma mampu memberikan performa yang baik dalam klasifikasi sentimen. *Naïve Bayes* menunjukkan akurasi sebesar 96% dengan performa yang konsisten pada ketiga kelas sentimen, meskipun masih terdapat kesalahan klasifikasi antar kelas. Sementara itu, algoritma LSTM berhasil mencapai akurasi lebih tinggi yaitu 98%, dengan keunggulan pada semua metrik evaluasi utama (*precision*, *recall*, dan F1-*score*). Hal ini membuktikan bahwa LSTM lebih efektif dalam menangkap konteks dan pola bahasa pada teks ulasan, sehingga menghasilkan prediksi sentimen yang lebih akurat. Dengan demikian, LSTM lebih direkomendasikan untuk digunakan dalam analisis sentimen berbasis teks yang membutuhkan pemahaman konteks kalimat secara lebih kompleks.

Untuk pengembangan penelitian selanjutnya, peningkatan kualitas data perlu dilakukan dengan memperluas sumber ulasan, memperbanyak jumlah data, serta melakukan pembersihan yang lebih mendalam terhadap kata-kata slang, singkatan, dan ejaan tidak baku. Selain itu, penggunaan algoritma berbasis Transformer seperti

BERT dapat dipertimbangkan agar diperoleh perbandingan performa yang lebih komprehensif dalam analisis sentimen Bahasa Indonesia. Hasil penelitian ini juga sebaiknya diterapkan dalam sistem nyata, misalnya dengan mengintegrasikan analisis sentimen ke dalam dashboard monitoring guna mendukung pengambilan keputusan oleh pengembang aplikasi dan tim layanan pelanggan. Bagi pihak PLN, hasil analisis sentimen dapat dijadikan bahan evaluasi berkala untuk meningkatkan kualitas layanan, dengan menjadikan kata-kata dominan dalam sentimen negatif sebagai indikator perbaikan, serta mempertahankan kata-kata positif sebagai kekuatan layanan. Dengan langkah tersebut, PLN dapat lebih responsif terhadap kebutuhan dan kepuasan pelanggan secara objektif dan berbasis data.

Daftar Pustaka

Akbar, A. T., Saifullah, S., Prapcoyo, H., Pembangunan, U., Veteran, N., Sleman, K., & Korespondensi, P. (2024). *Klasifikasi Ekspresi Wajah Menggunakan Covolutional Neural Facial Expression Classification Using Convolutional Neural*. *11*(6). https://doi.org/10.25126/jtiik.202411888

Ardana, M., Mayasari, R., & Maulana, I. (2025). *Bulletin of Computer Science Research Perbandingan Naïve Bayes dan SVM untuk Analisis Sentimen Ulasan Kompas . id pada Data Tidak Seimbang*. *6*(1), 399–408. https://doi.org/10.47065/bulletincsr.v6i1.929

Chandra, L. F., Prasetya, A., Rahmayuna, N., Hidayati, N., Informasi, S., Semarang, U., Informasi, S., Nusantara, U. B., Soekarno-hatta, J., Bayes, N., & Playstore, G. (2025). *Klasifikasi Sentimen Ulasan Aplikasi Motorku X di Google Playstore dengan Metode Naive Bayes dan Teknik NLP*. *9*(5), 7758–7765.

Cholis, F. M., Utomo, M. C. C., & Fadhliana, N. R. (2023). Analisis Sentimen Pada Twitter Terhadap Isu Penundaan Pemilu 2024 Dengan Membandingkan Metode Long Short-Term Memory Dan Naïve Bayes Classifier. *Journal Online Institut Teknologi Kalimantan*, *1*(2). https://journal.itk.ac.id/index.php/equiva/article/view/816

Farida, Z. W., & Rochmawati, N. (2024). Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di Twitter. *Journal of Informatics and Computer Science (JINACS)*, *5*(03), 369–376. https://doi.org/10.26740/jinacs.v5n03.p369-376

Hadiprakoso, R. B., Setiawan, H., Yasa, R. N., & Siber, P. (n.d.). *Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding*.

Ikhsan, M. (2023). Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory. *The Indonesian Journal of Computer Science Research*, *2*(1), 31–41. https://doi.org/10.59095/ijcsr.v2i1.29

Informatika, J. R. (2025). *Sentiment Analysis of PLN Mobile Application Services Using Naive Bayes, Support Vector Machine (SVM) and Decision Tree*. *7*(3).

Limbong, J. J. A., Sembiring, I., Hartomo, K. D., Kristen, U., Wacana, S., & Korespondensi, P. (2022). *Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Analysis ff Review Sentiment Classification on E-Commerce Shopee Word Cloud Based with Naïve Bayes and K-Nearest Neighbor Methods*. *9*(2), 347–356. https://doi.org/10.25126/jtiik.202294960

Naufal, L. E., Surojudin, N., & Afriantoro, I. (2025). *Analisis Sentimen Aplikasi Mobile JKN di Google Play Store Menggunakan Algoritma Naive Bayes*. *6*(4), 1999–2009. https://doi.org/10.47065/josh.v6i4.7846

Pasaribu, R., Ayu, I., & Suwiprabayanti, G. (2025). *Analisis Sentimen Ulasan Aplikasi dengan Multinomial Naïve Bayes , Logistic Regression , dan SVM*. *4*(November), 63–72.

Polgan, J. M., Silaban, N., Gumay, M. G., Bisnis, S. I., Pascasarjana, P., Manajemen, M., Informasi, S., Gunadarma, U., Online, P., & Vector, S. (2025). *Analisis Sentimen Pengguna Aplikasi Pinjaman Online Melalui Media Sosial X (Twitter) Menggunakan Metode Support Vector Machine*. *14*, 2116–2130.

Purnasiwi, R. G., Kusrini, & Hanafi, M. (2023). Analisis Sentimen Pada Review Produk Skincare Menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM). *Innovative: Journal Of Social Science Research*, *3*(2), 11433–11448.

Putra, A. R., & Ratnawati, D. E. (2025). *Analisis Sentimen Berbasis Aspek Pada Aplikasi Mobile Menggunakan Naïve Bayes Berdasarkan Ulasan Pengguna Playstore (Studi Kasus : Jconnect Mobile) Aspect-Based Sentiment Analysis on Mobile Applications Using Naïve Bayes Based on Playstore User Reviews (Case Study : Jconnect Mobile)*. *12*(2), 293–300. https://doi.org/10.25126/jtiik.2025127556

Roiqoh, S., & Zaman, B. (2023). *Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes*. *7*, 1582–1592. https://doi.org/10.30865/mib.v7i3.6194

Romadhoni, Y., & Holle, K. F. H. (2022). Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM. *Jurnal Informatika: Jurnal Pengembangan IT*, *7*(2), 118–124. https://doi.org/10.30591/jpit.v7i2.3191

Sandag, G. A., & Waworundeng, J. (2022). Analisis Sentimen Masyarakat Terhadap Exchange Tokocrypto Pada Twitter Menggunakan Metode LSTM. *CogITo Smart Journal*, *8*(2), 411–421. https://doi.org/10.31154/cogito.v8i2.418.411-421

Syah, M. P., Wardani, A. P., & Idhom, M. (2025). *Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi*. *5*(1), 47–59.

Waluyo, I. G. (2023). *Sentiment Analysis of Negative Comments on Social Media Using Long Short-Term Memory (LSTM) Method with TensorFlow Framework*. *2*(7), 1015–1030.

Wijaya, W. R., Informatika, T., Teknik, F., Nusantara, U., & Kediri, P. (2026). *Sistem Deteksi Dan Koreksi Otomatis Penulisan Bahasa Indonesia Menggunakan Dictionary Lookup dan Fuzzy Logic*. *5*, 114–121.