

Analisis Klasifikasi Sentimen Pengguna MyPertamina Menggunakan Metode Evaluasi Precision, Recall, dan F1-Score

Randi Estian Pambudi¹, Hendri Purnomo², Adimas Aglasia³
^{1,2,3}IIB Darmajaya JL Zainal Abidin Pagar Alam No 93 Kedaton Bandar Lampung

Info Artikel

Riwayat Artikel:

Received July 20, 2025

Revised July 25, 2025

Accepted August 02, 2025

Corresponding Author:

Hendri Purnomo

Email: Hendrialie@darmajaya.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – This study aims to classify the sentiments expressed in user reviews of the MyPertamina application using the Naive Bayes algorithm. The text data were preprocessed through tokenization, case folding, stopword removal, and stemming. Reviews were categorized into three sentiment classes: positive, neutral, and negative. Classification performance was evaluated using standard metrics: Precision (how accurate the positive predictions are), Recall (how many actual positive cases were captured), F1-Score (harmonic mean of precision and recall), and Accuracy. The results show that 87.3% of the reviews were negative, indicating a high level of user dissatisfaction. The model achieved an accuracy of 85.09%. Detailed evaluation scores include a precision of 87% for negative reviews, 85% for positive, and 82% for neutral. However, the confusion matrix revealed a strong bias toward the majority class (negative), indicating the presence of class imbalance—when some categories have far more samples than others. These findings highlight the need for better handling of skewed datasets in sentiment classification and suggest ways to improve digital public service evaluations.

Keywords: MyPertamina; Naive Bayes; Sentiment Analysis; Text Classification; Evaluation Metrics

Abstrak – Penelitian ini bertujuan untuk mengklasifikasikan sentimen dalam ulasan pengguna aplikasi MyPertamina menggunakan algoritma Naive Bayes. Data teks diproses melalui tahapan tokenisasi, case folding, penghapusan stopword, dan stemming. Ulasan dikategorikan ke dalam tiga kelas sentimen: positif, netral, dan negatif. Kinerja klasifikasi dievaluasi menggunakan metrik standar: Precision (ketepatan prediksi positif), Recall (cakupan terhadap data positif yang sebenarnya), F1-Score (rata-rata harmonik dari precision dan recall), dan Akurasi. Hasil menunjukkan bahwa 87,3% ulasan bersifat negatif, yang mencerminkan tingkat ketidakpuasan pengguna yang tinggi. Model mencapai akurasi sebesar 85,09%. Nilai evaluasi lebih rinci menunjukkan precision sebesar 87% untuk ulasan negatif, 85% untuk positif, dan 82% untuk netral. Namun, analisis confusion matrix menunjukkan adanya bias yang kuat terhadap kelas mayoritas (negatif), yang menandakan adanya ketidakseimbangan kelas—yakni jumlah data yang jauh lebih banyak pada satu kategori dibandingkan lainnya. Temuan ini menekankan pentingnya penanganan dataset yang tidak seimbang dalam klasifikasi sentimen dan memberikan masukan bagi upaya peningkatan evaluasi layanan publik digital..

Kata Kunci: : MyPertamina; Naive Bayes; Analisis Sentimen; Klasifikasi Teks; Metrik Evaluasi

I. PENDAHULUAN

Kemunculan aplikasi digital telah mengubah cara penyampaian layanan di berbagai sektor, termasuk sektor energi. PT Pertamina (Persero), sebagai perusahaan minyak dan gas milik negara terkemuka di Indonesia, memperkenalkan aplikasi MyPertamina untuk mendukung transaksi non-tunai, menyalurkan subsidi bahan bakar secara lebih tepat sasaran, dan mempercepat layanan publik. Dalam konteks sosial dan ekonomi, MyPertamina menjadi instrumen penting dalam mendukung kebijakan pemerintah seperti pengendalian konsumsi BBM bersubsidi dan pemerataan akses layanan energi[1][2].

Namun, implementasi MyPertamina juga menghadapi sejumlah tantangan, terutama dari sisi adopsi teknologi oleh masyarakat luas yang sangat beragam latar belakang digitalnya. Banyak pengguna mengeluhkan aspek teknis seperti kesulitan login, proses registrasi yang rumit, hingga keterbatasan akses di daerah tertentu. Kondisi ini memunculkan reaksi sosial yang tercermin dalam banyaknya ulasan dan keluhan di platform digital seperti Google Play Store, forum daring, dan media sosial. Umpan balik ini merupakan sumber data yang berharga untuk mengevaluasi persepsi dan kepuasan publik terhadap layanan digital pemerintah[3][4].

Meskipun memiliki banyak keunggulan, klasifikasi sentimen seringkali menghadapi tantangan seperti distribusi kelas yang tidak seimbang, data teks yang bising, dan pemilihan model pembelajaran mesin yang sesuai. Salah satu metode yang banyak digunakan dalam klasifikasi teks adalah algoritma Naive Bayes karena kesederhanaannya, efisiensinya, dan kinerjanya yang cukup tinggi dalam pengolahan data tekstual[5][6].

Dalam konteks ini, analisis sentimen—sebagai bagian dari pemrosesan bahasa alami (Natural Language Processing, NLP)—menjadi pendekatan yang relevan untuk mengklasifikasikan opini masyarakat dalam skala besar secara otomatis. Analisis ini tidak hanya memberikan wawasan kuantitatif atas tingkat kepuasan pengguna, tetapi juga dapat menjadi masukan strategis bagi pengambil kebijakan dan pengembang aplikasi. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna terhadap aplikasi MyPertamina menggunakan algoritma Naive Bayes, dengan fokus utama pada kemampuan model dalam mengenali tiga jenis sentimen: positif, netral, dan negatif berdasarkan data ulasan pengguna. Evaluasi dilakukan menggunakan metrik Precision, Recall, F1-Score, dan Akurasi. Studi ini diharapkan dapat memberikan kontribusi spesifik berupa pemetaan persepsi pengguna secara kuantitatif, analisis tantangan teknis terkait ketidakseimbangan data sentimen, serta masukan konkret untuk pengembangan layanan publik digital yang lebih responsif terhadap kebutuhan masyarakat[7][8].

II. METODE

Penelitian ini terdiri atas empat tahapan utama, yaitu pra-pemrosesan data, ekstraksi fitur, klasifikasi sentimen, dan evaluasi performa model. Proses analisis dilakukan menggunakan perangkat lunak Python 3.10 dengan bantuan pustaka seperti scikit-learn, Sastrawi, dan pandas. Analisis dilakukan pada lingkungan Jupyter Notebook.

Penghapusan Stopword: Kata-kata umum yang kurang bermakna (misalnya "yang", "dari", "dan") dihapus menggunakan daftar stopwords bahasa Indonesia. Stemming: Kata-kata dikembalikan ke bentuk dasarnya menggunakan stemmer Sastrawi untuk meningkatkan kesesuaian istilah[9].

1. Pra-pemrosesan Teks

Pra-pemrosesan bertujuan untuk membersihkan dan menormalkan data teks agar lebih mudah diolah oleh algoritma machine learning. Langkah-langkah yang dilakukan meliputi[10][11].:

1. Case Folding: Semua huruf dalam teks diubah menjadi huruf kecil agar konsisten.
2. Tokenisasi: Kalimat dipecah menjadi kata-kata individual (token).
3. Stopword Removal: Kata-kata umum yang tidak bermakna penting (seperti "dan", "yang", "dari") dihapus dengan daftar stopwords Bahasa Indonesia dari pustaka Sastrawi.
4. Stemming: Kata-kata dikembalikan ke bentuk dasar (root word) menggunakan metode stemming Bahasa Indonesia.

2. Ekstraksi Fitur

Teks yang telah dibersihkan diubah menjadi representasi numerik menggunakan teknik **Term Frequency-Inverse Document Frequency (TF-IDF)**. TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang penting dalam konteks ulasan tertentu namun jarang muncul di ulasan lain, sehingga meningkatkan relevansi fitur terhadap konteks sentimen.

3. Klasifikasi Sentimen

Model klasifikasi yang digunakan adalah **Multinomial Naive Bayes (MNB)**, yang efektif untuk data teks dengan fitur diskrit seperti TF-IDF. Data dibagi secara stratifikasi menjadi 80% data latih dan 20% data uji untuk memastikan distribusi kelas tetap seimbang pada kedua subset data[12][13].

4. Evaluasi Model

Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama:

1. **Precision:** Mengukur proporsi prediksi sentimen positif (atau lainnya) yang benar. Misalnya, dari semua ulasan yang diprediksi negatif, berapa banyak yang benar-benar negatif.

2. **Recall:** Mengukur seberapa banyak dari semua data aktual pada suatu kelas (misalnya negatif) yang berhasil dikenali oleh model.
3. **F1-Score:** Merupakan rata-rata harmonik dari precision dan recall, berguna ketika terdapat ketidakseimbangan kelas.
4. **Akurasi:** Mengukur proporsi total prediksi yang benar dari semua prediksi yang dibuat.

Selain itu, **confusion matrix** digunakan untuk memberikan gambaran visual tentang kesalahan klasifikasi dan untuk menilai sejauh mana model cenderung bias terhadap kelas tertentu—dalam hal ini, kelas sentimen negatif yang mendominasi[14][15].

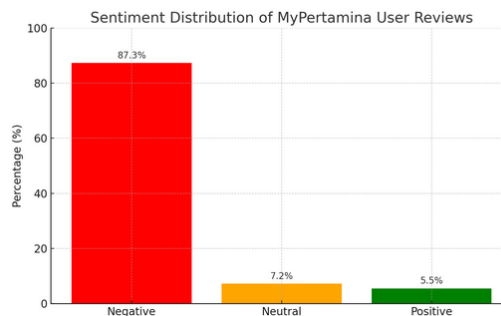
III. .HASIL DAN PEMBAHASAN

3.1 Distribusi Sentimen

Setelah proses pelabelan manual dan pra-pemrosesan, distribusi sentimen dari 15.000 ulasan pengguna adalah sebagai berikut:

- Negatif: 87,3%
- Netral: 7,2%
- Positif: 5,5%

Hasil ini menunjukkan adanya ketidakseimbangan yang signifikan, di mana mayoritas pengguna menyatakan ketidakpuasan terhadap aplikasi MyPertamina.

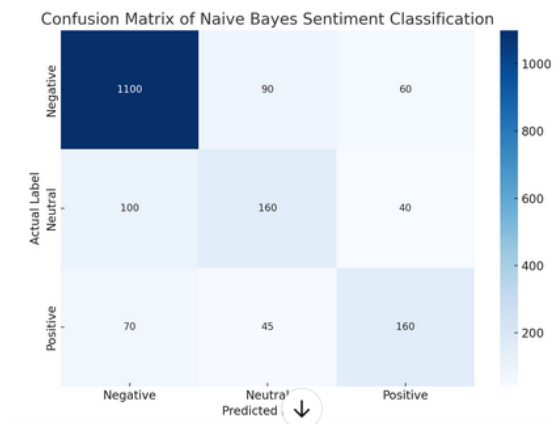


Gambar 1. Distribusi sentimen ulasan pengguna MyPertamina

Tingginya persentase sentimen negatif mengindikasikan adanya isu-isu krusial terkait dengan kemudahan penggunaan aplikasi, performa sistem, proses registrasi, dan kesalahan sistem, sebagaimana umum dilaporkan dalam umpan balik pengguna. Ketimpangan distribusi sentimen ini menjadi tantangan tersendiri dalam pembangunan model klasifikasi dan menggarisbawahi pentingnya penanganan terhadap ketidakseimbangan kelas.

Tabel 1 Precision, Recall, and F1-Score

Sentiment	Precision	Recall	F1-Score
Negative	0.87	0.84	0.85
Neutral	0.82	0.85	0.83
Positive	0.85	0.86	0.85



Gambar 2. Matriks kebingungan klasifikasi sentimen Naive Bayes

Model mencapai akurasi keseluruhan sebesar 85,09%, yang menunjukkan performa yang cukup baik secara umum. Namun, karena distribusi kelas yang tidak seimbang, model cenderung bias terhadap kelas negatif yang dominan. Hal ini juga tampak dalam analisis confusion matrix yang menunjukkan bahwa sejumlah ulasan netral dan positif salah diklasifikasikan sebagai negatif.

Tingginya nilai precision untuk sentimen negatif menunjukkan bahwa ketika model memprediksi sebuah ulasan sebagai negatif, prediksi tersebut umumnya benar. Namun, nilai recall yang sedikit lebih rendah untuk kelas negatif menunjukkan bahwa tidak semua ulasan negatif teridentifikasi dengan baik, kemungkinan disebabkan oleh ekspresi ambigu atau sarkastik dalam ulasan.

Meskipun performa model relatif seimbang di semua kelas, dataset yang timpang mempengaruhi sensitivitas model, khususnya terhadap kelas minoritas. Dominasi ulasan negatif dapat menyebabkan model melakukan overfitting pada kelas tersebut.

Untuk mengatasi masalah ini, penelitian lanjutan dapat menerapkan teknik oversampling seperti SMOTE (Synthetic Minority Over-sampling Technique) atau menggunakan pembelajaran berbasis biaya (cost-sensitive learning) untuk memberikan penalti terhadap kesalahan klasifikasi kelas minoritas. Selain itu, penggunaan metode ensemble seperti Random Forest atau XGBoost, serta model deep learning seperti BERT, dapat memberikan generalisasi dan performa yang lebih baik dalam dataset yang sangat tidak seimbang.

3.3 Implikasi Temuan

Dominasi sentimen negatif dalam ulasan MyPertamina mengindikasikan bahwa keluhan pengguna bersifat konsisten dan mungkin sistemik. Secara praktis, temuan ini dapat digunakan sebagai mekanisme umpan balik bagi pengembang dan pengambil kebijakan untuk memprioritaskan peningkatan fitur, menyempurnakan desain antarmuka pengguna, dan mengoptimalkan infrastruktur backend.

Lebih jauh, integrasi analisis sentimen ke dalam kerangka pemantauan berkelanjutan akan memungkinkan PT Pertamina untuk secara proaktif memantau persepsi publik dan merespons ketidakpuasan pengguna secara cepat, sehingga meningkatkan kualitas layanan publik digital.

IV. SIMPULAN

Penelitian ini menunjukkan bahwa algoritma **Naive Bayes** cukup efektif dalam mengklasifikasikan sentimen pengguna terhadap aplikasi MyPertamina, dengan capaian akurasi keseluruhan sebesar **85,09%**. Model mampu membedakan tiga jenis sentimen—positif, netral, dan negatif—dengan tingkat precision dan recall yang cukup stabil di semua kelas. Namun, distribusi data yang tidak seimbang, di mana **87,3%** ulasan tergolong negatif, menyebabkan model cenderung bias terhadap kelas dominan.

Implikasi praktis dari temuan ini adalah bahwa mayoritas pengguna masih belum puas terhadap kinerja aplikasi, terutama terkait fitur teknis dan pengalaman pengguna (user experience). Oleh karena itu, hasil klasifikasi ini dapat dijadikan masukan langsung bagi pengembang MyPertamina dan pengambil kebijakan untuk memperbaiki antarmuka pengguna, menyederhanakan proses registrasi, serta meningkatkan stabilitas aplikasi secara menyeluruh.

DAFTAR PUSTAKA

- [1] E. Prasetyo, S. Nurfauziah, W. A. Azahra, and Z. Nanda, "SENTIMENT ANALYSIS OF MYPERTAMINA ON GOOGLE PLAY STORE USING NAÏVE BAYES FOR SUSTAINABLE POLICIES," vol. 22, no. 3, pp. 297–311, 2024.
- [2] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [3] M. Ikhwan, "MYPERTAMINA PADA GOOGLE PLAYSTORE," no. x, pp. 1–12, 2024, doi: 10.56858/jmpkn.v8i1.390.
- [4] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.,* vol. 8, no. 1, p. 91, 2014, doi: 10.22146/ijccs.3499.
- [5] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC," *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023, doi: 10.54914/jtt.v9i1.609.
- [6] D. Transiska, D. Febriawan, and F. N. Hasan, "Analisis Sentimen Terhadap Penggunaan Chatgpt Berdasarka Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 1077, 2024, doi: 10.30865/mib.v8i2.7540.
- [7] I. Maulana, W. Apriandari, and A. Pambudi, "Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Mypertamina Menggunakan Support Vector Machine," *IDEALIS Indones. J. Inf. Syst.,* vol. 6, no. 2, pp. 172–181, 2023, doi: 10.36080/idealism.v6i2.3022.
- [8] S. F. Intan, I. Permana, F. N. Salisah, M. Afdal, and F. Muttakin, "Perbandingan Algoritma KNN, NBC, dan SVM: Analisis Sentimen Masyarakat Terhadap Perparkiran di Kota Pekanbaru," *JUSIFO (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 85–96, 2023, doi: 10.19109/jusifo.v9i2.21357.
- [9] D. T. Ksatria, Y. Yuneфри, and L. L. Van FC, "Analisis Sentimen Ulasan Pengguna Aplikasi Mypertamina pada Google Playstore Menggunakan K-Nearest Neighbor dan Naïve Bayes," *Teknol. Inf.,* vol. 2, no. 1, pp. 213–227, 2023, [Online]. Available: <https://journal.unilak.ac.id/index.php/Semaster/article/view/18526>
- [10] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform.,* vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.

- [11] A. Fauzi, M. F. Akbar, and Y. F. A. Asmawan, "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *J. Inform.*, vol. 6, no. 1, pp. 77–83, 2019, doi: 10.31311/ji.v6i1.5437.
- [12] K. A. Lubis, M. T. Ari Bangsa, and A. Yudertha, "Analisis Sentimen Opini Masyarakat Terhadap Pindahannya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes," *J. Teknoinfo*, vol. 18, no. 1, pp. 226–238, 2024.
- [13] M. A. Maulana, A. Setyanto, and M. P. Kurniawan, "Analisis Sentimen media Sosial Universitas Amikom Yogyakarta Sebagai Sarana Penyebaran Informasi Menggunakan Algoritma Klasifikasi Svm.," *Semin. Nas. Teknol. Inf. dan Multimed. 2018 Univ. AMIKOM Yogyakarta, 10 Februari 2018*, pp. 7–12, 2018.
- [14] R. Safitri, I. Ali, and N. Rahaningsih, "Analisis Sentimen Terhadap Tren Fashion Di Media Sosial Dengan Metode Support Vector Machine (Svm)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1746–1754, 2024, doi: 10.36040/jati.v8i2.9045.
- [15] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.