

Sentiment Classification of MyPertamina Reviews Using Naïve Bayes and Logistic Regression

Dwi Yuni Saraswati ^{1*}, Maya Rini Handayani ^{2*}, Khothibul Umam ^{3*}, Mokhammad Iklil Mustofa ^{4*}

* Teknologi Informasi, Fakultas Sains dan Teknologi, UIN Walisongo Semarang

2208096095@student.walisongo.ac.id ¹, maya@walisongo.ac.id ², khothibul_umam@walisongo.ac.id ³, iklil@walisongo.ac.id ⁴

Article Info

Article history:

Received 2025-06-02

Revised 2025-06-28

Accepted 2025-07-03

Keyword:

Google Play Store,
Logistic Regression,
Naïve Bayes,
Sentiment Analysis,
TF-IDF.

ABSTRACT

This research conducts a comparative evaluation of the effectiveness of the Naïve Bayes and Logistic Regression algorithms in mapping public perceptions of the MyPertamina application on the Google Play Store. The data consists of 2,000 user reviews obtained through a scraping technique. The research steps include labeling the reviews as positive or negative, followed by pre-processing and TF-IDF weighting. The dataset was systematically divided into two parts, with 80% allocated for model training and the remaining 20% for evaluation. The Naïve Bayes and Logistic Regression models were implemented using the Python programming language and evaluated based on accuracy, precision, recall, and F1-score metrics. The analysis shows that Logistic Regression achieved an accuracy of 86%, while Naïve Bayes achieved 81%. Logistic Regression demonstrated superior performance as it effectively captures linear relationships between features in TF-IDF representations and provides a more balanced outcome in terms of precision and recall. In contrast, Naïve Bayes is more influenced by high-frequency word distributions and does not account for feature correlations, which can limit its performance in certain contexts. Therefore, Logistic Regression is considered more suitable for sentiment classification tasks in this study. These findings emphasize the importance of selecting appropriate algorithms for sentiment analysis and suggest opportunities for future research using alternative methods to enhance predictive accuracy.



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

I. PENDAHULUAN

Teknologi digital telah mengubah banyak aspek kehidupan, termasuk sistem pembayaran dan transaksi keuangan. PT Pertamina (Persero) telah mengembangkan aplikasi MyPertamina, yang merupakan solusi pembayaran digital bagi konsumen bahan bakar minyak (BBM). Aplikasi ini memungkinkan pengguna mengelola program loyalitas pelanggan dan memungkinkan mereka untuk membeli BBM secara non-tunai. Meskipun demikian, MyPertamina telah menerima banyak ulasan dari pengguna sejak peluncurannya, yang mencakup ulasan positif dan negatif.

Analisis terhadap ulasan menjadi sangat penting untuk melihat ulasan pengguna untuk mengetahui masalah dan kepuasan pengguna dengan aplikasi[1]. Pengalaman pengguna dengan aplikasi biasanya diwakili oleh ulasan di Google Play Store. Menurut beberapa penelitian sebelumnya,

ulasan pengguna dapat dikategorikan berdasarkan sentimen; sentimen ini biasanya dikelompokkan menjadi dua, yakni sentimen positif serta sentimen negatif[2].

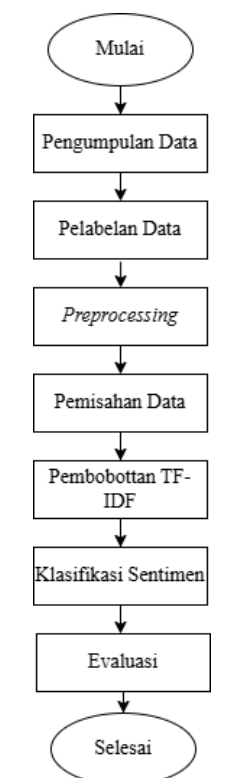
Penelitian sebelumnya tentang analisis sentimen dalam ulasan aplikasi MyPertamina telah dilakukan. Penelitian oleh Darmawan, dkk. melakukan analisis sentimen dengan algoritma Naive Bayes dan menemukan sebagian besar dari ulasan cenderung negatif, dengan akurasi model sebesar 91%[3]. Penelitian lain oleh Maulana, dkk. menggunakan algoritma Naive Bayes Classifier (NBC), menunjukkan akurasi 87%, ulasan negatif lebih banyak digunakan daripada ulasan positif[4]. Sementara itu, penelitian oleh Putri, dkk. melakukan perbandingan performa algoritma Naive Bayes dengan K-Nearest Neighbor, di mana Naive Bayes mendapatkan akurasi lebih tinggi dalam mengklasifikasikan sentimen pengguna aplikasi[5].

Algoritma Naïve Bayes dan Logistic Regression merupakan dua metode yang sering diterapkan, karena kemampuannya untuk mengklasifikasikan teks secara akurat dengan tingkat ketepatan yang relatif tinggi[6]. Algoritma Naïve Bayes menunjukkan kinerja yang memuaskan dalam pengolahan dataset dengan jumlah data latih yang relatif kecil serta mampu memberikan hasil yang cepat. Sementara itu, Logistic Regression sering digunakan karena kemampuannya dalam menangani data linear serta memberikan interpretasi probabilitas dari hasil klasifikasi[7].

Penelitian ini bertujuan untuk membandingkan tingkat akurasi algoritma Naïve Bayes dan Logistic Regression dalam menganalisis ulasan pengguna aplikasi MyPertamina yang diperoleh dari platform Google Play Store. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai persepsi pengguna, serta menjadi masukan yang relevan bagi pengembang dalam upaya peningkatan kualitas aplikasi.

II. METODE

Studi ini bertujuan untuk mengklasifikasikan ulasan pengguna ke dalam dua kategori sentimen, yaitu positif dan negatif. Data yang digunakan diperoleh dari penilaian aplikasi MyPertamina pada platform Google Play Store, dan telah melalui beberapa tahapan praproses sebelum diterapkan ke dalam klasifikasi model. Metode yang digunakan dalam penelitian ini adalah Naïve Bayes dan Logistic Regression, yang dipilih karena keduanya memiliki keunggulan dalam klasifikasi teks, khususnya dari segi akurasi yang memadai serta kemudahan dalam implementasinya.



Gambar 1. Alur Penelitian

Gambar 1 merupakan alur penelitian yang diawali dengan proses pengumpulan dataset, dilanjutkan dengan pelabelan dataset, dan *preprocessing* data. Setelah itu, data dipisahkan sebagai data untuk pelatihan dan data untuk pengujian. Selanjutnya, dilakukan tahap pembobotan fitur menggunakan TF-IDF untuk mengonversi data teks ulasan menjadi representasi numerik. Proses berikutnya adalah penerapan algoritma Naïve Bayes serta Logistic Regression. Terakhir, Evaluasi dilakukan untuk mengukur kinerja model sebelum menyatakan penelitian selesai.

A. Pengumpulan Data

Proses pengumpulan data dilakukan dengan teknik scraping dan menghasilkan data sebanyak 2.000 ulasan yang terdiri dari nilai bintang 1 hingga 5. Data yang dikumpulkan mencakup ulasan dan nilai bintang pengguna. Data yang diperoleh akan disimpan ke dalam sebuah basis data, sehingga dapat digunakan secara efektif dalam proses penelitian selanjutnya.

B. Pelabelan Data

Setelah data tinjauan dikumpulkan, library NLTK digunakan untuk memberi label positif dan label negatif. Setelah data berhasil dikumpulkan, ulasan kemudian diklasifikasikan ke dalam dua kategori. Kategori positif mencakup ulasan yang memiliki rating 3, 4, dan 5, sedangkan kategori negatif mencakup ulasan rating 1 dan 2[8]. Pemisahan data ini mempermudah proses penelitian yang akan dilanjutkan selanjutnya, serta fokus pada ulasan yang paling sesuai dengan kategori yang telah ditetapkan[9]. Oleh karena itu, lebih mudah dan efektif untuk memahami pendapat pengguna aplikasi MyPertamina.

C. Preprocessing

Setelah pelabelan dilakukan, data ulasan kemudian melewati tahapan *preprocessing*. Tahapan ini bertujuan untuk membuat data lebih bersih dan lebih sederhana sehingga algoritma pembelajaran mesin dapat mengolah data dengan lebih baik. Beberapa langkah teknis diperlukan dalam proses ini, seperti *case folding*, *tokenizing*, *stopword removal*, serta *stemming*.

Case Folding adalah suatu tahap dalam pengolahan data yang bertujuan untuk mengubah semua huruf dalam teks ulasan menjadi huruf kecil. Hal ini dilakukan untuk menyamakan penggunaan huruf kapital dan huruf kecil yang berpotensi menyebabkan ketidakkonsistenan dalam analisis[10]. Tahap berikutnya adalah *tokenizing*, yaitu langkah pemisahan kalimat menjadi unit-unit kata secara individu untuk mempermudah analisis teks pada level kata[11]. *Stopword removal*, yaitu menghilangkan kata yang dianggap tidak berkontribusi signifikan terhadap makna teks secara keseluruhan. Kata-kata tersebut, apabila tidak dihilangkan, berpotensi menimbulkan gangguan atau menyebabkan ketidaktepatan dalam proses analisis teks[12]. *Stemming* adalah proses untuk menghilangkan afiksasi dari kata berimbuhan dan mengembalikannya ke bentuk dasar. Dalam penelitian ini, proses *stemming* menggunakan pustaka

Sastrawi yang disesuaikan dengan ejaan bahasa Indonesia[13].

D. Pemisahan Data

Setelah *preprocessing* selesai, data dibagi menjadi dua bagian, dengan 80% digunakan sebagai data pelatihan dan 20% sisanya sebagai data pengujian[14], dengan menggunakan teknik pembagian *train-test split*. Pemisahan ini bertujuan mampu melakukan prediksi secara akurat terhadap data yang belum pernah ditemui sebelumnya, sehingga evaluasi yang diperoleh bersifat objektif dan mencerminkan performa riil dari model.

E. Pembobotan TF-IDF

TF-IDF merupakan teknik pembobotan yang mengonversi teks menjadi representasi numerik dengan menghitung hasil perkalian antara frekuensi kemunculan kata dalam suatu dokumen (TF) dan nilai kebalikan dari frekuensi kata tersebut di seluruh korpus (IDF)[15]. Dengan demikian, istilah yang sering muncul dalam suatu dokumen, namun jarang dijumpai pada dokumen lain, akan mendapat bobot yang lebih tinggi. Proses ini menghasilkan nilai bobot pada setiap kata yang mencerminkan tingkat kepentingan kata tersebut dalam masing-masing ulasan, sehingga dapat digunakan sebagai fitur dalam tahap analisis selanjutnya.

F. Klasifikasi Sentimen

Proses pemodelan dilakukan dengan menerapkan dua algoritma klasifikasi teks yang berbeda, yaitu Naïve Bayes dan Logistic Regression. Algoritma Naïve Bayes dipilih karena kemampuannya dalam mengestimasi probabilitas berdasarkan frekuensi kemunculan serta pola kombinasi nilai dalam dataset[16]. Selain itu, metode ini dikenal efisien dalam hal waktu transmisi, sehingga cocok digunakan untuk klasifikasi teks berukuran besar dengan representasi fitur yang bersifat sparse[17]. Sementara itu, Logistic Regression dipilih karena efektif dalam menangani hubungan linier antar fitur dan mampu memberikan performa yang tinggi serta stabil[18], khususnya pada data teks pendek seperti ulasan aplikasi, sehingga menjadikannya salah satu metode yang andal dalam analisis sentimen.

G. Evaluasi

Performa model dievaluasi dengan menggunakan sejumlah matriks evaluasi standar dalam klasifikasi, yaitu akurasi, presisi recall dan f-1 score. Metrik-metrik tersebut digunakan untuk menilai sejauh mana model mengklasifikasikan data dengan akurat dan konsisten. Dengan membandingkan hasil evaluasi antara kedua algoritma tersebut, melalui penelitian ini diharapkan mampu menyajikan gambaran objektif terhadap efektivitas masing-masing model dalam mengklasifikasikan persepsi masyarakat terhadap aplikasi MyPertamina, berdasar pada analisis ulasan yang diperoleh dari Google Play Store. Berikut persamaan (1) hingga (4) merupakan rumus untuk menghitung nilai akurasi, presisi, recall dan f1-score.

Akurasi adalah proporsi prediksi yang benar dari seluruh prediksi yang dilakukan.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

Presisi mengukur seberapa banyak prediksi positif yang benar dari seluruh prediksi positif.

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

Recall adalah rasio data positif yang berhasil dikenali dengan benar oleh model.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

f1-score merupakan rata-rata harmonis dari presisi dan recall

$$f1 - \text{score} = 2 \times \frac{\text{recall} \times \text{presisi}}{\text{recall} + \text{presisi}} \times 100\% \quad (4)$$

Persamaan (1) hingga (4) merupakan rumus confusion matrix yang digunakan untuk menghitung nilai akurasi, presisi, recall dan f1-score.

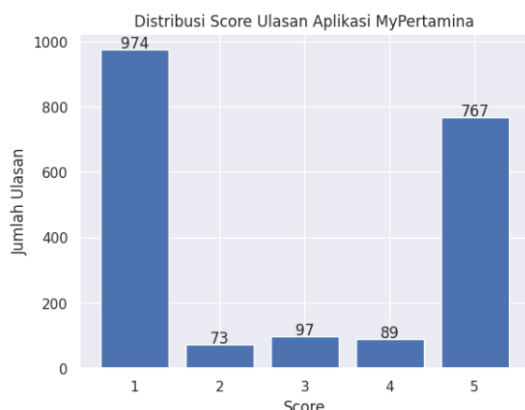
III. HASIL DAN PEMBAHASAN

A. Pengambilan Data

Proses pengambilan data menggunakan teknik *scraping* yang digunakan oleh Google Colab. Untuk mendapatkan data ulasan di Google Play Store, *install Google-play-scraper*. Hasil pengambilan data berisi ulasan, rating, tanggal, dan username sebanyak 2.000 data terbaru. Tabel 1 merupakan lima teratas dataset mentah hasil dari pengambilan data melalui metode *scraping*.

TABEL 1
DATASET MENTAH

Username	Score	at	Content
Frida Simamora	1	2025-05-11 16:01:58	Sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. Karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit BBMnya
Dwi Ariani	5	2025-05-11 15:48:56	Terbaik
Rut NATALIA PURBA	5	2025-05-11 13:09:04	Oke
Arif Maulana	1	2025-05-11 12:49:35	Aplikasi ribet, beli BBM dipersulit
Randi Randi	3	2025-05-11 12:13:56	Tidak bisa daftar barcode kendaraan



Gambar 2. Diagram Dataset

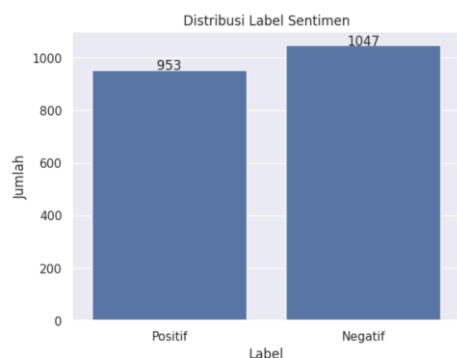
Gambar 2 merupakan diagram batang dari dataset mentah. Setelah proses ekstraksi data dilakukan, ditemukan bahwa aplikasi MyPertamina memperoleh berbagai macam ulasan dari pengguna dengan tingkat kepuasan yang bervariasi, yang tercermin melalui pemberian rating bintang 1 hingga 5.

B. Pelabelan Data

Setelah data berhasil dikumpulkan, tahap selanjutnya adalah pelabelan sentimen yang dilakukan secara otomatis berdasarkan skor rating yang diberikan oleh pengguna. Dalam proses ini, ulasan dengan rating 1 dan 2 dikategorikan sebagai sentimen negatif, sedangkan ulasan dengan rating 3, 4, dan 5 diklasifikasikan sebagai sentimen positif. Distribusi hasil pelabelan tersebut disajikan pada Tabel 2.

TABEL 2.
HASIL PELABELAN

Score	Content	Rate
1	Sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. Karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit BBMnya	Negatif
5	Terbaik	Positif
5	Oke	Positif
1	Aplikasi ribet, beli BBM dipersulit	Negatif
3	Tidak bisa daftar barcode kendaraan	Positif



Gambar 3. Hasil Pelabelan Data

Proses pelabelan memberikan gambaran mengenai persepsi pengguna terhadap aplikasi MyPertamina. Ulasan

dibedakan menjadi dua kategori utama, yakni positif dan negatif[4]. Hasil pelabelan menunjukkan bahwa terdapat 953 ulasan dengan sentimen positif dan 1.047 ulasan dengan sentimen negatif, seperti terlihat pada Gambar 3.

C. Preprocessing

Tahap berikutnya adalah *preprocessing*, yang merupakan langkah penting dan mendasar sebelum menerapkan algoritma klasifikasi pada dataset. Tahapan-tahapan yang diterapkan dalam proses *preprocessing* meliputi beberapa langkah sebagai berikut:

1) *Case Folding*: Proses mengonversi seluruh karakter huruf kapital pada data ke dalam huruf kecil guna menyamakan format teks dan meminimalkan perbedaan akibat kapitalisasi[19].

TABEL 3.
CASE FOLDING

Content	Case Folding
Sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. Karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit BBMnya	sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit bbmnya

Tabel 3 menunjukkan bagaimana data teks berubah sebelum dan sesudah dilakukan *case folding*.

2) *Tokenizing*: Proses memisahkan sebuah kalimat atau teks menjadi unit-unit kata yang berdiri sendiri. Setiap kalimat akan diuraikan menjadi kata-kata individual secara berurutan, sehingga memungkinkan analisis yang lebih mendalam terhadap struktur dan makna dari teks tersebut[20]. Dengan melakukan tokenisasi, sistem dapat mengenali setiap kata sebagai elemen terpisah yang nantinya akan dianalisis lebih lanjut dalam proses klasifikasi atau analisis sentimen. Tahapan ini merupakan dasar dalam analisis sentimen, sebab memungkinkan model agar dapat mengenali serta memahami teks dalam bentuk yang lebih terorganisir.

TABEL 4.
TOKENIZING

Content	Tokenizing
sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit bbmnya	['sekarang', 'setiap', 'daftar', 'akun', 'lagi', 'ada', 'masalah', 'system', 'tolonglah', 'jika', 'memang', 'tidak', 'bisa', 'lagi', 'daftar', 'dihapus', 'saja', 'sistem', 'barcode', 'karena', 'banyak', 'yang', 'tidak', 'mengerti', 'cara', 'daftar', 'akun', 'belum', 'lagi', 'daftar', 'unit', 'bbmnya']

Tabel 4 merupakan tabel hasil setelah melakukan *tokenizing* pada dataset.

3) *Stopword removal*: Proses penghapusan *stopword removal*, yang juga merupakan langkah penting dalam *preprocessing* data. *Stopword removal* mengacu pada kata-kata yang tidak memberikan kontribusi berarti dalam analisis,

seperti "di", "serta", "cara", "selalu", dan hal serupa, yang diidentifikasi berdasarkan daftar *stopword removal* dalam database[21]. Melalui tahapan ini, kata-kata kunci yang memiliki relevansi lebih tinggi dalam analisis sentimen dapat lebih ditekankan. Maka dari itu, proses *stopword removal* berkontribusi terhadap peningkatan akurasi analisis sentimen melalui penghilangan informasi yang tidak relevan, sehingga menghasilkan akurasi yang lebih tinggi.

TABEL 5.
STOPWORD REMOVAL

Content	Stopword removal
sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit bbmnya	['sekarang', 'daftar', 'akun', 'masalah', 'system', 'tolonglah', 'memang', 'daftar', 'dihapus', 'sistem', 'barcode', 'banyak', 'yang', 'mengerti', 'cara', 'daftar', 'akun', 'daftar', 'unit', 'bbmnya']

Tabel 5 di atas menyajikan hasil setelah penerapan teknik *stopword removal*, yang mengidentifikasi serta menghilangkan kata-kata yang tidak memberikan informasi yang relevan dalam proses analisis data.

4) *Stemming*: Data yang telah melalui tahap *stopword removal* akan dikonversi dari kata yang mengandung imbuhan diubah menjadi bentuk dasarnya[22]. *Stemming* menggunakan bantuan library Sastrawi sebagai alat utama dalam melakukan *stemming* terhadap kata-kata dalam Bahasa Indonesia.

TABEL 6.
HASIL STEMMING

Content	Stemming
sekarang setiap daftar akun lagi ada masalah system tolonglah jika memang tidak bisa lagi daftar dihapus saja sistem barcode. karena banyak yang tidak mengerti cara daftar akun belum lagi daftar unit bbmnya	['sekarang', 'daftar', 'akun', 'masalah', 'system', 'tolong', 'memang', 'daftar', 'hapus', 'sistem', 'barcode', 'banyak', 'yang', 'erti', 'cara', 'daftar', 'akun', 'daftar', 'unit', 'bbmnya']

Tabel 6 menampilkan hasil yang diperoleh setelah menerapkan proses *stemming* pada dataset, di mana kata-kata yang berimbuhan diubah ke bentuk dasarnya.

D. Pemisahan Data

Pemisahan data dilakukan dengan membagi dataset ke dalam dua kelompok utama, yaitu yakni 80% dialokasikan untuk data latih, sementara 20% sisanya digunakan sebagai data uji. Langkah ini bertujuan untuk menjamin kecukupan data yang diperlukan dalam proses pelatihan dan dapat dievaluasi secara optimal menggunakan data yang belum dipublikasikan. Pendekatan ini diharapkan dapat menghasilkan penilaian yang lebih objektif terhadap performa algoritma dalam mengklasifikasikan sentimen pada ulasan pengguna aplikasi MyPertamina

E. Pembobotan TF-IDF

Pada tahap pembobotan TF-IDF, data ulasan aplikasi MyPertamina dikonversi menjadi representasi numerik

dengan pemberian bobot pada setiap istilah berdasarkan frekuensi kemunculannya dalam dokumen (*Term Frequency*) serta kelangkaannya dalam keseluruhan korpus dokumen (*Inverse Document Frequency*)[23]. Metode ini menghasilkan fitur yang menonjolkan istilah-istilah signifikan sekaligus mereduksi kontribusi istilah yang bersifat umum dan kurang informatif. Representasi fitur yang diperoleh melalui TF-IDF selanjutnya digunakan dalam proses klasifikasi sentimen dengan menerapkan algoritma Naïve Bayes dan Logistic Regression. Hasil pembobotan TF-IDF dapat dilihat pada Gambar 4, yang menampilkan lima kata dengan bobot tertinggi berdasarkan rata-rata kemunculannya dalam dokumen.

TF-IDF Tertinggi Secara Rata-rata:

	Kata	Bobot TF-IDF
0	mantap	0.049158
1	ok	0.047082
2	aplikasi	0.042414
3	daftar	0.030951
4	bagus	0.024461

Gambar 4. Hasil Pembobotan TF-IDF

Berdasarkan hasil perhitungan rata-rata bobot TF-IDF, lima kata dengan nilai tertinggi adalah “mantap” (0,049158), “ok” (0,047082), “aplikasi” (0,042414), “daftar” (0,030951), dan “bagus” (0,024461).

F. Klasifikasi Sentimen

Pada tahapan ini, klasifikasi dilakukan dengan menerapkan algoritma Naïve Bayes yang diimplementasikan dengan pemrograman Python. Dataset yang digunakan dibagi secara acak oleh sistem menjadi dua subset, yakni 80% untuk data pelatihan dan 20% untuk data pengujian, guna memastikan proses pelatihan dan evaluasi model berjalan secara optimal.

G. Evaluasi

Evaluasi kinerja model klasifikasi sentimen dilakukan dengan menggunakan metrik evaluasi umum, yaitu akurasi, presisi, recall, dan F1-score. Metrik-metrik ini digunakan untuk menilai sejauh mana model mampu mengklasifikasikan data uji secara akurat. Selain itu, nilai presisi, recall, dan F1-score juga dianalisis secara terpisah untuk masing-masing kelas sentimen (positif dan negatif), guna memberikan pemahaman yang lebih rinci terhadap performa model dalam mengenali setiap kategori. Gambar 5 dan Gambar 6 menunjukkan hasil evaluasi terhadap data pengujian.

	precision	recall	f1-score	support
Negatif	0.75	0.95	0.84	209
Positif	0.92	0.66	0.77	191
accuracy			0.81	400
macro avg	0.84	0.80	0.80	400
weighted avg	0.83	0.81	0.81	400

Gambar 5. Klasifikasi Naive Bayes

Pada Gambar 5, akurasi Naive Bayes sebesar 0.81 mengindikasikan bahwa model berhasil melakukan klasifikasi yang tepat terhadap 81% dari seluruh data yang

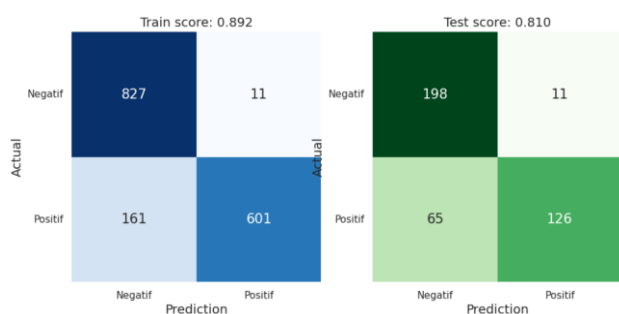
dianalisis. Sementara itu, label positif menunjukkan nilai presisi sebesar 0.92, recall sebesar 0.66, dan F1-Score sebesar 0.77. Sementara itu, label negatif memperoleh presisi sebesar 0.75, recall sebesar 0.95 dan f1-score sebesar 0.84.

	precision	recall	f1-score	support
Negatif	0.84	0.92	0.88	209
Positif	0.91	0.80	0.85	191
accuracy			0.86	400
macro avg	0.87	0.86	0.86	400
weighted avg	0.87	0.86	0.86	400

Gambar 6. Klasifikasi Logistic Regression

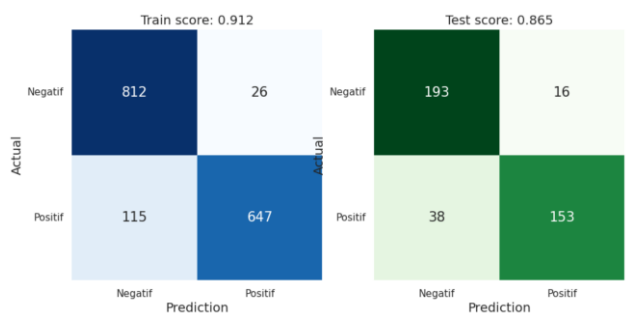
Gambar 6, akurasi Naive Bayes sebesar 0.86 mengindikasikan bahwa model berhasil melakukan klasifikasi yang tepat terhadap 86% dari seluruh data yang dianalisis. Sementara itu, label positif menunjukkan nilai presisi sebesar 0.91, recall sebesar 0.80, dan f1-score sebesar 0.85. Sementara itu, label negatif memperoleh presisi sebesar 0.84, recall sebesar 0.92 dan f1-score sebesar 0.88.

Untuk memperoleh gambaran yang lebih komprehensif mengenai pola kesalahan prediksi dari masing-masing model klasifikasi, disajikan *confusion matriks* sebagai representasi kinerja klasifikasi tiap algoritma, sebagaimana ditunjukkan pada Gambar 7 dan Gambar 8.



Gambar 7. *Confusion Matriks* Naive Bayes

Gambar 7 menunjukkan algoritma Naïve Bayes akurat dalam mengklasifikasikan ulasan negatif dengan 198 benar, namun kurang optimal pada ulasan positif dengan 65 *false negative*. Hal ini menunjukkan kecenderungan model terhadap kelas negatif.



Gambar 8. *Confusion Matriks* Logistic Regression

Gambar 8 menunjukkan bahwa Logistic Regression memberikan hasil yang lebih seimbang, dengan 193 ulasan negatif dan 153 positif diklasifikasikan dengan benar, serta

kesalahan klasifikasi yang lebih rendah. Model ini dinilai lebih stabil dalam mengenali kedua kelas sentimen.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terhadap 2.000 ulasan pengguna aplikasi MyPertamina yang diperoleh dari Google Play Store. Setelah labelling dan dilakukan normalisasi, data menunjukkan 953 ulasan positif dan 1.047 ulasan negatif. Tahapan-tahapan penting pada penelitian ini, seperti *case folding*, *stopword removal*, *tokenizing* dan *stemming*, berhasil meningkatkan kualitas data sebelum diterapkan ke dalam model klasifikasi. Selanjutnya, fitur teks diekstraksi menggunakan metode TF-IDF untuk mengonversi data ulasan menjadi representasi numerik. Penggunaan algoritma Naïve Bayes dalam klasifikasi sentimen memberikan hasil yang cukup memuaskan, dengan tingkat akurasi sebesar 81%, untuk data uji. Sedangkan algoritma Logistic Regression memberikan akurasi 86%. Hasil ini menunjukkan bahwa Logistic Regression lebih unggul dalam hal akurasi dan kemampuan mengenali data positif secara lebih luas.

Selain itu, Logistic Regression juga menunjukkan keunggulan dari sisi keseimbangan metrik evaluasi antar kelas. Hal ini mengindikasikan bahwa model ini lebih mampu menangkap relasi linier antara kata-kata dalam ulasan dan polaritas sentimen secara lebih konsisten. Sebaliknya, performa Naïve Bayes yang lebih rendah disebabkan oleh sensitivitasnya terhadap kata-kata dengan frekuensi tinggi, tanpa mempertimbangkan konteks dan hubungan antar kata yang lebih kompleks.

DAFTAR PUSTAKA

- [1] M. Suhendra, W. Swastika, and M. Subianto, "Analisis Sentimen Pada Ulasan Aplikasi Video Conference Menggunakan Naive Bayes," *Sainsbertek J. Ilm. Sains Teknol.*, vol. 2, no. 1, pp. 1–9, 2021, doi: 10.33479/sb.v2i1.145.
- [2] D. Angraina and A. Putri, "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 472–478, 2022, doi: 10.37859/coscitech.v3i3.4260.
- [3] G. Darmawan, S. Alam, and M. I. Sulisty, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naive Bayes," *STORAGE – J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023.
- [4] 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.
- [5] R. T. S. A. Putri, D. E. Ratnawati, and D. W. Brata, "Perbandingan Naive Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Aplikasi Gapura UB Berdasarkan Ulasan Pengguna pada Playstore," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 229–236, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] W. P. Ramadhan and D. Juardi, "Analisis Sentimen Ulasan Aplikasi BTN Mobile Menggunakan Algoritma Naive Bayes," vol. 13, no. 1, 2025.
- [7] A. C. Fauzan and K. Hikmah, "Implementasi Algoritma Naive Bayes Dalam Analisis Polarisasi Opini Masyarakat Terkait Vaksin Covid-19," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 122–128, 2022, doi: 10.36341/rabit.v7i2.2403.

- [8] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120340.
- [9] A. M. Putri, W. K. Nofa, and D. A. P. Hapsari, "Penerapan metode bert untuk analisis sentimen ulasan pengguna aplikasi segari di google play store," vol. 4, no. 1, pp. 89–104, 2025.
- [10] M. Apriliyani, M. I. Musyaffaq, S. Nur'Aini, M. R. Handayani, and K. Umam, "Implementasi analisis sentimen pada ulasan aplikasi Duolingo di Google Playstore menggunakan algoritma Naïve Bayes," *AITI*, vol. 21, no. 2, pp. 298–311, Sep. 2024, doi: 10.24246/aiti.v21i2.298-311.
- [11] M. Indra Buana and D. Brahma Arianto, "Analisis Sentimen Ulasan Pengguna Aplikasi ZenPro dengan Implementasi Algoritma Support Vector Machine (SVM)," *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 45–52, 2024, doi: 10.30872/atasi.v3i1.1092.
- [12] C. N. Adela, S. Karnila, S. Sutedi, and M. Agarina, "Analisis Ulasan Pengguna Aplikasi Seabank Dengan Support Vector Machine Dan Naïve Bayes," *J. Tekno Kompak*, vol. 18, no. 2, p. 441, 2024, doi: 10.33365/jtk.v18i2.4156.
- [13] D. A. N. Arifin, S. Pada, D. Teks, B. Indonesia, J. Pardede, and D. Darmawan, "Perbandingan Algoritma Stemming Porter , Sastrawi , Idris , Comparison Of Stemming Algorithms Porter , Sastrawi , Idris , And Arifin Setiono On Indonesian Text Documents," vol. 12, no. 1, 2025, doi: 10.25126/jtiik.2025128860.
- [14] S. Kusal *et al.*, "Sentiment Analysis of Product Reviews Using Deep Learning and Transformer Models: A Comparative Study," *Lect. Notes Networks Syst.*, vol. 843, no. February, pp. 183–204, 2024, doi: 10.1007/978-981-99-8476-3_15.
- [15] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks," *SINTESIA J. Sist. dan Teknol. Inf. Indones.*, vol. 1, no. 2, pp. 81–88, 2023.
- [16] H. Aprilianti, H. Mustofa, K. Umam, and M. R. Handayani, "Comparative Study of SVM , KNN , and Naïve Bayes for Sentiment Analysis of Religious Application Reviews," vol. 9, no. 3, pp. 920–927, 2025.
- [17] W. L. Azzahra, J. Indra, and S. Faisal, "Sentiment Analysis of User Reviews of the AdaKami Online Loan App from the App Store Using SVM and Naive Bayes," vol. 9, no. 3, pp. 838–850, 2025.
- [18] M. F. Zaenudin and Y. Sibaroni, "Combination Of Logistic Regression And Naïve Bayes," vol. 10, no. 2, pp. 1286–1298, 2025.
- [19] M. Tirta Nugraha, N. Nina Sulistiyowati, and U. Ultach Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat Pada Google Play Store Menggunakan Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 5, pp. 3593–3601, 2024, doi: 10.36040/jati.v7i5.7753.
- [20] K. Adib, M. R. Handayani, W. D. Yuniarti, and K. Umam, "Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM," *SINTECH (Science Inf. Technol. J.)*, vol. 7, no. 2, pp. 80–91, 2024, doi: 10.31598/sintechjournal.v7i2.1581.
- [21] Z. Firmansyah and N. F. Puspitasari, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes," *J. Tek. Inform.*, vol. 14, no. 2, pp. 171–178, 2021, [Online]. Available: <https://doi.org/10.15408/jti.v14i2.24024>
- [22] M. H. Aufan, M. R. Handayani, A. B. Nurjanna, and N. C. Hendro, "The Perceptions Of Semarang Five Star Hotel Tourists With Support Vector Machine On Google Reviews Persepsi Wisatawan Hotel Bintang Lima Semarang Dengan," vol. x, no. December, pp. 1–8, 2023.
- [23] M. D. Hendriyanto and B. N. Sari, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax," *J. Ilm. Inform.*, vol. 10, no. 02, pp. 80–84, 2022, doi: 10.33884/jif.v10i02.5477.