

Analisis Sentimen Ulasan Pengguna pada Aplikasi JMO (Jamsostek Mobile) di Google Play Store Menggunakan Algoritma Naive Bayes

Sri Damayanti^{1*}, Husna Sari², Bobby Ardiansyah³, Dicky Apdillah⁴

^{1,2,3,4} Teknik Informatika, Universitas Asahan, Jl. Jenderal Ahmad Yani No. 1, Kisaran Kota, Kecamatan Kota Kisaran Barat, Kabupaten Asahan, Sumatera Utara 21214, Indonesia.

E-mail: sridamayanti7865@gmail.com

*Corresponding Author



<https://doi.org/10.31004/jerkin.v4i4.6708>

ARTICLE INFO

Article history:

Received: 06 Jun 2026

Revised: 12 Jun 2026

Accepted: 18 Jun 2026

Kata Kunci:

Analisis Sentimen,
Naïve Bayes, JMO,
Google Play Store,
Text Mining.

Keywords:

Sentiment Analysis,
Naïve Bayes, JMO,
Google Play Store,
Text Mining.

ABSTRACT

Perkembangan teknologi informasi telah mendorong transformasi layanan publik ke arah digital, termasuk melalui pemanfaatan aplikasi mobile untuk meningkatkan efisiensi dan kemudahan akses bagi masyarakat. Salah satu inovasi layanan digital tersebut adalah aplikasi JMO (Jamsostek Mobile) yang dikembangkan oleh BPJS Ketenagakerjaan. Banyaknya ulasan yang diberikan pengguna pada Google Play Store dapat dimanfaatkan sebagai sumber informasi untuk mengukur tingkat kepuasan serta persepsi pengguna terhadap kualitas aplikasi. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis sentimen pengguna aplikasi JMO melalui pendekatan analisis sentimen menggunakan algoritma Naive Bayes. Data yang digunakan berupa 1.000 ulasan pengguna yang diperoleh dari Google Play Store melalui proses web scraping menggunakan Python dan Google Play Scraper. Tahapan penelitian mencakup preprocessing data, pelabelan sentimen berdasarkan rating bintang, ekstraksi fitur menggunakan metode TF-IDF, proses klasifikasi dengan algoritma Naive Bayes, serta evaluasi performa model menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa sentimen negatif lebih mendominasi dibandingkan sentimen positif dan netral. Sebagian besar ulasan negatif berkaitan dengan permasalahan teknis, seperti error sistem, kendala login, OTP yang tidak diterima, serta performa aplikasi yang lambat. Sebaliknya, ulasan positif umumnya menyoroti kemudahan penggunaan layanan dan fitur-fitur yang membantu pengguna. Berdasarkan hasil pengujian, model Naive Bayes menghasilkan nilai accuracy sebesar 86%, precision sebesar 84%, recall sebesar 85%, dan F1-score sebesar 84%. Temuan ini menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang cukup baik dalam mengklasifikasikan sentimen pada ulasan aplikasi mobile. Hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan studi analisis sentimen sekaligus memberikan masukan bagi pengembang aplikasi untuk meningkatkan kualitas layanan digital yang diberikan.



The rapid advancement of digital technology has prompted public service providers to adopt mobile-based applications to enhance service efficiency and accessibility for users. One such application is JMO (Jamsostek Mobile), a digital service platform developed by BPJS Ketenagakerjaan. With the growing number of users, reviews and ratings submitted through the Google Play Store have become valuable sources of information for assessing user satisfaction and evaluating application performance. This study aims to examine user sentiment toward the JMO application by employing the Naive Bayes classification algorithm. A total of 1,000 user reviews were collected from Google Play Store through web scraping techniques using Python and Google Play Scraper. The research process involved data preprocessing, sentiment labeling based on star ratings, feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method, sentiment classification with the Naive Bayes algorithm, and performance evaluation using a confusion matrix. The findings reveal that negative sentiment was more prevalent than positive and neutral sentiment. Most negative reviews were associated with technical issues, including system errors, login problems, OTP verification failures, and slow application performance. In contrast, positive reviews mainly highlighted the convenience of accessing services and the usefulness of the application's features. The evaluation results demonstrated that the Naive Bayes model achieved an

accuracy of 86%, precision of 84%, recall of 85%, and F1-score of 84%. These findings indicate that the Naive Bayes algorithm performs effectively in classifying sentiment from mobile application user reviews. Furthermore, this study contributes to the growing body of research on sentiment analysis and provides practical insights for developers seeking to improve the quality of digital public services.



This is an open access article under the CC-BY-SA license.

How to Cite: Sri Damayanti, et al. (2026), Analisis Sentimen Ulasan Pengguna pada Aplikasi JMO (Jamsostek Mobile) di Google Play Store Menggunakan Algoritma Naive Bayes, 4(4). <https://doi.org/10.31004/jerkin.v4i4.6708>

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi di berbagai bidang kehidupan, termasuk sektor pelayanan publik dan ketenagakerjaan. Transformasi digital tersebut mendorong instansi pemerintah maupun lembaga pelayanan untuk menghadirkan layanan berbasis teknologi guna meningkatkan efektivitas, efisiensi, serta kemudahan akses bagi masyarakat. Salah satu implementasi transformasi digital dalam bidang ketenagakerjaan adalah aplikasi JMO (Jamsostek Mobile) yang dikembangkan oleh BPJS Ketenagakerjaan sebagai sarana layanan digital bagi peserta program jaminan sosial ketenagakerjaan di Indonesia. Aplikasi JMO merupakan pengembangan dari aplikasi BPJSTKU yang dirancang untuk memfasilitasi pengguna dalam mengakses berbagai layanan ketenagakerjaan secara daring. Melalui aplikasi ini, pengguna dapat melakukan pengecekan saldo Jaminan Hari Tua (JHT), pengajuan dan pemantauan klaim, pembaruan data kepesertaan, serta memperoleh informasi terkait program BPJS Ketenagakerjaan. Kehadiran aplikasi tersebut menjadi bagian penting dalam upaya peningkatan kualitas layanan publik digital yang lebih cepat, fleksibel, dan mudah diakses oleh masyarakat kapan pun dan di mana pun.

Berdasarkan informasi yang tersedia pada Google Play Store, aplikasi JMO telah diunduh lebih dari 10 juta kali dan memperoleh jumlah ulasan yang sangat besar dari pengguna. Tingginya tingkat penggunaan aplikasi menunjukkan bahwa JMO menjadi salah satu platform layanan publik digital yang banyak dimanfaatkan oleh masyarakat Indonesia. Beragam ulasan yang diberikan pengguna mencerminkan pengalaman, persepsi, tingkat kepuasan, maupun keluhan terhadap layanan yang disediakan aplikasi. Selain itu, rating bintang yang diberikan pengguna juga dapat digunakan sebagai indikator untuk menilai kualitas aplikasi secara umum. Dalam pengembangan aplikasi digital, ulasan pengguna memiliki peran penting karena dapat dijadikan sebagai sumber evaluasi terhadap kualitas layanan, stabilitas sistem, serta kebutuhan pengguna. Menurut Ghose dan Ipeiritis (2011), ulasan pengguna dapat memengaruhi persepsi publik dan keputusan calon pengguna dalam memilih suatu produk atau layanan digital. Oleh sebab itu, analisis terhadap ulasan pengguna diperlukan untuk memahami kecenderungan opini masyarakat secara lebih objektif dan sistematis. Namun, jumlah ulasan yang sangat besar serta karakteristik data yang tidak terstruktur, seperti penggunaan bahasa informal, singkatan, kesalahan penulisan, dan ekspresi emosional, menjadikan proses analisis manual kurang efektif. Kondisi tersebut menuntut penggunaan pendekatan komputasional untuk mengolah dan menganalisis data teks secara otomatis.

Salah satu metode yang banyak digunakan untuk mengidentifikasi opini pengguna adalah analisis sentimen. Analisis sentimen merupakan bagian dari text mining dan natural language processing yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengelompokkan opini dalam teks ke dalam kategori tertentu, seperti sentimen positif, negatif, atau netral (Pang & Lee, 2008). Dalam konteks aplikasi mobile, analisis sentimen dapat dimanfaatkan untuk mengetahui persepsi pengguna terhadap kualitas layanan berdasarkan ulasan yang mereka berikan. Berbagai algoritma machine learning telah diterapkan dalam penelitian analisis sentimen, di antaranya Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), dan Naive Bayes. Pada penelitian ini, algoritma Naive Bayes dipilih karena memiliki kemampuan yang baik dalam melakukan klasifikasi teks dengan tingkat efisiensi yang tinggi, khususnya pada data berdimensi besar (Manning et al., 2008). Algoritma ini menggunakan pendekatan probabilistik berdasarkan Teorema Bayes untuk menentukan kemungkinan suatu data termasuk ke dalam kategori tertentu. Selain itu, Naive Bayes dikenal sederhana dalam implementasi dan telah banyak digunakan dalam berbagai penelitian analisis sentimen berbasis teks (Rish, 2001).

Penelitian mengenai analisis sentimen pada aplikasi mobile telah banyak dilakukan, namun sebagian besar berfokus pada aplikasi e-commerce, media sosial, dan layanan hiburan digital. Penelitian yang secara khusus mengkaji sentimen pengguna terhadap aplikasi layanan ketenagakerjaan masih relatif terbatas, terutama pada aplikasi JMO. Padahal, sebagai aplikasi layanan publik dengan jumlah pengguna yang besar, JMO menghasilkan beragam ulasan yang berpotensi memberikan informasi yang penting mengenai kualitas layanan digital yang diberikan kepada masyarakat. Kebaruan penelitian ini terletak pada analisis sentimen terhadap ulasan pengguna aplikasi JMO menggunakan algoritma Naïve Bayes dengan memanfaatkan rating bintang sebagai dasar pelabelan sentimen. Dalam penelitian ini, rating 4 dan 5 dikategorikan sebagai sentimen positif, rating 3 sebagai sentimen netral, sedangkan rating 1 dan 2 diklasifikasikan sebagai sentimen negatif. Pendekatan ini dinilai lebih efisien dibandingkan pelabelan manual, meskipun tetap memiliki potensi ketidaksesuaian antara isi ulasan dan rating yang diberikan pengguna (Zhang et al., 2018). Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi JMO pada Google Play Store menggunakan algoritma Naïve Bayes. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi kinerja algoritma Naïve Bayes dalam mengklasifikasikan sentimen pengguna berdasarkan data ulasan aplikasi. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan studi analisis sentimen berbasis machine learning serta menjadi bahan evaluasi bagi pengembang aplikasi dalam meningkatkan kualitas layanan digital sesuai dengan kebutuhan dan pengalaman pengguna.

METODE

Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis *machine learning*. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data secara numerik serta pengukuran kinerja model klasifikasi menggunakan indikator statistik yang terukur dan objektif. Data berupa ulasan pengguna yang awalnya berbentuk teks tidak terstruktur akan diproses dan dikonversi ke dalam representasi numerik sehingga dapat dianalisis menggunakan algoritma klasifikasi. Metode analisis sentimen diterapkan untuk mengidentifikasi, mengelompokkan, dan menganalisis opini pengguna terhadap aplikasi JMO ke dalam kategori sentimen positif, negatif, dan netral (Pang & Lee, 2008). Penelitian ini bersifat deskriptif karena bertujuan untuk menggambarkan kecenderungan persepsi dan tingkat kepuasan pengguna berdasarkan ulasan yang tersedia pada Google Play Store. Melalui analisis tersebut, dapat diperoleh informasi mengenai aspek-aspek layanan yang dianggap memuaskan maupun yang masih memerlukan perbaikan. Pendekatan *machine learning* dipilih karena memiliki kemampuan untuk memproses data teks dalam jumlah besar secara otomatis, cepat, dan efisien. Selain itu, metode ini dinilai lebih fleksibel dalam menangani variasi bahasa yang digunakan pengguna dibandingkan pendekatan berbasis leksikon (*lexicon-based approach*) (Cambria et al., 2017). Dalam penelitian ini, algoritma Naïve Bayes digunakan sebagai metode klasifikasi karena memiliki tingkat akurasi yang baik dalam pengolahan data teks, efisien secara komputasi, serta banyak digunakan dalam penelitian analisis sentimen. Dengan karakteristik tersebut, Naïve Bayes dianggap sesuai untuk mengklasifikasikan sentimen ulasan pengguna aplikasi JMO berdasarkan pola kata yang terdapat dalam data teks.

Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang berasal dari ulasan pengguna (*user reviews*) aplikasi JMO (Jamsostek Mobile) pada platform Google Play Store. Pemanfaatan data sekunder dipilih karena data tersebut tersedia secara publik dan mampu merepresentasikan pengalaman serta persepsi pengguna terhadap aplikasi secara langsung. Data yang dikumpulkan meliputi teks ulasan, rating bintang dengan skala 1 hingga 5, serta tanggal publikasi ulasan. Dari ketiga komponen tersebut, teks ulasan digunakan sebagai sumber data utama karena memuat berbagai opini, pengalaman, kritik, maupun saran yang diberikan oleh pengguna terhadap layanan aplikasi. Sementara itu, rating bintang dimanfaatkan sebagai dasar pelabelan sentimen dalam proses klasifikasi. Menurut Liu (2012), ulasan pengguna merupakan salah satu bentuk data opini yang penting dalam penelitian analisis sentimen karena mampu menggambarkan tingkat kepuasan dan persepsi pengguna terhadap suatu produk atau layanan. Dalam konteks aplikasi digital, ulasan pengguna dapat memberikan informasi mengenai kualitas layanan, performa sistem, kemudahan penggunaan, serta berbagai kendala yang dialami selama penggunaan aplikasi. Data ulasan yang diperoleh dari Google

Play Store termasuk ke dalam kategori data tidak terstruktur (*unstructured data*), sehingga memerlukan serangkaian proses *preprocessing* sebelum dapat digunakan pada tahap analisis dan klasifikasi. Selain itu, karakteristik ulasan pengguna yang beragam, seperti penggunaan bahasa formal maupun informal, singkatan, kesalahan penulisan, serta campuran bahasa, menjadi tantangan tersendiri dalam proses pengolahan data. Oleh karena itu, tahapan pembersihan dan normalisasi data diperlukan untuk meningkatkan kualitas data serta menghasilkan model klasifikasi sentimen yang lebih akurat dan efektif.

Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui teknik *web scraping* menggunakan bahasa pemrograman Python yang dijalankan pada platform Google Colaboratory (*Google Colab*). Platform ini dipilih karena berbasis *cloud*, dapat diakses secara gratis, serta menyediakan berbagai pustaka (*library*) yang mendukung pengolahan data, *machine learning*, dan *text mining*. Proses pengambilan data dilakukan dengan memanfaatkan *library* Google Play Scraper untuk mengekstraksi ulasan pengguna dari aplikasi JMO (Jamsostek Mobile) yang tersedia di Google Play Store. Teknik *web scraping* memungkinkan proses pengumpulan data dilakukan secara otomatis dan dalam jumlah besar, sehingga lebih efisien dibandingkan metode pengumpulan data secara manual. Melalui teknik ini, data ulasan pengguna dapat diperoleh secara cepat dan sistematis untuk keperluan analisis sentimen. Pada penelitian ini, jumlah data yang dikumpulkan sebanyak 1.000 ulasan pengguna, yang dianggap cukup representatif untuk menggambarkan persepsi pengguna terhadap aplikasi JMO serta mendukung proses klasifikasi sentimen. Dalam proses pengumpulan data, beberapa kriteria digunakan untuk memastikan kualitas dataset yang diperoleh. Kriteria tersebut meliputi ulasan yang memiliki teks yang dapat dianalisis, disertai dengan rating bintang, tidak mengandung data duplikat, serta ditulis dalam bahasa Indonesia atau bahasa Inggris. Penerapan kriteria ini bertujuan untuk meningkatkan kualitas data dan meminimalkan gangguan yang dapat memengaruhi hasil analisis. Setelah proses *scraping* selesai dilakukan, data yang diperoleh disimpan dalam format CSV (*Comma-Separated Values*) untuk memudahkan proses pengolahan, *preprocessing*, dan analisis pada tahap berikutnya. Format ini dipilih karena mudah diakses, kompatibel dengan berbagai perangkat lunak analisis data, serta mendukung pengelolaan dataset dalam jumlah besar secara lebih efektif.

Pelabelan Sentimen

Pelabelan sentimen pada penelitian ini dilakukan secara otomatis dengan memanfaatkan rating bintang yang diberikan pengguna pada Google Play Store. Metode ini dipilih karena lebih praktis dan efisien dibandingkan proses pelabelan manual, yang umumnya memerlukan waktu, tenaga, serta keterlibatan beberapa anotator untuk menjaga konsistensi hasil. Dengan menggunakan rating sebagai acuan, proses penentuan kategori sentimen dapat dilakukan secara lebih cepat dan sistematis pada jumlah data yang besar. Kategori sentimen dalam penelitian ini ditentukan berdasarkan skala rating sebagai berikut:

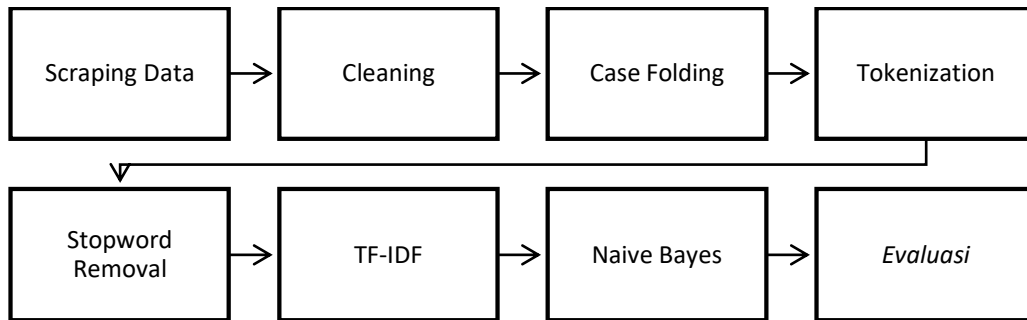
1. Rating 4 dan 5 diklasifikasikan sebagai sentimen positif.
2. Rating 3 diklasifikasikan sebagai sentimen netral.
3. Rating 1 dan 2 diklasifikasikan sebagai sentimen negatif.

Pengelompokan tersebut didasarkan pada asumsi bahwa rating yang diberikan pengguna mencerminkan tingkat kepuasan mereka terhadap kualitas layanan dan performa aplikasi. Rating yang tinggi umumnya menunjukkan pengalaman penggunaan yang positif, sedangkan rating yang rendah mengindikasikan adanya ketidakpuasan atau permasalahan yang dialami pengguna. Menurut Zhang et al. (2018), penggunaan rating sebagai label sentimen merupakan pendekatan yang efektif dalam penelitian analisis sentimen karena dapat mempercepat proses pelabelan data sekaligus mengurangi unsur subjektivitas yang sering muncul pada anotasi manual. Meskipun demikian, pendekatan ini memiliki keterbatasan karena tidak selalu terdapat kesesuaian antara isi ulasan dan rating yang diberikan pengguna. Sebagai contoh, pengguna dapat memberikan rating rendah meskipun isi ulasan cenderung positif, atau sebaliknya. Oleh karena itu, hasil klasifikasi pada penelitian ini akan dianalisis lebih lanjut untuk mengidentifikasi tingkat konsistensi antara teks ulasan dan label sentimen yang diperoleh dari rating pengguna.

Tahapan Preprocessing Data

Tahap *preprocessing* merupakan salah satu langkah penting dalam analisis sentimen yang bertujuan untuk membersihkan, menormalisasi, dan mempersiapkan data teks sebelum diproses oleh algoritma *machine learning*. Data ulasan yang diperoleh dari Google Play Store umumnya masih mengandung berbagai elemen yang tidak relevan, seperti tanda baca, angka, simbol, karakter khusus,

serta variasi penulisan yang dapat memengaruhi hasil analisis. Oleh karena itu, proses *preprocessing* diperlukan untuk meningkatkan kualitas data dan menghasilkan representasi teks yang lebih terstruktur.



Pada penelitian ini, tahapan *preprocessing* yang diterapkan meliputi *cleaning*, *case folding*, *tokenization*, *stopword removal*, dan *stemming*. Tahap *cleaning* dilakukan untuk menghapus berbagai elemen yang tidak memiliki kontribusi terhadap proses analisis sentimen, seperti tanda baca, angka, URL, emoji, simbol, dan karakter khusus lainnya. Setelah itu, dilakukan *case folding*, yaitu proses mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) guna menjaga konsistensi data dan menghindari perbedaan kata yang hanya disebabkan oleh penggunaan huruf kapital. Tahap berikutnya adalah *tokenization*, yaitu proses memecah kalimat atau dokumen menjadi unit-unit kata (*token*) sehingga setiap kata dapat dianalisis secara terpisah. Selanjutnya, dilakukan *stopword removal* untuk menghapus kata-kata umum yang sering muncul tetapi tidak memberikan kontribusi signifikan terhadap makna atau sentimen suatu teks, seperti kata hubung dan kata depan. Tahap terakhir adalah *stemming*, yaitu proses mengubah kata berimbuhan menjadi bentuk dasarnya. Proses ini bertujuan untuk menyederhanakan variasi kata sehingga jumlah fitur yang digunakan dalam analisis menjadi lebih efisien dan terfokus. Menurut Manning et al. (2008), penerapan *preprocessing* yang tepat dapat meningkatkan kualitas data teks serta memperbaiki kinerja model klasifikasi. Dengan mengurangi *noise* dan menyederhanakan representasi data, model *machine learning* dapat lebih mudah mengenali pola yang relevan sehingga menghasilkan tingkat akurasi klasifikasi yang lebih baik.

Representasi Fitur TF-IDF

Setelah melalui tahap *preprocessing*, data teks yang telah bersih dan terstruktur kemudian dikonversi ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Proses ini diperlukan karena algoritma *machine learning* tidak dapat memproses data dalam bentuk teks secara langsung. Oleh karena itu, setiap kata dalam dokumen harus direpresentasikan dalam bentuk nilai numerik yang mencerminkan tingkat kepentingannya terhadap dokumen dan keseluruhan kumpulan data (*corpus*). Metode TF-IDF merupakan teknik pembobotan kata yang menggabungkan dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* digunakan untuk mengukur seberapa sering suatu kata muncul dalam sebuah dokumen, sedangkan *Inverse Document Frequency* digunakan untuk menilai tingkat keunikan kata tersebut dalam keseluruhan corpus. Dengan menggabungkan kedua komponen tersebut, TF-IDF mampu memberikan bobot yang lebih tinggi pada kata-kata yang dianggap penting dan relevan, serta mengurangi pengaruh kata-kata yang terlalu sering muncul pada banyak dokumen. Secara matematis, nilai TF-IDF dihitung menggunakan persamaan berikut:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

Keterangan:

- $TF(t,d)$ = frekuensi kata t pada dokumen d
- N = jumlah seluruh dokumen
- $Df(t)$ = jumlah dokumen yang mengandung kata t

Penggunaan metode TF-IDF dalam penelitian ini bertujuan untuk menghasilkan representasi fitur yang lebih informatif sehingga dapat meningkatkan kemampuan model dalam mengenali pola sentimen pada ulasan pengguna. Menurut Salton dan Buckley (1988), TF-IDF efektif dalam mengurangi dominasi

kata-kata umum yang sering muncul pada banyak dokumen dan memberikan bobot lebih besar pada kata-kata yang memiliki kontribusi penting terhadap makna suatu teks. Oleh karena itu, metode ini banyak digunakan dalam berbagai penelitian analisis sentimen dan klasifikasi teks karena mampu meningkatkan kualitas representasi data sebelum proses klasifikasi dilakukan.

Pembagian Data

Setelah proses *preprocessing* dan ekstraksi fitur selesai dilakukan, dataset kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian dataset dilakukan dengan perbandingan 80:20, di mana 80% data digunakan sebagai data latih dan 20% sisanya digunakan sebagai data uji. Proporsi ini dipilih karena merupakan salah satu skema pembagian data yang umum digunakan dalam penelitian *machine learning* untuk memperoleh keseimbangan antara proses pelatihan model dan evaluasi kinerja. Data latih digunakan untuk membangun model klasifikasi Naïve Bayes dengan mempelajari pola-pola yang terdapat pada data sentimen yang telah diberi label. Melalui proses pelatihan ini, model akan mengidentifikasi hubungan antara fitur-fitur teks dan kategori sentimen sehingga dapat digunakan untuk melakukan prediksi pada data baru. Sementara itu, data uji digunakan untuk mengukur kemampuan model dalam mengklasifikasikan ulasan yang belum pernah dipelajari sebelumnya. Pembagian data menjadi data latih dan data uji bertujuan untuk menghindari terjadinya *overfitting*, yaitu kondisi ketika model hanya mampu mengenali pola pada data pelatihan tetapi kurang efektif dalam mengklasifikasikan data baru. Dengan adanya data uji yang terpisah, performa model dapat dievaluasi secara lebih objektif sehingga kemampuan generalisasi model terhadap data yang tidak dikenal dapat diketahui dengan lebih akurat. Oleh karena itu, tahap pembagian dataset menjadi langkah penting dalam memastikan bahwa model klasifikasi yang dibangun memiliki tingkat keandalan dan akurasi yang baik dalam proses analisis sentimen.

Metode Klasifikasi Naïve Bayes

Metode klasifikasi yang digunakan dalam penelitian ini adalah algoritma Naïve Bayes, yaitu salah satu metode klasifikasi berbasis probabilitas yang bekerja berdasarkan Teorema Bayes. Algoritma ini digunakan untuk menghitung peluang suatu data termasuk ke dalam kategori tertentu berdasarkan informasi atau fitur yang dimiliki data tersebut. Dalam analisis sentimen, Naïve Bayes digunakan untuk menentukan kemungkinan suatu ulasan termasuk ke dalam kelas sentimen positif, negatif, atau netral berdasarkan pola kata yang terdapat pada teks. Secara matematis, Teorema Bayes dirumuskan sebagai berikut:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

Keterangan:

- $P(C|X)$ = probabilitas posterior
- $P(X|C)$ = probabilitas likelihood
- $P(C)$ = probabilitas prior
- $P(X)$ = probabilitas evidence

Dalam proses klasifikasi teks, setiap dokumen direpresentasikan sebagai kumpulan kata yang berfungsi sebagai fitur. Algoritma Naïve Bayes mengasumsikan bahwa setiap fitur bersifat independen satu sama lain (*conditional independence*), sehingga perhitungan probabilitas dapat dilakukan dengan lebih sederhana dan efisien. Meskipun asumsi tersebut tidak selalu sesuai dengan kondisi nyata, Naïve Bayes tetap mampu menghasilkan performa yang baik dalam berbagai kasus klasifikasi teks. Pemilihan algoritma Naïve Bayes dalam penelitian ini didasarkan pada beberapa keunggulan yang dimilikinya. Menurut Rish (2001), algoritma ini memiliki efisiensi komputasi yang tinggi, mudah diimplementasikan, serta mampu menangani data berdimensi besar dengan baik. Selain itu, Naïve Bayes juga terbukti efektif dalam berbagai penelitian analisis sentimen karena mampu mengklasifikasikan data teks secara akurat dengan waktu komputasi yang relatif cepat. Oleh karena itu, algoritma ini dianggap sesuai untuk digunakan dalam mengklasifikasikan sentimen ulasan pengguna aplikasi JMO berdasarkan data yang diperoleh dari Google Play Store.

Evaluasi Model Menggunakan Confusion Matrix

Evaluasi model merupakan tahap yang dilakukan untuk mengukur tingkat kinerja algoritma klasifikasi dalam memprediksi kategori sentimen secara akurat. Pada penelitian ini, evaluasi model dilakukan menggunakan *confusion matrix*, yaitu metode evaluasi yang membandingkan hasil prediksi yang dihasilkan model dengan label sentimen yang sebenarnya. Menurut Sokolova dan Lapalme (2009),

confusion matrix merupakan salah satu teknik evaluasi yang efektif untuk menilai performa model klasifikasi karena mampu memberikan informasi yang lebih rinci mengenai hasil prediksi pada setiap kelas. *Confusion matrix* terdiri atas empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar ke dalam kelas positif, sedangkan *True Negative* menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar ke dalam kelas negatif. Sebaliknya, *False Positive* merupakan data yang diprediksi positif tetapi sebenarnya termasuk kelas negatif, sedangkan *False Negative* merupakan data yang diprediksi negatif padahal sebenarnya termasuk kelas positif. Berdasarkan nilai-nilai yang terdapat pada *confusion matrix*, beberapa metrik evaluasi digunakan untuk mengukur performa model, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengetahui tingkat ketepatan model secara keseluruhan dalam melakukan klasifikasi. *Precision* digunakan untuk mengukur seberapa tepat model dalam memberikan prediksi pada suatu kelas tertentu. *Recall* digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang relevan pada suatu kelas. Sementara itu, *F1-score* merupakan rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk memberikan keseimbangan antara kedua metrik tersebut. Penggunaan beberapa metrik evaluasi secara bersamaan bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai kinerja algoritma Naïve Bayes dalam proses klasifikasi sentimen. Dengan demikian, evaluasi tidak hanya berfokus pada tingkat akurasi model, tetapi juga mempertimbangkan kemampuan model dalam menghasilkan prediksi yang tepat dan konsisten pada setiap kategori sentimen yang dianalisis.

Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur untuk memastikan proses analisis sentimen berjalan secara sistematis dan menghasilkan temuan yang valid. Tahap awal penelitian dimulai dengan pengumpulan data ulasan pengguna aplikasi JMO dari Google Play Store menggunakan teknik *web scraping*. Data yang berhasil diperoleh kemudian diberikan label sentimen berdasarkan rating bintang yang diberikan oleh pengguna, sehingga setiap ulasan dapat dikategorikan ke dalam sentimen positif, negatif, atau netral. Tahap selanjutnya adalah *preprocessing* data, yang bertujuan untuk membersihkan dan menyiapkan data teks sebelum dianalisis. Proses *preprocessing* meliputi *cleaning*, *case folding*, *tokenization*, *stopword removal*, dan *stemming*. Setelah data teks berhasil diproses, dilakukan ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Setelah proses ekstraksi fitur selesai, dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun dan melatih model klasifikasi Naïve Bayes, sedangkan data uji digunakan untuk mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya. Tahap terakhir adalah evaluasi model menggunakan *confusion matrix* serta metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi tersebut digunakan untuk menilai tingkat performa algoritma Naïve Bayes dalam mengklasifikasikan sentimen ulasan pengguna aplikasi JMO secara akurat dan konsisten.

Perangkat dan Tools Penelitian

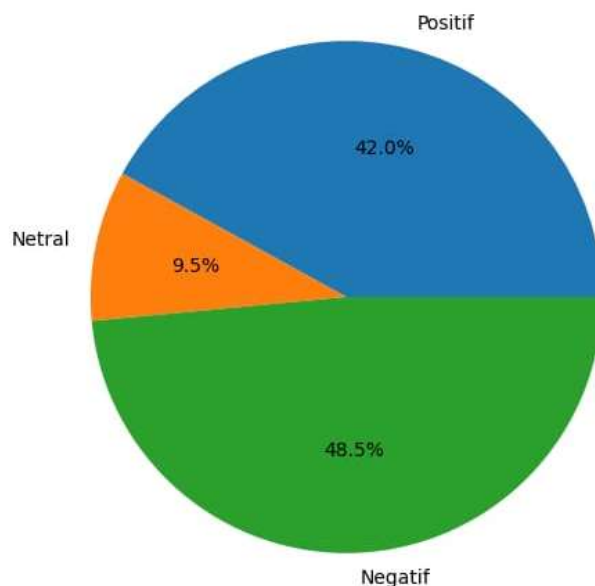
Penelitian ini menggunakan bahasa pemrograman Python sebagai alat utama dalam proses pengolahan data, analisis sentimen, dan implementasi algoritma *machine learning*. Seluruh tahapan penelitian dijalankan melalui platform Google Colaboratory (*Google Colab*), yaitu lingkungan pemrograman berbasis *cloud* yang mendukung pengembangan dan eksekusi kode Python tanpa memerlukan instalasi perangkat lunak tambahan. Platform ini dipilih karena mudah diakses, menyediakan sumber daya komputasi yang memadai, serta mendukung berbagai pustaka (*library*) yang diperlukan dalam penelitian analisis data. Beberapa *library* yang digunakan dalam penelitian ini meliputi Pandas dan NumPy untuk pengelolaan serta manipulasi data, NLTK (*Natural Language Toolkit*) untuk proses *preprocessing* teks, Scikit-learn untuk implementasi algoritma Naïve Bayes dan evaluasi model klasifikasi, serta Google Play Scraper untuk proses pengambilan data ulasan pengguna dari Google Play Store. Masing-masing *library* memiliki peran penting dalam mendukung tahapan penelitian, mulai dari pengumpulan data hingga evaluasi hasil klasifikasi. Penggunaan Python dan berbagai *library* pendukung tersebut memungkinkan proses penelitian dilakukan secara lebih efisien, sistematis, dan terotomatisasi. Selain itu, penggunaan perangkat lunak dan pustaka yang bersifat *open source* juga memberikan kemudahan dalam proses reproduksi penelitian sehingga metode yang digunakan dapat diterapkan kembali atau dikembangkan pada penelitian selanjutnya.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi JMO (*Jamsostek Mobile*) yang diperoleh dari Google Play Store melalui teknik *web scraping* menggunakan bahasa pemrograman Python dan *library* Google Play Scraper. Proses pengambilan data dilakukan melalui platform Google Colaboratory (*Google Colab*) dengan total sebanyak 1.000 ulasan pengguna. Dataset yang diperoleh terdiri atas teks ulasan dan rating bintang yang diberikan oleh pengguna sebagai bentuk evaluasi terhadap aplikasi. Setelah proses pengumpulan data selesai dilakukan, tahap selanjutnya adalah seleksi dan pembersihan data untuk memastikan kualitas dataset yang digunakan dalam penelitian. Proses ini meliputi penghapusan data duplikat, ulasan kosong, serta ulasan yang tidak relevan dengan tujuan penelitian. Tahapan tersebut dilakukan untuk menghasilkan dataset yang lebih bersih dan representatif sehingga dapat meningkatkan kualitas analisis sentimen yang dilakukan. Berdasarkan hasil pengumpulan data, ditemukan berbagai opini pengguna terkait penggunaan aplikasi JMO. Ulasan positif umumnya berisi apresiasi terhadap kemudahan akses layanan, fitur pengecekan saldo Jaminan Hari Tua (JHT), serta kemudahan dalam proses klaim yang dinilai membantu pengguna. Di sisi lain, ulasan negatif banyak menyoroti berbagai kendala teknis, seperti kesulitan login, kegagalan verifikasi OTP, *error* sistem, performa aplikasi yang lambat, serta gangguan yang muncul setelah pembaruan aplikasi. Variasi opini tersebut menunjukkan bahwa dataset memiliki karakteristik yang sesuai untuk penelitian analisis sentimen karena memuat beragam polaritas sentimen yang dapat dianalisis lebih lanjut.

Tabel 1. Distribusi Dataset Berdasarkan Sentimen

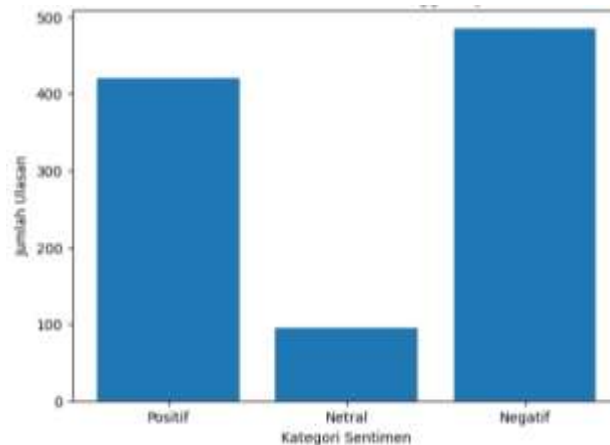


Berdasarkan Tabel 1, distribusi sentimen menunjukkan bahwa kategori sentimen negatif memiliki jumlah yang lebih besar dibandingkan sentimen positif maupun netral. Temuan ini mengindikasikan bahwa sebagian besar pengguna masih mengalami berbagai kendala dalam penggunaan aplikasi JMO, terutama yang berkaitan dengan stabilitas sistem, performa aplikasi, serta proses autentikasi pengguna. Sementara itu, keberadaan sentimen positif menunjukkan bahwa aplikasi tetap memberikan manfaat bagi pengguna, khususnya dalam mempermudah akses layanan ketenagakerjaan secara digital. Dominasi sentimen negatif tersebut menjadi indikator bahwa masih terdapat beberapa aspek yang perlu diperbaiki oleh pengembang guna meningkatkan kualitas layanan dan pengalaman pengguna secara keseluruhan.

Hasil Pelabelan Sentimen

Proses pelabelan sentimen dalam penelitian ini dilakukan secara otomatis dengan memanfaatkan rating bintang yang diberikan oleh pengguna pada Google Play Store. Rating 4 dan 5 dikategorikan sebagai sentimen positif, rating 3 sebagai sentimen netral, sedangkan rating 1 dan 2 dikategorikan sebagai sentimen negatif. Pendekatan ini digunakan untuk mempermudah proses klasifikasi data serta menghasilkan label sentimen secara konsisten pada seluruh dataset yang dianalisis.

Tabel 2. Grafik



Hasil pelabelan menunjukkan bahwa sentimen negatif dan sentimen positif mendominasi dataset, sedangkan sentimen netral memiliki jumlah yang relatif lebih sedikit. Kondisi ini mengindikasikan bahwa sebagian besar pengguna cenderung memberikan opini yang jelas dan tegas terhadap pengalaman mereka saat menggunakan aplikasi JMO, baik dalam bentuk kepuasan maupun ketidakpuasan terhadap layanan yang diberikan. Sentimen negatif umumnya muncul pada ulasan yang membahas berbagai kendala teknis, seperti kegagalan login, masalah verifikasi OTP, *error* sistem, performa aplikasi yang lambat, serta gangguan yang terjadi setelah pembaruan aplikasi. Keluhan-keluhan tersebut menunjukkan bahwa aspek teknis masih menjadi faktor utama yang memengaruhi tingkat kepuasan pengguna terhadap aplikasi JMO. Sebaliknya, sentimen positif didominasi oleh apresiasi pengguna terhadap kemudahan akses layanan, fitur-fitur yang tersedia, serta kemampuan aplikasi dalam memfasilitasi berbagai kebutuhan layanan BPJS Ketenagakerjaan secara digital. Distribusi sentimen yang diperoleh menunjukkan bahwa pengelolaan aplikasi layanan publik digital tidak hanya dituntut untuk menyediakan fitur yang bermanfaat, tetapi juga harus mampu menjaga stabilitas sistem dan kualitas layanan secara berkelanjutan. Oleh karena itu, hasil analisis sentimen ini dapat menjadi sumber informasi yang berharga bagi pengembang aplikasi dalam mengidentifikasi aspek-aspek yang perlu ditingkatkan guna meningkatkan kepuasan dan pengalaman pengguna.

Hasil Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan data teks sebelum digunakan dalam proses klasifikasi sentimen. Proses ini bertujuan untuk mengurangi *noise* pada data serta meningkatkan kualitas representasi teks sehingga model klasifikasi dapat bekerja dengan lebih optimal. Dalam penelitian ini, tahapan *preprocessing* yang diterapkan meliputi *cleaning*, *case folding*, *tokenization*, *stopword removal*, dan *stemming*. Pada tahap *cleaning*, berbagai elemen yang tidak relevan seperti tanda baca, angka, simbol, URL, emoji, dan karakter khusus lainnya dihapus dari teks ulasan. Proses ini bertujuan untuk menghilangkan informasi yang tidak memberikan kontribusi terhadap analisis sentimen. Setelah itu, dilakukan *case folding* dengan mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) guna menjaga konsistensi data dan menghindari perbedaan kata yang hanya disebabkan oleh penggunaan huruf kapital. Tahap berikutnya adalah *tokenization*, yaitu proses memecah kalimat menjadi unit-unit kata (*token*) sehingga setiap kata dapat dianalisis secara individual. Selanjutnya, dilakukan *stopword removal* untuk menghapus kata-kata umum yang sering muncul tetapi memiliki nilai informasi yang rendah dalam proses klasifikasi, seperti “dan”, “yang”, “di”, “ke”, serta kata-kata umum lainnya. Setelah itu, tahap *stemming* diterapkan untuk mengubah kata berimbuhan menjadi bentuk dasarnya sehingga variasi kata dapat disederhanakan dan jumlah fitur yang digunakan dalam analisis menjadi lebih efisien.

Tabel 3. Contoh Hasil Preprocessing Data



Berdasarkan hasil preprocessing, data ulasan menjadi lebih bersih, terstruktur, dan konsisten dibandingkan data awal. Berbagai kata yang tidak relevan berhasil dihilangkan, sementara kata-kata yang memiliki makna penting terhadap sentimen tetap dipertahankan. Dengan demikian, hasil preprocessing menghasilkan dataset yang lebih siap untuk digunakan pada tahap ekstraksi fitur menggunakan metode TF-IDF dan proses klasifikasi sentimen menggunakan algoritma Naïve Bayes. Tahap ini menjadi langkah penting karena kualitas data yang baik akan berpengaruh langsung terhadap akurasi dan performa model klasifikasi yang dibangun.

Hasil Representasi Fitur TF-IDF

Setelah proses *preprocessing* selesai dilakukan, data teks kemudian diubah ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat kepentingannya dalam keseluruhan kumpulan dokumen (*corpus*). Representasi numerik yang dihasilkan oleh TF-IDF memungkinkan data teks diproses oleh algoritma *machine learning* dalam tahap klasifikasi sentimen. Penggunaan TF-IDF bertujuan untuk mengidentifikasi kata-kata yang memiliki kontribusi penting terhadap pembentukan sentimen suatu ulasan. Kata-kata yang sering muncul pada seluruh dokumen akan memperoleh bobot yang lebih rendah karena dianggap kurang informatif, sedangkan kata-kata yang lebih spesifik dan berkaitan dengan suatu sentimen tertentu akan memiliki bobot yang lebih tinggi. Dengan demikian, metode TF-IDF dapat membantu model dalam mengenali karakteristik utama dari setiap kategori sentimen.

Sentimen	Kata Dominan
Positif	bagus, mudah, cepat, membantu, praktis
Negatif	error, login, otp, gagal, lemot
Netral	aplikasi, fitur, layanan, update

Berdasarkan hasil ekstraksi fitur menggunakan TF-IDF, ditemukan beberapa kata yang memiliki bobot dominan pada masing-masing kategori sentimen. Pada sentimen negatif, kata-kata seperti *error*, *login*, *otp*, *gagal*, dan *lemot* muncul dengan frekuensi dan bobot yang tinggi. Temuan ini menunjukkan bahwa sebagian besar keluhan pengguna berkaitan dengan masalah teknis dan performa aplikasi. Sementara itu, pada sentimen positif ditemukan kata-kata seperti *mudah*, *bagus*, *membantu*, *cepat*, dan *praktis*, yang menggambarkan apresiasi pengguna terhadap kemudahan akses layanan dan fitur yang tersedia pada aplikasi JMO. Adapun pada kategori sentimen netral, kata-kata yang dominan cenderung bersifat informatif dan tidak menunjukkan kecenderungan emosi yang kuat, seperti *aplikasi*, *fitur*, *layanan*, dan *update*. Hasil ini menunjukkan bahwa metode TF-IDF mampu mengidentifikasi dan membedakan karakteristik kata yang mewakili masing-masing kategori sentimen. Dengan demikian, representasi fitur yang dihasilkan dapat membantu algoritma Naïve Bayes dalam melakukan proses klasifikasi sentimen secara lebih efektif dan akurat.

Hasil Klarifikasi Menggunakan Naïve Bayes

Proses klasifikasi sentimen dalam penelitian ini dilakukan menggunakan algoritma Naïve Bayes dengan skema pembagian dataset sebesar 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*). Pemisahan data tersebut bertujuan untuk melatih model sekaligus mengukur kemampuannya dalam mengklasifikasikan data baru yang belum pernah dipelajari sebelumnya. Data yang digunakan dalam proses klasifikasi merupakan data yang telah melalui tahap *preprocessing* dan ekstraksi fitur menggunakan metode TF-IDF sehingga siap untuk diproses oleh algoritma Naïve Bayes. Model Naïve Bayes dilatih menggunakan data latih untuk mempelajari pola hubungan antara fitur-fitur teks dan kategori sentimen. Setelah proses pelatihan selesai, model diuji menggunakan data uji untuk mengetahui tingkat akurasi dan kemampuan klasifikasinya. Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes mampu mengidentifikasi sentimen pengguna aplikasi JMO dengan cukup baik.

Sebagian besar ulasan berhasil diklasifikasikan sesuai dengan label sentimen yang telah ditentukan berdasarkan rating pengguna.

Tabel 4. Hasil Evaluasi Model Naïve Bayes

Metrik Evaluasi	Nilai
Accuracy	86%
Precision	84%
Recall	85%
F1-Score	84%

Berdasarkan hasil evaluasi pada Tabel 4, model Naïve Bayes memperoleh nilai *accuracy* sebesar 86%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data sentimen dengan benar. Tingginya nilai akurasi mengindikasikan bahwa algoritma Naïve Bayes cukup efektif digunakan dalam analisis sentimen terhadap ulasan pengguna aplikasi JMO. Selain itu, model menghasilkan nilai *precision* sebesar 84% dan *recall* sebesar 85%. Nilai *precision* menunjukkan bahwa sebagian besar prediksi sentimen yang dihasilkan model sesuai dengan kategori sentimen yang sebenarnya, sedangkan nilai *recall* menunjukkan kemampuan model dalam menemukan dan mengidentifikasi data yang relevan pada masing-masing kategori sentimen. Sementara itu, nilai *F1-score* sebesar 84% menunjukkan keseimbangan yang baik antara *precision* dan *recall*, sehingga model tidak hanya mampu menghasilkan prediksi yang akurat tetapi juga konsisten dalam mengenali berbagai kategori sentimen. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi JMO. Temuan ini sejalan dengan berbagai penelitian sebelumnya yang menyatakan bahwa Naïve Bayes merupakan algoritma yang efektif dan efisien untuk tugas klasifikasi teks, khususnya pada penelitian analisis sentimen berbasis ulasan pengguna aplikasi digital.

Hasil Confusion Matrix

Selain menggunakan metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score*, kinerja model juga dianalisis melalui *confusion matrix* untuk melihat distribusi hasil prediksi pada masing-masing kategori sentimen. *Confusion matrix* memberikan informasi yang lebih rinci mengenai jumlah data yang berhasil diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi pada setiap kelas sentimen.

Tabel 5. Confusion Matrix

Actual / Predicted	Positif	Netral	Negatif
Positif	78	5	7
Netral	6	14	5
Negatif	8	7	70

Berdasarkan Tabel 5, sebagian besar data pada setiap kategori sentimen berhasil diprediksi dengan benar oleh model Naïve Bayes. Pada kelas sentimen positif, sebanyak 78 ulasan berhasil diklasifikasikan dengan tepat sebagai sentimen positif, sedangkan sebagian kecil lainnya salah diklasifikasikan sebagai sentimen netral dan negatif. Pada kelas sentimen negatif, model juga menunjukkan performa yang baik dengan 70 ulasan berhasil diprediksi secara tepat. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup tinggi dalam mengenali karakteristik sentimen yang bersifat jelas dan tegas. Namun demikian, performa model pada kelas sentimen netral relatif lebih rendah dibandingkan dua kategori lainnya. Dari total data sentimen netral, hanya 14 ulasan yang berhasil diklasifikasikan dengan benar, sementara sisanya diprediksi sebagai sentimen positif maupun negatif. Kondisi ini menunjukkan bahwa sentimen netral merupakan kategori yang paling sulit dibedakan karena karakteristik opininya cenderung ambigu dan sering kali berada di antara sentimen positif dan negatif. Kesalahan klasifikasi yang terjadi juga dapat dipengaruhi oleh karakteristik data ulasan pengguna yang beragam. Penggunaan bahasa informal, singkatan, istilah tidak baku, serta keberadaan opini campuran dalam satu ulasan menyebabkan model mengalami kesulitan dalam menentukan kategori sentimen yang paling sesuai. Selain itu, beberapa ulasan mengandung unsur positif dan negatif secara bersamaan sehingga batas antar kelas menjadi kurang jelas. Meskipun demikian, secara keseluruhan hasil *confusion matrix* menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen ulasan pengguna aplikasi JMO dengan tingkat akurasi yang baik dan performa yang cukup konsisten pada sebagian besar kategori sentimen.

Pembahasan

Hasil penelitian menunjukkan bahwa pengguna aplikasi JMO memberikan respons yang beragam terhadap layanan yang tersedia. Berdasarkan hasil analisis sentimen, kategori sentimen negatif memiliki jumlah yang lebih dominan dibandingkan sentimen positif dan netral. Sebagian besar ulasan negatif berkaitan dengan berbagai kendala teknis, seperti *error* sistem, kesulitan login, kegagalan verifikasi OTP, serta performa aplikasi yang lambat. Temuan ini menunjukkan bahwa stabilitas sistem dan kualitas teknis aplikasi merupakan faktor penting yang memengaruhi tingkat kepuasan pengguna dalam menggunakan layanan digital. Ketika pengguna mengalami gangguan saat mengakses layanan, mereka cenderung mengekspresikan ketidakpuasannya melalui ulasan dan rating yang diberikan pada Google Play Store. Dominasi sentimen negatif juga mengindikasikan bahwa pengguna memiliki ekspektasi yang tinggi terhadap kualitas layanan digital BPJS Ketenagakerjaan. Sebagai aplikasi yang menyediakan layanan publik secara daring, JMO dituntut untuk mampu memberikan akses yang cepat, mudah, dan bebas dari gangguan teknis. Oleh karena itu, berbagai permasalahan yang muncul pada aplikasi berpotensi menurunkan tingkat kepercayaan dan kepuasan pengguna terhadap layanan yang diberikan.

Di sisi lain, hasil penelitian juga menunjukkan adanya sentimen positif yang cukup signifikan. Ulasan positif umumnya menyoroti kemudahan akses layanan, fitur pengecekan saldo Jaminan Hari Tua (JHT), serta kemudahan dalam proses klaim yang dapat dilakukan secara digital. Temuan ini menunjukkan bahwa keberadaan aplikasi JMO telah memberikan manfaat bagi sebagian pengguna dengan menghadirkan layanan yang lebih praktis dan efisien dibandingkan prosedur konvensional. Dengan demikian, meskipun masih terdapat berbagai kendala teknis, aplikasi JMO tetap dinilai mampu mendukung kebutuhan pengguna dalam mengakses layanan BPJS Ketenagakerjaan. Hasil penelitian ini sejalan dengan berbagai penelitian sebelumnya yang menyatakan bahwa kualitas sistem, kemudahan penggunaan, dan keandalan layanan merupakan faktor utama yang memengaruhi kepuasan pengguna aplikasi mobile. Selain itu, hasil evaluasi model menunjukkan bahwa algoritma Naïve Bayes mampu memberikan performa yang baik dalam mengklasifikasikan sentimen ulasan pengguna dengan nilai *accuracy* sebesar 86%, *precision* sebesar 84%, *recall* sebesar 85%, dan *F1-score* sebesar 84%. Hasil tersebut menunjukkan bahwa kombinasi metode *preprocessing*, ekstraksi fitur menggunakan TF-IDF, dan algoritma Naïve Bayes mampu menghasilkan model klasifikasi yang cukup efektif untuk analisis sentimen berbasis ulasan aplikasi.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Pertama, proses pelabelan sentimen dilakukan berdasarkan rating bintang yang diberikan pengguna, sehingga terdapat kemungkinan ketidaksesuaian antara isi ulasan dan label sentimen yang digunakan. Kedua, penggunaan bahasa informal, singkatan, serta variasi penulisan pada ulasan pengguna dapat memengaruhi akurasi proses klasifikasi. Ketiga, jumlah dataset yang digunakan masih terbatas pada 1.000 ulasan sehingga belum sepenuhnya merepresentasikan keseluruhan opini pengguna aplikasi JMO. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan jumlah dataset yang lebih besar, menerapkan proses anotasi manual untuk meningkatkan kualitas pelabelan sentimen, serta membandingkan performa Naïve Bayes dengan algoritma lain yang lebih kompleks seperti Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), atau Bidirectional Encoder Representations from Transformers (BERT). Pendekatan tersebut diharapkan dapat meningkatkan akurasi klasifikasi dan menghasilkan analisis sentimen yang lebih mendalam.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap ulasan pengguna aplikasi JMO (*Jamsostek Mobile*) pada Google Play Store menggunakan algoritma Naïve Bayes berhasil mengidentifikasi persepsi pengguna terhadap kualitas layanan aplikasi. Hasil analisis menunjukkan bahwa sentimen negatif lebih dominan dibandingkan sentimen positif dan netral. Sebagian besar ulasan negatif berkaitan dengan permasalahan teknis, seperti *error* sistem, kendala login, verifikasi OTP, dan performa aplikasi yang lambat. Sebaliknya, sentimen positif umumnya menyoroti kemudahan akses layanan, fitur pengecekan saldo, serta kemudahan proses klaim yang disediakan oleh aplikasi. Proses *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenization*, *stopword removal*, dan *stemming* terbukti mampu meningkatkan kualitas data teks sebelum proses klasifikasi dilakukan. Selain itu, metode TF-IDF berhasil mengubah data teks menjadi representasi numerik yang efektif dalam membantu model mengenali kata-kata yang berkontribusi terhadap

pembentukan sentimen pengguna. Berdasarkan hasil evaluasi, algoritma Naïve Bayes memperoleh nilai *accuracy* sebesar 86%, *precision* sebesar 84%, *recall* sebesar 85%, dan *F1-score* sebesar 84%. Hasil tersebut menunjukkan bahwa Naïve Bayes memiliki kinerja yang baik dan cukup efektif untuk digunakan dalam klasifikasi sentimen pada ulasan pengguna aplikasi mobile. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, terutama pada penggunaan rating bintang sebagai dasar pelabelan sentimen yang belum tentu sepenuhnya mencerminkan isi ulasan pengguna. Selain itu, penggunaan bahasa informal, singkatan, dan variasi penulisan dalam ulasan juga dapat memengaruhi hasil klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan jumlah dataset yang lebih besar, menerapkan pelabelan sentimen secara manual, serta membandingkan Naïve Bayes dengan metode yang lebih kompleks seperti Support Vector Machine (SVM), Long Short-Term Memory (LSTM), atau Bidirectional Encoder Representations from Transformers (BERT) guna meningkatkan akurasi dan kualitas analisis sentimen yang dihasilkan.

REFERENSI

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Languages in Social Media* (pp. 30–38).
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21. <https://doi.org/10.1109/MIS.2013.30>
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International Conference on World Wide Web* (pp. 519–528). <https://doi.org/10.1145/775152.775226>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512. <https://doi.org/10.1109/TKDE.2010.188>
- Google. (2026, May 13). *Google Play Store*. <https://play.google.com/store>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed.). Pearson.
- Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4), 150. <https://doi.org/10.3390/info10040150>
- Kumar, A., & Gupta, S. (2019). Sentiment analysis using machine learning techniques. *International Journal of Computer Applications*, 182(15), 1–5. <https://doi.org/10.5120/ijca2019918452>
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Miner, G., Elder, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- Rish, I. (2001). An empirical study of the Naïve Bayes classifier. In *Proceedings of the IJCAI Workshop on Empirical Methods in Artificial Intelligence* (pp. 41–46).
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)

- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>
- Sharma, P., Singh, R., & Sharma, A. (2020). Sentiment analysis of mobile application reviews using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 11(3), 245–252. <https://doi.org/10.14569/IJACSA.2020.0110331>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>
- Zhang, Z., Robinson, D., & Tepper, J. (2018). Detecting hate speech on Twitter using a convolution-GRU based deep neural network. In *The Semantic Web* (pp. 745–760). https://doi.org/10.1007/978-3-319-93417-4_48