

KINERJA MULTINOMIAL NAÏVE BAYES PADA ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MOBILE JKN

Rahmalia Putri ¹, Nining Rahaningsih ², Irfan Ali ³, Willy Prihartono⁴

Teknik Informatika STMIK Ikmi Cirebon

Email: rhmalia Putri30@gmail.com

ABSTRACT

The Mobile JKN application is a key digital service provided by BPJS Kesehatan to support access to national health insurance. User reviews on Google Play Store contain rich information about satisfaction and technical issues, but these data are unstructured and difficult to interpret at scale. This study aims to evaluate the performance of a Multinomial Naïve Bayes model for classifying sentiment in Indonesian-language reviews of the Mobile JKN application. Approximately 10,000 recent reviews were collected via web scraping using the google-play-scraper library and processed through several text preprocessing stages, including cleaning, case folding, tokenization, stopword removal, and stemming. The sentiment labels (positive, negative, neutral) were automatically derived from the rating scores using a distant supervision approach. Text features were represented using TF-IDF and used to train and test a Multinomial Naïve Bayes classifier. Model performance was evaluated using accuracy, precision, recall, and F1-score, complemented by a confusion matrix, sentiment distribution visualization, and wordclouds for each sentiment class. The results show that the model achieves good overall accuracy and performs particularly well in identifying positive and negative sentiments, while the neutral class remains more challenging due to shorter review length and semantic overlap with the other two classes. Sentiment distribution indicates that negative reviews still dominate, highlighting persistent technical issues such as login failures, verification problems, and application errors. These findings demonstrate that Naïve Bayes combined with TF-IDF is effective for large-scale sentiment analysis of user feedback on public service applications.

Keywords: sentiment analysis, Naïve Bayes, Mobile JKN, TF-IDF, user reviews

Riwayat Artikel :

Tanggal diterima : 27-11-2025

Tanggal revisi : 04-12-2025

Tanggal terbit : 08-12-2025

DOI :

<https://doi.org/10.31949/j-ensitec.v12i01.16656>

1. PENDAHULUAN

Transformasi digital di sektor kesehatan mendorong lembaga publik seperti BPJS Kesehatan untuk menyediakan layanan berbasis aplikasi, salah satunya Mobile JKN yang dirancang untuk mempermudah peserta dalam mengakses informasi kepesertaan,

pendaftaran fasilitas kesehatan, hingga pengurusan klaim secara daring. Sejumlah kajian sebelumnya telah menelaah program Jaminan Kesehatan Nasional (JKN) maupun penggunaan aplikasi Mobile JKN dari sisi kualitas layanan, kepuasan, literasi kesehatan,

This is an open access article under the CC BY-4.0 license.



dan aksesibilitas layanan kesehatan di berbagai wilayah Indonesia [1][2].

Keberhasilan layanan digital tersebut sangat bergantung pada pengalaman dan kepuasan pengguna, yang banyak tercermin dalam ulasan pada platform distribusi aplikasi seperti Google Play Store. Ulasan pengguna Mobile JKN umumnya berbentuk teks bebas dan nonformal, banyak mengandung singkatan, emotikon, serta variasi gaya bahasa yang tinggi, sehingga informasi sentimen (positif, negatif, netral) sulit dianalisis secara manual ketika jumlah ulasan mencapai ribuan. Pendekatan analisis sentimen dengan teknik machine learning diperlukan agar opini pengguna dapat diolah secara sistematis dan berskala besar [3],[4][5].

Algoritma Multinomial Naïve Bayes (MNB) merupakan salah satu metode yang banyak digunakan dalam analisis teks karena sederhana, cepat, dan cukup andal pada data berukuran besar dengan banyak fitur kata. MNB memodelkan probabilitas kelas berdasarkan frekuensi kemunculan kata dalam dokumen dan secara luas diadopsi pada tugas klasifikasi sentimen untuk ulasan aplikasi, e-commerce, dan layanan publik digital [3],[6], [7].

Beberapa penelitian terdahulu menunjukkan bahwa Naïve Bayes mampu menghasilkan akurasi tinggi pada analisis sentimen ulasan aplikasi transportasi, marketplace, maupun layanan publik digital lainnya [8], [9],[10][11],[12]. Namun, sebagian besar pekerjaan sebelumnya berfokus pada domain e-commerce atau aplikasi transportasi dan belum secara spesifik memanfaatkan korpus ulasan Mobile JKN berbahasa Indonesia dalam jumlah besar yang diambil langsung dari Google Play Store. Di sisi lain, penelitian terkait Mobile JKN yang ada cenderung menitikberatkan pada aspek penerimaan pengguna, keberhasilan sistem, maupun literatur kualitatif tentang mutu layanan tanpa mengkaji kinerja kuantitatif model machine learning untuk analisis sentimen ulasan pengguna [1]–[2],[13]–[13].

Berdasarkan celah tersebut, penelitian ini mengusulkan evaluasi komprehensif terhadap algoritma Multinomial Naïve Bayes pada korpus ulasan Mobile JKN yang diperoleh dari Google Play Store. Skema pelabelan sentimen

dilakukan secara otomatis berbasis rating (distant supervision), dan kinerja model dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, serta analisis confusion matrix dan distribusi sentimen. Dengan demikian, kontribusi utama penelitian ini adalah memberikan gambaran empiris tentang sejauh mana Naïve Bayes efektif dalam mengklasifikasikan sentimen ulasan Mobile JKN dan bagaimana distribusi sentimen tersebut dapat digunakan sebagai masukan perbaikan layanan.

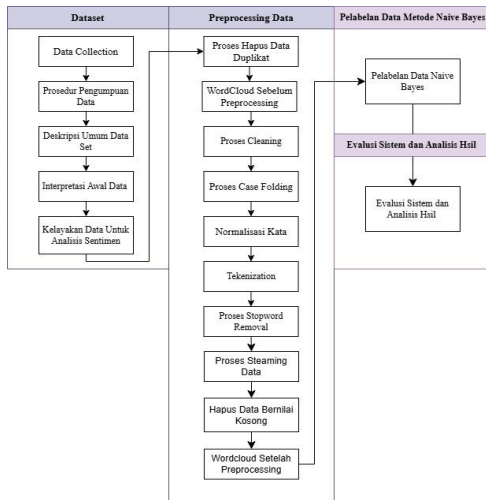
Rumusan masalah tunggal yang diangkat dalam artikel ini adalah: “Bagaimana performa algoritma Multinomial Naïve Bayes dalam mengklasifikasikan sentimen (positif, negatif, netral) pada ulasan pengguna Mobile JKN di Google Play Store, ditinjau dari metrik akurasi, precision, recall, dan F1-score serta distribusi sentimen yang dihasilkan?”

Tujuan penelitian adalah mengevaluasi kinerja model Multinomial Naïve Bayes untuk analisis sentimen ulasan Mobile JKN dan menginterpretasikan distribusi sentimen yang terbentuk sebagai dasar pemahaman kualitas layanan dari perspektif pengguna.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan rancangan eksperimen penerapan algoritma Multinomial Naïve Bayes pada dataset ulasan aplikasi Mobile JKN. Tahapan utama meliputi: pengumpulan data, preprocessing teks, pelabelan sentimen berbasis rating, ekstraksi fitur dengan TF-IDF, pelatihan dan pengujian model, serta evaluasi performa.

Tempat dan objek penelitian adalah ulasan publik aplikasi “Mobile JKN” (paket go.bpjs.mobile) pada Google Play Store dengan bahasa Indonesia.



Gambar 1. Tahapan penelitian

1. Pengumpulan Data

Data dikumpulkan melalui web scraping pada halaman aplikasi Mobile JKN di Google Play Store menggunakan library google-play-scraper pada Python. Fungsi reviews() digunakan dengan parameter bahasa Indonesia (lang='id'), negara Indonesia (country='id'), urutan ulasan terbaru, dan jumlah maksimum ±10.000 ulasan. Hasil scraping dikonversi ke dalam DataFrame pandas dan disimpan dalam format CSV.

Setelah proses pembersihan awal (penghapusan duplikasi berdasarkan reviewId, pembuangan nilai kosong pada kolom teks, dan penghapusan karakter non-standar seperti emoji dan simbol HTML), diperoleh sekitar 9.842 ulasan valid yang siap untuk tahap preprocessing.

Dataset memuat beberapa atribut penting: reviewId (ID unik), userName, score (rating 1–5), content (isi ulasan), at (tanggal ulasan), dan thumbsUpCount (jumlah suka). Atribut teks content dan rating score menjadi fokus utama untuk proses analisis sentimen.

Tabel 1. Deskripsi Umum Dataset Ulasan Mobile Jkn

Atribut	Keteranga	Contoh Nilai
reviewId	ID unik	gp:AOqpTO...
userName	Nama	“Andi Saputra”
score	Rating	1, 2, 3, 4, 5
content	Isi ulasan	“Aplikasi sering
at	Tanggal	2025-03-14
thumbsUpCount	Jumlah	4

2. Preprocessing Teks

Ulasan yang telah terkumpul menjalani serangkaian tahap preprocessing untuk mengurangi noise dan menstandarkan bentuk teks, meliputi:

- Cleaning**
 Menghapus duplikasi ulasan berdasarkan content.
 Menghapus entri dengan teks kosong.
 Menghilangkan angka, simbol, URL, dan karakter non-alfabet.
- Case Folding**
 Mengubah seluruh teks menjadi huruf kecil (lowercase) agar representasi kata konsisten.
- Tokenization**
 Memecah kalimat ulasan menjadi token kata menggunakan pustaka pemrosesan bahasa alami.
- Stopword Removal**
 Menghapus kata-kata umum berfrekuensi tinggi yang tidak memberikan kontribusi berarti pada sentimen (misalnya “yang”, “dan”, “di”).
- Stemming**
 Mengubah kata ke bentuk dasar menggunakan stemmer bahasa Indonesia, sehingga variasi bentuk kata seperti “membantu”, “membantunya” direduksi menjadi bentuk dasar yang sama [14].

Wordcloud awal dibuat sebelum preprocessing untuk menggambarkan kata-kata mentah yang masih mengandung noise. Setelah preprocessing, wordcloud kedua menampilkan kata-kata yang lebih bersih dan informatif sebagai representasi distribusi kata penting pada korpus ulasan.



Gambar 2. Wordcloud Ulasan Mobile Jkn Sebelum Dan Sesudah Preprocessing

3. Pelabelan Sentimen

Pelabelan sentimen dilakukan secara otomatis (distant supervision) dengan memanfaatkan nilai rating sebagai indikator tidak langsung dari polaritas sentimen. Skor tinggi diasosiasikan dengan sentimen positif, skor rendah dengan sentimen negatif, sedangkan skor menengah merepresentasikan kecenderungan netral. Pendekatan ini mengeliminasi kebutuhan anotasi manual yang memakan waktu dan rawan ketidakkonsistenan antar-pelabel, sekaligus selaras dengan praktik pada penelitian analisis ulasan aplikasi dan layanan publik lainnya [7].

Setiap ulasan dipasangkan dengan kelas sentimen (positif, negatif, netral) sehingga terbentuk ground truth yang diperlukan untuk melatih model supervised learning seperti Naïve Bayes.

4. Ekstraksi Fitur dan Pemodelan

Teks ulasan yang sudah diproses diubah menjadi representasi numerik menggunakan skema Term Frequency–Inverse Document Frequency (TF–IDF). Representasi ini menekankan kata-kata yang penting dalam suatu dokumen tetapi jarang muncul di keseluruhan korpus.

Algoritma yang digunakan adalah Multinomial Naïve Bayes, yang secara khusus cocok untuk data diskret berbasis frekuensi kata. Metode ini menghitung probabilitas setiap kelas sentimen berdasarkan distribusi kata dalam ulasan, dengan asumsi independensi antar-fitur yang dalam praktiknya terbukti cukup efektif untuk tugas text mining berskala besar [3], [6].

Dataset kemudian dibagi menjadi data latih dan uji menggunakan skema hold-out, dan model MNB dilatih pada vektor TF–IDF data latih.

5. Evaluasi Model

Kinerja model dievaluasi menggunakan empat metrik utama: accuracy, precision, recall, dan F1-score. Accuracy mengukur persentase prediksi yang benar; precision dan recall mengevaluasi ketepatan serta kelengkapan deteksi masing-masing kelas; sedangkan F1-score memberikan rata-rata harmonik antara precision dan recall [4], [5].

Selain itu, digunakan pula:

Confusion matrix untuk memetakan prediksi benar/salah pada kelas positif, negatif, dan netral.

Visualisasi distribusi sentimen hasil prediksi model.

Wordcloud per kelas sentimen untuk menonjolkan kata-kata yang dominan pada masing-masing kategori.

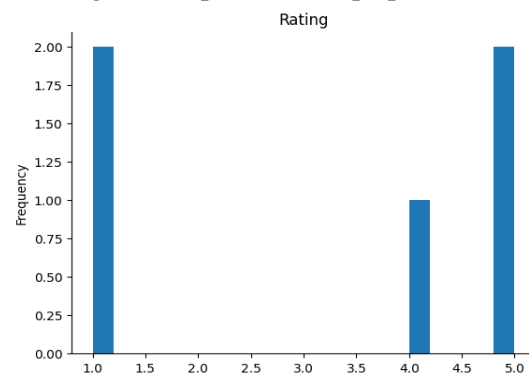
Tabel 2. Skema Evaluasi Kinerja Model Naïve Bayes

No	Kondisi Sebelum	Kondisi Setelah
1	“Aplikasi JKN bagus	“aplikasi jkn bagus
2	“Sudah coba login 3x	“sudah coba login
3	“Lemooooottt bgt	“lemot banget
4	“@bpjskesehatan	“tolong perbaiki
5	“Pelayanannya bagus,	“pelayanannya bagus

3. HASIL DAN PEMBAHASAN

Gambaran Data dan Distribusi Rating

Hasil scraping menghasilkan ribuan ulasan pengguna Mobile JKN dengan atribut rating, teks ulasan, dan informasi waktu. Visualisasi distribusi rating menunjukkan bahwa skor 1 dan 5 mendominasi, sedangkan skor menengah (2, 3, 4) relatif lebih sedikit. Pola ini menandakan kecenderungan penilaian yang ekstrem—pengguna cenderung sangat puas atau sangat tidak puas terhadap aplikasi.

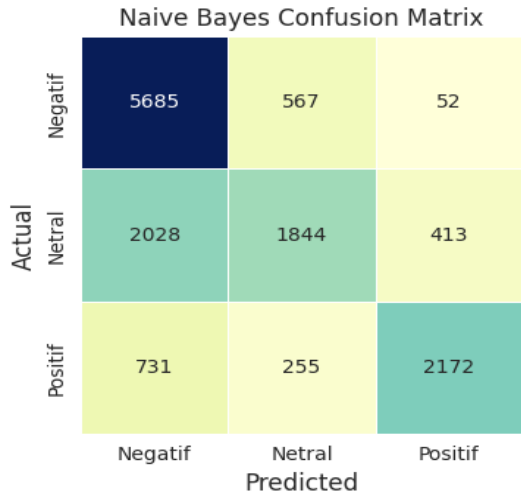


Gambar 3. Distribusi Rating Ulasan Mobile Jkn Di Google Play Store

Interpretasi awal menunjukkan bahwa rating 1 menjadi kategori dengan jumlah ulasan terbanyak, merefleksikan dominasi persepsi negatif yang umumnya berkaitan dengan gangguan teknis seperti gagal login, aplikasi crash, dan instabilitas server. Sebaliknya, rating tinggi (4–5) menggambarkan apresiasi terhadap kemudahan akses layanan tanpa perlu datang ke kantor BPJS serta peningkatan kinerja setelah pembaruan aplikasi.

Performa Model Berdasarkan Confusion Matrix

Confusion matrix digunakan untuk menilai pola prediksi model terhadap tiga kelas sentimen.



Gambar 4. Confusion Matrix Hasil Klasifikasi Sentiment Ulasan Mobile Jkn

Berdasarkan interpretasi pada confusion matrix:

Kelas positif menunjukkan jumlah prediksi benar paling tinggi. Hal ini menunjukkan bahwa model cukup baik dalam mengenali kosakata yang mengekspresikan kepuasan, seperti “bagus”, “membantu”, “cepat”, dan “mudah”.

Kelas negatif juga terdeteksi dengan akurasi yang baik; model mampu menangkap pola kata bernada keluhan seperti “error”, “tidak bisa”, “gagal”, atau “login” yang sering muncul dalam ulasan yang menyoroti kendala teknis.

Kelas netral memiliki tingkat misklasifikasi tertinggi, banyak ulasan netral yang salah diklasifikasikan sebagai positif atau negatif. Hal ini disebabkan oleh ulasan yang sangat singkat (“cukup”, “lumayan”) dan kemiripan struktur bahasa dengan dua kelas lain, diperparah oleh jumlah data netral yang lebih sedikit sehingga representasi fiturnya kurang kuat.

Temuan ini mengindikasikan bahwa Multinomial Naïve Bayes sangat efektif untuk membedakan ekspresi puas dan tidak puas yang eksplisit, namun lebih terbatas ketika harus menangani ekspresi ambivalen atau

netral, sejalan dengan laporan beberapa studi sejenis [8], [15],[10],[11].

Hasil Classification Report

```

***
Classification Report for Naive Bayes:

```

	precision	recall	f1-score	support
Negatif	0.673	0.902	0.771	6304.000
Netral	0.692	0.430	0.531	4285.000
Positif	0.824	0.688	0.750	3158.000
accuracy	0.706	0.706	0.706	0.706
macro avg	0.730	0.673	0.684	13747.000
weighted avg	0.714	0.706	0.691	13747.000

Gambar 5. Classification Report Model Naïve Bayes

Gambar 5. Classification Report Model Naïve Bayes

(bisa juga disajikan sebagai Tabel 2 berisi nilai precision, recall, F1-score per kelas)

Classification report menunjukkan bahwa:

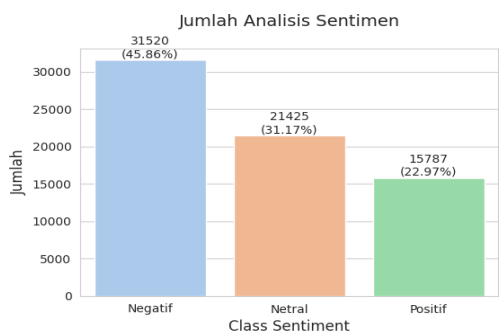
Akurasi keseluruhan berada pada kategori baik, merefleksikan kemampuan model membedakan dua kelas dominan (positif dan negatif).

Precision dan recall tertinggi terdapat pada kelas positif dan negatif, menguatkan kesimpulan bahwa model robust dalam mendeteksi ekspresi apresiasi maupun keluhan.

F1-score terendah muncul pada kelas netral, konsisten dengan temuan pada confusion matrix bahwa kelas ini paling banyak mengalami misklasifikasi.

Secara konseptual, hasil tersebut memperlihatkan bahwa ketimpangan jumlah data antar kelas dan tumpang tindih kosakata ikut mempengaruhi kinerja model, terutama terhadap kategori netral. Untuk peningkatan di masa depan, diperlukan strategi penyeimbangan data (misalnya resampling, class weighting) atau pengayaan fitur yang lebih peka terhadap konteks, sebagaimana disarankan pada studi tentang pemilihan algoritma dan skema preprocessing untuk analisis sentimen[3] [3].

Distribusi Sentimen Hasil Prediksi



Gambar 6. Visualisasi Distribusi Sentiment Hasil Klasifikasi

Visualisasi distribusi sentimen menggambarkan proporsi hasil prediksi model sebagai berikut: sekitar 22% ulasan bernuansa positif, 45% negatif, dan 31% netral.

Interpretasi utama dari distribusi ini:

Ulasan negatif yang dominan ($\pm 45\%$) mengindikasikan masih banyak persoalan teknis yang dirasakan pengguna, terutama terkait akses login, proses verifikasi peserta, dan kestabilan aplikasi.

Ulasan positif ($\pm 22\%$) menunjukkan bahwa aplikasi tetap dianggap membantu dan bermanfaat oleh sebagian pengguna, khususnya dalam mempermudah akses layanan tanpa tatap muka langsung.

Ulasan netral ($\pm 31\%$) cenderung berisi penilaian sedang atau deskriptif tanpa ekspresi emosi yang kuat; proporsi kelas ini wajar lebih rendah dibanding positif + negatif karena pengguna layanan publik sering menyampaikan opini ketika merasa sangat puas atau sangat kecewa.

Temuan ini melengkapi hasil-hasil studi terdahulu mengenai penerimaan dan penggunaan Mobile JKN [1][2],[13][16] dengan perspektif baru dari sisi sentimen tekstual yang berasal langsung dari ulasan pengguna di Google Play Store.

Analisis Wordcloud per Kelas Sentimen



Gambar 7. Sentiment Positif



Gambar 8. Sentiment Negatif



Gambar 9 Sentimen Netral

Wordcloud per kelas menguatkan temuan kuantitatif:

Wordcloud positif didominasi kata seperti “bagus”, “membantu”, “cepat”, “mudah”, yang menggambarkan apresiasi terhadap kemudahan dan kecepatan layanan.

Wordcloud negatif menonjolkan istilah “error”, “tidak bisa”, “login”, “verifikasi”, menunjukkan bahwa keluhan utama berkaitan dengan masalah teknis, bukan semata-mata konten layanan.

Wordcloud netral biasanya menampilkan kata-kata deskriptif atau komentar singkat yang tidak secara jelas condong ke positif maupun negatif.

Kombinasi analisis kuantitatif (metrik evaluasi) dan kualitatif (visualisasi teks) memberikan gambaran menyeluruh: model mampu menangkap pola sentimen dominan yang selaras dengan karakteristik masalah pada aplikasi Mobile JKN, dan hasilnya konsisten dengan interpretasi manual terhadap kata-kata kunci dalam ulasan.

4. KESIMPULAN

Berdasarkan hasil pemodelan dan evaluasi, dapat disimpulkan bahwa:

Algoritma Multinomial Naïve Bayes dengan fitur TF-IDF mampu mengklasifikasikan sentimen ulasan Mobile JKN dengan akurasi keseluruhan yang baik, terutama pada kelas positif dan negatif,

yang tercermin dari nilai precision dan recall yang relatif tinggi.

Kelas netral merupakan kategori yang paling sulit diprediksi, dengan F1-score terendah dan tingkat misklasifikasi tertinggi akibat panjang ulasan yang pendek, kemiripan struktur bahasa dengan kelas lain, serta jumlah data yang lebih sedikit.

Distribusi sentimen hasil prediksi menunjukkan bahwa ulasan negatif masih mendominasi (sekitar 45%) dibandingkan positif dan netral, menandakan bahwa masih terdapat persoalan teknis signifikan pada aplikasi Mobile JKN, terutama pada aspek login, verifikasi, dan stabilitas aplikasi.

Visualisasi wordcloud dan distribusi sentimen memperkuat interpretasi bahwa model tidak hanya mampu mengukur performa secara numerik, tetapi juga membantu mengidentifikasi fokus perbaikan layanan berdasarkan kata-kata kunci yang paling sering muncul di setiap kategori sentimen.

Untuk pengembangan ke depan, disarankan peneliti berikutnya mengeksplorasi teknik class imbalance handling, menambah fitur berbasis konteks (misalnya word embeddings), serta membandingkan performa Naïve Bayes dengan arsitektur yang lebih kompleks seperti SVM atau model berbasis transformer pada korpus ulasan Mobile JKN.

5. REFERENSI

- [1] D. Yuliani, "Analysis of public perception of JKN services using text mining," *Jurnal/Konferensi Tidak Diketahui*, 2022.
- [2] R. Ramadhani, "User satisfaction analysis of digital health applications based on reviews," *Jurnal/Konferensi Tidak Diketahui*, 2023.
- [3] E. D. Madyatmadja, B. N. Yahya, and C. Wijaya, "Contextual text analytics framework for citizen report classification: A case study using the Indonesian language," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3158940.
- [4] K. J. Tan, "Text preprocessing and algorithm selection for sentiment analysis reliability," *Jurnal/Konferensi Tidak Diketahui*, 2023.
- [5] N. A. M. Razali *et al.*, "Opinion mining for national security: Techniques, domain applications, challenges and research opportunities," *J Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00536-5.
- [6] A. Gasparetto, A. Zangari, M. Marcuzzo, and A. Albarelli, "A survey on text classification: Practical perspectives on the Italian language," *PLoS One*, vol. 17, no. 7, 2022, doi: 10.1371/journal.pone.0270904.
- [7] Q. Shen, S. Han, Y. Han, and X. Chen, "User review analysis of dating apps based on text mining," *PLoS One*, vol. 18, no. 4, 2023, doi: 10.1371/journal.pone.0283896.
- [8] M. Hidayatullah, "Sentiment analysis on digital public service reviews using Naïve Bayes," *Jurnal/Konferensi Tidak Diketahui*, 2023.
- [9] W. Utami and D. Kurniawan, "Evaluation of PeduliLindungi application reviews using Naïve Bayes," *Jurnal/Konferensi Tidak Diketahui*, 2022.
- [10] S. A. Rizky and L. Sulastri, "Sentiment analysis on digital health service applications," *Jurnal/Konferensi Tidak Diketahui*, 2021.
- [11] E. S. Lestari and A. Hidayat, "Sentimen pengguna aplikasi e-commerce menggunakan Naïve Bayes," *Jurnal/Konferensi Tidak Diketahui*, 2021.
- [12] R. Nasrullah and A. Fauzi, "Sentiment analysis in digital public service evaluation," *Jurnal/Konferensi Tidak Diketahui*, 2023.
- [13] F. Damanik, A. W. Widayanti, and C. Wiedyaningsih, "User acceptance of Mobile-JKN: Insights from the Technology Acceptance Model," *Jurnal Administrasi Kesehatan Indonesia*, vol. 12, no. 2, 2024, doi: 10.20473/jaki.v12i2.2024.206-217.
- [14] A. W. Wibowo, "The importance of stemming for Indonesian language text analysis," *Jurnal/Konferensi Tidak Diketahui*, 2020.
- [15] H. Nabilah, P. Permatasari, A. V El-Tsana, R. P. Anggitya, and A. R. 'Aisy, "The implementation of the JKN mobile application as an effort to improve the quality of health services in Indonesia: Literature review," *Jurnal Kesehatan*, vol. 15, no. 3, 2024, doi: 10.35730/jk.v15i3.1181.
- [16] A. P. Zulfa, A. Makmun, Z. K. Novriansyah, F. Sommeng, and Dahlia, "The relationship between the use of Mobile JKN application and health services at Majene Regency Hospital," *Journal of Health Policy and Management*, vol. 10, no. 2, 2025, doi: 10.26911/thejhp.m.s2025.10.02.01.