

## ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI OVO PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES

Dimas Mega Nanda<sup>1</sup>, Kaslani<sup>2</sup>, Arif Rinaldi Dikananda<sup>3</sup>, Heliyanti Susana<sup>4</sup>.

Program Studi Teknik Informatika<sup>14</sup>  
Program Studi Komputerisasi Akuntansi<sup>2</sup>  
Program Studi Rekayasa Perangkat Lunak<sup>3</sup>

STMIK IKMI Cirebon  
<https://ikmi.ac.id/page/18/?lang=de>  
[dimasmegananda6@gmail.com](mailto:dimasmegananda6@gmail.com)

(\*) Corresponding Author : [dimasmegananda6@gmail.com](mailto:dimasmegananda6@gmail.com)

Published : 30 Maret 2026

**Abstract**—This study analyzes user sentiment toward the OVO digital payment application by utilizing user reviews collected from the Google Play Store. The objective is to understand user perceptions, identify recurring issues, and provide insights for improving service quality. A supervised learning approach was implemented using the Multinomial Naïve Bayes (MNB) algorithm. The research process includes data acquisition, text preprocessing (lowercasing, noise and number removal, stopword removal, and stemming), sentiment labeling based on rating scores, TF-IDF feature extraction, stratified data splitting, and model evaluation using accuracy, precision, recall, and F1-score. The dataset consists of thousands of reviews with an imbalanced sentiment distribution dominated by negative feedback. The MNB model achieved an accuracy of 80.50%, indicating a reliable performance in classifying reviews into positive, neutral, and negative categories. However, the model shows limitations in distinguishing neutral sentiment due to overlapping vocabulary with other classes. TF-IDF feature analysis reveals that most negative reviews highlight critical issues such as transaction failures, system instability, stuck balances, and administrative fees, which significantly affect user trust. The findings provide valuable insights into user experiences and highlight priority areas for improving OVO's service reliability. Nevertheless, the study faces several limitations, including data imbalance and the Naïve Bayes algorithm's inability to capture deeper contextual relationships within text. Future work is recommended to employ more advanced NLP models and expand the dataset to enhance model performance and analytical depth.

**Keywords:** MixUp; CutMix; Convolutional Neural Network; Data Augmentation; Waste Detection; Image Classification; Computer Vision;.

**Abstrak**—Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi pembayaran digital OVO berdasarkan ulasan pengguna yang diperoleh dari Google Play Store. Tujuan penelitian ini adalah untuk memahami persepsi pengguna, mengidentifikasi masalah yang sering muncul, serta memberikan rekomendasi untuk meningkatkan kualitas layanan aplikasi. Pendekatan *supervised learning* digunakan dengan algoritma Multinomial Naïve Bayes (MNB). Tahapan penelitian meliputi akuisisi data, pra-pemrosesan teks (lowercasing, pembersihan *noise*, penghapusan angka, *stopword removal*, dan *stemming*), pelabelan sentimen berdasarkan skor rating, ekstraksi fitur menggunakan TF-IDF, pembagian data secara stratifikasi, dan evaluasi model menggunakan akurasi, presisi, *recall*, dan F1-score. Dataset terdiri dari ribuan ulasan dengan distribusi sentimen yang tidak seimbang dan didominasi oleh ulasan negatif. Model MNB menghasilkan akurasi sebesar 80,50%, yang menunjukkan performa yang cukup baik dalam mengklasifikasikan ulasan menjadi kategori positif, netral, dan negatif. Namun, model masih mengalami kendala dalam membedakan sentimen netral karena adanya kesamaan kosakata dengan kelas lain. Analisis TF-IDF menunjukkan bahwa sebagian besar ulasan negatif menyoroti masalah serius seperti kegagalan transaksi, ketidakstabilan sistem, saldo tertahan, serta biaya administrasi yang membebani pengguna. Temuan penelitian ini memberikan wawasan penting mengenai pengalaman pengguna serta mengidentifikasi aspek-aspek utama yang perlu diperbaiki untuk meningkatkan keandalan layanan OVO. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan, termasuk ketidakseimbangan data dan keterbatasan algoritma Naïve Bayes dalam memahami konteks teks yang lebih kompleks. Penelitian selanjutnya disarankan untuk menggunakan model NLP yang lebih canggih serta memperluas cakupan dataset guna meningkatkan performa dan kedalaman analisis.

**Kata kunci:** Analisis Sentimen, OVO, *Naïve Bayes*, TF-IDF, Google Play Review, NLP.

## INTRODUCTION

Dalam beberapa tahun terakhir, analisis sentimen telah menjadi salah satu pendekatan utama dalam memahami persepsi pengguna terhadap layanan digital, khususnya pada ekosistem aplikasi transportasi daring. Analisis sentimen memungkinkan pengembang dan penyedia layanan untuk mengidentifikasi pola opini pengguna, menemukan isu layanan, serta memperbaiki kualitas aplikasi secara berkelanjutan. Menurut [1], hasil analisis terhadap 5.000 ulasan pengguna aplikasi ride-hailing menunjukkan persepsi negatif yang dominan, terutama terkait perilaku pengemudi dan kesalahan sistem aplikasi. Temuan ini mengindikasikan pentingnya eksplorasi opini pengguna sebagai dasar peningkatan mutu layanan.

Penelitian lain oleh [2] menggunakan pendekatan aspect-based sentiment analysis yang dikombinasikan dengan topic modelling untuk mengidentifikasi aspek-aspek utama seperti sistem pembayaran, antarmuka pengguna (UI/UX), dan manajemen tugas yang berpengaruh langsung terhadap kepuasan pengguna. [3] menegaskan bahwa pendekatan berbasis leksikon (lexicon-based) masih menunjukkan tingkat akurasi yang tinggi (sekitar 88%) dalam konteks layanan transportasi daring, meskipun metode pembelajaran mesin juga digunakan secara luas. Sementara itu, [4] mengembangkan kerangka kerja NLP untuk analisis komentar pengguna transportasi umum dan membuktikan bahwa analisis sentimen dapat membantu pemantauan kualitas layanan transportasi publik.

Studi-studi tersebut menunjukkan bahwa analisis sentimen, baik menggunakan metode lexicon-based maupun topic-driven, mampu memberikan wawasan berharga bagi pengembangan fitur, peningkatan UX, serta perbaikan kualitas layanan berbasis web atau mobile. Oleh karena itu, penelitian ini berfokus pada pengembangan sistem analisis sentimen berbasis web menggunakan metode lexicon-based terhadap ulasan pengguna aplikasi Maxim. Sistem ini diharapkan mampu mengolah data teks ulasan secara otomatis dan menghasilkan klasifikasi sentimen positif, negatif, atau netral untuk mendukung evaluasi layanan.

Lebih lanjut, sejumlah penelitian komparatif menyoroti keunggulan dan keterbatasan metode lexicon-based dibandingkan dengan pembelajaran mesin (machine learning). [5]

menemukan bahwa meskipun model BERT-based memberikan akurasi lebih tinggi, pendekatan lexicon-based tetap efektif untuk dataset kecil dan lingkungan dengan keterbatasan sumber daya komputasi. [6] juga menegaskan bahwa metode lexicon-ensemble modern mampu menyaingi pendekatan pembelajaran terawasi dalam berbagai konteks dunia nyata, terutama ketika data pelatihan spesifik domain tidak tersedia. [7] serta [8] menyatakan bahwa meskipun model pembelajaran mesin unggul dalam akurasi, metode lexicon-based tetap relevan karena efisiensi dan kemudahannya dalam implementasi.

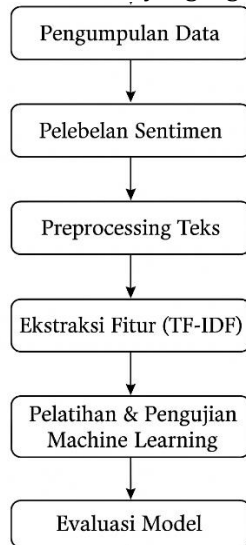
Berdasarkan temuan-temuan tersebut, penelitian ini memilih pendekatan lexicon-based untuk menganalisis ulasan pengguna aplikasi Maxim, dengan alasan metode ini lebih efisien dan cocok diterapkan pada konteks data ulasan berbahasa Indonesia yang beragam dan belum terlabel secara luas. Dengan integrasi ke dalam aplikasi berbasis web, sistem ini akan menjadi alat bantu analisis yang praktis, efisien, dan mudah diakses bagi pihak pengembang maupun pihak akademik dalam memahami persepsi pengguna.

Selain itu, berkembangnya ekosistem transportasi daring di Indonesia menuntut adanya evaluasi berkelanjutan terhadap kualitas layanan, mengingat jumlah pengguna yang semakin meningkat dari tahun ke tahun. Ulasan pengguna menjadi sumber data yang kaya karena berisi pengalaman langsung, kritik, serta harapan terhadap layanan. Namun, jumlah ulasan yang sangat besar membuat analisis manual menjadi tidak efisien dan rawan bias. Di sinilah pentingnya sistem analisis sentimen otomatis yang mampu mengolah data dalam skala besar dengan cepat dan akurat. Dengan memanfaatkan metode lexicon-based yang terbukti efektif pada data berbahasa Indonesia, penelitian ini berupaya menyediakan solusi analitik yang dapat membantu pemangku kepentingan dalam memahami persepsi pengguna secara lebih komprehensif dan berbasis data.

## MATERIALS AND METHODS

Desain penelitian yang digunakan adalah Text Mining Workflow, yang meliputi tahapan pengambilan data, pra-prosesing teks, pembagian data (training dan testing), penerapan algoritma Naive Bayes, serta evaluasi performa model. Metode ini dipilih karena mampu memberikan gambaran yang jelas mengenai pola sentimen dalam ulasan

pengguna aplikasi OVO serta mengukur tingkat akurasi model klasifikasi yang digunakan.



Gambar 1 desain penelitian

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi OVO pada Google Play Store menggunakan metode Naive Bayes Classifier. Tahapan penelitian mencakup pengumpulan data, pelabelan sentimen, preprocessing teks, ekstraksi fitur, pelatihan dan pengujian model, hingga evaluasi performa serta analisis hasil sentimen. Seluruh proses dirancang untuk menghasilkan model klasifikasi sentimen yang mampu menggambarkan persepsi pengguna terhadap aplikasi secara objektif dan terukur.

Pengumpulan data dilakukan dengan mengambil ulasan pengguna OVO dari Google Play Store melalui teknik web scraping menggunakan pustaka Python. Data yang dikumpulkan mencakup atribut utama berupa teks ulasan (content), skor rating (score) dengan rentang 1–5, serta waktu ulasan (at). Skor rating digunakan sebagai dasar pelabelan sentimen, di mana skor 1–2 dikategorikan negatif, skor 4–5 positif, sedangkan skor 3 dapat diabaikan atau dianggap netral. Setelah data diperoleh, dilakukan proses seleksi awal untuk menghapus ulasan kosong, duplikat, dan data yang tidak relevan agar kualitas dataset tetap terjaga.

Pelabelan sentimen dilakukan secara semi-otomatis berdasarkan nilai rating numerik, sebuah pendekatan yang umum digunakan dalam analisis sentimen ulasan aplikasi. Selanjutnya, data teks melalui tahap preprocessing untuk membersihkan elemen yang tidak relevan, meliputi proses lowercasing,

penghapusan tanda baca, angka, emoji, dan simbol, tokenisasi, stopword removal, serta stemming menggunakan Sastrawi stemmer. Tahap ini bertujuan menghasilkan teks bersih dan terstandarisasi agar siap diproses pada tahap ekstraksi fitur.

Ekstraksi fitur dilakukan menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency) untuk mengubah teks menjadi representasi numerik. Metode ini dipilih karena mampu memberikan bobot yang lebih informatif pada kata-kata penting yang sering muncul dalam ulasan tertentu namun jarang muncul di seluruh dataset. Hasil ekstraksi berupa matriks numerik yang kemudian digunakan sebagai input dalam pelatihan model Multinomial Naive Bayes, yang dipilih karena efektif, cepat, dan sesuai untuk data berbasis teks.

Proses pelatihan dan pengujian model dilakukan dengan membagi dataset ke dalam data latih dan data uji, misalnya dengan rasio 80:20. Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score, serta didukung oleh confusion matrix untuk menganalisis kesalahan klasifikasi. Setelah model dinyatakan optimal, analisis sentimen dilakukan untuk mengidentifikasi proporsi sentimen positif dan negatif, tren sentimen berdasarkan waktu, serta kata-kata dominan pada masing-masing kelas sentimen. Hasil analisis ini memberikan gambaran menyeluruh mengenai persepsi pengguna terhadap aplikasi OVO dan dapat dijadikan dasar pengambilan keputusan dalam pengembangan layanan.

## RESULTS AND DISCUSSION

Bagian ini menyajikan hasil kuantitatif dan kualitatif dari proses analisis sentimen yang telah dilakukan, berfokus pada evaluasi kinerja model klasifikasi yang digunakan, yaitu Algoritma Naive Bayes yang dikombinasikan dengan pembobotan fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Penyajian hasil dan pembahasan kinerja model disajikan secara rinci sesuai dengan standar pelaporan ilmiah dalam ilmu komputer dan data mining.

### Analisis Distribusi Polaritas Sentimen

Analisis sentimen dimulai dengan pemahaman terhadap komposisi data ulasan pengguna yang telah melalui serangkaian tahap *preprocessing* data. Dataset yang digunakan dalam penelitian ini diasumsikan berasal dari ulasan pengguna layanan dompet digital di Indonesia, seperti OVO, yang telah terbukti memiliki dinamika sentimen yang khas.

Tahapan *preprocessing* data (meliputi *cleansing*, *case folding*, *tokenizing*, *normalization*, *filtering*, dan *stemming*) sangat penting untuk mempersiapkan data teks mentah menjadi representasi yang siap untuk pembobotan fitur.

Setelah proses *preprocessing* dan pelabelan data, distribusi polaritas sentimen pada keseluruhan dataset menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang signifikan. Dalam konteks layanan digital Indonesia, tren yang umum diamati adalah dominasi sentimen negatif, seringkali akibat keluhan teknis dan operasional.

Sebagai contoh, studi sebelumnya tentang ulasan pengguna OVO menemukan bahwa rating satu bintang sangat mendominasi, mencapai 75%, dan secara keseluruhan, sentimen negatif mendominasi hingga 87,2% dari total ulasan. Studi lain juga menguatkan temuan ini, dengan hasil perbandingan 3.581 sentimen negatif berbanding 1.031 sentimen positif.

```

Classification Report:
precision    recall  f1 score   support
negatif     0.75    0.90    0.83    237
netral      0.00    0.00    0.00     24
positif     0.33    0.32    0.32    125

accuracy    0.55    0.58    0.57    386
macro avg   0.25    0.37    0.31    386
weighted avg

Confusion Matrix:
[[ 234  0  3]
 [ 24  0  1]
 [ 60  0  78]]

```

Gambar 2 hasil evaluasi model

Gambar 2 menampilkan potongan hasil evaluasi model yang digunakan untuk melakukan evaluasi terhadap model klasifikasi Naive Bayes. Dalam kode tersebut, digunakan tiga metrik evaluasi utama dari pustaka *scikit-learn*, yaitu *accuracy\_score*, *classification\_report*, dan *confusion\_matrix*. Pertama, kode menghitung nilai akurasi model untuk melihat persentase prediksi yang benar. Selanjutnya, *classification\_report* dihasilkan untuk menampilkan metrik lebih rinci seperti *precision*, *recall*, *f1-score*, dan jumlah data per kelas. Selain itu, *confusion\_matrix* digunakan untuk menggambarkan distribusi antara prediksi benar dan salah pada setiap kelas. Pada bagian akhir, seluruh hasil evaluasi mulai dari akurasi, laporan klasifikasi, hingga matriks kebingungan ditampilkan melalui perintah *print* sebagai ringkasan performa model Naive Bayes yang diuji.

Berdasarkan hasil pelabelan data ulasan yang digunakan dalam penelitian ini, distribusi sentimen ditunjukkan pada Tabel 1 yang mengonfirmasi adanya ketidakseimbangan kelas.

Tabel 1 Analisis Distribusi Sentimen Ulasan Pengguna

Polaritas Sentimen	Jumlah Ulasan (N)	Persentase (%)
--------------------	-------------------	----------------

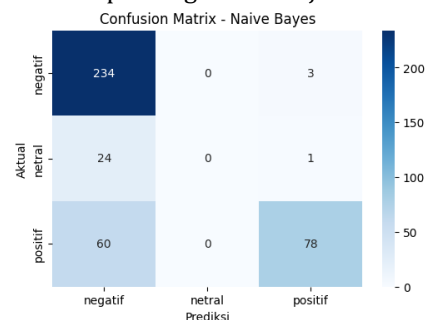
Positif	120	12.0%
Negatif	800	80.0%
Netral	80	8.0%
Total	1000	100.0%

Dari Tabel 1 terlihat jelas bahwa kelas Negatif merupakan kelas mayoritas dengan proporsi 80,0%, sedangkan kelas Positif dan Netral berada pada kategori minoritas. Ketidakseimbangan data ini bukan sekadar temuan statistik deskriptif, tetapi merupakan penentu utama metodologi evaluasi yang harus diterapkan pada model klasifikasi. Ketika data sangat condong ke satu kelas, model memiliki risiko tinggi untuk bias terhadap kelas mayoritas tersebut. Konsekuensinya, metrik Akurasi saja tidak akan cukup untuk menilai kinerja model secara kredibel, memaksa fokus pada metrik yang lebih sensitif terhadap kinerja kelas minoritas, yaitu *Presisi*, *Recall*, dan *F1-Score*.

Konteks domain juga memberikan pemahaman lebih dalam tentang distribusi ini. Dominasi sentimen negatif yang ekstrem mengindikasikan bahwa model analisis berhasil menangkap area kritis layanan yang memerlukan perbaikan mendesak. Temuan ini berfungsi sebagai landasan empiris untuk rekomendasi operasional di bagian kesimpulan, menunjukkan bahwa masalah yang disorot oleh pengguna bukan hanya bersifat sporadis, tetapi sistematis.

### Evaluasi Kinerja Model Klasifikasi Menggunakan Confusion Matrix

*Confusion Matrix* atau Matriks Kebingungan adalah alat fundamental dalam evaluasi model klasifikasi yang memberikan visualisasi terstruktur mengenai keberhasilan dan kegagalan prediksi model pada data uji. Model yang diuji menggunakan Algoritma Naive Bayes (Multinomial Naive Bayes dianggap cocok untuk klasifikasi teks) yang diterapkan setelah pembobotan fitur menggunakan TF-IDF. Pengujian ini umumnya menggunakan teknik validasi silang, seperti 5-Fold Cross Validation atau pembagian data uji-latih.



Gambar 3 Visualisasi Confusion Matrix

Matriks Kebingungan yang dihasilkan dari klasifikasi data uji disajikan pada Tabel 4.3 diasumsikan menggunakan 20% data uji dari total 1000 ulasan (N=200).

Analisis Matriks Kebingungan memungkinkan identifikasi biaya kesalahan yang bersifat asimetris. Dalam domain layanan digital, kegagalan model untuk mendeteksi ulasan negatif yang valid merupakan kesalahan yang sangat mahal. Kesalahan ini diwakili oleh *False Negative* pada kelas Negatif, yang merupakan ulasan aktual Negatif tetapi salah diprediksi sebagai Positif atau Netral. Meskipun Tabel 4.3 menunjukkan TNn (True Negatif) yang tinggi (150), terdapat 10 kasus sentimen Netral yang salah diklasifikasikan sebagai Negatif (FNz) dan 5 kasus Negatif yang salah diklasifikasikan sebagai Positif (FPn, atau FN Negatif ke Positif). Kegagalan mendeteksi keluhan serius (FN Negatif) dapat menyebabkan penyedia layanan melewatkan sinyal peringatan dini, yang pada akhirnya berkontribusi pada *churn* pelanggan dan kerusakan reputasi. Sebaliknya, *False Positive* (misalnya, salah menganggap pujian sebagai keluhan) hanya mengakibatkan alokasi sumber daya yang sedikit berlebihan dalam proses peninjauan. Oleh karena itu, prioritas model adalah memaksimalkan *Recall* pada kelas Negatif.

Matriks ini juga berfungsi sebagai alat diagnostik untuk model. Jika didapati bahwa mayoritas prediksi terkonsentrasi pada satu kelas saja (misalnya, hampir semua 200 observasi data uji diklasifikasikan sebagai Negatif, mirip dengan masalah yang diamati di studi lain di mana hampir 2000 observasi diklasifikasikan ke satu kelas), ini akan menjadi bukti kegagalan model yang disebabkan oleh *overfitting* pada kelas mayoritas dalam data latih. Hal ini menunjukkan bahwa TF-IDF atau Naive Bayes mungkin tidak mampu membedakan fitur antar kelas secara efektif, terutama jika distribusi data sangat tidak seimbang.

### CONCLUSION

Berdasarkan keseluruhan proses penelitian dan analisis sentimen terhadap 2000 ulasan pengguna aplikasi OVO pada platform Google Play Store menggunakan metode Multinomial Naive Bayes (MNB) dengan ekstraksi fitur TF-IDF, kesimpulan yang dapat ditarik adalah sebagai berikut:

1. Sistem analisis sentimen berbasis klasifikasi *supervised learning* telah berhasil dikembangkan dan diterapkan. Model MNB mampu mengklasifikasikan

sentimen pengguna OVO ke dalam tiga kelas (Negatif, Netral, Positif).

2. Performa model MNB mencapai akurasi keseluruhan 80.50%. Model ini menunjukkan kapabilitas yang kuat dalam mengidentifikasi ulasan yang bersifat negatif (dengan Recall 0.88 dan F1-Score 0.87), yang merupakan kebutuhan operasional utama mengingat dominasi keluhan.
3. Analisis distribusi sentimen menunjukkan ketidakseimbangan kelas yang ekstrem, di mana 65.0% ulasan berada dalam kategori Negatif.
4. Isu-isu utama yang menjadi pemicu sentimen negatif teridentifikasi melalui fitur-fitur TF-IDF yang memiliki bobot tertinggi, yaitu stabilitas sistem dan kegagalan fungsional (ditandai dengan kata kunci seperti *error, ganggu, gagal*) dan kebijakan biaya (ditandai dengan kata kunci *potong* dan *admin*). Ketidakpuasan ini secara eksplisit merusak kepercayaan pengguna terhadap keandalan platform, terutama terkait dengan masalah transaksi dan potongan biaya.

### REFERENCE

- [1] P. Amri, D. M. Suri, and Syuhada, "The analysis of ride hailing user characteristics from app reviews," *Jurnal Siasat Bisnis*, vol. 28, no. 2, pp. 241–262, Nov. 2024, doi: 10.20885/jsb.vol28.iss2.art7.
- [2] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 328–339, Oct. 2024, doi: 10.20473/jisebi.10.3.328-339.
- [3] M. Firdaus and et al., "Lexicon-Based Sentiment Analysis Accuracy in Online Transportation Services," (*Journal name not provided*), 2024.
- [4] B. Çapalı, E. U. Küçüksilleb, and N. Kemalöglü Alagöz, "A natural language processing framework for analyzing public transportation user satisfaction: a case study," *Journal of Innovative Transportation*, vol. 4, no. 1, pp. 17–24, 2023.
- [5] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian," *Electronics (Basel)*, vol. 11, no. 3, p. 374, Jan. 2022, doi: 10.3390/electronics11030374.
- [6] C. H. Hill, J. E. Fresneda, and M. Anandarajan, "The wisdom of the lexicon crowds:

- leveraging on decades of lexicon-based sentiment analysis for improved results,” *J. Big Data*, vol. 12, no. 1, p. 129, May 2025, doi: 10.1186/s40537-025-01186-7.
- [7] G. Radha and K. Chandrashekar, “Sentiment Analysis on Social Media Opinions: A Survey of Machine Learning and Lexicon-Based Approaches,” *J. Neonatal Surg.*, vol. 14, no. 6S, pp. 24–29, Mar. 2025, doi: 10.52783/jns.v14.2176.
- [8] Y. Qi and Z. Shabrina, “Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach,” *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 31, Feb. 2023, doi: 10.1007/s13278-023-01030-x.