

Optimasi Metode Naïve Bayes Classifier Menggunakan Pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) Pada Analisis Sentimen

¹Ardiansyah, ²Kurniawan

^{1,2}Universitas Bina Darma, Indonesia

ardi48298@gmail.com; kurniawan@binadarma.ac.id

Article Info

Article history:

Received, 2024-09-26

Revised, 2024-10-10

Accepted, 2024-11-06

Kata Kunci:

analisis_sentimen
naïve_bayes_classifier
web_scraping
tf-idf
rsud_siti_fatimah

Keywords:

sentiment_analysis
naïve_bayes_classifier
web_scraping
tf-idf
siti_fatimah_general_hospital

ABSTRAK

Tujuan utama dari penelitian ini adalah untuk melakukan analisis sentimen publik yang diarahkan terhadap RSUD Siti Fatimah, menggunakan metodologi Naïve Bayes Classifier. Pendekatan analitis ini digunakan untuk secara sistematis mengkategorikan ulasan ke dalam sentimen positif dan negatif. Data yang berkaitan dengan ulasan diperoleh melalui teknik pengikisan web dari Google Maps, diikuti oleh serangkaian prosedur prapemrosesan teks, yang mencakup sanitasi teks, tokenisasi, dan penerapan TF-IDF untuk pembobotan. Berdasarkan nilai Klasifikasi positif Precision menunjukkan 83%, Recal 1,00 dan F-1 Score 0,91 yang berarti Model ini menunjukkan kinerja yang sangat baik dalam mengidentifikasi sentimen positif. Namun, model kurang efektif dalam mengidentifikasi sentimen negatif, dengan recall yang sangat rendah.

ABSTRACT

The main objective of this research is to conduct an analysis of public sentiment directed toward RSUD Siti Fatimah, using the Naïve Bayes Classifier methodology. This analytical approach was used to systematically categorize reviews into positive and negative sentiments. Data relating to the reviews was obtained through web scraping techniques from Google Maps, followed by a series of text preprocessing procedures, which included text sanitization, tokenization, and the application of TF-IDF for weighting. Based on the positive Classification values Precision shows 83%, Recal 1.00, and F-1 Score 0.91 which means the Model shows excellent performance in identifying positive sentiments. However, the model is less effective in identifying negative sentiments, with very low recall.

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



Penulis Korespondensi:

Ardiansyah,
Universitas Bina Darma,
Email: ardi48298@gmail.com

1. PENDAHULUAN

Rumah Sakit Umum Daerah (RSUD) berfungsi sebagai institusi kesehatan di bawah kepemilikan pemerintah daerah, yang meliputi Rumah Sakit Siti Fatimah yang berlokasi di Palembang. Sebagai penyedia layanan kesehatan terkemuka di Sumatera Selatan, Rumah Sakit Siti Fatimah memberikan kontribusi signifikan terhadap penyediaan layanan kesehatan kepada penduduk sekitarnya. Evaluasi publik mengenai layanan yang diberikan oleh rumah sakit ini sering disampaikan melalui berbagai platform, termasuk Google Maps.

Evaluasi semacam itu sangat penting karena menawarkan wawasan komprehensif tentang persepsi masyarakat tentang layanan rumah sakit [1]–[3]. Sentimen publik yang berlaku mengenai layanan rumah sakit dapat berfungsi sebagai indikator tingkat kepuasan atau ketidakpuasan, yang pada gilirannya mempengaruhi citra publik lembaga dan kepercayaan Masyarakat [1], [4]–[6].

Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen tinjauan publik yang berkaitan dengan RSUD Siti Fatimah dengan menggunakan metodologi Naïve Bayes Classifier, algoritma pembelajaran mesin yang digunakan untuk klasifikasi teks. Pendekatan Naïve Bayes diakui sebagai salah satu teknik unggulan dalam analisis sentimen karena keterusterangannya dan kemampuannya untuk mengelola kumpulan data berbasis teks yang luas dengan mahir [7]–[11]. Dalam ranah analisis sentimen, teknik ini digunakan untuk mengkategorikan teks ulasan menjadi dua klasifikasi berbeda: positif dan negatif.

Analisis Sentimen merupakan proses sistematis memahami, mengekstraksi, dan mengkategorikan pendapat atau ekspresi emosional dari teks yang dibuat pengguna, seperti ulasan atau komentar [12]–[15]. Tujuan utama analisis sentimen adalah untuk memastikan apakah teks yang dianalisis menyampaikan sentimen positif, negatif, atau netral. Dalam konteks penelitian ini, analisis sentimen digunakan untuk menjelaskan persepsi masyarakat mengenai layanan yang diberikan oleh Rumah Sakit Siti Fatimah, yang berfungsi sebagai umpan balik penting bagi institusi kesehatan yang berusaha meningkatkan kualitas layanan.

Naïve Bayes Classifier (NBC) beroperasi berdasarkan Teorema Bayes untuk meramalkan klasifikasi teks berdasarkan probabilitas terkait. Algoritma ini sering digunakan dalam analisis sentimen karena kemampuannya dalam mengklasifikasikan data teks, ditambah dengan kemampuannya dalam memproses kumpulan data substansial. Sejumlah penelitian sebelumnya, seperti yang dilakukan oleh Rahman et al. (2024), menunjukkan bahwa Naïve Bayes menunjukkan kinerja yang terpuji dalam klasifikasi teks, terutama dalam menilai kelayakan pelanggan dalam domain perbankan. Dalam studi saat ini, Naïve Bayes disandingkan dengan metodologi Random Forest, menghasilkan hasil yang menunjukkan Naïve Bayes lebih unggul dalam mengklasifikasikan data berdasarkan probabilitas [16].

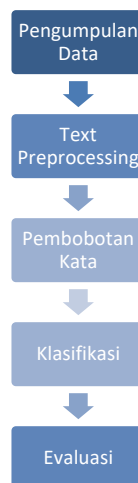
Penelitian lebih lanjut yang dilakukan oleh Khofifah dkk. (2022) menguatkan kemandirian Naïve Bayes dalam melakukan analisis sentimen pada ulasan pengguna mengenai tempat wisata di Google Maps. Temuan penyelidikan ini mengungkapkan bahwa Naïve Bayes mencapai akurasi yang terpuji dalam klasifikasi teks berorientasi sentimen, menjadikannya metodologi yang cocok untuk menganalisis ulasan layanan di RSUD Siti Fatimah [17].

Investigasi oleh Maulana et al. (2023), dalam menjalankan analisis sentimen aplikasi MyPertamina menggunakan algoritma NBC, menunjukkan bahwa metodologi ini efektif dalam menafsirkan sentimen ulasan pengguna [18]. Selanjutnya, penelitian oleh Darwis et al. (2021), yang menerapkan teknik Naïve Bayes untuk analisis sentimen data Twitter, juga menunjukkan bahwa algoritma ini dapat berperan dalam membedakan opini publik mengenai layanan public [5].

Mengambil wawasan dari berbagai penelitian sebelumnya, penerapan Naïve Bayes Classifier dalam analisis sentimen di Rumah Sakit Siti Fatimah diantisipasi untuk menghasilkan hasil yang tepat, yang pada akhirnya dapat berfungsi sebagai sumber daya yang berharga bagi administrasi rumah sakit dalam upaya mereka untuk meningkatkan kualitas layanan kesehatan.

2. METODE PENELITIAN

Metode penelitian menggunakan metodologi Naïve Bayes Classifier yang dipergunakan untuk melakukan analisis sentimen komprehensif dari ulasan yang dikumpulkan secara sistematis dari platform Google Maps, menggunakan teknik scraping yang dirancang untuk mengekstrak data tekstual yang relevan.

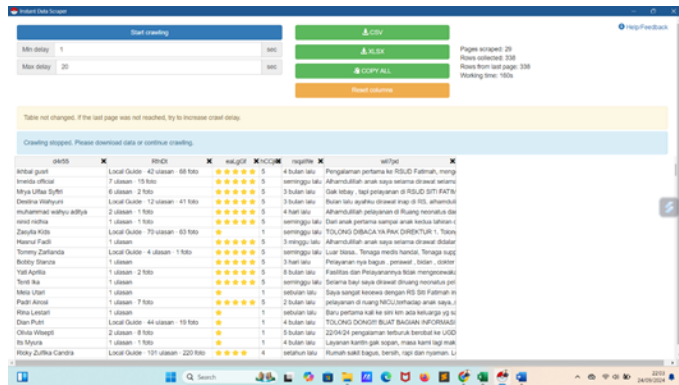


Gambar 1 Langkah metodologis

Langkah-langkah metodologis yang dilakukan selama penelitian ini digambarkan sebagai berikut:

1. Pengumpulan Data

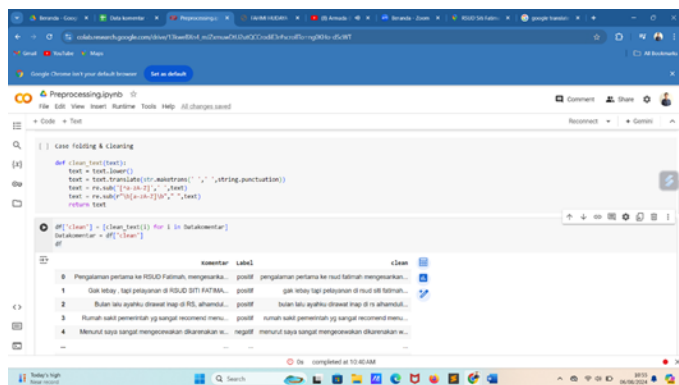
Tahap awal penelitian melibatkan pengumpulan data sistematis yang berasal dari ulasan buatan pengguna yang tersedia di platform Google Maps, dicapai melalui pemanfaatan ekstensi browser Instant Data Scraper, yang memfasilitasi proses ekstraksi data yang efisien dan otomatis.



Gambar 2 Scraping data

2. Teks preprocessing

Tahap selanjutnya penelitian ini mencakup serangkaian proses diantaranya mencakup case folding, cleaning, tokenization, stopwords removal, dan stemming yang bertujuan untuk membersihkan dan menyederhanakan teks.



Gambar 3 Hasil case folding dan cleaning

3. Pembobotan Kata

Fase berikutnya melibatkan penerapan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF), yang berfungsi untuk menetapkan bobot yang dapat diukur untuk setiap kata individu dalam kumpulan data, sehingga memungkinkan pemahaman bernuansa tentang pentingnya setiap istilah relatif terhadap keseluruhan korpus.

```

import numpy as np
import pandas as pd
import sklearn
import sklearn.feature_extraction.text
import sklearn.metrics
import sklearn.pipeline
import sklearn.preprocessing
import sklearn.model_selection

# mengimport library yang dibutuhkan
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, precision_score, recall_score, f1_score

# TF-IDF
tfidf = TfidfVectorizer()
x = tfidf.fit_transform(df['ulasan'])

# split data
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.1, random_state=42)

# Naive Bayes

```

Gambar 4 Proses TF-IDF

4. Klasifikasi

Dalam langkah penting ini, model klasifikasi Naïve Bayes digunakan untuk secara sistematis mengkategorikan sentimen yang diungkapkan dalam ulasan ke dalam klasifikasi positif dan negatif yang berbeda, sehingga memfasilitasi interpretasi terstruktur dari nada emosional yang disampaikan oleh pengguna.

5. Evaluasi

Tahap terakhir, kinerja model klasifikasi dinilai secara ketat melalui penerapan berbagai metrik evaluatif, termasuk akurasi, presisi, recall, dan skor F1, yang secara kolektif memberikan gambaran komprehensif tentang efektivitas model sehingga secara akurat membedakan sentimen yang ada dalam data ulasan.

```

# Akurasi, presisi, recall, f1 score dan confusion matrix
print('akurasi:', accuracy)
print('hasil presisi', precision_score(y_test, prediction, average='binary', pos_label=1))
print('hasil recall', recall_score(y_test, prediction, average='binary', pos_label=1))
print('hasil f1 score', f1_score(y_test, prediction, average='binary', pos_label=1))
print('hasil confusion matrix', confusion_matrix(y_test, prediction))
print('hasil classification_report', classification_report(y_test, prediction))

akurasi: 0.8423529320987
hasil presisi: 0.8113333333333334
hasil recall: 0.8
hasil f1 score: 0.8055555555555556
hasil confusion matrix:
[[ 0  3]
 [ 0 38]]
-----
              precision    recall  f1-score   support

0             1.00      0.25      0.60         4
1             0.81      1.00      0.90        38

accuracy: 0.82
macro avg: 0.92      0.62      0.65        42
weighted avg: 0.87      0.84      0.80        42

```

Gambar 5 Hasil proses evaluasi model

3. HASIL DAN ANALISIS

Hasil dari penelitian ini menunjukkan bahwa mayoritas ulasan yang diterima oleh RSUD Siti Fatimah Palembang bersifat positif, mengindikasikan adanya tingkat kepuasan yang cukup tinggi dari masyarakat terkait pelayanan yang diberikan oleh rumah sakit. Ulasan-ulasan ini sering kali memuji aspek-aspek seperti fasilitas yang memadai, kebersihan rumah sakit, serta profesionalisme tenaga medis. Hal ini menunjukkan bahwa RSUD Siti Fatimah telah mampu memenuhi ekspektasi pasien dalam beberapa aspek penting, yang secara tidak langsung memperkuat citra positif rumah sakit di mata masyarakat.

Namun demikian, ada sejumlah ulasan negatif yang memfokuskan pada beberapa kelemahan dalam pelayanan. Ulasan negatif ini umumnya berkaitan dengan waktu tunggu yang lama, baik untuk proses administrasi, pelayanan medis, serta sikap tenaga medis yang dinilai kurang ramah dan tanggap terhadap keluhan pasien. Masalah ini menjadi perhatian penting, karena persepsi negatif dapat mempengaruhi kepuasan pasien secara keseluruhan dan menurunkan reputasi institusi kesehatan di mata masyarakat.

Metode Naïve Bayes Classifier digunakan untuk mengklasifikasi ulasan dalam kategori positif dan negatif. Naïve Bayes terbukti sangat efektif dalam mengidentifikasi pola dalam teks ulasan, yang kemudian digunakan untuk mengukur sentimen secara keseluruhan. Penggunaan metode ini melibatkan beberapa tahap

penting, seperti text preprocessing mencakup pembersihan data, tokenisasi, dan pembobotan menggunakan TF-IDF. Proses ini membantu memastikan data yang digunakan dalam klasifikasi relevan dan bersih dari noise.

Setelah tahap preprocessing, model Naïve Bayes dilatih menggunakan data ulasan yang telah dilabeli, dan hasil klasifikasi dievaluasi menggunakan confusion matrix. Evaluasi menunjukkan tingkat akurasi model cukup tinggi, dengan nilai precision, recall, dan F1-score yang memadai untuk kedua kategori sentimen. Tingkat akurasi ini menunjukkan bahwa Naïve Bayes mampu mengklasifikasikan ulasan dengan baik, terutama dalam membedakan ulasan positif dan negatif berdasarkan pola kata yang muncul.

Dalam pengukuran precision dan recall, analisis menunjukkan bahwa model ini lebih baik dalam mengidentifikasi ulasan positif dibandingkan ulasan negatif. Hal ini dapat diakibatkan oleh dominasi ulasan positif dalam dataset, yang menyebabkan bias dalam pelatihan model. Meski demikian, model masih mampu mendeteksi sejumlah ulasan negatif dengan cukup baik, yang menjadi dasar bagi identifikasi aspek-aspek layanan yang perlu ditingkatkan.

Tabel 1 Hasil nilai precision, recall dan f-1 score

Jenis Klasifikasi	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
Positif	0,83	1,00	0,91
Negatif	1,00	0,25	0,40

Hasil perhitungan Naive Bayes dapat disimpulkan sebagai berikut:

Untuk Klasifikasi Positif:

- Precision: 0,83, yang berarti 83% dari prediksi positif yang dilakukan oleh model benar-benar positif.
- Recall: 1,00, yang menunjukkan model berhasil menangkap semua data positif yang sebenarnya.
- F-1 Score: 0,91, yang merupakan harmonisasi antara precision dan recall. Angka ini menunjukkan keseimbangan yang sangat baik antara keduanya.

Untuk Klasifikasi Negatif:

- Precision: 1,00, yang berarti setiap prediksi negatif yang dilakukan oleh model adalah benar.
- Recall: 0,25, yang mengindikasikan bahwa model hanya berhasil menemukan 25% dari data negatif yang sebenarnya.
- F-1 Score: 0,40, yang menunjukkan bahwa keseimbangan antara precision dan recall untuk klasifikasi negatif cukup rendah.

Secara keseluruhan, hasil analisis ini memberikan gambaran yang komprehensif tentang persepsi masyarakat terhadap RSUD. Temuan ini memberikan wawasan penting bagi manajemen rumah sakit untuk memahami kekuatan dan kelemahan layanan, serta area yang perlu diperbaiki, seperti efisiensi waktu tunggu dan sikap tenaga medis. Dengan melakukan perbaikan tersebut, RSUD diharapkan dapat meningkatkan kualitas pelayanan yang lebih baik dan responsif terhadap kebutuhan masyarakat.

Selain itu, temuan ini juga menunjukkan potensi metode Naïve Bayes Classifier dalam analisis sentimen di bidang pelayanan kesehatan. Keberhasilan metode ini dalam mengklasifikasikan ulasan publik berdasarkan sentimen positif dan negatif memperlihatkan bahwa teknik ini dapat diterapkan di berbagai sektor layanan publik lainnya untuk mendapatkan pemahaman yang lebih baik mengenai persepsi masyarakat.

4. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen terhadap ulasan masyarakat mengenai RSUD Siti Fatimah Palembang menggunakan metode Naïve Bayes Classifier. Berdasarkan nilai Klasifikasi positif Precision menunjukkan 83%, Recall 1,00 dan F-1 Score 0,91 yang berarti Model ini menunjukkan kinerja yang sangat baik dalam mengidentifikasi sentimen positif. Namun, model kurang efektif dalam mengidentifikasi sentimen negatif, dengan recall yang sangat rendah.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada segenap pihak yang telah memberi bimbingan dan dukungan kepada penulis sehingga tulisan ini bisa selesai dan bermanfaat.

REFERENSI

- [1] A. A. Firdaus, A. Yudhana, and I. Riadi, “DECODE : Jurnal Pendidikan Teknologi Informasi,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 236–245, 2023, [Online]. Available: <http://journal.umkendari.ac.id/index.php/decode>
- [2] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and ..., “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment ...,” ... *Indones. J. ...*, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/195>
- [3] D. Pratmanto and F. F. D. Imaniawan, “Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors,” *Comput. Sci.*, 2023, [Online]. Available: <http://103.75.24.116/index.php/co-science/article/view/1917>
- [4] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, “Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naive Bayes (Studi Kasus: Facebook Page BerryBenka),” *4th Appl. Bus. Eng. Conf.* researchgate.net, 2016. [Online]. Available: https://www.researchgate.net/profile/Muhammad-Ihsan-Zul/publication/311107797_Analisis_Sentimen_Terhadap_Toko_Online_di_Sosial_Media_Menggunakan_Metode_Klasifikasi_Naive_Bayes_Studi_Kasus_Facebook_Page_BerryBenka/links/583e787b08ae8e63e6158698/Analisis-Se
- [5] D. Darwis, N. Siskawati, and Z. Abidin, “PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [6] P. A. Permatasari, L. Linawati, and L. Jasa, “Survei Tentang Analisis Sentimen Pada Media Sosial,” *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, p. 177, 2021, doi: 10.24843/mite.2021.v20i02.p01.
- [7] J. Fitriana and J. Triloka, “Prediksi Stok Bahan Baku Minuman Pada Cafe Kiyu Menggunakan Machine Learning (Studi Kasus Pada: Cafe Kiyu),” *Tek. J. Ilm. Bid. Ilmu ...*, 2023, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/7598>
- [8] N. R. Nuraeda, M. Liebenlito, and T. E. Sutanto, “Explainable Sentiment Analysis pada Ulasan Aplikasi Shopee Menggunakan Local Interpretable Model-agnostic Explanations,” *Indones. J. ...*, 2024, [Online]. Available: <http://ijcs.net/ijcs/index.php/ijcs/article/view/3870>
- [9] A. U. Haspriyanti and ..., “Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearst Neighbor,” ... *Syst. Artif. ...*, 2021, [Online]. Available: <https://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/view/17>
- [10] D. Suhardinata, A. K. Ningsih, and ..., “Klasterisasi Data Penduduk Untuk Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan K-Means (Studi Kasus: Desa Tanimulya Bandung Barat),” ... *(International J. ...*, 2023, [Online]. Available: <http://ijespgjournal.org/index.php/ijespg/article/view/55>
- [11] J. Fitriana and J. Triloka, “Menu Sales Prediction at Kiyu Café Using Machine Learning,” *J. Comput. Networks ...*, 2024, [Online]. Available: <https://jurnal.itscience.org/index.php/CNAPC/article/view/3556>
- [12] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and ..., “Comparison Of Classification Algorithm And Feature Selection In Bitcoin Sentiment Analysis,” *Jurnal Teknik ...* researchgate.net, 2022. [Online]. Available: https://www.researchgate.net/profile/Dede-Kurniadi/publication/362093077_Comparison_of_Classification_Algorithm_and_Feature_Selection_in_Bitcoin_Sentiment_Analysis/links/62d68339de26053523a6256b/Comparison-of-Classification-Algorithm-and-Feature-Selection
- [13] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN,” *J. KomtekInfo*, 2023, [Online]. Available: <https://jkomtekinfo.org/ojs/index.php/komtekinfo/article/view/330>
- [14] H. Ardian and S. Kosasi, “Analisis Sentimen Pada Review Produk Kosmetik Bahasa Indonesia Dengan Metode Naive Bayes,” *ENTER*, 2019, [Online]. Available: <http://www.sisfotenika.stmikpontianak.ac.id/index.php/enter/article/view/862>
- [15] A. Pangestu, Y. T. Arifin, and R. A. Safitri, “Analisis Sentimen Review Publik Pengguna Game Online Pada Platform Steam Menggunakan Algoritma Naive Bayes,” *JATI (Jurnal Mhs. Tek. ...*,

- 2023, [Online]. Available: <https://www.ejournal.itn.ac.id/index.php/jati/article/view/8829>
- [16] A. A. Rahman, S. S. Prasetyowati, and ..., "Performance analysis of the imbalanced data method on increasing the classification accuracy of the machine learning hybrid method," *JUPI (Jurnal ...)*, 2023, [Online]. Available: <http://www.jurnal.stkipppgritulungagung.ac.id/index.php/jipi/article/view/3286>
- [17] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28-38, 2022, doi: 10.35969/interkom.v16i4.192.
- [18] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC," *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42-48, 2023, doi: 10.54914/jtt.v9i1.609.