JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.478-489

COMPARISON OF THE PERFORMANCE OF SVM, RANDOM FOREST, AND NEURAL NETWORK ALGORITHMS IN SENTIMENT ANALYSIS OF OPENAI APPLICATION REVIEWS ON THE GOOGLE PLAY STORE

Ahmad Latif *1), Fandy setyo utomo2), Taqwa Hariguna3)
1. Magister of Computer Science of Universitas Amikom Purwokerto, Indonesia
2. Magister of Computer Science of Universitas Amikom Purwokerto, Indonesia
3. Magister of Computer Science of Universitas Amikom Purwokerto, Indonesia

Article Info
Keywords: Sentiment Analysis; Machine Learning; Support Vector Machine (SVM); Random Forest (RF); Neural Network (NN)

Article history:
Received 7 May 2025
Revised 3 June 2025
Accepted 5 July 2025
Available online 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7793

* Corresponding author.
Corresponding Author
E-mail address:
23MA41D033@students.amikompurwokerto.ac.id

ABSTRACT
This study compares the performance of three machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN)—in sentiment analysis of user reviews for the OpenAI application on the Google Play Store. The primary objective of this study is to evaluate the effectiveness of each algorithm in classifying user reviews into three sentiment categories: positive, negative, and neutral. The dataset used consists of user reviews of the OpenAI application, collected directly from the Google Play Store. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The results indicate that the Neural Network algorithm achieved the best overall performance in terms of accuracy and F1-score. SVM demonstrated competitive performance, particularly in classifying positive and neutral sentiments, while Random Forest showed an advantage in terms of precision but performed lower overall, especially in classifying negative sentiments. Therefore, the Neural Network is considered the most effective algorithm for sentiment analysis tasks in this study.

.

I. INTRODUCTION

Application reviews on digital platforms such as the Google Play Store provide valuable insights into user experiences. Through both positive and negative reviews, users convey feedback regarding necessary features and areas that require improvement. Previous studies have shown that approximately 49% of application developers consider constructive feedback from users when implementing updates, with many adopting changes based on the reviews received [1] [2] . This feedback not only supports improvements in application functionality but also reflects users' emotional responses to their usage experience [3].

User reviews often contain critical information that can be used to enhance application quality. For instance, reviews may reveal the presence of bugs or highlight the need for features that have not yet been implemented [4]. [1] emphasized that sentiment analysis of these reviews is essential for understanding how user experience influences the software development process. In addition, user behavior can be analyzed to assess satisfaction levels and identify barriers faced during application usage, as illustrated in a study on hypnotherapy applications [5]. With the advancement of technology—particularly in the field of artificial intelligence (AI)—OpenAI has emerged as a major player, offering popular AI-based products such as the GPT model. Nevertheless, one of the main challenges in AI application development lies in understanding user perceptions, as reflected in their reviews on platforms like the Google Play Store. Therefore, sentiment analysis becomes crucial for evaluating the strengths and weaknesses of applications from the user's perspective.

In conducting sentiment analysis, machine learning algorithms have experienced significant development. A variety of techniques have been employed to process and classify text based on the sentiment it contains [6]. One widely used method is Support Vector Machine (SVM), which has proven effective in solving classification problems, particularly when a clear margin exists between two classes. Research has shown that SVM performs well for text classification, especially on high-dimensional and sparse data [7][8]

In addition to SVM, Random Forest is another algorithm commonly applied in sentiment analysis. As an ensemble method, it combines multiple decision trees to improve the stability and accuracy of predictions [9]. Random Forest is highly adaptive to complex and heterogeneous data, making it more reliable than single-method models. Meanwhile, Neural Network-based algorithms, especially deep learning models such as Recurrent Neural Networks (RNN) and Transformers, offer exceptional capabilities in identifying complex patterns in text data. For example, the BERT model (Bidirectional Encoder Representations from Transformers) has been shown to excel in understanding linguistic context and nuances, making it highly effective for sentiment analysis across various domains [10][11] [12]. BERT is not only a text classification model but also employs an advanced architecture capable of capturing sentence-level context deeply, which makes it highly relevant for a wide range of Natural Language Processing (NLP) tasks [13].

Reviews of the OpenAI application on the Google Play Store often use inconsistent language, mixing formal and informal styles, and include slang, abbreviations, and emoticons that are difficult for automated systems to interpret. Additionally, the presence of ambiguity, sarcasm, and implicit emotional expressions makes the sentiment identification process more complex, as the model must be able to deeply understand the context to provide accurate classification. Furthermore, spelling errors and irregular sentence structures also pose challenges during preprocessing and model development, requiring special approaches to ensure more effective sentiment analysis.

Regarding the imbalance of sentiment classes in the dataset, reviews are generally dominated by positive sentiments, while reviews with negative or neutral sentiments are far fewer in number. This imbalance causes models to tend to favor predictions for the majority class, which is positive sentiment, resulting in suboptimal performance in classifying minority sentiments. This also affects model evaluation, where accuracy metrics alone are insufficient to fully represent overall performance. Therefore, the use of other metrics such as precision, recall, and F1-score is essential to provide a more comprehensive and fair evaluation across all sentiment classes.

In sentiment analysis of OpenAI application reviews on the Google Play Store, algorithm selection is a crucial factor in achieving accurate and efficient results. Support Vector Machine (SVM) is widely used in NLP due to its ability to handle high-dimensional data and prevent overfitting on moderately sized datasets. Random Forest, an ensemble method based on decision trees, is highly effective in managing noisy data and correlated features, while also providing a good level of interpretability. On the other hand, Neural Networks can identify non-linear patterns and complex interactions among text features by utilizing word representations such as word embeddings, often demonstrating superior performance in sentiment classification tasks.

The application of neural networks in sentiment analysis has demonstrated outstanding performance across tasks such as named entity recognition, text classification, and emotion prediction [7] [14]. Approaches that combine Long Short-Term Memory (LSTM) with Transformer architectures are often recommended for sequential data, as they allow the model to retain important long-term information in text processing [15] [16]. In this era of big data, integrating traditional approaches such as SVM and Random Forest with cutting-edge deep learning models has become increasingly critical to improving the accuracy and efficiency of sentiment analysis. Choosing the right algorithm plays a key role in producing accurate and reliable analyses [17][18]. Furthermore, ensemble learning approaches that combine multiple algorithms have proven effective in building more robust models by leveraging the precision of SVM, the reliability of Random Forest, and the pattern recognition power of Neural Networks [19].

Sentiment analysis provides valuable insights that support strategic decision-making, such as setting priorities for improvements, developing new features, and adjusting the user interface, ultimately enhancing user satisfaction and loyalty. Additionally, in AI-based applications, these analysis results can be used to train models to be more sensitive to user preferences, for example, by adapting the chatbot’s communication style to make interactions more natural and satisfying. Furthermore, regular monitoring of negative sentiment helps developers anticipate potential declines in the app’s reputation and take preventive actions to maintain user trust through more effective and transparent communication.

In the context of OpenAI applications on the Google Play Store, this study aims to develop and evaluate various sentiment analysis models to determine the most effective approach for classifying user reviews based on sentiment. By utilizing review data from the Google Play Store, this research is expected to provide a clearer understanding of user perceptions of OpenAI applications and to identify areas for improvement to enhance user satisfaction.

II. METHOD

The approach used in this study is a quantitative approach, aimed at analyzing the sentiment of user reviews of the OpenAI application on the Google Play Store platform. In this study, three machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN)—were employed to classify

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.478-489

review sentiments into three categories: positive, neutral, and negative. To obtain accurate and reliable results, the study was conducted through several key stages, including a literature review, data collection, data preprocessing, and evaluation and comparison of model performance It can be seen in Figure 1.

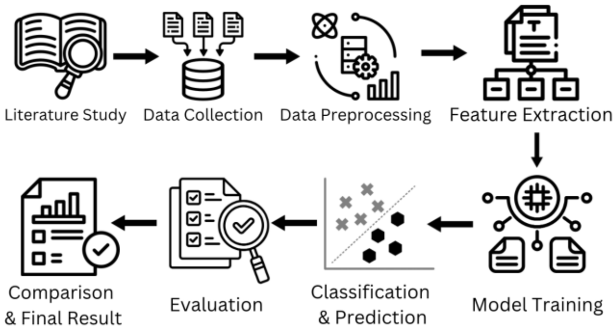


Figure 1 Research Methodology

The first stage of this study was a literature review, aimed at understanding the concept of sentiment analysis and the methods that have been employed in previous research. Sentiment analysis is a technique within Natural Language Processing (NLP) that seeks to extract, identify, and categorize opinions in text based on the sentiments they express [20]. This review also included various machine learning approaches, including text processing techniques, as well as the strengths and limitations of the algorithms used in this study.

In conducting web scraping on the Google Play Store, tools such as BeautifulSoup, Selenium, and google-play-scraper (for Python or Node.js) are commonly used. Selenium is particularly useful for handling content that is dynamically loaded via JavaScript. The scraping process involves accessing the application page, collecting review data including text, ratings, dates, and usernames, and addressing the infinite scroll feature by performing automatic scrolling or repeatedly calling internal APIs. The collected data is then stored in a format that is easy to process, such as CSV, followed by a cleaning process to remove duplicates, empty data, and spam. To ensure the data obtained is valid and complete, the scraper must download all reviews through pagination or full scrolling, ensure the data originates from the official page, avoid duplicate sources, and verify that the dates and ratings match the desired time period.

The next stage was data collection, where review data were obtained from the Google Play Store using web scraping techniques with the BeautifulSoup and Scrapy libraries in the Python programming language. A total of 10,000 recent reviews written in Indonesian were collected as the primary dataset. The use of web scraping for data collection has been widely applied in similar studies, such as the work by [21], which gathered over 5,000 social media reviews to analyze sentiment toward government policies.

Once the data were collected, a preprocessing stage was conducted to clean the text from irrelevant elements. This process consisted of several steps, including case folding (converting all letters to lowercase), tokenization (breaking the text into individual words), stopword removal (eliminating common words that carry little semantic weight in sentiment analysis), and stemming (reducing words to their root form) using the Sastrawi library. This preprocessing step is critical for improving model accuracy in identifying sentiment patterns in text [22].

Following preprocessing, feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) to convert text into numerical representations. TF-IDF is frequently used in text analysis because it assigns higher weights to words that appear frequently in a document but are rare in other documents, thereby enhancing model performance in sentiment detection [23].

The resulting numerical vectors were then split into training data (80%) and testing data (20%) using stratified sampling to ensure balanced sentiment distribution across subsets. Model training was then conducted using the three main algorithms: SVM, Random Forest, and Neural Network.

Support Vector Machine (SVM) is a classification algorithm that works by finding the optimal hyperplane that separates data into distinct classes with the maximum margin [24]. This algorithm is highly effective for high-dimensional data and is widely used in text classification due to its ability to clearly distinguish between features. Random Forest, on the other hand, is an ensemble method that builds multiple decision trees and aggregates their results to improve accuracy and reduce overfitting. In sentiment analysis, Random Forest constructs a classification model based on features extracted from text. Each decision tree predicts a sentiment class, and the final prediction is determined by majority vote. Key advantages of Random Forest include its ability to handle large datasets, its flexibility with unstructured data, and its capacity to assess feature importance, helping identify which elements most influence classification outcomes[25].

Meanwhile, Neural Networks (NN) utilize artificial neural architectures to detect complex patterns in text [26]. This study used a Multilayer Perceptron (MLP) model with hidden layers to process deeper information and enhance classification accuracy. The MLP model is relatively interpretable and effective in identifying the main factors that determine sentiment in user reviews.

After training, the models were tested using the test data to evaluate their ability to accurately categorize sentiment in the reviews. The models' performance was assessed using four main evaluation metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions; precision assesses the model's ability to avoid false positives; recall evaluates the model's capacity to correctly identify all relevant instances; and the F1-score balances precision and recall [25].

The final stage of this study involved comparing the results and conducting a conclusive analysis. The performance of the three algorithms was compared based on their evaluation metrics to determine which method was most effective in classifying the sentiment of Indonesian-language reviews of the OpenAI application. Previous studies have shown that SVM often outperforms other methods in sentiment analysis tasks [27]. Thus, this research is expected to provide deeper insights into user perceptions of AI-based applications and contribute to the development of sentiment analysis technologies for the Indonesian language.

III. RESULTS AND DISCUSSION

In this study, the first stage involved collecting user review data for the OpenAI application available on the Google Play Store platform. A total of 10,000 recent reviews written in Indonesian were collected using web scraping techniques, implemented through the BeautifulSoup and Scrapy libraries in the Python programming language. The collected data included several key pieces of information, such as review text, user ratings, review dates, and review sources, which were subsequently used as the foundation for the sentiment analysis process It can be seen in Figure 2 and 3.

```
[3] from google_play_scraper import app, reviews
    import pandas as pd
    import datetime
```

```
from google_play_scraper import reviews, Sort

app_id = 'com.openai.chatgpt'

def get_reviews(app_id, lang='id', count=10000, sort=Sort.NEWEST, filter_score_with=None, filter_device_with=None, continuation_token=None):
    try:
        result, continuation_token = reviews(
            app_id,
            lang=lang,
            country='id',
            sort=sort,
            count=count,
            filter_score_with=filter_score_with,
            filter_device_with=filter_device_with,
            continuation_token=continuation_token
        )
        return result, continuation_token
    except Exception as e:
        print("Error:", e)
        return None, None

reviews, continuation_token = get_reviews(app_id)

if reviews is not None:
    print("Jumlah Ulasan:", len(reviews))
    if len(reviews) > 0:
        print("Contoh ulasan:")
        print(reviews[0])
else:
    print("Tidak ada ulasan ditemukan.")
```

Figure 2. Dataset Scraping Process

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.478-489

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Review ID    10000 non-null  object
 1   Username     10000 non-null  object
 2   Rating       10000 non-null  int64
 3   Review Text  10000 non-null  object
 4   Date         10000 non-null  object
dtypes: int64(1), object(4)
memory usage: 390.8+ KB
```

	Review ID	Username	Rating	Review Text	Date
0	d93e2e21-1047-4dd7-8c8a-0714ecda0a7b	Pengguna Google	5	mantab oke jossss	2025-05-02 02:08:11
1	f758462d-2459-4477-a2c8-2d130b6122cf	Pengguna Google	5	BAGUS BANGET PUAS BANGET	2025-05-02 01:53:59
2	7b8fe1f3-f1ad-4499-9713-91d6b4aa4a37	Pengguna Google	2	sekarang terlalu Lambat untuk membuat gambar	2025-05-02 01:53:46
3	ffacfd27-3333-4d8b-bc6e-d8abd1ae4db7	Pengguna Google	5	mantap luar biasa, sangat membantu	2025-05-02 01:49:50
4	db6e4044-3358-460a-9238-82c50547e88f	Pengguna Google	5	membantu saya setiap ada masalah	2025-05-02 01:43:22

Figure 3. Data Scraping Results

After the data has been successfully collected, a data preprocessing stage is carried out to ensure the quality of the data before sentiment classification is performed. The purpose of preprocessing is to clean and prepare the text data so that it can be effectively processed by machine learning models [28]. The following are the data preprocessing steps used in this study, which include:

1. Cleaning

   Cleaning is the process of cleaning the 'Review Text' column in the DataFrame by removing unwanted elements such as URLs, HTML, emojis, symbols, and numbers, so that the text becomes cleaner and ready for further text analysis It can be seen in Figure 4.

Review Text	cleaning
mantab oke jossss	mantab oke jossss
BAGUS BANGET PUAS BANGET	BAGUS BANGET PUAS BANGET
sekarang terlalu Lambat untuk membuat gambar	sekarang terlalu Lambat untuk membuat gambar
mantap luar biasa, sangat membantu	mantap luar biasa sangat membantu
membantu saya setiap ada masalah	membantu saya setiap ada masalah
sangat bagus aplikasinya dan membantu🙏	sangat bagus aplikasinya dan membantu
bagus sekali 10/100	bagus sekali
Aplikasi ini bagus dan sangat membantu sekali	Aplikasi ini bagus dan sangat membantu sekali
ok	ok
sangat membantu	sangat membantu

Figure 4. Data Cleaning

2. Case Folding

   Case folding is the process of converting text to lowercase, removing URLs from the content column, eliminating mentions, deleting hashtags, removing special characters, removing punctuation, and eliminating extra spaces [29] It can be seen in Figure 5.

cleaning	case_folding
mantab oke jossss	mantab oke jossss
BAGUS BANGET PUAS BANGET	bagus banget puas banget
sekarang terlalu Lambat untuk membuat gambar	sekarang terlalu lambat untuk membuat gambar
mantap luar biasa sangat membantu	mantap luar biasa sangat membantu
membantu saya setiap ada masalah	membantu saya setiap ada masalah

Figure 5. Case Folding Data

3. Tokenization

   Tokenization is the process of dividing text or sentences into smaller units called 'tokens'. Tokens can be words, phrases, or other entities such as numbers or symbols. The main purpose of tokenization is to facilitate text analysis and processing [30] It can be seen in Figure 6.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.478-489

tokenize	case_folding
[mantab, oke, jossss]	mantab oke jossss
[bagus, banget, puas, banget]	bagus banget puas banget
[sekarang, terlalu, lambat, untuk, membuat, ga...	sekarang terlalu lambat untuk membuat gambar
[mantap, luar, biasa, sangat, membantu]	mantap luar biasa sangat membantu
[membantu, saya, setiap, ada, masalah]	membantu saya setiap ada masalah

Figure 6. Tokenization Data

4. Stopword Removal

Filtering (Stopword Removal) is the stage where important words are extracted from the tokenized results using a stoplist algorithm to discard less important words. Stopwords are common words that frequently appear in the text and are considered to have no meaningful contribution. By removing stopwords from the text, we can focus on more important words [31] It can be seen in Figure 7.

tokenize	stopword removal
[mantab, oke, jossss]	[mantab, oke, jossss]
[bagus, banget, puas, banget]	[bagus, banget, puas, banget]
[sekarang, terlalu, lambat, untuk, membuat, ga...	[lambat, gambar]
[mantap, luar, biasa, sangat, membantu]	[mantap, membantu]
[membantu, saya, setiap, ada, masalah]	[membantu]

Figure 7. Stopword Removal Data

5. Stemming

Stemming is a method in natural language processing that converts words to their root form, aiding in the preprocessing of text, words, and documents for text normalization [32] It can be seen in Figure 8.

stopword removal	stemming_data
[mantab, oke, jossss]	mantab oke jossss
[bagus, banget, puas, banget]	bagus banget puas banget
[lambat, gambar]	lambat gambar
[mantap, membantu]	mantap bantu
[membantu]	bantu

Figure 8. Stemming Data

After preprocessing, the data ready for classification consists of 6,104 reviews, after removing empty or irrelevant reviews It can be seen in Figure 9.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.478-489

```
<class 'pandas.core.frame.DataFrame'>
Index: 6104 entries, 0 to 9998
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Date               6104 non-null   object
 1   Username           6104 non-null   object
 2   Rating             6104 non-null   int64
 3   Review Text        6104 non-null   object
 4   cleaning           6104 non-null   object
 5   case_folding       6104 non-null   object
 6   normalisasi        6104 non-null   object
 7   tokenize           6104 non-null   object
 8   stopword removal   6104 non-null   object
 9   stemming_data      6104 non-null   object
dtypes: int64(1), object(9)
memory usage: 653.6+ KB
```

Figure 9. Number of Data Ready

Next, a sentiment distribution analysis was conducted to determine the proportion of each sentiment category in the dataset. The analysis revealed the sentiment analysis results of a number of tweets with three main sentiment categories: Neutral, Positive, and Negative. In the Neutral category, there were 2,629 tweets, representing 43.07% of the total data. The Positive category recorded 2,567 tweets, or 42.05%, while the Negative category had 908 tweets, amounting to 14.88%. This chart provides a clear picture of the sentiment distribution in the dataset, with neutral and positive sentiments dominating, while negative sentiment makes up only a small portion of the total tweets analyzed It can be seen in Figure 10.

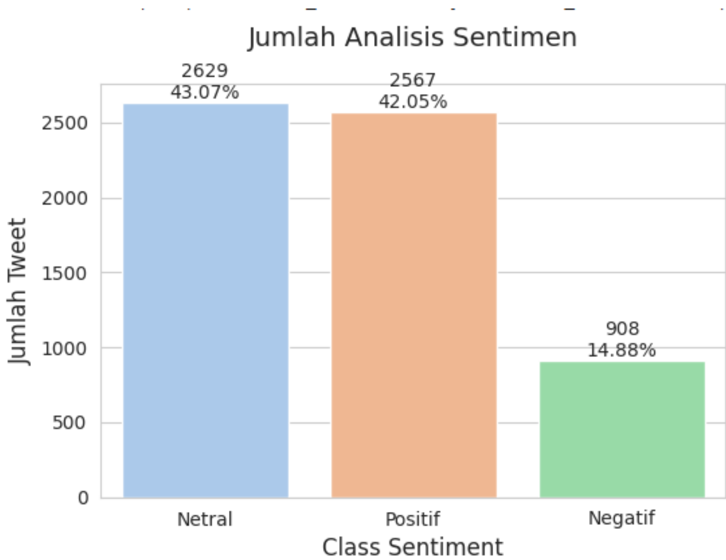


Figure 10. Sentiment Distribution of User Reviews

After analyzing the data distribution, the text is converted into a numerical format using the TF-IDF (Term Frequency-Inverse Document Frequency) method. This technique gives more weight to words that frequently appear in a review but are rarely found in other reviews, thus enhancing the model's ability to capture the sentiment meaning from the text.

The dataset is then split into 80% for training data and 20% for testing data, using stratified sampling techniques to ensure that the sentiment category distribution remains balanced in each data subset. In this stage, the processed data is used to train and test three sentiment classification models: Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN) It can be seen in Figure 11.

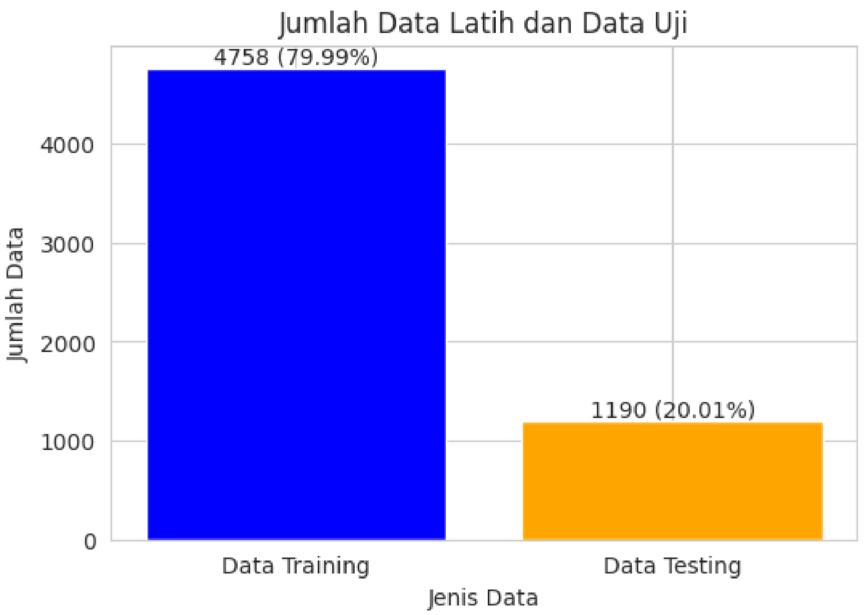


Figure 11. Training and Testing Data Split

To clarify the patterns of words that frequently appear in user reviews, a WordCloud visualization is created from the dataset It can be seen in Figure 12.

Figure 12. General WordCloud

Words such as help, good, application, like, and really helpful appear larger, indicating that these words are frequently used in comments or reviews. Overall, this suggests a positive sentiment toward the application or service, with many users feeling helped and satisfied with the functionality or performance of the application It can be seen in Figure 13.

Figure 13. Sentiment WordCloud

WordClouds depicting words that frequently appear in reviews of the application with positive and negative sentiments. In the positive sentiment, words such as 'good', 'help', 'application', and 'yes' dominate, indicating that users are very satisfied with the performance of the application, especially with features like 'chat GPT' and 'photos'. Meanwhile, in the negative sentiment, although the words 'application' and 'help' still appear, words like 'bad', 'wrong', 'lacking', and 'limit' reflect complaints about technical issues and feature limitations that hinder the user experience.

In addition to the comparison of evaluation metrics, a confusion matrix analysis is also conducted to gain a deeper understanding of the performance of each model in classifying positive, neutral, and negative sentiments It can be seen in Figure 14.

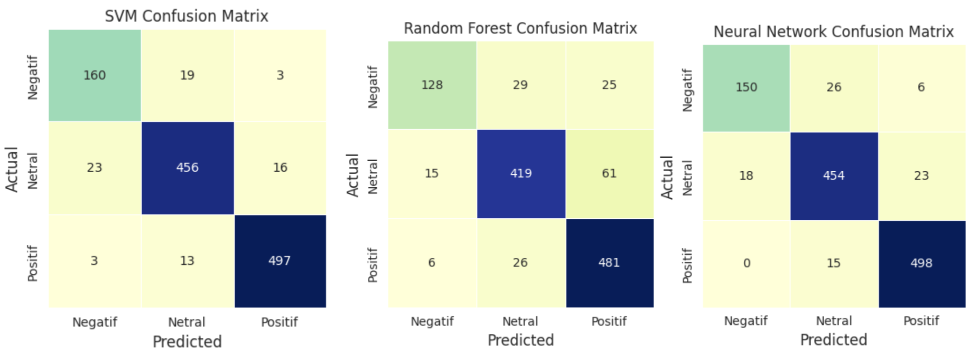


Figure 14. Confusion Matrix Results for SVM, RF, and NN

The results of the confusion matrix show that the Neural Network is the most effective model among the three tested models, with excellent classification ability and few errors, particularly in the Positive category. SVM also demonstrates good performance, although there are some errors in classifying the Neutral category. On the other hand, Random Forest shows poorer results, with more errors in the Neutral and Negative categories, although it performs reasonably well in the Positive category. Overall, the Neural Network provides the best results in sentiment classification compared to SVM and Random Forest It can be seen in Table 1.

TABLE I.
COMPARISON BETWEEN MODELS

Model	Sentiment	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
SVM	Negative	86.0	87.9	87.0	93.5
	Neutral	93.4	92.1	92.8	
	Positive	96.3	96.9	96.6	
Random Forest	Negative	85.9	70.3	77.3	86.4
	Neutral	88.4	84.6	86.5	
	Positive	84.8	93.8	89.1	
Neural Network	Negative	89.3	82.4	85.7	92.6
	Neutral	91.7	91.7	91.7	
	Positive	94.5	97.1	95.8	

Based on the table above, it can be concluded that the Neural Network provides the best results with an accuracy of 92.6% and very high values for precision, recall, and f1-score in all sentiment categories. SVM also shows good performance with an accuracy of 93.5%, with very good results in the Positive and Neutral categories. Random Forest has a lower performance with an accuracy of 86.4%, especially in the Negative category, where recall and f1-score show poorer performance.

Support Vector Machine (SVM) excels at separating classes with an optimal margin, making it highly effective for high-dimensional data such as text, especially in classifying clear positive and negative sentiments. However, SVM struggles with ambiguous, neutral, or sarcastic data due to its reliance on a fixed separating margin. To improve its performance, more complex kernels can be applied, parameters like C and gamma can be tuned, contextual features such as word embeddings can be used, and data balancing techniques like SMOTE can be implemented.

Random Forest is known for its robustness to noisy data and correlated features, while also providing interpretability through feature importance. This algorithm demonstrates fairly stable performance across various sentiment categories. Nevertheless, Random Forest tends to be biased toward the majority class if the data is imbalanced and is less effective at capturing complex feature relationships, especially for neutral sentiments. To address these issues, adjustments to the number and depth of trees, the use of resampling techniques to handle data imbalance, and improved feature engineering are necessary.

Neural Networks have strong capabilities in capturing nonlinear patterns and complex interactions between features, making them effective at recognizing implied sentiments and emotional contexts, with good performance across all sentiment classes. However, Neural Networks are prone to overfitting, particularly with small datasets, and are often difficult to interpret (black-box). Errors can also occur if the training data is not sufficiently representative. To improve this, regularization, dropout, data augmentation, pretrained embeddings like BERT or Word2Vec, fine-tuning of models, and enhancing dataset quality are highly recommended.

With the results above, this research contributes to the development of sentiment analysis technology in the Indonesian language. Most previous studies have focused more on analysis in English, while this research demonstrates that machine learning can be effectively applied to sentiment classification in Indonesian, despite challenges such as non-standard word variations and code-mixing. With the increasing use of AI in human interactions (Aristanto et al., 2023), this research can serve as a foundation for the development of more advanced sentiment analysis models for other applications, such as virtual assistants, customer service chatbots, and opinion-based recommendation systems.

Neural Networks in sentiment analysis face challenges such as the need for large training datasets to reduce the risk of overfitting, a "black-box" nature that is difficult to interpret, and higher computational resource consumption and training time compared to SVM and Random Forest. To address these issues and improve accuracy, hybrid or ensemble approaches can be employed by combining Neural Networks with conventional methods. This approach leverages the strengths of each algorithm while reducing classification errors through techniques such as stacking, boosting, or voting, resulting in more accurate and reliable classification

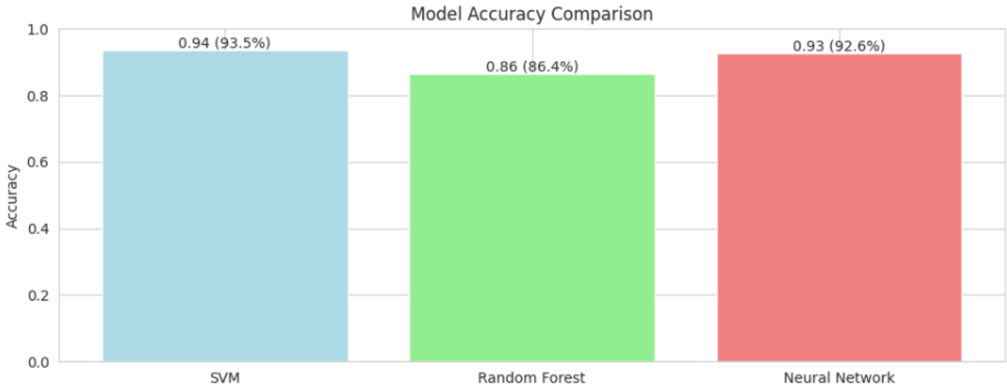
This research plays a crucial role in advancing sentiment analysis technology, particularly for the Indonesian language, which has distinctive linguistic features such as flexible syntactic structure, the use of loanwords, dialectal variations, as well as informal language like slang and abbreviations. These unique characteristics require the development of models that can accurately capture the specific context and meaning of Indonesian. By adapting and evaluating algorithms such as SVM, Random Forest, and Neural Network specifically for this language, the study is expected to produce methods that are more precise and tailored to the needs of the local market. Furthermore, the findings have the potential to accelerate the adoption of more inclusive and adaptive NLP technologies, thereby opening opportunities for developing more effective and personalized AI-based services across various sectors such as e-commerce, public services, and social media.

The results of this study reveal the performance of the SVM, Random Forest, and Neural Network algorithms in conducting sentiment analysis on OpenAI app reviews in the Google Play Store, with specific adaptations for the Indonesian language. Compared to previous studies that mostly focused on English or other international languages, this research makes an important contribution by taking into account the unique linguistic characteristics of Indonesian, such as flexible syntactic structures, the use of loanwords, as well as slang and abbreviations found in user reviews. Unlike earlier studies, such as [33] [34][35], which reported high performance of Neural Networks on English-language data, this study emphasizes that the selection and tuning of algorithms must be adjusted to the language characteristics and local context to achieve optimal results. Furthermore, this research also addresses issues related to handling informal data and dialectal variations, which have been less considered in previous studies, thereby opening opportunities for the development of more inclusive NLP models tailored to the needs of the Indonesian market.

IV. Conclusion

This study compares the performance of three classification algorithms, namely Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN), in analyzing user reviews' sentiment toward the Open AI application on the Google Play Store. Based on the analysis results and comparison of the three sentiment classification models—SVM (Support Vector Machine), Random Forest, and Neural Network—it can be concluded that the Neural Network provides the best results among the three models. This model demonstrates an accuracy of 92.6%, with excellent precision, recall, and f1-score across all sentiment categories (Negative, Neutral, and Positive). The Neural Network is particularly effective in classifying Positive sentiment tweets, with the highest precision and recall values among the other models. On the other hand, SVM also shows solid performance with an accuracy of 93.5%, particularly in classifying Positive and Neutral tweets, though there is a slight decline in the Negative category. Nevertheless, SVM remains a very competent model for sentiment classification tasks.

Random Forest, although showing quite good performance in the Positive category, has an accuracy of 86.4%, which is lower compared to SVM and Neural Network. This model shows poorer performance in the Negative category, with recall and f1-score values much lower than the other two models. Overall, Neural Network proves to be the best choice for sentiment classification, followed by SVM, while Random Forest requires further refinement to improve its performance, particularly in handling the Negative category.



Although this study shows that the Neural Network provides the best accuracy, there are still several challenges that need to be addressed, such as sentiment data imbalance and informal language variations, as well as language mixing in user reviews. Therefore, future research could develop models with deep learning approaches, such as Bidirectional LSTM or Transformer-based models (BERT), which are more effective in understanding the context of natural language. Additionally, aspect-based sentiment analysis could be used to recognize user opinions regarding specific features of the application. Further research could also examine trends in sentiment changes over time, providing deeper insights into user reactions to updates or changes in the Open AI application service.

ACKNOWLEDGMENT

The author would like to express their deepest gratitude to all parties who have provided support and contributions in completing this research. Without their help and guidance, this research would not have been successfully carried out.

REFERENCES

[1] D. Gunawan, "Keputusan pembelian konsumen marketplace Shopee berbasis social media marketing," 2022, Accessed: May 07, 2025. [Online]. Available: https://books.google.com/books?hl=id&lr=&id=K79mEAAAQBAJ&oi=fnd&pg=PA6&dq=Gunawan+et+al.+(2022)+ulasan&ots=SBi0pL5mw9&sig=TBu7Txv8ET-wazhAU0fcLuvaR2A

[2] Z. Yang, Y. Zhang, D. G.-6th I. C. on, and undefined 2022, "Analysis of aquaculture insurance demand based on CNN and sentiment dictionary," *spiedigitallibrary.orgZ Yang, Y Zhang, D Guo6th Int. Conf. Mechatronics Intelligent, 2022•spiedigitallibrary.org*, doi: 10.1117/12.2644664.SHORT.

[3] N. Hicham, H. N.-M. T. and Applications, and undefined 2024, "Improving emotion classification in e-commerce customer review analysis using GPT and meta-ensemble deep learning technique for multilingual system," *Springer*, Nov. 2024, doi: 10.1007/S11042-024-19965-4.

[4] C. Ossai, N. W.-I. for H. and Social, and undefined 2023, "Automatic user sentiments extraction from diabetes mobile apps–An evaluation of reviews with machine learning," *Taylor Fr.*, vol. 48, no. 3, pp. 211–230, 2023, doi: 10.1080/17538157.2022.2097083.

[5] C. Ossai, N. W.-A. I. in Medicine, and undefined 2023, "Sentiments prediction and thematic analysis for diabetes mobile apps using Embedded Deep Neural Networks and Latent Dirichlet Allocation," *Elsevier*, Accessed: May 07, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365723000234

[6] H. Hermawan, F. Mahardika, I. Darmayanti, R. B. B. Sumantri, D. I. S. Saputra, and A. Aminuddin, "New Media as a Tools to Improve Creative Thinking: A Systematic Literature Review," *Proc. - 2023 IEEE 7th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2023*, pp. 64–69, 2023, doi: 10.1109/ICITISEE58992.2023.10404556.

[7] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "Bert: Pre-training of deep bidirectional transformers for language understanding," *aclanthology.org*, pp. 4171–4186, Accessed: May 07, 2025. [Online]. Available: https://aclanthology.org/N19-1423/?utm_campaign=The Batch&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-_m9bbH_7ECE1h3lZ3D61TYg52rKpifVNjL4fvJ85uqggrXsWDBTB7YooFLJeNXHWqhvOyC

[8] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 380–385, 2019, Accessed: May 07, 2025. [Online]. Available: https://arxiv.org/pdf/1903.09588

[9] R. A.-T. E. Letters and undefined 2024, "A Formulation of Investor Sentiment of Cryptocurrencies and Cryptocurrency Futures and Options," *scirp.org*, 2024, doi: 10.4236/tel.2024.142032.

[10] K. Tanaka, R. Sasano, K. T. preprint arXiv:2408.04293, and undefined 2024, "Are Social Sentiments Inherent in LLMs? An Empirical Study on Extraction of Inter-demographic Sentiments," *arxiv.org*, Accessed: May 07, 2025. [Online]. Available: https://arxiv.org/abs/2408.04293

[11] D. Intan *et al.*, "IoT-Based Smart Air Conditioner as a Preventive in the Post-COVID-19 Era: A Review," *journal.umy.ac.idDIS Saputra, IPD Suarnatha, F Mahardika, A Wijanarko, SW HandaniJournal Robot. Control (JRC), 2023•journal.umy.ac.id*, vol. 4, no. 1, 2023, doi: 10.18196/jrc.v4i1.17090.

[12] E. Frederick, N. Swim, … A. K.-S. of S., and undefined 2022, "Racist state and the state of race: An analysis of Instagram commentary pertaining to LeBron James," *journals.humankinetics.com*, Accessed: May 07, 2025. [Online]. Available: https://journals.humankinetics.com/view/journals/ssj/40/3/article-p259.xml
[13] J. Chu, … M. A.-U.-J. of, and undefined 2010, "An empirical analysis of shopping behavior across online and offline channels for grocery products: The moderating effects of household and product characteristics," *journals.sagepub.comJ Chu, M Arce-Urriza, JJ Cebollada-Calvo, PK ChintaguntaJournal Interact. Mark. 2010•journals.sagepub.com*, Accessed: Jul. 12, 2024. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1016/j.intmar.2010.07.004
[14] D. Ghoul, J. Patrix, G. Lejeune, J. V.-N. L. Processing, and undefined 2024, "A combined AraBERT and Voting Ensemble classifier model for Arabic sentiment analysis," *Elsevier*, Accessed: May 07, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949719124000487
[15] Z. Khaerunnisa, K. Muhammad, and F. Mahardika, "Indonesian Journal of Digital Business Optimization of Cloud-Based Digital Archiving System Using Golang and the ICONIX Process," vol. 5, no. April, pp. 87–96, 2025.
[16] T. Tundo and F. Mahardika, "Fuzzy Inference System Tsukamoto–Decision Tree C 4.5 in Predicting the Amount of Roof Tile Production in Kebumen," *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 7, no. 2, p. 533, 2023, doi: 10.31764/jtam.v7i2.13034.
[17] T. I. Hariyanto, D. A. Halim, J. Rosalind, C. Gunawan, and A. Kurniawan, "Ivermectin and outcomes from Covid-19 pneumonia: a systematic review and meta-analysis of randomized clinical trial studies," *Rev. Med. Virol.*, vol. 32, no. 2, Mar. 2022, doi: 10.1002/RMV.2265.
[18] K. Tanaka, R. Sasano, and K. Takeda, "Are Social Sentiments Inherent in LLMs? An Empirical Study on Extraction of Inter-demographic Sentiments," Aug. 2024, Accessed: May 07, 2025. [Online]. Available: https://arxiv.org/pdf/2408.04293
[19] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
[20] P. Sari, R. S.-J. Mnemonic, and undefined 2024, "Komparasi algoritma Support Vector Machine dan Random Forest untuk analisis sentimen metaverse," *ejournal.itn.ac.id*, vol. 7, no. 1, 2024, Accessed: May 07, 2025. [Online]. Available: https://www.ejournal.itn.ac.id/index.php/mnemonic/article/view/8977
[21] E. Junita, R. S.-J. (Jurnal, and undefined 2024, "Analisis Sentimen Hate Speech Mengenai Calon Wakil Presiden Indonesia Menggunakan Algoritma BERT," *jurnal.stkippgritulungagung.ac.id*, vol. 9, no. 4, pp. 2042–2053, 2024, doi: 10.29100/jipi.v9i4.5544.
[22] E. Eskiyaturrofikoh, R. S.-J. (Jurnal, and undefined 2024, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (SVM)," *jurnal.stkippgritulungagung.ac.id*, vol. 9, no. 3, pp. 1408–1419, 2024, doi: 10.29100/jipi.v9i3.5392.
[23] S. Rahayu, J. Jaya Purnama, A. Hamid, and N. K. Hikmawati, "Analisis Sentimen AicoGPT (Generative Pre-trained Transformer) Menggunakan TF-IDF," *ojs.uajy.ac.id*, Accessed: May 07, 2025. [Online]. Available: https://ojs.uajy.ac.id/index.php/jbi/article/view/7039
[24] D. Ningsih, R. S.-J. T. I. (Jutif), and undefined 2024, "COMPARISON OF NAÏVE BAYES AND INFORMATION GAIN ALGORITHMS IN CYBERBULLYING SENTIMENT ANALYSIS ON TWITTER," *jutif.if.unsoed.ac.id*, vol. 5, no. 4, pp. 1085–1091, 1908, doi: 10.52436/1.jutif.2024.5.4.1908.
[25] S. Agung Lestari, P. Studi Sistem Informasi, S. Tinggi Ilmu Komputer Cipta Karya Informatika, K. Jakarta Timur, and D. Khusus Ibukota Jakarta, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *e-journal.hamzanwadi.ac.id*, vol. 5, no. 3, 2024, Accessed: May 07, 2025. [Online]. Available: https://e-journal.hamzanwadi.ac.id/index.php/edumatic/article/view/25667
[26] K. Najibulloh, I. Tahyudin, D. Intan, and S. Saputra, "Analisis Sentimen Ulasan Co-Pilot Google Play dengan SVM, Neural Network, dan Decision Tree," *e-journal.hamzanwadi.ac.id*, vol. 9, no. 1, pp. 275–283, 2025, doi: 10.29408/edumatic.v9i1.29673.
[27] R. Umar, I. Riadi, and D. A. Faroek, "A Komparasi Image Matching Menggunakan Metode K-Nearest Neightbor (KNN) dan Support Vector Machine (SVM)," *J. Appl. Informatics Comput.*, vol. 4, no. 2, pp. 124–131, 2020, doi: 10.30871/jaic.v4i2.2226.
[28] M. Ayu Asri Oktarini Putri, I. Wayan Sumarjaya, and I. Gusti Ngurah Lanang Wijayakusuma, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *ojs.cbn.ac.id*, vol. 9, no. 2, p. 305, 2025, Accessed: May 07, 2025. [Online]. Available: https://ojs.cbn.ac.id/index.php/jukanti/article/view/836
[29] R. Hidayah, … R. I.-J. of D. D., and undefined 2024, "Sentiment Classification of User Reviews for KAI Access Application Using Naive Bayes Method," *journal.ittelkom-pwt.ac.id*, vol. 4, no. 2, pp. 91–97, 2024, Accessed: May 07, 2025. [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/dinda/article/view/1565
[30] B. S. Atikah, Y. Sibaroni, D. Puspandari, and A. Info, "Impact of Feature Extraction on Multi-Aspect Sentiment Classification for Livin'byMandiri Using BiLSTM," *newinera.com*, vol. 05, p. 2024, 2024, doi: 10.37899/journallamultiapp.v5i5.1541.
[31] M. Ayu Asri Oktarini Putri, I. Wayan Sumarjaya, and I. Gusti Ngurah Lanang Wijayakusuma, "Penerapan Algoritma Support Vector Machine Pada Review Aplikasi Tiktok," *ojs.stmik-banjarbaru.ac.id*, vol. 9, no. 2, p. 305, 2025, Accessed: May 07, 2025. [Online]. Available: https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/2247
[32] M. Ayu Asri Oktarini Putri, I. Wayan Sumarjaya, and I. Gusti Ngurah Lanang Wijayakusuma, "Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine," *rumahjurnal.or.id*, vol. 9, no. 2, p. 305, 2025, Accessed: May 07, 2025. [Online]. Available: https://www.rumahjurnal.or.id/index.php/JEKIN/article/view/824
[33] M. Mathebula, A. Modupe, and V. Marivate, "ChatGPT As a Text Annotation Tool to Evaluate Sentiment Analysis on South African Financial Institutions," *IEEE Access*, vol. 12, no. July, pp. 144017–144043, 2024, doi: 10.1109/ACCESS.2024.3464374.
[34] G. Jeffson Sagala and Y. T. Samuel, "Sentiment Analysis on ChatGPT App Reviews on Google Play Store Using Random Forest Algorithm, Support Vector Machine and Naïve Bayes," *Int. J. Eng. Bus. Soc. Sci.*, vol. 2, no. 04, pp. 1194–1204, 2024, doi: 10.58451/ijebss.v2i04.148.
[35] D. Trihapningsari, A. Widyasuri, M. A. Putri, and A. Fatihin, "Sentiment Analysis of ChatGPT Exploration Based on Opinions on Platform X Using Naïve Bayes Algorithm," vol. 4, no. July 2024, pp. 94–101, 2025.