Journal of Information Technology and Applications Research
Volume 1 Nomor 2 November 2025
e-ISSN: 3089-8013; p-ISSN:3089-8021, Page 01-17
DOI: https://doi.org/10.63956/jitar.v1i2.31
Available online at: https://jitar.org/index.php/jitar/article/view/31

Sentiment Analysis of Netflix App Reviews on Google Play Store using the Naive Bayes

Fritadila Shafira[1], Adam Huda Nugraha[2]
Department of Information Systems, Faculty of Computer Science and Information Technology, Gunadarma University

Abstract
The rapid growth of movie streaming apps like Netflix in Indonesia has generated millions of user reviews on the Google Play Store. However, the high volume of reviews makes it difficult for developers to understand user perceptions. This study aims to analyze sentiment towards 1,000 user reviews of the Netflix application collected through data scraping in the period August 16 to October 7, 2025. Labeling was conducted using a lexicon-based approach, where each word was assigned a score based on positive and negative dictionaries. The total score determined the sentiment polarity; if the score was greater than -1 (if score > -1), the review was categorized as positive, otherwise as negative. This approach helps reduce bias that may occur when labeling relies solely on user ratings. The method used is Naïve Bayes with preprocessing stages including case folding, data cleaning, word normalization, tokenizing, stopword removal, and stemming. Furthermore, the data is weighted using TF-IDF and divided into training and test data with a ratio of 80:20. The implementation results show high model performance with accuracy of 77.96%, precision of 77.97%, recall of 77.96%, and an F1-score of 76.49%. The sentiment classification indicated that 36.5% of the reviews were positive and 63.5% were negative, indicating dissatisfaction of the majority of users. This study proves that the Naïve Bayes method is quite effective in classifying the sentiment of reviews. For further research, it is recommended to use a larger amount of data, a longer period, and explore other algorithms such as SVM, Random Forest, K-Nearest Neighbor, Decision Tree, and Rule-Based to improve the quality of sentiment analysis.
Keywords: Naive Bayes Implementation, Sentiment Analysis, Netflix, Google Play Store.

INTRODUCTION

The development of information technology has influenced people's activities, particularly in the use of film access through OTT streaming services such as Netflix, which has been available in Indonesia since 2016. Netflix offers ad-free films and a range of subscription packages.[1] Netflix has become a threat to the pay TV industry and conventional cinemas in Indonesia. [2]. Netflix has a strong appeal for movie enthusiasts, as evidenced by the number of downloads exceeding 1 billion and a rating of 3.7 out of 5 based on over 13 million reviews on the Google Play Store 2025. The large number of reviews indicates significant attention and enthusiasm from users, containing various opinions ranging from praise to criticism, providing insight into user sentiment toward

Received: September 15, 2025; Accepted: October 22, 2025; Published: November 19, 2025
*Corresponding author : adam_huda@staff.gunadarma.ac.id

the app. In this study, the review labeling process was not based solely on ratings, as this could cause bias. Instead, a lexicon-based approach was used, referring to a dictionary of positive and negative words. This approach was considered more representative in determining text polarity because it took into account the meaning of the words contained in the sentence. Therefore, sentiment analysis of Netflix user reviews is important to understand the general perspective and feedback that can serve as evaluation material for app developers. Sentiment analysis using the Naïve Bayes method was chosen because it offers high accuracy and fast text classification, and this research was conducted to assist Netflix's development by classifying the sentiment of Netflix user reviews on Google Play.

This study focuses on analyzing user reviews for the Netflix app on Google Play using the Naïve Bayes method. The objective of this study is to determine how effective and accurate this method is in categorizing sentiment in reviews for the Netflix app, as well as to provide an overview of positive and negative sentiments emerging from user reviews as a source of information for Netflix app developers. Previous studies related to sentiment analysis of the Netflix app have used the Naïve Bayes method. Research conducted by Raisa Priskilla and Suparni achieved an accuracy of 67.08% using 1,000 tweets. [3]. Subsequent research by Jessica Athalia Rieuwpassa et al. achieved an accuracy of 87.33% from 529 Netflix reviews on the Google Play Store. [4]. Additionally, research conducted by Ananda Bagas Pranata and colleagues showed an accuracy of 93.39%, with a positive precision rate of 85.52% based on 1,000 Netflix user reviews. [5]. However, other studies show that there are comparison algorithms with lower accuracy than Naïve Bayes. Zulkarnain (2024) compared the K-Nearest Neighbor (KNN), Decision Tree, and Naïve Bayes algorithms on the Netflix dataset. The results show that Naïve Bayes achieved an accuracy of 72%, higher than Decision Tree (70%) and KNN (61%). This confirms that Naïve Bayes is still relevant for use in text classification, especially for movie streaming review data. [6].

Based on the results of this study, it can be concluded that the Naïve Bayes method is quite effective for sentiment classification in Netflix application review data. Thus, the state of the art of this study is the combination of a reliable Naïve Bayes classification methodology with the latest data and communicative visualization techniques, thereby

making a significant contribution to the field of Netflix application review sentiment analysis.

RESEARCH METHOD

This study began by collecting review data automatically from the Google Play app using Python and then entering the Netflix app link. The data to be analyzed consisted of 1,000 reviews that met the criteria, with the dependent variable being user sentiment (positive or negative), including information about the app, user reviews, ratings, and other features related to the app currently available on Google Play.

In this study, the knowledge discovery from data (KDD) method was used. This process, also known as the knowledge discovery from data process, consists of a series of systematic steps used to transform unused data into useful and beneficial knowledge.

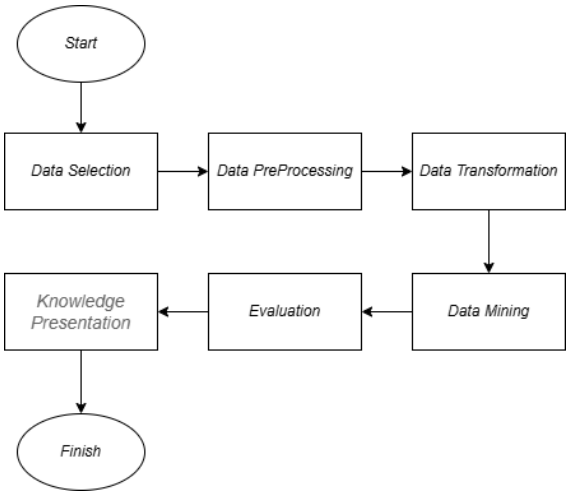

Figure 1. Research stages

The research process began with data collection using web scraping techniques. Data pre-processing aimed to convert all letters in the document to lowercase, clean the data, change or correct words in the text to make them more standard and easier for the system to understand, divide the data into separate words, and remove common words that did not provide information, such as “and,” “in,” “from,” etc., and converting words to their base form. The processed data is then labeled or assigned a sentiment score based on its context (positive or negative). Word frequency in the dataset is weighted using TF-IDF. The processed data is divided into two categories: training data and testing data.

Sentiment probability is modeled using the Naive Bayes algorithm based on the characteristics of words found in the training data. The model's performance is evaluated using a confusion matrix to calculate accuracy metrics, precision values, recall values, and F1-scores. The final results are visualized in a clearer and more understandable form regarding the model's performance, using visual representations of the most frequently occurring words in the reviews, word frequency bar charts, sentiment diagrams, and rating bar charts.

RESULTS

Data Selection

In this study, the first process carried out was data collection. The data collection technique used web scraping with the Google Play Scrapper library using the Python programming language and automatically retrieved data from the Google Play app. Using 1,000 user reviews of the app containing criticism, suggestions, and user experiences when using the app between August 16 and October 07, 2025, the results will be saved in CSV file format.

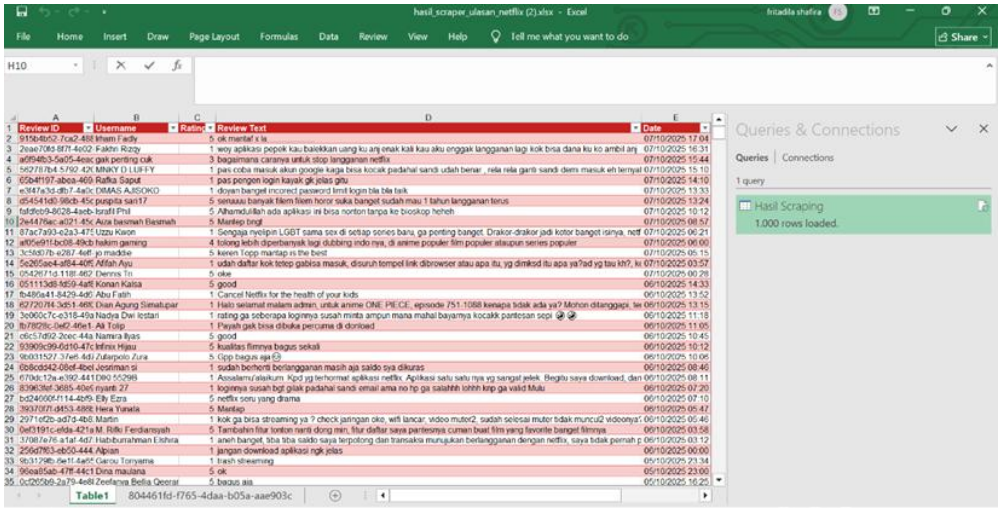

Figure 1. Data Seelction

Data Pre-processing

After selecting data taken from the Google Play app, which is still unstructured, it will undergo data pre-processing, including Case Folding, data cleaning, word normalization, tokenization, stopwords removal, stemming, and data labeling.

1. Case Folding

The first stage of pre-processing is letter matching. This stage aims to make the review text analysis more consistent by changing every letter in the text to lowercase.

Table 1. Case Folding Results

Preliminary Review	case folding results
Halo selamat malam admin, untuk anime ONE PIECE, episode 751-1088 kenapa tidak ada ya? Mohon ditanggapi, terima kasih 🙏	halo selamat malam admin untuk anime one piece episode 751-1088 kenapa tidak ada ya? mohon ditanggapi, terima kasih 🙏.
Alhamdulillah ada aplikasi ini bisa nonton tanpa ke bioskop heheh	alhamdulillah ada aplikasi ini bisa nonton tanpa ke bioskop heheh
Tambahin fitur tonton nanti dong min, fitur daftar saya pantesnya cuman buat film yang favorite banget filmnya	tambahin fitur tonton nanti dong min, fitur daftar saya pantesnya cuman buat film yang favorite banget filmnya

Description: Capital letters in the initial review are converted to lowercase letters after being processed into the equation result.

Source: Netflix app review on Google Play.

2. Data cleaning

The second stage in pre-processing is data cleaning. The aim is to remove characters that are not important to the machine learning process, such as symbols, numbers, URLs, emojis, special characters, double spaces, and punctuation marks.

Table 2. Data Cleaning Results

Preliminary Review	Data Cleaning Results
Halo selamat malam admin, untuk anime ONE PIECE, episode 751-1088 kenapa tidak ada ya? Mohon ditanggapi, terima kasih 🙏	halo selamat malam admin untuk anime one piece episode kenapa tidak ada ya mohon ditanggapi terima kasih
Gpp bagus aja 😊	gpp bagus aja
APK NONTON FILM FAVORIT❤❤	apk nonton film favorit

Description: Characters such as periods, commas, numbers, plus signs, and emojis are lost after processing.

Source: Netflix app reviews on Google Play.

3. Word normalization

The third stage in pre-processing is word normalization. The purpose of this process is to change non-standard words or words that do not conform to official spelling into standard words based on the rules of Indonesian.

Table 3. Word Normalization Results

Preliminary Review	Word Normalization Results
loginnya susah bgt gilak padahal sandi email ama no hp ga salahhh lohhh knp ga valid Mulu	loginnya susah banget gila padahal sandi email sama no hp tidak salah loh kenapa tidak valid mulu
ngga bisa login	tidak bisa login
ini gimana min nggak bisa masuk/login eror atau gimana	ini bagaimana min tidak bisa masuklogin eror atau bagaimana

Description: replace every non-standard word with a standard word based on the dictionary entries that have been added previously.

Source: Netflix app reviews on Google Play.

4. **Tokenization**

The fourth stage of pre-processing is tokenization. The purpose of this stage is to distinguish each word in the review sentence that has undergone the word normalization process previously.

Table 4. Tokenization Results

Preliminary Review.	Tokenization Results
Halo selamat malam admin, untuk anime ONE PIECE, episode 751-1088 kenapa tidak ada ya? Mohon ditanggapi, terima kasih 🙏	['halo', 'selamat', 'malam', 'admin', 'untuk', 'anime', 'one', 'piece', 'episode', 'kenapa', 'tidak', 'ada', 'ya', 'mohon', 'ditanggapi', 'terima', 'kasih']
netflix seru yang drama	['netflix', 'seru', 'yang', 'drama']
Belum bisa kebuka aplikasinya..	['belum', 'bisa', 'kebuka', 'aplikasinya']

Description: Review sentences from normalized words are separated one by one.

Source: Netflix app reviews on Google Play.

5. **Stopwords Removal**

The fifth stage of the pre-processing process is the stopwords removal. The purpose of this stage is to focus on important words by removing words that contain little information. Conjunctions such as “in,” “to,” “this,” “that,” “and,” “from,” and repeated words are considered irrelevant here and make the review sentence simpler.

Table 5. Stopwords Removal Results

Preliminary Review	Stopwords Removal Results
seruuuu banyak filem filem horor suka banget sudah mau 1 tahun langganan terus	['seru', 'banyak', 'filem', 'horor', 'suka', 'langganan']
mudah dan update vidionya	['mudah', 'vidionya']
kecewa banget udah daftar udah bayar malah ga bisa log in alasan gangguan terus	['kecewa', 'bayar', 'malah', 'alasan', 'gangguan']

Description: Words lacking information and repeated words were processed into simpler review sentences.

Source: Netflix app reviews on Google Play.

6. **Stemming**

The sixth stage in pre-processing is the stemming. The purpose of this stage is to convert words into their root forms. This process reduces words that have affixes at the beginning, middle, end, or a combination of affixes.

Table 6. Stemming Results

Preliminary Review	Stemming Results
Netflix sangat bagus banget untuk menonton film dengan lengkap. Film yang kita cari semua ada di netflix	netflix bagus banget tonton film lengkap film cari netflix
cukup menarik tapi kalau mau premium bukan 1 bayar aku maunya sekali bayar.	cukup tarik premium bukan bayar aku mau sekali bayar
Disuruh bayar ., Daftar aja suruh bayar. Gw kira ada yang free watch.	suruh bayar suruh bayar gue kira free watch

Description: Sentences that have affixes at the beginning, middle, end, or a combination of affixes after being processed into words in their basic form.

Source: Netflix app reviews on Google Play.

Data labeling

The labeling stage aims to assign sentiment labels to each user review on the Netflix application. In this study, the method used is a lexicon-based approach, which utilizes a dictionary of positive and negative words. The labeling process is carried out by counting the number of words included in the positive lexicon and negative lexicon lists. Each

word that matches the positive dictionary will add a positive score, while words found in the negative dictionary will add a negative score.

Table 7. Data Labeling Results

Preliminary Review	*Polarity*	*Polarity Score*
pilih metode bayar.	-2	Negatif
daftar ribet	-1	Negatif
langgan	1	Positif

Description: User reviews of the app are labeled as negative sentiment if they contain emotional elements, and reviews containing praise are labeled as positive.

Source: Netflix app reviews on Google Play.

Data Transformation

The first step in the data transformation stage is data division, where the data is divided into two parts: training data and testing data. In 1000 review data, it is divided into 80:20, which means 80% is used as training data and 20% as testing data.

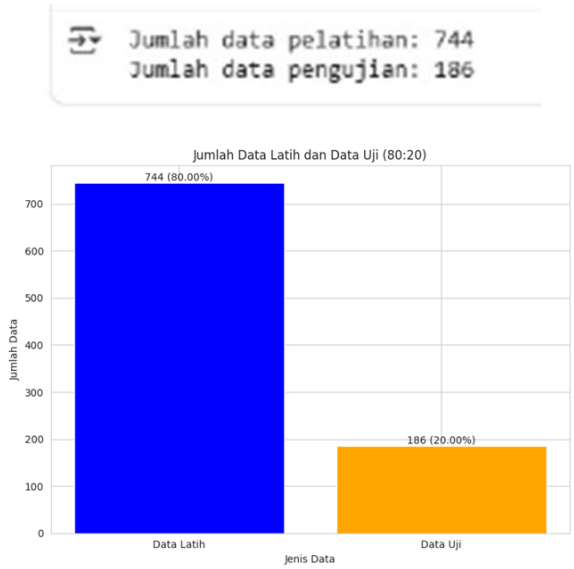

Figure 2. Data Distribution Results

The second step is to weight the words using the tf-idf method, which ensures that each word is weighted according to its frequency. The following is an overview of the manual review data calculation:

Review: "The application is very good," data #170 out of 1,000 review data.

Number of words in the data:

- "aplikasi" ⟶ 124 times
- "nya" ⟶ 144 times
- "sangat" ⟶ 50 times
- "bagus" ⟶ 138 times

1. ['aplikasi']

- Term Frequency (TF)

  $= \frac{1}{4} = 0.25$
- Inverse Document Frequency (IDF)

  = log $(\frac{1000}{124})$

  = log (8.06) = 0.90
- TF – IDF

  = 0.25 x 0.90 = **0.225**

2. ['nya']

- Term Frequency (TF)

  $= \frac{1}{4} = 0.25$
- Inverse Document Frequency (IDF)

  = log $(\frac{1000}{144})$

  = log (6.94) = 0.84
- TF – IDF

  = 0.25 x 0.84 = **0.21**

3. ['sangat']

- Term Frequency (TF)

  $= \frac{1}{4} = 0.25$

- Inverse Document Frequency (IDF)

  = log ($\frac{1000}{50}$)

  = log (20) = 1.30

- TF – IDF

  = 0.25 x 1.30 = **0.325**

4. [‘bagus’]

- Term Frequency (TF)

  $= \frac{1}{4} = 0.25$

- Inverse Document Frequency (IDF)

  = log ($\frac{1000}{138}$)

  = log (7.24) = 0.85

- TF – IDF

  = 0.25 x 0.85 = **0.212**

Data Mining

The next process is the data mining stage. The goal is to transform raw data into useful information. In this study, 1,000 review data were used to determine the system's performance on the training data and test data that had been prepared to classify user reviews. The classification method used was Naive Bayes, which refers to Bayes' Theorem and has a reference for classifying sentiment sentences into positive or negative types.

```
Review: netflix gobs lalajo kudu aya kouta unggal potong wae strees
Prediksi: Negatif
Nilai y_pred: negatif

Review: bulan langgan malah masuknonton suruh up ribu cukup p malah suruh p jelas
Prediksi: Negatif
Nilai y_pred: negatif

Review: bayar malah terus suruh bayar jelas jelas bayar slado rekening kuga potong sama netflix malah suruh bauar aneh
Prediksi: Negatif
Nilai y_pred: negatif

Review: paket premium eh tau tonton cuma doang bayar ribu tagih tanggal oktober malah tangguh
Prediksi: Negatif
Nilai y_pred: negatif

Review: aok sampah ribet tahi
Prediksi: Negatif
Nilai y_pred: negatif
```

Figure 3. Data Mining Results

Model prediction results for each review by displaying the review text, sentiment prediction, and prediction value. by showing how the model categorizes each review.

Evaluation

The evaluation stage aims to determine the performance of the model that was previously trained and tested using the naïve bayes algorithm. The method used for evaluation in this study is the confusion matrix. This testing method is used to calculate the performance level or accuracy of a classification process. The results will include accuracy, precision, recall, and F1-score values. Using these values, the quality and accuracy of the classification process can be assessed. The results of positive and negative sentiment reviews will also be displayed.

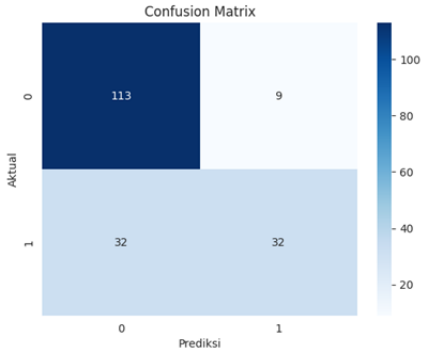

Figure 4. Evaluation Results

The confusion matrix results are as follows:

- True Positives (TP) : 32
- False Positives (FP) : 32
- False Negatives (FN) : 9
- True Negatives (TN) : 113

True Positive (TP) and True Negative (TN) values are correct classification results. False Positive (FP) values are values that are predicted to be positive but are actually negative, while False Negative (FN) values are values that are predicted to be negative but are actually positive.

The results of the classification evaluation based on the confusion matrix are described below:

1. Accuracy

$$Accuracy \ = \frac{32 + 113}{32 + 32 + 113 + 9} = 0.78 \ x \ 100\% = 78\%$$

2. Precision

$$Class \ Negative = \frac{113}{113 + 32} = 0.78 \ x \ 100\% = 78\%$$

$$Class \ Positive = \frac{32}{32 + 9} = 0.78 \ x \ 100\% = 78\%$$

3. Recall

$$Class \ Negative = \frac{113}{113 + 9} = 0.93 \ x \ 100\% = 93\%$$

$$Class \ Positive = \frac{32}{32 + 32} = 0.50 \ x \ 100\% = 50\%$$

4. F1-Score

$$Class \ Negative = \ 2(\frac{0.78 \ x \ 0.93}{0.78 + 0.93}) = 0.85 \ x \ 100\% = 85\%$$

$$Class \ Positive = \ 2(\frac{0.78 \ x \ 0.50}{0.78 + 0.50}) = 0.61 \ x \ 100\% = 61\%$$

Based on the results of testing using the Naïve Bayes algorithm on sentiment analysis of Netflix app reviews on Google Play Store, the results obtained were 78% accuracy, 78% precision, 78% recall 78%, and an F1 score of 76%.

Knowledge Presentation

The result of this stage is a visualization of the research results to obtain an overview or general information about the Netflix app user review data. The following is a word cloud of the overall positive and negative review data for the Netflix app.

Figure 5. Overall Wordcloud

Based on the image above, there are words that appear frequently, such as “bayar”, “Netflix”, “tonton”, “langgan”, “film”, dan ”bagus”. Words that appear frequently will be larger in the WordCloud visualization.

Figure 6. Wordcloud Positive Review

Based on the image above, the larger the size of a word in the WordCloud, the higher the frequency of that word's appearance in the discussion topic in positive reviews. The words that frequently appear are “langgan”, “bayar”, “Netflix”, “saldo”, “tonton” dan “film”. This indicates that positive reviews from Netflix app users express the ease of using the Netflix app for smooth movie streaming, the quality of the movies, and gratitude for the app's performance.

Figure 7. Wordcloud Negative Review

Based on the image above, the words that frequently appear in negative reviews of the Netflix app are “film”, “bayar”, “tonton”, “susah” dan “Netflix”. It can be concluded that in negative reviews, users mention that the challenges they face when using the Netflix app include frequent login failures, difficulties in creating a Netflix account, and issues or problems related to payment for subscription packages within the Netflix app. These topics can be utilized as a reference for Netflix Company to evaluate the performance of the Netflix app for user convenience.

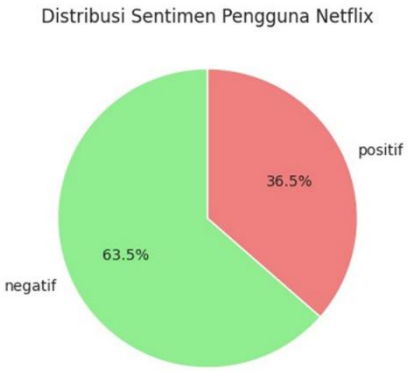

Figure 8. Diagam Pie Sentimen

Based on the sentiment diagram above, the distribution of user sentiment toward the Netflix app is based on sentiment analysis using the Naive Bayes method. A total of 63.5% of user reviews fall into the negative category, while 36.5% fall into the positive category. This indicates that most users who leave reviews on the Google Play app are more likely to express dissatisfaction or criticism toward Netflix's services. These reviews can serve as a valuable source of feedback for Netflix app developers to improve service quality, app features, and the overall user experience.

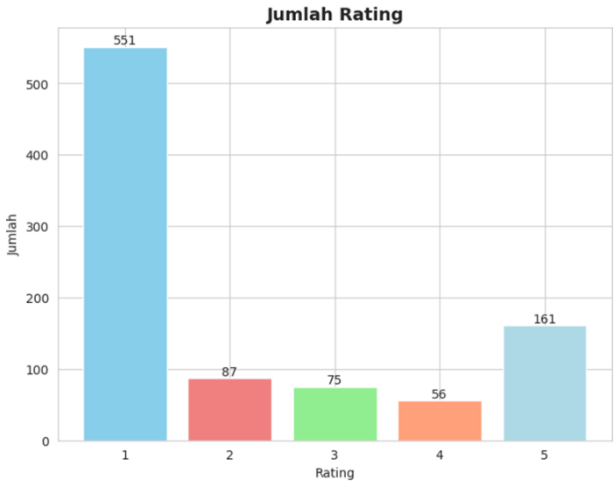

Figure 9. Barchart Ratings

Based on the rating diagram above, the most ratings were given to a score of 1, with more than 551 reviews, indicating that most users are dissatisfied with the Netflix app. This is due to technical issues such as login difficulties, videos not playing, or slow app performance. The 5-star rating also has a fairly high number, around 161 reviews, indicating that some users are still very satisfied with the service provided. Ratings of 2, 3, and 4 have relatively few and balanced numbers, each ranging from 56 to 87 reviews. This indicates that users tend to give extreme ratings, either very poor or very good, and rarely give moderate ratings.

CONCLUSION

Sentiment analysis research on user reviews of the Netflix app on Google Play Store using the Naïve Bayes method with a lexicon-based approach has been successfully conducted. Sentiment labeling was performed by calculating the total word score from the positive and negative lexicon dictionaries. If the score value was greater than -1, the review was categorized as positive, while a score ≤ -1 was categorized as negative. From the classification results, 36.5% of reviews were categorized as positive and 63.5% were categorized as negative. The model evaluation results showed an accuracy rate of 77.96%, precision of 77.97%, recall of 77.96%, and an F1-score of 76.49%. Thus, the Naïve Bayes method demonstrated satisfactory performance in classifying user review sentiment.

REFRENCES

Arsi, P. and Waluyo, R., Analysis of Sentiment in Discourse on the Relocation of Indonesia's Capital Using the Support Vector Machine (SVM) Algorithm, Journal of Information Technology and Computer Science, 8(1), p. 147. DOI: https://doi.org/10.25126/jtiik.0813944, 2021.

Djamzuri, M.I. and Mulyana, A.P., The Phenomenon of Netflix Premium Video Streaming Platform in Building Cyber Ethics Awareness from a Communication Science Perspective. JISIP (Journal of Social Sciences and Education), 6(1). Doi: https://doi.org/10.36312/jisip.v6i1.2804/http, 2022.

Priskilla, R. & Suparni, Analysis of Public Sentiment Toward the Netflix Online Movie Streaming Application Using the Naïve Bayes Method, JATI (Journal of Computer Science Students), 8(2), pp.1339–1342. Bina Sarana Informatika University, 2023.

Rieuwpassa, J.A., Sugito & Widiharih, T, Implementation of the Naive Bayes Classifier Method for Classifying User Review Sentiment on the Netflix Application on Google Play, Jurnal Gaussian, 12(3), pp.362–371. Doi: https://doi.org/10.14710/j.gauss.12.3.362-371, 2023.

Pranata, A.B., Abdillah, A.R., & Irwiensyah, F., Sentiment Analysis of User Reviews of the Netflix Application on Google Play Using the Naïve Bayes Algorithm. KLIK: Scientific Journal of Informatics and Computer Science, 4(6), pp.3091–3098. Doi: https://doi.org/10.30865/klik.v4i6.1964, 2024.

Zulkarnain, Z., Mutia, R., Ariani, J.A., Barik, Z.A. and Azmi, H., 2024. Performance Comparison K-Nearest Neighbor, Naive Bayes, and Decision Tree Algorithms for Netflix Rating Classification. IJATIS: Indonesian Journal of Applied Technology and Innovation Science, 1(1), pp.16-22. Doi: doi.org/10.57152/ by Crossref.