JEET: Journal of Engineering and Emerging Technology
Vol. 3, No. 1, Mei 2025, Page 43~52
ISSN: 2964-4178

Analisis Sentimen pada Ulasan Aplikasi Tiktok di *Google Play Store* Menggunakan Pendekatan *Naive Bayes*
(Sentiment Analysis on Tiktok App Reviews on Google Play Store Using The Naive Bayes Approach)

Baiq Dwi Zulianti K*, Joselina Rizki Bimantari, Heri Wijayanto

Magister of Information Technology, Mataram University Jl. Majapahit 62, Mataram, NTB, Indonesia

*Corresponding author: dwizulianti11@gmail.com

(Received: May 2025 / Revised: May 2025 / Accepted: June 2025)

ABSTRACT

The TikTok application has gained global popularity, leading to a surge of user reviews on platforms such as the Google Play Store. This study aims to analyze user sentiment using the Naive Bayes algorithm by classifying reviews into positive and negative categories. The dataset, consisting of 1,914 user reviews, underwent preprocessing through case folding, stopword removal, and lemmatization. Sentiment labeling was carried out using TextBlob, and classification performance was evaluated using K-Fold Cross Validation and a confusion matrix. Results indicate that the Naive Bayes model achieved an average accuracy of 82.76%, demonstrating its effectiveness in identifying sentiment in user-generated content. Despite certain limitations, such as the absence of neutral sentiment and class imbalance, this research provides valuable insights into user perceptions of the TikTok app and highlights the importance of textual context in sentiment classification. For future work, it is recommended to expand the sentiment categories by including neutral labels and to experiment with other classification algorithms or deep learning approaches to improve accuracy and generalizability.

Keywords:

Sentiment Analysis;
Naive Bayes;
TikTok App Review;

ABSTRACT

Aplikasi TikTok telah meraih popularitas global, yang menyebabkan lonjakan ulasan pengguna di platform seperti Google Play Store. Penelitian ini bertujuan untuk menganalisis sentimen pengguna menggunakan algoritma Naive Bayes dengan mengklasifikasikan ulasan ke dalam kategori positif dan negatif. Dataset yang terdiri dari 1.914 ulasan pengguna menjalani tahap praproses melalui case folding, penghapusan stopword, dan lemmatisasi. Pelabelan sentimen dilakukan menggunakan TextBlob, dan kinerja klasifikasi dievaluasi menggunakan K-Fold Cross Validation serta confusion matrix. Hasil penelitian menunjukkan bahwa model Naive Bayes mencapai rata-rata akurasi sebesar 82,76%, yang membuktikan efektivitasnya dalam mengidentifikasi sentimen pada konten buatan pengguna. Meskipun terdapat beberapa keterbatasan, seperti tidak adanya kategori sentimen netral dan ketidakseimbangan kelas, penelitian ini memberikan wawasan berharga tentang persepsi pengguna terhadap aplikasi TikTok dan menyoroti pentingnya konteks teks dalam klasifikasi sentimen. Untuk penelitian selanjutnya, disarankan untuk memperluas kategori sentimen dengan menambahkan label netral serta mengeksplorasi algoritma klasifikasi lain atau pendekatan pembelajaran mendalam guna meningkatkan akurasi dan kemampuan generalisasi.

Keywords:

Analysis Sentiment;
Naive Bayes;
TikTok App Review;

1. Introduction

Technology is currently experiencing very rapid development. There are various media available for communication, and ease of access is increasing with the availability of

JEET: Journal of Engineering and Emerging Technology
Vol. 3, No. 1, Mei 2025, Page 43~52
ISSN: 2964-4178

widespread internet connections. Research on sentiment analysis in this context is still a relatively new and developing field. In the past, when the internet was not as widespread as it is now, data collection was more limited and often only depended on the opinions of those around you, such as friends and family. However, with the rapid development of technology today, especially with the existence of smartphones, people can easily access various communication media, including Facebook, Instagram, Twitter, and so on (Kusuma, A., & Nugroho, A. (2021). Each media has different features, creating a variety of experiences in communicating. One application that is currently stealing the attention of many people is TikTok.

The TikTok application provides users with the opportunity to quickly create and share short videos. Even though the majority of users are elementary school children, the average age of TikTok users is almost the same as that of minors. Many underage TikTok users use devices that do not support the features or filters available in the application. The negative impact of this is that video uploaders often receive negative criticism because they think the uploaded video is of low quality (Apriliani et al., 2024).

In the comments section, users can find comments in various languages that are often rude and can even lead to bullying of the video uploader. Not only that, this application also provides a like button which is used to show appreciation for the videos that have been watched. Many negative reviews of the TikTok application were found on the Google Play Store, so sentiment analysis is needed to identify and sort these reviews into negative or positive opinions. By conducting this sentiment analysis, it is hoped that it can help solve problems around review grouping and provide deeper insight into the user experience of the TikTok application (Apriliani et al., 2024).

In response to the aforementioned issues, the author performed sentiment analysis using public feedback sourced from TikTok application reviews on the Google Play Store. The objective of this sentiment analysis is to categorize the text as positive, negative, or neutral based on the existing dataset. To classify and determine the nature of the reviews, the Naive Bayes method is employed. This study aims to identify the sentiment of user reviews and offer recommendations for enhancing the TikTok application based on user feedback from the Google Play Store.

The Naive Bayes method, a widely used classification algorithm in sentiment analysis, is utilized in this study. This algorithm determines the sentiment of the text—positive, negative, or neutral—by calculating the probability of each word (Apriani, R, et al, 2019). The use of the Naive Bayes method in this study was motivated by its simplicity, transparency, and computational efficiency, especially when dealing with structured textual data such as application reviews. While it is true that more recent and sophisticated methods like Convolutional Neural Networks (CNN) or Transformer-based models such as BERT have demonstrated superior performance in various natural language processing tasks, this study aims to establish a baseline using classical machine learning techniques. The results obtained can serve as a point of comparison for future research involving deep learning-based approaches. Incorporating models such as LSTM or BERT is part of our future research agenda, with the goal of improving the model's ability to understand contextual meaning and handle class imbalances more effectively (A. Kurniawan, 2019). The goal of this research is to identify patterns in TikTok user reviews on the Google Play Store. By analyzing these reviews, the study aims to provide insights into user preferences and issues, offering developers valuable information to enhance the user experience.

2. Methods

2.1. Dataset

The dataset used in this research was taken from Kaggle and consists of 7472 user comments regarding the TikTok application on the Google Play Store. This dataset has 10 parameters that provide various information regarding user reviews. These parameters include a unique ID for each review (reviewId), user name (userName), URL of the user's profile image (userImage), content or text of the review (content), star rating given (score), number of "likes" or "thumbs up " received reviews (thumbsUpCount), the version of the app when the review was created (reviewCreatedVersion), the date and time the review was created (at), the reply from the app developer if any (replyContent), and the date and time of the reply from the developer (repliedAt). This dataset provides comprehensive and relevant information to analyze user sentiment towards the TikTok application. By using these 10 parameters, research can leverage rich data to produce accurate sentiment analysis. This dataset can be accessed here.

2.2. Data Labeling

To analyze the sentiment of user reviews, this research uses TextBlob, a Python library for text processing. TextBlob is used for word labeling, which involves determining the polarity and subjectivity of each review. Polarity indicates how positive or negative a text is, with values ranging from -1 (very negative) to 1 (very positive), while subjectivity indicates how subjective or objective a text is, with values ranging from 0 (very objective) to 1 (very subjective). In this study, the sentiment classification includes two categories: positive and negative. Reviews with a polarity value greater than 0 are labeled as positive and less than 0 as negative. This rule-based threshold ensures that only strongly opinionated texts are classified as positive or negative. These sentiment labels are then used as the basis for training and testing the Naive Bayes model in classifying the sentiment of TikTok user reviews. This process enables automatic identification of sentiment patterns and provides deeper insight into user perceptions of the TikTok app.

2.3. Text Preprocessing

The next stage in this research is text preprocessing, which is an important step to prepare the data before further analysis. Text preprocessing involves several steps, namely the removal of irrelevant punctuation, numbers and special characters; converting text to lowercase for consistency; removal of stop words, namely common words that do not have important meaning such as "and", "or", "but"; and stemming or lemmatization, which aims to change words to their basic form so as to reduce word variability. These steps aim to clean and simplify the text, so that the resulting features are more representative and relevant for sentiment analysis. This text preprocessing process is important to improve the accuracy and efficiency of the Naive Bayes model used in user review sentiment classification.

2.4. K-Fold Cross Validation

After the evaluation stage using the separation of the dataset into training and test data, the K-Fold Cross Validation stage was also conducted to validate the model's performance more robustly. K-Fold Cross Validation involves dividing the dataset into k equal-sized folds. Each fold is used once as a validation set, while the remaining $k - 1$ folds are combined to form the training set. This process is repeated k times, ensuring that every data point is used for both training and validation exactly once.

For example, with $k = 5$, the dataset is divided into five parts:

- In Iteration 1, Fold 1 is used for testing, and Folds 2–5 are used for training.
- In Iteration 2, Fold 2 is used for testing, and Folds 1, 3–5 are used for training.
- This continues until all folds have been used as the test set.

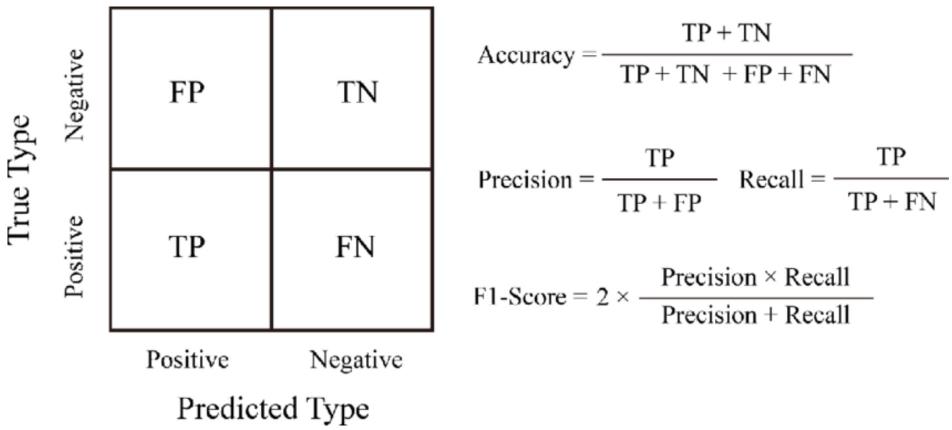
The performance metrics from each iteration such as accuracy, precision, recall, and F1-score—are then averaged to provide a more stable and generalized evaluation of model performance. This method helps prevent overfitting and gives a more reliable estimate of how the Naive Bayes model will perform on unseen data. By applying K-Fold Cross Validation, the research ensures that the classification model is consistently evaluated across the entire dataset, making the results more trustworthy (Putri et al., 2023).

2.5. Confussion Matrix

After completing the K-Fold Cross Validation stage, an analysis was performed using a confusion matrix to gain a more detailed understanding of the model's classification performance. A confusion matrix is a table that is commonly used to evaluate the performance of a classification model by comparing the predicted labels with the actual labels. It consists of four key components:

- True Positive (TP): The number of correctly predicted positive samples.
- True Negative (TN): The number of correctly predicted negative samples.
- False Positive (FP): The number of negative samples incorrectly predicted as positive.
- False Negative (FN): The number of positive samples incorrectly predicted as negative.

From these four values, several evaluation metrics can be derived:



In this study, the confusion matrix is used to evaluate how well the Naive Bayes model distinguishes between positive and negative sentiments in TikTok user reviews. By analyzing the distribution of these four components, researchers can identify the types of errors the model tends to make (e.g., high false positives or false negatives), which can inform future improvements to the classification system (N. M. A. J. Astari, et al, 2020).

3. Results and Discussion

3.1. Data Collection

The dataset used in this study is secondary data obtained from the Kaggle platform, consisting of 1,914 publicly available user reviews of the TikTok application on the Google Play Store. Each entry contains two attributes: the review text and its corresponding sentiment label. Prior to sentiment labeling, the dataset underwent a preprocessing phase, including case folding, removal of special characters and numbers, stopword elimination, and lemmatization. These steps were conducted to ensure that the extracted features were cleaner and more representative. After preprocessing, the dataset was labeled into two sentiment categories—positive (1,580 reviews) and negative (334 reviews) using the TextBlob library. Although neutral sentiment was not considered in this study, its inclusion is planned for future work to enable a more comprehensive sentiment analysis.

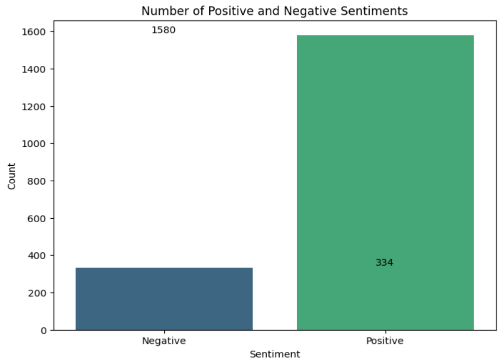


Figure 1. Dataset Proportion

To analyze user sentiment, this study utilizes the TextBlob library, which calculates the *polarity* of each review on a scale from -1 (very negative) to +1 (very positive). For the purpose of labeling, we defined clear thresholds: reviews with a polarity score greater than 0 were labeled as *positive*, while those with a score less than 0 were labeled as *negative*. Reviews with a polarity score equal to 0 were not included in this study but are considered *neutral* and will be incorporated in future research. This thresholding approach allows for a consistent and objective labeling process.

Table 1. Example Dataset.

Text	Sentiment
Great Fun App Far	Positive
App would get higher rating literally cant sign second open app try tap either sign sign literally anything screen completely freeze cant anything ive uninstalled app restarted phone nothing fixing dont know phone new problem thats experience	Negative
Wish Could Give 100 Percent Rating	Positive
best one	Positive
Love Amazing App	Positive
nice app video going	Positive
Love Tik Tok	Positive
App wont stop crashing horrible im making video crash deletes progress get app fixer	Negative

3.2. Preprocessing

The dataset used in this study comprises real-world user reviews collected from the Google Play Store, reflecting authentic user experiences with the TikTok application. These reviews often contain informal language, typos, abbreviations, and mixed expressions. Prior to conducting the preprocessing steps, a manual inspection of the data was performed to understand the general characteristics of both positive and negative reviews. For example, positive reviews often contained words such as “love”, “fun”, and “amazing”, while negative reviews typically included expressions like “crash”, “banned”, or “bad experience”. This preliminary insight guided the preprocessing phase and helped inform the interpretation of sentiment results later in the analysis. Data processing involves preparing the data beforehand, which includes pre-processing steps and removing duplicate data. These processing steps are executed using the stages in text mining.

JEET: Journal of Engineering and Emerging Technology
Vol. 3, No. 1, Mei 2025, Page 43~52
ISSN: 2964-4178

3.2.1. Case Folding

In this process, the dataset that has been previously cleaned will take steps to change all letters to lower case.

3.2.2. Stopword

In this stopword process, the dataset will be filtered for words that do not have meaning, such as conjunctions.

3.2.3. Lemmatize

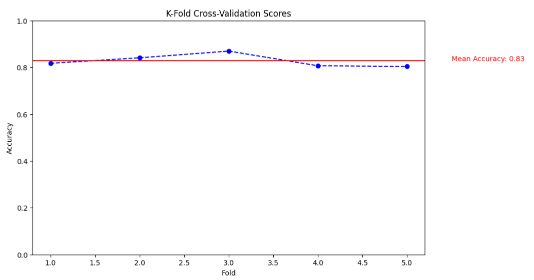
Lemmatization is a pre-processing method in natural language processing (NLP) that simplifies a word to its root form to recognize similarities. For instance, a lemmatization algorithm would convert the word "better" to its base form, "good."

Table 2. Comparison of Raw Data and Preprocessed Data

Content	Pre-Processed Text
No words	word
Great fun app so far!	great fun app
The app would get a higher rating but I literally keep getting banned for no reason at all.	app get high rate keep get ban reason
I WISH I COULD GIVE THIS A 100 PERCENT RATING	wish could give percent rate
Pictures and record	picture record

3.3. Implementation

The Naive Bayes model was evaluated using the 5-Fold Cross Validation method to ensure the robustness and generalization capability of the classifier. The individual accuracy scores obtained from each fold were as follows: Fold 1 = 0.8172 (81.72%), Fold 2 = 0.8407 (84.07%), Fold 3 = 0.8695 (86.95%), Fold 4 = 0.8068 (80.68%), and Fold 5 = 0.8037 (80.37%). Based on these values, the mean accuracy achieved by the model was 0.8276 (82.76%), indicating good overall performance. Additionally, the standard deviation of the accuracy scores was 0.0246 (2.46%), which reflects a relatively low variance across different folds. This suggests that the Naive Bayes model exhibits stable and consistent performance in classifying user sentiments from TikTok reviews on the Google Play Store.




```
Cross-validation scores: [0.81723238 0.84073107 0.8694517  0.80678851 0.80366492]
Mean accuracy: 0.8275737153636898
Standard deviation: 0.02464743299383115
```

Figure 2. K-Fold Value

JEET: Journal of Engineering and Emerging Technology
Vol. 3, No. 1, Mei 2025, Page 43~52
ISSN: 2964-4178

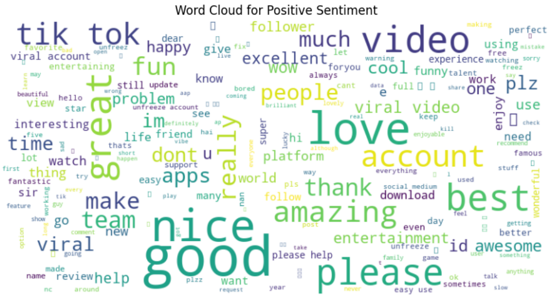


Figure 3. Word Cloud Positive

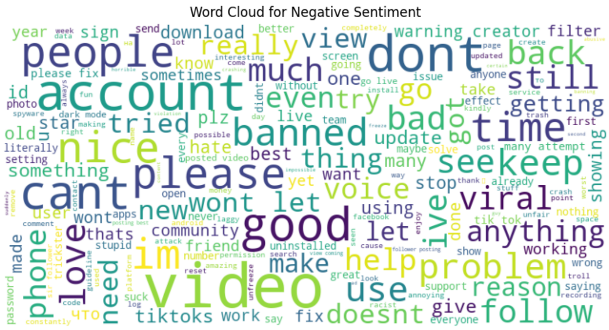


Figure 4. Word Cloud Negative

Table 3. Word Frequency

Word	Frequency
app	95,150
good	64,055
love	42,202
nice	41,961
tiktok	39,335
best	20,725
like	18,868
great	17,690

The analysis of reviews with ratings of 4 and 5 reveals that the most frequently used word is “app”, appearing 95,150 times, which indicates that users often directly refer to the application in their feedback. Other highly frequent words such as “good”, “love”, “nice”, and “great” suggest a generally positive user sentiment and satisfaction. The word “tiktok” appears 39,335 times, highlighting that many users mention the app by name when expressing positive experiences. Terms like “best” and “like” further reflect user appreciation and approval toward various features of the app. These frequently occurring words collectively indicate strong user engagement and positive perception of the TikTok platform.

The same words appear in the wordcloud for both negative and positive sentiments because these words are often used in both positive and negative contexts. Here are some reasons why this happens. Common words like "video," "account," "people," and "please" are words that are used in many contexts, whether users are giving praise or complaining about a feature or their experience. Moreover, the same words can be used in different contexts by users. For example, "good" may appear in a positive review to emphasize good quality, while in a negative review, it may be used to suggest improvements, such as "could be good if...".

The same user may also have a mixed experience with the app, praising certain aspects while criticizing others. For example, someone could say "I love the videos, but the account

JEET: Journal of Engineering and Emerging Technology
Vol. 3, No. 1, Mei 2025, Page 43~52
ISSN: 2964-4178

management is bad". Additionally, words frequently used in general discussions about the app tended to appear more frequently in both sentiment categories. This includes words directly related to the app's main features, such as "video" for video-based apps like TikTok.

The word "video" can be used in a positive context such as "The video quality is amazing" or in a negative context such as "The video keeps freezing and buffering". The word "account" can be used in a positive context such as "Creating an account was easy and quick" or in a negative context such as "I have issues logging into my account". Likewise, the word "people" can be used in a positive context such as "People here are so creative and fun" or in a negative context such as "People are leaving negative comments on my posts".

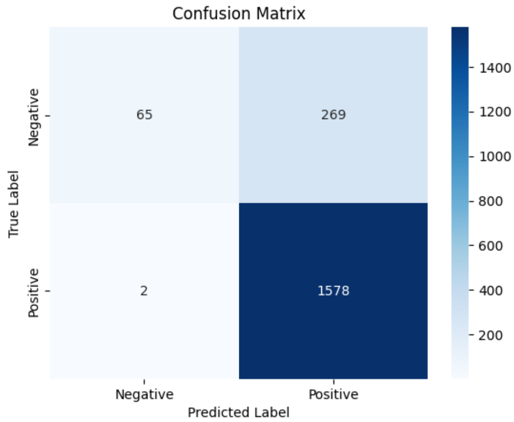


Figure 5. Confussion Matrix

Of the total 334 reviews indicating negative sentiment towards TikTok, the model was only able to correctly identify 65 reviews, while 269 reviews were incorrectly classified as positive sentiment. This results in a low recall value for negative sentiment of 19%. However, for positive sentiment, the model achieved perfect recall of 100%, with only 2 reviews incorrectly classified as negative sentiment.

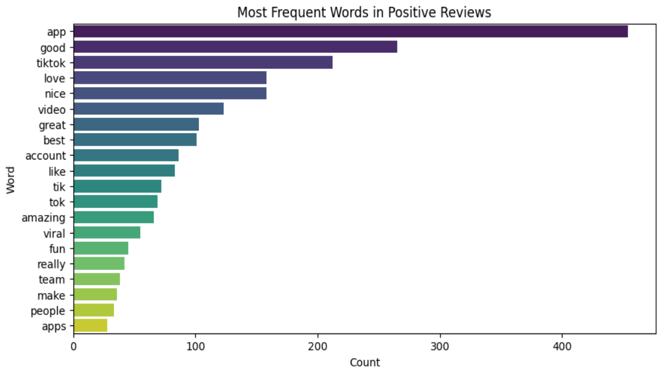


Figure 6. The most frequently occurring positive words

The word "app" is the word that appears most frequently, with the highest number of occurrences, more than 400 times. This shows that positive reviews often speak directly about the app. The words "good" and "tiktok" also appeared frequently, more than 200 times each. "Good" indicates that many users gave positive reviews about the quality of the app, while "tiktok" as the name of the app indicates that many users specifically mentioned the app in a positive context. Words like “love,” “nice,” “video,” “great,” and “best” also appear quite frequently, with a frequency of between 100 and 200 times. These words show that users often use very positive terms in their reviews. Other words like "account", "like", "tik", "tok", "amazing", "viral", "fun", "really", "team", "make", "people", and "apps", although lower in frequency than the previously mentioned words, still indicates specific aspects of an application that users like. In Figure 6, it can be observed that the bias between the precision and recall of the positive and negative classes remains relatively high. This discrepancy indicates that the model is less consistent in correctly identifying certain sentiment classes, particularly when dealing with imbalanced data. To address this issue, future improvements may include applying data balancing techniques such as SMOTE or adjusting class weights during model training. These approaches can help reduce bias by providing the model with more representative samples from each class. Although this bias may not drastically change the overall accuracy, it could significantly affect the reliability of predictions for minority classes, thereby influencing the practical application of the model.

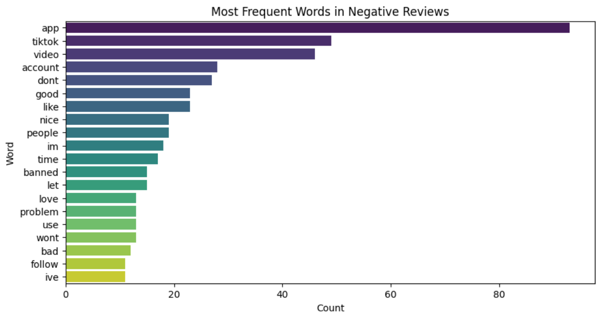


Figure 7. The most frequently occurring negative words

The word that appeared most frequently was "app," with over 80 occurrences, indicating that many negative reviews mentioned apps specifically. "tiktok" and "video" also come up frequently, which makes sense considering this review is probably about the TikTok app. Words like "account" and "dont" indicate that account-related issues and things that don't work or that users don't like are also widely discussed.

User experience related words like "good", "nice", "like", and "love" while usually having positive connotations, appear in a negative context here, perhaps in phrases like "not good", "don't like ", or "I love TikTok but...". "people" and "time" may refer to problems experienced by other users or application usage time issues.

There are also specific issues identified, such as "banned" which indicates that there are a lot of negative reviews associated with a blocked account. "problem" and "use" indicate various problems using the application, while "bad" and "wont" indicate negative experiences. The app's interactions and features are also highlighted, with the word "follow" possibly indicating an issue with the app's following feature. "ive" is likely part of a longer phrase like "I've been banned" or "I've had problems.".

4. Conclusions

From the results of this research, it can be concluded that the Naive Bayes approach is effective for classifying the sentiment of reviews of the TikTok application on the Google Play Store with the results that the word "app" appears most often in negative reviews, indicating that many users specifically mention the application in their complaints. Words like "tiktok", "video", "account", "dont", and "banned" indicate problems with the app's core features and account management. Common words like "video," "account," "people," and "please" appeared frequently in both positive and negative reviews, demonstrating the importance of the context of word use in determining sentiment. Additionally, some words such as "good," "nice," "like," and "love" can appear in both positive and negative contexts, depending on how users express their experiences. The duality of user experience was also revealed, with some users praising certain aspects while criticizing others. Overall, this analysis shows that despite significant complaints, there are also aspects of the app that users appreciate, highlighting the importance of understanding context in user reviews.

References

Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes. Jurnal Ilmiah Teknologi Informasi Asia, 15(2), 137-146.

Apriliani, N., Suarna, N., & Prihartono, W. (2024). Analisis Sentimen Review Penggunaan Tiktok Melalui Pendekatan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3725–3731. https://doi.org/10.36040/jati.v7i6.8299

Putri, K. S., Setiawan, I. R., & Pambudi, A. (2023). Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier. *Technologia : Jurnal Ilmiah*, *14*(3), 227. https://doi.org/10.31602/tji.v14i3.11259

Lasulika, M. E. (2019). Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel. ILKOM Jurnal Ilmiah, 11(1), 11-16.

E. Kartika and J. Gondohanindijo, "Rancang Bangun Model Sentimen Analisis Review Produk Pada Toko Online Menggunakan Naive Bayes," Seminar Nasional Hasil Penelitian dan Pengabdian Kepada Masyarakat, pp. 201–212, 2020.

M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," SMATIKA JURNAL, vol. 10, no. 02, pp. 71–76, 2020.

N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," Jurnal Sistem dan Informatika (JSI), vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.

A. Kurniawan and S. Adinugroho, "Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features," 2019. [Online]. Available: http://j-ptiik.ub.ac.id

A. A. Nabhan, B. Rahayudi, and D. E. Ratnawati, "Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru (New Normal) dengan Metode Naïve Bayes," 2021. [Online]. Available: http://j-ptiik.ub.ac.id