Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

June 2026. Vol. 5. No. 3; e-ISSN: 2808-4519

Analysis of Hate Speech Againts Gojek Drivers using the Naïve Bayes Algorithm on the Facebook Platform

Nazwa Putri Ananda[1*], Firahmi Rizky[2]

[1,2] *Muhammadiyah University of North Utara, Medan 20238, Indonesia*
nazwaputriananda123@gmail.com [1*]

Abstract

Social media has become a space where many individuals express their opinions freely, including negative comments that may lead to hate speech. One group often targeted by such speech is Gojek drivers. This study aims to classify user comments on Facebook into two sentiment categories: positive and negative, with a primary focus on negative comments. The data was collected from public Facebook posts using the APIFY scraping tool. After the data was gathered, several preprocessing stages were carried out, including case folding, cleaning, tokenization, normalization, stopword removal, and stemming. The text data was then converted into numerical form using CountVectorizer. The classification algorithm used in this research is Naive Bayes with the MultinomialNB model, as the input data consists of word frequency. The results of the model evaluation show that this algorithm performs well in classifying negative comments, especially in identifying word patterns that commonly appear in hate speech directed toward Gojek drivers.

Keywords: *Sentiment, Hate Speech, Facebook, Gojek, Naïve Bayes, MultinomialNB*

1. Introduction

Hate speech is an act of intimidation or abusive treatment that has intrinsic characteristics such as bullying, which is an aggressive, deliberate, and repeated action over time by an individual or group against another person, carried out continuously on electronic media. To understand the patterns and levels of hate speech that occur, a method is needed that can analyze the sentiment of comments directed at Gojek drivers on the Facebook platform. One effective approach is to use sentiment analysis based on the Naïve Bayes algorithm, which is a probabilistic classification method in machine learning. This algorithm is able to classify comments or reviews into positive, negative, or neutral sentiment with a high degree of accuracy [1]. Sentiment analysis is an automated process for processing text data to analyze people's opinions in written language. This is done to identify reviews of subjects and objects such as individuals, organizations, and products and services. Sentiment analysis can help understand public opinion and communication on social media [2]. Sentiment analysis is performed in two ways: the first is machine learning-based, which trains a classifier on a manually labeled dataset. The second is lexical-based, which doesn't require training on a dataset. It measures the polarity of documents or sentences based on the sentiment of words and phrases using specific linguistic rules. To do this, reviews must be classified into neutral, positive, and negative categories. Negative reviews contain elements of hate speech, while positive reviews contain praise or support.

In this study, the data collection process was carried out using APIFY, a cloud-based automation platform used for web scraping, crawling, and process automation on the web. This service allows users to extract data from various websites in an automated and structured manner. APIFY provides a JavaScript-based working environment (Node.js) and supports the use of popular libraries such as Puppeteer and Cheerio. One of APIFY's advantages is its ability to run agents (called actors) that can run in the background and be scheduled regularly, and the data extraction results can be directly exported in formats such as JSON, CSV, or Excel [3]. The scraping results are displayed in a table consisting of several columns: post author (username), comment, which contains the user's comment, number of likes, which records the number of likes on the comment, and post URL, which is a link to the source post on Facebook. This structure allows hate speech comments against Gojek drivers to be identified and further analyzed. The scraped data is then exported to CSV format and used as raw data in preprocessing before being analyzed with the Naïve Bayes algorithm [4], [5].

2. Research Methodology

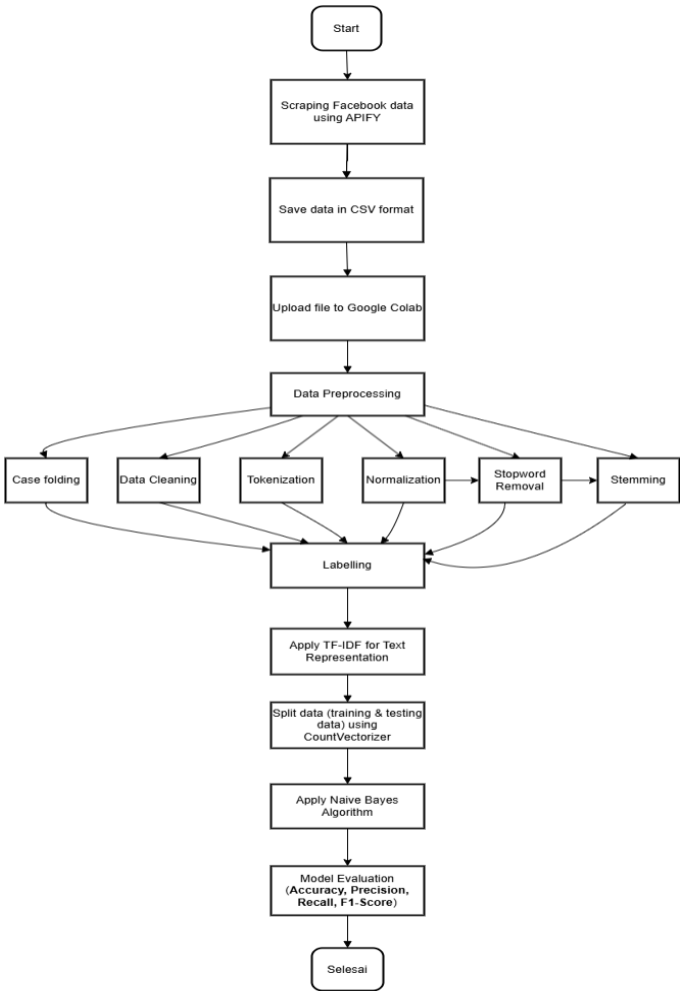

Fig. 1*:* Flowchart Design

2.1. Type of Research

This study applies a quantitative, exploratory approach. The research employs machine learning-based sentiment analysis using the Naïve Bayes algorithm, specifically the MultinomialNB model, since the dataset consists of word frequencies . The research is also descriptive in nature, aiming to reveal the intensity and forms of negative comments directed at Gojek drivers. Several stages of text mining were performed, including case folding, data cleaning, tokenization, normalization, stopword removal, and stemming, before classification was carried out [6].

2.2. Data Collection

The data was collected from Facebook public posts and groups related to Gojek services. Web scraping was conducted using APIFY, a cloud-based automation platform that extracts comments systematically . The collected dataset contained 2,217 comments. After scraping, the data was exported into CSV format for further processing. Only relevant comments related to Gojek drivers were included, while spam, duplicates, and irrelevant comments were removed. The final dataset was then labeled into positive and negative categories based on lexical and manual approaches.

3. Result and Discussion

3.1. Result

The dataset was split into 80% training data (1,731 comments) and 20% testing data (433 comments) using CountVectorizer. The dimensionality of the vectorized data reached 3,667 unique features (words). The Naïve Bayes algorithm successfully classified comments into positive and negative categories. Evaluation using accuracy, precision, recall, and F1-score showed that the model could recognize negative comments and detect hate speech patterns with a reasonable degree of reliability.

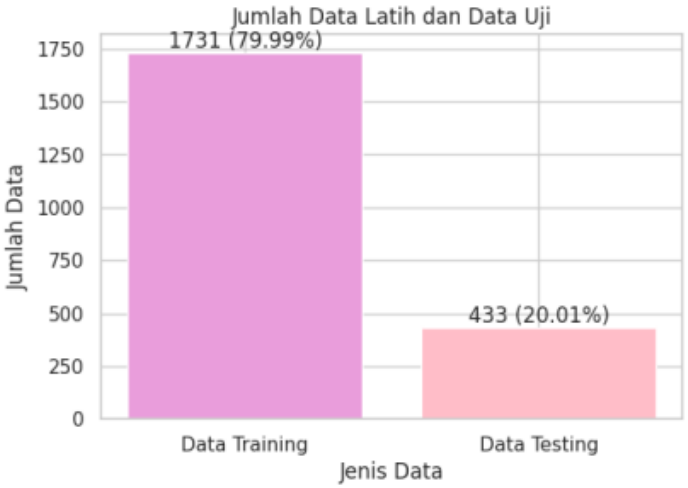

Fig.2: Data Training and Testing

These results demonstrate a visualization of the data division into two parts: training data and testing data. The bar chart shows that 1,731 data points were used for training, representing 79.99% of the total available data. Meanwhile, 433 data points were used for testing, representing 20.01%.

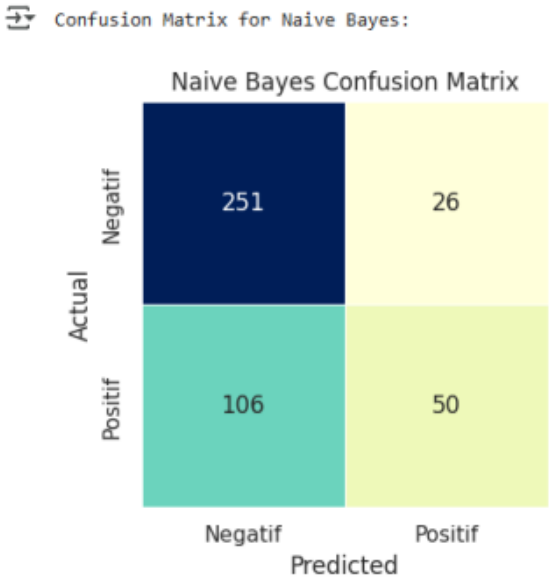

Fig. 3: Confusion Matrix

Based on the confusion matrix above, it can be seen that from all the test data, the Naïve Bayes model successfully classified 251 negative comments correctly, while 26 negative comments were misclassified as positive. Fifty positive comments were correctly recognized as positive, but 106 positive comments were misclassified as negative. From these results, it can be concluded that the model tends to be more accurate in recognizing positive comments than negative comments.

Fig. 4: Negative Comments

The negative sentiment word cloud above shows a collection of words frequently appearing in negative comments. The words "kalang," "abang," "driver," "kamu," and "ojol" appear most dominant because they are displayed in the largest size. This indicates that these words are most frequently used in negative comments. Other words such as "jangan," "order," "kerja," "polis," and "akun" also appear quite prominently. Generally, these words are used in the context of complaints, sarcasm, or protests against events perceived as detrimental or not meeting expectations. Words like "dog" and "pig" are commonly used as insults. Meanwhile, "smell," "fat," and "sombong" can be forms of physical or personality insults. The word "pecat" indicates comments suggesting or demanding action against someone (for example, a driver or Gojek employee). The presence of these words reinforces the fact that in sentiment analysis, negative sentiment does not only mean dissatisfaction but can also include expressions of hatred, insults, or verbal abuse.

Fig. 4: Positive Comments

Meanwhile, the positive sentiment word cloud above shows the most frequently occurring words in comments categorized as positive. The most dominant words are “sama”, “buat”, “orang”, “abang”, and “kerja”, which are displayed in the largest font size. This indicates that in positive comments, users frequently use words related to a sense of togetherness, work activities, and greetings or forms of respect such as "abang". Additionally, words like “polisi”, “jasa”, “pak”, “customer”, dan “driver" also frequently appear, which may reflect support for a particular party or service. This word cloud provides an overview of the word patterns frequently used in positive comments without having to read each comment individually.

3.2. Discussion

1. Device Requirements
   This research was conducted entirely online using a laptop with sufficient specifications to handle data scraping, preprocessing, and machine learning analysis. The hardware used included:
   a. Processor: AMD Ryzen 3
   b. RAM: 8 GB
   c. Storage: 512 GB SSD
   d. Operating System: Windows 10 (64-bit)

2. Results Used
   The dataset collected consisted of 2,217 Facebook comments related to Gojek drivers. After preprocessing and filtering, the data was transformed into numerical vectors using CountVectorizer, resulting in 3,667 unique features. The dataset was divided into 80% training data (1,731 comments) and 20% testing data (433 comments). The Naïve Bayes algorithm (MultinomialNB) was applied to classify the comments into positive and negative sentiment categories. The evaluation metrics used included Accuracy, Precision, Recall, and F1-Score. The results demonstrated that the model was able to classify negative comments with good reliability, especially in detecting recurring patterns of hate speech in the dataset.

3. Testing in This Research
   Testing was carried out using the testing dataset (433 comments) that had not been included in the training process. This approach ensured that the model’s performance was evaluated fairly on unseen data. The evaluation produced the following outcomes:
   a. Accuracy: measured the proportion of correctly classified comments against the total test data.
   b. Precision: indicated the percentage of comments predicted as negative that were truly negative.
   c. Recall: measured the ability of the model to detect all actual negative comments.
   d. F1-Score: provided a balanced measure between Precision and Recall.

The results showed that the Naïve Bayes model was effective in text classification for sentiment analysis of Facebook comments, demonstrating stable performance across the evaluation metrics.

4. Conclusion

This research analyzed hate speech and negative sentiment against Gojek drivers on the Facebook platform using the Naïve Bayes algorithm. A dataset of 2,217 comments was collected through web scraping with APIFY, preprocessed, and transformed into numerical

representation using CountVectorizer. The dataset was then divided into training and testing sets to evaluate model performance. The results show that the Naïve Bayes algorithm (MultinomialNB) is effective in classifying Facebook comments into positive and negative categories, with strong capability in detecting recurring patterns of hate speech. The model evaluation using accuracy, precision, recall, and F1-score confirmed the reliability of the approach. The findings demonstrate that social media contains a significant portion of negative comments toward Gojek drivers, which can affect their professional and psychological well-being. From a technical perspective, the device requirements and software environment used were sufficient to process and analyze the dataset efficiently. The testing stage confirmed the applicability of the Naïve Bayes algorithm in sentiment analysis for large-scale text data.

References

[1] I. S. Arfan, S. Fauziah, and I. Nawangsih, "Analisis Sentimen Terhadap Cyber Bullying di X Menggunakan Algoritma Naïve Bayes: Sentiment Analyst of Cyber Bullying in X Using Naïve Bayes Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1411–1419, 2024.

[2] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan rekayasa perangkat lunak*, vol. 3, no. 2, pp. 145–160, 2022.

[3] O. Belghaouti, W. Handouzi, and M. Tabaa, "Improved traffic sign recognition using deep convnet architecture," *Procedia Comput. Sci.*, vol. 177, pp. 468–473, 2020, doi: 10.1016/j.procs.2020.10.064.

[4] Sawitri, S., Simanjuntak, M., & Pardede, A. M. H. (2024). Application of Naive Bayes Method to Diagnose FMD Disease in Goats. *Journal of Mathematics and Technology (MATECH)*, *3*(2), 140-148.

[5] Lestari, S. A. M., Pardede, A. M., & Simanjuntak, M. (2024). Prediksi Disleksia pada Anak menggunakan Metode Naive Bayes. *Jurnal Kajian dan Penelitian Umum*, *2*(5), 37-51.

[6] BOYKE, G. M., AKIM, M. H. P., & RUSMIN, S. (2024). DIAGNOSA PENYAKIT PARU-PARU DENGAN METODE NAIVE BAYES. *SATURNUS: JURNAL TEKNOLOGI DAN SISTEM INFORMASI Учредители: Asosiasi Riset Ilmu Manajemen dan Bisnis Indonesia*, *2*(4), 268-276.