Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 06, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Tiktok Social Media Sentiment Analysis Using the Nave Bayes Classifier Algorithm

Putri Suci Rahmadani[1)*], Fenny Chintya Tampubolon[2)], Adelia Nurfattul Jannah[3)], Novia Lucky Halen Hutabarat[4)] , Allwin M. Simarmata[5)]
[1,2,3,4,5)] Prima Indonesia of Universitas, Medan, Indonesia
[1)]psuci1512@gmail.com, [2)]fennych19@gmail.com, [3)]adeliajannah00@gmail.com,
[4)]novialuckyhalenhutabarat101@gmail.com [5)]allwinsimarmata@unprimdn.ac.id

Submitted : July 25, 2022 | **Accepted** : July 25, 2022 | **Published** : July 28, 2022

Abstract: Social media is a computer application designed to make it simpler to communicate with others without having to do it face-to-face, as well as a tool for having fun and reducing feelings of isolation. Existing social media applications include games, music, and media for communicating with distant individuals, among others. These social media are utilized by parents, adolescents, and even young children. The application Tik-Tok is frequently used by children as a social networking platform. Tik-Tok has succeeded in grabbing the interest of youngsters, such that children are curious about creating short movies on the platform. Due to the fact that this application is used by children, the researcher seeks to use the *Naïve Bayes Classifier* Algorithm to recognize and differentiate unfavorable remarks on TikTok's social media. The rising number of negative remarks in the TikTok comments column can hinder the mental development of youngsters, and it is hoped that this algorithm would encourage users to post positive comments on this application. Based on the data gathering until the results of classification are obtained. There are 600 comments data randomly collected from TikTok users, gathered through the export comments website. After evaluating, the accuracy of the application of the *Naïve Bayes Classifier* algorithm in conducting sentiment analysis is 80% while the result of the AUC is 46%.

Keywords: social media; sentiment analysis; tiktok; negative comments; naïve bayes classifier;

INTRODUCTION

Social media is a program on a computer that makes it simpler for users to contact with others who are not next to them, because with social media, users may converse or send messages without having to meet or face each other. In addition, it can be utilized for recreation or relaxation. There are numerous applications available on social media, including music, gaming, remote communication, and how-to guides. Not only teenagers but even children and parents utilize this application (Mawengkang & Sitompul, 2022). Tik Tok is a popular application among young people. Tik Tok is an application that debuted in China before making its way to Indonesia in late 2017. (Zulqornain & Adikara, 2021).

Sentiment analysis is a computer investigation of the thoughts, feelings, and emotions expressed in text (Wilandini & Purwantoro, 2022). Various algorithms, notably the Naive Bayes algorithm, can be used to perform sentiment analysis. The Naive Bayes Classifier is a classification technique based on Bayes' theorem that employs statistical and probabilistic techniques, notably H. Prediction of probabilities based on prior experience (Bayes theorem) with a very strong (naive) assumption about each condition/event always being dependent (Pinka et al., 2022). Natural language processing (NLP) is the study of the interactions between computers and human (natural) language. It is a subfield of computer science, specifically computational linguistics. The *Naïve Bayes* method is used to calculate points for each available sentence in grouped comments to determine whether they are truly favorable, neutral, or even negative (Laurensz & Sediyono, 2021). This topic is extremely intriguing (Ahuja & Dubey, 2017) on the classification of film opinions based on positive sentiments, neutral sentiments, and even negative sentiments derived from Twitter comments (Indrayuni & Nurhadi, 2022). Using the deep learning methodology, sentiment analysis is performed. It is proposed to apply the Naive Bayes Classifier, the method used to detect unfavorable comments on Instagram, to ensure that the comments categorized previously fall into the negative or positive group (Yennimar & Rizal, 2019). Against

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 06, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

using Nave Bayes and Categorical Proportional Difference to forecast bad comments on tiktok. This research use 5-Cross Validation, with variants on the word, for its evaluation.

We propose, based on the findings of earlier research, to apply the Naive Bayes algorithm to analyze Tik Tok sentiment on social media using negative Tik Tok comment data. A range of qualitative research words are employed in this study's examination. good sentiment, negative sentiment, and neutral sentiment.

LITERATURE REVIEW

Naïve Bayes Classifier

Naïve Bayes a commonly used methodology for conducting sentiment analysis utilizing prediction algorithms based on Bayes' theorem. This method use supervised learning to do classification.

Confussion Matrix

The Confusion Matrix is utilized to determine the accuracy value. The confusion matrix is a frequently-utilized performance measurement technique. The confusion matrix compares the outcomes of the system's analysis against the expected results. This Confusion Matrix is appropriate for calculations with a dataset including both positive and negative classes. In the confusion matrix, the equation for calculating the values of accuracy, precision, recall, and f-1 score may be observed.

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+TP+FN+FP}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (4)$$

METHOD

Research methodology comprises of the activities or processes utilized by researchers to demonstrate the research process. This study employs the *Naïve Bayes Classifier* method to examine public sentiment regarding tiktok comments. This study's data processing procedure is depicted in the diagram below.

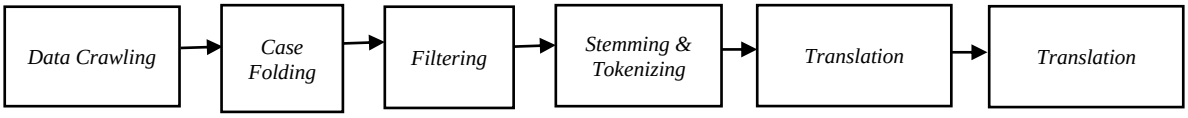

Fig. 1 Stages of Research Using Naïve Bayes Classifier Algorithm

Data Crawling

This study's data set was comprised of 600 random comments from tiktok users. The dataset is acquired from the webpage for export comments. These tiktok comments are derived from videos/content uploaded to tiktok that feature both negative and good remarks. After obtaining a sample of the negative and positive comment datasets on December 14, 2021, the data was entered into the naïve Bayes classifier (NBC) to calculate the score of the categorized commentary lines in order to determine whether the comments were negative or positive.

Case Folding

The case folding technique converts all characters in comments to lowercase (lowercase). The goal of case folding is to create identical character shapes, hence aiding sentiment analysis. In addition to converting characters to lowercase, the case folding procedure uses the regular expression library to remove digits, punctuation, and emoticons. After removing duplicate numerals, punctuation, and emojis, the researchers eliminated duplicate comments. Remove duplicate numbers, symbols, emoticons, and comments from the source code.

Filtering

The purpose of the filtering procedure is to remove unnecessary words by removing stopwords from the Sastrawi and nltk libraries. This is accomplished during the retrieval of key words. In addition to the stopwords already present in the Sastrawi collection, the researchers added the following words that, according to the researcher, have no significance.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 06, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Stemming&Tokenizing

Following filtration, the stemming procedure is executed. Using the Sastrawi dictionary, the stemming procedure reduces inflected words. Reducing inflection is intended to return words to their base form. Like the words "met" and "found," "find" will become "temu." The tokenization procedure consists of splitting the text into tokens. This token will be examined in the future. To perform tokenization, using the nltk.tokenize library.

Translation

Because the textblob library can only recognize English sentences, the comment classification will be translated into English prior to classification using the textblob library. To translate comments, the function = googletranslate(column name, "id", "en") is utilized.

RESULT

Marking comments on posts for your page on the social networking platform Tiktok, which were collected on December 14, 2021, was used to collect data. In this study, successfully gathered community comments are then imported into CSV format and classified according to two main groups, namely positive and negative comments. unfavorable comments. In the initial phase, all uppercase letters are changed to lowercase letters and emoticons are deleted. Table 1 shows the results of the case folding procedure.

Table 1. Hasil Proses Case Folding

Before Processing to Case Folding	After Processing to Case Folding
Love the way it makes you smile, bro. Jes	Love the way it makes you smile, bro. Jes
Brother Justin is so handsome. Bismillah. I can get his sister 😵	Brother Justin is so handsome. Bismillah. I can get his sister
Why does Brother Jes seem to be frozen with emotion? "😵" WKWKWKWKWK	Why does Brother Jes seem to be frozen with emotion? WKWKWKWKWK

Case-folded comments that have been further filtered to eliminate unnecessary punctuation and sentences. Table 2 displays the results of the procedure of comment filtering.

Table 2. Filtering Process Results

Before processing to filtering	After Processing to filtering
how come jes really freezes with expression wkwkwkwkwk	jes really freezes the expression
wkwkwkw I'm truly sorry; I used to rush.	I'm truly sorry; I used to rush.
eh no cut	no cut

After the filtering process, the next step is stemming and tokenizing. Stemming aims to change affixed words into basic words. Meanwhile, tokenizing acts as a separator between words to be used as tokens for analysis.

Table 3. The result processing of **Stemming&Tokenizing**

The comment before stemming dan tokenizing	The comment after stemming dan tokenizing
Jess, I simply adore the smile effect.	['Give', 'effect', 'Smile', 'so', 'bro', 'jes']
why is justin so handsome bismillah got his sister	['justin', 'Handsome', 'very', 'get', 'his sister']
Why does Jes actually look expressionless?	['jes', 'really', 'really', 'Freeze', 'expression']
bro jess cracks a little smile.	['bro', 'jess', 'smile', 'little]

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 06, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Following the filtering process are stemming and tokenizing. Stemming seeks to transform affixes into root words. In the meanwhile, tokenization serves as a delimiter between words that will be used as tokens for analysis.

Table 4. The Result processing of **Translation**

The comment which hasn't been translated yet	The comment which has already been translated
love the way it makes you smile, bro	just give the effect of smile bro
Justin has a sister and is such a lovely guy.	justin is so handsome to get the brother
Jes is an actual expression statue	jes is really a statue of expression
bang jess cracks a smile	bang jess smile a little

After feature extraction, the modeling is carried out using the Nave Bayes Classifier algorithm and the sklearn.naive bayes import MultinomialNB package. At the stage of modeling, the researcher uses 80% of the data for training and 20% for testing. There are six comments with negative sentiment, seventy-six comments with neutral sentiment, and eight comments with positive sentiment. The final step of the data process is the evaluation of the model, which follows the completion of data modeling. There are numerous metrics for evaluating models, including accuracy, precision, recall, and F1-Score. To determine the evaluation criteria for the model in this challenge, a confusion matrix is required. The Confusion Matrix, also known as the Error Matrix, is a specialized table that summarizes the performance of an algorithm or model.

Table 5. Precision Value Results, Recall, F1-Score

Sentiment	Precision	Recall	F1-Score
Negative	1.00	0.50	0.67
Netral	0.78	1.00	0.87
Positive	0.89	0.32	0.42

Table 6. Result Value Accuracy and AUC

Accuracy	0.80
AUC	0.46

After completing the evaluation phase, the researcher implemented the naive Bayes model that had been constructed using 565 tiktok comments that had been cleaned and translated. The application of Naive Bayes yielded 464 neutral comments, 76 positive remarks, and 26 negative comments. Figure 2 depicts the graph of the number of comments per category.

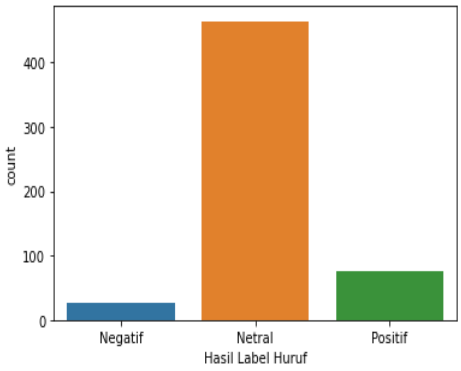

Fig. 2 Hasil Labeling

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

Using the Nave Bayes Classifier Algorithm, this study successfully analyzed the sentiment of Tiktok comments. Positive comments and negative remarks are utilized as test data in this research. After evaluation, the accuracy of the application of the Nave Bayes Classifier algorithm for sentiment analysis is 80%. While the findings of the AUC value are 46%, Based on the execution of the Nave Bayes algorithm on 565 comments that have through the cleaning process, 26 negative sentiment comments, 464 neutral sentiment comments, and 76 positive sentiment comments have been identified. Based on the obtained accuracy findings, the naive Bayes classifier algorithm can be evaluated for analyzing public sentiment on the social networking platform tiktok. To improve the effectiveness of naive Bayes in analyzing public sentiment, it is suggested that future study boost the accuracy of the results by increasing the amount of datasets or by employing additional approaches, such as gain ratios.

CONCLUSIONS

Using the *Naïve Bayes Classifier* technique, tiktok comments can be analyzed for sentiment. The accuracy of the application of the Nave Bayes Classifier algorithm for conducting sentiment analysis is 80%, with an AUC of 46%, according to the evaluation procedure. The *Naïve Bayes Classifier* algorithm was used to look at 565 comments that had already been cleaned up. The results were 26 negative comments, 464 neutral comments, and 76 positive comments. The use of the *Naïve Bayes Classifier* algorithm to the analysis of public sentiment on social media has been shown to be accurate.

REFERENCES

Ahuja, S., & Dubey, G. (2017). Clustering and sentiment analysis on Twitter data. *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, 1–5. https://doi.org/10.1109/TEL-NET.2017.8343568

Andika, L. A., Amalia, P., & Azizah, N. (2019). *Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier*. *2*(1), 34–41.

Fauzi, A., Akbar, M. F., Ferdi, Y., & Asmawan, A. (2019). *Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes*. *6*(1), 77–83.

Indrayuni, E., & Nurhadi, A. (2022). *Sentiment Analysis About COVID-19 Booster Vaccine on Twitter Using Deep Learning*. *7*(3), 900–905.

Kurniasari, I., & Fatta, H. Al. (2021). *Analisis Sentimen Opini Publik pada Instagram mengenai Covid-19 dengan SVM*. *1*(1), 67–74.

Kusuma, A., & Nugroho, A. (2021). *Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes*. *15*(2), 137–146.

Laurensz, B., & Sediyono, E. (2021). *Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the*. *10*(2), 118–123.

Mawengkang, H., & Sitompul, O. S. (2022). *Performance of Naive Bayes method with data weighting*. *7*(3), 817–821.

Nurhafida, S. I., Sembiring, F., Raya, J., & No, C. (2022). *Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)*. *6*, 317–327.

Nurhuda, F., & Sihwi, S. W. (2014). *Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier*. *2*(2).

Pinka, R., Dakhi, B., & Maezar, A. (2022). *Sentiment Analysis Of Public Opinions On The Effectiveness Of Online Learning Using Naïve Bayes Algorithm*. *6*(1), 273–279. https://doi.org/10.52362/jisicom.v6i1.822

Pintoko, B. M., & L, K. M. (2018). *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*. *5*(3), 8121–8130.

Putri, D. D., Nama, G. F., & Sulistiono, W. E. (2022). *Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (Dpr) Pada Twitter Menggunakan Metode Naive Bayes Classifier*. *10*(1), 34–40.

Sari, F. V. (2019). *Analisis Sentimen Pelanggan Toko Online Jd . Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi*. *10*(2), 681–686.

Verawardina, U., Edi, F., & Watrianthos, R. (2021). *Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes*. *5*(1), 157–163. https://doi.org/10.30865/mib.v5i1.2604

Wilandini, D., & Purwantoro. (2022). *Penerapan Algoritma Naïve Bayes Dalam Mengklasifikasikan Media Sosial Untuk Mengamati Trend Kuliner*. *8*(1), 31–39.

Yennimar, & Rizal, R. A. (2019). *Comparison of Machine Learning Classification Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency*. *4*(1).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 06, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Zulqornain, J. A., & Adikara, P. P. (2021). *Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorial Propotional Difference (CPD)*. *5*(7), 2886–2890.

*name of corresponding author