

Sentiment Analysis of Tiktok Shop Closure using Naïve Bayes Algorithm and Support Vector Machine

Ahmad Zamsuri^{*1}, Wileks²

¹²Informatics Engineering, Faculty of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia.
 * Corresponding Author.
 Email: ahmadzamsuri@unilak.ac.id

Received 13/06/2024, Revised 20/06/2024, Accepted 25/06/2024, Published 30/06/2024

This work is licensed under a <u>Creative Commons Attribution 4.0 International License</u>.

Abstract

 \odot

This research explores the closure of TikTok Shops in Indonesia triggered by the implementation of Minister of Trade Regulation (MOT) 31/2020. It focuses on analyzing user responses and sentiment towards this policy by utilizing Naïve Bayes and Support Vector Machine algorithms. While some netizens supported the move for security reasons, while others criticized it for limiting business opportunities, an analysis of 1000 datasets from Twetter with the keyword "close TikTok Shop" revealed that neutral sentiment dominated, indicating a lack of clarity or confidence regarding the reason for the closure. The results also showed that Support Vector Machine (SVM) had higher accuracy (94.4%) than Naïve Bayes (89.1%), signaling the superiority of SVM in classifying sentiment in this dataset. These findings provide deep insights into the public's perceptions and attitudes regarding the closure of TikTok Shop, providing an important basis for government and corporate understanding of the public's response to the policy.

Keywords: Sentiment Analysis, Tiktok, Tiktok Shop, Naïve Bayes, Support Vector Machine

Introduction

With the rapid advancement of technology along with the development of human civilization, internet access is increasingly widespread, enabling various media and platforms to support global communication. According to the Indonesian Internet Service Providers Association (APJII) report in 2022, internet penetration in Indonesia continues to increase, reaching 77.02 percent in 2021-2022, with Java Island as the largest user (43.92%) followed by Sumatra (16.63%) (Zulfiandri & Subiyakto, 2023).

TikTok, launched on September 20, 2016 by ByteDance, has become a social media platform that changes the way users interact, especially through the TikTok Shop feature. As an innovative E-Commerce platform, TikTok Shop contributes significantly to the online business ecosystem with its voucher policy, fast shipping, and free shipping. However, a fundamental change occurred when the Indonesian government on September 25, 2023 banned social media with multiple platforms, including TikTok Shop, based on the revised Minister of Trade Regulation (Yanti et al., 2023).



The closure of TikTok Shop on October 4, 2023 became a controversial event that triggered a variety of responses from Indonesian netizens. Sentiment analysis using Naive Bayes and Support Vector Machine algorithms (Sunaryanto et al., 2022) is expected to provide an in-depth understanding of the different user responses to this policy. This research aims to provide in-depth insight into the impact of the closure on the MSME market and public attitudes (Kurnianto & Febriawan, 2023).

Materials and Methods

Figure 1 explains the stages of research starting from Data Crowling, Dataset, Pre-Processing, Labeling Words, Weight Words, Naïve Bayes and SVM Classification, and Evaluation.



Fig 1. Phases of Research



1. Crowling Data

The author accessed data from the Twitter social media platform using the Twitter API with the Python programming language. The data retrieval process involves crawling Tweets that include the keyword "close TikTok Shop" in csv format.

2. Dataset

Data was obtained from *Twitter* through a crawling process with the keyword "close TikTok *Shop*" with a total of 1011 data. The *crawling* process involves retrieving information about users and *Tweets*, with a step of removing duplicate content to ensure data uniqueness. The results are automatically saved in a .csv file.

3. Preprocessing

The *pre-prosecution* step involves a series of steps, such as cleaning, typesetting, word splitting, filtering, and root pruning.

- a. Cleaning involves removing attributes such as usernames, hashtags, URLs, punctuation marks, and symbols.
- b. Typesetting is done to equalize text in lowercase.
- c. Word splitting involves the process of dividing each word in a sentence into separate word units.
- d. Filtering removes common but irrelevant words (stopwords).
- e. Word root pruning is done to convert words into their base form by removing affixes.

4. Data Labeling

The data labeling process is an important step to identify positive, negative, and neutral sentiments in the text. It allows the machine learning model to understand and learn the patterns associated with each label.

5. Weighting

The use of *CountVectorizer* helps to count the frequency of occurrence of each word in the document, resulting in a numerical vector. Each dimension represents a word from the vocabulary, and the value of each dimension indicates the frequency of the word in the document.

6. Algorithms

This research uses two algorithms *Naïve Bayes* and *Support Vector Machine* (SVM) in testing this research.

- *Naïve Bayes* algorithm is used to classify based on the probability of each class. This model is very effective for sentiment analysis and has been proven in various studies (Aldisa & Maulana, 2022). *Naive Bayes*, a classification method based on Bayes' theorem, is used in sentiment analysis because of its ability to predict class membership under the assumption of independence. Its advantages lie in efficiency (Lisnawita et al., 2022), making the sentiment analysis process shorter, and its good accuracy even with limited training data (Tanggraeni & Sitokdana, 2022).
- b. Support Vector Machine (SVM) functions to select the best separating function to separate two types of objects. SVM is used to classify data in two classes, with the One-Against-One method for multi-class classification (Styawati et al., 2021). Support Vector Machine (SVM) is a learning system that utilizes linear



functions in large features with optimization theory-based learning algorithms. Introduced by Vapnik in 1992, SVM became a collection of leading ideas in pattern recognition (Agustini et al., 2022).

c. Confusion Matrix is used to evaluate the performance of classification algorithms in sentiment analysis. It consists of four components: *True Positive* (TP) for correct prediction and correct actual value, *True Negative* (TN) for correct prediction and correct actual value, *False Positive* (FP) for incorrect prediction but correct actual value, and *False Negative* (FN) for incorrect prediction but correct actual value (Kusuma & Cahyono, 2023).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} X \ 100\%$$

$$Precision = \frac{TP}{FP + TP} X \ 100\%$$

$$Recall = \frac{TP}{FN + FP} X \ 100\%$$

7. Evaluation

Evaluation in this research aims to measure the performance or accuracy of sentiment analysis methods. This assessment helps understand the extent to which a technique or algorithm can recognize and interpret sentiment or opinion from text.

Results and Discussion

Retrieving data from Twitter (Crowling)

Data collection in this study was carried out by crowling data from Indonesian twitter comments with the keyword close tiktok shop using google colab with data totaling 1011 data.

Pre-Processing

This pre-processing stage consists of 5 stages, namely:

- 1. Cleansing, is a procedure for cleaning attributes from hastags, URLs, punctuation marks and symbols.
- 2. Case folding, is the process of homogenizing the contents of tweets into lower case alphabetic characters.
- 3. Tokenizing, is the process of breaking each word of a sentence into word units.
- 4. Filtering, is the process of removing words that often arise but are not fundamental or have no influence (stopwords).
- 5. Stemming, At this stage, the words in the document will be converted into their basic form b removing affix words.

Table 1. Pre-processing Results



Full text	Cleansing	Case folding	Tokenizing	Filtering	Steaming
@guk_guk_gukk waktu sebelum tutup tiktokshop aku pernah dapet fitlife isolate cm 500rb an padahal aslinya bisa 700rbanbeneran bakar duit sih pas itu	Waktu Sebelum Tutup TiktokShop aku pernah dapet fitlife isolate cm rb an padahal aslinya bisa rb an beneran bakar duit sih pas itu	waktu sebelum tutup tiktokshop aku pernah dapet fitlife isolate cm rb an padahal aslinya bisa rb an beneran bakar duit sih pas itu	[waktu, sebelum, tutup, tiktokshop, aku, pernah, dapet, fitlife, isolate, cm, rb, an, padahal, aslinya, bisa, rb, an, beneran, bakar, duit, sih, pas, itu]	[tutup, tiktokshop, dapet, fitlife, isolate, aslinya, beneran, bakar, duit, sih, pas]	tutup tiktokshop dapet fitlife isolate asli beneran bakar duit sih pas



Gambar 1. Word Cloud Sentimen

In Figure 2, is a sentiment *Word Cloud* that displays words that often appear in data that has been completed in stemming where the bigger the word eats it shows that the word appears most often in sentiment data and vice versa what if the word that appears is small then it shows that the word rarely appears in sentiment data.

1. Labeling

At this stage the dataset that has gone through the *pre-processing* stage will be labeled with sentiment (positive, negative, or neutral) using the text *blob library*. the labeling process uses the *textblob library*, where the first thing to do is create a textblob object with text and then calculate the sentiment polarity of the text and then assign a sentiment category based on the polarity value, which if the polarity is greater than 0 is included in the



positive sentiment, smaller than 0 is included in the negative sentiment and if neither will be included in the neutral sentiment.



Figure 3. Sentiment Count

Figure 3 is a graphic image of the number of percentages of all data that has been labeled, namely neutral sentiment data 877, positive 103, negative 31.

Weighting

After the labeling process the next process is CountVectorizer weighting, CountVectorizer calculates the frequency of occurrence of each word in each document from the vocabulary that has been built. This produces a numeric vector for each document, where each dimension of the vector represents one word from the vocabulary, and the value in each dimension is the frequency of that word in that document.

Algorithm Implementation

Before getting into the implementation of the algorithm the first thing to do is to divide between training data and test data. Models that are trained with a larger volume of training data will have a higher ability to recognize patterns, capture relationships between features, and produce accurate predictions. This contributes to improved model performance when faced with test data. The proportion of 70% training data and 30% test data is a general rule used, therefore the data has been divided into training data and test data where 70% training data and 30% test data and 30% test data and 30% test data then the result is 707 training data and 304 test data.

Naïve Bayes Algorithm

In the case of the TikTok Shop closure, the Naive Bayes Algorithm can be used to categorize the sentiment of user reviews or comments related to the closure. For example, this algorithm can model the likelihood that a review will fall into the positive, negative, or neutral category based on the words contained in it.





Gambar 2. Confusion Matrix Naïve Matrix

Figure 4 is the result of testing the Confusion Matrix of the Naïve Bayes algorithm where the Negative word results in 2, for the Neutral word 256 and the Positive word 13. From this test it can be concluded that the Neutral word is more than the Positive and Negative words.

Support Vector Machine

Support Vector Machine (SVM) is a learning system with linear functions in broad features, trained using optimization algorithms. Introduced by Vapnik in 1992 as a leading idea in pattern recognition (Herlinawati et al., 2020). SVM can be used to build a model capable of distinguishing positive, negative, and neutral reviews when TikTok Shop is closed. To maximize the distance between sentiments, the SVM model will search for the ideal hyperplane. SVM can help categorize reviews into appropriate sentiment categories by using relevant user review features.







Figure 5 is the result of the Confusion Matrix test of the Support Vector Machine Algorithm where the Negative word results are 5, for the Neutral word 262 and the Positive word 20. From this test it can be concluded that the Neutral word is more than the Positive and Negative words.

Evaluation

By conducting an evaluation, research can provide more reliable results about the extent to which the Naive Bayes and Support Vector Machine (SVM) models can predict user sentiment about the closure of TikTok Shop. This analysis can help researchers and readers understand the advantages and disadvantages of the two classification models used.



Figure 6. Evaluation of Naïve Bayes and Support Vector Machine

In Figure 6, the accuracy, precision, and recall values of the *Naïve Bayes* and *Support Vector Machine* algorithms are visualized where the accuracy of *Naïve Bayes* is 89.1% and *Support Vector Machine* is 94.4%. Precision of *Naïve Bayes* 75.3% and *Support Vector Machine* 91.1%. *Recall of Naïve Bayes* 53.4% and *Support Vector Machine* 72.6%.

Conclusion

Sentiment analysis of the TikHub Shop closure helps one to make certain inferences. First, neutral sentiment rules over positive and negative emotion, indicating that many people are not very delighted or outraged and could therefore be perplexed or puzzled about the cause of the closure. Second, with an accuracy of 94.4% compared to Naïve Bayes which only achieved 89.1%, the Support Vector Machine (SVM) technique demonstrated a greater performance in categorizing the sentiment in the utilized dataset. Thirdly, SVM precision (91.1%) is greater than Naïve Bayes' (75.3%) demonstrating that SVM has better capacity to precisely identify different emotions. Finally, SVM's recall of 72.6% is also higher than Naïve Bayes' 53.4%, suggesting that SVM is better at identifying and classifying instances that ought to belong in positive, negative, and neutral emotions. These findings highlight SVM's excellence on the dataset used in sentiment analysis.



References

- Putra, A. E. Analisis Sentimen Tanggapan Masyarakat Kepada Video Youtube Mengenai Resesi 2023 Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor (Knn) (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta)
- S. D. Yanti, S. Astuti, and C. Safitri, "Pengaruh Pengalaman Belanja Online Dan Kepercyaan Terhadap Minat Beli Ulang Di Tiktok Shop (Studi Kasus Mahasiswa Fkip Uhamka 2018)," J. EMT KITA, vol. 7, no. 1, pp. 47–61, 2023, doi: 10.35870/emt.v7i1.728.
- E. Kurnianto and D. Febriawan, "Analisis Sentimen Perbedaan Pendapat Netizen Indonesia Terhadap Penutupan Tiktok Shop Menggunakan Algoritma Naïve Bayes," vol. 5, pp. 404–414, 2023, doi: 10.30865/json.v5i2.7170.
- R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," Build. Informatics, Technol. Sci., vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," J. Inform. J. Pengemb. IT, vol. 6, no. 3, pp. 150– 155, 2021, doi: 10.30591/jpit.v6i3.2870.
- N. W. S. Agustini, D. Priadi, and R. V. Atika, "Profil Kimia dan Aktivitas Antibakteri Fraksi Aktif Nannochloropsis sp. sebagai Senyawa Penghambat Bakteri Penyebab Gangguan Kesehatan Mulut," J. Pascapanen dan Bioteknol. Kelaut. dan Perikan., vol. 17, no. 1, p. 19, 2022, doi: 10.15578/jpbkp.v17i1.781.
- Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor. Jurnal Informatika/Jurnal Informatika: Jurnal Pengembangan IT, 8(3), 302–307. <u>https://doi.org/10.30591/jpit.v8i3.5734</u>.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. CESS (Journal of Computer Engineering, System and Science), 5(2), 293. <u>https://doi.org/10.24114/cess.v5i2.18186</u>.
- Sunaryanto, H., Hasan, M. A., & Guntoro, G. (2022). Classification Analysis of Unilak Informatics Engineering Students Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3), Random Forest and K-Nearest Neighbors (KNN). IT Journal Research and Development, 7(1), 36–42. <u>https://doi.org/10.25299/itjrd.2022.8912</u>
- Lisnawita, L., Guntoro, G., & Musfawati, M. (2022). Implementation of Naïve Bayes for Classification of Learning Types. Digital Zone/Digital Zone, 13(1), 44–54. https://doi.org/10.31849/digitalzone.v13i1.9825