Journal of Theoretical and Applied Information Technology
15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SENTIMENT ANALYSIS FOR TIKTOK SHOP'S CLOSURE IN INDONESIA USING NAIVE BAYES MODELS AND NLP.

HENOCH JULI CHRISTANTO[1], STEVEN SONDRA ALLEN WIDODO[2], CHRISTINE DEWI[3], YERIK AFRIANTO SINGGALEN[4], DALIANUS RIANTAMA[5], ANDRI DAYARANA K. SILALAHI[6]

[1]Atma Jaya Catholic University of Indonesia, Information System Department, Faculty of Engineering, Jakarta, Indonesia
[2,3]Satya Wacana Christian University, Faculty of Information and Technology, Salatiga, Indonesia
[4]Atma Jaya Catholic University of Indonesia, Tourism Department, Faculty of Business Administration and Communication, Jakarta, Indonesia
[5]Bina Nusantara University, Management Department, BINUS Business School Undergraduate Program, Jakarta 11480, Indonesia
[6]Chaoyang University of Technology, Department of Business Administration, Taiwan

E-mail: [1]henoch.christanto@atmajaya.ac.id, [2]672019362@student.uksw.edu, [3]christine.dewi@uksw.edu, [4]yerik.afrianto@atmajaya.ac.id, [5]dalianus.riantama@binus.ac.id [6]andridksilalahi@gmail.com

ABSTRACT

Sentiment Analysis (SA), or opinion mining, is a task in natural language processing (NLP) that entails identifying the sentiment conveyed in a text, such as positive, negative, or neutral. Multiple methodologies and strategies exist for conducting sentiment analysis, from conventional procedures to more sophisticated machine-learning techniques. This study applies Sentiment Analysis (SA) techniques with NLP approaches to gauge sentiments related to TikTokShop’s closure in Indonesia. The study uses Twitter data to analyze sentiments using different algorithms such as the Multinomial Naive Bayes, the Bernoulli Naive Bayes, and the Complement Naive Bayes. Moreover, it utilizes a Count Vectorizer and TF-IDF Vectorizer to enhance sentiment analysis. Furthermore, using TextBlob with the CountVectorizer approach is the most accurate at 86.60% in sentiment classification. The analysis sheds light on sentiment analysis techniques applicable to TikTokShop closure as well as which algorithm and vectorization approach can be used to measure sentiments derived from the Twitter data.

Keywords: *Sentiment Analysis, TikTokShop Closure, Twitter Data, TextBlob*

1. INTRODUCTION

The current landscape has seen a lot of technological advancements and digital innovation which has revolutionized different business areas. Concerning the prolonged pandemic, digitalization is crucial in helping businesses survive and grow during these times. Established entrepreneurs and aspiring must understand these aspects, especially ones concerning digital data utilization. There is a vast amount of data available in the digital world that helps companies grow their businesses and can also help one identify business opportunities. Another important point is using social media as a treasure trove for brand preferences, product choices, and opinionated opinions about brands, products, and events. In this digital era, social media is the place where people are allowed to say anything about anything freely. Nevertheless, for all its worth, social media data analysis is paramount to realize the real benefits of this wealth of information. This refers to a process that entails careful acquisition and extraction of available data, organizing it in line with tangible patterns, and finally utilizing this arranged info for making smart presumptions [1].

E-commerce growth in Indonesia, one of the biggest countries in SouthEast Asia, gave rise to an abundance of e-commerce platforms offering different advantages and creating stiff competition among significant companies across this sector. For instance, the emergence of TikTok shop in the TikTok app illustrates this trend [2]. TikTok shop is a social media shopping platform that mixes e-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

commerce with social media experiences. They offer new lifestyle buying options [3]. TikTok shop is a game-changing new shopping solution that allows creators, brands, and merchants to promote and sell items directly through TikTok apps. In-feed videos, live streaming, and a page for product showcases are all available for sale activities [4]. Their user experience and interactivity are innovative. Currently, TikTok Shop has closed in Indonesia because of numerous issues such as the privacy of personal data and problems within the system [5]. As a result, this has brought about huge disruptions in the country's E-commerce industry with these trade statutes implemented by the Ministry of Trade [6].

The effect of closing down the Tiktok shop exceeds regulatory challenges, affecting marketing firms and businesses involved leading to the loss of jobs. Furthermore, sellers and businesses heavily reliant on TikTok Shop as their primary sales channel are now compelled to seek alternative platforms for their operations. Moreover, this closure is not limited to the economy, as it may change the very way people interact or conduct businesses in a society, especially those who are actively involved on the platform. However, the closure brings along several complicated consequences in different areas of life, potentially evoking diverse emotional responses and sentiments within the community. A case study of this event portrays a connection between regulatory decisions, business practices, and societal consequences of the Indonesian digital commerce domain.

Sentiment analysis, a research field that delves into how individuals express their thoughts, emotions, evaluations, attitudes, and emotional responses toward entities and their attributes through written text, involves employing natural language processing (NLP) techniques to categorize textual data into three primary sentiments: positive, negative, and neutral [7, 8]. It also analyses textual data retrieved from the sites, news, reviews, opinions, and the product's description [9].

The use of sentiment analysis is very important across different sectors of trade, especially when it comes to the management of some parts of strategic plans as well as customer relations. This analytical framework is broadly applied in rating consumer insights, reviewing customer feedback, and improving services by modifying the existing product elements. Businesses can evaluate consumers' sentiments through their reviews. In this way, they can deduce satisfaction in various aspects of business processes [10, 11]. Moreover, sentiment analysis is an important tool in market research that helps to understand what consumers think and feel about particular goods, services, or brands. Market research, competitive intelligence, and supporting product development strategy, three are incomparable components that offer priceless tips [12, 13]. Businesses use sentiment analysis for monitoring brand sentiments in social media and internet forums. The approach is proactive and allows for timely detection of problems involving brand image to take quick action and mitigate such risks [14, 15]. Furthermore, Sentiment analysis is a full-scale in-depth review of feelings, opinions, attitudes, or perspectives communicated through various social media sites. At the same time, in the sphere of social networks, people act as social sensors, sharing materials that reflect their feelings, views, and opinions [16].

This research focuses on analyzing sentiments linked to the closure of TikTokShop in Indonesia using a Twitter dataset. Its main contributions lie in utilizing Twitter data to explore sentiments surrounding TikTokShop's closure, identifying the most accurate model compared to Naive Bayes Algorithm through experimental investigation, and employing various classifiers (like VADER and Text Blob) and methodologies (including feature selection and extraction techniques such as Multinomial NB, Bernoulli NB, Complement NB, Count Vectorizer, and TFIDF Vectorizer).

The study is divided into five sections: the second reviews Sentiment Analysis literature, the third details the methodology, the fourth presents and discusses experimental results, and the fifth concludes with a summary and outlines potential future research directions.

2. RELATED WORK

2.1 Sentiment Analysis and Naïve Bayes

Sentiment analysis is a natural language processing process that captures and categorizes sentiments from textual data. The mood, statements as well as subjective data are reviewed in the text. In essence, sentiment analysis has found application in different sectors, such as finance, and determining schools, among others, and in automated business analytics [17–19]. Wongkar et al.[20] conducted a sentiment analysis on Twitter during the electoral politics around the presidential election of the Republic of Indonesia. For instance, a naïve Bayes, KNN & SVM algorithm classification was used to compare research findings. Friska et al [21] were involved in using the Naïve Bayes and SVM algorithms to analyze sentiment within the TikTok app. Sentiment analysis has grown a lot. It uses different methods like dictionary guides and machine learning systems. All these improvements aim to make sentiment analysis ways more precise [22].

Naive Bayes is a popular machine learning technique that is often utilized for tasks involving classification, particularly in the realm of natural language processing and text classification. It is based on Bayes' theorem, a mathematical concept that calculates the likelihood of an event occurring based on prior knowledge of similar events. The name "naive" comes from the algorithm's assumption of independence among features, which means it assumes that the presence of one feature in a class has no bearing on the presence of other features. Despite this simplified approach, Naive Bayes is widely praised for its simplicity, speed, and effectiveness in a variety of classification tasks, such as sentiment analysis, spam filtering, and document categorization.

Naive Bayes in sentiment analysis examines the occurrences of words or features within the text to make predictions about the sentiment expressed. It can distinguish between positive, negative, and neutral sentiments by learning from the frequencies and patterns of words associated with each sentiment class in the training data. The Naive Bayes (NB) algorithm falls under supervised learning, implying that it requires prior labeled data to make predictions or decisions. One key advantage of employing the Naive Bayes algorithm is its approach to decision-making without the need for numerical optimization methods [23]. Various advancements in Naive Bayes (NB) classifiers have led to improved discrimination capabilities, with one notable development being the Regularized Naive Bayes (RNB) method. RNB demonstrates excellent performance by effectively balancing discrimination power and generalization capability. Notably, data discretization plays a crucial role in the effectiveness of Naive Bayes classifiers [24].

The Bayes theorem, specifically in Naive Bayes classifiers, calculates the conditional probability of a label given the observed features. These pre-processing techniques assist in creating feature representations from the text, enabling the Naive Bayes classifier to estimate the likelihood of a label given the observed features more effectively. This calculation is represented by the Equation (1) [25].

$$P(label|features) = \frac{P(label)*P(features|label)}{P(features)} \quad (1)$$

where P(label | features) denotes the probability of a label given the observed features. P(label) signifies the prior probability of the label. P(features | label) represents the probability of observing the features given the label, and P(features) is the probability of observing the features. This formula is fundamental in the context of Naive Bayes classifiers for making predictions or classifications based on observed data.

2.2 TikTok Shop

TikTok Shop is a feature within the TikTok social media platform that integrates e-commerce functionalities with social media experiences. It allows creators, brands, and merchants to showcase and sell products directly within the TikTok app. This feature enables users to seamlessly discover, explore, and purchase various items while engaging with TikTok's content [26].

The platform's serious commitment to TikTok Shop is evident through robust support, offering discount vouchers and promotional assistance not only to buyers but also extending benefits to sellers. These benefits include video promotion and live selling features, demonstrating TikTok's dedication to fostering a thriving e-commerce environment [27]. Public sentiment surrounding TikTok Shop extends beyond the platform itself, with discussions and opinions frequently expressed on other social media platforms like Twitter [28]. This broad feedback serves as the basis for authors to conduct sentiment analysis regarding public perception of TikTok Shop, which they then share on their Twitter page [29].

The interplay between TikTok Shop's features, TikTok's commitment, and public sentiments expressed across various social media platforms forms a dynamic ecosystem, influencing both user engagement and the platform's evolution in the realm of social commerce.

2.3 Twitter

Twitter is one of the most common platforms for social networking and enables people to share their views in the form of short messages not exceeding 280 characters long which may consist of any combination of texts, images, videos, links, and tags. It enables customers to follow the relevant profiles, make posts or comments, give likes, and retweet others' statements among others. As a result, Twitter has become a platform for sharing real-time

Journal of Theoretical and Applied Information Technology
15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

information about news, opinions, and various content with all people worldwide. Active internet users aged 25 to 34 prefer to use this popular social networking platform as compared to other options. As a medium for sharing information, in 2021, Twitter had around 397 million monetizable active users and 187 million daily followers, proving that it is still very popular [30].

To facilitate data collection, Twitter offers Application Programming Interfaces (APIs) that require users to obtain four keys: the consumer key, the consumer secret, the access token, and the access secret [31]. These keys provide proof of who a user is so that they can have secure access to Twitter's data such as tweets, profile information, and other confidential pieces of information. Twitter's API is an essential instrument in acquiring user-generated information.

2.4 Performance Evaluation

The aim of evaluating performance was to gauge how well the model could accurately understand and classify sentiments expressed on Twitter regarding the closure of the TikTok Shop. The study used the F1 score and accuracy score, alongside their respective class support divisions, as key evaluation metrics. Accuracy score and precision were defined using specific formulas (Formula 2 and Formula 3), while the formulas (Formula 4 and Formula 5) determined the recall and F1 metrics, according to the reference provided, these formulas were utilized or derived for performance assessment [32].

$$Accuracy\ score = \frac{TP+TN}{TP+TN+FP+F} \quad (2)$$

$$Precision = \frac{TP}{TP+F} \quad (3)$$

$$Recall/Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (5)$$

The accuracy formula evaluates the overall correctness, by finding the ratio of all accurate predictions made concerning positive items as well as negative ones by total dataset. Precision comes after and measures the relationship between the correctly predicted positive instances called true positives, TP, and predictions in general. Recall is another measure that expresses true positive instances relative to the amount of correctly predicted positive instances. Additionally, the F1-Score constitutes an even indicator that compares the weighted mean of precision and recall [33].

True Positive (TP) represents the count of reviews correctly categorized into their respective sentiment classes, while False Positive (FP) indicates reviews incorrectly assigned to a sentiment category they don't belong to. Conversely, False Negative (FN) denotes reviews mistakenly labeled as not belonging to a sentiment class when they do [34].

3. METHODOLOGY

3.1 Research Design

As seen in Figure 1, this research entails phases used for an extensive sentiment analysis. First, the Twitter dataset is collected from Twitter API where specific pages are scraped for the data to be gathered [35]. The data is very comprehensive as it passes through rigorous cleansing and preparation that comprises lower casing, tokenization; punctuation removal; elimination of numbers, specials, etc. Data labeling is done using tools such as Vader Sentiment and TextBlob after the preprocessing phase. Processing of the preprocessed data into subsequent stages is carried out using feature extraction and selection methods like TF-IDF Vectorizer, and Count Vectorizer among others. This entails identifying various Naïve Bayes algorithms such as Multinomial Naïve Bayes, Bernoulli Naïve Bayes, and Compliment Naïve Bayes. The last stage of this study is to compare these Naive Bayes algorithms which makes up the concluding section of this research methodology. The method adopted is compatible with the process of cleaning, pre-processing, feature extraction, and algorithmic analysis presented at the beginning part describing the research design section.

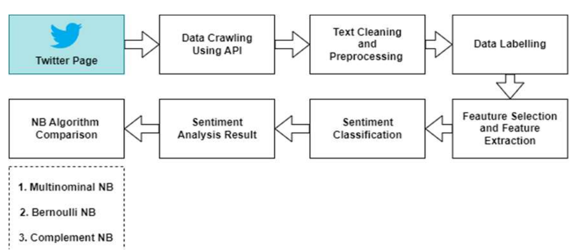


Figure 1: The research workflow

Figure 2 presents the Comparison of Sentiment Classification in this research study. The experimentation process was divided into different libraries for data labeling, namely TextBlob and VADER Sentiment. TextBlob is a Python library used for processing textual data to determine sentiment polarity, while VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically

Journal of Theoretical and Applied Information Technology
15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

designed for social media text. Additionally, this research delved into diverse feature selection and extraction methods, employing TF-IDF Vectorizer and Count Vectorizer. These methods were integrated with the implementation of various Naive Bayes algorithms, including Multinomial NB, BernoulliNB, and ComplementNB, to gauge their effectiveness in sentiment classification.

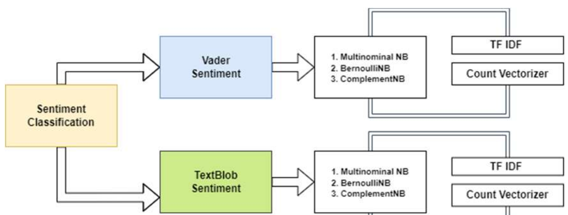


Figure 2: Comparison of Sentiment Classification

3.2 Datasets

Our research utilized Twitter as our primary dataset, collecting data via the API from specific pages using tweet harvest, and filtering content based on specified keywords and a defined date range. The scraped tweets were subsequently stored in a CSV file for further analysis. To execute this script, a valid Twitter account is required, along with an Access Token obtained by logging into Twitter via a web browser and extracting the auth_token cookie. Employing the keyword for example "TikTok shop Ban in Indonesia, TikTok shop Tanah Abang, TikTok shop UMKM" we initially gathered over 5000 tweets related to the closure of TikTok Shop in Indonesia. However, after rigorous cleaning and pre-processing steps, our dataset was refined to approximately 3000 tweets, which served as the foundation for our sentiment analysis regarding the TikTok Shop closure in Indonesia. Table 1 shows the example of the tweet of TikTok Shops closure.

Table 1: Tweet Of Tiktokshop's Closure

No	Tweet	Sentiment
1	*This is my opinion. I personally agree that the government should shut down TikTok Shop before tackling loan services' accounts. The reason being, TikTok Shop is both disturbing and threatening to our micro, small, and medium enterprises (UMKM). Remember the news about the Tanah Abang market incident where buyers were cheated, despite attempting live streaming to verify the products?*	Positive
2	*The closure of TikTok Shop fundamentally does not significantly impact the resurgence of Tanah Abang or other trading centers. Its impact primarily affects online stalls whose market share was taken over by TikTok Shop.*	Neutral
3	*Due to TikTok Shop, many Micro, Small, and Medium Enterprises (UMKM) entrepreneurs, especially millennials like myself, who were just starting out, experienced a lack of income. Moreover, Indonesia is not just about Tanah Abang.*	Negative

Figure 3 portrays the Twitter dataset overview of the positive and negative classes. when the different categories in the dataset (like positive, negative, and neutral sentiments) are equally represented, it ensures that the performance measurements are trustworthy across all these categories. However, if the dataset is heavily imbalanced, with one category dominating (for instance, mostly positive or negative reviews and very few neutral ones), relying solely on accuracy metrics might not provide valuable insights. This imbalance makes it easier for the model to predict the more frequent categories accurately, affecting the overall reliability of the accuracy metric.

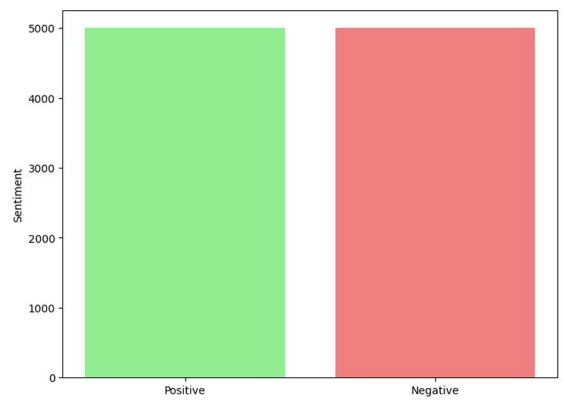


Figure 3: Twitter Dataset overview of Positive and Negative class

4. EXPERIMENT AND RESULT

4.1 Experiment Results

The experiment results for Sentiment Classification using CountVectorizer with the VADER and TextBlob libraries are in line with the findings showcased in Table 2. When leveraging the VADER sentiment library, the accuracy metrics observed were as follows: Multinomial NB achieved 67.20%, BernoulliNB reached 66.24%, and ComplementNB obtained 70.56%. Conversely, utilizing TextBlob demonstrated higher accuracy rates: Multinomial NB achieved 86.60%, BernoulliNB attained 85.92%, and ComplementNB reached 83.56%.

The use of Sentiment Classification accuracy metrics evaluated for the integration of CountVectorizer with VADAR and TestBlob libraries showed clear differences. In this particular case, text blob generally outperformed all Naive

Bayes algorithms than Vader. The Multinomial NB model proved to be the most accurate in overall performance with either sentiment database, indicating the possible usefulness of this algorithm for sentiment analysis within this specific scope.

Table 2: Performance Result of Sentiment Classification Using Count Vectorizer

Items	Vader Sentiment			TextBlob		
	Count Vectorizer			Count Vectorizer		
Metrics	MNB	BNB	CNB	MNB	BNB	CNB
Accuracy	0.67	0.66	0.71	0.87	0.86	0.83
Sensitivity	0.56	0.77	0.69	0.96	0.98	0.95
Precision	0.69	0.68	0.63	0.95	0.95	0.96
F1	0.62	0.72	0.66	0.97	0.97	0.96

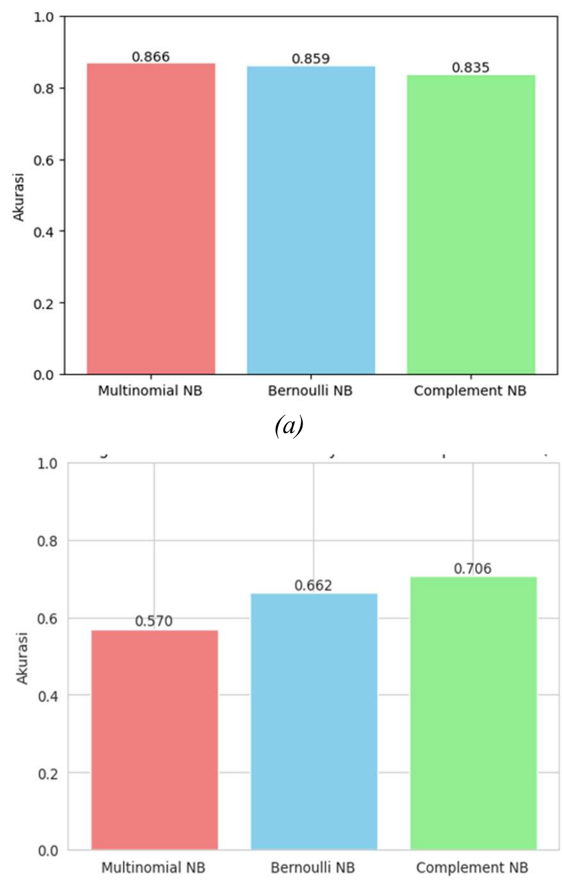
The experiment results for Sentiment Classification using TF-IDF Vectorizer with the VADER and TextBlob libraries align with the findings illustrated in Table 3. When utilizing the VADER sentiment library, the observed accuracy metrics were as follows: Multinomial NB achieved 56.48%, BernoulliNB reached 66.72%, and ComplementNB attained 70.24%. Conversely, employing TextBlob with TF-IDF Vectorizer demonstrated higher accuracy rates: Multinomial NB achieved 84.04%, BernoulliNB attained 83.41%, and ComplementNB reached 81.51%.

The Sentiment Classification accuracy metrics obtained, as outlined in the experiment results, showcase discernible differences between the VADER and TextBlob libraries when integrated with TF-IDF Vectorizer. Notably, TextBlob consistently outperformed VADER across all Naive Bayes algorithms assessed in this study within the TF-IDF Vectorizer setup. The Multinomial NB model again displayed the highest accuracy rates among the evaluated models, highlighting its potential effectiveness for sentiment analysis tasks in this specific experimental framework.

Table 3: Performance Result of Sentiment Classification Using TF-IDF Vectorizer

Items	Vader Sentiment			TextBlob		
	TF – IDF Vectorizer			TF – IDF Vectorizer		
Metrics	MNB	BNB	CNB	MNB	BNB	CNB
Accuracy	0.57	0.67	0.71	0.84	0.83	0.81
Sensitivity	0.62	0.71	0.71	0.98	0.95	0.95
Precision	0.69	0.77	0.70	0.95	0.95	0.96
F1	0.75	0.71	0.77	0.97	0.96	0.96

Figure 4(a) illustrates the comparison between Naive Bayes algorithms using TextBlob, while Figure 4(b) showcases the comparison using the VADER Sentiment library. The findings indicate that TextBlob consistently achieves higher accuracy scores for Naive Bayes algorithms compared to VADER Sentiment. This difference in accuracy can be explained by the dataset mainly consisting of tweets from Indonesia. TextBlob's superior performance can be attributed to its capability to handle sentiment analysis for informal and brief text, which are common traits found in tweets. TextBlob employs a pre-trained sentiment analysis model equipped with a more extensive lexicon and ruleset, specifically tailored to capture nuanced sentiments in informal and conversational language often seen in social media content like Indonesian tweets. Conversely, VADER Sentiment can perform well in sentiment analysis for texts from social media; however, because it relies on a lexicon and rule-based approach its ability to detect subtitles and complicated details relevant only to the Indonesian language might be limited in comparison to TextBlob.



(a)

(b)

Figure 4: Naive Bayes Algorithm Bar Comparison. (a) NB Comparison using TextBlob, (b) NB Comparison using VADER

Journal of Theoretical and Applied Information Technology
15th April 2024. Vol.102. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This study classifies performance using NB, which demonstrates ease of use, capacity to efficiently process multidimensional texts, as well as successful natural language processing applications. For various kinds of textual data, the three models Multinomial NB, BernoulliNB, and ComplementNB were used as algorithm options due to the individual assumptions of the models. Multinomial NB has found wide applications in document classification processes whereas BernoulliNB works well with binary and Boolean features. Therefore, it was necessary to include ComplementNB which is known for handling imbalanced data, and assess its efficiency in such a context.

These methods of feature extraction included TF-IDF Vectorizer and Count Vectorizer. TF-IDF is a method that gives weight to words depending on their occurrence within a document and through their total count for the whole corpus. The use of Count Vectorizer is different as it only counts the number of words within a single document. These are both forms of converting textual data into numerals format to facilitate algorithms such as NB in processing the data efficiently.

Accuracy scores indicated significant discrepancies between TextBlob and VADER Sentiment using Naive Bayes algorithms. As illustrated in Figure 4, TextBlob's NB algorithm was substantially more accurate with an average of about 86% while VADER just managed to obtain an average of approximately 64%.

Word clouds in Figure 5 portray the sentiments associated with shut down of the TikTok shop. The word cloud shown in Figure 5(a) generated by using TextBlob depicts significant keywords including "TikTok shop," "Tanah Abang", and "UMKM" (Micro, Small, and Medium Enterprises) in the form of big fonts This signifies that TextBlob's sentiment analysis was focused on Indonesia. However, Figure 5(b) shows a word cloud constructed by using VADER Sentiment with words "TikTok shop", "UMKM", Tanah Abang" and "closed". Both word clouds provide specific settings (width=800, height=500, max_words=400, min_font_size=5, interpolation=bilinear).

The word clouds elucidate the emotional and economic consequences of closing down TikTokShop. The keywords for TextBlob and Vader Sentiment analysis are related to "TikTok Shop", "UMKM" and "TanahAbang". The appearance of these words in their respective cloud emphasizes how essential TikTok shop has been towards small businesses Such a presentation shows that the consequences of banning or closing TikTok shop are constantly seen as negative across all those involved. It means that there is an understanding among small entrepreneurs and vendors within the Tanah Abang market in Indonesia that the TikTok shop closure would be bad indeed.

(a)

(b)

Figure 5: Word cloud of TikTok shop's Closure. (a) Word cloud using TextBlob, (b) Word cloud using VADER

5. CONCLUSIONS

In conclusion, this study examined feelings about TikTok Shop's shutdown in Indonesia utilizing sentiment analysis of tweets. Using different Naive Bayes models including Multinomial, Bernoulli, and Complement Naive Bayes, coupled with the feature selection technique of CountVectorizer, uncovered a critical discovery. The investigation showed that the best results were achieved when using TextBlob as a sentiment classification library, together with CountVectorizer. Notably, the Multinomial Naive Bayes model reached a considerable precision of 86.60% exceeding the efficacy realized by VADER as the sentiment classification library. This again underlines the importance of selecting an appropriate library together with appropriate techniques for sentiment analysis showing that TextBlob and CountVectorizer combination in this context has given excellent sentiment analysis results to the closure of TikTok shop in Indonesia.

In potential future research, we aim to delve into sentiment analysis to elevate performance by leveraging Support Vector Machine (SVM) as an alternative Sentiment Classification algorithm. Our objective involves utilizing libraries like BERT to discern sentiment polarity, thereby enhancing our understanding and accuracy in sentiment analysis. The focus will be on combining SVM with BERT and exploring its potential for achieving improved sentiment classification outcomes.

ACKNOWLEDGMENTS

This research is supported by the Vice-Rector of Research, Innovation, and Entrepreneurship at Satya Wacana Christian University.

REFERENCES:

[1] S. Sharma and A. Jain, "Role of sentiment analysis in social media security and analytics," *WIREs Data Mining and Knowledge Discovery*, vol. 10, 2020, doi: 10.1002/widm.1366.

[2] R. A. Baroroh, Kussudyarsana and Soepatini, "The Effect Of Price, Product, And Flash Sale On Repurchase Decision With Customer Satisfaction As Moderation On Tiktokshop Features," *Jurnal Ekonomi dan Bisnis*, vol. 11, 2022.

[3] M. Reza Putra, "Literature Review: Competitive Strategy and Competitive Advantages on E-Commerce Shopee Indonesia," *Journal of Law, Politic and Humanities*, vol. 2, p. 165–174, 2022, doi: 10.38035/jlph.v2i4.106.

[4] J. Liang, H. Lu, Y. Zhao, J. Li, Z. Liu and Y. Feng, "A Study on Traffic Analysis of Cross-Border E-commerce Shops on the TikTok Platform," Presented at the, 2023.

[5] S. Muslim, M. Muktar and S. Diansah, "Implikasi Hukum Penutupan TikTok Shop terhadap Regulasi Hukum Bisnis di Indonesia," *Jurnal Hukum dan HAM Wara Sains*, vol. 2, 2023, doi: 10.58812/jhhws.v2i10.713.

[6] Bhwana Garda, "TikTok Shop Closure Gives Rise to Competition in E-Commerce Industry, Says Expert."

[7] K. P. Gunasekaran, "Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review," 2023, doi: 10.17148/IJARCCE.2019.8126.

[8] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *J Informetr*, vol. 3, 2009, doi: 10.1016/j.joi.2009.01.003.

[9] M. Subramanian, V. Easwaramoorthy Sathiskumar, G. Deepalakshmi, J. Cho and G. Manikandan, "A survey on hate speech detection and sentiment analysis using machine learning and deep learning models," *Alexandria Engineering Journal*, vol. 80, p. 110–121, 2023, doi: 10.1016/J.AEJ.2023.08.038.

[10] H. J. Christanto and Y. A. Singgalen, "Sentiment Analysis of Customer Feedback Reviews Towards Hotel's Products and Services in Labuan Bajo," *Journal of Information Systems and Informatics*, vol. 4, 2022, doi: 10.51519/journalisi.v4i4.294.

[11] A. Taparia and T. Bagla, "Sentiment Analysis: Predicting Product Reviews' Ratings using Online Customer Reviews," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3655308.

[12] H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research," 2023.

[13] H. J. Christanto, "Game Theory Analysis on Marketing Strategy Determination of KAI Access and Traveloka based on Usability of HCI (Human-Computer Interaction)", Journal of Information Systems and Informatics, Vol. 4, No. 3. September 2022. pp. 665-672. doi:.10.51519/journalisi.v4i3.300

[14] X. Liu, A. C. Burns and Y. Hou, "An Investigation of Brand-Related User-Generated Content on Twitter," *J Advert*, vol. 46, 2017, doi: 10.1080/00913367.2017.1297273.

[15] H. J. Christanto, S. A. Sutresno, A. Denny and C. Dewi, "USABILITY ANALYSIS OF HUMAN COMPUTER INTERACTION IN GOOGLE CLASSROOM AND MICROSOFT TEAMS", Journal of Theoretical and Applied Information Technology, Vol. 101, No. 16. August 2023. pp. 6425-6436.

[16] B. Wang, E. Wang, Z. Zhu, Y. Sun, Y. Tao and W. Wang, "An explainable sentiment prediction model based on the portraits of users sharing representative opinions in social sensors," *Int J Distrib Sens Netw*, vol. 17, 2021, doi: 10.1177/15501477211033765.

[17] M. Schmitt, "Automated machine learning: AI-driven decision making in business analytics," *Intelligent Systems with Applications*, vol. 18, 2023, doi: 10.1016/j.iswa.2023.200188.

[18] M. Nauman, N. Akhtar, A. Alhudhaif and A. Alothaim, "Guaranteeing Correctness of Machine Learning Based Decision Making at Higher Educational Institutions," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3088901.

[19] K. Sharma and R. Bhalla, "Decision Support Machine- A Hybrid Model for Sentiment Analysis of News Headlines of Stock Market," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, 2022, doi: 10.32985/ijeces.13.9.7.

[20] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," in: Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019. *Institute of Electrical and Electronics Engineers Inc*., 2019.

[21] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, 2023, doi: 10.37373/tekno.v10i2.419.

[22] C. Dewi, R.-C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam Journal of Computer Science*, p. 1–14, 2023, doi: 10.1142/S2196888823500100.

[23] P. N. Andono, E. H. Rachmawanto, N. S. Herman and K. Kondo, "Orchid types classification using supervised learning algorithm based on feature and color extraction," *Bulletin of Electrical Engineering and Informatics*, vol. 10, 2021, doi: 10.11591/eei.v10i5.3118.

[24] S. Wang, J. Ren and R. Bai, "A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes," *Expert Syst Appl*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120094.

[25] A. Aninditya, M. A. Hasibuan and E. Sutoyo, "Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy," in: 2019 *IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, p. 112–117, IEEE, 2019.

[26] H. J. Christanto, S. A. Sutresno, V.S. Simi, C. Dewi and G. Dai, "ANALYSIS OF GAME THEORY IN MARKETING STRATEGIES OF TIKTOK AND INSTAGRAM", Journal of Theoretical and Applied Information Technology, Vol. 101, No. 22. November 2023. pp. 7100-7109.

[27] C. I. Ratnapuri, M. Karmagatri, D. Kurnianingrum, I. D. Utama and A. Darisman, "USERS OPINION MINING OF TIKTOK SHOP SOCIAL MEDIA COMMERCE TO FIND BUSINESS OPPORTUNITIES FOR SMALL BUSINESSES," *J Theor Appl Inf Technol*, vol. 101, 2023.

[28] H. J. Christanto, S. A. Sutresno, Y. A. Singgalen and C. Dewi, "Analyzing Benefits of Online Train Ticket Reservation App Using Technology Acceptance Model", Ingénierie des Systèmes d'Information, Vol. 29, No. 1, 2024. pp. 107-115. doi: 10.18280/isi.290112

[29] J. Mantik, R. Dwi Ayu Lestari, B. Setya Rintyarna, M. Dasuki, G. Kerang, K. Sumbersari, J. Jember and J. Timur, "Application Of N-Gram On K-Nearest Neighbor Algorithm To Sentiment Analysis Of TikTok Shop Shopping Features," *Jurnal Mantik*, vol. 6, 2022.

[30] K. A. F. A. Samah, N. M. N. Azharludin, L. S. Riza, M. N. H. H. Jono and N. A. Moketar, "Classification and visualization: Twitter sentiment analysis of Malaysia's private hospitals," *IAES International Journal of Artificial Intelligence*, vol. 12, 2023, doi: 10.11591/ijai.v12.i4.pp1793-1802.

[31] A. Karami, M. Lundy, F. Webb and Y. K. Dwivedi, "Twitter and Research: A Systematic Literature Review through Text Mining," *IEEE Access*, vol. 8, p. 67698–67717, 2020, doi: 10.1109/ACCESS.2020.2983656.

[32] Z. Nurdalia, D. Permana and A. Salma, "Comparison of Naïve Bayes and K-Nearest Neighbor for DKI Jakarta Air Pollution

Standard Index Classification," *UNP Journal of Statistics and Data Science*, vol. 1, 2023, doi: 10.24036/ujsds/vol1-iss2/29.

[33] S. Supangat, M. Z. Bin Saringat and M. Y. F. Rochman, "Predicting Handling Covid-19 Opinion using Naive Bayes and TF-IDF for Polarity Detection," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, 2023, doi: 10.30812/matrik.v22i2.2227.

[34] A. Asnimar, A. Achmad, Y. Yuyun, A. Iskandar and Mansyur, "Classification Of Prospective Scholarship Recipients Kartu Indonesia Pintar (KIP) With Decision Tree Algorithm And Naïve Bayes," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 12, pp. 118–129, 2022, doi: 10.35585/inspir.v12i2.6

[35] T. D. Dikiyanti, A. M. Rukmi and M. I. Irawan, "Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm," in: *Journal of Physics: Conference Series*, 2021.