

SENTIMENT ANALYSIS OF INDONESIAN PEOPLE'S RESPONSE AGAINST THE PAYLATER PAYMENT METHOD USING THE NAIVE BAYES ALGORITHM

Ani Nuraeni, Adhitia Erfina, Dede Sukmawan

Universitas Nusa Putra, Indonesia

*Corresponding Author:

Email: ani.nuraeni_si19@nusaputra.ac.id

Abstract.

Paylater is a payment method that uses an installment system of 1 to 12 months only with KTP and selfie. Paylater raises pros and cons in Indonesian society, causing several problems such as: users have the nature of impulse buying, low interest quite a large range of 2% - 4%, until the Paylater user data leak case. This research was conducted to determine positive, neutral, and negative sentiments using the Naive Bayes algorithm Classifier in the Instagram comment column found on the official marketplace accounts Shopee, GopayIndonesia, Akulakupaylater, Kredivo, and Homecredit using scraping techniques python programming language. The results of this study to determine the ranking method Paylater payments in the marketplace are based on the highest percentage of positive scores, the ranking results are as follows: Akulaku (43.37%), Kredivo (39.16%), Shopeepaylater (34.85%), Gopaylater (28.57%), and Homecredit (25.31%) with the aim of recommending the Paylater payment method in the marketplace based on the most positive scores to be considered for users.

Keywords: Paylater, Impulse Buying, Naive Bayes, Instagram, Python

INTRODUCTION

E-commerce is a marketing and buying and selling business model that has an electronic transaction system using internet or television media. One of the business models of e-commerce is the marketplace. Marketplace is a place for business activities of buying and selling goods and services through online applications or websites (Putri, Luthfia E. 2023). To help the transaction process in the marketplace, consumers can use various payment methods, one of which is the Paylater payment method.

According to the Kata Data Insight (KIC) survey in 2022, users of the Paylater payment method in various Indonesian marketplaces reached 17% (Kata Data Insight Center (KIC) and Kredivo.2022). Paylater is a payment method that makes it easy for consumers to transact on the marketplace with an installment system ranging from 1 to 12 months with only a KTP and a selfie. With all its convenience, Paylater raises pros and cons in today's Indonesian society. People who choose the pros and use the Paylater payment method think that this system can alleviate and assist in meeting various needs when they do not have income and people who choose cons tend not to use this payment method thinking that this system will be detrimental in the future when they are unable to pay their bills .

Apart from being easy to use, Paylater also has various problems. First, Paylater makes Generation Z (generation of digital natives) who were born and raised in the digital era behave consumptively because this generation mostly behaves spontaneously or impulse buying. Second, the interest offered is quite high, ranging from 2% - 4% depending on the application provider. Third, the security of consumer data in the hands of the Paylater provider company with the risk of data being spread (The Conversation.2022).

Based on the background of this problem, research was carried out by knowing the opinion of the Indonesian people regarding the Paylater payment method for the purpose of ranking the best Paylater application from the public opinion based on the most positive

percentage of the classification results. The author uses the Instagram comments column for data collection because Indonesia has 89.15 million Instagram users in 2023 (Sadya, Sarnita. 2023).

LITERATURE REVIEW

Related Research

Sentiment analysis about public opinion towards Paylater. The dataset taken is from Twitter with the account @worksfess. This study aims to determine the opinions of workers using Paylater who have impulse buying behavior by using two algorithms. The first uses the Naive Bayes Classifier algorithm which produces 405 data with 70.62% negative sentiment (268 data), 22.72% positive sentiment (92 data) and 6.67% neutral sentiment (27 data) and the second uses the TextBlob algorithm which produces 55.8% negative sentiment (226 data), 33.09% positive sentiment (134 data) and 11.11% neutral sentiment (45 data). From the data generated, it is concluded that the community is not good at using Paylater by testing the confusion matrix model showing that the Naive Bayes Classifier algorithm is more accurate by 91% compared to TextBlob which is only 61% (Safira, Alfandi, and Firman N. Hasan. 2023). Sentiment analysis on the response of the Indonesian people to Non-Fungible Tokens (NFT). The dataset taken is from the Youtube comments column. This study aims to understand the public's response to NFT, how enthusiastic the community is towards NFT and to determine the accuracy of the Naive Bayes Classifier algorithm in sentiment analysis regarding NFT. The results of this study obtained an accuracy of 64% positive, 63% neutral and 0% negative. 99% positive recall, 0.7% neutral and 0% negative (Erfina, Adhitia, and Dinda T. Mahardika. 2022).

Sentiment analysis regarding comments by netizens in the Instagram comments column, a case study on the Instagram case of West Java Governor Ridwan Kamil. The dataset was taken from the Instagram account column of West Java Governor Ridwan Kamil using the Naive Bayes Classifier and TF-IDF algorithms and obtained an accuracy rate of 92% (Maulidina, Mega K. 2020). Research conducted by Habibi Al Rasyid Harpizon, Rahmad Kurniawan, Iwan Iskandar, Roni Salambue, Elvia Budianita, and Fadhilah Syafria related to Ustadz Abdul Somad's Da'wah related to Ustadz Abdul Somad's video lecture discussing hadith and fiqh. This study uses the Naive Bayes Classifier algorithm from the Youtube comments dataset using 1000 comments from 10 videos. The results of this study were 67% positive, 6% negative and 27% neutral. The conclusion of this study is that it obtained an accuracy of 87%, 91% precision and 97% recall (Harpizon, Habibi A., Rahmad Kurniawan, Iwan Iskandar, Roni Salambue, Elvia Budianita, and Fadhilah Syafira. 2022). Analysis of public opinion sentiment in Playstore towards delivery service applications that started from the government's policy for WFH (work from home) efforts to reduce the effects of Covid-19 which triggered an increase in online shopping. This study uses datasets from Playstore comments, delivery service applications using the Naive Bayes algorithm which produces 100% accuracy for JNT and other delivery services. A common mistake that occurs is the delay in sending the package not according to the specified time. In conclusion, a high accuracy value of 90% of the test results (Erfina, Adhitia, and Muhamad F. Al-Shufi. 2022).

Theoretical basis

Paylater is a payment method with an installment payment system ranging from 1 to 12 months so that it can facilitate and ease the process of consumer transactions in the marketplace. The Paylater registration process can be done by uploading documents such as KTP and taking selfies in the marketplace application that you want to use (Prasetya, Audrey Nathania E. 2023). Sentiment analysis is a process for determining opinions, emotions, and attitudes through text. Sentiment analysis has three levels, in this study using the message level because the focus of this research will group responses or opinions into several categories such as positive sentiment, neutral sentiment, and negative sentiment (Nursyi'ah, Siti Y., Adhitia Erfina, and Cecep Warman. 2021). Naive Bayes Classifier is a classifier algorithm for predicting probabilities that has a high degree of accuracy for classification that is easy to implement (Maulidina, Mega P., and Dr. Enny

ISSi., M.Kom. 2020). Python is a high-level programming language that has commands or program code almost the same as human languages.

Python uses interpreted processing of program code which is processed line by line without having to compile the program (Sembiring, Falentino, and Adhithia Erfina. 2020). Scraping is a technique for automatically extracting information from a web page to obtain data in the form of a new HTML web page. The scraping process in this study uses the help of Python as a programming language and Visual Studio Code as a text editor (Ayani, Dhita D., Helen S. Pratiwi, and Hafiz Muhandi. 2019). Preprocessing is a technique of changing data in the form of raw data into structured data, usually raw data is often incomplete, inconsistent, and may contain errors. In this stage, the raw data will go through a process of cleaning, case folding, stemming, tokenizing, normalization, and stopwords removal, the results of which will affect the level of accuracy (Munthe, Mampe P., and Anton Siswo R. Ansori. 2021). TextBlob is a library in Python that is used to process textual data. In this study, TextBlob was used in the sentiment labeling process to determine positive sentiment, neutral sentiment, and negative sentiment, using the help of the Google Translate library to facilitate sentiment labeling in Indonesian (Parlika, Rizky, Sunu I. Pradika, Amir M. Hakim, and Kholilul RNM 2020). Sastrawi is part of the Python Library which is used to change or delete affixes into Indonesian base words automatically using the Python programming language. In this study Literature was used for the stemming process (Nugroho, Kuncahyo S. 2019). Word weighting aims to weight features based on their frequency of occurrence. Features that have been given a weight can be used for the classification process. The weighting stage in this study used the Matplotlib Library (Rofiqoh, Umi, Rizal S. Perdana, and MA Fauzi. 2017). Confusion Matrix is a method used to calculate accuracy, recall, precision, and error which aims to analyze the performance of the algorithms used in detecting damage. Recall, namely the percentage of all documents significantly with the query. Precision is to identify significant rankings. Calculation of accuracy, namely the comparison of the correct cases with the entire number of existing cases. The error rate is the comparison of cases that have been determined incorrectly to the number of existing cases (Harpizon, Habibi A., Rahmad Kurniawan, Iwan Iskandar, Roni Salambue, Elvia Budianita, and Fadhilah Syafria. 2022).

METHODS

Research methods

The research method used is sentiment analysis which is a process for determining public opinion through text, in this study the authors use the message level level because it focuses on grouping responses or opinions into categories of positive sentiment, neutral sentiment, and negative sentiment using the Naive Bayes Classifier algorithm to classify the data that has been obtained (Cindo, Mona, Dian P. Rini, and Ermatika. 2019). The steps for conducting this research are shown in the following figure:

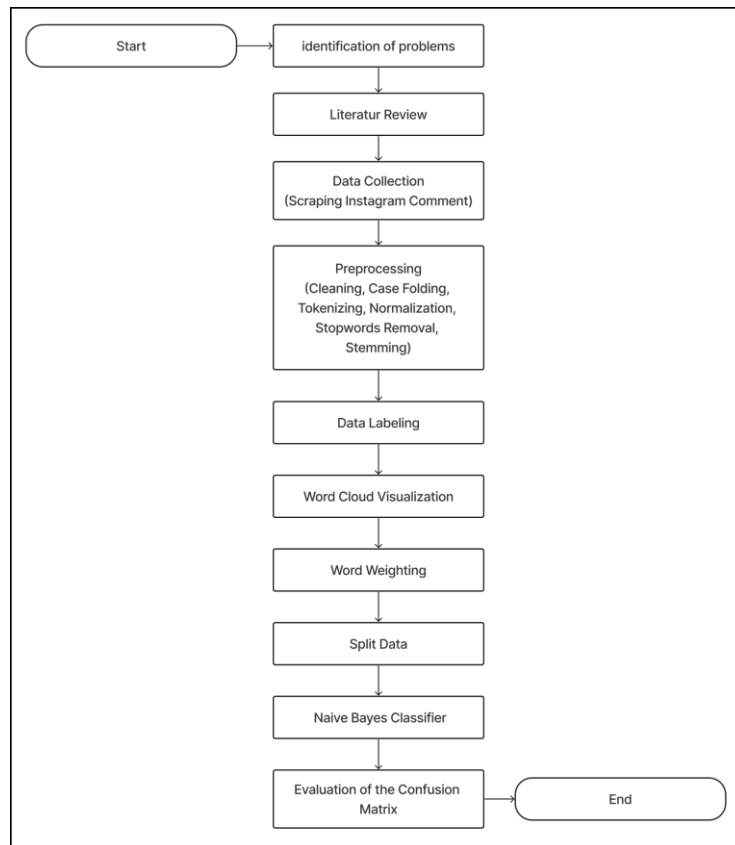


Figure 1. Stages of research

Data collection

The method of collecting data in this study is by scraping techniques using the help of the Python programming language. The dataset taken is comment column data from Paylater content on the Paylater provider's official account with Instagram usernames as follows: shopee_id, gopayindonesia, kredivo, akulakupaylater_id, and homecreditid.

RESULT AND DISCUSSION

Scraping Process

The Scraping process uses Selenium for the scraping process, web driver, BeautifulSoup, and prettytable. The web driver used is Firefox because it is a relatively lightweight browser so it can maximize the scraping process. In this study, the scraping process was carried out by entering the username and password for the Instagram account, the web link used from Instagram, and content links that will be taken for comments assisted by the BeautifulSoup library to convert the Instagram web into normal HTML. Its function is to retrieve classes in HTML so that authors can retrieve only username and comment data by obtaining data as much as Shopee paylater 959 data, Gopaylater 620 data, Akulaku 628 data, Kredivo 513 data, and Homecredit 883 data taken on February 2023 which were exported to CSV format.

Preprocessing

Preprocessing is the stage for cleaning data whose function is not to interfere with the results of accuracy as follows: Cleaning to remove emoticons, hashtags, URLs, punctuation, numbers, short words, and double spaces. Case folding to change text into standard letters or lowercase letters which is called lower case. Tokenizing to divide comment text data from sentences into word for word in order to facilitate the normalization process. Normalization in this study is to improve the words of the text comments which are shortened to basic words that are appropriate to the writing. Stopwords removal to clean words that lack meaning for example.

Stemming to remove affixes such as prefixes, suffixes or inserts and change them from tokenizing to regular sentences to facilitate the data labeling process. The result of preprocessing is as follows:

Table 1. Preprocessing Results

Preprocessing Process	Before	After
Cleaning	Udh ga kerjasama dengan samsung ya...padahal mau po s23u pake kredivo cuma ga ada promonya...jd malas	Udh ga kerjasama dengan samsung yapadahal mau po su pake kredivo cuma ga ada promonya jd malas
Case folding	HCI is cool but it's a shame the LIGHT OF INTEREST funds Until KALI LIPET, if you want a lot of interested interest, lower it, minklo for goods, the standard price is okay	hci is cool, but it's a shame that the interest rate loan is so high that it's multiple times if you want lots of interest, lower interest rates for goods, it's okay, the price is standard
Tokenizing	hci is cool, but it's a shame that the interest rate loan is so high that it's multiple times if you want lots of interest, lower interest rates for goods, it's okay, the price is standard	['hci', 'cool', 'but', 'honey', 'loan', 'fund', 'interest', 'height', 'until', 'times', 'lipets', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'minklo', 'tuk', 'goods', 'okay', 'standard', 'price']
Normalization	['hci', 'cool', 'but', 'honey', 'loan', 'fund', 'interest', 'height', 'until', 'times', 'lipets', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'minklo', 'tuk', 'goods', 'okay', 'standard', 'price']	['home credit indonesia', 'cool', 'but', 'honey', 'loan', 'fund', 'interest', 'height', 'until', 'if', 'fold', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'admin if', 'for', 'goods', 'okay', 'standard', 'price']
Stopwords Removal	['home credit indonesia', 'cool', 'but', 'honey', 'loan', 'fund', 'interest', 'height', 'until', 'if', 'fold', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'admin if', 'for', 'goods', 'okay', 'standard', 'price']	['home credit indonesia', 'cool', 'hope', 'loan', 'funds', 'interest', 'height', 'if', 'fold', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'admin if', 'goods', 'okay', 'standard', 'price']
Stemming	['home credit indonesia', 'cool', 'hope', 'loan', 'funds', 'interest', 'height', 'if', 'fold', 'if', 'want', 'many', 'enthusiasts', 'interest', 'lower', 'admin if', 'goods', 'okay', 'standard', 'price']	Indonesian home credit is cool, it's a shame to borrow funds with high interest if you want a lot of interest in interest lower the admin if the goods are okay standard price

Data Labeling

Data labeling to label sentiment into positive sentiment, neutral sentiment, and negative sentiment. Data labeling is done automatically using a Python script with the help of the Textblob library. Because the data is in Indonesian, it must be translated into English using the Googletrans library. To find out the percentage of data labeling as follows:

$$\frac{\text{the amount of data of a class}}{\text{total comment data}} \times 100\%$$

Equation (1)

The results of sentiment labeling are as follows:

Table 2. Sentiment Classification Results

Marketplaces	Positive	Neutral	Negative
Shopeepaylater (766 data)	34.85% (267 data)	43.99% (337 data)	21.14% (162 data)
Gopaylater (595 data)	28.57% (170 data)	59.49% (354 data)	11.93% (71 data)
Akulaku (528 data)	43.37% (229 data)	41.07% (217 data)	15.53% (82 data)
Kredivo (429 data)	39.16% (168 data)	37.52% (161 data)	23.31% (100 data)
Homecredit (881 data)	25.31% (223 data)	24.63% (217 data)	50.05% (441 data)

Naive Bayes Classifier

The classification used in this study to predict the classification was carried out on test data by studying information from the training data which in this study was taken from the results of the data labeling process. This classification process uses the Python programming language with the help of the sklearn library. The classification results show the following accuracy figures:

Table 3. Accuracy Results

Marketplaces	Accuracy
Shopeepaylater	66.88%
Gopaylater	74.79%
Akulaku	75.47%
Kredivo	59.30%
Homecredit	82.49%

Evaluation of the Confusion Matrix

The confusion matrix serves to evaluate how accurate the Naive Bayes classification model is known through the precision value which indicates the level of ability of the classification algorithm in finding accuracy, recall shows the success rate of the classification algorithm in retrieving information, and f-1 score shows the comparison of the average precision and weighted recall. The results of the confusion matrix evaluation process are as follows:

	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.49	0.61	0.54	31	negatif	0.83	0.31	0.45	16
netral	0.72	0.75	0.73	63	netral	0.79	0.87	0.83	69
positif	0.74	0.62	0.67	60	positif	0.65	0.71	0.68	34
accuracy			0.67	154	accuracy			0.75	119
macro avg	0.65	0.66	0.65	154	macro avg	0.76	0.63	0.65	119
weighted avg	0.68	0.67	0.67	154	weighted avg	0.76	0.75	0.73	119

(i)

	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.56	0.71	0.63	14	negatif	0.50	0.57	0.53	23
netral	0.83	0.71	0.76	41	netral	0.65	0.52	0.58	29
positif	0.77	0.80	0.79	51	positif	0.62	0.68	0.65	34
accuracy			0.75	106	accuracy			0.59	86
macro avg	0.72	0.74	0.73	106	macro avg	0.59	0.59	0.59	86
weighted avg	0.77	0.75	0.76	106	weighted avg	0.60	0.59	0.59	86

(ii)

	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.86	0.92	0.89	84	negatif	0.86	0.92	0.89	84
netral	0.73	0.75	0.74	44	netral	0.73	0.75	0.74	44
positif	0.86	0.73	0.79	49	positif	0.86	0.73	0.79	49
accuracy			0.82	177	accuracy			0.82	177
macro avg	0.82	0.80	0.81	177	macro avg	0.82	0.80	0.81	177
weighted avg	0.83	0.82	0.82	177	weighted avg	0.83	0.82	0.82	177

(iii)

	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.86	0.92	0.89	84	negatif	0.86	0.92	0.89	84
netral	0.73	0.75	0.74	44	netral	0.73	0.75	0.74	44
positif	0.86	0.73	0.79	49	positif	0.86	0.73	0.79	49
accuracy			0.82	177	accuracy			0.82	177
macro avg	0.82	0.80	0.81	177	macro avg	0.82	0.80	0.81	177
weighted avg	0.83	0.82	0.82	177	weighted avg	0.83	0.82	0.82	177

(iv)

	precision	recall	f1-score	support		precision	recall	f1-score	support
negatif	0.86	0.92	0.89	84	negatif	0.86	0.92	0.89	84
netral	0.73	0.75	0.74	44	netral	0.73	0.75	0.74	44
positif	0.86	0.73	0.79	49	positif	0.86	0.73	0.79	49
accuracy			0.82	177	accuracy			0.82	177
macro avg	0.82	0.80	0.81	177	macro avg	0.82	0.80	0.81	177
weighted avg	0.83	0.82	0.82	177	weighted avg	0.83	0.82	0.82	177

(v)

Figure 2. Shopeeaylater (i), Gopaylater (ii), Akulaku (iii), Kredivo (iv) and Homecredit (v).

CONCLUSION

This research discusses the response of the Indonesian people to the Paylater payment method in the Shopeeaylater, Gopaylater, Akulaku, Kredivo, and Homecredit marketplaces. This research originated from the existence of pro and con opinions in the community regarding Paylater. The dataset in this study uses Instagram comment column data from content about Paylater on the official account of the associated marketplace using the Naive Bayes classification algorithm. The results of this study are to determine the ranking of Paylater payment methods in the marketplace based on the highest percentage of positive scores, namely Akulaku (43.37%), Kredivo (39.16%), Shopeeaylater (34.85%), Gopaylater (28.57%), and Homecredit (25.31%) with the aim to recommend the Paylater payment method on the marketplace based on the most positive scores so that it becomes a consideration for users.

Comment data taken by Shopeeaylater 959 data, Gopaylater 620 data, Akulaku 628 data, Kredivo 513 data, and Homecredit 883 or 3,603 data comments from a total of 5 Paylater payment methods therefore for further research can add more data so as to produce accuracy values higher In the preprocessing process, normalization still uses manual by making a dictionary so it is possible for words that are not appropriate to still be in the final result, so suggestions for further research are to try to use automatic normalization. For the next research process, it might be possible to use a different algorithm besides Naive Bayes to be a comparison in the research results.

REFERENCES

- Ayani, Dhita D., Helen S. Pratiwi, and Hafiz Muhandi. 2019."Implementation of Web Scraping for Data Retrieval on Marketplace Sites e7 (October): 258.
- Cindo, Mona, Dian P. Rini, and Ermatika. 2019."Literature Review: Classification Methods on Sentiment Analysis", 66-67. 978-602-52720-1-1.
- Erfina, Adhitia, and Muhamad F. Al-Shufi. 2022."SENTIMENT ANALYSIS OF COURIER SERVICE APPLICATIONS IN THE PLAY STORE USING THE NAIVE BAYES ALGORITHM"5(2).

- Harpizon, Habibi A., Rahmad Kurniawan, Iwan Iskandar, Roni Salambue, Elvia Budianita, and Fadhilah Syafira. 2022."Analysis of Comment Sentiments on YouTube About Ustadz Abdul Somad's Lectures Using the Naïve Bayes Algorithm"5, no. 1 (February).
- Maulidina, Mega K. 2020. "SENTIMENT ANALYSIS OF CITIZENS' COMMENTS ON INSTAGRAM POSTS USING THE NAÏVE BAYES CLASSIFIER METHOD AND TF-IDF (Case Study: West Java Governor Ridwan Kamil's Instagram)", (December).
- Maulidina, Mega P., and Dr. Enny ISSi., M.Kom. 2020."SENTIMENT ANALYSIS OF CITIZENS' COMMENTS ON INSTAGRAM POSTS USING THE NAÏVE BAYES CLASSIFIER METHOD AND TFIDF (Case Study: West Java Governor Ridwan Kamil's Instagram)", 5.
- Sembiring, Falentino, and Adhitia Erfina. 2020."Snake Language For Python Programming. Kapolo Koto, West Sumatra: Independent Scholars".
- Munthe, Mampe P., and Anton Siswo R. Ansori. 2021."Analysis of Comment Sentiment on the Indonesian Language Food Vlogger Youtube Channel Using the Naïve Bayes Algorithm"8 (December): 2. 2355-9365.
- Nugroho, Kuncahyo S. 2019. "Basic Text Preprocessing with Python | by Kuncahyo Setyo Nugroho. (<https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe>).
- Nursyi'ah, Siti Y., Adhitia Erfina, and Cecep Warman. 2021."ANALYSIS OF ONLINE LEARNING SENTIMENTS DURING THE COVID-19 PANDEMIC ON TWITTER USING THE NAÏVE BAYES ALGORITHM", (August).
- Pahlevi, Reza. 2022. "KIC: The Most Frequently Used E-Wallet for E-Commerce Transactions." Databox. (<https://databoks.katadata.co.id/datapublish/2022/06/03/kic-e-wallet-paling-sering-digunakan-untuk-transactione-commerce>).
- Parlika, Rizky, Sunu I. Pradika, Amir M. Hakim, and Kholilul RNM 2020. "TextBlob's Python-based Twitter Sentiment Analysis of Bitcoin and Cryptocurrencies"2 (December): 34.
- Prasetya, Audrey Nathania E. 2023. "ANALYSIS OF THE EXISTENCE OF PAY LATER IN THE MARKETPLACE ON PUBLIC PURCHASING POWER"3 (February). 10.46306/rev.v3i2.125.
- Princess, Luthfia E. 2023. "The Effect of E-Commerce on Business Development in Indonesia"Vol. 2, no. 1 (January): 45.
- Rofiqoh, Umi, Rizal S. Perdana, and MA Fauzi. 2017."Sentiment Analysis of User Satisfaction Levels of Indonesian Cellular Telecommunications Service Providers on Twitter Using the Support Vector Machine and Lexicon Based Features Methods1 (December): 1727.
- Sadya, Sarnita. 2023. "The Fourth Largest RI Instagram Users in the World in Early 2023." Indonesian data. (<https://dataindonesia.id/digital/detail/user-instagram-ri-larger-fourth-di-dunia-padaawal-2023>).
- Safira, Alfandi, and Firman N. Hasan. 2023."ANALYSIS OF PUBLIC SENTIMENT ON PAYLATER USING THE NAIVE BAYES CLASSIFIER METHOD"5, no. 1 (January): 1.
- Erfina, Adhitia, and Dinda T. Mahardika. 2022."Indonesia Analysis Sentiment On Fungible Token (NFT)"9, no. 2 (December). 2355-0082.
- The Conversations. 2022. "PayLater helps facilitate transactions but triggers consumptive behavior and debt arrears, how do consumers react to this?" The Conversations. (<https://theconversation.com/paylater-bantu-facilitation-transaksi-tapi-trigger-behavior-consumptive-dantung-guns-debt-how-consumers-respond-to-it-184820>).