Sentiment Analysis Of E-Wallets on Twitter social media With Naïve Bayes and Lexicon-Based Methods

Ananda Dimas Sanjaya
Department of Informatics
Science and Informatics Faculty
Universitas Jenderal Achmad Yani
West Java, Cimahi, Indonesia
anadadimass17@if.unjani.ac.id

Tacbir Hendro Pudjiantoro, Ade Kania Ningsih, Faiza Renaldi
Department of Informatics
Science and Informatics Faculty
Universitas Jenderal Achmad Yani
West Java, Cimahi, Indonesia
thp@if.unjani.ac.id, ade.kanianingsih@lecture.unjani.ac.id,
faiza.renaldi@unjani.ac.id

Abstract

Sentiment analysis is one of the areas of application of text analysis and NLP (Natural Language Processing). It builds a system to identify and analyze personal information from textual sources, which are generally divided into positive and negative categories. Many case studies are chosen to measure the accuracy of the computational model of sentiment analysis. In this study, we use social media's perception of electronic wallets usage. The Naïve Bayes method was chosen. It can help classification because it is assumed to be an independent variable, and Lexicon is used to calculate the weight of each word. Although several studies have been conducted on sentiment analysis, none have used non-standard word correction. The use of non-standard word correction deals with non-standard words such as slang and word abbreviations. This research begins by digging up the required data, namely the keyword electronic wallet. The standard word dictionary normalizes non-standard words, preprocessing data with four stages: case folding, tokenizing, stop word, and stemming. The lexicon dictionary produces positive and negative labels, and nave Bayes is used to classifying. The data used in this study were 3878 tweets, with a distribution of 70% training data and 30% test data. Sentiment analysis obtained in this study shows that Twitter users in Indonesia are more likely to give negative comments to electronic wallets. The results of this study indicate that by using both methods and adding standard word corrections and testing using RapidMiner, the accuracy rate for classifying positive and negative sentiments reaches 88.56%. Further research can add Levenshtein Distance normalization to the classification results to better influence the accuracy value.

Keywords

Text Mining, Naïve Bayes Classification, Lexicon Based, Electronic Wallet, Non-Standard Word Correction

1. Introduction

The development of information technology is increasingly advanced in today's digital era. Not only as a medium of information and communication, but many people also use social media to express their opinions, feelings, experiences, and other things that concern them(Troussas et al. 2013). Not a few netizens also use Twitter as a place to express opinions, and these opinions contain positive and negative sentiments. Sentiment analysis is a method of collecting comments and comments from other people about something, such as an issue on a web-based social network(Alsaeedi and Khan 2019).

Social media data is one way to perform sentiment analysis. Social media can generate thousands of tweets or opinions every day(Medhat, Hassan, and Korashy 2014). People in Indonesia use social media to channel their opinions and opinions about various things. The case study taken by researchers is the trend of e-wallet applications that have been trending on Twitter regarding the problem of conceding balances in several electronic wallet applications to be used as sentiment analysis material because this electronic wallet application is being widely used by the public, especially in online transactions. Indonesia has provided several electronic wallets, such as Dana, OVO, and Go Pay. Public assessments of the ease of payment methods through electronic wallets can be obtained through social media, one of which is through Twitter, attracting many opinions to be discussed with many tweets and comments on social media, especially on Twitter(Salim and Mayary 2020).

The Twitter social media network is characterized by up-to-date tweet data with a large number and varied topics. Thus, Twitter has the potential to perform sentiment analysis(Najib et al. 2019). Every day social media can generate thousands of opinion posts or tweets, and everyone expresses opinions and opinions through social media freely. These opinions contain positive, negative, and neutral sentiments(Nugroho 2018). And sentiment analysis belongs to text mining and has additional challenges not encountered in data mining, such as complex and incomplete text structures, unclear and non-standard meanings, and different languages plus inaccurate translations(Andika, Azizah, and Respatiwulan 2019).

This research aims to see the potential of the community contained in Twitter social media for the case of electronic wallet applications, and this research wants to test using a combination of two methods, namely naïve bayes and lexicon-based, to increase the accuracy rate, with the addition of standard word normalization. Using the naïve bayes method alone can increase the accuracy rate for classification, but the pre-processing process has the disadvantage of miss-classifying the data(Parveen and Pandey 2017). Therefore, to improve and reduce data misclassification caused by the naïve bayes method, it will be combined using lexicon based. Lexicon-based can be used after the pre-processing stage; lexicon-based will have a dictionary containing a collection of words with positive or negative sentiment meaning(Zhang et al. 2011). However, the lexicon method also found several factors that affect the performance of sentiment analysis with a lexicon-based approach, including the use of non-standard language, the lack of words that are not in the dictionary, the appearance of words that have more positive values than words that are negative in a tweet that should be negative or vice versa(Halimi and Rudyanto Arief 2021). Therefore, to handle the problem of non-standard words only in modern language or slang, word abbreviations, and misspellings, a normalization process is performed by checking the tokens in the sentence using a standard word dictionary(Prananda Antinasari, Rizal Setya Perdana 2017).

Naïve Bayes as a classification method in text mining is used in sentiment analysis. This method has good potential in classification in terms of precision and data computation(Joshi and Vala 2014). Thus, naïve bayes as a method for sentiment analysis and lexicon weighting becomes renewable so that naïve bayes and lexicon are widely used to be applied in Text Mining(Xu, Pan, and Xia 2020). Sentiment analysis has been widely used for various studies, such as classifying Indonesian people towards online learning on Twitter social media using Lexicon and K-Nearest Neighbours methods. The results obtained from the comparison of sentiment data are K-Nearest Neighbours with an accuracy rate of 80.66% of K-Nearest Neighbours sentiment data, and Lexicon sentiment data produces 80.92%. However, the misclassification of data on pre-processing is not optimal because stop words and normalized data sets lack so many data sets. So a dataset is needed that can convert non-standard words into standard words based on KBBI(Halimi and Rudyanto Arief 2021). While sentiment analysis with a case study of online product reviews using the Naïve Bayes method produces the lowest accuracy value in testing 5 classes using 80% training data and 20% test data resulting in an accuracy value of 52.66%, while in testing 3 classes using a dataset of 90% training data and 10% test data has the highest accuracy of 77.78%. In other words, the amount of training data in the sentiment analysis system influences system predictions(Gunawan, Pratiwi, and Pratama 2018). The sentiment analysis research on the level of satisfaction of users of mobile telecommunications service providers on Twitter social media with the Support Vector Machine method and Lexicon Based Features resulted in an accuracy value of 79%, but there was a miss-classification of data. Namely, sentences or words with positive sentiments appeared in the test data that should have negative sentiments and vice versa(Perdana and Fauzi 2017).

This research aims to implement the Naïve Bayes method, which was chosen because it can help for classification because it is assumed to be an independent variable, and the lexicon to calculate the weight of each word. Only the variable's variance within a class is needed to determine the classification. In addition, this research wants to use nonstandard word correction to handle the problem of nonstandard words only in modern language or slang and abbreviated words. And reduce invalid data or miss-classified data. This research uses data sources from social media Twitter and uses a case study of electronic wallet applications in Indonesia with several service categories, promos & admin fees, and transactions & top-ups.

2. Methods

The system that will be built is an application with the concept of analysing sentiment towards electronic wallets on Twitter social media using Naïve Bayes and Lexicon-based methods and adding corrections from non-standard words to standard words according to the standard word dictionary. The purpose of this system is to provide information about public sentiment towards electronic wallets, which will be grouped into positive or negative sentiments. The description of the system to be built is shown in Figure 1.

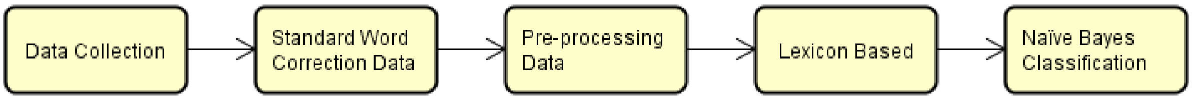


Figure 1. Research Methods

The first stage of this research is to data collection. After the required data is processed and has become data in the form of opinions, the data is changed from non-standard to standard words according to the Big Indonesian Language Dictionary (KBBI). After changing non-standard words to standard, the data is processed using pre-processing, namely case-folding, tokenizing, stop word, and stemming, which aims to clean data from raw or unused words. Then the dataset is labelled using a lexicon-based dictionary. After the data gets positive sentiment and negative sentiment, it enters the classification stage using naïve bayes.

2.1 Data Collection

In collecting data for this research using the Twitter API, the data collection steps are as follows:

- Create a developer Twitter account
- Request access to get an API (Application Program Interface) token to the Twitter developer app.
- Crawling tweet data is done using the RStudio application.

The data source used in this research is a collection of Indonesian tweets taken from each customer's Twitter account. In the crawling process, it will automatically retrieve tweets with keywords that use hashtags. The data used and cleaned as many as 3,639 comments.

2.2 Standard Word Correction

Non-standard word correction or language normalization is a process used to convert non-standard words into standard words, according to the Big Indonesian Dictionary (KBBI)(Buntoro et al. 2014). In the correction of non-standard words, there are various kinds of words that are considered non-standard words, such as writing in modern language or slang (ex: donk = dong), writing with guidelines (ex: pluang = opportunity), writing with short words (ex: yg = yang).

In this research, non-standard word repair is used to handle the problem of non-standard words in modern or slang language and word abbreviations only. The repair of non-standard words is done by the normalization process, which is carried out by checking the tokens in the sentence using the dictionary_katabaku. If there are non-standard words, they will be converted into standard words according to the dictionary so that the result is a sentence using standardized words. An example of a dictionary_katabaku table can be seen in Table 1.

Table 1. Example of dictionary_katabaku

Non-standard Words	Standard Word
cool	fun
bgt	really
cape	reach
donk	dong
cb	cashback
thanks	thank you
lbh	more

In this study, the standard word dictionary used was homemade with adjustments from the Big Indonesian Dictionary (KBBI). All text with non-standard words will be converted to standard words. An example of the word improvement implementation process can be seen in Table 2.

Table 2. Raw Word Process

Tweet	Standard Word
29 years old, happy to use DANA because of the Vouchers	29 years old, happy to use DANA because there are many Vouchers

2.3 Pre-Processing
Pre-processing cleans insufficient and unclear data and prepares the text for classification(Haddi, Liu, and Shi 2013). Reducing noise in the text can help improve the classifier's performance and speed up the classification process, thus helping in sentiment analysis (Kadhim 2018). The following steps must be taken in the pre-processing process:

2.3.1 Case Folding
Case Folding is converting all capital letters to lowercase. An example of the case folding implementation process is by checking the size of each character from the beginning to the end of the character. If a character is found that uses capital letters, it will be converted to lowercase. At this pre-processing stage, it can be seen in Table 3.

Table 3. Case Folding Process

Tweet	Case Folding
29 years old, happy to use DANA because there are many Vouchers	29-year-old, happy to use dana because there are many vouchers

2.3.2 Tokenizing
Tokenizing is cutting words based on each word that composes them into one part. An example of the tokenizing implementation process is by cutting all words in each sentence based on word separators such as periods (.), commas (,), and spaces, then parts that have only one non-alphabetic character and numbers will be discarded. At this pre-processing stage it can be seen in Table 4.

Table 4. Tokenizing Process

year	more	happy	happy	use
dana	because	many	voucher	

2.3.3 Stop word
Stop word is the process of eliminating words that do not match the document's topic because these words do not affect the accuracy of sentiment classification. An example of the stop word implementation process is to check whether the word is the same as the stop word list or not, and if the word is the same as the one in the stop word list, then the word will be deleted. At this pre-processing stage, it can be seen in Table 5.

Table 5. Stop word Process

happy	happy	use	dana	many
voucher				

2.3.4 Stemming
Stemming performs the process of returning various word formations to basic word formations by removing affixes. An example of the stemming implementation process is that every word in a tweet will be checked from the beginning to the end of the word. If there are words that contain affixes, then the affixes on the word will be removed at this pre-processing stage which can be seen in Table 6.

Table 6. Stemming Process

happy	happy	use	dana	many
voucher				

The purpose of data pre-processing is to convert raw data into datasets from the results of crawling tweets from Twitter social media.

2.4 Lexicon-Based

Lexicon-based features are word features that have positive or negative sentiments based on a dictionary or lexicon. Lexicon is a collection of sentiment words that have been known and collected. For the weighting process on this feature, a dictionary or lexicon containing words containing sentiment called sentiment dictionaries is needed(Perdana and Fauzi 2017).

The lexicon-based approach is a sentiment analysis approach model often used in a study to determine classification, using a dictionary of language words or corpus equipped with weights on each word as a language or lexical source(Asri and Wahyu 2021). The purpose of creating a dictionary here is to minimize misclassification when the test data has ambiguous sentiment values(Mowlaei, Saniee Abadeh, and Keshavarz 2020). Lexicon-based can improve the performance in classifying sentiment(Khare and Chougule 2012).

The sentiments used are positive sentiments and negative sentiments. The processing stage of the positive lexicon word dictionary and the negative lexicon word dictionary is to break the tweet text into fragments of words to make it easier to determine positive and negative scores. Lexicon-based scores are calculated using equations 1 & 2.

$$if\ \Sigma_k\ Score\ (k)\ >\ then\ 0\ positive \tag{1}$$

$$if\ \Sigma_k\ Score\ (k) <\ then\ 0\ negative \tag{2}$$

In this research, the sentiment dictionary was obtained from previous research and then added manually according to the researcher's needs.

2.4.1 Data Labelling

A lexicon is a word feature with a positive or negative sentiment based on a dictionary or lexicon. Lexicon is a collection of sentiment words that have been known and compiled. For weighting this feature, a dictionary or lexicon containing words is needed, called a sentiment dictionary. At this stage, the sentiment dictionary used in this study uses a dictionary made by itself. An example can be seen in the following Table 7.

Table 7. Data Labelling

29 ears old, happy to use DANA because of the vouchers

Lexicon Calculation Example There are 0 negative words and 5 positive words detected in the lexicon dictionary, namely "happy", "happy", "dana", "many", and "voucher" as positive words. Then the calculation results are obtained.

Score = (number of positive words) - (number of negative words) then the score = 5 - 0 = 5. The final value obtained from the calculation results in a score of 5 or > 0 then identified as positive.

2.5 Naïve Bayes

Naïve Bayes Classifier is the simplest and most used classification. Naïve Bayes calculates class probabilities based on the distribution of words in the document(Indrayuni 2019).
Naïve Bayes classification is built on training data to estimate the probability of each category in the document features under test. The system will be trained using new data (training data and test data) and then given the task of guessing the target function value of the data(Andika, Azizah, and Respatiwulan 2019). Classification is the process of finding a set of models or functions that describe and distinguish classes of data so that the model can be used to predict the class of an object whose class label is unknown. Naive Bayes classification also shows high accuracy and is good when used for large data sets(Mahayani et al. 2020). In general, the process of naïve bayes classification can be seen in Equation 3.

$$\boldsymbol{P(cj\ |wi) = \frac{P(cj)\ P(wi\ |\ cj)}{P(wi)}} \tag{3}$$

In the classification calculation, the probability of word occurrence can be eliminated, and this is because the probability does not affect the comparison of classification results from each category. So, the classification process can be simplified with Equation 4.

$$\boldsymbol{P(cj\ |wi) = P(cj)\ P(wi\ |\ cj)} \tag{4}$$

In the classification calculation, a prior is used to calculate the chances of category occurrence in all documents, and the prior calculation can be seen in Equation 5.

$$\boldsymbol{P(cj) = NcN} \tag{5}$$

3 Results And Discussion

This section presents the results of the analysis with a brief discussion. Section 3.1 describes the dataset to provide an overview of the data used in this study. Section 3.2 presents the results of implementing the Naïve Bayes and lexicon methods and adding standard word corrections to the system. Section 3.3 presents the distribution of the dataset, which is divided into training data and test data.

3.1 Sentiment Result

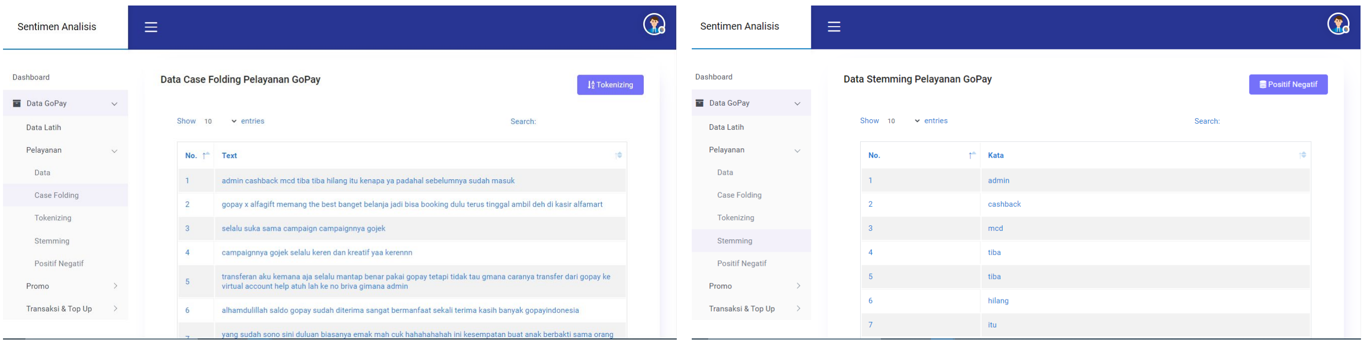
The data used and cleaned were 3,878 comments, and a non-standard word correction was carried out. This process helped correct words that were not standard Indonesian into standard Indonesian, then entered the pre-processing stage, then labelled based on the Lexicon dictionary using the Lexicon-Based method first. The data that becomes the dataset consists of two data, namely training data and test data, which are then classified using the Naïve Bayes method. Here are some examples of tweets that have been labelled in Table 8.

Table 8. Classified data

Tweet	Sentiment
The Gopay PayLater feature can really help my finances right now #WithPayLater	Positive
Wih cashback up to 250,000 OVO Points, not bad at all. The cashback can be used for a week's meal. #NoInvestDilemma @ovo_id	Positive
#DANADompetDigital is equipped with face verification or face verification for the secure opening of the application	Positive
Handling customer complaints is prolonged to respond, even though the need is very urgent!	Negative
Want to trf to bank mandiri Error, kenapan min? Again need fast, even error	Negative
I top up gopay. How come it doesn't go in. How is it min?	Negative

3.2 Implementation of Lexicon Based Software and Naïve Bayes

The results of this study are the results of accuracy using the Naïve Bayes and Lexicon methods, as well as the addition of standard word corrections based on tweet data about e-wallet applications in Indonesia and accurate results based on positive and negative classes. In this study, the results obtained are implemented in software. More details can be seen in Figure 2 below.



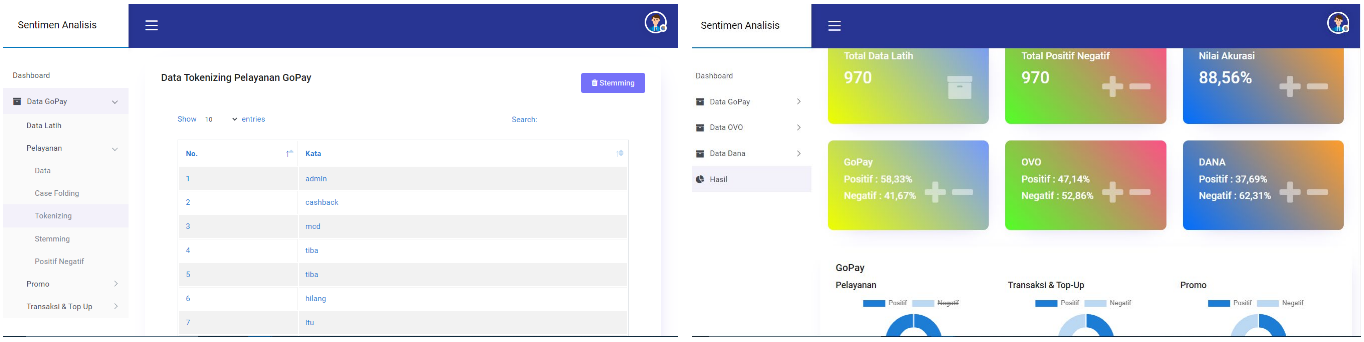


Figure 2. Software Implementation

The Figure 2 above is the result of the implementation of the software created. The software performs pre-processing, which has a case folding, tokenizing, and stemming process based on the category of e-wallet applications in Indonesia.

The image above is the result of tweets that have been processed, carried out a standard word repair process, and then classified with classification steps using a lexicon dictionary and classification using naïve bayes. It can be seen from the tweets that have been successfully classified into positive sentiment and negative sentiment.

3.3 Test Result

The data in this research is taken from a collection of Indonesian tweets taken from each customer's Twitter account using RStudio, then stored in CSV and excel. The keywords discussed use hashtags. The data used and cleaned as many as 3,878 comments. The ratio between training data and test data is 70:30, with 70% training data and 30% test data. Table 9 shows the distribution of training and test data.

Table 9. Distribution of Training and Test Data

	Training Data	Test Data	Total
Total Data	2908	970	3878

The accuracy calculation is carried out using the RapidMiner tool, as shown in Table 10. The accuracy calculation itself is needed to see whether the performance of the lexicon-based and naïve bayes methods is as expected, and it can be seen in the accuracy calculation itself in equation 6 below.

$$\text{Accuracy} = \frac{TP+TN}{Total} \qquad (6)$$

Table 10. Naïve Bayes Classification and Evaluation Results

Classification Category	True Positive	True Negative	Precision
Positive Prediction	439	86	83.62%
Negative Prediction	25	420	94.38%
Recall	94.61%	83.00%	

From the data results in the table above, with a total of 970 tweets, 464 true positives and 506 true negatives were obtained. These results show that the recall and precision presentations are between 83% and 94%. The use of naïve bayes produces an accuracy of 88.56% for the accuracy of classifying positive and negative sentiments.

4 Conclusions

Based on the results of the research that has been done, it can be concluded that the computational model classification mechanism in the sentiment analysis of e-wallet applications in Indonesia on Twitter social media shows that Twitter users in Indonesia are more likely to give negative comments on electronic wallet applications. By using the naïve bayes and lexicon methods and standard word repair, the accuracy rate is 88.56% for the accuracy of classifying positive and negative sentiment tweets on electronic wallets. This level of accuracy shows that sentiment analysis using the naïve bayes and lexicon methods and standard word repair is quite good because the resulting accuracy rate is 88.56%.

This study's findings are that using the lexicon method after pre-processing can reduce the number of data misclassification leaks. And by using a lexicon dictionary and standardized word repair to convert non-standard words into standard words according to the Big Indonesian Dictionary (KBBI), a significant and accurate increase in accuracy value can be obtained compared to using only one method.

Future research is expected to increase the amount of data used, modify the naïve bayes algorithm and add and refine the lexicon dictionary so that it can classify tweets into positive and negative sentiments more accurately and precisely. In addition, it is hoped that further research can improve this to facilitate the classification process in Indonesian by adding Lowenstein Distance normalization to the classification results to influence the accuracy value better. Besides that, online applications can be added to 4 to 5 electronic wallet applications, and it is hoped that this research can be used as a reference for further research.

Reference

Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A Study on Sentiment Analysis Techniques of Twitter Data." *International Journal of Advanced Computer Science and Applications* 10 (2): 361–74. 2019. https://doi.org/10.14569/ijacsa.2019.0100248.

Andika, Lingga Aji, Pratiwi Amalia Nur Azizah, and Respatiwulan Respatiwulan. "Analisis Sentimen Masyarakat Terhadap Hasil Quick Count Pemilihan Presiden Indonesia Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier." *Indonesian Journal of Applied Statistics* 2 (1): 34. 2019 https://doi.org/10.13057/ijas.v2i1.29998.

Asri, Jefry Sunupurwa, and Sawali Wahyu. "Analisis Sentimen Menerapkan Lexicon-Learning Based Untuk Melihat Opini Masyarakat Mengenai Protokol Kesehatan Dan Perkembangan Vaksin Covid-19 Di Indonesia Menggunakan Dataset Twitter." In *Konferensi Nasional Ilmu Komputer (KONIK)*, 530–36. 2021.

Buntoro, Ghulam Asrofi, Teguh Bharata Adji, Adhistya Erna Purnamasari, and Jurusan FT Teknik Elektro UGM Jln Grafika. "Sentiment Analysis Twitter Dengan Kombinasi Lexicon Based Dan Double Propagation Solidification Process of Metal View Project Sentiment Analysis Twitter Dengan Kombinasi Lexicon Based Dan Double Propagation." *CITEE*, 7–8. 2014.http://website.com.

Gunawan, Billy, Helen Sasty Pratiwi, and Enda Esyudha Pratama. "Sistem Analisis Sentimen Pada Ulasan Produk Menggunakan Metode Naive Bayes." *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)* 4 (2): 113. 2018. https://doi.org/10.26418/jp.v4i2.27526.

Haddi, Emma, Xiaohui Liu, and Yong Shi. "The Role of Text Pre-Processing in Sentiment Analysis." *Procedia Computer Science* 17: 26–32. 2013. https://doi.org/10.1016/j.procs.2013.05.005.

Halimi, Ahmad ;, and M Rudyanto Arief. "Analisis Sentimen Masyarakat Indonesia Terhadap Pembelajaran Online Dari Di Media Sosial Twitter Menggunakan Lexicon Dan K-Nearest Neighbor." *Jurnal Kecerdasan Buatan, Komputasi Dan Teknologi Informasi* 2 (1): 18–28. 2021. https://ejournal.unuja.ac.id/index.php/core/article/view/2283.

Indrayuni, Elly. "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes." *Jurnal Khatulistiwa Informatika* 7 (1): 29–36. 2019. https://doi.org/10.31294/jki.v7i1.1.

Joshi, Malini, and Hetal Vala. "Opinion Mining For Sentiment Data Classification." *International Journal of Research in Information Technology* 3 (1): 8–14. 2014.

Kadhim, Ammar Ismael. "An Evaluation of Preprocessing Techniques for Text Classification." *International Journal of Computer Science and Information Security* 16 (6): 22–32. 2018. https://sites.google.com/site/ijcsis/.

Khare, Vineet R., and Rahul Chougule. "Decision Support for Improved Service Effectiveness Using Domain Aware Text Mining." *Knowledge-Based Systems* 33: 29–40. 2012. https://doi.org/10.1016/j.knosys.2012.03.005.

Mahayani, Indira, Dewi Agushinta, R Muhammad, Edy Supriyadi, Fakultas Ilmu, Teknologi Informasi, Universitas Gunadarma, Jl Margonda, Raya No, and Jawa Barat. "ShopeePayLater Pada Aplikasi Belanja Online (Shopee) Menggunakan Metode Lexicon Based Dan Naïve Bayes Classi Er Pendahuluan Naïve Bayes Classi- Analisis Sentimen Dan Klasi-." *Jurnal Ilmiah KOMPUTASI* 19: 545–58. 2020.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment Analysis Algorithms and Applications: A Survey." *Ain Shams Engineering Journal* 5 (4): 1093–1113. 2014. https://doi.org/10.1016/j.asej.2014.04.011.

Mowlaei, Mohammad Erfan, Mohammad Saniee Abadeh, and Hamidreza Keshavarz. "Aspect-Based Sentiment Analysis Using Adaptive Aspect-Based Lexicons." *Expert Systems with Applications* 148: 113234. 2020. https://doi.org/10.1016/j.eswa.2020.113234.

Najib, Ahmad Choirun, Akhmad Irsyad, Ghiffari Assamar Qandi, and Nur Aini. "Perbandingan Metode Lexicon-Based Dan SVM Untuk Analisis Sentimen Berbasis Ontologi Pada Kampanye Pilpres Indonesia Tahun

2019 Di Twitter Abstrak." *Fountain of Informatics Journal* 4 (2). 2019. http://dx.doi.org/10.21111/fij.v4i2.3573.

Nugroho, Agung. "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram." *J-SAKTI (Jurnal Sains Komputer Dan Informatika)* 2 (2): 200. 2018. https://doi.org/10.30645/j-sakti.v2i2.83.

Parveen, Huma, and Shikha Pandey. "Sentiment Analysis on Twitter Data-Set Using Naive Bayes Algorithm." *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, ICATccT 2016*, 416–19. 2017. https://doi.org/10.1109/ICATCCT.2016.7912034.

Perdana, Rizal Setya, and Muhammad Ali Fauzi. "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine Dan Lexicon Based Features Twitter Event Detection View Project Human Detection and Tracking View Project." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1(12) (October): 1725–32. 2017. https://www.researchgate.net/publication/320234928.

Prananda Antinasari, Rizal Setya Perdana, M. Ali Fauzi. "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1 (12): 1733–41. 2017.

Salim, Siti Saidah, and Joanna Mayary. "Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighb." *Jurnal Ilmiah Informatika Komputer* 25 (1): 1–17. 2020. https://doi.org/10.35760/ik.2020.v25i1.2411.

Troussas, Christos, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno, and Jaime Caro. "Sentiment Analysis of Facebook Statuses Using Naive Bayes Classifier for Language Learning." *IISA 2013 - 4th International Conference on Information, Intelligence, Systems and Applications*, 198–205. 2013. https://doi.org/10.1109/IISA.2013.6623713.

Xu, Feng, Zhenchun Pan, and Rui Xia. "E-Commerce Product Review Sentiment Classification Based on a Naïve Bayes Continuous Learning Framework." *Information Processing and Management* 57 (5): 1022212020.. https://doi.org/10.1016/j.ipm.2020.102221.

Zhang, Lei, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu. "Combining Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis." *HP Laboratories Technical Report*, no. 89. 2011.

Biography

Tacbir Hendro Pudjiantoro., S.Si., M.T., is a lecturer in the Department of Information Systems, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani, Indonesia. Obtained his S.Si. at Universitas Padjadjaran, Indonesia, in 1995 and obtained his M.T. at the Bandung Institute of Technology, Indonesia, in 2001.

Dra., Ade Kania Ningsih, M.Stat., is a lecturer in the Department of Informatics, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani, Indonesia. Obtained a Dra. at Universitas Padjadjaran, Indonesia, in 1987 and obtained an M.Stat.

Faiza Renaldi, S.T., M.Sc., is a lecturer in the Department of Information Systems, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani Indonesia. Obtained his M.Sc., at Universiteit Utrecht, Netherlands in 2006.

Ananda Dimas Sanjaya is an undergraduate student in Informatics at Universitas Jenderal Achmad Yani, West Java, Cimahi, Indonesia, and joined the field of Informatics at the Faculty of Science and Informatics in 2017.