E-ISSN: 2622-304X, P-ISSN: 2622-3031
Available online at:
http://proceedings.itbwigalumajang.ac.id/index.php/progress

Sentiment Analysis Based on Review on Flip Application in Google Playstore

Fadhel Akhmad Hizham[1], Marita Qori'atunnadyah[2], Jesi Irwanto[3]

Departmen of Informatics, Institut Teknologi dan Bisnis Widya Gama Lumajang, Indonesia[1,2,3]

Corresponding Author: hizhamfadhel@gmail.com

Abstract

Money is one of the legal means of exchange for payment, both in the form of cash and electronic money which is allocated in the form of balances. Along with the development of technology, the use of electronic money is increasingly being used. Electronic money also makes it easier to send transfer-based money. However, transfer-based remittances incur an additional fee as an administration fee. In Google Play Store, there are several transfer applications without administration fees. Flip application is one of application which has a high review score and the highest number of downloads on transfer-based applications. Almost all users before using the application rely on comments as a consideration in using the application, because these comments provide the latest information from the point of view of users who have used the application. Comments that number in the hundreds of thousands certainly have difficulty analyzing each comment manually, so a sentiment analysis process is carried out. This study used Decision Tree, Naïve Bayes, and Backpropagation Neural Network methods for classifying sentiment. The Backpropagation Neural Network method has the best classification evaluation in terms of accuracy, recall, and F1-Score, with an average value of 87.85%, 97.49% and 91.39% respectively.

Keywords: Backpropagation Neural Network, Decision Tree, Naïve Bayes, Sentiment Analysis

INTRODUCTION

Money is a legal means of exchange for payment, both in the form of cash and electronic money allocated in the form of balances. As technology develops, electronic money is increasingly being used. Electronic money is an innovation formulated and implemented by Agus D. W. Martowardojo as Governor of Bank Indonesia called the National Non-Cash Movement (GNTT) (Anjelina, 2018).

Electronic money has a tremendous impact on digital transactions and the other aspects (Habibah et al., 2023). Electronic money also makes it easier to send money based on transfers. However, transfer-based money are subject to additional fees as administration fees. Therefore, many applications have been developed to transfer money without additional administration fees.

In Google Play Store, there are several transfer applications without administration fees. In this study, researchers chose the Flip application because this application has a high review score and the highest number of downloads when compared to other free transfer-based applications. Almost some users before using the application rely on comments as a consideration in using the application, because these comments provide the latest information from the point of view of users who have already used the application. There are hundreds of thousands of comments, of course it is difficult to analyze each comment manually, so a sentiment analysis process is carried out (Masturoh & Pohan, 2021).

Sentiment analysis is an analytical technique for processing various opinions expressed in text form by utilizing media related to a product or application. Sentiments contain textual information and have polarities that are positive, negative or neutral (Romadloni et al., 2019). This polarity is used as a reference for making decisions and for developing a product or application for the future (Habibah et al., 2023).

Sentiment analysis is using a various method of classification to determine polarities from reviews or comments. Some research about sentiment analysis include Backpropagation Neural Network (BNN) method (Anugerah & Sibaroni, 2022), (Abiyyu Musa & Sibaroni, 2022), (Wiguna et al., 2022), Naive Bayes Classification (NBC) method (Izunnahdi et al., 2023), (Ulya et al., 2022), Support Vector Machine (SVM) method (Qinanda & Nilogiri, 2023), Decision Tree method (Asshiddiqi & Lhaksmana, 2020), (Wahid & Saputri, 2022), and K-Nearest Neighbor method (Romadloni et al., 2019), (Anam et al., 2021).

From this description, researchers analyzed sentiment on the Flip application based on user reviews of the application, using the Decision Tree, Naïve Bayes, and Backpropagation Neural Network methods. The main goals of this research is to compare the three methods of classification method to determine the best method for classifying sentiment analysis on Flip application.

METHODS

Collecting Data

The data obtained in this research are user reviews of the Flip application from the Google Play Store. The amount of data used is 1,000 reviews, with a 1-star score to a 5-star score for 200 reviews each.

Text Preprocessing

Text preprocessing is the process of changing the form of unstructured textual data into structured data. In the text preprocessing process, there are 4 stages, namely case folding, tokenizing, stopwords, and stemming (Rahman et al., 2017).

Case folding is a text preprocessing process that changes all letters in a document to lowercase. The case folding process only uses the characters 'a' to 'z' which are acceptable, while characters other than letters are removed and are considered delimiters. For example, a user who wants to get the information "ANALYSIS" and types "ANALYSIS", "AnaLisis", or "analysis", will still be given the same value, namely "analysis". An example of a review that goes through the case folding stage is shown in Figure 1.

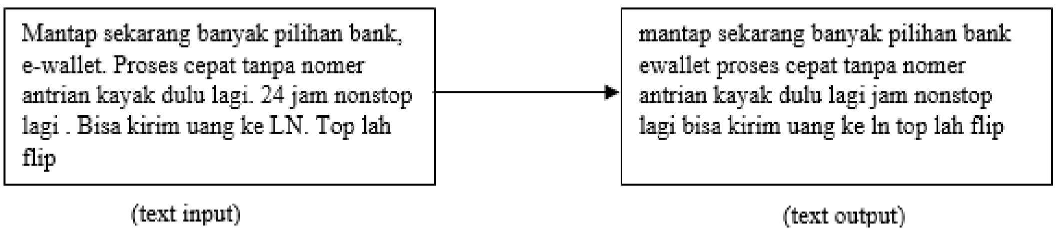

Figure 1. Case Folding

Tokenizing is a text preprocessing process that cuts the input string or sentence based on the words that make it up. An example of a review that goes through the tokenizing stage is shown in Figure 2.

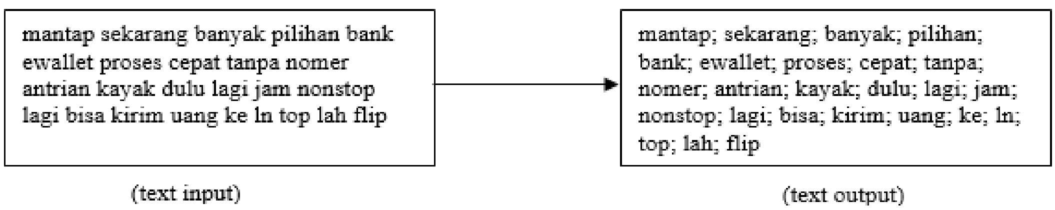

Figure 2. Tokenizing

Stopwords is a text preprocessing process that takes important words from the tokenization results or deletes words that are too common. Before carrying out stopword removal, there must be a stopword list (stoplist). If it is included in the stoplist, the word is deleted and only the words that characterize the contents of the document are left (Manalu, 2014). Examples of stop words are "which", "from", "to", "in", and so on. These words are examples of high frequency words that can be found in almost every document. Stopword removal can reduce index size and processing time. In this research, the stop words used come from Sastrawi (Chandra et al., 2016). An example of a review that goes through the stopwords stage is shown in Figure 3.

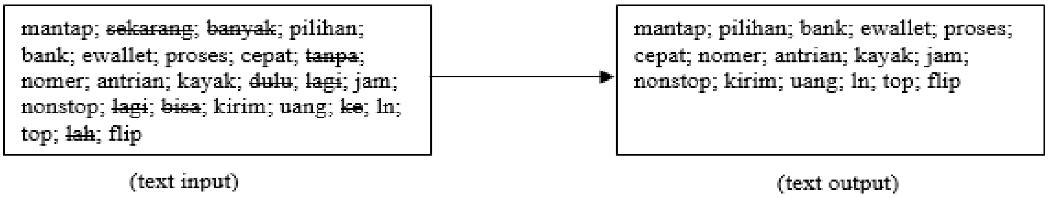

Figure 3. Stopwords

Stemming is a text preprocessing process that groups words that have similar base words and meanings but have different meanings because they have different affixes. For example, the words election, choice, voter will be returned to their basic word form, namely "choose". In this research, stemming comes from the Sastrawi Stemmer library in Python. An example of a review that goes through the stemming stage is shown in Figure 4.

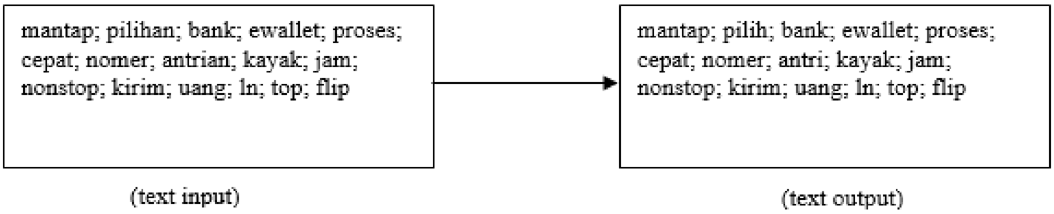

Figure 4. Stemming

Word Weighting

After the text preprocessing stage, the next stage is word weighting. This stage gives a score to the appearance of words, which uses the Process Document from Data operator which uses the term frequency – inverse document frequency (TF-IDF) method (Ginantra et al., 2022). Term Frequency (TF) is used to measure the frequency with which a word appears in a document (Pratama et al., 2023).

The TF calculation is divided by the length of the document as a way to normalize it or by using logarithmics which aims to minimize the relative influence of the number of frequencies of words that appear in one document. The formula for calculating TF can be defined as:

$$tf(t,d) = \begin{Bmatrix} 1 + \log(f_{t,d}) & \text{if } f_{t,d} > 0 \\ 0 & \text{if } f_{t,d} = 0 \end{Bmatrix} \quad (1)$$

Note that $tf(t,d)$ is the Term Frequency and $f_{t,d}$ is the world of world *t* in document *d.*

Inverse Documents Frequency (IDF) is used to measure the importance of the word. When calculating TF, all words are considered important, but adverbs and words that have a high frequency of occurrence or are found in almost all documents need to be balanced with those that are rare (low frequency of occurrence). The formula for calculating IDF can be defined as:

$$idf(t,D) = \log \frac{N}{df_t} \quad (2)$$

Note that N is the total number of documents in a set of documents and df_t is the number of documents that contain the word *t*, if the word value does not exist in all documents (0) then the value is 1.

Based on the TF and IDF calculations that have been described in (1) and (2), the TF-IDF value is calculated by multiplying the equation of TF and IDF. The formula for calculating TF-IDF can be defined as:

$$tf.idf(t, d, D) = \left(1 + \log\left(f_{t,d}\right)\right)\left(\log \frac{N}{df_t}\right) \tag{3}$$

The value of TF-IDF can affect according to the frequency of occurrence of the term in a document and balanced by the number of documents in the corpus containing the term. Values can increase proportionally which helps to adjust for the fact that some terms can appear frequently (Fikri, 2019).

Naive Bayes Classification

Naive Bayes is a classification method with sample probability. This method calculates probability by adding up the frequencies and mixture of values from the dataset. In other words, the presence or absence of one feature of a class is not related to other features (Krithiga et al., 2021). The formula of Naive Bayes Classification can be defined as:

$$P(Y \mid X) = \frac{P(Y) \prod_{i=1}^{q} P(X_i|Y)}{P(X)} \tag{4}$$

While Naive Bayes Classification with continuous feature (Gaussian Naive Bayes) can be defined as:

$$P(X \mid Y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{5}$$

Note that $P(X \mid Y)$ is data probability with attribute X at class Y (posterior probability), $P(Y)$ is probability of class Y (priot probability), σ is standard deviation, μ is mean of attributes in continouous feature, and $\prod_{i=1}^{q} P(X_i|Y)$ is independent probability class Y from all features in vector X.

Decision Tree

Decision Tree with the C4.5 algorithm approach is a classification technique used to extract relevant relationships with data. This algorithm divides the training data with the help of information gain. Attributes that have a high frequency are considered to separate the data from the information available in the dataset (Hardiani, 2021). The formula of Naive Bayes Classification can be defined as (Hozairi et al., 2021):

$$Entropy(S) = \sum_{i=1}^{n} pi * log_2 pi \tag{6}$$

Note that *S* is case set, *n* is number of *S* partitions, and *pi* is proportion from S_i on *S*. After determining a value of entropy, the next step is counting Information Gain, with the formula:

$$Gain\ (S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_1|}{|S|} * Entropy(S_1) \tag{7}$$

Note that *A* is attribute, $|S_i|$ is number of case at i-th partition, and $|S|$ is number of case at S.

Backpropagation Neural Network (BNN)

BNN is a supervised training algorithm that has many layers. BNN uses the output error to change the value of its weights in a backward direction. To get this error, the forward propagation stage must be carried out first.

There are 3 phases in BNN training, namely the forward phase (feed forward), the backward phase (back propagation), and the weight modification phase. In the feed forward phase, the input pattern is calculated forward starting from the input layer to the output layer. In the back propagation phase, each output unit receives a target pattern that is related to the input pattern to calculate the error value. The error will be propagated backwards. Meanwhile, the weight modification phase aims to reduce errors that occur. These three phases are repeated continuously until the termination condition is met.

Evaluating Classification

In this research, classification testing was carried out by implementing the Backpropagation Neural Network (BNN) method, which then determined the classification performance using a confusion matrix. From the confusion matrix, accuracy, precision and recall measurement results are obtained. Accuracy is an assessment of the accuracy of classification obtained from comparing correctly classified positive and negative data with all data. Precision is an assessment of the relevance of the data sought to the information needed. Recall is an assessment of the system's success in retrieving information. The confusion matrix is presented in Table 1 (Saputra et al., 2018).

Table 1. Confusion Matrix

		Actual Class	
		True	*False*
Predicted Class	*True*	TP (*True Positive*)	FP (*False Positive*)
	False	FN (*False Negative*)	TN (*True Negative*)

True Positive (TP) means if the predicted data is positive and matches the actual (positive) value. False Positive (FP) means if the predicted data does not match the actual value. False Negative (FN) means if the predicted value is negative and the actual value is positive. True Negative (TN) means if the prediction is negative and the actual negative is true. The measurement of accuracy, precision, recall and F-Measure values based on the confusion matrix is formulated in equations 8 to 11:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{F-Measure} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{11}$$

RESULTS AND DISCUSSION

Collecting Data Results

Before implementing the method, the initial stage in this research is collecting data. In this research, the data collected is user reviews of the Flip application obtained from the Google Play Store with a total of 1,000 reviews, with a 1 star score to a 5 star score for 200 reviews each.

Text Preprocessing

After the data is collected, the text preprocessing stage is carried out first, before classification is carried out to analyze review sentiment. The text preprocessing stages consist of case folding, tokenizing, stopwords, and stemming.

After the stemming process is carried out, the text preprocessing is complete. The next stage is to assign a weight to each word for each dataset, namely by using TF-IDF. Of the 1,000 reviews, there are 9,429 words that have been processed to the stemming stage and there are 1,918 different words if words containing duplicates are removed. TF-IDF weighting begins by looking for TF, which is the number of occurrences of words from each dataset. Details of the calculation are shown in Table 2.

Table 2. Calculation of TF

Data	TF(flip)	TF(transfer)	TF(aplikasi)	TF(transaksi)	TF(bantu)
1	0	1	0	1	0
2	0	2	0	1	0
3	0	0	0	0	0
4	0	0	1	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	1	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮
1000	0	1	0	0	0

Source: Processed Data (2023)

After looking for the TF value of each word for each dataset, the next step is to look for the DF value. The DF value is obtained from the number of occurrences of words in each dataset, whose Tf value is more than 0. The DF value is in table 3.

Table 3. DF Value

DF(flip)	DF(transfer)	DF(aplikasi)	DF(transaksi)	DF(bantu)
255	189	135	119	101

Source: Processed Data (2023)

After finding the DF value, the next step is to find the IDF value for each word in that keyword. The IDF value functions as a balance for words that appear too often in all documents by calculating the base 10 log value of the number of documents divided by the DF value. The greater the DF value, the smaller the IDF value, and vice versa. Based on the DF value that has been obtained, the IDF value is as in Table 4.

Table 4. IDF Value

IDF(flip)	IDF(transfer)	IDF(aplikasi)	IDF(transaksi)	IDF(bantu)
0,5935	0,7235	0,8697	0,9245	0,9957

Source: Processed Data (2023)

After finding the IDF value, the next step is to change the TF-IDF value by multiplying the TF by the IDF. For example, the TF value for the word "transfer" for dataset number 2 is 2, while the IDF value for the word "transfer" is 0.7235. So the TF-IDF value is 2 x 0.7235 = 1.4470. The results of the TF-IDF calculation are shown in Table 5.

Table 5. Calculation of TF-IDF

Data	TF-IDF (flip)	TF-IDF (transfer)	TF-IDF (aplikasi)	TF-IDF (transaksi)	TF-IDF (bantu)
1	0	0,723538	0	0,924453	0
2	0	1,447076	0	0,924453	0
3	0	0	0	0	0
4	0	0	0,869666	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0,869666	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0,869666	0	0
⋮	⋮	⋮	⋮	⋮	⋮
1000	0	0,723538	0	0	0

Source: Processed Data (2023)

Classification Method Implementation

The number of attributes used in this research is 1,918 attributes containing words after text preprocessing with the value of each attribute based on the TF-IDF calculation that has been obtained. To determine the class, it is determined from the review score from each dataset. If the review score is 1 to 3 stars, it is counted as a "negative" class or 0, while for a review score of 4 to 5 stars, it is counted as a "positive" class or 1 In this research, there are 600 reviews that have a negative class, and 400 reviews that have a positive class.

Evaluating Classification

Evaluation on this research used confusion matrix, including actual and prediction class. From confusion matrix, the score of accuracy, precision, recall, and F1-score are obtained by using formula 8 to 11. In this research, trials were carried out with various experiments for dividing the dataset, including 60:40 (600 train data and 400 test data), 70:30 (700 train data and 300 test data), 75:25 (750 train data and 250 test data), 80:20 (800 train data and 200 test data), and 90:10 (900 train data and 100 test data). The results of accuracy, precision, recall and F1-score for each test set using the Naïve Bayes, Decision Tree and Backpropagation Neural Network methods are shown in Table 6 to 9.

Table 6. Accuracy Result

Number of Test Set	Classification		
	Naive Bayes	Decision Tree	Backpropagation
400	68,00%	84,00%	96,50%
300	68,33%	82,00%	92,67%
250	66,00%	79,20%	85,60%
200	68,00%	78,50%	97,50%
100	69,00%	69,00%	67,00%
Average	67,86%	78,54%	87,85%

Source: Processed Data (2023)

Table 7. Precision Result

Number of Test Set	Classification		
	Naive Bayes	Decision Tree	Backpropagation
400	80,00%	95,24%	95,45%
300	77,85%	93,29%	98,79%
250	77,05%	91,60%	80,75%
200	77,00%	89,47%	96,69%
100	80,00%	80,00%	64,89%
Average	78,38%	89,92%	87,31%

Source: Processed Data (2023)

Table 8. Recall Result

Number of Test Set	Classification		
	Naive Bayes	Decision Tree	Backpropagation
400	66,14%	78,74%	99,21%
300	67,21%	75,96%	89,07%
250	62,25%	72,19%	100,00%
200	65,25%	72,03%	99,15%
100	65,57%	65,57%	100,00%
Average	65,28%	72,90%	97,49%

Source: Processed Data (2023)

Table 9. F1-Score Result

Number of Test Set	Classification		
	Naive Bayes	Decision Tree	Backpropagation
400	72,41%	86,21%	97,30%
300	72,14%	83,73%	93,68%

250	68,86%	80,74%	89,35%
200	70,64%	79,81%	97,91%
100	72,07%	72,07%	78,71%
Average	71,22%	80,51%	91,39%

Source: Processed Data (2023)

Based on Table 6 to 9, it can be concluded that in analyzing the sentiment of Flip user reviews, the Backpropagation Neural Network method has the best classification evaluation in terms of accuracy, recall and F1-Score, with an average value of 87.85%, 97.49% and 91.39% respectively. Meanwhile, the Decision Tree method has the best classification evaluation in terms of precision, with an average of 89.92%. Therefore, applying the Backpropagation Neural Network classification method is the best method for conducting sentiment analysis from Flip user reviews on the Google Play Store.

CONCLUSION

Sentiment analysis on the Flip application based on user reviews is carried out by collecting review data from Google Play with total 1,000 reviews, 200 reviews each for a score of 1 to 5. Before the sentiment analysis process is carried out, the data that has been collected is processed first through the text preprocessing stage, such as case folding, tokenizing, stopwords, and stemming. After the data is processed, the weight of each word per review is calculated using TF-IDF. After getting the weights with TF-IDF, the data is classified using 3 classification methods, namely Naïve Bayes, Decision Tree, and Backpropagation Neural Network. The dataset that is classified consists of 1,918 attributes in the form of words that have been given TF-IDF weights and the class of the dataset consists of 600 reviews with the "negative sentiment" class or 0 for reviews with a score of 1 to 3 and 400 reviews with the "sentiment" class. positive” or 1 for reviews with a score of 4 to 5. Based on the results of the confusion matrix, the application of the Naïve Bayes classification method obtained an average accuracy value of 67.86%, precision 78.38%, recall 65.28%, and F1-score 71.22%. The application of the Decision Tree classification method obtained an average accuracy value of 78.54%, precision 89.92%, recall 72.90%, and F1-score 80.51%. The application of the Backpropagation Neural Network classification method obtained an average accuracy value of 87.85%, precision 87.31%, recall 97.49%, and F1-score 91.39%. Of the three classification methods, the Backpropagation Neural Network method is the best classification method for analyzing user sentiment for the Flip application on the Google Play Store.

The future work of this research is on the text preprocessing stage, in addition to the four text preprocessing stages, the Part-Of-Speech (POS) Tagging or n-gram algorithm can be added. Apart from that, to increase the accuracy percentage, you can add the reviews used for classification, and increase the comparison of training data to testing data. The more training data, the testing data formulation can improve and be more accurate in classifying sentiment.

REFERENCES

Abiyyu Musa, M. R., & Sibaroni, Y. (2022). Radicalism Speech Detection in Indonesia on Twitter Using Backpropagation Neural Network Method. *Journal of Computer System and Informatics (JoSYC)*, *3*(4), 356–362. https://doi.org/10.47065/josyc.v3i4.2146

Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na¨ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 139–150. https://doi.org/10.30812/matrik.v21i1.1092

Anjelina, A. (2018). PERSEPSI KONSUMEN PADA PENGGUNAAN E-MONEY. *JOURNAL OF APPLIED MANAGERIAL ACCOUNTING*, *2*(2). https://doi.org/10.30871/jama.v2i2.934

Anugerah, A. P., & Sibaroni, Y. (2022). Sentiment Analysis Of Development Jakarta-Bandung High-Speed Train Using Twitter Social Media With BNN Method. *JURIKOM (Jurnal Riset Komputer)*, *9*(3), 543–549. https://doi.org/10.30865/jurikom.v9i3.4143

Asshiddiqi, M. F., & Lhaksmana, K. M. (2020). Perbandingan Metode Decision Tree Dan Support Vector Machine Untuk Analisis Sentimen Pada Instagram Mengenai Kinerja Pssi. *EProceedings of Engineering*, *7*(3).

Chandra, D. N., Indrawan, G., & Sukajaya, I. N. (2016). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmiah Teknologi Informasi Asia*, *10*(1).

Ginantra, N. L. W. S. R., Yanti, C. P., Prasetya, G. D., Sarasvananda, I. B. G., & Wiguna, I. K. A. G. (2022). Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *11*(3), 205–215. https://doi.org/10.23887/janapati.v11i3.49450

Habibah, N., Budianita, E., Fikry, M., & Iskandar, I. (2023). Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor. *JURIKOM (Jurnal Riset Komputer)*, *10*(1), 192–200. https://doi.org/10.30865/jurikom.v10i1.5429

Hardiani, T. (2021). Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta. *International Journal of Health Science and Technology*, *2*(1). https://doi.org/10.31101/ijhst.v2i1.1829

Hozairi, H., Anwari, A., & Alim, S. (2021). IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES. *Network Engineering Research Operation*, *6*(2). https://doi.org/10.21107/nero.v6i2.237

Izunnahdi, M., Aburrahman, G., & Wardoyo, A. E. (2023). Sentimen Analisis pada Data Ulasan Aplikasi KAI Access di Google Play Store menggunakan Metode Multinomial Naïve Bayes. *Jurnal Smart Teknologi*, *4*(2), 192–198.

Krithiga, B., Sabari, P., Jayasri, I., & Anjali, I. (2021). Early detection of coronary heart disease by using naive bayes algorithm. *Journal of Physics: Conference Series*, *1717*(1). https://doi.org/10.1088/1742-6596/1717/1/012040

Manalu, B. U. (2014). ANALISIS SENTIMEN PADA TWITTER MENGGUNAKAN TEXT MINING. *Teknologi Informasi Fak. ILKOM UNSUT*.

Masturoh, S., & Pohan, A. B. (2021). Sentiment Analysis Against the Dana E-Wallet on Google Play Reviews Using The K-Nearest Neighbor Algorithm. *Pilar Nusa Mandiri Journal of Computing and Information System*, *17*(1), 53–58. https://doi.org/10.33480/pilar.v17i1.2182

Pratama, Y., Januar Miraswan, K., & Selatan, S. (2023). Sentiment Analysis Using Pseudo Nearest Neighbor and TF-IDF Text Vectorizer. *Sriwijaya Journal of Informatic and Applications*, *4*(2), 20–26.

Qinanda, M. D., & Nilogiri, A. (2023). Akurasi Sentimen Pada Komentar Youtube Untuk Permendikbud Tentang Pencegahan Dan Penanganan Kekerasan Seksual Berbasis Naïve Bayes Dan Support Vector Machine. *Jurnal Smart Teknologi*.

Rahman, A., Wiranto, & Doewes, A. (2017). Online News Classification Using Multinomial Naive Bayes. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, *6*(1).

Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, KNN Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *IKRA-ITH Informatika*, *3*(2), 1–9.

Saputra, D. D., Pratama, B., Akbar, Y., & Gata, W. (2018). Penerapan Text Mining untuk Assignment Complaint Handling Customer Terhadap Divisi Tterkait menggunakan Metode Decision Tree Algoritma C4.5 (Studi Case: PT. XL AXIATA, TBK). *Journal CKI On SPOT*, *11*(2), 207–216.

Ulya, S., Ridwan, A., Cholid Wahyudin, W., & Hana, F. M. (2022). Text Mining Sentimen Analisis Pengguna Aplikasi Marketplace Tokopedia Berdasar Rating dan Komentar Pada Google Play Store. *Jurnal Bisnis Digital Dan Sistem Informasi*, *3*(2), 33–40.

Wahid, A., & Saputri, G. (2022). Analisis Sentimen Komentar Youtube Tentang Relawan Patwal Ambulance Menggunakan Algoritma Naïve Bayes dan Decision Tree. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2). https://doi.org/10.30865/json.v4i2.4941

Wiguna, I. K. A. G., Sugiartawan, P., Sudipa, I. G. I., & Pratama, I. P. Y. (2022). Sentiment Analysis Using Backpropagation Method to Recognize the Public Opinion. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems*, *16*(4), 423–434. https://doi.org/10.22146/ijccs.78664