Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.2, April 2025, pp. 526~533
e-ISSN: 2548-6861

Twitter Sentiment Analysis on Digital Payment in Indonesia Using Artificial Neural Network

Siska Febriani [1], Vera Wati [2], Yuli Wijayanti [3], Irwan Siswanto [4]
[1,3,4] Informatika, Institut Teknologi Bisnis Dan Kesehatan Bhakti Putra Bangsa Indonesia
[2] Sistem Informasi Kota Cerdas, Politeknik Negeri Indramayu
siska17feb@gmail.com [1], vera.w@polindra.ac.id [2], yuliwijayanti118@gmail.com[3], irwan.siswanto00@gmail.com[4]

Article Info

Article history:

Received 2024-12-02
Revised 2025-01-22
Accepted 2025-02-20

Keyword:

Artificial Neural Network,
Digital Payment,
Sentiment Analysis,
Twitter

ABSTRACT

In the rapid development of technology, the need for big data processing is increasingly important, especially in the context of digital transactions such as e-wallets in Indonesia. On the other hand, sentiment analysis of digital payment platforms via Twitter requires fast and accurate data processing, but often faces challenges in managing big data and optimal classification quality. This study uses the Term TF-IDF method for text preprocessing and Artificial Neural Network (ANN) for sentiment classification. The preprocessing process includes case folding, removing numbers and punctuation, tokenization, filtering, and stemming. For classification, ANN is used which is optimized with the Backpropagation and K-fold Cross Validation algorithms to improve the accuracy of the model in grouping positive and negative sentiments from tweets about digital payment platforms. Through this approach, the study produces a sentiment classification model in analyzing big data. The results in this study are Gopay gets a positive value and gets the first value in sentiment assessment with an accuracy rate of 72% using ANN. Of the 5 digital payments that received a negative value and ranked last, namely Link Aja with an achievement rate of 43%. Based on these results, it shows that this approach contributes to identifying consumer sentiment towards e-wallet platforms, which is useful for developing digital marketing strategies. The contribution given is in improving sentiment analysis of digital payment platforms by utilizing Big Data processing technology and machine learning, so that it can be used to improve services and marketing strategies based on user data.

I. INTRODUCTION

In 2019 this technological development has developed rapidly. One area of technology that has the opportunity to develop is the Big Data field. Big Data is a term given to data sets that are very large and complex in size, making it impossible to process them using conventional database management tools or other data processing applications. The rampant habits of the Indonesian people to shop online have an impact on the need for fast and efficient payment methods. Nowadays there are many developing digital payments in Indonesia. Digital payment often called e-wallet is a new method of transaction tool that no longer requires banknotes or checks to transact [1]. Research on the Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification examined by Sanjay Yadav shows that to a certain extent, cross-validation k-fold with var-Indonesian people's rampant habits to shop online impactor enduring validation to classification quality. In the results of his report, he studied the differences between two validation schemes and analyzed the possibility of using k-fold cross-validation, which exceeds hold-out validation even on large datasets [2].

The application uses deep convolution neural networks for the classification of sentiments in Twitter tweets in this work. Our approach combines pre-trained word embedding features generated using word sentiment polarity features based on lexicon features and n-gram sentiment features as sentiment feature vectors from tweets, and incorporating feature sets into deep convolution neural networks that have also been investigated by Zhao Jianqiang [3]. In his research comparing the performance of our model with the base model which is the n-gram word model on five Twitter datasets, the results show that our model performs better on the accuracy and classification of F1-Measure for Twitter sentiments [3].

The Word-Level and Sentences-Leves Language Identification study: Application to Algerian and Arabics Dialects by Mohamed Lichouri concluded that by combining three methods for text classification, this study gives results that the binary classification model produces better scores on dialects Arabic specifically, with an average accuracy for Algerian dialects between 62% at the word level and 76% at the sentence level [4]. This is from several processing processes by comparing several methods for the identification of textual Arabic dialects by considering word and sentence-level approaches. This study uses three methods for text classification, namely Linear Support Machine L-SVM, sentence-level approach Bernoulli Naive Bayes BNB, and Multinomial Naive Bayes MNB [4].

In sentiment analysis research, the use of Artificial Neural Networks (ANN) has significant advantages compared to other algorithms such as Naïve Bayes, SVM, and Random Forest. Several studies have shown that machine learning, especially ANN, is more effective in handling complex and unstructured text data, such as those found in social media sentiment analysis. Similar studies that have been conducted on Neural Networks for sentence classification show that their performance can capture contextual dependencies that are difficult to find by linear models such as Naïve Bayes [5]. In addition, Bengio et al. emphasize that ANNs are able to handle non-linear relationships in data, an important capability in sentiment analysis which often involves highly variable sentence contexts [6]. ANN models are also better able to handle big data, where other methods often struggle. The performance of machine learning can also process very large and complex data volumes better than traditional methods [3][7][8]. Other studies have also highlighted the use of CNNs at the character level to address language variations, such as spelling or slang that frequently appear in social media data [9], while Mikolov et al. showed that word representation using Word2Vec in ANN enhances the understanding of word context [10]. Another advantage is that the use of ANN can help overcome the problem of vanishing gradients or problems that are often encountered when training neural networks [11]. When dealing with the problem of word order and context in text, the sequence-to-sequence model is very effective, giving better results compared to other algorithms that do not take word order into account [12][13][14]. With the ability to handle large, complex, and non-linear data, as well as the flexibility to understand the context of sentences and relationships between words, ANN is a great choice for sentiment analysis. The advantages of ANN in processing highly varied and complex data make it much more effective than other algorithms, especially in the context of social media sentiment analysis.

In addition, ANN can also handle large and complex data more effectively, thanks to its ability to learn from large amounts of data and find deeper and more relevant feature representations, which is very useful in analyzing Twitter text data that have high diversity and complexity. ANN is also more flexible in handling various types of features, whether numeric, categorical, or text, making it very suitable for sentiment analysis on social media that often has unstructured data. Previous studies have shown that ANN, especially in the form of deep learning and transformers, excels in handling text data that has complex contextual dependencies [15][16]. Emphasis on the importance of selecting the right number of layers and neurons, as well as the appropriate activation function, to achieve the best performance in deep learning applications, including sentiment analysis [17][18]. Thus, to ensure the validity and transparency of the research results, it is very important to include a more in-depth explanation of the ANN network structure used, so that the evaluation results and use of the model can be accounted for and optimized optimally. Based on some of these studies, the authors try to research to classify text data using a neural network by making comparisons during preprocessing using term frequency and term frequency-inverse document frequency.

II. METHOD

2.1 Term Frequency – Inverse Document Frequency

The TF-IDF (Term Frequency - Inverse Document Frequency) method is a method used to calculate the weight of each word that is most commonly used in information retrieval in a document. This method is used by many people because it has characteristics that are efficient, easy and have fairly good accuracy results [19]. This method will calculate the value of Term Frequency (TF) and Inverse Document Frequency (IDF) on each token (word) in each document in the corpus or collection of words in a particular document that is entered into the system. This method will calculate the weight of each token t in document d using the formula :

$$W_{dt} = tf_{dt} * IDF_t$$

Where d is the d-th document of all documents to be selected, t is the t-word of the keyword, W is the weight of the d-th document of the t-word, tf (term frequency) is the number of words searched for a particular document, IDF is an Inversed Document Frequency where IDF value is obtained from log2 (D / df) which is (D) total document and df is a lot of documents containing words that are searched for in a particular document. After the weight (W) of each document is known, then the sorting process is carried out where the greater the value of W, the greater the level of similarity of the document to keywords, and vice versa. So that values can be obtained from each word to be included in the training that neural networks will carry out [20].

2.2 Artificial Neural Network

An Artificial Neural Network (ANN) is a network that models the human brain's nervous system using neurons arranged according to a particular architecture. The human brain is very complex, non-linear and processes information in parallel, and can organize neurons to recognize patterns effectively [21]. The neurons in ANN are modeled by artificial nerve cells called perceptrons. Perceptron is the simplest form of a neural network that is used to classify patterns [22]

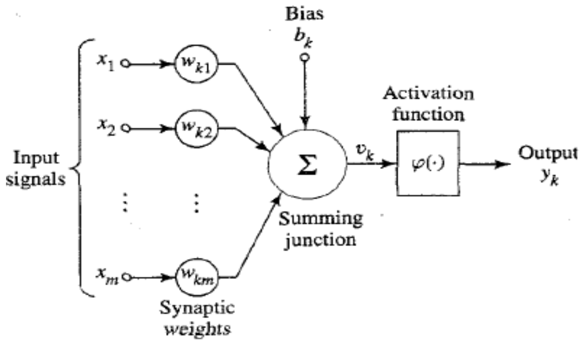


Figure 1. Neuron Model [22]

In Figure 1 there is a bias weight which is defined as b_k which gets a value of input +1 which is useful to avoid a linear combiner that is worth 0. The input signal is defined as $x_1, x_2, \ldots, x_m$. Weights are defined as $wk_1, wk_2, \quad wk_m$. Then the net input can be written with the equation:

$$v_k = \sum_{j=1}^{m} w_{kj} x_j + b_k \quad (1)$$

$$y_k = \varphi(v_k) \quad (2)$$

where φ (.) is an activation function. Various kinds of activation functions can be used, such as hard limiter, sigmoid, logistic, ReLu, etc. which function to limit the output value to a certain range.

2.3 Backpropagation

The Backpropagation algorithm was first introduced by Rummelhart et all in 1980 to overcome the learning delta rule which cannot be applied to Multi-Layer Perceptron (MLP) due to hidden layers [22].

The basis for calculating changing the weight to error is as follows:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial net_j}\frac{\partial net_j}{\partial w_{ij}} \quad (3)$$

Where E is the error and w_{ij} is the weight that connects the neurons i and j, o_j is the output value of the neuron j while net_j is the weighted number of inputs. So to calculate the change in weights connected to the output layer can be done with the following formula:

$$\frac{\partial E}{\partial w_{ij}} = (t - y)f'(net_j)x_i \quad (4)$$

Where, t is the target output, y is the output of the neural network, f '(net_j) is a derivative of the activation function, while x_i is the output of the neuron to i. In order to facilitate changing the weights on the hidden layer, the sensibility error values are calculated as follows:

$$\delta_j = \frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial net_j} \quad (5)$$

So the error sensibility of hidden neurons can be written as:

$$\delta_i = f'(net_i)\sum_l \delta_l w_{il} \quad (6)$$

That way the weight can be changed using the following formula:

$$\Delta w_{ij} = \mu \delta_j x_i \quad (7)$$

Where μ is the learning rate.

One of the problems faced by neural networks is getting stuck in local minima. To avoid this, neural networks usually use momentum m. With momentum m, the change in weight at time t can be stated as follows:

$$\Delta w_{ij}(t) = \mu \delta_j x_i + m \Delta w_{ij}(t-1) \quad (8)$$

2.4 K-fold cross-validation

The k-fold cross-validation method is used to validate the classifier model. K-Fold Cross Validation divides the set into k parts randomly and independently of each other. As much data (k-1) fold is used to conduct training models while 1 fold is used to do testing. Validation is performed k times until all data in the dataset is tested on the model [2].

III. DISCUSSION

3.1 Research Flow

The flow of research is a description of the research to be conducted, to find out the flow of research can be seen in Figure 2.

3.2 Research Step

3.2.1. Data Collection

In the Twitter data collection stage, the first thing to do is to obtain the Twitter Key API which is obtained from the https://developer.twitter.com/en/apps page. To get the Twitter API Key, you must fill out the API Key submission form first. After successfully registering it will get an API Key that can be used as authentication in Twitter data retrieval.

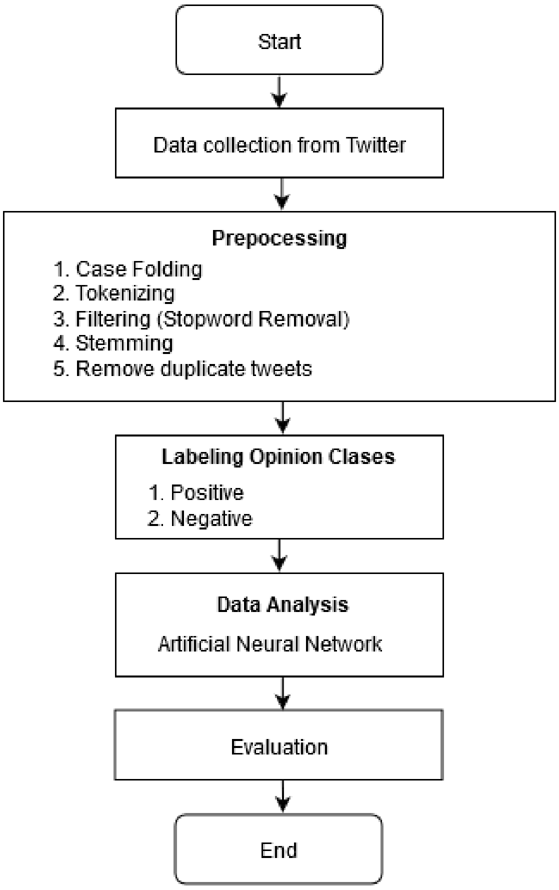
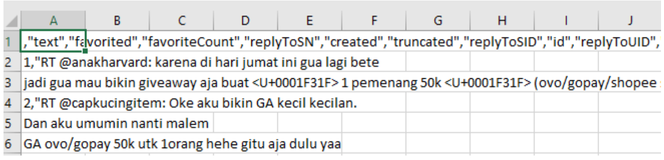


Figure 2. Research Flow

If Twitter Key API authentication is successful, In the process of crawling datasets from Twitter, detailed information about the dataset is very important to ensure valid and representative analysis. The dataset consists of approximately 1.350 tweets each for each digital payment platform, namely Gopay, OVO, LinkAja, Doku, and Jenius. The distribution of positive and negative data from each platform needs to be identified to understand the tendency of public sentiment in more depth. This data was taken over one month, so it reflects the current opinions of Twitter users. The data selection criteria are based on the use of keywords such as "gopay," "ovo," "LinkAja," "Doku," and "Jenius," which represent how often these words appear in conversations among the public, especially Twitter users. Data that is successfully pulled from Twitter will be saved in .csv format. Figure 3 is an example of the results of Twitter data crawling.



	A
1	,"text","favorited","favoriteCount","replyToSN","created","truncated","replyToSID","id","replyToUID",
2	1,"RT @anakharvard: karena di hari jumat ini gua lagi bete
3	jadi gua mau bikin giveaway aja buat <U+0001F31F> 1 pemenang 50k <U+0001F31F> (ovo/gopay/shopee
4	2,"RT @capkucingitem: Oke aku bikin GA kecil kecilan.
5	Dan aku umumin nanti malem
6	GA ovo/gopay 50k utk 1orang hehe gitu aja dulu yaa

Figure 3. Crawling Result

3.2.2. Preprocessing

Data obtained from the crawling process still cannot be processed, because the raw data still contains a lot of noise and is not structured. To eliminate noise in words and convert sentences to be structured and easy to understand, the process of preprocessing is done. The preprocessing stage includes:

a. Case Folding

Case folding is one of the simplest and most effective forms of text preprocessing, although it is often ignored. The purpose of case folding is to change all letters in a document into lowercase letters. Several ways can be used in the case folding step :

1) Change the text to lowercase

The process of converting letters in sentences into lowercase letters is important for search engines. This is also because of the use of capital letters that are inconsistent in their placement, making it difficult for search engines to work. All text, including URLs, mentions (@username), and hashtags, is converted to lowercase to maintain data consistency. The results of this step are shown in table 1.

Table 1. Process Change Text to Lowercase Letters

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	rt @sagitaariuss: "cashbacknya gedean pake ovo apa dana ya?" https://t.co/n3sme8loew ...

2) Delete number

Table 2. The Process of Deleting Numbers

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	rt @sagitaariuss: "cashbacknya gedean pake ovo apa dana ya?" https://t.co/n3smeloew ...

Delete numbers are done in sentiment analysis. What is needed in analyzing sentiments is that the processed data is ensured to be free of various kinds of noise. Removing a number will not change the meaning of a sentence. This process is shown in Table 2.

3) Remove punctuation
In doing sentiment analysis punctuation is also not important and is considered disturbing so it is eliminated, and will not change the meaning of a sentence. Remove punctuation marks like [! ”# $% &’ () * +, -. / :; <=>? @ [\] ^ _ `{|} ~]. The results of this step are shown in Table 3.

Table 3. The Process of Remove Punctuation

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	cashbacknya besar pakai ovo apa dana ya

4) Delete whitespace (blank character)
This stage aims to remove unnecessary whitespace in the text. This process will reduce the irregularity in the data and ensure that relevant words can be processed better.

b. Tokenizing

Tokenizing is the process of separating the text into pieces called tokens for later analysis. This process is done to get words that have value. Tokenizing aims to make it easier to calculate the frequency of occurrence of words in the corpus. This process is seen in Table 4. Tokenizing requires identifying and separating words in sentences, as well as paying attention to Twitter-specific elements such as mentions (@username) and hashtags (#hashtag). URLs can be treated as a separate token or removed depending on the purpose of the analysis.

Table 4. The Process of Tokenizing

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	[‘cashbacknya’, ‘gedean’, ‘pake’, ‘ovo’, ‘apa’, ‘dana’, ‘ya’]

c. Filtering

The filtering stage is to filter out tweet sentences or subtraction of words in the corpus called stopword. Stopword is to eliminate words that have no influence or eliminate words that have no meaning. In this stage, the keywords are also eliminated. The stopword used is from the Literature Library. At this stage, text that is not important for sentiment analysis can be removed, such as stop words (common words that do not provide significant information, for example "is", "the", "and"), numbers, punctuation, and URLs. The results of this stage are shown in table 5.

Table 5. The Process of Filtering

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	[‘cashbacknya’, ‘gedean’, ‘pake’]

d. Stemming

Stemming is used to calculate the index number of a document, but also to group basic words. The stemming process is seen in Table 6.

Table 6. The Process of Stemming

Input	Output
RT @sagitaariuss: "Cashbacknya gedean pake ovo apa dana ya?" https://t.co/N3SmE8loeW ...	[‘cashback’, ‘besar’, ‘pakai’]

e. Remove duplicate tweets

3.2.3. Labeling Opinion Classes

The next stage of the research is the labeling stage of the opinion class, in this stage, each data will be labeled according to the opinion class (positive and negative).

3.2.4. Data Analysis

The next stage is to preprocess data before entering the training and testing phase of data that will be carried out by an artificial neural network (ANN), data validation with k-fold cross-validation. The number of neurons in the input layer of an artificial neural network (ANN) is determined by the number of features extracted through techniques such as TF-IDF, and word embeddings, which represent important information from the text. This adjustment ensures that the input data is maintained without losing detail. The hidden layers are designed with 64 and 32 neurons. The first layer with 64 neurons captures complex patterns from the data, while the second layer with 32 neurons surrounds the representation to minimize overfitting and improve generalization. This configuration is often used because it provides a balance between efficiency and performance, although the final optimization is done through experiments, as explained in Figure 4.

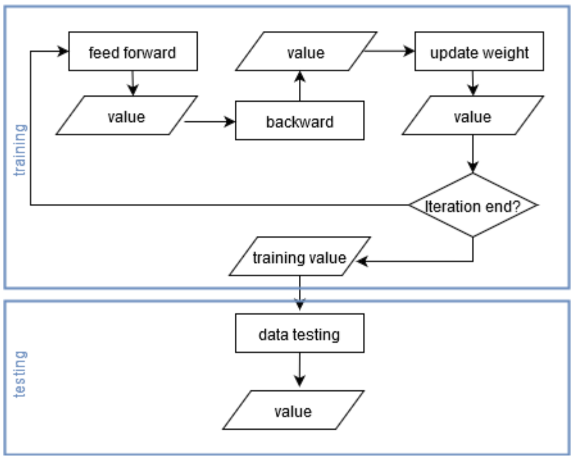


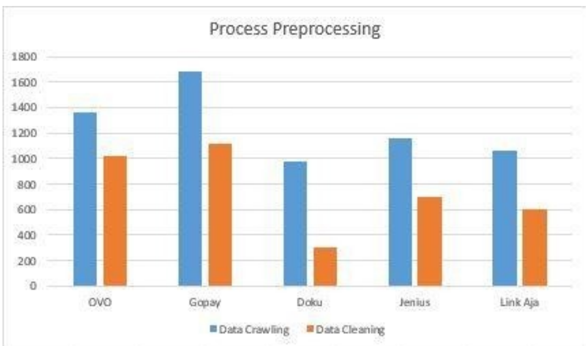
Figure 4. Data Analysis Using ANN

IV. RESULT

4.1 Crawling Data From Twitter

The crawling that the researchers did in this study is looking for data via Twitter social media. The process is carried out to search for keywords namely "gopay", "ovo", "Link Aja", "Doku", and "Jenius" data obtained approximately 6256 data. The data taken in this study is the tweet data of the Indonesian people from November to December 2019. Data successfully retrieved from Twitter will be saved in .csv format. The results of crawling as in the previous figure 3.

The results of the crawling are then processed and suspended by the preprocessing that is requested to remove junk such as emoticons, punctuation, and words that lack meaning such as words and, me, di, too, and others that have been requested in the discussion of the methodology section. Graph 1 is the result of the preprocessing process.

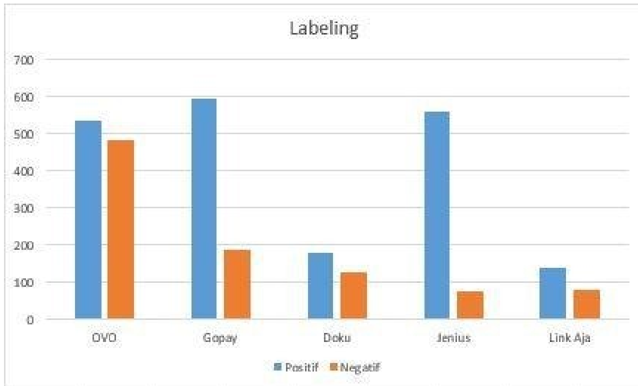


Graph 1. Preprocessing Process Results

In graph 1, the difference in results between crawling data from Twitter and data from preprocessing results, the initial data amounted to 6256 reduced to around 3745. From the graph above the data that experienced the most shrinkage during the cleaning process is Doku with the final dataset number 308. This means that the data that is crawled from Twitter with the Doku keyword is too much garbage data. While the best crawling results are when searching for OVO.

4.2 Feature Extraction

The dataset is clean and then goes through building the dataset features and testing, namely the labeling process by the variables determined by the researcher, in this study the researchers labeled 3 variables: positive, negative, and neutral. This labeling process researchers used a manual process by labeling the data one by one. Labeling results for five digital payments are presented in the graph below.



Graph 2. Results of the Labeling Process

From the labeling, we will get the frequency of occurrence of each word that is used as a key. In this case, each digital payment will be analyzed for the number of times the keywords can be used as a reference for ANN classification. Below is an example of the results of LinkAja and Jenius frequency trams.

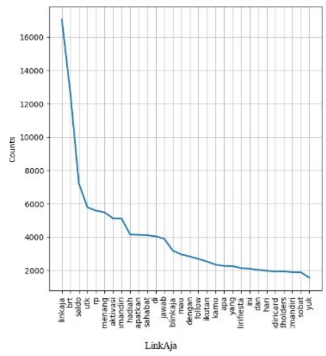


Figure 5a. Term Frequency (TF)

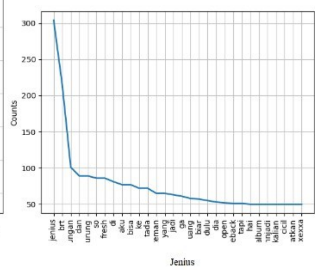


Figure 5b. Term Frequency (TF)

From Figure 5a and 5b Term Frequency (TF) It can be seen that the appearance of the word Link is more than 1600 times and for keywords the term Jenius has 300 times more frequency.

4.3 Modeling

The next stage is building the architecture or modeling process. In this case, the experiment uses 3 input layer nodes and 2 output layers. After adjusting the hidden layer repeatedly, it is concluded that the hidden layer used is 30 neurons. The results of the experiment are presented in the table below:

Table 6. Architecture or Modeling

input				Target	Output	
x1	**x2**	**x3**	**bias**	**y**	**w1**	**w2**
1	1	0	1	1	1	1
1	0	1	1	1	1	1
0	1	1	1	1	1	1
1	0	0	1	0	1	0
0	0	1	1	0	0	1
0	1	0	1	0	1	1

The results of the modeling will be used later to process the classification and produce negative and positive sentiments. Experiments carried out at this stage are 5 kinds of architecture. Eporch itself is obtained from how many nodes will be used for sharing the testing process data. Backpropagation will process all dataset features to get a classification by repeating the process until it gets the best accuracy.

Table 5. Backpropagation

Arc	Training			Testing	
	eporch	**time**	**MSE**	**MSE**	**Accuracy**
1	5	02:09	0.00089	0.00077	98.89%
2	8	01:32	0.00054	0.00083	87.98%
3	11	00:48	0.00067	0.00067	87.86%
4	17	00:57	0.00074	0.00081	89.89%
5	**15**	**00:43**	**0.00098**	**0.00074**	**99.23%**

With five backpropagation experiments, it was found that the best accuracy was in the 5th experiment with the epoch loading of 15 nodes, with a long process of 00.43. The MSE value in the fifth experiment is greater than previous experiments but it gets the best accuracy rate which is 99.23%.

4.4. Evaluation

To get the best results, the validation stage is carried out using the k-fold validation method. The trick is to divide all the datasets into 5 nodes. Training data will be compared with testing data. From this comparison, we can determine which dataset schema is higher than the results of the other schemes.

Table 6. K-Fold Validation

Scema	Amount of Data		Accuracy
	Training	**Testing**	
1	3496	249	87.89%
2	3496	249	89.56%
3	3496	249	89.78%
4	3496	249	90.34%
5	**3496**	**249**	**92.89%**

4.5 Results of Sentiment Analyis

After all the processes that have been carried out in the previous stages, the results of sentiment analysis on digital payment in Indonesia are presented in Table 6.

Table 6. The Results of Sentiment Analysis

Digital Payment	Sentimen	Score	Sentimen	Score
OVO	Positive	58%	Negative	42%
Gopay	**Positive**	**72%**	Negative	28%
Doku	Positive	56%	Negative	44%
Jenius	Positive	69%	Negative	31%
LinkAja	Positive	43%	**Negative**	**57%**

V. CONCLUSION

5.1 Conclusion

The conclusions obtained from the results of this study are as follows:

a. From the table of analysis sentiment results, we can see that each digital payment receives various positive and negative sentiments.
b. The most positive sentiment results were received by e-money Gopay with the acquisition of 72% of all datasets.
c. For digital payments that get the most negative sentiment values, namely LinkAja at 57%.

5.2 Suggestion

This study is expected to provide insight for further research to explore the factors that influence sentiment towards digital payments. We are also aware of the shortcomings in the crawling data cleaning process, so in future research, it is recommended to update the dataset to make it better. In addition, it is important to use additional evaluation metrics such as precision, recall, F1-score, and confusion matrix. However, the results of this study remain relevant and accountable, because the use of the Artificial Neural Networks (ANN) model in sentiment analysis with case studies of Gopay, OVO, Doku, Jenius, and LinkAja showed good performance.

DAFTAR PUSTAKA

[1] Budi Maryanto, "Big Data dan Pemanfaatannya Dalam Berbagai Sektor," *Media Inform.*, vol. Vol. 16 No, no. 2, p. 14, 2017.

[2] S. Yadav and S. Shukla, "Analysis of k-Fold Cross- Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, no. Cv, pp. 78–83, 2016, doi: 10.1109/IACC.2016.25.

[3] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," *IEEE Access*, vol. 6, no. c, pp. 23253– 23260, 2018, doi: 10.1109/ACCESS.2017.2776930.

[4] M. Lichouri, M. Abbas, A. A. Freihat, and D. E. H. Megtouf, "Word-Level vs Sentence-Level Language Identification: Application to Algerian and Arabic Dialects," *Procedia Comput. Sci.*, vol. 142, no. November, pp. 246–253, 2018, doi: 10.1016/j.procs.2018.10.484.

[5] H. Kim and Y.-S. Jeong, "Sentiment Classification Using Convolutional Neural Networks", doi: 10.3390/app9112347.

[6] Y. Bengio, "Learning Deep Architectures for AI." Now Publishers Inc, 2009.

[7] Jones and B. Learning, *Artificial Intelligence: A Systems Approach: A Systems Approach*. Jones & Bartlett Learning, 2015. [Online]. Available: https://books.google.co.id/books?id=ekUHwvRP7n UC&dq=best+first+search&hl=id&source=gbs_navl inks_s

[8] T. Chen and S. Zhong, "Privacy-preserving backpropagation neural network learning," *IEEE Trans. Neural Networks*, vol. 20, no. 10, pp. 1554–1564, 2009, doi: 10.1109/TNN.2009.2026902.
X. Zhang, J. Zhao, and Y. LeCun, "Character-level Convolutional Networks for Text Classification," in *Advances in Neural Information Processing Systems*, 2015, vol. 28. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/250cf8b51c773f3f8dc8b4be867a9a02-Paper.pdf

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, 2013.

[10] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.

[11] I. Sutskever, "Sequence to Sequence Learning with Neural Networks," *arXiv Prepr. arXiv1409.3215*, 2014.

[12] H. Yousuf, M. Lahzi, S. A. Salloum, and K. Shaalan, "A systematic review on sequence-to-sequence learning with neural network and its models.," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, 2021.

[13] P. Liu, S. Chang, X. Huang, J. Tang, and J. C. K. Cheung, "Contextualized non-local neural networks for sequence learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, no. 01, pp. 6762–6769.

[14] J. Shin, Y. Kim, S. Yoon, and K. Jung, "Contextual- CNN: A novel architecture capturing unified meaning for sentence classification," in *2018 IEEE international conference on big data and smart computing (BigComp)*, 2018, pp. 491–494.

[15] B. S. Atikah, Y. Sibaroni, and D. Puspandari, "Impact of Feature Extraction on Multi-Aspect Sentiment Classification for Livin'byMandiri Using BiLSTM," *J. La Multiapp*, vol. 5, no. 5, pp. 524–539, 2024.

[16] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[17] A. Mathew, P. Amudha, and S. Sivakumari, "Deep learning techniques: an overview," *Adv. Mach. Learn. Technol. Appl. Proc. AMLTA 2020*, pp. 599–608, 2021.

[18] L. M. Qaisi and I. Aljarah, "A twitter sentiment analysis for cloud providers: A case study of Azure vs. AWS," *Proceedings - CSIT 2016: 2016 7th International Conference on Computer Science and Information Technology*. 2016. doi: 10.1109/CSIT.2016.7549473.

[19] S. Kumar and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets," *J. Big Data*, vol. 6, no. 1, pp. 1–16, 2019, doi: 10.1186/s40537-019-0224-1.

[20] S. Nooriansyah, C. Fatichah, and K. Sambodho, "Analisis Kinerja Metode Artificial Neural Network Dan Support Vector Regression Untuk Prediksi Significant Wave Height," *SCAN - J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, 2018, doi: 10.33005/scan.v13i1.1054

[21] Simon Haykin, "Neural Networks - A Comprehensive Foundation - Simon Haykin.pdf." p. 823, 2005.