Sentiment Analysis Model Development on E-Money Service Complaints

Imam Tahyudin [1], Andhika Rafi Hananto [1], Silvia Anggun Rahayu [1], Rayinda Maya Anjani [1], Ade Nurhopipah [2]

[1] *Department of Information System, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia*
[2] *Department of Informatics, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia*

Abstract – **Technology provides various conveniences for users in many aspects, such as in the commercial business world with the development of financial technology or fintech. In Indonesia, there are fintech services with the most users, namely OVO and DANA. Various kinds of criticisms and suggestions related to deficiencies and weaknesses in OVO and DANA services were submitted by users via social media Twitter. This study aims to analyze and develop a sentiment identification model for user complaints against OVO and DANA services on Twitter. This study uses the four classification algorithm methods: Random Forest, K-NN, Decision Tree, and Naïve Bayes Classifier. The data from social media Twitter is taken from user tweets regarding criticism and complaints on OVO and DANA services. The data obtained was 5,000 instances consisting of 20% positive and 80% negative reviews. Accuracy values represent the study results with 87% for Naive Bayes, 86% for Decision Tree, 91% for K-NN, and 87% for Random Forest. Based on these results, it can be concluded that testing of sentiment analysis on user complaints on OVO and DANA services using the K-NN algorithm is superior to the other algorithms. This research results in a system that combines the four prediction models. The evaluation shows that the developed model can accurately classify user sentiment.**

Keywords – **Sentiment analysis, K-NN, naïve Bayes, decision tree, random forest.**

1. Introduction

Along with the growth of technology and Internet networks, the use of technology and the Internet has helped various activities in everyday life. In addition, the development of increasingly modern technology raises many alternatives in various sectors, one of which is the field of financial transactions. Technological developments in financial transactions can be seen from the means of payment used [1].

OVO and DANA are financial technology service providers established in 2017 and 2018. One of the communication channels used by OVO and DANA with customers is through social media, Twitter. Along with the increase in social media users in Indonesia, social media is used as one of the company's communication channels with customers. Users use Twitter as an online news platform and social media to communicate, express themselves, and give their views. On Twitter, the user posts short sentences with a maximum of 280 characters, called "tweets" [2], [3]. This limitation in conveying user opinions on tweets or User Generated Content distinguishes Twitter from other social media, thus forming unique characteristics that cause users to be asked to explain something quickly, briefly, and concisely so that it is more quickly understood. Presentation of data on social media Twitter is more informative than Facebook and Instagram [4]. Currently, Twitter users are talking about OVO and DANA through tweets to express their opinions. One of the analysis processes that can be used is text data mining, which searches for information automatically from different text sources. The data source on Twitter is tweets from users. Text analysis to get information related to positive or negative portrayal of an opinion is known as sentiment analysis.

The increase in OVO and DANA users makes OVO and DANA one of Indonesia's largest fintech payment providers. Along with this fact, Indonesians often use Twitter to express opinions or complain about the quality of their services.

In addition, Twitter data nowadays can be obtained easily, so the phenomenon is attractive to study [5]. Utilizing Twitter data submitted for OVO and DANA has led to the phenomenon of big data. This collection of large and complex datasets is too challenging to analyze in traditional ways.

With the increasing number of fintech service providers in Indonesia, users are also becoming more selective in choosing which fintech services to use. Therefore, the quality of service is one of the essential factors that must be analyzed by utilizing data mining involving methods from several fields, for instance, statistics, database systems, and machine learning [6]. This research utilizes data mining to perform data collection, data processing and sentiment analysis to classify sentiments into positive and negative classes using the K-NN (K-Nearest Neighbor) method and other algorithms.

Sentiment analysis is a process carried out to analyze or identify opinions, sentiments, judgments, and emotions from a person's statement on a domain. In the process, sentiment analysis can classify text based on several groups or sentences, such as positive and negative sentiments [7]. The methods used in this research are several classifiers, including the K-NN. K-NN is a data classification method based on test data's distance or similarity value with training data. K-NN is an appropriate method for classifying data into predetermined classes based on the closest distance or level of similarity of the data with existing datasets and training data.

From this explanation, this research has three goals. The first is to test the K-NN and several other algorithms in categorizing the number of positive comments from Twitter social media users on E-money in OVO and DANA. The second is to find out the performance differences of several other algorithms in the classification. The last aim is to develop a model for handling the dataset and knowing the results of public sentiment to determine the quality of service provided by OVO and DANA. The result can be used to find the dimensions of service quality that need to be improved or receive more attention from the data obtained from user tweets through tweets on Twitter social media. Finally, it can help companies improve the quality of service provided to users.

2. Literature Review

Data mining aims to process datasets to find useful hidden patterns. The data processing results can be useful in making future decisions. Data mining is also known as pattern recognition [1], [8]. This study generally discusses several methods, for instance, regression, classification, clustering, variable selection, and market basket analysis [2], [9]. Data mining can be applied in multiple fields. For instance, it can be utilized in business areas to develop more effective strategies and leverage resources more optimally and insightfully [10].

Data mining algorithms are also performed to address the sentiment analysis problem. Sentiment analysis is the intersection fields of Natural Language Processing (NLP), Text Mining, and Linguistic Computing. This task analyses people's opinions, sentiments, attitudes, judgments, and emotions about a particular topic, activity, organization, individual, product, or service. Sentiment analysis can produce two values, namely positive and negative sentiment, or can be deepened to find who or which group is the source of positive and negative views [11].

Research conducted by [12] applying sentiment analysis through Instagram data to determine culinary tourism's reputation in Bandung using Naive Bayes classification. This study uses all 864 user comment data. The experiment was implemented with three data split ratios, namely 70:30, 75:25, and 80:20, using three classifiers: Naïve Bayes, Decision Tree and K-NN. The result concluded that the Naïve Bayes reached the highest performance by obtaining an accuracy of 86.87%, which was tested using RapidMiner tools. This study got satisfactory evaluation results, namely Precision of 93% and Recall of 57.50%, resulting in an F1-Measure of 75.39%.

Research performed by [11] entitled Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers explores big data by applying an Artificial Intelligence (AI) machine learning approach. This research results in a system that analyzes the business level, crime rate, education level, and health level in Malaysia, Singapore, Vietnam, and Myanmar. This application is intended to measure the behavioral impact of using social media on ASE citizens and compare the rate of change in these sectors.

Research by [13] discusses Sentiment Analysis for Beauty Product Reviews using K-NN and Information Gain. This research was applied to a dataset taken from the female daily website. In feature selection, an IG value of more than 0.5 has a high level of relevance to the class. The optimal k value is 23. So, from the scenarios carried out, the highest accuracy was obtained in the dataset that applied stemming, normalization, IG with a threshold of 0.5, and k = 23 with an accuracy of 74.21%.

A study by [14] discusses the KNN Method on the Sentiment of Android Game product review Analysis. This study used log data from students to compare Decision Trees, Neural Networks, Naive Bayes, and Support Vector Machine (SVM). The processed review data can be appropriately classified

into positive and negative forms. This study showed that the K-NN accuracy on product review data on the App Store for Android Games reaches 75.50%, and the AUC value is 0.825.

Research [15] conducted sentiment analysis regarding the homecoming ban policy using the K-NN and applied the Cross-Industry Standard Process for Data Mining (CRIPS-DM) procedure. In this study, the data used was 4,799 tweets taken from Twitter social media on April 04, 2021 - May 17, 2021, with positive sentiment totaling 834 tweets and 3,965 tweets of negative sentiment. This research shows that K-NN can be implemented well because it achieves an accuracy of 86.67% with a recall of 39.52%, precision of 70.97% and specificity of 96.60%. This experiment used a split data ratio of 80 for training and 20 for testing with k=3. One could argue that the K-NN algorithm can classify data correctly and well.

3. Research Methods

The research was carried out for sentiment analysis using the K-NN Algorithm and several other algorithms on the DANA and OVO applications using the Jupiter Notebook to measure the study's accuracy. To conduct experiments, researchers used review comment data on Twitter social media. The data was 5,000, consisting of 20% positive and 80% negative reviews. The text processing implemented Tokenize, Transform Case, and Remove Stop word based on a dictionary. Testing is done by algorithms, which results in a confusion matrix. Next, the accuracy value, the Curve of Receiver Operating Characteristic (ROC), and the AUC (Area Under Curve) are obtained.

This research consists of several stages: dataset collection, preprocessing, model training, and result evaluation. Figure 1 describes the steps of the research method.

Figure 1. Research method

1) *Data collection:* Figure 1 describes that the study begins by reviewing Twitter user comments or posts. The data used are user reviews consisting of 20% positive and 80% negative reviews. From this stage, selected and classified negative and positive reviews have resulted.
2) *Dataset:* The collected datasets will be studied by training and testing (supervised machine learning) through a data preprocessing process to classify

which reviews represent positive or negative sentiments. This research will use a comparison between the text preprocessing performed and implemented on each dataset.

3) *Preprocessing Data:* Preprocessing is a stage for processing newly obtained raw words. In this process, filtering is carried out related to existing data to make a data match in the machine learning algorithms application. In the preprocessing process, there are three stages carried out including:
   a) *Tokenization* is a process that breaks a sentence into parts or words [16]. The token is defined as an instance of a character's sequence grouped as a functional semantic unit.
   b) *Cleansing* is a step where unneeded characters and punctuation in the text are removed, for example, exclamation marks, question marks, commas, and periods.
   c) *Filtering* removes words that appear in large numbers but are considered meaningless. These words are called stopwords. Filtering enables users to focus on other unique words that are much more influential [17].
4) *Training and testing data:* This research uses Python with Jupiter Notebook software for data training and testing.
5) *Performance evaluation:* The next step is to evaluate some of the performance by comparing the four algorithms: K-NN, Decision Tree, Random Forest, and Naive Bayes.
6) *Model testing:* The last step is the process of testing the model that has been developed with the four algorithms that have been developed.

4. Result and Discussion

This research contains three outline results: dataset evaluation, model development, and model evaluation. The data evaluation consists of the data analysis and the text preprocessing result. Next, in the model development part, the performance of each model was investigated. Finally, in the model evaluation section, sentiment analysis was implemented based on combined predictions of all models.

4.1. Dataset Evaluation

At the initial stage, this research used the data composition as seen in Table 2, with 710 positive and 4463 negative reviews. This data comparison can be categorized as unbalanced, seeing that the positive data is only less than 25% of the negative data. However, this research aims to develop a model for categorizing negative sentiment on complaint sentiment. Therefore, more dominant negative data is needed.

Table 1. Sentiment data comparison

Sentiment Comparison	
Positive	Negative
710	4463

After the dataset has been identified, the next step is preprocessing. The preprocessing stage is the normalization, stemming, and stopword removal of the data. The purpose of some of these processes is to simplify the model, and the computation of the data becomes easier. For preprocessing, what needs to be removed are links (www or https), hashtags, symbols, and emoticons. To see the data before and after the preprocessing process, see Figures 2 and 3.

Figure 2. Word cloud before data cleaning

Figure 3. Word cloud after data cleaning

The next stage is to display the most frequent words in the dataset and use them as a reference to see the results of data preprocessing. Figure 4 illustrates the popular terms in the dataset.

Figure 4. The top 10 frequent word

4.2. Model Development

This research uses four algorithms commonly used in classifying data: K-NN, random forest, Decision Tree, and Naive Bayes. These four algorithms will be used in the model development to identify the positive and negative values of the text jointly. Figure

5 explains how the accuracy performance of the four algorithms used in this study. The KNN algorithm excels with an accuracy value of 91.8% compared to Random Forest, Decision Tree, and Naive Bayes.

Figure 5. Accuracy comparation

(a)

(b)

(c)

(d)

Figure 6. ROC curve of (a) K-NN, (b) Random Forest,

(c) Decision Tree, and (d) Naïve Bayes

Figure 6 describes the ROC Curves, representing how the algorithm performs in the developed models. The results in the ROC picture show that Random Forest has an AUC higher than other algorithms. The conclusion of the different evaluation results of the four algorithms can be seen in Table 2.

Table 2. Model results

No	Model	Accuracy	Precision	Recall
1	K-Nearest Neighbors	0.91	0.47	0.25
2	Random Forest	0.87	0.00	0.00
3	Decision Tree	0.86	0.43	0.30
4	Naive Bayes	0.87	0.44	0.06

Table 2 presents the results of the accuracy, precision, recall, and ROC values obtained from the four classification algorithm methods. The best accuracy algorithm is K-Nearest Neighbors, 91%. After that, random forest, naïve Bayes, and decision tree, respectively.

However, the precision and recall values are not so good. It is indicated because of the imbalanced dataset.

4.3. Model Evaluation and Testing

The final stage of this research uses the model that has been developed to test sentiment sentences that are not included in the dataset and classify them. Tables 3 to 5 are the results of testing the model on three sentiment sentence samples. The model identifies the results by matching the prediction with the target. If the words in a sentence are dominantly identified as positive, then the algorithm that detects it will score 1. The four algorithms work together to identify sentiment per sentence. The more algorithms that identify sentiment as positive, the sentence or sentiment will be categorized as a positive review. If the voting between the four is balanced, then the system will assess the sentence with a neutral status. Likewise, if most models identify the sentence as having negative sentiments, the system will mark them as negative reviews.

Table 3. Model simulation positive

Preprocessed Review: ['membantu bangek gak punya reken atm transaksinya mudah semoga aman indomaret adakan penarikan uang lewat dana jadi jauh alfa indomaret ambil']			
Prediction by Classifiers:			
Naive Bayes	[1]		
Decision Tree	[1]		
K-Nearest Neighbors	[1]		
Random Forest	[0]		
Comment impact	(0: Negative | 4: Positive)	:	3
TextBlob Sentiment Polarity (UnPreprocessed)		:	0.0

Table 4. Model simulation neutral

Preprocessed Review: ['cs cepat balas email terkait pengaduan saldo ovo cash top saldo top tf bank cimb niaga notif top berhasil saldo estimasi pemotong biaya top']			
Prediction by Classifiers:			
Naive Bayes	[1]		
Decision Tree	[1]		
K-Nearest Neighbors	[0]		
Random Forest	[0]		
Comment impact	(0: Negative | 4: Positive)	:	2
TextBlob Sentiment Polarity (UnPreprocessed)		:	0.5

Table 5. Model simulation negative

Preprocessed Review: ['aplikasi lambat systemnya payah beli paket data pake ovo saldonya berkurang paket datanya tidak terisi']			
Prediction by Classifiers:			
Naive Bayes	[0]		
Decision Tree	[0]		
K-Nearest Neighbors	[0]		
Random Forest	[0]		
Comment impact	**(0: Negative | 4: Positive)**	**:**	**0**
TextBlob Sentiment Polarity (UnPreprocessed)		:	0.0

5. Conclusion

This research has been done using the user comments data on Twitter, with 20% positive reviews and 80% negative reviews on OVO and DANA services. This research aims to analyse and develop a sentiment identification model of user complaints about OVO and DANA services on user tweets using four classification algorithms: Random Forest, K-NN, Decision Tree, and Naïve Bayes. The results of this study have proven that the developed model is accurate in classifying user sentiment, with an average accuracy of 85%. The highest accuracy resulting from the K-NN algorithm reaches 91.80%. Based on these values, testing sentiment analysis of user complaints about OVO and Dana's services using the K-NN is superior to the other classifier algorithms. Future research can be conducted to create a new system for obtaining better accuracy, such as combining deep learning algorithms. Furthermore, improving the precision and recall values is necessary to solve the imbalanced dataset problem.

References:

[1]. Atalaya, O. F. C., *et al*. (2022). K-NN supervised learning algorithm in the predictive analysis of the quality of the university administrative service in the virtual environment. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(1), 521–528.

[2]. Raturi, R. (2018). Machine Learning Implementation for Business Development in Real Time Sector. *International Journal of Pure and Applied Mathematics*, *119*(15), 1289–1300.

[3]. Shallal, Q. M., Hussien, Z. A., & Abbood, A. A. (2020). Method to implement K-NN machine learning to classify data privacy in IoT environment. *Indonesian Journal of Electrical Engineering and Computer Science*, *20*(2), 985–990.

[4]. Bose, D. (2020). Movie recommendation system with Collaborative Filtering using K-NN. *Big Data Analiytics Report SPRING*, *April*, 0–4.

[5]. Jalla, H., & Girija, P. N. (2019). *Privacy-Preserving K-NN Classi fi cation Using Vector Operations*. Springer Singapore.

[6]. Alamsyah, A., Sudrajat, F. D., & Irawan, H. (2017). Property Business Classification Model Based on Indonesia E-Commerce Data. *The 8th International Conference on Sustainable Collaboration in Business, Information, and Innovation*, *arXiv:2102*.

[7]. Ehsani, R., & Drabløs, F. (2020). Robust Distance Measures for kNN Classification of Cancer Data. *Cancer Informatics*, *19*, 1–9.

[8]. Al-Khowarizmi, Syah, R., Nasution, M. K. M., & Elveny, M. (2021). Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor. *International Journal of Electrical and Computer Engineering*, *11*(3), 2696–2703.

[9]. Virjonen, P., Nevalainen, P., Pahikkala, T., & Heikkonen, J. (2018). Ship Movement Prediction Using k-NN Method. *Proceedings - 2018 Baltic Geodetic Congress, BGC-Geomatics 2018*, *June*, 304–309.

[10]. Patel, H., & Verma, D. (2021). Data Classification in Web Usage Mining using SVM , SVR and K-NN. *Journal of Innovative Engineering and Research (JIER)*, *2*(October), 18–22.

[11]. Naw, N. (2018). Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers. *International Journal of Scientific and Research Publications (IJSRP)*, *8*(10), 407–411.

[12]. Finandra, S., Murahartawaty, & Hamami, F. (2021). Penerapan Analisis Sentimen Melalui Data Instagram Untuk Mengetahui Reputasi Wisata Kuliner Di Kota Bandung Menggunakanmetode Klasifikasi Naïve Bayes. *E-Proceeding of Engineering*, *8*(5), 9382–9391.

[13]. Prastika S., E. Y., Faraby, S. Al, & Purbolaksono, M. D. (2021). Analisis Sentimen pada Ulasan Produk Kecantikan Menggunakan K-Nearest Neighbor dan Information Gain. *EProceedings of Engineering*, *8*(5), 10091–10105.

[14]. Sahara, S., & Agung Permana, R. (2022). Metode Knn Pada Sentiment Analisis Review Produk Game Android. *Ijns.Org Indonesian Journal on Networking and Security*, *11*(2), 2302–5700.

[15]. Lestari, D. A., & Mahdiana, D. (2021). Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021. *Informatik : Jurnal Ilmu Komputer*, *17*(2), 123–131.

[16]. Purnawansyah, Haviluddin, B., Fanany, A., Gafar, O., & Tahyudin, I. (2018). K-NN supervised learning algorithm in the predictive analysis of the quality of the university administrative service in the virtual environment. *Proceedings of the Eleventh International Conference on Management Science and Engineering Management*, *2*.

[17]. Tahyudin, I., Fitriyani, N. A., Dewiyanti, N., Amin, M. S., Firdaus, M. Y., & Utama, F. P. N. (2015). Inovasi Promosi Obyek Wisata Menggunakan Teknologi Augmented Teality (AR) Melalui Layar berbasis Android. *Jurnal Telematika*, *8*(1), 1–13.