SENTIMENT ANALYSIS AGAINST THE DANA E-WALLET ON GOOGLE PLAY REVIEWS USING THE K-NEAREST NEIGHBOR ALGORITHM

Siti Masturoh[1*)] ; Achmad Baroqah Pohan[2];

[1]Program Studi Sistem Informasi
Universitas Nusa Mandiri
www.nusamandiri.ac.id
siti.uro@nusamandiri.ac.id

[2]Program Studi Sistem Informasi
Fakultas Teknik & Informatika
Universitas Bina Sarana Informatika
www.bsi.ac.id
achmad.abq@bsi.ac.id

(*) Corresponding Author

Abstract—DANA e-Wallet or digital wallet application can be downloaded on the Android platform via Google Play, and google play itself provides a review column. The public will usually see reviews on Google Play before they download an application because the information obtained through these reviews is considered effective in providing information, problems regarding reviews or sentiment analysis of the application must be processed using text mining. Text mining in this study uses k-nearest neighbor by testing 3 classes based on star rating, the first class consists of 1-5 stars, the second class consists of (1 & 5 stars, 3rd class consists of labeling stars (1 & 2) negative label, 3 neutral labels, as well as 4 & 5 stars positive labels) and testing the value of k 1-10 so that the highest accuracy value is obtained with class 2 (1 star and 5 stars) and the best test at k 1 value is obtained the accuracy result of 86.64%

Keywords: E-wallet, DANA, K-Nearest Neighbor

Abstrak—*E-Wallet DANA atau aplikasi dompet digital bisa di unduh di platform andoid melalui google play, dan google play sendiri menyediakan kolom ulasan. Masyarakat biasanya akan melihat ulasan pada google play sebelum mereka mengunduh sebuah aplikasi karena informasi yang didapat melalui ulasan tersebut dianggap efektif memberikan sebuah informasi, permasalahan mengenai ulasan atau analisa sentimen terhadap aplikasi harus diolah menggunakan teks mining. Text mining pada penelitian ini menggunakan k-nearest neighbor dengan pengujian 3 kelas berdasarkan pemberian bintang, kelas pertama terdiri dari bintang 1-5, kelas ke 2 terdiri dari (bintang 1 & 5, kelas ke 3 terdiri dari pemberian label pada bintang (1 &2 label negatif, 3 label netral, serta bintang 4 & 5 label postif) dan pengujian nilai k 1-10, sehingga didapatkan nilai accuracy tertinggi dengan kelas ke 2 (bintang 1 dan bintang 5) dan pengujian terbaik pada nilai k 1 didapatkan hasil accuracy sebesar 86.64%*

Kata Kunci*: E-wallet, DANA, K-Nearest Neighbor*

INTRODUCTION

Digital wallets (E-wallets) are currently developing so rapidly, Moreover, there are many attractive offers so that people have started to switch to e-wallets to carry out financial transactions. Digital wallets (e-wallets) can be used to make payments for various transactions that are already available (Puspita, 2019). Users can make non-cash financial transactions without cards, both online and offline (Mahendrajaya et al., 2019). The application of the non-cash system in Indonesia is a direct influence of a development in financial technology (Fintech) (Kesumastuti, 2020).

The use of e-wallets is considered to have a positive impact because it is more efficient and results in reduced cash circulation (Inggiharti, 2020). In Indonesia, there are many digital wallet services (E-wallet) such as OVO, DANA, Go-Pay, LinkAja, T-cash, ShopeePay and so on. In addition, e-wallets often offer many attractive discounts for making transactions so that it becomes the trigger for the community to switch to choosing to make payments via e-wallets. (Kusnawan et al., 2019). Payment of financial transactions via E-wallets is also considered to be one of the solutions during the Covid-19 pandemic to reduce the spread of the virus (Aulia, 2020). From several e-wallet products, the researcher took a case study, namely the DANA product.

On the Android platform, you can download electronic wallet applications such as DANA

through Google Play. Google Play is equipped with a comment function so that users can provide comments on users' satisfaction with the apps they downloaded (Aaputra et al., 2019). The purpose is to evaluate the application performance of the creator of the application so that improvements can be made (Gunawan et al., 2017). Application users usually they often see reviews of an application before they download the application (Aaputra et al., 2019) Therefore, a technique for monitoring and organizing is needed that can be used to process text data, which is commonly called text mining, which is one of the fields of Natural Language Processing. (NLP) (Salam et al., 2018).

Sentiment analysis or opinion mining is a computational study of the opinions, behaviors, and emotions of entities. These entities can describe a person's behavior in the form of information about events and topics. Sentiment analysis is the process of understanding, extracting and processing text data to obtain the emotional information contained in the opinion sentence (Buntoro, 2017). Sentiment analysis can also be performed to determine opinions or comments on an entity, whether it has a positive or negative trend, so that it can be used as a reference for improving services and improving product quality. (Adhine Salsabila et al., 2019)

This is not the first time that the research on sentiment analysis of electronic wallets has been carried out. Other researchers have done the same thing. This is described in Table 1 of "Peper Reference". The content is as follows:

Table.1 Peper Reference

Research Title	Research Methods & Results
Gopay User Sentiment Analysis Using the Lexicon Based Method and Support Vector Machine	Use dictionary-based methods and support vector machines. There are as many as 1,210 data on social media Twitter, and the results of using Lexicon-based tags are 923 (positive) and 287 (negative) respectively. The results obtained by the SVM method classification The yield using linear core is 98.17% The polynomial kernel is 84.38%
E-Wallet Sentiment Analysis on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization	Using the naive Bayes algorithm based on particle swarm optimization can get 83.60% cross-validation and AUC value of 0.801, while the cross-validation value without particle swarm optimization is 82.30% and the AUC value is 0.780.
Sentiment Analysis Of Digital Wallet Service Users Using Naive Bayes Classifier And Particle Swarm Optimization	Using the naive Bayes algorithm based on particle swarm optimization, by checking DANA and iSaku, using confusion matrix and ROC curve test results show that the use of PSO can improve accuracy, in which the value of DANA digital wallet increased from 60.00% to 91.67% The iSaku digital wallet increased from 53.23% to 85.00%
Analysis of Electronic Wallet Sentiment on Twitter Social Media Using the Naive Bayes Classifier	Using the naive Bayesian classifier to use the hashtag #Gopay #OVO #dana on Twitter as search data, the results of the GoPay research got higher positive reviews from Twitter users, which were 46.67%, 37.50% of funds and 16.67% of OVO.
E-Wallet Sentiment Analysis Using Naive Bayes And Support Vector Machine Algorithm	Use naive Bayesian methods and support vector machines. The data used are Google Play reviews on OVO and DANA. The results show that by measuring the accuracy of use, the community needs OVO the most. For the SVM algorithm, the confusion matrix reaches 91.00%. ROC curve shows the best AUC result 0.986

Source : (Masturoh & Pohan, 2021)

Based on Table 1, We can see the research using Lexicon-based methods and support vector machines to analyze GoPay user sentiment. This research explains the sentiment analysis of comments on Twitter on the GO-PAY app, which is part of the Gojek app and is in high demand. There are as many as 1210 data obtained on social media twitter, and the results of using Lexicon-based tags are 923 (positive) and 287 (negative) respectively. The results obtained by the SVM classification method using linear kernels are 98.17% and 84.38% for polynomial kernels, respectively (Mahendrajaya et al., 2019). In a study called

"Google Play E-Wallet Sentiment Analysis Using Naive Bayes Algorithm Based on Particle Swarm Optimization", a case study of Google Play's OVO sentiment was carried out, and the result was a cross-validation rate of 83.60% and AUC value. Is 0.801 without using particle swarm optimization, the cross-validation value is 82.30%, and the AUC value is 0.780 (Aaputra et al., 2019).

Subsequent research called Naive Bayes classifier service user sentiment analysis and particle swarm optimization, the method is to study DANA and iSaku by studying 490 Tweet data. The test results using the confusion matrix and ROC curve show that the use of PSO can improve accuracy, where the proportion of DANA digital wallets increased from 60.00% to 91.67%, and the proportion of iSaku digital wallets increased from 53.23% to 85.00% (Cahyani & Mardiana, 2020). The next research entitled Electronic Wallet Sentiment Analysis on Twitter Social Media Using the Naive Bayes Classifier with search data taken 100 data on Twitter with the hashtag #Go-Pay #OVO #DANA and the results of GoPay research were more positively rated by Twitter users, namely 46.67%, then Funds of 37.50% and OVO 16.67% (Putra et al., 2020). The data algorithm used in the next research of e-wallet sentiment analysis using Naive Bayes and support vector machines is a review of Google Play on OVO and DANA. The results show that by measuring the SVM algorithm using the confusion matrix achieve 91.00% accuracy, the community needs OVO most. ROC curve shows the best AUC result of 0.986(Kristiyanti et al., 2020).

Therefore, in this study, we used the k-nearest neighbor (KNN) algorithm to analyze the sentiment of the DANA e-wallet application in the Google Play reviews based on the star award (*), thereby finding the best result with the highest accuracy value K value, 3 classes.

MATERIALS AND METHODS

In this study we used a dataset of reviews on google play for the DANA e-wallet seen from the star award (*) on the link https://play.google.com/store/apps/details?id=id.dana. The dataset we use is 500 testing data, where for each 1 -5 star there are 100 testing data based on Table 1 testing data on the number of star rating reviews in the DANA application comment column.

Table 2 Data Testing Number of FUND Reviews

	Number of Stars (*)				
	1	**2**	**3**	**4**	**5**
Amount of data	100	100	100	100	100
amount					500

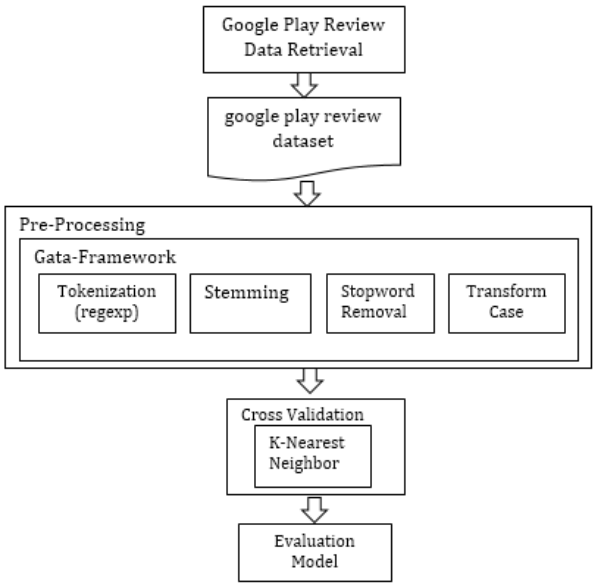
Source : (Masturoh & Pohan, 2021)

Based on Table 2, Data Testing Number of FUND Reviews, researchers took testing data based on star rating in the reviews in the Google Play comments column with details of the amount of data taken, namely 1 to 5 stars, the researcher took testing data for each 100 data on star awards. And here is an example of a DANA user review in the Google Play review column.

Source : (Masturoh & Pohan, 2021)

Figure 1 Example of DANA User Reviews

In Figure 1, we can see that DANA users provide their reviews in the comment column and provide stars (*) on the review regarding their opinion about the DANA application. The data is obtained from a manual process one by one from the user comments of the DANA application, then the data is preprocessed using the web gata framework on the link http://www.gataframework.com/ furthermore managed with the help of the rapidminer application. The research stages are described based on the picture. 2 The stages of research are described as follows:



Source : (Masturoh & Pohan, 2021)

Figure 2 Stages of Research

Based on Figure 2 above explains how the stages of the research were carried out, first the dataset was taken from a google play review on the DANA application based on the giving of stars (*) from each of the 1-5 stars, the data was taken as much as 100 then grouped into 3 classes, namely:

a. The first class consists of (1 to 5 stars).
b. 2nd class consists of (stars 1 and 5)
c. Class 3 is obtained from the grouping of stars 1 and 2 as negative labels, star 3 is labeled neutral and stars 4 and 5 are labeled positive.

Furthermore, after data collection, data preprocessing is carried out which aims to prepare the text into data that will be ready to undergo processing at the next stage by cleaning the entities that can interfere with the analysis proces. (Saidah & Mayary, 2020)

The data preprocessing was carried out using the web gata framework with several stages such as:

1. Tokenization regexp which has the function of breaking down sentences in a file into words as well as deleting unnecessary characters (Giovani et al., 2020).
2. Indonesian Stemming whose task is to find the root word of a word. By removing all affixes, both consisting of prefixes, inserts, suffixes and confixes (combinations of prefixes and suffixes) in derivative words (Ernawati & Wati, 2018).
3. Indonesian Stopword Removal has the function of removing conjunctions in a sentence that is entered (Aaputra et al., 2019).
4. Transform Case which functions to change all capital letters to lowercase all. This is so that at the time of classifying data there is no diversity of letters and there is no error in the tokenize process (Aaputra et al., 2019).

After the data is preprocessed, the next step is the classification process using the K-Nearest Negihbor (KNN) algorithm. The K-Nearest Neighbor (KNN) algorithm is used because it has simplicity to a process because the process is carried out based on a simple weighting approach and ease of implementation, adaptation and e learning proces and has a high accuracy value (Pristiyanti et al., 2018). *K-Nearest Neighbor* (KNN) is also a proces for grouping data based on the closest distance or the level of similarity of the data to the existing training dataset / data so that later the data will be grouped into a class by looking at the number of k values closest to the training data (Deviyanto & Wahyudi, 2018).

Then the results of the research phase carried out after performing the data classification process using the k-nearest neighbor algorithm will get the best results from the 3 classes with the best accuracy value based on the value of k 1 to 10.

RESULTS AND DISCUSSION

Based on the results of the research methodology carried out on sentiment analysis on the DANA application using the k-nearest neighbor algorithm with 500 data taken from user reviews based on star awards (*) with the amount of data from each star 1 to 5 as many as 100 data which are then obtained 3 classes and the best results from the 3 classes with testing the value of k 1 - 10 are described in Table 3, the test results are as follows :

Table 3 Test Results

value k	Accuracy Value		
	Class 1 (1-5 stars)	Grade 2 (1 & 5 stars)	Grade 3 (positive, neutral, negative)
1	74.30%	86.64%	75.27%
2	54.82%	77.98%	65.82%
3	57.08%	82.31%	57.05%
4	52.37%	81.36%	48.98%
5	48.20%	80.83%	46.91%
6	44.22%	76.52%	44.67%
7	41.77%	78.48%	4.16%
8	38.75%	75.62%	40.03%
9	39.12%	78.00%	41.58%
10	39.13%	76.57%	38.83%

Source : (Masturoh & Pohan, 2021)

From Table 3, the test results can be seen that the test values from the 3 classes with the first class using 1-5 stars, the highest accuracy is 74.30% obtained from giving the value of k 1, while in the second class test the highest accuracy value is 86.64% with a value of k 1, and the last test with

class 3 obtained the highest accuracy value of 75.27% with the value of k 1. So that from the comparison of the three classes that have been tested, we can see that the highest accuracy value is obtained in the second class by giving the k 1 value and the highest accuracy value is 86.64%.

CONCLUSION

Sentiment analysis on the DANA e-wallet on google play reviews using the k-nearest neighbor (KNN) algorithm which is grouped into 3 classes based on the number of stars (*) on the google play review menu and the highest accuracy value is obtained by testing the second class on (1 star and 5 star) and the k 1 value is obtained with the best accuracy result of 86.64% when compared to the first class and the third class and when compared to K 2 to 10 values.

REFERENCES

Aaputra, S. A., Didi Rosiyadi, Windu Gata, & Syepry Maulana Husain. (2019). Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(3), 377–382. https://doi.org/10.29207/resti.v3i3.1118

Adhine Salsabila, K. D., Ghiffarie, A., Baistama, R. P., Variadi, M. I., & Rhajendra, M. D. (2019). Analisis Sentimen Terhadap Produk The Body Shop Tea Tree Oil. *Jurnal Teknologi Dan Manajemen Informatika*, *5*(2). https://doi.org/10.26905/jtmi.v5i2.3684

Aulia, S. (2020). Pola Perilaku Konsumen Digital Dalam Memanfaatkan Aplikasi Dompet Digital. *Jurnal Komunikasi*, *12*(2), 311. https://doi.org/10.24912/jk.v12i2.9829

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, *2*(1), 32–41. https://ejurnal.itats.ac.id/integer/article/view/95

Cahyani, A. D., & Mardiana, T. (2020). Sentiment Analysis of Digital Wallet Service Users Using Naïve Bayes Classifier and Particle Swarm Optimization. *Jurnal Riset Informatika*, *2*(4), 241–250. https://doi.org/10.34288/jri.v2i4.160

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1. https://doi.org/10.14421/jiska.2018.31-01

Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, *VI*(1), 64–69.

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, *14*(2), 115. https://doi.org/10.33365/jti.v14i2.679

Gunawan, F., Fauzi, M. A., & Adikara, P. P. (2017). Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). *Systemic: Information System and Informatics Journal*, *3*(2), 1–6. https://doi.org/10.29080/systemic.v3i2.234

Inggiharti, N. (2020). Pengaruh Electronic Wallet Terhadap Kegiatan Keuangan Indonesia (Perbandingan Aplikasi Electronic Wallet Milik Perusahaan Financial Technology terhadap Aplikasi Electronic Wallet Milik BUMN). *University Of Bengkulu Law Journal*, *5*(1), 74–88.

Kesumastuti, T. M. (2020). The Process of Adoption Interest in Using Digital Wallet in Central Jakarta (Case Study on Go-Pay Users). *International Journal of Multicultural and Multireligious Understanding*, *7*(2), 277–286.

Kristiyanti, D. A., Putri, D. A., Indrayuni, E., Nurhadi, A., & Umam, A. H. (2020). E-Wallet Sentiment Analysis Using Naïve Bayes and Support Vector Machine Algorithm. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012079

Kusnawan, A., Diana, S., Andy, A., & Tjong, S. (2019). Pengaruh Diskon pada Aplikasi e-Wallet terhadap Pertumbuhan Minat Pembelian Impulsif Konsumen Milenial di Wilayah Tangerang. *Sains Manajemen*, *5*(2), 137–160. https://doi.org/10.30656/sm.v5i2.1861

Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. *Komputek*, *3*(2), 52. https://doi.org/10.24269/jkt.v3i2.270

Masturoh, S., & Pohan, A. B. (2021). *Analisis Sentimen Terhadap E-Wallet DANA Pada Ulasan Google Play Menggunakan Algoritma K-Nearest Neighbor (KNN)*.

Pristiyanti, R. I., Fauzi, M. A., & Muflikhah, L. (2018). Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *2*(3), 1179–

1186.

Puspita, Y. C. (2019). Analisis Kesesuaian Teknologi Penggunaan Digital Payment Pada Aplikasi Ovo. *Jurnal Manajemen Informatika*, *09*(02), 121–128.

Putra, M. W. A., Susanti, Erlin, & Herwin. (2020). Analisis Sentimen Dompet Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier. *IT Journal Research and Development*, *5*(1), 72–86. https://doi.org/10.25299/itjrd.2020.vol5(1).5159

Saidah, S., & Mayary, J. (2020). Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor. *Jurnal Ilmiah Informatika Komputer*, *25*(1), 1–17. https://doi.org/10.35760/ik.2020.v25i1.2411

Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia). *Prosiding SINTAK*, 480–486.