VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.3997

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

APPLICATION OF THE K-NEAREST NEIGHBOR (KNN) ALGORITHM IN SENTIMENT ANALYSIS OF THE OVO E-WALLET APPLICATION

Siti Masturoh[1*]; Risca Lusiana Pratiwi[2]; M Rangga Ramadhan Saelan[3], Ummu Radiyah[4]

Sistem Informasi[1, 2]
Sains Data[3]
Informatika[4]
Universitas Nusa Mandiri
https://nusamandiri.ac.id/
siti.uro@nusamandiri.ac.id[1], risca.ral@nusamandiri.ac.id[2], rangga.mgg@nusamandiri.ac.id[3], ummu.urd@nusamandiri.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *The OVO application can be downloaded on the Android platform via Google Play, Google play has a review feature on the application product to be downloaded, so that the review can be viewed or accessed by anyone, With these reviews, potential users of the application will see how important it is to consider using an application, problems regarding reviews or sentiment analysis of applications processed using text mining. The purpose of this study is to provide information to prospective OVO application users before using the application which can be seen from the results of giving reviews based on rating or stars (*) in the OVO application review column on Google Play and the authors categorize them into 3 classes, the first class (1 to 5 stars, second class (1 and 5 stars) third class by providing labeling grouping (1&2 stars are negative labels, 3 stars are neutral labels and 4&5 stars are positive labels) testing using the k-nearest neighbor method by finding the value of k from the k value of 1-10 to get the highest accuracy value, in order to obtain the highest accuracy value of 84.86% in the 2nd class test and giving a value of k 1 which means that the 1st and 5th star tests get positive values so that they can give a good impression to prospective application users OVO.*

Keywords: *E-Wallet, OVO, K-Nearest Neighbor*

Intisari— *Aplikasi OVO bisa di unduh pada platform android melalui google play, Google play terdapat fitur ulasan terhadap produk aplikasi yang akan diunduh, sehingga ulasan tersebut dapat dilihat atau diakses oleh siapa saja, dengan ulasan tersebut calon pengguna aplikasi akan melihat betapa pentingnya mempertimbangkan dalam menggunakan suatu aplikasi, permasalahan mengenai ulasan atau analisa sentimen terhadap aplikasi diolah menggunakan text mining. Metode text minning yang digunakan pada penelitia ini yaitu menggunakan algoritma K-Nearest Neighbor (KNN). Tujuan penelitian ini untuk memberikan informasi kepada para calon pengguna aplikasi OVO sebelum menggunakan aplikasi tersebut yang dapat dilihat dari hasil pemberian ulasan berdasarkan pemberian rating atau bintang (*) pada kolom ulasan aplikasi OVO di google play lalu penulis mengkategorikan ke dalam 3 kelas, kelas pertama (bintang 1 sampai bintang 5, kelas ke dua (bintang 1 dan bintang 5) kelas ke tiga dengan memberikan pengelompokan pelabelan (bintang 1&2 label negatif, bintang 3 label netral dan bintang 4&5 label positif) penggujian menggunakan metode k-nearest neighbor dengan mencari nilai k dari nilai k 1-10 untuk mendapatkan nilai accuracy tertinggi, sehingga didapatkan nilai accuracy tertinggi 84.86% pada penggujian kelas ke 2 dan pemberian nilai k 1 yang artinya pengujian bintang 1 dan bintang 5 mendapatkan nilai positif sehingga bisa memberikan kesan yang baik bagi para calon pengguna aplikasi OVO.*

Kata Kunci: *E-Wallet, OVO, K-Nearest Neighbor*

INTRODUCTION

The industrial revolution 4.0 is a phenomenon that collaborates cyber technology and automation technology which is marked by the digitization of almost all economic sectors including the financial sector (fintech). [1]. One of the fintech products is a digital wallet (E-wallet). [2], Currently, many people have switched to making payment transactions using e-wallets, moreover, there are many benefits and offers that can be obtained [3]. Apart from that, for the security of the existence of

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

digital wallets (e-wallets) in Indonesia, of course, you must obtain permission from Bank Indonesia (BI) so that the data of e-wallet users is maintained. [4].

E-wallets also have a positive impact, apart from being more efficient they can also reduce cash circulation [5], as well as being the best solution when there is a covid-19 virus to reduce the spread of the virus [6]. In Indonesia, the e-wallet products that we can use are OVO, DANA, LinkAja and Go-Pay [7]. From several e-wallet products, researchers took case studies on the OVO application.

We can download the OVO application on the Android platform on Google Play. Google Play has a review feature for application products that will be downloaded, so that these reviews can be seen or accessed by anyone [8], These reviews can affect potential users as a material consideration in using an application [9].

We can also process review data to see how important the sustainability of an application including OVO is, so it would be good if sentiment analysis was carried out on OVO application review data on Google Play. Sentiment analysis can determine whether an opinion or opinion on an entity has a positive or negative tendency so that it can be used as a reference for improving service and product quality. [10].

Research on e-wallet sentiment analysis has been carried out by previous researchers, and the following are several reference papers which are described in table.1 Reference paper is as follows :

Table. 1 Reference Paper

Research Title	Research Titles
Implementation of Naïve Bayes Classifier and Information Gain for sentiment analysis of e-wallet users during a pandemic	Using the naïve Bayes classifier algorithm and information gain by researching e-wallet users and showing classification results for 92% accuracy, 92% precision and 100% recall
Sentiment analysis based on marketing mix aspects of access to digital wallet application reviews (case study: the LinkAja application on Twitter)	Using the SVM algorithm with the undersampling method to balance classes into datasets. The classification results get negative sentiment on the product aspect with a value of 98% of the total reviews and from the aspect of the place with 100% of the total reviews, while neutral sentiment on the product aspect gets a value of 89% of the total reviews, and positive aspects as much as 98% of the total review
The effect of perceived convenience and service features on the interest in using e-wallets in the DANA application in Surabaya	significant and positive influence on the significant and positive interest in using (Y) with a value of Sig. 0.000 <0.05, while simultaneously (X1) and service features (X2) have an influence on interest (Y) with a presentation value of 26.2%, but 73.8% affect other variables that are not in the research model

Based on Table 1. Referral Paper, we can see research on sentiment analysis of e-wallet users using the naïve Bayes and information gain method. get the classification results using naïve bayes classifier produce 84% accuracy, 91% precision, and 91% recall. Meanwhile, the classification results using naïve Bayes and information gain are 92% accuracy, 92% precision, 100% recall [11]. Furthermore, for sentiment analysis research based on marketing mix aspects of access to digital wallet application reviews (case study: the LinkAja application on Twitter) this research focuses on aspect-based sentiment analysis to determine positive, negative or neutral aspects of the reviews given by consumers.

To classify aspects using string matching and using the thefuzz library, then for sentiment classification using the SVM algorithm then using undersampling to balance classes in the dataset. The classification results show that the LinkAja

application gets a negative sentiment value on the product aspect with a value of 98%, 89% neutral and 98% positive from the total reviews [12].

Subsequent research entitled The effect of perceived convenience and service features on the interest in using e-wallets in the DANA application in Surabaya, in his research using multiple linear regression analysis techniques which aim to determine the effect of convenience and service features on the interest in using e-wallets in applications with a sample that used a total of 214 respondents with a population of all DANA application users.

The results of this study show that ease (X1) has a significant and positive influence on the significant and positive interest in using (Y) with a Sig value. 0.000 <0.05, while simultaneously (X1) and service features (X2) have an influence on interest (Y) with a presentation value of 26.2%, but 73.8% affect other variables that are not in the research model [13].

Based on the references above, the author tries to conduct research with a different object, namely the OVO application and tries to use a different method, namely in this study using the KNN method, where data is collected based on reviews on Google Play to provide information to potential OVO application users.

MATERIALS AND METHODS

This study uses the results of user reviews of the OVO application on Google Play which gives a star rating (*) on the link https://play.google.com/store/apps/details?id=ovo.id. The data used in this study were 5000 data which were divided into two types of data with a ratio of 90% for training data and 10% for test data. The dataset The data used in this study were 5000 data which were divided into two types of data with a ratio of 90% for training data and 10% for test data [14], the number of reviews giving stars in the comments column on the OVO application.

Table 2 Data Testing Number of OVO Reviews

	Number of Stars (*)				
	1	2	3	4	5
Amount Data	100	100	100	100	100
Total					500

Based on Table 2. Data Testing Number of OVO Reviews, researchers took data testing based on star rating in the Google Play comments column with each amount of data taken, namely from 1 to 5 stars of 100 data. And the following is picture 1. An example of an OVO user review in the Google Play review column.

Fig.1 Example of an OVO User Review

Based on figure 1, we can see the reviews and stars (*) of the OVO application users regarding their opinion of the OVO application. We get the data from manual processes one by one from OVO application user reviews, after that the data is pre-processed using the web gata framework on the link http://www.gataframework.com/ selanjutnya managed with the help of rapidminer application. The research stages are described based on Figure 2. The research stages are explained as follows:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

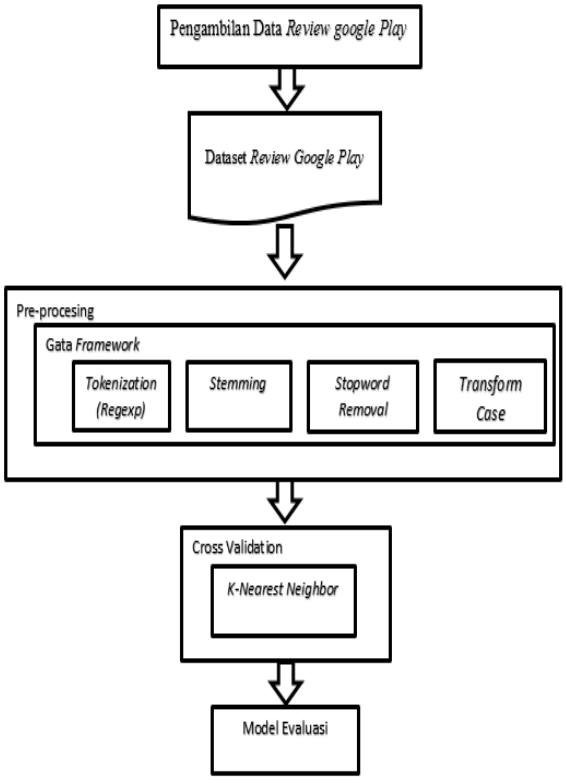

Fig. 2 Research Stages

Based on Figure 2 above, it explains how the stages of the research were carried out, first the dataset was taken from Google Play reviews on the OVO application based on giving stars (*) from each star 1-5, 100 data were taken then grouped into 3 classes, namely :

a. the first class consists of (stars 1 to 5).

b. class 2 consists of (1 and 5 stars)

c. class 3 is obtained from grouping stars 1 and 2 as negative labels, stars 3 are 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 first cleaning the entities that can interfere with the analysis process. [15]

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

1. 1. Tokenization regexp which has the function of breaking sentences in a file into words while removing unnecessary characters [16]
2. 2. Indonesian Stemming whose job is to find the basic words of a word. By eliminating all good affixes consisting of prefixes, infixes, endings and confixes (combinations of prefixes and suffixes) in derived words [17]
3. 3. Indonesian Stopword Removal has the function of removing connecting words in an inserted sentence [14]
4. 4. Transform Case which functions to change capital letters to all lowercase letters. This is intended so that when classifying data there is no diversity of letters and no errors occur in the tokenize process [14]

After the data has been preprocessed, the next stage is the classifier process using the K-Nearest Neighbor (KNN) algorithm. The K-Nearest Neighbor (KNN) algorithm is used because it has simplicity for a process because the process is carried out based on a simple weighting approach and is easy to implement, adapt and elerning process and has a high accuracy value [18]. K-Nearest Neighbor (KNN) is also a process for grouping data based on the closest distance or the level of similarity of the data with 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 [19].

Then the results of the research phase carried out after carrying out 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 giving k values 1 to 10.

RESULTS AND DISCUSSION

Based on the results of research that has been carried out on the application of the KNN algorithm (K-Nearesrt Neighbor) in sentiment analysis of the OVO application with a total data of 500 user reviews based on star rating (*) with a total of 1 to 5 stars for each of 100 data which is then test using 3 classes and testing based on the value of k 1-10. The test results can be seen based on table 3 of the test results as follows :

Table 3. Test Results

value *k*	Accuracy Value		
	class 1 (1-5 stars)	Class 2 (1 & 5 stars)	class 3 (positive, neutral, negative)
1	76.56%	84.86%	82.45%
2	53.50%	74.71%	73.32%
3	46.88%	79.12%	72.63%
4	40.65%	71.33%	71.26%
5	41.78%	72.76%	72.12%
6	34.79%	65.48%	71.08%

value *k*	Accuracy Value		
	class 1 (1-5 stars)	Class 2 (1 & 5 stars)	class 3 (positive, neutral, negative)
7	34.40%	66.93%	71.25%
8	29.30%	60.17%	67.46%
9	28.72%	61.17%	69.52%
10	24.76%	58.76%	58.35%

Based on table 3, the test results explain that from testing 3 classes to get the highest accuracy value by giving a value of k 1 with a reduction in the first class consisting of a review giving 1 star to 5 stars the highest accuracy is 76.56%, then testing in the second class consisting of 1 star and 5 star reviews have the highest accuracy score of 84.86%, then for the third class which consists of positive, neutral and negative values the labeling results get the highest score of 82.45%. So that from the comparison of the three class tests, the highest accuracy value was obtained in the second class test, namely scoring 1 star and 5 star with the highest accuracy value of 84.86%.

CONCLUSION

The application of the KNN (K-Nearest Neighbor) algorithm to the OVO application sentiment analysis based on reviews on Google Play based on star rating has been grouped into 3 classes based on the number of stars (*) and is obtained with the highest accuracy value obtained by testing using class 2 on stars 1 and 5 stars and a value of k 1 is obtained by obtaining the best accuracy of 84.86%. So that it can give a good impression to potential OVO application users before they decide to use the application..

REFERENCE

[1] R. Sulistyowati, L. Paais, and R. Rina, "Persepsi Konsumen Terhadap Penggunaan Dompet Digital," *ISOQUANT J. Ekon. Manaj. dan Akunt.*, vol. 4, no. 1, p. 17, 2020, doi: 10.24269/iso.v4i1.323.

[2] T. M. Kesumastuti, "The Process of Adoption Interest in Using Digital Wallet in Central Jakarta (Case Study on Go-Pay Users)," *Int. J. Multicult. Multireligious Underst.*, vol. 7, no. 2, pp. 277–286, 2020, [Online]. Available: https://ijmmu.com/index.php/ijmmu/article/download/1463/1159.

[3] A. B. P. Siti Masturoh, "Sentiment Analysis Against the Dana E-Wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm," *Ejournal.Nusamandiri.Ac.Id*, pp. 53–58, 2020, [Online]. Available: www.bsi.ac.id.

[4] L. Angelica and U. Soebiantoro, "Analisa Menggunakan Dompet Digital," *J. Manaj.*, vol. 14, no. 2, pp. 232–238, 2022, doi: 10.29264/jmmn.v14i2.11209.

[5] N. Inggiharti, "Pengaruh Electronic Wallet Terhadap Kegiatan Keuangan Indonesia (Perbandingan Aplikasi Electronic Wallet Milik Perusahaan Financial Technology terhadap Aplikasi Electronic Wallet Milik BUMN)," *Univ. Bengkulu Law J.*, vol. 5, no. 1, pp. 74–88, 2020, [Online]. Available: https://ejournal.unib.ac.id/index.php/ubelaj/article/view/10368.

[6] S. Aulia, "Pola Perilaku Konsumen Digital Dalam Memanfaatkan Aplikasi Dompet Digital," *J. Komun.*, vol. 12, no. 2, p. 311, 2020, doi: 10.24912/jk.v12i2.9829.

[7] S. Watmah, S. Fauziah, and N. Herlinawati, "Identifikasi Faktor Pengaruh Penggunaan Dompet Digital Menggunakan Metode TAM Dan UTAUT2," *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 261–269, 2020, doi: 10.31294/ijse.v6i2.8833.

[8] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[9] M. R. Fahlevvi, "Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine," *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no. 1, pp. 1–13, 2022, [Online]. Available: http://ejournal.ipdn.ac.id/JTKP,.

[10] K. D. Adhine Salsabila, A. Ghiffarie, R. P. Baistama, M. I. Variadi, and M. D. Rhajendra, "Analisis Sentimen Terhadap Produk The Body Shop Tea Tree Oil," *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 2, 2019, doi: 10.26905/jtmi.v5i2.3684.

[11] A. Isnanda, Y. Umaidah, and J. H. Jaman, "Implementasi Naïve Bayes Classifier Dan Information Gain Pada Analisis Sentimen Penggunaan E-Wallet Saat Pandemi," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, pp. 144–153, 2021, doi: 10.37012/jtik.v7i2.648.

[12] R. M. Turjaman and I. Budi, "Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter)," *J. Darma Agung*, vol. 30, no. 2, p. 266, 2022, doi:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

10.46930/ojsuda.v30i2.1672.

[13] N. D. Abrilia and S. Tri, "Pengaruh Persepsi Kemudahan Dan Fitur Layanan Terhadap Minat Menggunakan E-Wallet Pada Aplikasi Dana Di Surabaya," *J. Pendidik. Tata Niaga*, vol. 8, no. 3, pp. 1006–1012, 2020.

[14] S. Aditya, D. Rosiyadi, W. Gata, and S. M. Husain, "Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization," *J. Resti*, vol. 3, no. 3, pp. 377–382, 2019.

[15] S. Saidah and J. Mayary, "Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.

[16] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[17] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626.

[18] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, "Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 3, pp. 1179–1186, 2018, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1140.

[19] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.