JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 4, No. 2, December 2021, pp. 232~239
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 36/E/KPT/2019
DOI: 10.36378/jtos.v4i2.1776

Particle Swarm Optimization-based Support Vector Machine Method for Sentiment Analysis in OVO Digital Payment Applications

Retno Sari[1], Ratih Yulia Hayuningtyas[2]
[1,2]Teknik Informatika, Universitas Nusa Mandiri, Indonesia

Article Info

Article history:

Received 10 16, 2021
Revised 12 05, 2021
Accepted 12 20, 2021

Keywords:

PSO
SVM
OVO
Sentiment
Analysis

ABSTRACT

Sentiment analysis is used to analyze reviews of a place or item from an application or website that then classified the review into positive reviews or negative reviews. reviews from users are considered very important because it contains information that can make it easier for new users who want to choose the right digital payment. Reviews about digital payment ovo are so much that it is difficult for prospective users of ovo digital payment applications to draw conclusions about ovo digital payment information. For this reason, a classification method is needed in this study using support vector machine and PSO methods. In this study, we used 400 data that were reduced to 200 positive reviews and 200 negative reviews. The accuracy obtained by using the support vector machine method of 76.50% is in the fair classification, while the accuracy obtained by using the support vector machine and Particle Swarm Optimization (PSO) method is 82.75% which is in good classification.

Corresponding Author:

Retno Sari
Teknik Informatika
Universitas Nusa Mandiri
Indonesia

1. Introduction

Digital payments are currently one of the payment options used among the public. Digital payment is a non-cash payment that has utilized technology [1]. Digital payment is considered more practical because it can control spending, practical and many promos are given. With digital payment, people do not need to worry about forgetting non-cash payment tools because digital payment applications can be installed on everyone's smartphone that is always taken anywhere.

The covid-19 pandemic made an increase in digital payment users as a means of payment in shopping places, this was done to reduce the spread of the covid-19 virus. This digital payment has the same function as a wallet in general, which is to store some money.

Ovo is one of the digital payment applications that can be installed on smartphones through the play store. Ovo users have enough of this can be seen in the number of people who download the ovo application as many as more than 10 million app downloads. With ovo makes it easier for people to make payments by scanning

barcodes. To save or add money to the ovo application by transferring money to the virtual account number given to the application.

Classification sentiment analysis divides the three forces: documents, sentences, and aspects [2]. Sentiment analysis is used to analyze reviews of a place or item from an application or website that then classified the review into positive reviews or negative reviews [3].

Users of the ovo application are many, making users give reviews to the application both positive reviews and negative reviews. The review was made as a means of communication between users and ovo application makers so that ovo application makers can improve facilities that are felt less by users and as a form of user assessment of ovo applications.

Before someone decides to use a digital payment application will certainly see reviews from users. The information is taken into consideration to decide what application to use as a means of payment.

To get a whole conclusion from the review selected classification methods. The classification method selected is a PSO-based Support Vector Machine. Support Vector Machine method is one of the methods that is considered to have a kernel that has the diffusion of separating high-dimensional non-linear data[4], but Support Vector Machine has shortcomings in the selection of appropriate parameters or features [5]. Particle Swarm Optimization is one of the optimization techniques to improve accuracy [6]. A review of studies on related research can be seen in table 1.

Table 1 Review of Related Research Studies

Previous Research	Methods Used	Research Results
Comparison of Classification Algorithm for Sentiment Analysis of Foreign Language Film Review Naive Bayes	Support Vector Machine and k-Nearest Neighbour	SVM accuracy was 91.92% and AUC 0.981 in English datasets, while in German datasets it produced 90.96% accuracy and AUC 0.978 using 10 fold cross-validations[7].
Comparison of Support Vector Machine And Naïve Bayes Algorithm With Genetic Algorithms on Sentiment Analysis of Gubernatorial Candidates West Java 2018-2023	Support Vector Machine, Naive Bayes, SVM-based Genetic Algorithm, and Naïve Bayes based Genetic Algorithm	Using validation using 10 fold cross-validation. It found that the data value for support vector machines produced 92.61% with an AUC of 0.950, Naive Bayes produced an accuracy average of 93.29% with an AUC of 0.525, support vector machines based on genetic algorithms produced an average accuracy of 93.03% with an AUC of 0.869, Naive Bayes based on genetic algorithms produced an average accuracy of 92.85% with an AUC of 0.543. In this study, the ALgorithm-based Genetic Vector Machine Support Method has the greatest accuracy.[8].
Comparison of Vector Machine And Decision Tree Support Methods for Sentiment Analysis Review comments on online transportation applications	Decision Tree and Support Vector Machine	Support Vector Machine Method the accuracy value is higher with an accuracy value of 90.20% contained in the k-fold value of 3 kernel type radial [9].
Sentiment Analysis With SVM, Naive Bayes and KNN For The Study of Indonesian People's Responses to the Covid-19 Pandemic On Social Media Twitter	Support Vector Machine, Naïve Bayes and k-Nearest Neighbour	Classifying for the response of the Indonesian people to the Covid-19 pandemic through social media twitter with data taken as much as 10,000 response data. This study used 10 fold cross validation where SVM method by 90.1%, Naïve Bayes by 79.2%m KNN by 62.1%[10].

This research was conducted to find out how influential Particle Swarm Optimization is to improve the optimization of Support Vector Machine methods for sentiment analysis in ovo digital payment applications. This is done to make it easier for prospective users of the digital payment application ovo to determine reliable digital payment.

2. Research Method

In this study, some steps were taken as follows:

a. Data collection

This study used data obtained from play store. In this application was taken reviews about digital payments ovo, data used as many as 200 positive reviews and 200 negative reviews.

b. Initial data processing

At the initial data processing, stage/preprocessing through five processes namely transform case, tokenize, token by length filter, stopword filter, and stemming.

c. Proposed methods

The methods proposed in this study are the Support Vector Machine and PSO methods. Support Vector Machine method is one of the methods that is considered to have a kernel that can separate high-dimensional non-linear data[4].

Particle Swarm Optimization is one of the optimization techniques to improve accuracy [6].

d. Experiments and Testing Methods

Experiments and testing with the Support Vector Machine method use RapidMiner where the data used for experiments and testing is training data from play store that has been grouped into negative and positive classifications. Results from the assessment and testing will be improved accuracy with Particle Swarm Optimization to get the best results.

e. Evaluation and Validation of Results

In the evaluation and validation stage of the results of this study using 10 fold cross-validation, where accuracy is measured by confusion matrix in the form of ROC curves used to measure AUC values. Classification of AUC values as follows [11]:

0,90-100=*Excellent Classification;*
0,80-0,90=*Good Classification;*
0,70-0,80=*Fair Classification;*
0,60-0,70=*Poor Classification;*
0,50-0,60=*Failure.*

3. Result and Discussion

The results of this study start from the initial data processing, experimentation, and testing methods as well as evaluation and validation of results. Here are the stages:

A. Early Data Processing

At this stage, data cleaning is carried out by:

1) Transform Case

In the preprocessing stage of the transform case feature, all reviews both positive reviews and negative reviews have changed all letters of the text into lowercase all. Here are the results of the transforming case.

Review	*Tansform Case*
1st installation, my phone hanging. It requires me to did a force reboot on my phone... And when I tried to taking pics of my KTP for premier upgrade, one major problem happen... GODDAMNIT THIS APP AUTOFOCUS ARE MESSED MY DAMN PHONE CAMERA REAL BAD, TURNS ALL BLURRY AND UNREADABLE FOR FOR UPGRADE!! PUT A BETTER SUPPORT ON OLD, SINGLE CAMERA PHONE WILL YOU!? OR AT LEAST LET US SENT A SCANNED KTP INSTEAD A BLURRY PHOTO, COULDN'T WE?	*1st installation, my phone hanging. it requires me to did a force reboot on my phone... and when i tried to taking pics of my ktp for premier upgrade, one major problem happen... goddamnit this app autofocus are messed my damn phone camera real bad, turns all blurry and unreadable for for upgrade!! put a better support on old, single camera phone will you!? or at least let us sent a scanned ktp instead a blurry photo, couldn't we?*

2) Tokenize

At the tokenize preprocessing stage, all reviews eliminate certain characters such as punctuation marks and numbers in the review. Here's an example of tokenize.

Review	*Tokenize*
1st installation, my phone hanging. it requires me to did a force reboot on my phone... and when i tried to taking pics of my ktp for premier upgrade, one major problem happen... goddamnit this app autofocus are messed my damn phone camera real bad, turns all blurry and unreadable for for upgrade!! put a better support on old, single camera phone will you!? or at least let us sent a scanned ktp instead a blurry photo, couldn't we?	*st installation my phone hanging it requires me to did a force reboot on my phone and when i tried to taking pics of my ktp for premier upgrade one major problem happen goddamnit this app autofocus are messed my damn phone camera real bad turns all blurry and unreadable for for upgrade put a better support on old single camera phone will you or at least let us sent a scanned ktp instead a blurry photo couldn t we*

3) Filter token by length

At the preprocessing stage of the token by length filter feature, all reviews are filtered by length. Here are the results of the token by length filter.

Review	*Filter token by length*
st installation my phone hanging it requires me to did a force reboot on my phone and when i tried to taking pics of my ktp for premier upgrade one major problem happen goddamnit this app autofocus are messed my damn phone camera real bad turns all blurry and unreadable for for upgrade put a better support on old single camera phone will you or at least let us sent a scanned ktp instead a blurry photo couldn t we	*installation phone hanging requires did force reboot phone and when tried taking pics ktp for premier upgrade one major problem happen goddamnit this app autofocus are messed damn phone camera real bad turns all blurry and unreadable for for upgrade put better support old single camera phone will you least let sent scanned ktp instead blurry photo couldn*

4) Filter Stopword

In the preprocessing stage of the stopword filter feature, all reviews eliminate common words that do not have the required meaning or information, this feature eliminates some verbs, adjectives, adverbs, and connecting words. Here are the results of the stopword filter.

Review	*Filter Stopword*
installation phone hanging requires did force reboot phone and when tried taking pics ktp for premier upgrade one major problem happen goddamnit this app autofocus are messed damn phone camera real bad turns all blurry and unreadable for for upgrade put better support old single camera phone will you least let sent scanned ktp instead blurry photo couldn	*installation phone hanging requires force reboot phone tried taking pics ktp premier upgrade major problem happen goddamnit app autofocus messed damn phone camera bad turns blurry unreadable upgrade put support old single camera phone sent scanned ktp instead blurry photo*

5) Stemming

In the preprocessing stage of the stemming feature, all reviews change the word that is connected to the basic word by eliminating all the appending in the word. Here are the results of stemming.

Review	*Stemming*
installation phone hanging requires force reboot phone tried taking pics ktp premier upgrade major problem happen goddamnit app autofocus messed damn phone camera bad turns blurry unreadable upgrade put support	*instal phone hang requir forc reboot phone tri take pic ktp premier upgrad major problem happen goddamnit app autofocu mess damn phone camera bad turn blurri unread upgrad*

old single camera phone sent scanned ktp instead blurry photo	*put support old singl camera phone sent scan ktp instead blurri photo*

B. Experiments and Testing Methods

Testing methods using Support Vector Machine with fold cross-validation 2-9 and population 1-10. Here are the results of experiments and testing of Support Vector Machine (SVM) methods with fold cross-validation and population.

Table 2 Accuracy with Support Vector Machine

Cross Validation	***Accuracy***	***Support Vector Machine***	
		Precission	***Recall***
2	72.75% +/- 2.25% (mikro: 72.75%)	65.50% +/- 2.10% (mikro: 65.42%) (positive class: positif)	96.50% +/- 0.50% (mikro:96.50%) (positive class: positif)
3	72.26% +/- 3.31% (mikro: 72.25%)	65.33% +/- 2.98% (mikro: 65.19%) (positive class: positif)	95.50% +/- 0.03% (mikro: 95.50%) (positive class: positif)
4	76.25% +/- 1.48% (mikro: 76.25%)	68.59% +/- 1.42% (mikro: 68.55%) (positive class: positif)	97.00% +/- 1.73% (mikro: 97.00%) (positive class: positif)
5	75.50% +/- 4.91% (mikro: 75.50%)	68.50% +/- 5.40% (mikro: 67.96%) (positive class: positif)	96.50% +/- 3.39% (mikro: 96.50%) (positive class: positif)
6	75.22% +/- 6.28% (mikro: 75.25%)	68.20% +/- 6.35% (mikro: 67.72%) (positive class: positif)	96.51% +/- 2.01% (mikro: 96.50%) (positive class: positif)
7	74.00% +/- 3.99% (mikro: 74.00%)	67.22% +/- 4.39%% (mikro: 66.90%) (positive class: positif)	95.00% +/- 3.66% (mikro: 95.00%) (positive class: positif)
8	76.50% +/- 5.72% (mikro: 76.50%)	69.40% +/- 5.49% (mikro: 68.93%) (positive class: positif)	96.50% +/- 2.40% (mikro: 96.50%) (positive class: positif)
9	74.52% +/- 3.85% (mikro: 74.50%)	67.53% +/- 3.76% (mikro: 67.25%) (positive class: positif)	95.52% +/- 2.97% (mikro: 95.50%) (positive class: positif)
10	75.50% +/- 6.78% (mikro: 75.50%)	68.64% +/- 6.57% (mikro: 67.96%) (positive class: positif)	96.50% +/- 3.91% (mikro: 96.50%) (positive class: positif)

The accuracy table with Support Vector Machine found cross-validation 8 which has the greatest accuracy value of 76.50%, with the accuracy value still included in the fair classification. This study added Particle Swarm Optimization to improve accuracy by using cross-validation 8. Here's a table of accuracy results with Support Vector Machine and Particle Swarm Optimization.

Table 3 Accuracy with Vector Machine Support and Particle Swarm Optimization

Population	*Accuracy*	*Support Vector Machine and Particle Swarm Optimization*	
		Precission	***Recall***
1	77.50% +/- 8.05% (mikro: 77.50%)	70.63% +/- 7.79% (mikro: 69.93%) (positive class: positif)	96.50% +/_ 3.12% (mikro: 96.50%) (positive class: positif)
2	78.75% +/- 4.47% (mikro: 78.75%)	72.00% +/_ 4.71% (mikro: 71.70%) (positive class: positif)	95.00% +/- 3.32% (mikro: 95.00%) (positive class: positif)
3	81.75% +/- 3.07% (mikro: 81.75%)	74.51% +/- 3.23% (mikro: 74.33%) (positive class: positif)	97.00% +/- 5.2% (mikro: 97.00%) (positive class: positif)
4	79.50% +/- 2.78% (mikro: 79.50%)	72.36% +/- 3.38% (mikro: 72.18%) (positive class: positif)	96.00% +/- 2.00% (mikro: 96.00%) (positive class: positif)
5	79.25% +/- 4.68% (mikro: 79.25%)	72.03% +/- 4.79% (mikro: 71.75%) (positive class: positif)	96.50% +/_ 3.12% (mikro: 96.50%) (positive class: positif)
6	81.00% +/- 6.00% (mikro: 81.00%)	74.16% +/- 6.94% (mikro: 73.46%) (positive class: positif)	97.00% +/- 3.32% (mikro: 97.00%) (positive class: positif)
7	78.25% +/- 5.04% (mikro: 78.25%)	71.02% +/- 5.72% (mikro: 70.55%) (positive class: positif)	97.00% +/- 2.65% (mikro: 97.00%) (positive class: positif)
8	82.75% +/- 3.31% (mikro: 82.75%)	76.81% +/- 4.50% (mikro: 76.25%) (positive class: positif)	94.50% +/- 2.76% (mikr: 94.50%) (positive class: positif)
9	82.25% +/- 3.80% (mikro: 82.25%)	75.49% +/- 5.03% (mikro: 75.10%) (positive class: positif)	96.50% +/- 3.71% (mikro: 96.50%) (positive class: positif)
10	81.00% +/- 5.29% (mikro: 81.00%)	74.43% +/- 5.08% (mikro: 74.22%) (positive class: positif)	95.00% +/- 5.20% (mikro: 95.00%) (positive class: positif)

C. Evaluation and validation of results

By using Support Vector Machine using 8 fold cross-validation found the accuracy of 76.50%. Confusion matrix Support Vector Machine can be seen in the following table.

Table 4 Confusion Matrix Support Vector Machine

***Accuracy* : 76.50% +/- 5.72% (mikro: 76.50%)**		
	True Negatif	**True Positif**
Pred Negatif	113	7
Pred Positif	87	193

The test data above will be assessed the results of predictions using ROC charts. It can be seen in the ROC graphic image for Support Vector Machine and figure 2 ROC graphics for Support Vector Machine and PSO.

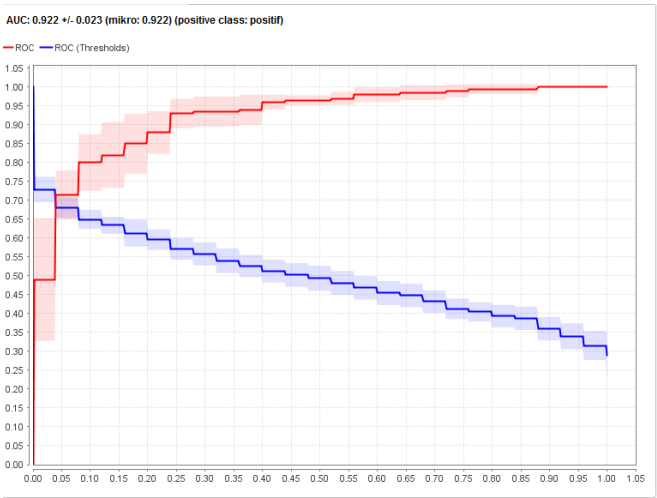

Figure 1 ROC Support Vector Machine Curve

Experiments using the Support Vector Machine and Particle Swarm Optimization methods using 8 fold cross-validation and population 8 found an accuracy of 82.75%. Confusion matrix Support Vector Machine and PSO can be seen in the following table.

Table 5 Confusion Matrix Support Vector Machine and Particle Swarm Optimization

***Accuracy* : 82.75% +/- 3.31% (mikro: 82.75%)**		
	True Negatif	**True Positif**
Pred Negatif	142	11
Pred Positif	58	189

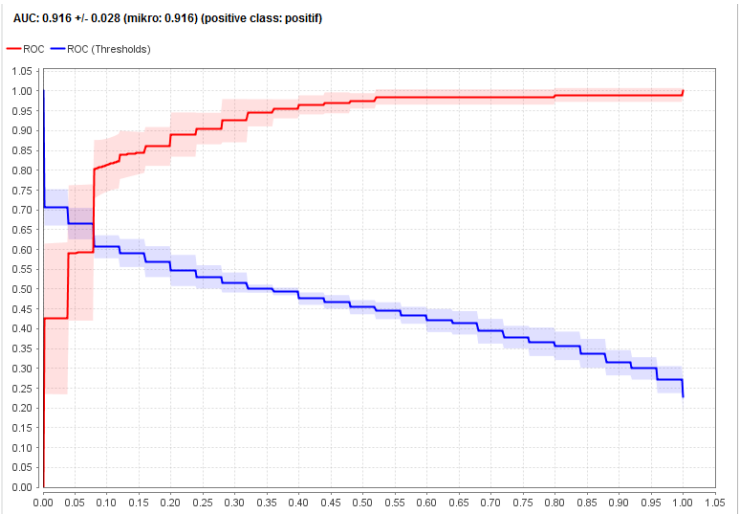

Figure 2 ROC Support Vector Machine and Particle Swarm Optimization Curve

4. Conclusion

This study entitled Particle Swarm Optimization-based Support Vector Machine Method compares the accuracy obtained from the use of 2 methods, namely the support vector machine method and support vector machine with PSO. After experimenting for sentiment analysis with English-language reviews of digital payment ovo, it was found that using support vector machine using 8 fold cross-validation had an accuracy of 76.50%, while the results of experiments using Particle Swarm Optimization-based vector machine support method with 8 fold cross-validation were found to be 82.75% accuracy. It can be seen that there is an increase when added optimization using Particle Swarm Optimization by 6.25%.

References

[1] A. Agrani and B. Rikumahu, "Perbandingan Analisis Sentimen Terhadap Digital Payment 'go-pay' Dan 'ovo' Di Media Sosial Twitter Menggunakan Algoritma Naive Bayes Dan Word Cloud," *eProceedings Manag.*, vol. 7, no. 2, pp. 2534–2542, 2020.

[2] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020.

[3] E. S. Basryah, A. Erfina, C. Warman, D. Digital, and P. Store, "ANALISIS SENTIMEN APLIKASI DOMPET DIGITAL DI ERA 4 . 0 PADA MASA PENDEMI COVID-19 DI PLAY STORE," pp. 189–196, 2021.

[4] M. Tri Anjasmoros and dan Fitri Marisa, "Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode Svm Dan Nbc (Studi Kasus: Komentar Pada Play Store)," *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020.

[5] N. Yunita, "Analisis Sentimen Berita Artis Dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization," *J. Sist. Inf. STMIK Antar Bangsa*, vol. 5, no. 2, pp. 104–112, 2016.

[6] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020.

[7] S. W. Yudha and M. Wahyudi, "Komparasi Algoritma Klasifikasi Untuk Analisis Sentimen Review Film Berbahasa Asing," *Semin. Nas. Inform. Sist. Inf. Dan Keamanan Siber*, pp. 180–185, 2018.

[8] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," *J. Tek. Komput. AMIK BSI*, vol. VI, no. 1, pp. 121–129, 2020.

[9] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021.

[10] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Prisma*, vol. 4, pp. 628–634, 2021.

[11] Gorunescu, *(2011). Data Mining: Concepts, Models, and Techniques.* Verlag Berlin Heidelberg: Springer, 2011.