



Sentiment Analysis of Google Play Store Reviews using Support Vector Machines

Mochamad Idris

Information System Department,
Faculty of Information Technology and
Communications
Semarang University, Semarang Indonesia

Mussalimun

Diponegoro University,
Semarang Indonesia

ABSTRACT

The development of Fintech Lending-crowdfunding through the Play Store with easy access to online loan services has got a lot of attention of market segments in Indonesia in meeting their financial needs. Data mining can be used to process the reviews contained in the Fintech Igrow comments column on Google Play. The feedbacks are in the form of comments or reviews represent positive or negative sentiments. This study aims to identify and analyze by classifying public opinions into positive and negative reviews. Support Vector Machine (SVM) algorithm was chosen as a classification method. The results show that the Support Vector Machine (Linear Kernel) has the same accuracy value of 77% as the Support Vector Machine (RBFKernel). This SVM model with RBF kernel performs well in classifying positive reviews, but there is still room for improvement in terms of precision for negative sentiment classification.

General Terms

The study uses Support Vector Machine (Linear Kernel) and Support Vector Machine (RBFKernel) with the Python 3.0 programming language.

Keywords

Sentiment Analysis; Classification; Support Vector Machine; Google Play Store

1. INTRODUCTION

There is particular interest among Indonesians in the development of Fintech Lending - crowdfunding on Play Store allowing easy access to the application and online loan services. This is a market segmentation where people's financial needs can be fulfilled with ease and without many conditions obtained from information about this loan application process. Igrow is an agricultural crowdfunding platform that currently has public trust to fund farmers' cultivation projects through www.igrow [13].

Igrow with its updated feature, of course, causes a lot of reviews which will increase every day in terms of changes and problems in the Google Play Store which can be seen from several reviews from users who have installed it on cellphones or installed the application on Android-based cellphones, of which 3 thousand the review has been recorded or recorded and of course the user has installed the application. Of course, the Igrow application attracts a lot of reviews with the development of feature updates and is updated day by day in terms of changes and problems in the Google Play Store, as shown by several reviews from users who have installed it on their mobile phones and Android. Has increased over 3,000. The review has been stored or recorded and of course the user has installed the

application. Application assessment in the form of reviews shows various reviews and comments from users who provide a form of assessment in the form of various responses and opinions both in terms of features and services, other opinions and complaints about bugs or difficulties or user suggestions and ideas regarding application development from users of this application. Reviewing the opinions and criticisms and ideas of these users will provide both positive and negative views and sentiments. Based on the description of the actual problems contained in the reviews and comments of this application as discussed above, this is a description of the order of positions in the Playstore and is one of the reasons this research must be carried out to help provide market/user sentiment analysis on the platform. Based on OJK data and the ranking of fintech lending companies with agricultural capital services as of January 5 2023, Igrow is third after Crowde.co and TaniFund and the market segmentation of these fintech applications is towards the Crowde.co and TaniFund applications for various reasons of convenience and level of security, thus influencing the level of congestion which has an impact on the level of credibility of the fintech application.

The results of previous studies on sentiment analysis on the Google Play Store showed that the accuracy, precision and recall values of K-NN and Naïve Bayes Classification are (73.85%, 76.60%, 85.71%) and (75.38%, 80.95%, 80.95%). Naïve Bayes classification produces slightly better predictions than K-NN by getting slightly better accuracy and precision values, but K-NN has higher values on recall [6]. The results of this indicate mostly used algorithms variant in sentiment analysis are K-NN and Naïve Bayes Classification, as well as SVM.

Support Vector Machine (SVM) is a method of machine learning algorithms. Support Vector Machine (SVM) works on basic principles by defining a hyperplane which is a function that can be used to separate classes which can maximize the margin between two different classes. In classification modeling, Support Vector Machine (SVM) has the advantage that this algorithm has a more mature and clearer concept than other algorithms. This makes SVM widely used in research [14]. As the results of research conducted by Budi and Mude (2020) analyzed the application of Support Vector Machine (SVM) and Naïve Bayes to group text reviews [2]. This research compares the level of accuracy of each method for classifying text reviews taken from the Google Play Store platform. From the application, this shows that the Support Vector Machine (SVM) is superior to Naïve Bayes in text reconstruction with correct test results of 81.46% and 75.41% respectively. Due to the characteristics mentioned above, Naïve Bayes and K-Nearest Neighbor classification methods are suitable for this study as a way to compare their performance



with support vector machine (SVM) classification methods.

Based on the problems above, it is necessary to carry out research to classify public opinions in the form of positive and negative reviews. Accuracy values will be calculated using the Support Vector Machine (SVM) classification approach. The results of this Support Vector Machine (SVM) method approach can be used as an evaluation of the Igrow platform to improve other features and services expected by users.

2. LITERATURE REVIEW

Sentiment analysis or opinion mining is the process of understanding, extracting and processing textual data automatically to get sentiment information contained in a comment sentences [3]. Sentiment analysis can be used to predict, analyze the public atmosphere, mood and views of internet users and technology services contained in a text [1]; therefore, this study is expected to find out the users' sentiment towards Fintech Igrow. The purpose of sentiment analysis is to provide valuable information to someone within unstructured data collection [5].

Many users express their emotions through writing in the comments column, including their satisfaction and disappointment. Text mining (text data mining) is the process of obtaining information from textual data including emails, documents, HTML files and reviews (Pamungkas and Kharisudin, 2020). Applications of text mining can assist in several fields with a lot of textual data, such as risk management, fraud detection, business intelligence, and social media analytics (Sihombing et al., 2021). This method also applies Natural Language Processing (NLP) techniques with the aim to help the system understand natural human language in the form of datasets or comment data. Sentiment analysis powered by machine Learning methods in sentiment analysis has Sentiment analysis has the advantage of building an analytical model in the context that will be analyzed, by classifying outputs into various sentiments, such as “very positive, positive, neutral, negative and very negative”, “angry, happy and sad”, etc. In this study, researchers developed a sentiment analysis model to analyze Igrow fintech comments on Playstore. We aim to classify product reviews as positive, negative or neutral by preparing training data and text preprocessing, as well as training models. In this case, sentiment analysis will study and detect patterns in the data using algorithms. Among the algorithms that can be applied to Sentiment Analysis include the K-Nearest Neighbor (K-NN), Naive Bayes, and Support Vector Machine (SVM).

Sentiment Analysis or opinion mining is a computational study to identify or recognize opinions, sentiments, evaluations, attitudes, emotions, subjective information, appraisals, or views expressed in texts. Sentiment analysis can get the percentage of positive and negative sentiment towards a person, company, institution, product or certain condition. Sentiment analysis can be divided into 3 polarity values, namely, positive, negative, and neutral, or determining which group is the source of positive or negative sentiment.

Support Vector Machine (SVM) is actually a harmonious combination of computational theories that have existed decades before, such as hyperplane margins [4]. Kernel was introduced by Aronszajn in 1950, and many supporting concepts. However, until 1992, there had never been an attempt to assemble these components. In contrast to the neural network strategy which tries to find a hyperplane that best separates the two classes, SVM tries to find an optimal hyperplane in the input space. SVM has the basic principle of a linear classifier,

and further develops to solve nonlinear problems. by incorporating the concept of kernel tricks in high-dimensional workspaces. This development provides research interest in pattern recognition to investigate the potential capabilities of SVM theoretically and in terms of applications. Following the feature extraction process, the important features of facial data or images will be used for classification. The classification method used was the Support Vector Machine (SVM). The SVM classifier uses a function or hyperplane to separate two classes. SVM will try to find the maximum hyperplane to separate two classes.

Training classification occurs after special features have been added to the training data. These special features are feature vectors in smaller dimensional spaces. In the classification process, the hyperplane variables for each classifier obtained will be stored and will later be used as data for each classifier in the testing process. In other words, the training classification process is to find support vectors from the input data.

The testing process used for classification were the extracted features from the testing and training process. This process yields the index value of the largest decision function that classifies the test data. Recognition is declared correct if the class resulting from the test classification process matches the test data class. Based on the classification process, the final result is a facial image that matches the index value of the decision function with the greatest result [4].

Evaluation is the interpretation stage of text mining results. A detailed evaluation is conducted to ensure that the results at the modeling stage are in line with the executed targets at the review stage, or based on the comment data on Playstore reviews.

The built classification model needs to be evaluated to check how well the model performed on the desired classification. Precision and recall are commonly used to measure the performance of a classifier performance [9].

The accuracy value represents how the entire documents accurately classified. In the case of imbalanced classes/categories in training datasets where there are majority and minority data classes, the accuracy value often cannot accurately represent the model performance [7]. With a classification value of 97%, the model can be considered to be very accurate. There is a possibility, however, that 97% of the detections are correct, while the remaining 3% are incorrect. To overcome this, precision and recall measurements are usually carried out in evaluating classification models. Apart from knowing and showing the overall accuracy of the model, this measurement is also able to show how the model performs in each class.

Precision and recall measurements are two evaluation metrics often used in text classification [11]. Taking the example of class R samples, precision is the number of samples in class R that are correctly guessed as R, compared to the total amount of sample data guessed as R. Additionally, recall is the percentage of class R samples that are correctly classified as R compared to the total number of R samples. In this study, the mechanism that can be used to measure the validity of classification results is by calculating precision, recall and f-score values.

When calculating the precision value, we are measuring the degree of certainty (exactness) or the percentage of data that is classified correctly by the classification model. Recall is the opposite of precision. Recall measures the sensitivity or ratio of



data for each label that is correctly classified to data that is incorrectly classified for another label (misclassified). F-score is a trade-off between precision and recall. The f-score value is obtained by calculating the harmonic mean of precision and recall.

According to Lancaster (1979) in Sokolova & Lapalme (2009), recall and precision are calculated using the following formula:

$$\text{Recall} = \frac{\text{number of relevant documents retrieved (a)}}{\text{total number of documents in the database (a+c)}} \times 100 \quad (1)$$

$$\text{precision} = \frac{\text{number of relevant documents retrieved (a)}}{\text{total number of documents retrieved (a+b)}} \times 100 \quad (2)$$

F-measure or F1 is an evaluation calculation in information retrieval that combines recall and precision. The recall and precision values in a situation can have different weights. Using the harmonic mean of recall and precision, the F-measure displays the reciprocity between recall and precision..

According to Manning (2009), separating similar documents is sometimes worse than putting pairs of dissimilar documents into the same cluster [9]. Thus, we can use the F- with a false negative value that is stronger than a false positive value. Next, selecting a value of $\beta >$, thus giving more weight to recall. A balanced FMeasure gives the same weighting to recall and precision, with a value of $\alpha = \frac{1}{2}$ or $\beta = 1$. The range of F-Measure values is 0 to 1.

3. METHODOLOGY

The data was collected from comments made by users of Fintech Igrow's Google Play store platform. This study only used comments, not data on commenters' names, comment dates, satisfaction ratings, etc. Reviews and collections of data are integrated into the review data set.

The data is then subjected to pre-processing, including:

1. *Case folding is the process of converting all letters in a text to lowercase.*
2. Tokenization is collecting all words and removing unnecessary characters including punctuation marks, other symbols " , . / ;) and etc.
3. Stop word Removal is the process of checking the list of stopwords and the stop word lists are collections of connecting words to be removed. But if there are no conjunctions, the process will continue without removing the words in the document.
4. Stemming is a process of transforming affixes into important words by removing all affixes. The importance of stemming process in creating a system is eliminating prefixes and suffixes.

Sentiment Support Vector Machine (SVM) training aims to find the vector α , W value and constant b to get the best hyperplane. For this study, comments were used as training data. In SVM training, each classification model was trained on data from two classes i and class j. The input data that will be used for the training process, namely comment data P1 to P6, has been given a class and has gone through the pre-processing stage. In accordance with the input data, the comment data was given a positive class and a negative class, then given a class 1 or -1 label, where class -1 was the negative class while class 1 was the positive class.

In this process, input data that has gone through the pre-

processing stage is converted into vector data. In the training process, comment data P1 to P6 with labels 1 and 0 will be used.

Input Space Feature Mapping to carry out the RBF kernelization calculation process, calculate the training data based on the function on the RBF $x_i - x_j$ first. Then kernelization can be implemented by entering equation 1, namely

$$(xix) = \exp(-\gamma||x_i - x||^2), \gamma > 0 \quad (\text{Equation 1})$$

Then do the same calculations for the next iteration. The next stage is to carry out calculations for the y value, which is the value of the label or value of the class that has been given. Next, $y_i y_j$ is calculated for 6 data where $i, j = 1, 2, \dots, n$.

Calculating the Lagrange Matrix kernel above each element is the result $K(xix) = \exp(-\gamma||x_1 - x||^2)$ which will correlate with α_i, α_j . By using kernels instead of dot-product $xixj$ in the duality equation.

$$Ld = \Sigma i=16 \alpha i - 12 \Sigma i=1, j=16 \alpha i \alpha j y_i y_j x_i x_j \quad (\text{Equation 2})$$

$$\text{Condition 1} : -\alpha_1 + \alpha_2 - \alpha_3 + \alpha_4 + \alpha_5 - \alpha_6 = 0$$

$$\text{Condition 2} : \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6 \geq 0$$

In the objective function, the second term has been multiplied by $y_i y_j$. This equation meets Quadratic Programming standards so that it can be solved using a commercial solver for Quadratic Programming (QP).

Separator Field This result shows that all training data is a support vector. Because the value $\alpha > 0$ while b was obtained from the training process carried out. After all α and b were obtained, the Support Vector Machine (SVM) model is ready to be used for predictions as in equation 3, namely:

$$f(\emptyset(x)) = (\text{sign}(w \cdot \emptyset(x)) + b)$$

Where

$$\begin{aligned} &= \sum_{i=1, x_1 \in sv}^n (\text{sign}(a_1 y_1 \emptyset(x) \cdot \emptyset(x_i)) + b) \\ &= \sum_{i=1, x_1 \in sv}^n (\text{sign}(a_1 y_1 K(x, x_i)) + b) \end{aligned}$$

where $i = 1, 2, 3, \dots, n =$ number of support vectors. Then we get the dividing field:

$$f(\emptyset(x)) = \sum_{i=1, x_1 \in sv}^n (\text{sign}(a_1 y_1 K(x_i, x_{testing})) + b)$$

Evaluation and validation are carried out after getting the α and b values as feature models, and the gamma parameter value 0.5 (γ) from the training process, then test the test data into class +1 or -1 with the feature model that has been obtained. The data used for testing is training data P1 – P6 as the result of training data.

The validation stage was carried out by applying 6 folds cross validation. The validation process involved two parts, namely: training set and testing set. The training data sub-process is used to teach model determined at the modeling stage with existing training data. After the algorithm model is trained in the training subprocess stage, testing will then be carried out. A Confusion Matrix was used to evaluate or test of classification results.

4. RESULTS AND DISCUSSION

The research data used in this analysis was the Igrow review



This classification provides a comprehensive assessment of the performance of a linear kernel-based Support Vector Machine (SVM) in classifying sentiment in customer reviews. This linear kernel-based SVM achieved 77% accuracy. It indicated that 77% of the model predictions were correct. This reflects a reasonable level of overall accuracy in sentiment classification.

For the “Negative Sentiment” class, the model achieved a high precision of 90%, which indicates that 90% of the data predicted as negative sentiment was indeed negative. However, the recall value is lower, namely 62%, which implies that the model is able to recognize 62% of negative sentiments in the dataset. The F1 score for negative sentiment is 73%, indicating balanced performance between precision and recall for this class. Support, or the amount of data included in the negative sentiment class, is 112.

Meanwhile, for the "Positive Sentiment" class, SVM shows a precision of 71%, indicating that 71% of the data predicted as positive sentiment is actually positive. Recall for positive sentiment was higher at 93%, which means the model correctly identified 93% of the positive sentiment in the dataset. The F1 score for positive sentiment reached 80%, indicating good model performance. The support value for positive sentiment is 111. Overall, the Support Vector Machine with a linear kernel performs quite well, with a balanced comparison between precision and recall for negative and positive sentiment.

In a comparative study of sentiment analysis algorithms on Google Play Store reviews for the “Igrow” application, Support Vector Machine (SVM) with RBF Kernel showed good performance. The model achieved an overall accuracy of 77%, indicating that about 77% of its predictions were correct, making it quite effective in classifying customer review sentiment.

In classifying negative sentiment, this SVM with the RBF kernel shows a significant precision of 89%, indicating that 89% of reviews classified as negative are truly negative. However, the recall for negative sentiment was 61%, indicating that the model was only able to identify 61% of all negative sentiment in the data. The harmonized mean of precision and recall, known as the F1 score, is 72% for negative sentiment.

For positive sentiment, the SVM with this RBF kernel achieved a precision of 70%, indicating 70% accuracy in predicting positive sentiment, and an impressive recall of 93%, indicating its ability to identify the majority of true positive sentiment. The F1 score for positive sentiment is 80%, which highlights the model’s effectiveness in balancing between precision and recall. It is important to note that SVM with RBF kernel shows potential for increased precision for negative sentiment classification.

Overall, this SVM model with RBF kernel performs well in classifying positive reviews, but there is still room for improvement in terms of precision for negative sentiment classification.

Additionally, the multinomial naive Bayes model showed superior performance in classifying negative sentiment, as evidenced by a high recall of 90%, and the model correctly identified 90% of all existing negative reviews. Additionally, the model also achieved a good precision of 70% for negative sentiment, and the F1 value was 79%, indicating solid performance overall. A support value of 112 showed the number of reviews classified as negative.

On the positive sentiment side, this model achieved a fairly

high precision of 86%, indicating that 86% of reviews labeled as positive were actually positive. However, recall for positive sentiment was relatively lower at 61%, indicating that the model correctly identified 61% of all true positive reviews. The F1 score for positive sentiment was 72%, providing a balanced measure of performance. The support value for positive sentiment was 111.

Overall, the Multinomial Naive Bayes model provides good performance in predicting negative sentiment and maintains a good balance of precision and recall for negative sentiment, and shows high precision for positive sentiment. This model achieved an accuracy of 76%, which reflects its prediction accuracy for both positive and negative sentiment categories. However, this model has slightly lower recall for positive sentiment.

5. CONCLUSION

The calculation results of the Support Vector Machine (Linear Kernel) method has the same accuracy value as the Support Vector Machine (RBF Kernel) with an accuracy value of 77%, but the values are different for Precision, Recall and F1 Score.

The performance Support Vector Machine (SVM) based linear kernel for Negative Sentiment class, the model achieved a high precision of 90%, which indicates that 90% of the data predicted as negative sentiment was indeed negative. Meanwhile, for the "Positive Sentiment" class, SVM shows a precision of 71%, indicating that 71% of the data predicted as positive sentiment is actually positive.

The performance Support Vector Machine (SVM) SVM with the RBF kernel shows In classifying negative sentiment a significant precision of 89%, indicating that 89% of reviews classified as negative are truly negative. For positive sentiment, the SVM with this RBF kernel achieved a precision of 70%, indicating 70% accuracy in predicting positive sentiment

It is necessary to carry out trials with other algorithm methods so that better levels of performance can be identified and it is also necessary to optimize the accuracy results that have been obtained from various methods.

6. REFERENCES

- [1] Basryah E. S., Erfina, A. dan Warman, C. 2021. *Analisis sentimen aplikasi dompet digital di era 4.0 pada masa pandemi covid-19 di play store menggunakan algoritma Naive Bayes Classifier*. In: Prosiding SISMATIK 2021, Sukabumi, Agustus 2021.
- [2] Budi dan Mude. 2020. Perbandingan metode klasifikasi Support Vector Machine dan Naive bayes untuk analisis sentimen pada ulasan tekstual di google play store. *Jurnal Ilmiah Ilkom*, 12 (2) : pp.154-161.
- [3] Buntoro. 2017. Analisis sentimen calon gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal*, 2(1), pp: 32-41
- [4] Burges, Christopher J. C. 1998. "A Tutorial on Support Vector Machine for Pattern Recognition", *Data Mining and Knowledge Discovery* 2:121-167, 1998.
- [5] Herlinawati. 2019. The effect of entrepreneurial orientation on SMEs business performance Indonesia. *Journal of Entrepreneurship Education*, Volume 22, Issue 5.
- [6] Ikhsan, R. B., Saraswati, L. A., Muchardie, B. G., Vional and Susilo, A, 2019. The determinants of students' perceived learning outcomes and satisfaction in binus



- online learning. In : 2019 5th International Conference on New Media Studies (CONMEDIA), October 9-11, 2019.
- [7] Jiawei, Kamber, and Pei. 2012. *Data Mining: Concepts and Techniques*. Elsevier.
- [8] Manning, C., D., Raghavan, P., & Schütze, H. 2009. *An Introduction to Information Retrieval*. Cambridge University Press Online Edition
- [9] Miner, G., Delen, D., Nisbet, R. A. 2012. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Elsevier.
- [10] Pamungkas, F. S., dan Kharisudin. I, 2020. Analisis sentimen dengan SVM, NAIVE BAYES dan knn untuk studi tanggapan masyarakat Indonesia terhadap pandemi covid-19 pada media sosial twitter, PRISMA, Prosiding Seminar Nasional Matematika, vol. 4, p. 633,
- [11] Saksonova & Merlino. 2017. Fintech as financial innovation – the possibilities and problems of implementation. *European Research Studies Journal* Volume XX, Issue 3A, 2017, pp.: 961-973.
- [12] Sihombing, D. A., James, D. D., Massie, Merinda, H. C., Pandowo. 2021. Pengaruh brand image dan kualitas layanan terhadap keputusan pembelian layanan J&T express selama pandemi Covid-19. *Jurnal EMBA* Vol.10 No.1 Januari 2022, Hal. 1794-1802.
- [13] Stefany dan Sari. 2019. Strategi Marketing Public Relations PT Crowde Membangun Bangsa dalam Meningkatkan Minat Investasi. *PROLOGIA*, 2(2):
- [14] Zhou B., Wu X, Lv Z, Zhang L, Guo X. 2016. A fully automated trial selection method for optimization of motor imagery based brain-computer interface. *PLoS ONE* 11(9). <https://doi.org/10.1371/journal.pone.0162657>.