PAPER NAME

58119941582023G1kinerja_Paper.pdf

AUTHOR

Bernardinus Harnadi

WORD COUNT

5100 Words

CHARACTER COUNT

26856 Characters

PAGE COUNT

6 Pages

FILE SIZE

317.4KB

SUBMISSION DATE

Apr 3, 2024 7:49 AM GMT+7

REPORT DATE

Apr 3, 2024 7:50 AM GMT+7

## 17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 16% Internet database
- 9% Publications database
- Crossref Posted Content database

## Excluded from Similarity Report

- Crossref database
- Submitted Works database
- Bibliographic material
- Cited material

# Evaluating the Performance and Accuracy of Supervised Learning Models on Sentiment Analysis of E-Wallet

Bernardinus Harnadi
Information Systems Department
Soegijapranata Catholic University
Semarang, Indonesia
bharnadi@unika.ac.id

Albertus Dwiyoga Widiantoro
Information Systems Department
Soegijapranata Catholic University
Semarang, Indonesia
yoga@unika.ac.id

***Abstract*—This study has purpose to evaluate the performance and accuracy of supervised learning Models used on Sentiment Analysis of E-Wallet. User comment data was taken using the web scraping method on 4 major E-wallet applications in Indonesia namely, Ovo, Dana, Doku, and LinkAja. The data was taken in January-May 2023 with the amount of data after preprocessing are 11267 with 6349 negative labels and 4918 positive labels. Data labeling uses a star scale that has been pinned by the user, 1-3 stars are labelled negative and 4-5 stars are labelled positive labels. The labeling results were tested using supervised learning model including SVM (Support Vector Machine), Multinomial Naive Bayes, Bagging with Multinomial Naive Bayes, and Random Forest algorithms. The performance of these algorithms is measured using Precision, Recall, F1-Score, and Support. The accuracy of the algorithms is also evaluated using train accuracy score, test accuracy score, train ROC-AUC Score, test ROC-AUC score, area under precision-recall curve, and area under ROC-AUC. This study shows that labeling generates a significant value, which means that the user's negative and positive comments need to be considered by the E-wallet manager in order to improve the quality of the system and services.**

***Keywords— sentiment analysis, e-wallets, SVM, Naïve Bayes, Indonesia***

## I. Introduction

Cashless transactions are becoming commonplace and have the potential to make conventional transactions with currency obsolete. Strategic industrial markets, such as tourism, business and healthcare have currently adopted these digital transactions [1]. E-wallet is a type of FinTech that is safe, mobile and easy to access. E-wallet is defined as a digital payment method where the available funds are stored on a server and not on a chip [2]. Another definition states that it is an electronic card that allows digital transactions via smartphones [3]. The most significant contribution of the FinTech invention is the virtualization of debit and credit cards, which eliminates the need for consumers to carry physical financial media and offers a new level of innovation in transactions. [4]. In recent years, e-wallets have evolved into a transaction tracking method with a focus on cost-effectiveness [5]. In addition, organizations and companies have a need to develop business strategies by capturing the intentions of prospective clients in the e-wallet market share [5]. Therefore, e-wallets are proper to studied considering the modern structural adjustments to the digital economy as a shaper of the global economic landscape. [6].

E-wallets offer the advantages of easy, cheap and flexible on cashless transactions, especially for individuals who do not have banking access. The main benefits of e-Wallets for merchants and customers including minimizing cash risks, faster payments, and saving effort and time [7] [8]. Currently, mobile payments using e-Wallets have grown rapidly in developing countries. They are making significant efforts to improve its applicability [9] [10]. However, there are still many problems with e-wallet applications experienced by users, as evidenced by the large number of user comments on each e-wallet application included in the Play Store.

In general, consumer surveys are a common approach used to determine consumer demand and preferences but are usually time consuming and expensive [11]. Many of e-commerce consumers who purchase products through online platforms such as Amazon.com want to express their opinions about products online. Reviewing product by users truly reflect consumers' experiences and feelings in using the product which usually contain positive and negative sentiments towards the product [12]. These reviews provide valuable information to service providers regarding customer reactions, attitudes, needs and preferences towards related products and services that can help them to improve their product design and after-sales service quality [13]. By utilizing the large number of online reviews with sentiment orientation (positive or negative), consumer preferences can be explored for the products and services they purchase [14].

This study has purpose to evaluate the performance and accuracy of supervised learning Models used on Sentiment Analysis of E-Wallet. There are four supervised learning models naming SVM, Multinomial Naive Bayes, Bagging with Multinomial Naive Bayes and Random Forest used in this study. The performance of these algorithms is measured using Precision, Recall, F1-Score, and Support. The accuracy of the algorithms is also evaluated using train accuracy score, test accuracy score, train ROC-AUC Score, test ROC-AUC score, area under precision-recall curve, and area under ROC-AUC. Evaluation of the performance and accuracy of the models are carried out to test whether the model can be used for sentiment analysis purposes, especially for e-wallet sentiment analysis.

## II. Review of Literatures

This study employs several steps naming pre-processing, building model, analyzing accuracy level of model, and verifying the accuracy of model.

### *A. Pre-modelling.*

This step starts with collecting data using web scraping in one year period (January-May 2023) and then labelling data. Furthermore, selection and balancing data must be conducted to attain the ripe dataset. Pre-modelling step consists of pre-text processing, word vectorization, and evaluation.

Pre-text preprocessing

The pre-text preprocessing stage is carried out before the reviewing data is converted into a numeric vector. The process includes case folding, stop-word removal, tokenization, and stemming [15].

Case folding is the process of characters uniformity into lowercase letters. This process is necessary with the reason that the same word with different fonts will be considered as two different features. This process can increase the dimensions of the generated token without giving new meaning to the variety of features.

Stop-word removing is the process of removing word lines that only contain one and duple word and deleting lines that contain symbols.

Tokenization is the process of breaking down a sentence (comment) into its constituent single words. Stop word removal is the process of removing formal words (no negative or positive sentiments). In this study, the process of removing stop words uses a built-in function. In addition, slang word conversion is the process of converting slang words into standard words.

Stemming is the process of converting tokens (words) into its basic forms. The next process is combining several words into one sentence.

Data labeling can be done in 2 ways automatic and manual. E-wallet customer review data includes reviews and service ratings from 1 to 5 given by consumers. Data labeling is based on review score data. Scores 4-5 are labeled positive or 1. Scores 1-3 are labeled negative or 0.

Word Vectorization

Vectorization has purpose to convert each token in the dataset into a vector value. The method to conduct this vectorization is TF-IDF. TF-IDF is a word weight value. TF-IDF represents the distribution of each word in a whole document or a corpus [16]. The first step calculates the TF value using equation (1), calculate the IDF value using equation (2), and calculate the TF-IDF value using equation (3) [17].

$$TF\ (t,d) = 0.5 + 0.5 + \frac{f(t,d)}{max(\{f(w,d):w \in d)} \tag{1}$$

$$IDF\ (t,d) = log\frac{N}{Df(t,d)} \tag{2}$$

$$TF - IDF\ (t,d,D) = tf(t,d)x\ idf(t,D) \tag{3}$$

### B. Building Models

The models are employed in this study by means of TF-IDF, SVM with TF-IDF, Multinomial Naive Bayes, Gaussian Naive Bayes, and Random Forest. The models employed in this study are to classify user comment review.

### C. Calculating performance of the model

The confusion matrix is a matrix for storing information used as an indicator of the performance of the model and a reference for the performance of the classification algorithm in the evaluation phase. The confusion matrix resulted at this process shows on Table 1.

The classification of data from the confusion matrix is used to generate meaningful data to calculate performance of the model, including Accuracy, Precision, Recall (Sensitivity / True Positive Rate), and F1-Score.

TABLE I. CONFUSION MATRIX

| | | Predicted Values | |
|---|---|---|---|
| | | Positive | Negative |
| Actual Values | Positive | TP | FP |
| | Negative | FN | TN |

Note: TP = how much of the data is actual positive, and the model also predicts positive. TN = how much of the data is actual negative, and the model also predicts negative. FP = how much of the data is actually negative, but the model predicts positive. FN = how much data is actual positive, but the model predicts negative.

Accuracy is defined as the total number of times the model can perform classifications correctly. The accuracy formula can be written using the equation (4) [18].

$$Accuracy = \frac{TP+TN}{Total} \tag{4}$$

Precision measures how many of the positive predictions are actually correct. The precision formula can be written using the equation (5).

$$Precision = \frac{TP}{FP+TP} \tag{5}$$

Recall measures how much of the total positive of data are successfully identified by the model. The recall formula can be written using the equation (6).

$$Recall = \frac{TP}{FN+TP} \tag{6}$$

F1-score is the harmonic average between precision and recall. This score can give a better view of model performance if precision and recall are in different ranges. The F1-score formula can be written using the equation (7).

$$F1 - Score = 2\ x\ \frac{Precision*Recall}{Precision+Recall} \tag{7}$$

In addition, there are several others performance model naming Support, Macro Average, and Weighted Average. The first, Support is the total amount of data in each class. Meanwhile, Macro Average is the average of metrics (precision, recall, F1 score) calculated for each class separately, without taking into account class proportions. The last, Weighted Average is an average metric calculated for each class by assigning weights based on the amount of data in each class. The Weighted Average Score provides a picture of overall model performance by considering how much influence each class has on the average.

Performance evaluation of a model provides good insight into the model's performance against each class as well as the overall model performance.

### D. Verifying the accuracy of the model.

In the context of model evaluation, the metrics of evaluation mentions several evaluation naming train accuracy score, test accuracy score, train ROC-AUC Score, test ROC-AUC score, area under precision-recall curve, and area under ROC-AUC [19].

The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). TPR (8) is placed on the vertical (y) axis and FPR (10) is placed on the horizontal (x) axis.

TPR (True Positive Rate) or Sensitivity:

$$TPR = \frac{TP}{TP+FN} \tag{8}$$

Specificity:

$$Specificity = \frac{TN}{TN+FP} \tag{9}$$

FPR:

$$FPR = 1 - Specificity = \frac{FP}{TN+FP} \quad (10)$$

Train Accuracy Score is defined as the extent to which the model actually predicts the correct class on the training data. The score calculates what percentage of the total predictions in the training data are proper.

Test Accuracy Score is the same with train accuracy score with difference in the measured data is test data. The score indicates how well the model performs on data it has never calculated before.

Train ROC-AUC Score. ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) measures a model's ability to differentiate between two classes by plotting a ROC curve and calculate the area under it. The ROC-AUC score reflects how well the model separates classes on the training data.

Test ROC-AUC Score is a ROC-AUC score measured on test data, and the score reflects how well the model differentiates classes on data never been calculated before.

Area Under Precision-Recall Curve is an area under the Precision-Recall curve describing the balance between precision and sensitivity of the model at various prediction thresholds. A higher score indicates that the model has good precision on data with minority or less balanced class.

Area Under ROC-AUC Curve is an area under the ROC-AUC curve having focus on more general class separation rather than just a specific sensitivity. This score provides an overall picture of the model's ability to separate classes.

All of evaluation method including the accuracy score, ROC-AUC score, and area under the Precision-Recall and ROC-AUC curves provide insight into the model's performance in classifying data. The higher the score, the better the model performance. However, it is important to consider the context and purpose of implementation in evaluating the model as a whole.

## III. Methodology

Dataset for this study is comment data collected from Indonesian e-wallet users conducting payment on OVO, Dana, Doku, and LinkAja facilitated by google app store.

User comment data was taken using the web scraping method. The data was taken in January-May 2023. After preprocessing the final data are 11267 with 6349 negative labels and 4918 positive labels. Labeling of data use a star scale that has been pinned by the user, 1-3 stars are labelled negative and 4-5 stars are labelled positive.

The labeling results are tested using TF-IDF, SVM-TF-IDF, Multinomial Naive Bayes, Bagging with Multinomial Naive Bayes, and Random Forest algorithm. For evaluating the algorithms, performance of these algorithms is measured using Precision, Recall, F1-Score, and Support. The accuracy of the algorithms is also evaluated using train accuracy score, test accuracy score, train ROC-AUC Score, test ROC-AUC score, area under precision-recall curve, and area under ROC-AUC.

## IV. Result and Discussion

The main process of this study is pre-modelling, building SVM model, calculating accuracy level of the model, and verifying the accuracy of the model. Pre-modelling process involve collecting data using web scraping and labeling process naming positive and negative labels. Furthermore, word vectorization process is conducted with TF-IDF to determine a word weight value.

### *A. Labeling Process and Wordcloud*

This study collects 11267 rows × 8 columns data. The labeling process delivers 6349 reviews as negative (0) labels and 4918 reviews as positive (1) labels. The result of labeling process presents on Figure 1.

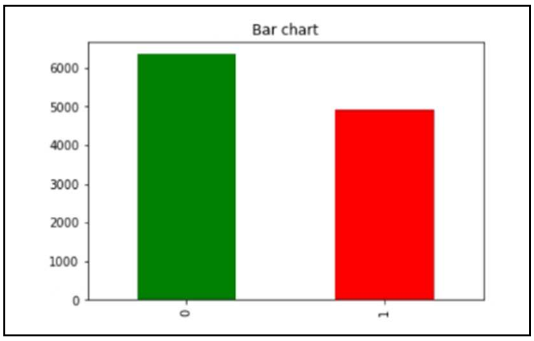


Fig. 1. Result of Labeling Process

The result on Figure 1 states that the reviews of product with negative rating is greater than positive rating. This means that the number of e-wallet users feeling dissatisfied with the transactions and giving negative feedback is greater than they are satisfy with the transactions and giving positive feedback.

In addition, Figure 2 and 3 present word clouds for negative and positive reviews.


Fig. 2. Negative Wordcloud

The negative wordcloud in Figure 2 presents the words that come up most frequently in negative reviews. The first meaningful word is “me”. It means that the problem in users is individual problem in their transaction. The second meaningful word is “dana”. The word “Dana” in e-wallet context relates to an operator of e-wallet in Indonesia or the value of the money. It means that the users have negative sentiment relating to the service or the product of “Dana” or they have problem relating to their money value in their transactions. The others meaningful words are “transaction”, “application”, and “balance”. It means the users have problem on their transaction, e-wallet application, and balance in e-wallet.

The result of positive wordcloud presents on Figure 3. The meaningful words on Figure 3 including “top”, “dana”, “me”, "this application," "excellent", "pay the bill", "CS", "very helpful", "doku application" and "fast". The words "top", "excellent", and "fast" are the user appreciation of the service and quality of e-wallet their use. The word "me" means that the positive feeling and appreciation are an individual expression in using e-wallet. Furthermore, the words doku application and dana relate to the popular e-wallet used by Indonesian consumers. The last words naming "pay the bill", "CS", and "very helpful" relate to the positive feeling of e-wallet user regarding to the service and assistance of the customer service.


Fig. 3. Positive Wordcloud

## B. Word vectorization

Word vectorization is conducted after labeling process to get an integer representation of a word or word weighting. Word vectorization uses TF-IDF (term frequency–inverse document) to weight the words. There is the code of this process:

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(decode_error='replace', encoding='utf-8')
```

### Word Vectorization of TF-IDF

The result of word vectorization with TF-IDF must be trained and tested and the output of this process is in the matrix:

(10140, 9698)
(1127, 9698)

The matrix has meaning:

- The number 10140 on first row and column of matrix indicates the number of documents in the train dataset (X_train). Each document in X_train will be represented by one row in the matrix.
- The number 9698 on first row and second column of matrix indicates the number of unique features or words learned by the vectorizer from the training dataset. Each of these features represents a word in the dictionary formed by the vectorizer.
- The number 1127 on second row and first column of matrix indicates the number of documents in the test dataset (X_test). Each document in X_test will be represented by one row in the results matrix.
- The number 9698 on second row and column of matrix indicates the same number of unique features or words that the vectorizer has learned from the training dataset. Even though it is possible that some words in X_test are not in the dictionary that the vectorizer learns during training, the number of features remains the same for consistency purposes.

## C. Evaluating Performance of Supervised Learning Models

### TF-IDF Model

The calculating of confusion matrix of TF-IDF on train and test processes is presented on Table 2. A confusion matrix usually consists of four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

TABLE II. CONFUSION MATRIX OF TF-IDF

| Train Process | | Predicted Values | |
|---|---|---|---|
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 3688 TP | 5 FP |
| | ***Negative*** | 2 FN | 4755 TN |
| Test Process | | **Predicted Values** | |
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 1016 TP | 113 FP |
| | ***Negative*** | 212 FN | 1476 TN |

Note: TP = how much of the data is actual positive, and the model also predicts positive. TN = how much of the data is actual negative, and the model also predicts negative. FP = how much of the data is actually negative, but the model predicts positive. FN = how much data is actual positive, but the model predicts negative.

The Confusion Matrix of train process on Table 2 shows that the model has a very low number of FP (5), which indicates that the model has a very good performance in classifying negatives (first class). The number of FN (2) is also low, indicating good performance in classifying some positive cases (second class).

Interpretation of Confusion matrix results should always be considered in the context of the specific application and purpose. Evaluation of model performance should be analyzed by considering the relative impact of FP and FN on the objectives.

The Confusion matrix of test process on Table 2 shows that the model has a slightly higher number of FP (113). The value indicates that the model has less difficulty in classifying negatives (first class). The number of FN (212) is also slightly higher. The value indicates slightly lower performance in classifying some positive cases (second class).

Based on the result of Confusion Matrix, performance evaluation matrix can be calculated using equation 4 – 7 such as accuracy, precision, sensitivity (recall), specificity, and others. The model performance evaluation matrix presents on Table 3.

TABLE III. PERFORMANCE EVALUATION MATRIX FOR TF-IDF MODEL

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| **0** | 0.87 | 0.91 | 0.89 | 635 |
| **1** | 0.87 | 0.83 | 0.85 | 492 |
| **accuracy** | 0.87 | 1127 | | |
| **macro avg** | 0.87 | 0.87 | 0.87 | 1127 |
| **weighted avg** | 0.87 | 0.87 | 0.87 | 1127 |

### SVM-TF-IDF Model

Calculating the SVM - TF-IDF model is started with creating object of TF-IDF vectorizer. The calculating of confusion matrix of SVM-TF-IDF on train and test process is presented on Table 4.

TABLE IV. CONFUSION MATRIX OF SVM-TF-IDF (TRAIN PROCESS)

| Train Process | | Predicted Values | |
|---|---|---|---|
| | | *Positive* | *Negative* |
| **Actual Values** | ***Positive*** | 3439 TP | 44 FP |
| | ***Negative*** | 251 FN | 4716 TN |
| Test Process | | **Predicted Values** | |
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 1006TP | 89 FP |
| | ***Negative*** | 222 FN | 1500TN |

The Confusion matrix of train process on Table 4 shows that the model has a relatively low number of False Positives (44), which means the ability of model to classify negatives (first class) is quite good. However, the number of False Negatives (251) is slightly higher, it means that the model has difficulty to classify some positive cases (second class).

The confusion matrix of test process on Table 4 shows that the model has a fairly low number of False Positives (89). The model has a good ability to classify negatives (first class). However, the number of False Negatives (222) is slightly higher. It means that the model has difficulty to classify some positive cases (second class). Based on the Confusion matrix, performance evaluation matrix can be calculated and the result is presented on Table 5.

TABLE V. PERFORMANCE EVALUATION MATRIX FOR SVM-TF-IDF MODEL

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| **0** | 0.87 | 0.94 | 0.91 | 1589 |
| **1** | 0.92 | 0.82 | 0.87 | 1228 |
| **accuracy** | 0.89 | 2817 | | |
| **macro avg** | 0.89 | 0.88 | 0.89 | 2817 |
| **weighted avg** | 0.89 | 0.89 | 0.89 | 2817 |

According to Table 5, the accuracy score of the model is 0.89 (89%). The result has meaning 89% of all predictions are correct. The precision score of class 0 (negative label) is 0.87. It means that 87% of predictions classified as negative by the model are actually negative cases. Meanwhile, the score of class 1 (positive label) is 0.92 (92%). It means that 92% of positive predictions are actually positive cases.

Furthermore, recall score for class 0 is 0.94. It means that 94% of the negative data was identified correctly by the model. For class 1, recall score is 0.82, which means 82% of positive data was identified. F1-score for class 0 is 0.91 and for class 1 is 0.87. The macro average and weighted average of precision, recall, and F1-scores are around 0.89.

### Multinomial Naive Bayes Model

The calculating of confusion matrix of Multinomial Naive Bayes model are presented on Table 6 .

TABLE VI. CONFUSION MATRIX OF MULTINOMIAL NAIVE BAYES

| **Train Process** | | **Predicted Values** | |
|---|---|---|---|
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 3055TP | 94 FP |
| | ***Negative*** | 635 FN | 4666 TN |
| **Test Process** | | **Predicted Values** | |
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 919 TP | 57 FP |
| | ***Negative*** | 309 FN | 1532 TN |

Table 7 presents the performance evaluation matrix for Multinomial Naive Bayes model.

TABLE VII. PERFORMANCE EVALUATION MATRIX FOR MULTINOMIAL NAIVE BAYES MODEL

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| **0** | 0.83 | 0.96 | 0.89 | 1589 |
| **1** | 0.94 | 0.75 | 0.83 | 1228 |
| **accuracy** | 0.87 | 2817 | | |
| **macro avg** | 0.89 | 0.86 | 0.86 | 2817 |
| **weighted avg** | 0.88 | 0.87 | 0.87 | 2817 |

### Bagging with Multinomial Naïve Bayes

The calculating of confusion matrix of Bagging with MultinomialNB model on train and test processes are presented on Table 8.

TABLE VIII. CONFUSION MATRIX OF BAGGING WITH MULTINOMIALNB

| **Train Process** | | **Predicted Values** | |
|---|---|---|---|
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 3041 TP | 99 FP |
| | ***Negative*** | 649 FN | 4661 TN |
| **Test Process** | | **Predicted Values** | |
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 921 TP | 55 FP |
| | ***Negative*** | 307 FN | 1534 TN |

The performance evaluation matrix for Bagging with MultinomialNB model is presented on Table 9.

TABLE IX. PERFORMANCE EVALUATION MATRIX FOR BAGGING WITH MULTINOMIALNB MODEL

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| **0** | 0.83 | 0.97 | 0.89 | 1589 |
| **1** | 0.94 | 0.75 | 0.84 | 1228 |
| **accuracy** | 0.87 | 2817 | | |
| **macro avg** | 0.89 | 0.86 | 0.87 | 2817 |
| **weighted avg** | 0.88 | 0.87 | 0.87 | 2817 |

### Random Forest

The calculating of confusion matrix of Random Forest model on train and test processes are presented on Table 10.

TABLE X. CONFUSION MATRIX OF RANDOM FOREST

| **Train Process** | | **Predicted Values** | |
|---|---|---|---|
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 3689 TP | 6 FP |
| | ***Negative*** | 1 FN | 4754 TN |
| **Test Process** | | **Predicted Values** | |
| | | ***Positive*** | ***Negative*** |
| **Actual Values** | ***Positive*** | 977 TP | 107 FP |
| | ***Negative*** | 251 FN | 1482 TN |

The performance evaluation matrix for Random Forest model is presented on Table 11.

TABLE XI. PERFORMANCE EVALUATION MATRIX FOR RANDOM FOREST MODEL

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| **0** | 0.86 | 0.93 | 0.89 | 1589 |
| **1** | 0.90 | 0.80 | 0.85 | 1228 |
| **accuracy** | 0.87 | 2817 | | |
| **macro avg** | 0.88 | 0.86 | 0.87 | 2817 |
| **weighted avg** | 0.88 | 0.87 | 0.87 | 2817 |

### *D. Evaluating the accuracy of of Supervised Learning Models*

The calculation of the accuracy of supervised learning models including SVM-TF-IDF, Multinomial Naïve Bayes, Bagging with MultinomialNB, and Random Forest are presented on Table 12.

TABLE XII. EVALUATING THE ACCURACY OF SUPERVISED LEARNING MODELS

| Accuracy score | SVM | Mult. Naive Bayes | Bagging with Mult. NB | Random Forest |
|---|---|---|---|---|
| **Train accuracy** | 0.965 | 0.914 | 0.911 | 0.999 |
| **Test accuracy** | 0.890 | 0.870 | 0.871 | 0.873 |
| **Train ROC-AUC** | 0.995 | 0.958 | 0.956 | 0.999 |
| **Test ROC-AUC** | 0.930 | 0.927 | 0.927 | 0.931 |
| **Area under Precision-Recall curve** | 0.866 | 0.834 | 0.836 | 0.845 |
| **Area under ROC-AUC** | 0.927 | 0.933 | 0.933 | 0.922 |

According to Table 12, the training accuracy score of SVM-TF-IDF is about 0.965, which means about 96.5% correct predictions on the training data. The test accuracy score is about 0.890, which means about 89.0% correct predictions on the test data. A ROC-AUC score closes to 1 indicates good performance in distinguishing classes on the training data. The test ROC-AUC score is about 0.930, which indicates that the model has good performance in distinguishing classes on the test data.

## V. CONCLUSION

This study has established a classification model for collecting data of comments from E-wallet users in Indonesia to classify the comments regarding application services so that customer complaints can be minimized and can be responded to immediately.

It is proven that the labelling process at the preprocessing stage which has been carried out based on a score of 1-5 given by the user can be used for classification.

Examining on the SVM, Multinomial Naive Bayes, Bagging with MultinomialNB, and Random Forest algorithms obtained a high level of accuracy with a value of 0.87.

The area under the Precision-Recall curve above 0.80 indicates that the classification model determined at the testing stage has good performance in separating positive and negative classes. This means that the models have high precision and good recall at different threshold levels.

The ROC- AUC score for SVM, Multinomial Naive Bayes, Bagging with MultinomialNB, and Random Forest is more than 0.90, indicating that the classification models have good performance in distinguishing positive and negative classes and have a high True Positive Rate compared to False Positive Rate at different threshold levels.

It is concluded that the supervised learning algorithms naming SVM, Multinomial Naive Bayes, Bagging with MultinomialNB, and Random Forest can be used to classify E-wallet comment data in Indonesia.

## ACKNOWLEDGMENT

The authors gratefully acknowledges the Directorate General of Higher Education, Research, and Technology of the Ministry of Education, Culture, Research, and Technology Indonesia, which provides the research grant No. 076/E5/PG.02.00.PL/2023.

## REFERENCES

[1] M. S. Rosli, N. S. Saleh, A. Md. Ali, and S. Abu Bakar, "Factors Determining the Acceptance of E-Wallet among Gen Z from the Lens of the Extended Technology Acceptance Model," Sustain., vol. 15, no. 7, pp. 1–23, 2023, doi: 10.3390/su15075752.

[2] H. M. Aji, I. Berakon, and A. F. Riza, "The effects of subjective norm and knowledge about riba on intention to use e-money in Indonesia," J. Islam. Mark., vol. 12, no. 6, pp. 1180–1196, 2020, doi: 10.1108/JIMA-10-2019-0203.

[3] A. Sikri, S. Dalal, N. . Singh, and D. Le, "Mapping of e - Wallets With Features," Cyber Secur. Parallel Distrib. Comput., no. May, pp. 245-261, 2019, doi: 10.1002/9781119488330.ch16.

[4] A. Daragmeh, J. Sági, and Z. Zéman, "Continuous intention to use e-wallet in the context of the covid-19 pandemic: Integrating the health belief model (hbm) and technology continuous theory (tct)," J. Open Innov. Technol. Mark. Complex., vol. 7, no. 2, 2021, doi: 10.3390/joitmc7020132.

[5] M. Yang, A. Al Mamun, M. Mohiuddin, N. C. Nawi, and N. R. Zainol, "Cashless transactions: A study on intention and adoption of e-wallets," Sustain., vol. 13, no. 2, pp. 1–18, 2021, doi: 10.3390/su13020831.

[6] A. Zhavoronok, O. Popelo, R. Shchur, N. Ostrovska, and N. Kordzaia, "The Role of Digital Technologies in the Transformation of Regional Models of Households' Financial Behavior in the Conditions of the National Innovative Economy Development," Ing. des Syst. d'Information, vol. 27, no. 4, pp. 613–620, 2022, doi: 10.18280/isi.270411.

[7] B. Shaw and A. Kesharwani, "Moderating Effect of Smartphone Addiction on Mobile Wallet Payment Adoption," J. Internet Commer., vol. 18, no. 3, pp. 291–309, 2019, doi: 10.1080/15332861.2019.1620045.

[8] D. Chatterjee and K. Bolar, "Determinants of Mobile Wallet Intentions to Use: The Mental Cost Perspective," Int. J. Hum. Comput. Interact., vol. 35, no. 10, pp. 859–869, 2019, doi: 10.1080/10447318.2018.1505697.

[9] K.-L. Chan, C.-M. Leong, and B. L. C. Yiong, "Sharing economy through e-wallet: Understanding the Determinants of User Intention in Malaysia," J. Mark. Adv. Pract., vol. 2, no. 2, pp. 1–18, 2020, [Online]. Available: jmaap.org/.

[10] D. Chawla and H. Joshi, "Consumer attitude and intention to adopt mobile wallet in India – An empirical study," Int. J. Bank Mark., vol. 37, no. 7, pp. 1590–1618, 2019, doi: 10.1108/IJBM-09-2018-0256.

[11] D. W. Caves, J. A. Herriges, and R. J. Windle, "Customer Demand for Service Reliability in the Electric Power Industry: a Synthesis of the Outage Cost Literature," Bull. Econ. Res., vol. 42, no. 2, pp. 79–121, 1990, doi: 10.1111/j.1467-8586.1990.tb00294.x.

[12] N. Pappas and A. Popescu-Belis, "Adaptive sentiment-aware one-class collaborative filtering," Expert Syst. Appl., vol. 43, pp. 23–41, 2016, doi: 10.1016/j.eswa.2015.08.035.

[13] C. S. Tucker and H. M. Kim, "Trend mining for predictive product design," J. Mech. Des. Trans. ASME, vol. 133, no. 11, 2011, doi: 10.1115/1.4004987.

[14] C. N. Dellarocas, "The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms," SSRN Electron. J., no. March, 2005, doi: 10.2139/ssrn.393042.

[15] K. Bastani, H. Namavari, and J. Shaffer, "Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints," Expert Syst. Appl., vol. 127, pp. 256–271, 2019, doi: 10.1016/j.eswa.2019.03.001.

[16] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," Procedia Comput. Sci., vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.

[17] A. Mee, E. Homapour, F. Chiclana, and O. Engel, "Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit[Formula presented]," Knowledge-Based Syst., vol. 228, p. 107238, 2021, doi: 10.1016/j.knosys.2021.107238.

[18] A. F. Pathan and C. Prakash, "Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory," J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.09.011.

J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," J. Thorac. Oncol., vol. 5, no. 9, pp. 1315–1316, 2010, doi: 10.1097/JTO.0b013e3181ec173d.

## 17% Overall Similarity

Top sources found in the following databases:

- 16% Internet database
- 9% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| | | |
|---|---|---|
| 1 | dspace.daffodilvarsity.edu.bd:8080<br>Internet | 3% |
| 2 | dspace.cvut.cz<br>Internet | 3% |
| 3 | jurnal-ppi.kominfo.go.id<br>Internet | 2% |
| 4 | researchbank.swinburne.edu.au<br>Internet | 1% |
| 5 | mdpi.com<br>Internet | 1% |
| 6 | datascience.stackexchange.com<br>Internet | <1% |
| 7 | repository.unika.ac.id<br>Internet | <1% |
| 8 | ispe.org<br>Internet | <1% |
| 9 | fpsp.edu.rs<br>Internet | <1% |

| | | |
|---|---|---|
| 10 | biorxiv.org<br>Internet | <1% |
| 11 | export.arxiv.org<br>Internet | <1% |
| 12 | Jannat, Fatima. "Novice Programmers' Emotion and Competency Asse...<br>Publication | <1% |
| 13 | Ton Duc Thang University<br>Publication | <1% |
| 14 | dokumen.pub<br>Internet | <1% |
| 15 | icicelb.org<br>Internet | <1% |
| 16 | blog.csdn.net<br>Internet | <1% |
| 17 | codeavail.com<br>Internet | <1% |
| 18 | researchgate.net<br>Internet | <1% |
| 19 | jurnal.iaii.or.id<br>Internet | <1% |
| 20 | docplayer.net<br>Internet | <1% |
| 21 | duo.uio.no<br>Internet | <1% |

| | | |
|---|---|---|
| 22 | universalai.in<br>Internet | <1% |
| 23 | bip.us.edu.pl<br>Internet | <1% |
| 24 | journals.plos.org<br>Internet | <1% |
| 25 | ojs.wiserpub.com<br>Internet | <1% |
| 26 | open.metu.edu.tr<br>Internet | <1% |
| 27 | fatalerrors.org<br>Internet | <1% |
| 28 | frontiersin.org<br>Internet | <1% |
| 29 | Abedzadeh, Najmeh. "Implementing a New Algorithm to Balance and C...<br>Publication | <1% |
| 30 | Hanoi National University of Education<br>Publication | <1% |
| 31 | doaj.org<br>Internet | <1% |