Khazanah Journal of Management Vol (1) No (1) 25-41 e-ISSN: xxxx-xxxx
DOI: https://doi.org/10.xxxxx/ Management.v1i1.4
https://journal.azkia.id/index.php/Management

Zakat Recipient Recommendation System Based on Machine Learning Approach

Ayuni Tia Sari [1*], Tito Bastian[1]
[1]Department of Informatics, UIN Sunan Gunung Djati Bandung Bandung, Indonesia
*Corresponding Author Email: ayunitiasari@gmail.com

Received: March 9, 2026; Revised: March 29, 2026, Accepted: April 2, 2026

ABSTRACT

Background: Zakat is an Islamic obligation aimed at reducing economic inequality and providing assistance to those in need. However, the distribution of zakat is often complex and requires an efficient and objective decision-making process. The development of information technology, particularly machine learning, offers opportunities to improve the effectiveness and accuracy of zakat recipient selection.

Purpose: This study aims to examine the application of a machine learning-based recommendation system in determining zakat recipients and to highlight its potential in supporting objective, accurate, and efficient zakat distribution.

Methods: This study employs a literature review approach by analyzing previous studies related to zakat distribution, decision support systems, data mining, and machine learning methods, particularly recommendation systems and the Random Forest algorithm.

Results: The findings indicate that machine learning-based approaches can assist in identifying eligible zakat recipients more objectively and accurately. The use of recommendation systems and Random Forest algorithms reduces subjectivity in decision-making, accelerates the selection process, and minimizes errors in zakat distribution. In addition, system-based approaches can improve the overall efficiency of zakat management.

Conclusions: The implementation of machine learning in zakat recipient recommendation systems has significant potential to improve the quality of zakat distribution. Nevertheless, the application of such technology should remain aligned with the religious and social values of zakat to ensure that its spirit and objectives are preserved.

Research Contribution: This study contributes to the development of technology-based zakat management by providing a conceptual understanding of how machine learning can be integrated into zakat recipient selection systems to support fairer, more effective, and more accountable distribution.

Keywords: Zakat, Mustahik, Random Forest.

INTRODUCTION

Zakat is a religious obligation in Islam that encourages Muslims to share their wealth with those in need. As a social and religious obligation, zakat aims to reduce economic inequality and provide assistance to the less fortunate. However, with the increasing number of zakat recipients and the complexity of the socio-economic conditions of society, it is crucial for zakat organizations to ensure their zakat funds are distributed effectively and efficiently (Hidayat & Mukhlisin, 2020).

Zakat has four important meanings. First, zakat is related to the blessing of wealth and sustenance. By paying zakat, a person hopes for blessings from Allah SWT, which will enable their wealth to provide greater benefits. Second, zakat contributes to the growth and development of society. Through zakat, people make positive contributions to help those who are entitled to it, especially in productive ways such as business capital or education, so they can develop and improve their standard of living. Third, zakat involves the purification of wealth. This means cleansing it of negative traits such as greed and injustice. By paying zakat, a person cleanses themselves of stinginess and materialism, which can undermine morals and spirituality. Finally, zakat is related to order. Through zakat, individuals create abundance in society by ensuring basic needs are met and reducing social inequality (Welinda et al., 2016). Zakat provides benefits and well-being for individuals and society as a whole. By paying zakat, individuals hope to receive blessings, contribute to societal growth, cleanse their wealth of negative traits, and create abundance for themselves and others (Ahmad et al., 2020).

In the ever-evolving digital era, machine learning approaches have become an interesting field for application in various contexts, including zakat management. Machine learning-based recommendation systems have emerged as a potential tool to assist zakat organizations in identifying the neediest zakat recipients (Rahman, 2021).

At the At-Taqwa Mosque in Gadel Village, Indramayu Regency, one of the frequently encountered problems is the manual selection process for eligible recipients (mustahik). This leads to a lengthy selection process and the potential for miscalculations, reducing the accuracy of the selection results. Furthermore, the lack of access to information on the list of zakat recipients also results in some recipients receiving assistance more than once a year. In an effort to address this issue, the At-Taqwa Mosque has established criteria for selecting eligible recipients, including residence status, income, employment status, number of dependents, and family vehicle ownership. Zakat is then distributed to 30 eligible recipients selected by the mosque's management committee annually during the holy month of Ramadan. Therefore, a system was designed to support mosque management in making decisions regarding zakat recipient recommendations for eligible recipients. This system is expected to make the selection process more efficient and transparent, and to avoid miscalculations and repeated donations to the same eligible recipients.

This research methodology includes three main components: data description, implementation procedures, and the application of machine learning algorithms. The data used comes from the 2019 Southeast Sulawesi Statistics Center, totaling 8,710 data points. The dataset contains several attributes representing household socio-economic conditions, namely wall type, floor type, floor ownership, drinking water source, lighting source, fuel, and number of household members. All these attributes are used as the basis for the analysis process to more objectively identify the characteristics of potential zakat recipients.

The implementation procedure for this research refers to the CRISP-DM (Cross Industry Standard Process for Data Mining) model, which consists of six stages: business understanding, data understanding, data preparation, modeling, validation, and evaluation (Navisa et al., 2021). In the business understanding stage, the research focuses on formulating the objectives and needs of stakeholders in determining mustahik. Next, the data understanding stage explores the characteristics, quality, and relationships between variables in the dataset. The data preparation stage includes attribute selection, handling missing values, transformation, and data adjustment to ensure it is ready for use in the modeling process. Following this, the modeling stage is carried out by applying an appropriate machine learning algorithm. The resulting model is then tested in the validation stage to measure its performance, accuracy, and reliability. The final stage, evaluation, aims to assess the suitability of the model results for the research objectives to determine the feasibility of its implementation in supporting the zakat recipient recommendation system. See Figure 1.

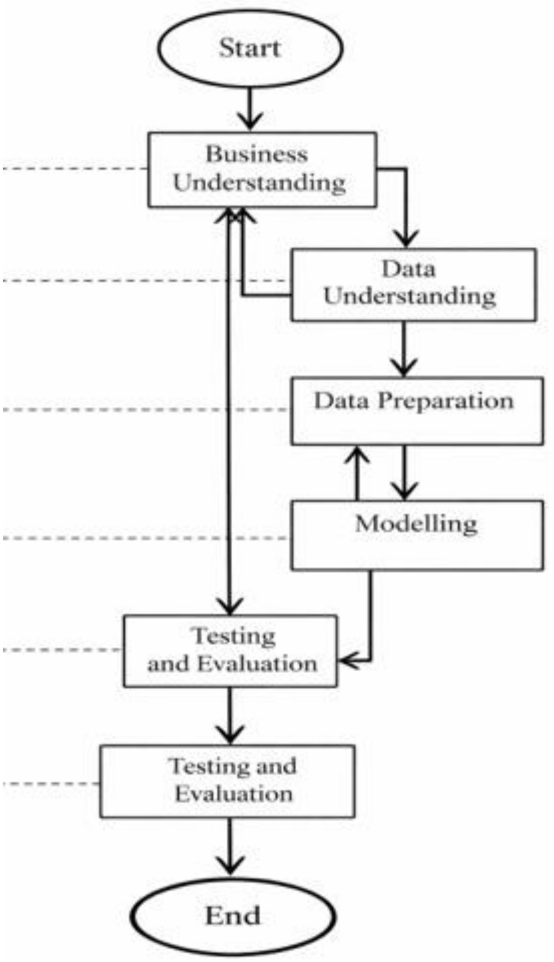

Figure 1. Implementation Procedure

The Random Forest algorithm is a machine learning method belonging to the Classification and Regression Tree (CART) group. This algorithm works using the concept of a decision tree, but instead of forming a single tree, it creates a collection of trees resembling a forest. The Random Forest model is formed through two main mechanisms: bootstrap aggregation and random feature selection. In the bootstrap aggregation stage, training data is randomly sampled with replacement to form each tree. Next, in the random feature selection stage, only a subset of features is randomly selected for each tree. This ensures that each tree has distinct characteristics, so that when all the predictions from these trees are combined, the model can produce more accurate and stable predictions, while reducing the risk of overfitting, which is common with single decision trees (Danoedoro et al., 2022).

In the decision tree formation process in the Random Forest algorithm, the first step is to check whether all data labels are identical. If all labels are identical, the tree will immediately form leaves with those label values without the need for further separation. However, if the data labels are still diverse, then the information value is calculated using a specific formula according to the decision tree method, such as ID3, C4.5, or the Gini Index. This calculation is related to the concept of entropy, which is a measure of the level of uncertainty or diversity in a data set. If an attribute has discrete values, then the data will be separated based on each value contained in that attribute to obtain more homogeneous branches. Next, each attribute is evaluated based on the information value it produces, where the attribute with the most optimal separation value will be selected as the tree branch. For continuous-valued attributes, the data is first sorted, then the split point is determined from the average of two adjacent values, then each candidate split point is calculated for information value to obtain the best separation. After the tree branches are formed, the same process is repeated recursively until it reaches a certain limit or until a leaf is formed with the majority value of the data. With this mechanism, Random Forest is able to build a robust, adaptive, and effective classification model for various types of data (Wuisan, 2020). See figure 2.

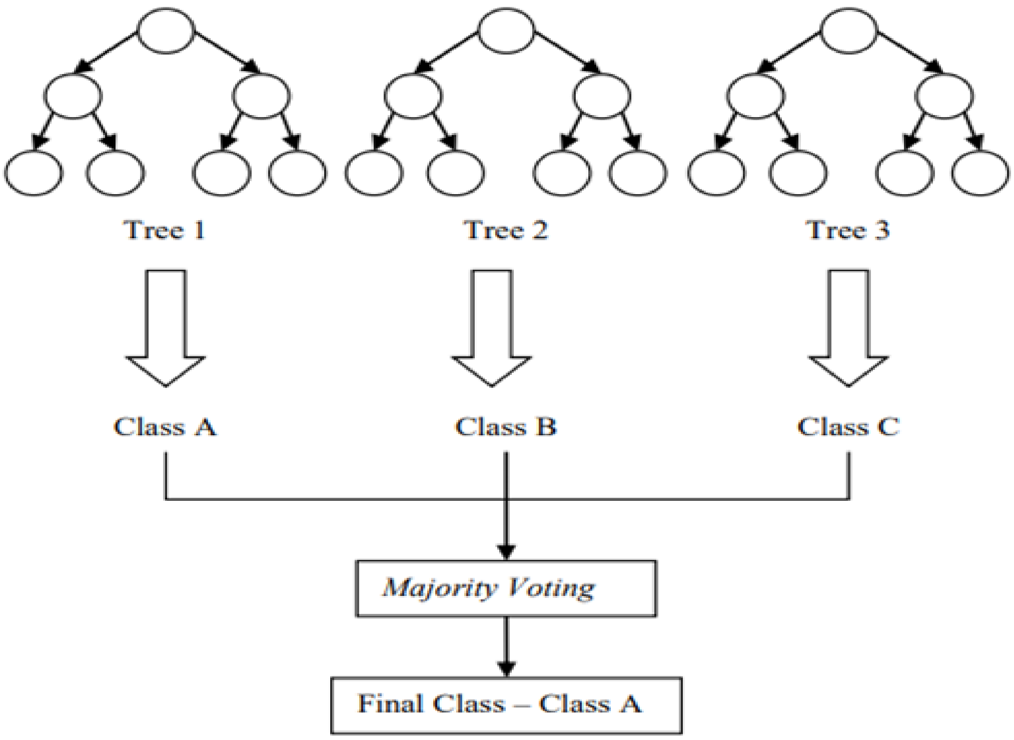

Figure 2 Random Forest Algorithm

RELATED WORKS

A common problem during zakat distribution is the inaccurate determination of eligible recipients (mustahiq), which results in errors in aid distribution. This can result in aid not reaching those in need. The method used in this study is the K-Means algorithm, implemented through a GUI-assisted Matlab application. The method's stages include data collection, data processing, data clustering, and analysis. This means that the average value of each poverty variable is high, so the government should prioritize this area for zakat distribution or other government aid programs compared to clusters one and two (Arif & Christyanti, 2022). Previous research has identified that the manual process for determining zakat recipient eligibility at the Kendari City National Zakat Agency is complex and ineffective. This study introduces the Fuzzy C-Means (FCM) method as a more efficient and effective alternative. The results show that the FCM method successfully produces decisions similar to the results of the manual determination by the Kendari City National Zakat Agency through deliberation. However, the FCM method offers advantages in efficiency and time effectiveness in determining zakat recipients (Welinda et al., 2016). In this study (Asa, 2019), the authors used data mining techniques utilizing the C4.5 algorithm to classify data on the distribution of community funds at the Agam Regency BAZNAS. The data was classified into two categories: distribution (consumptive) and utilization (productive). Classification was performed using RapidMiner Studio software, and the results demonstrated a high level of accuracy. The accuracy of the results achieved in this study was based on the selection of appropriate attributes and the transformation of data into selected attributes. This enabled the C4.5 algorithm to generate rules that could assist the Agam Regency BAZNAS in classifying the types of zakat to be distributed more efficiently and quickly.

The Weighted Product method was implemented in decision-making regarding zakat recipients. The calculation was performed by multiplying the attribute rating of each alternative by the weight of the relevant attribute. Before multiplication, the rating of each attribute must be raised to the power of the weight of the relevant attribute. The final result of the calculation using the Weighted Product method is a vector value V. The alternative with the highest V value will be prioritized as a zakat recipient. Using this method allows for a more structured and objective selection process in determining zakat recipients. This approach helps produce more accurate and transparent decisions in zakat distribution, considering various relevant criteria (Eliyen & Efendi, 2019).

In LAZ Al Kahfi Peduli, zakat distribution is carried out through subjective and manual assessments of zakat recipient criteria. A Decision Support System (DSS) can help more objectively determine who is entitled to and prioritized to receive zakat. The method used in this DSS is the Simple Multi-Attribute Rating Technique (SMART). By implementing the SMART method, decision-makers can evaluate various relevant attributes in prioritizing zakat recipients. The use of the SMART method in the DSS for zakat distribution at LAZ Al Kahfi Peduli will provide significant benefits in improving more accurate and equitable decision-making (Setiyaningsih et al., 2020).

Based on this research, the random forest classification algorithm has the highest accuracy value, at 97.88%. This study also used naive Bayes and support vector machine classification algorithms in comparison with random forest. The experiment was conducted using 10-fold cross-validation and accuracy measurements were performed using Accuracy, F-Measure, Recall, Precision, and ROC metrics. The results showed that the random forest algorithm performed better than other algorithms (Apriliah et al., 2021).

Based on this research, it can be concluded that the analysis of Twitter user sentiment regarding PSBB yielded less than satisfactory results. The Random Forest algorithm had an accuracy of 0.578, while the Support Vector Machine algorithm had an accuracy of 0.557. This indicates that both algorithms were unable to effectively classify user sentiment regarding PSBB (Adrian et al., 2021).

The results of this research indicate that the Random Forest algorithm has a higher level of accuracy in analyzing problematic loans and non-problematic debtors. Based on this research, the Random Forest algorithm achieved an accuracy level of 87.88% (Zailani & Hanun, 2020).

The results of the study, which implemented Random Forest to classify Palembang songket motifs using the Scale-Invariant Feature Transform (SIFT), can be concluded as follows: The Cantik Manis songket motif achieved the best accuracy per class, with a percentage of 100%. The Bunga Cina songket motif and the Pulir songket motif achieved 91.11% accuracy per class. The highest precision values were found for the Bunga Cina songket motif and the Cantik Manis songket motif, both achieving 100%. Meanwhile, for the Pulir songket motif, the precision value reached 78.94%. The highest recall values were found for the Cantik Manis songket motif and the Pulir songket motif, both reaching 100%. Meanwhile, the Bunga Cina songket motif had a recall value of 69.23%. Thus, the research results indicate that implementing Random Forest using the Scale-Invariant Feature Transform (SIFT) method is capable of providing good levels of accuracy, precision, and recall in classifying Palembang songket motifs (Devella et al., 2020).

Based on the research results, it can be concluded that a classification system has been successfully developed to predict consumer acceptance of cars. The method used in this system is Random Forest with a gain ratio algorithm. The resulting decision tree can vary depending on the selected attributes and the random data collection. In the decision tree, the variable placed as the root node has the most significant influence on consumer acceptance of cars. This means that this variable is a key factor influencing consumer decisions. The classification system can use information from this variable to identify patterns and relationships between other variables that impact consumer acceptance. By applying the Random Forest method and the gain ratio algorithm, this classification system can provide accurate and reliable results in predicting consumer acceptance levels (Nugroho & Emiliyawati, 2017).

From the results of the research conducted, it can be concluded that the application of resampling methods, such as Random Over-Under Sampling, to the Random Forest algorithm can improve classification accuracy in credit assessment. Other research also shows that the use of resampling techniques, such as SMOTE, can improve classification performance in cases of data imbalance. Thus, approaches involving data-level resampling are an effective solution to address the problem of class imbalance in credit assessment. In this context, resampling methods, such as Random Over-Under Sampling or SMOTE, can help generate a more balanced dataset between different classes, thereby improving the performance of classification algorithms, such as Random Forest. By using this approach, classification systems can overcome the challenges arising from class imbalance in credit assessment data and produce more accurate and reliable predictions (Syukron & Subekti, 2018).

A classification study using the multiclass Random Forest method to classify education quality based on National Examination (UNBK) results and accreditation scores per instrument item at the high school level in 2018 concluded that the model achieved an overall accuracy rate of 83.49%. The results showed that the Random Forest classification model successfully predicted with significant accuracy in classifying education quality based on UNBK data and accreditation scores. The 83.49% accuracy rate indicates that the model provided quite good results in differentiating education quality based on observed factors. By using the multiclass Random Forest method, this study presents an effective approach to classifying education quality. The model combines information from UNBK results and accreditation scores per instrument item to produce accurate predictions for determining the level of education quality at the high school level. Thus, this study makes an important contribution to understanding and monitoring education quality using the Random Forest classification method and demonstrates a reliable level of accuracy in classifying education quality based on available data. (Ramadhan et al., 2019) In this study, to measure the severity of apple leaf disease using the Random Forest classification method, a dataset of 467 apple leaf images was used. The classification results using Random Forest showed that in the training process, the model achieved the highest accuracy of 100%. Meanwhile, in the testing process, the model achieved the highest accuracy of 75.3191%. In this study, Random Forest was used as a classification algorithm to predict the severity of disease in apple leaves based on the given image. The training results that achieved 100% accuracy indicate that the model was able to perfectly learn the patterns and characteristics of the training data used. However, in the testing process, the accuracy achieved was slightly lower, namely 75.3191%. This may be due to differences between the training data and the testing data, as well as variations in the characteristics and conditions of the apple leaves tested. Although the testing accuracy value did not reach the same level as the training, the results still showed a fairly good level of accuracy in classifying and predicting the severity of disease in apple leaves. Thus, this study contributes to the development of a classification method using Random Forest for measuring disease severity in apple leaves. Despite the differences between the training and testing results, the model was still able to provide accurate classification results in the context studied (Ratnawati & Sulistyaningrum, 2019).

Based on the review of previous research, the novelty of this research lies in the application of the Random Forest algorithm in the zakat recipient recommendation system in the context of determining mustahik, because previous studies in the field of zakat are still dominated by the K-Means, Fuzzy C-Means, C4.5, Weighted Product, and SMART approaches, while Random Forest is more widely applied to other domains such as health, credit, sentiment, image, and education classification. Thus, this research not only offers a more modern approach based on machine learning to improve objectivity, accuracy, and consistency in determining zakat recipients, but also presents a contribution in the form of combining the concept of aid recipient recommendations with the classification capabilities of Random Forest which have proven superior in various previous studies. Therefore, this research is expected to be a new alternative in the development of a zakat decision support system that is more adaptive, scalable, and relevant to the needs of digital transformation of zakat management today.

RESULTS AND DISCUSSION

Result

This study used the random forest classification method to analyze zakat recipients. After analyzing the modeling implementation using the random forest algorithm, the following results were obtained:

Data Processing

After processing the data, the results can be seen from several descriptive statistics of this dataset, and it can be concluded that all data has no anomalies, so it can be continued for the training and testing process. See Table 1.

Table 1 Data Processing

Statistics	Floor area (m2)	Type of wall	Floor Type	FTBAB Ownership	Source of drinking water	Source of light	Fuel	Many household assistants	exp_cap
Count	8710.000000	8710.000000	8710.000000	8710.000000	8710.000000	8710.000000	8710.000000	8710.000000	8.710000e+03
Mean	78.183008	2.096211	4.692767	1.790930	4.640643	1.313777	6.289093	4.151550	1.024341e+06
Std	50.684331	1.131769	1.763156	1.710816	2.235971	0.673758	2.631676	1.935222	9.018126e+05
Min	5.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	1.328234e+05
25%	45.000000	1.000000	2.000000	1.000000	2.000000	1.000000	4.000000	3.000000	4.747547e+05
50%	66.000000	3.000000	5.000000	1.000000	5.000000	1.000000	7.000000	4.000000	7.972101e+05
75%	98.000000	3.000000	6.000000	1.000000	7.000000	1.000000	10.000000	5.000000	1.287970e+06
Max	516.000000	7.000000	9.000000	6.000000	11.000000	4.000000	10.000000	16.000000	1.581974e+07

Table 1 shows the descriptive statistics of the dataset used in this study, which consists of 8,710 observations for each variable. The table presents summary measures including count, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum for all attributes, namely *luas lantai (m²), jenis dinding, jenis lantai, kepemilikan FTBAB, sumber air minum, sumber penerangan, bahan bakar, banyak ART,* and *exp_cap*. Based on the table, the average floor area is 78.18 m² with values ranging from 5 m² to 516 m², indicating substantial variation in household housing conditions. The average number of household members (*banyak ART*) is 4.15, with a minimum of 1 and a maximum of 16, showing differences in household size among the observed data. In addition, the variable *exp_cap* has a mean of 1,024,341, with values ranging from 132,823.4 to 15,819,740, reflecting a wide disparity in expenditure per capita. Overall, these descriptive statistics indicate that the dataset has considerable variability across socioeconomic indicators, which is important for supporting the classification process in determining zakat recipients.
Adding the "Outcome" column

The Outcome column is a response change and contains values 0, 1. A value of 1 indicates mustahik and a value of 0 indicates not mustahik. Table 2 shows the results obtained after adding the outcome column. Of the total 8710 data, 7769 data have the label Outcome 1, while the remaining 941 have the label 0. The ratio of the number of data between Outcome 1 (mustahik) and 0 (not mustahik) is 90% : 10%.

Table 2. Addition of the "Outcome" Column

.No	Floor area (m2)	Type of wall	Floor Type	FTBAB Ownership	Source of drinking water	Source of light	Fuel	Many household assistants	exp_cap	Outcome
0	91	1	2	1	2	1	7	6	5.071845e+05	1
1	72	1	2	1	7	1	7	9	2.649656e+05	0
2	72	1	6	1	7	1	10	8	2.274009e+05	0
3	128	1	2	1	2	1	7	5	9.077881e+05	1
4	135	1	2	1	2	1	7	4	6.157560e+05	1
...	...	...	...	...	...	...	...	...	...	...
8705	44	3	5	1	7	1	10	4	3.285493e+05	1
8706	30	3	6	1	8	1	7	2	5.978062e+05	1
8707	63	1	2	1	2	1	7	7	2.749833e+05	0
8708	84	1	6	1	7	1	10	6	2.980794e+05	0

Table 3. Results of Adding the Outcome Column

1.	**Class**	2.	**Number of Data**
3.	Mustahik (1)	4.	7769
5.	Non-Mustahik (0)	6.	941
7.	**Total**	8.	**8710**

Table 3 shows the distribution of the Outcome variable in the dataset. The results indicate that 7,769 data points are classified as mustahik (1), while 941 data points are classified as non-mustahik (0). This distribution shows that the dataset is imbalanced, with the majority of observations belonging to class 1. Such a condition needs to be considered in the modelling stage because class imbalance may affect the performance of the classification algorithm.

Distribution of Training Data and Testing Data

Table 4. Distribution of Data Variables

Variables	Information	Data Entry
X	Independent variable / feature	All columns from index 0 to 7
Y	Dependent variable / target	The last column, namely Outcome

Table 5. Details of variable x (features)

No	Attribute Name
1	Floor area (m2)
2	Wall type
3	Floor type
4	Food and beverage ownership
5	Drinking water source
6	Lighting source
7	Fuel
8	Number of household members

Tabel 5. Variable Y (target):

Target Variable	Information
Outcome	Classification labels for mustahik and non-mustahik

Table 3,4 and 5 shows the data splitting process into independent variables and dependent variables. The independent variable X consists of the first eight attributes, namely luas lantai, jenis dinding, jenis lantai, kepemilikan FTBAB, sumber air minum, sumber penerangan, bahan bakar, and banyak ART. Meanwhile, the dependent variable Y is the last column, namely Outcome, which is used as the target class in the classification process.

Separating the data into training and testing data is a crucial step in machine learning modeling. Training data is used to train the model and build the model itself, while testing data is used to evaluate its performance. Using testing data is crucial because it allows the model to be assessed on previously unseen data.

To split the data, we can use the "train_test_split" function provided in the "sklearn.model_selection" library. In this example, we will split the data with 70% training data and 30% testing data. The data split is done randomly and proportionally using a "stratified random sampling" method. This method ensures that both classes (outcomes) in the data are well represented in both the training and testing datasets.

Using these steps, we can split the data with an appropriate proportion between the training and testing data. This allows us to train the model using the training data and test its performance using the previously unseen testing data. The use of the stratified random sampling method also ensures that balanced class representation is maintained in both data sets, resulting in a more accurate evaluation of the model built.

```
[ ] X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.3, random_state=1)
```

```
[ ] X_train.shape
```

```
(6097, 8)
```

```
[ ] X_test.shape
```

```
(2613, 8)
```

Figure 3 Data Distribution

Figure 3 shows that there are 6097 training data and 2613 testing data.

Random Forest Model Training with Scikit-Learn

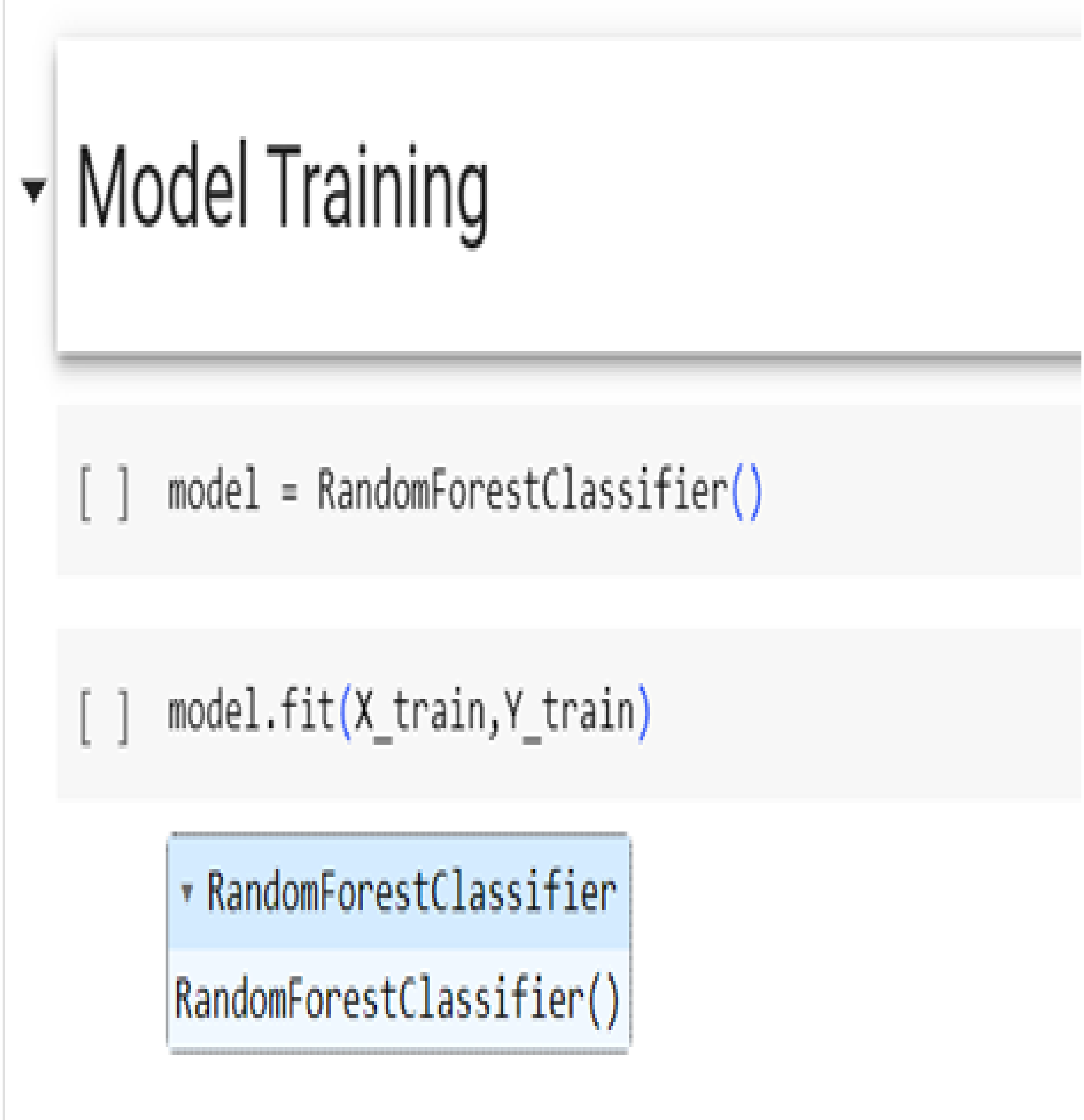

Figure 4 Training Model

Figure 4 shows the model training stage in the Random Forest algorithm using Scikit-Learn. This stage begins with model initialization using RandomForestClassifier(), followed by the training process using the training data X_train and Y_train using the fit() function. Through this process, the model learns the characteristics of the data to form classification patterns that can be used to predict the target class. The appearance of the RandomForestClassifier() output indicates that the model has been successfully built and is ready for use in the prediction and evaluation stages.

Model Evaluation

After training the model with Random Forest, the model is evaluated using the results from the testing data. Using data the model has never seen during the training process provides a fair evaluation and avoids overfitting. The predicted results are then compared to the actual values to measure the model's efficiency at predicting the data. Several factors can be applied in developing a classification model, including accuracy, balanced accuracy, f1 score, ROC AUC, and others. In this context, the metric used is the accuracy score.

```
[ ]  train_data_acc = accuracy_score(X_train_pred, Y_train)

     print(train_data_acc)

     0.9937674266032475

[ ]  X_test_pred = model.predict(X_test)

[ ]  test_data_acc = accuracy_score(X_test_pred,Y_test)

     print(test_data_acc)

     0.878300803673938
```

Figure 5 Model Evaluation

Figure 5 shows the evaluation stage of the Random Forest model after the training process was completed. The evaluation was conducted by measuring the model's accuracy on the training and test data using the accuracy score metric. The test results showed that the accuracy on the training data reached 0.9937674266032475, or approximately 99.38%, while the accuracy on the test data was 0.878300803673938, or approximately 87.83%. These values indicate that the model performed very well on the training data and still produced fairly high classification results on the test data.

The difference in accuracy values between the training and test data indicates that the Random Forest model excels at learning patterns in the training data. However, there was a decrease in performance when the model was tested on previously unseen data. Nevertheless, the accuracy on the test data was still quite good, so the model can be considered quite effective for use in classifying zakat recipients. Based on these results, the Random Forest model is able to provide relatively accurate predictions in distinguishing mustahik and non-mustahik classes based on the attributes used in the study.

Discussion

Zakat management is a crucial aspect in realizing the socio-economic function of Islam, particularly in ensuring that zakat funds are properly distributed to those entitled to receive them appropriately, fairly, and efficiently. In practice, the process of determining mustahik (recipients of zakat) is not always simple, as it involves numerous socio-economic indicators, limited data, and the potential for subjectivity in decision-making. At the same time, the development of digital technology has driven significant changes in the governance of zakat institutions, both in terms of collection and distribution. Firdaus et al (2023) show that the use of online zakat platforms contributes to the growth of zakat collection, while Salleh et al (2022) emphasizes that digital innovation makes it easier for the public to pay zakat and helps optimize zakat collection. These findings demonstrate that digital transformation in the zakat ecosystem is not only relevant at the fund collection stage but also has significant potential for application at the distribution stage, particularly in determining zakat recipients, making it more measurable and data-driven.

In this context, the application of machine learning is particularly appealing because it can help zakat organizations classify or make recommendations based on patterns emerging from historical data. Unlike manual approaches that rely heavily on individual officer judgment, machine learning-based approaches enable the selection process for recipients of zakat (mustahiq) to be carried out more consistently, systematically, and objectively. This is crucial because poorly targeted zakat distribution can reduce the effectiveness of zakat institutions' social programs. Previous research has shown that zakat recipient data management can be assisted by various computational approaches. Imantoyo (2024), for example, highlighted the importance of web-based information systems for managing zakat, infaq, and sadaqah recipients, demonstrating that digitizing mustahik data is the foundation for better decision-making. With the availability of structured data, advanced analysis processes using machine learning become increasingly feasible to support zakat distribution decisions.

The results of this study demonstrate that the use of the Random Forest algorithm in zakat recipient recommendation systems has good prospects. Conceptually, Random Forest works by constructing multiple decision trees from different data samples and feature subsets, then combining the predictions from all the trees to obtain a more stable final decision. These characteristics make Random Forest known for its ability to reduce the risk of overfitting, which often occurs with single decision trees, and for its robust classification performance across a wide range of data types. These advantages are also reflected in various studies outside the zakat domain. Apriliah et al. (2021) used Random Forest for early diabetes prediction, Adrian et al. (2021) compared Random Forest with SVM in PSBB sentiment analysis, and Ramadhan et al. (2019) demonstrated the application of Random Forest in identifying important factors for assessing education quality. These diverse application domains demonstrate Random Forest's adaptability in handling classification problems with varying data characteristics.

In this study, the advantages of Random Forest are relevant because data on prospective zakat recipients generally consist of a number of household indicators, such as house floor area, wall type, flooring type, ownership of a toilet facility, drinking water source, lighting source, fuel, number of household members, and per capita expenditure. These variables do not always have a linear relationship and often interact in complex ways. In such situations, a decision tree-based approach, which can capture interactions between factors without overly stringent statistical assumptions, is highly appropriate. Furthermore, the use of Random Forest offers the advantage of robustness to data variation, allowing the model to maintain relatively good predictions despite varying household characteristics. Therefore, the application of Random Forest in determining eligible beneficiaries can be understood not simply as the use of modern technology, but as the selection of a method appropriate to the nature of the problem and the structure of the data being studied.

When compared with previous research in the field of zakat, this research's position becomes clearer. Studies on determining zakat recipients have used various methods, such as K-Means for grouping potential zakat recipients, Fuzzy C-Means for decision support systems, C4.5 for identifying zakat distribution, and Weighted Product for determining zakat recipients (Arif & Christyanti, 2022; Asa, 2019; Eliyen & Efendi, 2019; Welinda et al., 2016). Each method has important contributions, but also has a different focus. K-Means emphasizes data clustering, Fuzzy C-Means emphasizes flexibility in membership degrees, C4.5 relies on a single decision tree, and Weighted Product tends to be oriented towards criterion weighting. Compared to these methods, Random Forest offers an ensemble classification approach that is theoretically more robust in maintaining predictive stability, especially when the data has significant variation. Thus, this research presents novelty not only in terms of the zakat theme but also in the methods used to support zakat recipient recommendations.

The findings of this study also demonstrate that a machine learning-based approach can help reduce subjectivity in determining mustahik. In practice, subjectivity often arises when decisions rely heavily on individual judgment, the experience of officers, or considerations that vary across regions. Through a classification model, the assessment process can be linked to clear indicators and can be replicated across data sets using the same procedures. This is crucial for zakat institutions seeking to build a more accountable and transparent distribution system. However, the results of this study should not be interpreted as meaning that automated systems completely replace humans. Instead, recommendation systems should be viewed as tools to strengthen decision quality, not as the sole determinant. Rahman (2021) emphasizes that digital innovation in zakat management needs to be directed toward optimizing services, while social and religious values remain the primary foundation. In other words, the final decision regarding the eligibility of mustahik still needs to consider the local context, field verification, and ethical-religious considerations that are not always fully captured by numerical data.

Another important aspect to discuss is the efficiency of zakat institutions. With large amounts of data, manual selection processes require significant time, effort, and administrative consistency. A machine learning-based recommendation system can expedite this process, particularly during the initial screening of potential recipients. With a model, officers can more easily map priority groups, check for potential inconsistencies in data, and avoid duplication of aid. At this point, the use of machine learning also aligns with the growing digitalization of zakat management in recent years. Hidayat and Mukhlisin (2020) demonstrate that digital technology has opened up growth opportunities in online zakat systems, while Imantoyo (2024) highlights the importance of information systems in managing zakat recipients. If zakat collection has benefited from digitalization, then zakat distribution should logically be guided by the same principles: efficiency, accuracy, and traceability.

However, the results of this study also need to be read critically. One common challenge in classification modeling is the possibility of class imbalance, which occurs when the amount of data in one class is significantly greater than in another. This condition can impact model performance, especially if the evaluation relies solely on accuracy. Therefore, in further research, it would be beneficial to evaluate models using not only accuracy but also other metrics such as precision, recall, F1-score, and ROC-AUC. This need for more comprehensive evaluation aligns with broader machine learning research practices. Adrian et al. (2021), for example, compared classification methods on sentiment data, while Navisa et al. (2021) applied the CRISP-DM framework to a comparison of music genre classification algorithms. This demonstrates that model success is not determined solely by a single performance figure, but also by the appropriateness of the evaluation method to the data characteristics and research objectives.

From a methodological perspective, the use of the CRISP-DM framework in this study also strengthens the quality of the analysis process. The stages of business understanding, data understanding, data preparation, modeling, validation, and evaluation help ensure that models are built not simply to achieve a certain accuracy value, but to address the real needs of zakat management. This approach is crucial because zakat issues are not merely technical classification issues, but also social policy and institutional governance. In this context, machine learning must remain positioned as a tool that supports the goals of public welfare. Findings from other studies on the application of Random Forest in health, education, and sentiment analysis indicate that this method is flexible enough to be adapted to various fields, but its success is greatly influenced by data quality, variable selection, and clarity of analysis objectives (Apriliah et al., 2021; Ramadhan et al., 2019). Therefore, the contribution of this research lies not only in the model results but also in providing a more systematic implementation framework for the zakat context.

This discussion confirms that the application of machine learning in zakat management has significant potential to improve the effectiveness, efficiency, and objectivity of zakat distribution. This study demonstrates that the Random Forest algorithm can be a promising alternative in developing a zakat recipient recommendation system, particularly due to its ability to process diverse household socio-economic data and produce relatively stable predictions. Furthermore, this study also emphasizes that technology cannot be separated from the core values of zakat itself. Recommendation systems should be positioned as decision-making support that strengthens the professionalism of zakat institutions, while final decisions still require field verification, socio-religious considerations, and local wisdom. Therefore, the integration of technology, institutional governance, and zakat values is key to achieving a more targeted, transparent, and equitable zakat distribution.

CONCLUSION

Zakat plays a strategic role in reducing economic disparities and assisting those in need. However, its distribution process often faces challenges, particularly in determining zakat recipients accurately, objectively, and efficiently. Research shows that the application of machine learning techniques through a recommendation system with the Random Forest algorithm demonstrates significant potential in supporting the process of determining eligible recipients. This approach can help zakat organizations identify deserving recipients more accurately, reduce subjectivity in decision-making, and increase transparency and efficiency in zakat fund allocation. By utilizing data mining and data analysis, the mustahik selection process can be conducted in a more measurable and systematic manner.

The application of technology in zakat management cannot completely replace human judgment. Social, religious, and local wisdom values remain essential for the zakat distribution process to ensure that decisions are not only data-accurate but also socially just. Therefore, integrating a machine learning-based recommendation system with stakeholder considerations can provide a more comprehensive solution in zakat management. Overall, this study shows that the use of the Random Forest algorithm can be an effective alternative in supporting the digital transformation of zakat management towards a more targeted, transparent, and accountable distribution system.

ACKNOWLEDGMENT
The author would like to express sincere gratitude to all parties who have contributed to the completion of this study.

CONFLICTS OF INTEREST
The author declares no conflict of interest.

AUTHOR CONTRIBUTIONS
All authors were involved in the development and design of the study. The processes of material preparation, data collection, data analysis, and manuscript drafting were carried out jointly by all authors. Each author reviewed and approved the final version of the manuscript.

FUNDING
The authors received no financial support for the research, authorship, and/or publication of this article.

DATA AVAILABILITY
Data sharing is not applicable to this article because no new data were created or analyzed in this study.

BIBLIOGRAPHY
Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan metode klasifikasi Random Forest dan SVM pada analisis sentimen PSBB. *Jurnal Informatika UPGRIS*, *7*(1). https://doi.org/10.26877/jiu.v7i1.7564

Ahmad, A. A. M., Saharuna, Z., & Raharjo, M. F. (2020). Pemanfaatan data mining dalam penentuan rekomendasi mustahik (penerima zakat). *Elektron Jurnal Ilmiah*, *12*(2), 67–73. https://doi.org/10.30630/eji.12.2.182

Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi kemungkinan diabetes pada tahap awal menggunakan algoritma klasifikasi Random Forest. *Sistemasi: Jurnal Sistem Informasi*, *10*(1), 163–171. https://doi.org/10.32520/stmsi.v10i1.1129

Arif, A., & Christyanti, R. (2022). Clustering calon penerima zakat menggunakan metode K-means (Studi kasus di Provinsi Kalimantan Utara). *SMARTICS Journal*, *8*(2), 73–79. https://doi.org/10.21067/smartics.v8i2.7531

Asa, R. S. (2019). Identifikasi penyaluran zakat menggunakan algoritma C4.5 (Studi kasus di BAZNAS Kabupaten Agam). *Jurnal Sains Dan Informatika*, *5*(1), 50–55. https://doi.org/10.22216/jsi.v5i1.4048

Danoedoro, P., Murti, S. H., & Zulfajri. (2022). Klasifikasi penutup/penggunaan lahan data Landsat-8 OLI menggunakan metode Random Forest. *Jurnal Penginderaan Jauh Indonesia*, *3*(1), 1–7. https://doi.org/10.12962/jpji.v3i1.266

Devella, S., Yohannes, Y., & Rahmawati, F. N. (2020). Implementasi Random Forest untuk klasifikasi motif songket Palembang berdasarkan SIFT. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 310–320.

Eliyen, K., & Efendi, F. S. (2019). Implementasi metode weighted product untuk penentuan mustahiq zakat. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, *4*(1), 146–150.

Hidayat, A., & Mukhlisin. (2020). Analisis pertumbuhan zakat pada aplikasi zakat online Dompet Dhuafa. *Jurnal Ilmiah Ekonomi Islam*, *6*(3), 675–684. https://doi.org/10.29040/jiei.v6i3.1435

Navisa, S., Hakim, L., & Nabilah, A. (2021). Komparasi algoritma klasifikasi genre musik pada Spotify menggunakan CRISP-DM: Indonesia. *Jurnal Sistem Cerdas*, *4*(2), 114–125. https://doi.org/10.37396/jsc.v4i2.162

Nugroho, Y. S., & Emiliyawati, N. (2017). Sistem klasifikasi variabel tingkat penerimaan konsumen terhadap mobil menggunakan metode Random Forest. *Jurnal Teknik Elektro*, *9*(1), 24–29.

Rahman, H. (2021). Inovasi pengelolaan zakat di era digital (Studi akses digital dalam pengumpulan zakat). *Dirosat: Journal of Islamic Studies*, *6*(2), 53–63. https://doi.org/10.28944/dirosat.v6i2.412

Ramadhan, A., Susetyo, B., & Indahwati. (2019). Penerapan metode klasifikasi Random Forest dalam mengidentifikasi faktor penting penilaian mutu pendidikan. *Jurnal Pendidikan Dan Kebudayaan*, *4*(2), 169–182. https://doi.org/10.24832/jpnk.v4i2.1327

Ratnawati, L., & Sulistyaningrum, D. R. (2019). Penerapan Random Forest untuk mengukur tingkat keparahan penyakit pada daun apel. *Jurnal Sains Dan Seni ITS*, *8*(2), A71–A77.

Setiyaningsih, T., Mafiroh, W., & Novianti, E. (2020). Menentukan penerima zakat menggunakan metode simple multi attribute rating technique (SMART). *Jurnal Sains Dan Teknologi Fakultas Teknik Universitas Darma Persada*, *10*(2), 40–50.

Syukron, A., & Subekti, A. (2018). Penerapan metode random over-under sampling dan Random Forest untuk klasifikasi penilaian kredit. *Jurnal Informatika*, *5*(2), 175–185.

Welinda, R., Sarita, M. I., & Dewi, A. P. (2016). Implementasi metode fuzzy C-means pada sistem pendukung keputusan penentuan mustahik di BAZNAS Kendari. *SemanTIK*, *2*(1), 155–168.

Wuisan, D. S. S. (2020). Analisis sistem informasi pengelolaan data transaksi pembelian consumer untuk manajemen pembelian perusahaan. *Infotech: Journal of Technology Information*, *6*(1), 29–34. https://doi.org/10.37365/jti.v6i1.75

Zailani, A. U., & Hanun, N. L. (2020). Penerapan algoritma klasifikasi Random Forest untuk penentuan kelayakan pemberian kredit di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, *6*(1), 7–14.