Prediction of Supporting Factors for the Success of BAZNAS RI Digital Fundraising Using the C4.5 Algorithm

Yuliana[1], Binti Mamluatul Karomah[1], Mohammad Reza Fahlevi[1]
[1]Nahdlatul Ulama University of Indonesia, Indonesia

ABSTRACT

The C4.5 algorithm as a prediction method in zakat fundraising. The methodology adopts CRISP-DM (Cross-Industry Standard Process for Data Mining) by involving the data preparation stage to process raw data into data that can be processed and implementing the C4.5 algorithm to build a decision tree model. The prediction model formed can predict the success of zakat fundraising from prospective muzaki. The main objective of this study is to develop a predictive model that can help the National Amil Zakat Agency of the Republic of Indonesia (BAZNAS RI) in planning a more effective Zakat fundraising strategy. BAZNAS RI can optimize zakat fundraising strategies and allocate resources more efficiently through this prediction model. Muzaki data including information about age, gender, occupation group, transaction period, nominal, and nominal category became data input in this study. The model evaluation results show that the model has an accuracy rate of 92% in making predictions, giving hope that this model can be an effective tool in supporting zakat fundraising activities by BAZNAS RI.

Keywords: *Fundraising, CRISP-DM, C4.5 algorithm, prediction, BAZNAS RI.*

INTRODUCTION

The utilization of technology in fundraising Zakat funds holds significant potential in enhancing awareness and participation among Muslims worldwide. Fundraising, as an activity aimed at gathering funds or other resources from the community, whether individuals or organizations, to support the financing of programs and operational activities, plays a crucial role in achieving specific objectives (Perdana & Zen, 2020). Several institutions have collaborated with e-commerce platforms, online applications, and similar avenues to leverage the benefits of financial technology (Febiana et al., 2022).

Zakat, one of the pillars of Islam, is incumbent upon Muslims to fulfill. According to the World Giving Index 2022, a Global View of Giving Trends, Indonesia stands out as the most generous nation globally, maintaining the highest World Giving Index score for five consecutive years. Indonesia retains its top ranking with a 68% Index, aiding strangers (58%), exhibiting the highest donation rates (84%), and volunteering (63%) globally. Zakat, Infaq, and Sadaqah (ZIS) funds acquired through fundraising significantly impact the existence of Zakat management institutions. This is because these funds enhance community welfare and alleviate poverty levels in the country(Alfian & Widodo, 2022).

The National Amil Zakat Agency (BAZNAS) is the official and sole body established by the government based on Presidential Decree Number 8 of 2001(Keputusan Presiden Republik Indonesia Nomor 8, 2001), tasked with collecting and distributing Zakat, Infaq, and Sadaqah (ZIS) at the national level. The enactment of Law Number 23 of 2011 on Zakat Management further solidifies BAZNAS's role as the institution authorized to manage Zakat nationally. In this Law, BAZNAS is designated as a non-structural government agency that operates independently and is accountable to the President through the Minister of Religious Affairs (BAZNAS, 2019). One of BAZNAS's goals is the effective distribution of Zakat, Infaq, and Sadaqah (ZIS), and other religious social funds (DSKL), to alleviate poverty, enhance community welfare, and reduce social disparities.

Based on the target fundraising goals for Zakat funds next year, amounting to 33 billion, it is crucial to predict and understand the factors influencing fundraising success, especially for a State Institution focusing on social and humanitarian projects such as the National Amil Zakat Agency.

According to the Gartner Group, as cited in Larose, data mining is a process of discovering meaningful relationships, patterns, and trends by examining large datasets stored in repositories using pattern recognition techniques such as statistical and mathematical methods (Dewi, 2020). The C4.5 algorithm is a data mining algorithm used for classification and estimation. The C4.5 algorithm belongs to the decision tree induction algorithms developed by J. Ross Quinlan, known as ID3 (Iterative Dichotomiser 3) (Suweleh et al., 2020). This algorithm can be used to predict whether a fundraising project will successfully achieve its fundraising target or not. Before implementing the C4.5 Algorithm model, an evaluation is conducted using a confusion matrix. The confusion matrix can calculate various evaluation metrics, including accuracy, precision, recall, and etc. These metrics provide insights into the model's performance in classifying data and help understand the types of errors made by the model (Paskalis et al., 2019).

By implementing the C4.5 algorithm on fundraising projects' data, factors influencing digital fundraising success can be identified, and how these factors can be optimized to enhance the National Amil Zakat Agency's digital fundraising success (Undang-Undang Nomor 23, 2011). Hence, this study aims to discuss the implementation of the C4.5 algorithm-based prediction as a supportive factor for digital fundraising success with the case study of BAZNAS RI. This research is expected to assist BAZNAS in fundraising for Zakat, Infaq, and Sadaqah, as per BAZNAS's vision of becoming the primary institution in uplifting the community's welfare. This research focuses on philanthropic cases of Zakat, Infaq, and Sadaqah (ZIS) fundraising.

LITERATURE REVIEW

The fintech phenomenon has significantly impacted the financial sector, changed the traditional paradigm, and created innovation opportunities for related institutions, including zakat organizations (Hudaefi et al., 2020). Digital zakat fundraising is one of the innovations that take advantage of advances in information and communication technology to bring the zakat digital fundraising process closer to the community. By utilizing digital platforms such as applications, websites, and social media, digital zakat fundraising has opened the door for more potential *muzaki* or zakat givers to participate in helping others (Doddy et al., 2022).

In zakat digital fundraising, the process of collecting zakat funds becomes more accessible and more efficient. Previously, *muzaki* had to come directly to zakat institutions or send zakat funds conventionally, but now they can give zakat quickly and conveniently through digital platforms. This appealed to a broader audience, especially the younger generation, which is very tech-oriented. In addition to facilitating *muzaki*, digital zakat fundraising also benefits the zakat institution. By implementing a digital platform, zakat institutions can manage and monitor the funds collected more transparently and accurately. Information about the use of zakat funds can also be easily accessed by *muzaki*, so they feel more confident because their donations are used appropriately (Alim & Z. Basri, 2020).

One of the main purposes of paying zakat is to distribute wealth to achieve economic equality and social justice. By setting aside a portion of their wealth, more fortunate people significantly contribute to helping people in need, thus creating a more just and empathetic social environment. The payment of zakat also serves as "al-Barakah" i.e., a blessing that brings abundance and luck to those who carry it out. Zakat can cleanse one's heart and soul from greed and materialism, and open the door of fortune and blessings from Allah (Karim et al., 2022).

Python, the main programming language used in Jupyter Notebook, has many advantages. Python is a programming language best known for its ability to support object-oriented programming and various other programming paradigms. In addition, Python can be run on multiple operating system platforms such as PCs, Macintosh, and UNIX making it a flexible choice for various purposes (Ginting et al., 2020). Some of the advantages of Python include faster program development with less code, multi-platform support that facilitates program portability, ease of learning Python language for beginners, an automated memory management system that reduces user workload, and full support for object-oriented programming paradigms that help in building more modular and easy-to-understand programs.

Data mining is divided into two words. Data is a collection of recorded facts or an entity that has no meaning and has been neglected. Mining is a mining process, so data mining can be interpreted as a process that produces an output in the form of knowledge. Data mining is a semi-automated process using statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify potential and useful information stored in large databases (Zai, 2022). Data mining is part of the KDD (Knowledge Discovery in Databases) process which consists of several stages such as data selection, pre-processing, transformation, data mining, and evaluation of results. According to (Takdirillah, 2020), data mining is divided into several groups based on the tasks that can be done, namely classification, association, clustering, prediction, and estimation.

Data mining usually processes data from databases with large sizes, and from the data is carried out looking for patterns or trends for the purpose of applying data mining (Liliana et al., 2021). The results of the data mining processing can then be used to make the required prediction decisions. Prediction is a function that can find specific patterns from data. These patterns can be known from various variables in the data. When you have found a pattern, the pattern obtained can be used to predict other variables whose value or type is unknown.

Decision tree algorithms include supervised learning, which means they require previous goals or target variables as training data. Using the C4.5 algorithm, we can study large amounts of data and build learning models in the form of decision trees that can be used to predict data that has never appeared before(Febriarini & Astuti, 2019). Decision trees are one alternative for solving problems taken from these data. A set of training data must be prepared to form an algorithm with the values of the target variables. This method evaluates all attributes using statistical measures such as information gain and entropy calculations. Information gain is the acquisition of information or a measure of the effectiveness of an attribute in classifying data. Equation (i) is the information gain formula, which is (Widaningsih, 2019).

$$\text{Gain}\,(S, A) = \text{Entropy}(S) - \sum \frac{|Sv|}{|S|}\,\text{Entropy}\,(Sv) \text{--- (i)}$$

Information:
A: attribute
|Sv|: number of samples for v value
|S|: the sum of all sample data

Entropy is the diversity of data. Equation (2) is the entropy formula:

$$\text{Entropy}\,(S) = -\sum p_i \log_2 p_i \qquad \text{--- (ii)}$$

Information:
pi = portion or ratio between the number of class I samples and the sum of all samples in the dataset

The confusion matrix is a very important tool in the field of classification for evaluating the predictive efficiency of classification models. A confusion matrix is a table that describes the performance achieved by classifiers in the test dataset. The table consists of rows and columns that represent the actual and predicted classes respectively. All other evaluation metrics used in the classification field (accuracy, precision, and recall) are derived from matrix confusion. The confusion matrix for binary classification problems contains four values, including: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TN and TP indicate the correct frequency of the model in classifying data instances as negative and positive, respectively; whereas FN and FP show the frequency of the model incorrectly classifying data instances as negative and positive (Ryoba et al., 2020).

Visual Studio Code (VS Code) is a lightweight and powerful text editor developed by Microsoft for multiplatform operating systems, which means it is available for Linux, Mac, and Windows versions. This text editor directly supports JavaScript, TypeScript, and Node.js programming languages, as well as other programming languages with the help of plugins that can be installed through the Visual Studio Code marketplace (such as C++, C#, Python, Go, Java, and others). In addition, VS Code is also open source, which means that its source code can be viewed and contributions can be made to its development. The source code of VS Code can be found on GitHub. This makes VS Code a favorite choice for application developers because they can participate in the development process in the future. (Permana & Romadlon, 2019).

Deploying models into systems often requires recoding the model to make it faster or more compatible with existing production environments. This process can take time and effort if the right tools are not used. To overcome this challenge, a framework is needed that can turn data scripts into website applications quickly and easily. The important role of Streamlit, an opensource framework from Python that allows data professionals to create web applications using the Python language, is to apply models from machine learning or data science (Ferdyandi et al., 2022).

Streamlit is an ideal choice for data scientists because of its excellent features. One of them is the ability to refresh the view automatically when changes occur to the script. This allows users to see changes in the model directly without performing a manual refresh. That way, the process of developing and iterating the model becomes faster and more practical. In addition, Streamlit also offers a variety of interactive widgets that make it easier for users to control and manipulate data(Dani, 2022).

RESEARCH METHODOLOGY

This research applies the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as the research approach used. CRISP-DM is a method that provides a standard general problem-solving process from existing research units. The data mining process based on CRISP-DM consists of 6 phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Hasanah et al., 2021). This methodological process consists of 6 phases which can be explained as follows.

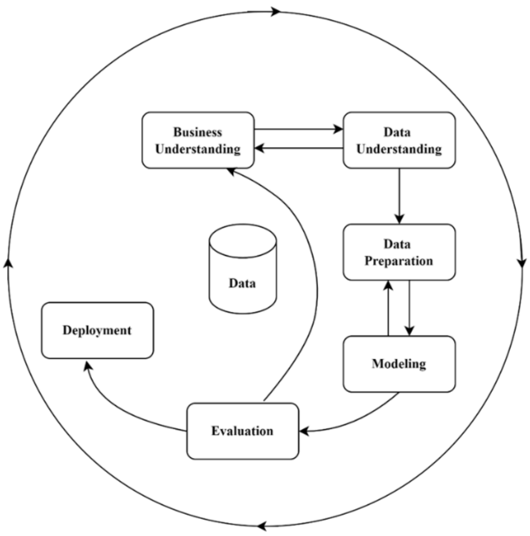


Figure 1 CRISP-DM Methodologists
Source: (Hasanah et al., 2021)

Business Understanding

An important initial stage to fully understand the business context is zakat fundraising which aims to maximize the distribution and utilization of ZIS-DSKL to alleviate poverty, improve the welfare of the ummah, and reduce social inequality. The modeling results will be used to make business decisions in implementing an efficient Zakat fundraising strategy. The data needed to achieve the research objectives is *muzaki* data and does not use sensitive data to maintain the confidentiality of the data. The list of stakeholders involved in the research are

1. Directorate of Study and Development of ZIS-DSKL,
2. Directorate of Services, Promotion and Data Optimization, and
3. Directorate of Information Security, Data and Digital Services.

Data Understanding

The next stage after Business Understanding in data mining projects. The purpose of data understanding is to get an overview of data characteristics, and data quality, and identify problems or shortcomings that need to be addressed before continuing the prediction process.

The dataset used in this experiment consists of information about *muzaki* that includes various variables such as age, gender, occupation group, transaction period, and nominal category. Here are the details:

Table 1. Final Dataset

Variable	Data Type	Description
Age	Numerical	Age of *muzaki* in years
Gender	Categorical	*Muzaki* sex
Occupation Group	Categorical	Types of *muzaki* work
Transaction Period	Categorical	Transaction time
Category Nominal	Categorical	Category nominal zakat

Source: Author's Computations (2024)

Data Preparation

In this stage, it is to build the final dataset from raw data into data ready for processing. Several things will be done as through:

1. Data cleaning, identifying, and handling missing values in datasets.
2. Data splitting, separating data into two subsets, namely training data (training data) and testing data (test data).
3. Data transformation, the conversion of categorical data into forms suitable for predictions such as one-hot encoding or label encoding to be used as input in the modeling stage.

Modeling

At this stage, build a predictive model based on pre-prepared data. C4.5 algorithm development is used to make decisions or make predictions based on existing data. Flowchart of the C4.5 algorithm in the figure below:

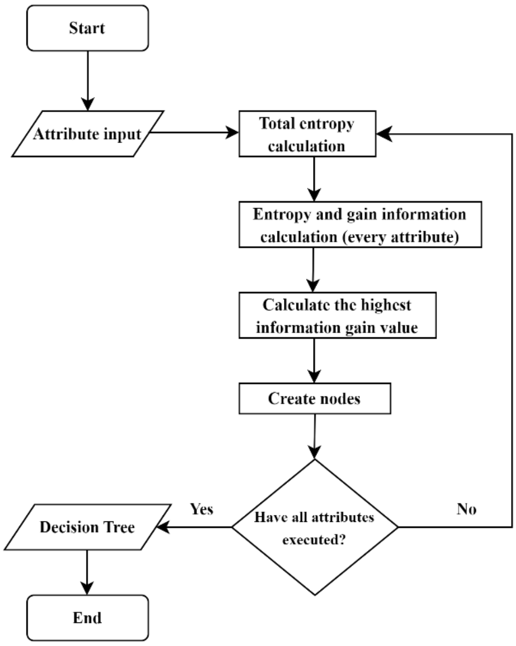


Figure 2. Flowchart Algoritma C4.5
Source: (Mardi, 2019)

After inputting the dataset, entropy calculations are carried out according to formula (ii). Calculate entropy and information gain according to formula (i), and calculate the highest value of information gain. Next, create a node and run all the attributes to produce a decision tree model. Interpretation of results from prediction models to gain insight into factors that affect results and used to make business decisions or fundraising zakat and provide recommendations to BAZNAS RI.

Evaluation

This stage is done by looking at the level of performance of the model generated by the algorithm. Evaluation is a critical process of understanding the extent to which models can reliably make decisions or make accurate predictions. The parameter used for the evaluation of algorithm comparison is the Confusion Matrix with rules of accuracy, precision, and recall values. The results of this model are the findings of testing data that has been prepared for model creation. Therefore, evaluation of the initial model formed is important to test the quality and effectiveness of the model before deciding whether the model found is feasible to implement (deploy). Identify errors made by the model, such as the types of errors that often occur and their causes. Therefore, corrective efforts are made that can be made based on the identification of the error. Further testing the stability of the model to ensure that the model provides consistent results so that it can be relied upon on different data or future data.

Deployment

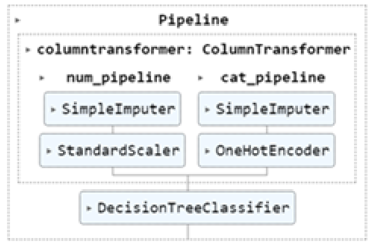
In the final stage, the predictive model that has been developed and evaluated is applied in a production or operational environment. The model is implemented and integrated into the system used by BAZNAS RI to make decisions or support business processes (zakat fundraising). The stages are divided into two, namely deploy simply and deploy thoroughly. Comprehensive deployment is a continuous intensive research using larger data and the data mining process is carried out in parallel. In this research, it is deployed continuously, because BAZNAS RI will use it in predicting digital fundraising in the future.

RESULTS & DISCUSSIONS

At this stage, testing will be carried out using the C4.5 algorithm on the dataset that has been prepared. This process aims to identify patterns and relationships between variables that can be used to predict the results of digital fundraising from prospective muzak. Before testing using the C4.5 algorithm, a machine learning pipeline was created to help organize data processing workflows and models, making it easier to develop and customize models.

Testing using the c4.5 algorithm

```
pipefinal.fit(X_train, y_train)
```




```
pipefinal.score(X_test, y_test)
0.9227848101265823
```

Figure 3. Pipeline machine learning
Source: Author (2024)

From the image above, it can be seen that an accuracy score of 0.92 can be obtained from creating a machine learning pipeline, which means that the model has better performance. Furthermore, a predictive model is built based on previously prepared data. The result of this stage is a ready-to-use predictive model to predict the outcome of digital fundraising.

```
def calculate_entropy(series):
    counts = series.value_counts()
    probabilities = counts / counts.sum()
    entropy = -np.sum(probabilities * np.log2(probabilities))
    return entropy
```

Figure 4. Calculate the entropy of every variable
Source: Author (2024)

The entropy calculation of each variable in the dataset starts by calculating the total number of samples in the data and then calculating the number of occurrences of each value in the categorical variable using the 'value_counts' method. It then calculates the probability of occurrence of each value by dividing the number of occurrences by the total number of samples. Probabilities are the portion or ratio between the number of class I samples and the sum of all samples in the dataset. Then, calculate the entropy of each variable in the dataset using the formula in equation (ii) that has been described. The results of calculating the entropy of each variable are obtained in Table 2 below. This value indicates that the variable with greater variation is age.

Table 2. *The entropy value of each variable*

Variable	Nilai Entropy
Age	4.90513981626614
Gender	0.7960833381430548
Occupation Group	3.4041416302684167
Transaction Period	1.2569797128181075

Source: Author's Computations (2024)

Next, calculate weighted entropy to calculate the value of information gain, so that the variable with the highest information gain is obtained, which is used as the root node. The first thing to do is extract unique values from columns on the data frame and initialize 'weighted_entropy_sum' to store the weighted amount of entropy. Then, it iterates through the unk values and calculates the entropy of the subset using the 'calculate_entropy' function.

Then the weighting and accumulation by multiplying the result of 'calculate_entropy' by the proportion of the subset to the whole dataset (the number of samples in the subset divided by the total sample) and adding to the 'weighted_entropy_sum'. Then returns the total value of 'weighted_entropy_sum'.

```
def weighted_entropy(df, column, target):
    unique_values = df[column].unique()
    weighted_entropy_sum = 0
    total_samples = len(df)

    for value in unique_values:
        subset = df[df[column] == value]
        subset_entropy = calculate_entropy(subset[target])
        weighted_entropy_sum += (len(subset) / total_samples) * subset_entropy

    return weighted_entropy_sum
```

Figure 5. Calculate weighted entropy
Source: Author (2024)

Then the calculation of information gain from each variable in the dataset is carried out using the formula in equation (1) described in Chapter 2. The first thing to do is calculate the total entropy of the entire target column and the weighted entropy. It iterates through each column and the code calculates the entropy of that column and the gain information based on age.

```
def information_gain(df, column, target):
    total_entropy = calculate_entropy(df[target])
    w_entropy = weighted_entropy(df, column, target)
    return total_entropy - w_entropy

# Hitung entropy untuk setiap variabel kategorikal dan Information Gain
columns_to_analyze = ['umur', 'gender', 'occupation_group',
                      'periode_transaksi']

for column in columns_to_analyze:
    entropy_value = calculate_entropy(df[column])
    ig = information_gain(df, column, 'umur')
    print(f"Entropy for {column}: {entropy_value}")
    print(f"Information Gain for {column}: {ig}")
```

Figure 6. Calculate information gain
Source: Author (2024)

The results of calculating each variable's information gain are obtained in Table 3 below. The age variable's highest gain will be used as the root node.

Table 3. Value *of information gain* of each variable

Variable	Nilai Information Gain
Age	4.905139816266144
Gender	0.10551167788908788
Occupation Group	0.8829641831609703
Transaction Period	0.033769493514653526

Source: Author's Computations (2024)

Evaluation and Validation of Results

In this section, we will evaluate and validate the results of the prediction model that has been built using the C4.5 algorithm. Evaluation is done by checking how well the model can predict the results of zakat fundraising based on predetermined attributes.

```
y_pred = model_c45.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average="weighted")
recall = recall_score(y_test, y_pred, average="weighted")
conf_matrix = confusion_matrix(y_test, y_pred)

print("Akurasi:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("Confusion Matrix:")
print(conf_matrix)
```

```
Akurasi: 0.9177215189873418
Precision: 0.9013219796428049
Recall: 0.9177215189873418
Confusion Matrix:
[[   4    7    3    4]
 [   3 1415   32    0]
 [   2   69   22    3]
 [   2    5    0    9]]
```

Figure 7. Confusion Matrix

Source: Author (2024)

The explanation of the results of the model evaluation is described as follows:

1. The accuracy of the model is about 91.77%, which indicates that most of the predictions made by the model correspond to the actual data.
2. The precision is about 90.13%, which means a fraction of the positive predictions made by the model may be wrong.
3. The recall had has a value of about 91.77%, indicating that the model is able to identify most of the positive labels present.
4. In this matrix, columns represent the classes predicted by the model, while rows represent the actual classes of data. From the matrix, it can be seen that the model has some errors in predicting some classes, especially in classes with a lower number of samples. For example, models tend to have errors in predicting the second and third classes, which is indicated by a higher number of false negatives (FN) and false positives (FP) in the matrix.

Overall, the evaluation results show that the model has high accuracy and a good ability to identify positive labels, but still needs improvement, especially in reducing errors in predicting classes with lower sample numbers.

```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.36      0.22      0.28        18
           1       0.95      0.98      0.96      1450
           2       0.39      0.23      0.29        96
           3       0.56      0.56      0.56        16

    accuracy                           0.92      1580
   macro avg       0.56      0.50      0.52      1580
weighted avg       0.90      0.92      0.91      1580
```

```
df.shape
```

```
(7896, 5)
```

Figure 8. Classification Report

Source: Author (2024)

Based on the validation results, the model's performance in classifying data into four different classes can be seen from various evaluation metrics. For class 0, the precision score was recorded at 0.36, while the recall was 0.22, and the F1 score was 0.28. For class 1, a high precision of 0.95 indicates a high level of prediction accuracy, supported by a recall of 0.98 and an F1 score of 0.96. Class 2 shows a precision of 0.39, a recall of 0.23, and an F1 score of 0.29. Meanwhile, for class 3, precision reached 0.56, recall also 0.56, and F1-score 0.56. The overall accuracy of the model is about 92%. Using averages from evaluation matrices, such as the macro average of about 0.52 and the weighted average of about 0.91, you can get an idea of the model's overall performance in predicting different classes.

The results of the evaluation and validation of the model showed good performance in identifying majority classes but still needed improvement in better classifying minority classes to improve overall precision and recall.

Experimental Results

In this section, the results of experiments conducted on two types of data, group data and single data, will be displayed. Experiments on group data aim to identify overall patterns or trends among larger data groups, while experiments on single data focus on more detailed predictions of individuals within that group.

1. Group Data Experiments

Before conducting experiments on cohort data, a data input file in CSV format consisting of 1580 data records was prepared for prediction. In experiments using group data can be accessed via a link https://prediksifundraisingdigital.streamlit.app/. After accessing the link, users can make predictions by uploading input files. The results of those predictions will be displayed along with visualizations that support the interpretation of the results. To give a clearer picture, the author also includes the view of the web application in Figures 9 and 10 below.

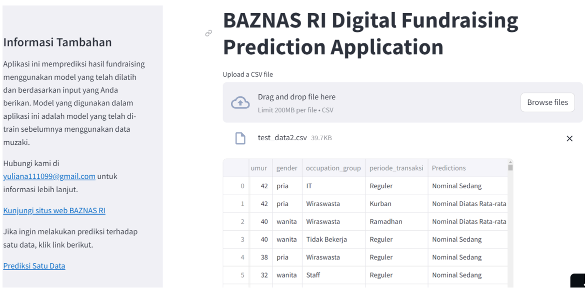


Figure 9. Results of Group Data Experiments

Source: Author (2024)

In Figure 10, the display of the digital fundraising prediction application with its tone feature to upload files in CSV format. In the sidebar, there is additional information, researcher contacts, the BAZNAS RI website, and links to make predictions of single data or one data.

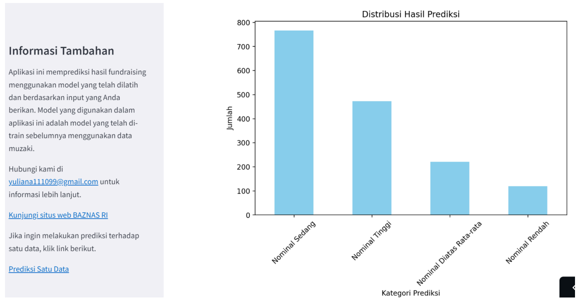


Figure 10. Visualization of Experimental Results

Source: Author (2024)

It can be seen that the result of the predictions that have been made is that most *muzaki* will transact with a medium nominal category, so it can be predicted that the collection of zakat funds will reach the target.

2. Single Data Experiment

In the experiment, use a single data through the following link https://prediksifundraisingdigitalv2.streamlit.app/. After accessing the platform, users are asked to set parameters by sliding the age slide to 51 years, choosing gender as a woman, profession as a civil servant, and transaction period as Ramadan. By setting these parameters, users can get a prediction result that is a nominal amount above the average that may be collected from fundraising campaigns. You can see the view of the Streamlit web application in the following image.

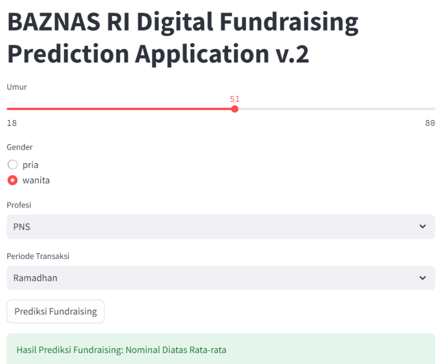


Figure 11. Single Data Experiment Results

Source: Author (2024)

A description of the nominal data range of each category is obtained from the nominal clustering that has been carried out. With information on the nominal categories as follows (Rupiah):

1. Low Nominal: 10,000 to 2,146,000
2. Medium Nominal: 2.175.089 to 8.000.000
3. Nominal Above average: 8,058,000 to 18,306,000
4. High Nominal: 19,165,165 to 37,936,944

The results of creating a fundraising prediction model show an accuracy rate of 92%, obtained through Machine Learning techniques such as the Decision Tree Classifier. With a careful model-building process, this model managed to predict fundraising results with satisfactory accuracy, based on pre-prepared data. The analysis also reveals important factors that influence fundraising results, such as age, gender, transaction period, and nominal categories, which are reflected in the contribution of these variables in the model decision-making process. For example, it can be seen that the age of the *muzaki*, gender, and nominal category of donations have a significant impact on the number of funds raised in a certain period.

By understanding the key variables that affect fundraising results, BAZNAS RI can allocate resources more efficiently, and optimize strategies and effectiveness of zakat collection. These results provide insight for BAZNAS to improve its performance based on the information provided by the prediction model. Thus, the creation of a predictive fundraising model significantly contributes to supporting the achievement of BAZNAS' objectives in improving the effectiveness of digital fundraising campaigns, as well as allowing to optimization of fundraising strategies and efforts based on more in-depth data analysis.

Research Limitations

This research succeeded in producing a fundraising prediction model that has a satisfactory level of accuracy. However, there are some limitations to note. First, limitations in the data, such as lack of relevant variables or low data quality, can affect the validity and generalizability of the model.

Second, the assumptions underlying the model also need to be considered to ensure the interpretation of the results and their applicability in various situations. Third, the technical limitations of the model can affect its ability to handle complex patterns in the data. Therefore, the results of this model need to be interpreted carefully in the context of actual decision-making.

CONCLUSION

From the results of the research that has been done, conclusions can be drawn about the main findings and contributions that have been obtained as follows.

1. The results of making a fundraising prediction model showed an accuracy rate of 92% using the C4.5 algorithm.
2. Important factors that affect *fundraising* results are age, gender, transaction period, and nominal category, which is reflected in the contribution of these variables in the model decision-making process. For example, it can be seen that the age of the *muzaki*, gender, and nominal category of donations significantly impact the amount of zakat collected in a certain period.
3. By understanding the variables that affect fundraising results, BAZNAS RI can allocate resources more efficiently, and optimize strategies and effectiveness of zakat collection. These results provide

insight for BAZNAS to improve its performance based on the information provided by the prediction model.

REFERENCES

Alfian, M., & Widodo, N. M. (2022). Faktor Penentu Keberhasilan Penggunaan Fintech dalam Pembayaran Zakat , Infaq , Shodaqoh (ZIS). *Jurnal Ilmiah Ekonomi Islam*, *8*(03), 2651–2656. https://jurnal.stie-aas.ac.id/index.php/jie Jurnal

Alim, M. N., & Z. Basri, M. T. (2020). Financial Determinants In Zakat Institution Management Effecting muzakky loyalty In Indonesia. *International Journal of Business and Management*, *9*(2), 35–47.

Dani, H. (2022). Review on Frameworks Used for Deployment of Machine Learning Model. *International Journal for Research in Applied Science and Engineering Technology*, *10*(2), 211–215. https://doi.org/10.22214/ijraset.2022.40222

Doddy, M., Ali, J., Hindardjo, A., & Sani, A. (2022). Boost Zakat Fundraising Through E-Customer Relationship Management in Digital Era. *Technium Business and Management*, *2*(2), 61–70. https://doi.org/10.47577/business.v2i2.6762

Febiana, N., Tanjung, H., & Hakiem, H. (2022). Pengaruh literasi zakat, infaq, shadaqah (ZIS), kepercayaan, dan brand awareness terhadap keputusan menyalurkan zakat dan donasi melalui Tokopedia : studi pada mahasiswa FAI Universitas Ibn Khaldun Bogor pada angkatan 2017-2018. *El-Mal Jurnal Kajian Ekonomi & Islam*, *5*(2), 291–313.

Febriarini, A. S., & Astuti, E. Z. (2019). Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang. *Eksplora Informatika*, *8*(2), 95–103. https://doi.org/10.30864/eksplora.v8i2.156

Ferdyandi, M., Setiawan, N. Y., & Abdurrachman Bachtiar, F. (2022). *Prediksi Potensi Penjualan Makanan Beku berdasarkan Ulasan Pengguna Shopee menggunakan Metode Decision Tree Algoritma C4.5 dan Random Forest (Studi Kasus Dapur Lilis)*. *6*(2), 588–596. http://j-ptiik.ub.ac.id

Ginting, V. S., Kusrini, K., & Taufiq, E. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, *10*(1), 36–44. https://doi.org/10.35585/inspir.v10i1.2535

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2).

Hudaefi, F. A., Beik, I. S., Zaenal, M. H., Choiri, M., Farchatunnisa, H., & Junari, U. L. (2020). How Does Zakat Institution Respond To Fintech? Evidence From Baznas, Indonesia. *International Journal of Zakat and Islamic Philanthropy*, *2*(1), 32–40.

Karim, A., Mufakhidin, A., Kusuma, H. H., Adeni, A., & Fitri, F. (2022). Mitigating Poverty: the Clustering of Potential Zakat in Indonesia. *Analisa: Journal of Social Science and Religion*, *7*(1), 109–126. https://doi.org/10.18784/analisa.v7i1.1641

Keputusan Presiden Republik Indonesia Nomor 8. (2001). https://peraturan.bpk.go.id/Details/57030/keppres-no-8-tahun-2001

Liliana, D. Y., Maulana, H., & Setiawan, A. (2021). Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes. *Multinetics*, *7*(1), 48–53. https://doi.org/10.32722/multinetics.v7i1.3786

Mardi, Y. (2019). Data Mining : Klasifikasi

Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika. *Jurnal Edik Informatika*, *2*(2), 213–219.

Perdana, B. A., & Zen, M. (2020). Fundraising Dana Infak dan Sedekah dalam Meningkatkan Kepercayaan Jamaah Masjid. *Tadbir: Jurnal Manajemen Dakwah*, *5*(2), 137–146. https://doi.org/10.15575/tadbir.v5i2.2099

Permana, A. Y., & Romadlon, P. (2019). Perancangan Sistem Informasi Penjualan Perumahan Mengunakan Metode SDLC Pada PT. Mandiri Land Prosperous Berbasis Mobile. *SIGMA – Jurnal Teknologi Pelita Bangsa*, *10*(2), 153–167.

Ryoba, M. J., Qu, S., Ji, Y., & Qu, D. (2020). The right time for crowd communication during campaigns for sustainable success of crowdfunding: Evidence from Kickstarter platform. *Sustainability (Switzerland)*, *12*(18). https://doi.org/10.3390/su12187642

Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, *4*(1), 37–46. https://doi.org/10.29408/edumatic.v4i1.2081

Undang-Undang Nomor 23. (2011). https://peraturan.bpk.go.id/Details/39267/uu-no-23-tahun-2011

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Portal Data*, *2*(3), 1–12.

Yuliana

Nahdlatul Ulama University
of Indonesia
Indonesia

yuliana111099@gmail.com