ITEJ
Information Technology Engineering Journals
eISSN : 2548-2157

Url : https://syekhnurjati.ac.id/journal/index.php/itej
Email : itej@syekhnurjati.ac.id

Classification of Family Hope Program Assistance Recipients Using the C4.5 Algorithm with Z-Score Normalization (Case Study in Atu Lintang District)

Siti wahyuni
Informatics Engineering
Malikussalaeh University
Lhokseumawe, Indonesia
siti.210170130@mhs.unimal.ac.id

Asrianda
Informatics Engineering
Malikussalaeh University
Lhokseumawe, Indonesia
asrianda@unimal.ac.id

Sujacka Retno
Informatics Engineering
Malikussalaeh University
Lhokseumawe, Indonesia
sujacka@unimal.ac.id

Abstract—One of the challenges in distributing social assistance is determining recipients who are truly eligible objectively and efficiently. This study develops a classification system for Family Hope Program (PKH) recipients by utilizing the C4.5 algorithm combined with Z-Score normalization to group citizen data into Eligible or Ineligible categories. The data used came from 551 residents of Atu Lintang District and included attributes such as house status, wall type, toilet facilities, occupation, and income. The research stages started from data preprocessing, attribute normalization, training the model, to evaluating its performance through metric such as accuracy, precision, recall, and F1-score. The evaluation results showed that the model achieved an accuracy of 94%, precision 0.96, recall 0.90, and F1-score 0.93 for the Eligible category. Based on the confusion matrix, the model was able to correctly classify 47 Eligible residents and 57 Ineligible residents. Analysis of the attributes showed that occupation was the most influential feature in the classification process. These results prove that the application of the C4.5 algorithm can be applied effectively to build a decision support system in the distribution of social assistance, and provide accurate and easy-to-understand results. This study also opens up opportunities for improving model performance by adding more data and testing with alternative algorithms going forward.

Keywords—C4.5 Algorithm, Classification, Confusion Matrix, Family Hope Program, Z-Score Normalization

I. INTRODUCTION

Poverty is still a serious problem in Atu Lintang District, where many families have difficulty meeting basic needs such as food, education, and health services. This condition has a direct impact on the decline in the quality of life of the community and slows down the development process in various sectors [1]. The inability to meet basic needs also risks exacerbating social inequality and reducing public trust in government-run assistance programs.

Even though the government has attempted various aid programs, the reality on the ground shows that there are still many families who have not received aid that is appropriately targeted [2]. Inaccurate aid distribution not only causes injustice, but also hampers the effectiveness of social programs as a whole. Therefore, a more accurate and objective decision support system is needed to identify families who are truly eligible for assistance[3].

One approach that can be used to answer this challenge is to apply data mining techniques, especially the C4.5 algorithm. This algorithm works by forming a decision tree based on the relationship between variables, so that it is able to display classification rules clearly and easily understood [4]. In addition, the C4.5 algorithm can handle numeric and categorical data in one processing process with the help of the gain ratio in its entropy calculation [5]. Through this technique, aid recipient data can be mapped and grouped based on relevant attributes to produce more targeted decisions [6][7].

The application of the C4.5 algorithm in the context of social assistance has been proven effective in various studies. Pratiwi et al. (2022) successfully used this algorithm to classify recipients of Cash Social Assistance (BST) in Keramas Village with an accuracy rate of 97.83%, indicating that C4.5 is reliable in determining the eligibility of recipients of assistance based on socio-economic data [8]. Likewise, research by Rizal et al. (2022) which applied the C4.5 algorithm to predict recipients of Non-Cash Food Assistance (BPNT) and obtained a high accuracy of 98.5%, supports the role of C4.5 in data-based decision modeling. [9].

To improve the performance of the algorithm in the classification process, pre-processing steps such as data normalization are needed. Differences in scale between attributes are often obstacles that affect the performance of the classification algorithm. [10]. One commonly used normalization method is the Z-Score, which standardizes the data by transforming the attribute values into a normal distribution with a mean of zero and a standard deviation of one [11]. With this normalization, the attributes have an equivalent scale so that the algorithm can work more optimally.

In addition to normalization, several studies have also shown that combining the C4.5 algorithm with attribute selection methods such as Chi-Square can improve classification accuracy. For example, in a study on tea aroma classification, the C4.5 algorithm was combined with Chi-Square to select the most relevant attributes. As a result, the model accuracy increased from 93.65% to 94.27% after applying the combination [12]. These findings suggest that preprocessing steps such as attribute selection are critical in improving the performance of the C4.5 algorithm, especially when used on data with many variables.

In the application of classification for determining social assistance recipients, the C4.5 algorithm has been widely used because it is able to produce clear and accurate decision trees. Research conducted by Fatricia & Irawan (2023) [13] showed that the C4.5 algorithm was effective in classifying the eligibility of social assistance recipients with higher accuracy than other methods such as K-NN and Naïve Bayes.

On the other hand, research conducted by Safitri et al. (2024) [14] proves that the use of Z-Score normalization can improve the classification accuracy of aid recipient data. However, these studies still use a single approach, either only applying the C4.5 algorithm without normalization, or only applying normalization without strategic combination with the classification method.

Based on these shortcomings, this study combines the C4.5 algorithm with Z-Score normalization to maximize the accuracy of social assistance recipient data classification. This study aims to develop a website-based system that combines the C4.5 algorithm and Z-Score normalization, so that it can assist the government in determining eligible and targeted aid recipients. By implementing this combination, this system is expected to be able to classify PKH recipients in Atu Lintang District more efficiently, transparently, and effectively.

II. RELATE AND WORKS

A. C4.5 Algorithm

The C4.5 algorithm is a classification method that forms a model in the form of a decision tree. This decision tree presents classification rules in the form of branches that are easy to understand and effective in the decision-making process. The advantage of this algorithm lies in its ability to handle a large number of attributes and select the best attribute based on the highest information gain value. Before determining the best attribute, the algorithm calculates the entropy value to measure the level of data diversity [15]. This algorithm has also been used in various classification scenarios, such as in the classification of social assistance recipients and sensor data processing, and has been proven to produce high accuracy in determining data-based decisions [8][9][12].

In the process of forming a decision tree, the C4.5 algorithm uses several calculations, including:

1. Entropy is used to measure the level of data diversity:

   $$Entrophy(S,A) = \sum_{i=1}^{n} -pi \times log2pi \qquad (1)$$

   Information :
   S : Class Set
   N : Number of Partitions S
   Pi : Proportion of Si to S

2. Gain is used to select the best attribute in data separation:

   $$Gain\ (S,A) = Entrophy(S) \sum_{i=1}^{n} \frac{|si|}{|s|} \times Entrophy(S_i) \qquad (2)$$

   Information:
   S : Case Set
   A : Attribute
   n : Number of Attribute Partitions A
   (Si) : Number of Cases of Partition i
   (S) : Number of Cases In S

3. Split Info calculates how much data is split:

   $$SplitInfo\ (S,A) = \sum_{i=1}^{n} \frac{|Si|}{|S|} * log_2 \frac{|Si|}{|S|} \qquad (3)$$

   Information:
   S : Case Set
   A : Attribute
   |Si| : Number of cases in partition k-i
   |S| : Number of cases in S

4. Gain Ratio is the ratio between gain and split info which is used to determine the root of the tree:

   $$Gain\ Ratio = \frac{Gain\ (A)}{SplitInfo\ (S,A)} \qquad (4)$$

B. Z-Score Normalization

Z-Score Normalization is a data preprocessing technique used to standardize numeric attribute values to be on a comparable scale. This process is done by changing each data value into a standard value based on the average and standard deviation of all data. Data that has a value above the average will be positive, while data below the average will be negative. In this way, the attribute values will have a distribution with a mean of zero and a standard deviation of one[16].

Z-Score normalization is very important to be applied before the classification process, because it can eliminate the influence of scale differences between attributes. This will make classification algorithms such as C4.5 work more fairly on all attributes, simplify the comparison process between data, and increase the accuracy of the classification results produced[17]. The application of this technique in the context of social assistance classification has also shown that Z-Score normalization can improve model accuracy, especially when used in conjunction with the C4.5 algorithm [14].

The formula used in the Z-Score method is as follows:

$$Z = \frac{X - \mu}{\sigma} \quad (5)$$

Information:

Z : Z-Score value (normalization result)
X : The original value of the data to be normalized
μ (mu) : Average attribute
σ (sigma) : Standard deviation of attributes

C. Confusion Matrix

In this study, the evaluation of the classification result performance was carried out using a confusion matrix. This matrix functions to compare the model prediction results with actual data [18]. From this matrix, evaluation metrics such as accuracy, precision, recall, and F1-score can be obtained, which are used to measure the extent to which the C4.5 algorithm model with Z-Score normalization is able to classify PKH recipient data correctly [19]. These evaluation metrics are very important because they provide a comprehensive picture of the model's performance, not only in terms of prediction accuracy, but also the model's balance in handling eligible and unfit data. The use of confusion matrices like this is also commonly applied in other classification studies to measure the effectiveness of the methods used, including in studies based on the C4.5 algorithm [12]. The following is a confusion matrix table.

Table 1. Confusion Matrix

	Class	
	Positive	**Negative**
Positive	Positive True	False Positive
Negative	Negative False	Negative True

Where TP is the number of positive data correctly classified by the system. TN is the number of negative data correctly classified by the system. FN is the number of positive data classified as negative by the system. FP is the number of negative data classified as positive by the system. Here are the formulas and their explanations:

1. Accuracy : Measures how often the model makes correct predictions out of the total predictions.

$$Accuracy\ (\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

2. F1-Score : Provides a balance between precision and recall in one value.

$$F1 - Score\ (\%) = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \times 100\% \quad (7)$$

3. Precision : Measures how many positive predictions are actually correct.

$$Precision\ (\%) = \frac{TP}{TP+FP} \times 100\% \qquad (8)$$

4. Recall : Measures the ability of a model to find all true positive examples.
$$Recall\ (\%) = \frac{TP}{TP+FN} \times 100\% \qquad (9)$$

In this study, the C4.5 algorithm is used as the main method for classifying aid recipients. As a comparison, the author also considered other algorithms such as Random Forest and SVM, but this comparison was based on literature studies, not through direct testing.

The main advantage of C4.5 lies in its ability to produce a decision tree model that can be explained visually and is easy to understand, especially by non-technical users such as social workers. While Random Forest and SVM tend to be more complex and do not produce classification rules that are easy to trace. This is also in line with the findings of Maulana et al. (2025), which showed that although Random Forest has a higher accuracy (87.02%) than C4.5 (78.52%), the C4.5 model is easier to interpret because it produces a clear and transparent decision tree [20]. Based on the test results and supported by the literature, C4.5 is considered the most balanced and appropriate algorithm for use in this study because it is able to combine competitive accuracy with easy-to-understand interpretation of results, and supports the integration of web-based systems that have been built.

III. METHODOLOGY

A. Research Flow

This research uses a method consisting of literature study, needs analysis, data collection, system design, system implementation, with a flow of stages using the waterfall approach illustrated in Figure 1.

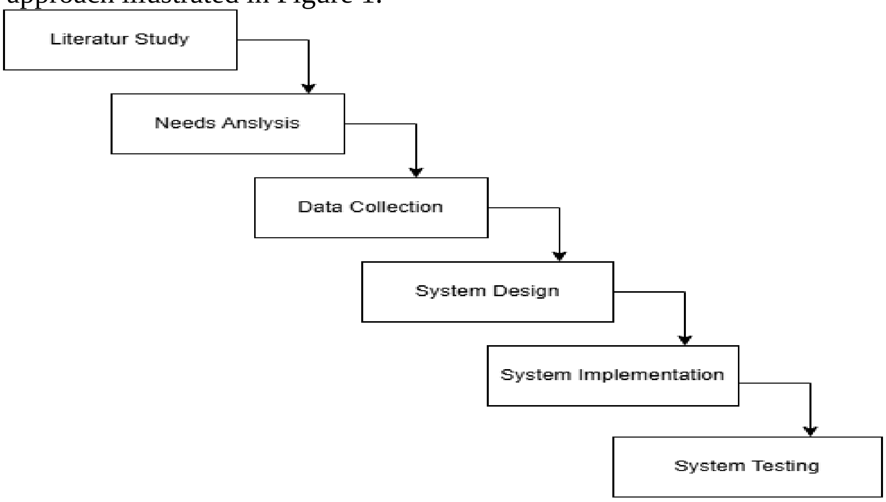

Figure .1. Research Flowchart

B. System Scheme

The stages in the system scheme show the process that will be carried out in the implementation of the system, in this system it will go through several stages, namely data collection, data processing, Z-score normalization, criteria selection, C4.5 model development, model evaluation, model performance decisions, new data classification, and displaying the classification results output, the system scheme is illustrated in Figure 2.

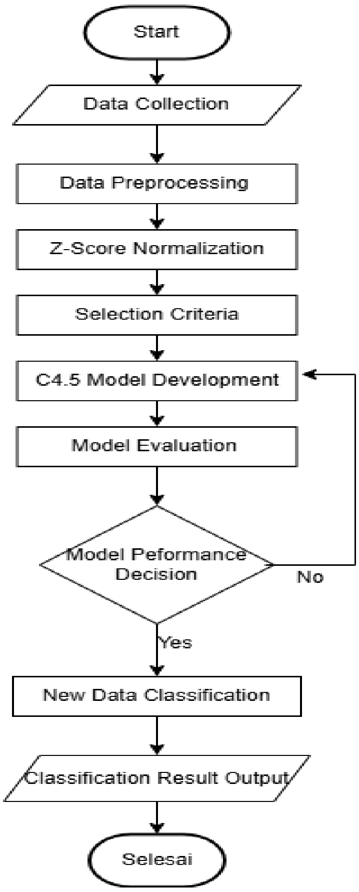

Figure .2. System Scheme

IV. RESULT AND DISCUSSION

This study produces a classification system for recipients of the Family Hope Program assistance in Atu Lintang District using the C4.5 algorithm, which is built on a website with Python and Flask, with the application of Z-Score normalization in the data processing process to equalize the variable scale and increase classification accuracy, so that it is hoped that the distribution of PKH assistance can be carried out more precisely.

A. Training Dataset Analysis

Before the classification process is carried out, the data is first collected and arranged in a table using Google Collab to facilitate the processing and analysis process. The Dataset display used in the study can be seen in the following image:

```python
import pandas as pd
import numpy as np

df = pd.read_csv('dataset.csv')
df
```

	NAMA	ALAMAT	AUD	SD	SMP	SMA	DISABILITAS	LANSIA	HAMIL	NOMINAL	STATUS RUMAH	JENIS DINDING	FASILITAS JAMBAN	KENDARAAN	PEKERJAAN
0	KASTRI	ATU LINTANG	0	0	0	0	0	2	0	1200000	1	2	2	2	3
1	NURUL FATIMAH	ATU LINTANG	1	1	0	0	0	0	0	975000	2	2	1	1	1
2	FITRI SURYANI	ATU LINTANG	0	1	0	0	0	0	0	225000	2	2	1	2	1
3	SUPARNI	ATU LINTANG	0	0	0	0	0	1	0	600000	3	1	1	1	4
4	RUMINI	ATU LINTANG	0	1	1	0	0	0	0	600000	1	1	1	1	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
546	NUR NAFIDAH	TANOH ABU	0	0	0	0	0	0	1	750000	2	1	1	1	1
547	MUJIEM	TANOH ABU	0	0	0	0	0	1	0	600000	1	1	1	2	1
548	SALMA	TANOH ABU	0	1	1	0	0	0	0	600000	1	1	1	1	1
549	DARMILAWATI	TANOH ABU	0	1	0	1	0	0	0	725000	1	1	1	1	2
550	SUMARMI	TANOH ABU	0	0	0	0	0	1	0	600000	1	1	1	2	1

551 rows × 15 columns

Figure .3. Research Dataset

Figure 1 shows data from 551 residents of Atu Lintang District covering various attributes, such as number of AUD, SD, SMP, SMA, status disabilitas, lansia, hamil, jenis dinding rumah, fasilitas jamban, dan pekerjaan. Most of the data shows limited economic conditions, with the focus of assistance on improving quality of life. This data will be analyzed using the C4.5 algorithm and normalized using the Z-Score method so that each attribute has an equivalent scale.

```python
# 1. Membaca dan Persiapan Data
print("Langkah 1: Persiapan Data")
data = pd.read_csv("data_bantuan_with_target.csv")  # Pastikan file ini ada
print("Data Awal (5 baris pertama):")
print(data.head())


# Pilih fitur (atribut) dan target
features = ['STATUS RUMAH', 'JENIS DINDING', 'FASILITAS JAMBAN', 'KENDARAAN', 'PEKERJAAN',
            'AUD', 'SD', 'SMP', 'SMA', 'DISABILITAS', 'LANSIA', 'HAMIL', 'NOMINAL']
X = data[features]
y = data['TARGET']
```

Figure .4. Data Preparation and Feature Selection

In the first step in this analysis, the data is loaded into the system using the pandas library via the read_csv function to read the data_bantuan_with_target.csv file. The first five rows are displayed with data.head() to verify the data structure. Next, all variables are entered as features, namely STATUS RUMAH, JENIS DINDING, FASILITAS JAMBAN, KENDARAAN, PEKERJAAN, AUD, SD, SMP, SMA, DISABILITAS, LANSIA, HAMIL, and NOMINAL. Meanwhile, the TARGET variable is used as the classification target. This process ensures that the data used is in accordance with the analysis objectives and is ready to be used for model training.

```python
# 2. Normalisasi Z-Score
print("\nLangkah 2: Normalisasi Z-Score")
scaler = StandardScaler()
X_normalized = scaler.fit_transform(X)

# Konversi ke DataFrame untuk kemudahan interpretasi
X_normalized_df = pd.DataFrame(X_normalized, columns=features)
print("Data Setelah Normalisasi Z-Score (5 baris pertama):")
print(X_normalized_df.head())

# Simpan data yang dinormalisasi (opsional)
X_normalized_df['TARGET'] = y
X_normalized_df.to_csv("data_normalized.csv", index=False)
print("Data yang dinormalisasi disimpan ke 'data_normalized.csv'")
```

Figure .5. Z-Score Normalization

Then, data normalization is performed using the Z-Score method to equalize the scale of all features so that none dominates the classification process. This process is carried out with StandardScaler from the sklearn.preprocessing library using the fit_transform function, which calculates the mean and standard deviation of each feature, then converts its value to a Z score (with a mean of 0 and a standard deviation of 1). The normalization results are stored in the X_normalized variable, then converted back to DataFrame form using pandas to make it easier to view. The normalized data is also saved in the data_normalized.csv file for further analysis.

```
Langkah 2: Normalisasi Z-Score
Data Setelah Normalisasi Z-Score (5 baris pertama):
   STATUS RUMAH  JENIS DINDING  FASILITAS JAMBAN  KENDARAAN  PEKERJAAN  \
0     -0.697445       0.855982          1.264538    0.99457   0.900398
1      0.636903       0.855982         -0.790803   -1.00546  -0.937081
2      0.636903       0.855982         -0.790803    0.99457  -0.937081
3      1.971251      -1.168249         -0.790803   -1.00546   1.819137
4     -0.697445      -1.168249         -0.790803   -1.00546   0.900398

        AUD        SD       SMP       SMA  DISABILITAS    LANSIA     HAMIL  \
0 -0.433510 -0.769771 -0.536081 -0.522032    -0.264392  2.711508 -0.287254
1  2.191371  0.940486 -0.536081 -0.522032    -0.264392 -0.608583 -0.287254
2 -0.433510  0.940486 -0.536081 -0.522032    -0.264392 -0.608583 -0.287254
3 -0.433510 -0.769771 -0.536081 -0.522032    -0.264392  1.051462 -0.287254
4 -0.433510  0.940486  1.865389 -0.522032    -0.264392 -0.608583 -0.287254

    NOMINAL
0  1.671063
1  0.866988
2 -1.813264
3 -0.473138
4 -0.473138
Data yang dinormalisasi disimpan ke 'data_normalized.csv'
```

Figure .6. Z-score Normalization Output Results

After Z-Score normalization is performed, all features such as STATUS RUMAH, JENIS DINDING, FASILITAS JAMBAN, KENDARAAN, PEKERJAAN, AUD, SD, SMA, DISABILITAS, to NOMINAL have been scaled to have a mean of 0 and a standard deviation of 1. This process ensures that no feature dominates due to differences in scale, and helps the model analyze the data proportionally. The normalization results are stored in the data_normalized.csv file for use in the training and subsequent analysis stages.

```python
# 3. Pemisahan Data dan Pelatihan Model
# Bagi data menjadi set pelatihan dan pengujian (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X_normalized_df[features], y, test_size=0.2, random_state=42)

model = DecisionTreeClassifier(criterion='entropy', random_state=42)
model.fit(X_train, y_train)

# Prediksi pada data pengujian
y_pred = model.predict(X_test)
```

Figure .7. Data Separation and Model Training

At this stage, the data is divided into two parts, namely 80% for training and 20% for testing using the train_test_split function. The goal is for the model to be trained and tested with different data to avoid overfitting and produce objective evaluations. The DecisionTreeClassifier model with the 'entropy' criterion is used for training using X_train and y_train, then tested with X_test using model.predict(X_test). This step is important to determine the extent to which the model is able to classify the status of PKH aid recipients well on data that has never been seen before.

```
# 4. Evaluasi Model
print("\nLangkah 4: Evaluasi Model")
# Hitung metrik evaluasi
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, pos_label='Layak')
recall = recall_score(y_test, y_pred, pos_label='Layak')
f1 = f1_score(y_test, y_pred, pos_label='Layak')

# Tampilkan hasil evaluasi
print("Hasil Evaluasi Model:")
print(f"Akurasi: {accuracy:.2f}")
print(f"Precision (Layak): {precision:.2f}")
print(f"Recall (Layak): {recall:.2f}")
print(f"F1-Score (Layak): {f1:.2f}")

# Tampilkan classification report untuk detail lebih lanjut
print("\nClassification Report:")
print(classification_report(y_test, y_pred))

# Tampilkan confusion matrix
cm = confusion_matrix(y_test, y_pred, labels=['Layak', 'Tidak Layak'])
cm_df = pd.DataFrame(cm, index=['Layak', 'Tidak Layak'], columns=['Layak', 'Tidak Layak'])
print("\nConfusion Matrix:")
print(cm_df)

# Visualisasi Confusion Matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm_df, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.show()
```

Figure .8. Model Evaluation

After the model is trained, an evaluation is conducted to assess its ability to classify the status of PKH recipients. This evaluation includes calculating accuracy, Precision, Recall, and F1-Score, especially in the Eligible category, to measure the accuracy and success of the model's predictions. A Classification report is presented to see the model's performance in detail in both categories. In addition, a Confusion Matrix is calculated and visualized using a heatmap to illustrate the number of correct and incorrect predictions, thus providing a clearer picture of the model's accuracy and errors in classification.

```
Langkah 4: Evaluasi Model
Hasil Evaluasi Model:
Akurasi: 0.94
Precision (Layak): 0.96
Recall (Layak): 0.90
F1-Score (Layak): 0.93

Classification Report:
              precision    recall  f1-score   support

       Layak       0.96      0.90      0.93        52
 Tidak Layak       0.92      0.97      0.94        59

    accuracy                           0.94       111
   macro avg       0.94      0.93      0.94       111
weighted avg       0.94      0.94      0.94       111


Confusion Matrix:
             Layak  Tidak Layak
Layak           47            5
Tidak Layak      2           57
```

Figure .9. Model Evaluation Results and Confusion Matrix

After training, the model was evaluated to assess its performance in classifying the status of PKH recipients. As a result, the model achieved a high accuracy of 94%. The precision of the Eligible category was 0.96 and Not Eligible 0.92 indicating good prediction accuracy. The recall was also high, namely 0.90 for Eligible and 0.97 for Not Eligible, indicating the model's ability to recognize each category. The F1-Score of 0.93 for Eligible showed a balance between Precision and Recall. The Classification report strengthened the results with macro and weighted average which were also 0.94. The Confusion Matrix showed that out of 52 Eligible data, 47 were classified correctly and 5 incorrectly; while out of 59 Not Eligible data, 57 were correctly and 2 were incorrectly predicted.

B. Classification Result

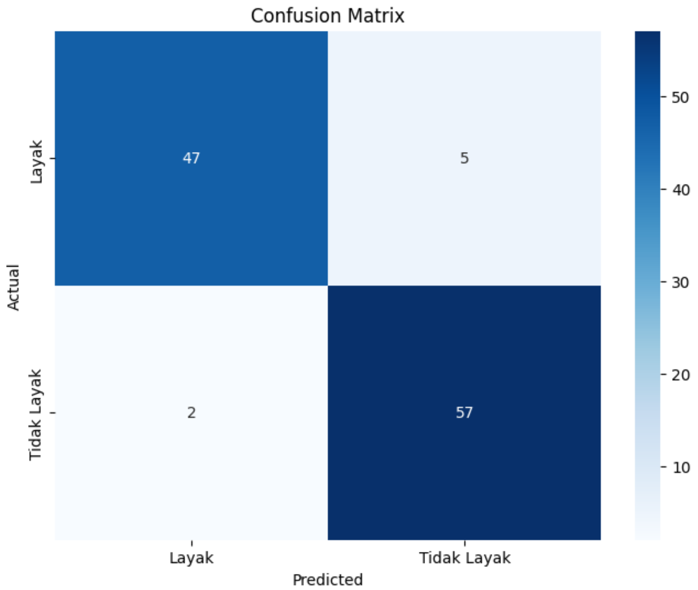

Figure .10. Confusion Matrix Visualization

. In this Confusion Matrix visualization, the vertical axis (Actual) shows the actual labels of the data, while the horizontal axis (Predicted) shows the labels of the model's predicted results. The colors on the heatmap represent the amount of data in each combination of actual and predicted, with darker colors indicating a larger amount. This visualization reinforces the previous evaluation results that the model has high classification ability, especially in recognizing the Tidak Layak category, with very low prediction errors in both classes. This kind of visualization is also very helpful in analyzing model performance intuitively and quickly.

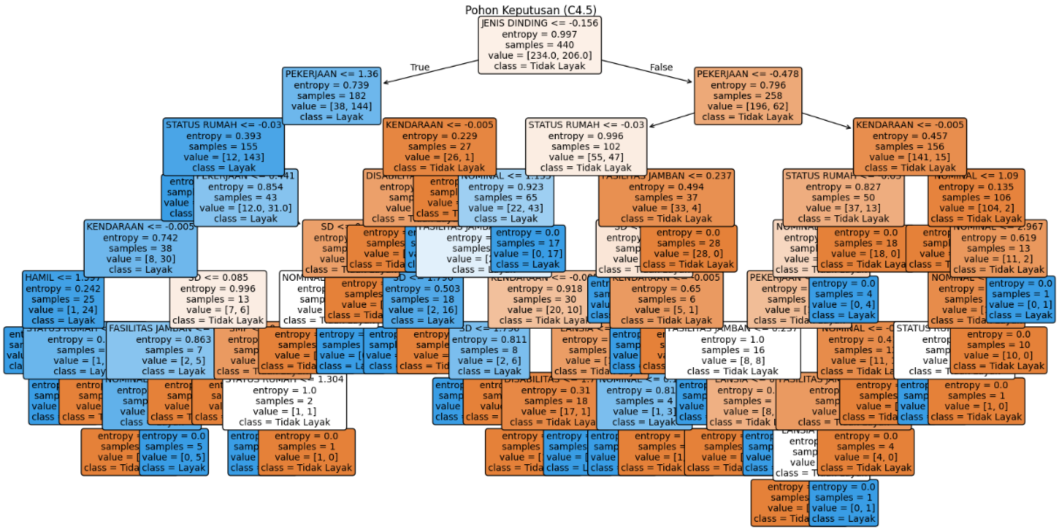

Figure .11. Decision Tree Visualization

The decision tree visualization of the C4.5 algorithm shows how the model makes decisions based on features in the data. Each node represents a split based on the value of a feature, for example, the feature * PEKERJAAN* is used in the initial split with a threshold of 0.479. The nodes display information such as the number of data points, predicted class, and entropy value. Other features such as STATUS RUMAH*, *KENDARAAN*, *JAMBAN*, and *HAMIL* also play a role in subsequent splits. This visualization helps understand the classification flow and improves the interpretability of the model.

C. Website Implementation

The data page on this system is designed to make it easier for admins to manage information on PKH recipient families. Admins can add data manually via the "Tambah Data Keluarga" button or upload bulk data using a CSV file.

A filter feature based on address is also provided to facilitate searching for specific data. The interface of this page is made simple and intuitive, so that admins can add, edit, or delete data easily and maintain smooth system operations.

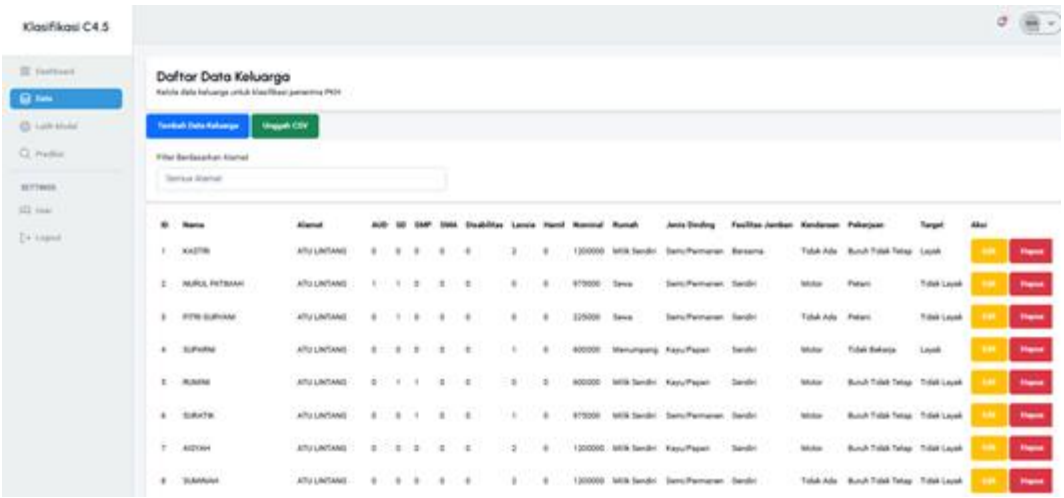

Figure .12. Family Date Page

The "Prediksi kelayakan Bantuan" page allows users to enter family data to be processed by the C4.5 model. Users fill in all important attributes and then press the "Prediction" button to see the results of the PKH recipient eligibility classification. This feature is useful for helping social workers make decisions more quickly, objectively, and on target based on data that has been processed systematically.

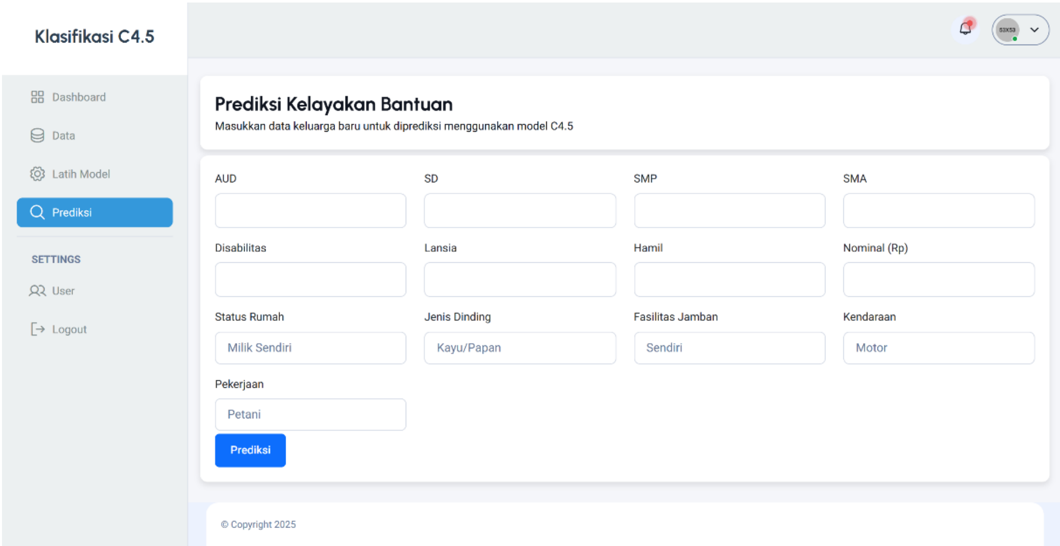

Figure .13. Prediction Page

The implementation of the website is a visual representation of the output of the PKH recipient eligibility classification system that has been developed. This implementation also aims to evaluate whether the system is functioning well or still requires improvements or additions.

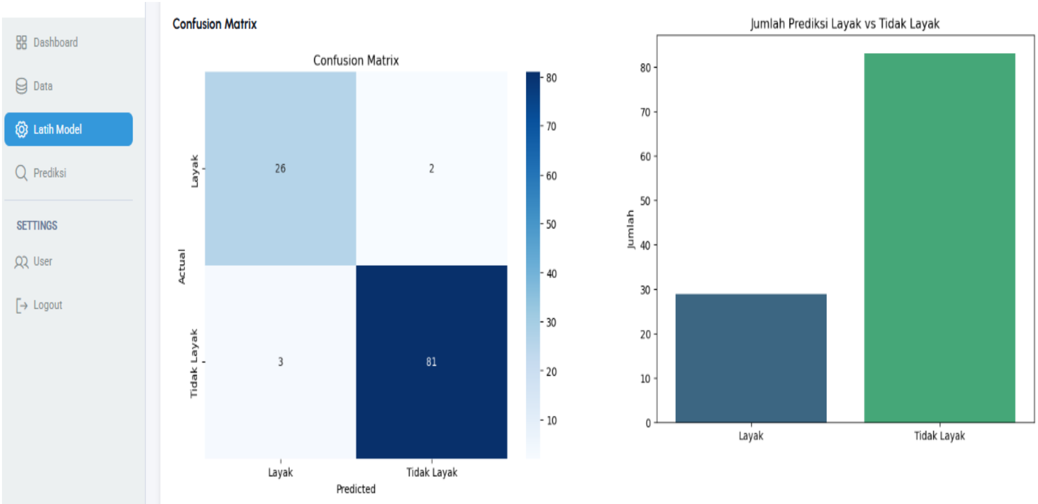

Figure .14. Visualization of Confusion Matrix and Prediction Distribution

Figure 10 shows the visualization of the Confusion Matrix results and the diagram of the classification results of aid recipients who are considered eligible and ineligible on the Algorithm C.4.5 classification website. The colors on the heatmap represent the amount of data in each combination of actual and predicted, with darker colors indicating a larger amount.

D. Discussion

In developing a social assistance recipient classification system, it is important to pay attention to ethical aspects in automated decision-making. The data used in the system must be kept confidential so that it is not misused. In addition, the classification process needs to be designed so that it does not discriminate against certain groups. This system is not intended to completely replace the role of humans, but rather as a tool in the decision-making process. Therefore, the classification results produced still need to be reviewed by social workers before being used as the basis for final decisions.

V. CONCLUSION

This study shows that the application of the C4.5 algorithm is able to classify the status of Family Hope Program (PKH) recipients with an accuracy of 94%, effectively distinguishing Eligible and Ineligible recipients. The model evaluation resulted in a Precision of 0.96, Recall of 0.90, and F1-Score of 0.93 for the Eligible category, indicating a strong balance between model accuracy and sensitivity. The confusion matrix revealed that the model successfully classified 47 recipients as Eligible with 5 errors, and 57 recipients as Ineligible with only 2 classification errors, so the error rate is relatively low. The factors PEKERJAAN, JENIS DINDING, and STATUS RUMAH were found to be the most influential features in the classification process. In addition, the model not only provides classification results, but also produces probability predictions, where testing on new data shows a 100% confidence level in the Eligible prediction, demonstrating the reliability of the model in supporting decision making related to aid distribution.

REFERENCES

[1] Erfain *et al.*, “Efektifitas Pelaksanaan Program Keluarga Harapan dalam Upaya Pengentasan Kemiskinan di Kecamatan Batuatas Kabupaten Buton Selatan,” *Arus J. Sos. dan Hum.*, vol. 1, no. 3, pp. 108–117, 2021, doi: 10.57250/ajsh.v1i3.40.

[2] M. Qamal, I. Sahputra, N. Nurdin, M. Maryana, and M. Mukarramah, “Sistem Pendukung Keputusan Penentuan Penerimaan Bantuan PKH Menggunakan Metode Naïve Bayes,” *TECHSI - J. Tek. Inform.*, vol. 14, no. 1, p. 21, 2023, doi: 10.29103/techsi.v14i1.6960.

[3] U. S. Mangunsong, Arridha Zikra Syah, and Endra Saputra, “Implementasi Sistem Seleksi Calon Penerima Subsidi Program Keluarga Harapan Dengan Metode SAW Di Kelurahan Sijambi,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 4, no. 3, pp. 737–747, 2024, doi: 10.51454/decode.v4i3.559.

[4] J. Rohmadan Wahari and Y. Reswan, “Implementation of the C45 Algorithm for Predicting the Visit Rate of Patients at dr. Daniel Bengkulu Implementasi Algoritma C45 untuk Memprediksi Tingkat Kunjungan Pasien pada Klinik dr. Daniel Bengkulu,” *J. Kom.*, vol. 2, no. 1, pp. 1–8, 2022, [Online]. Available: https://doi.org/10.53697/jkomitek.v2i1

[5] Z. Situmorang, S. Mandasari, Y. Franciska, K. Andriyani, and P. S. Ramadhan, “Algoritma C45 Dalam Memprediksi Minat Calon Mahasiswa,” *J. Sci. Soc. Res.*, vol. 5, no. 1, p. 125, 2022, doi: 10.54314/jssr.v5i1.809.

[6] J. Amarda, N. Suarna, and O. Nurdiawan, “Klasifikasi Industri Kerajinan Berdasarkan Provinsi Menggunakan Metode Decision Tree Untuk Membandingkan Pengujian Algoritma Id3 Algoritma C45 Dan Algoritma Cart,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 790–796, 2022, doi: 10.36040/jati.v6i2.5736.

[7] Asrianda, L. Rosnita, B. Jange, and Zulfadli, “Multi-criteria K-nearest neighbor in the classification of eye diseases at Dr. Fauziah Bireuen Hospital,” *Instal J. Komput.*, vol. 15, no. 12, pp. 439–446, 2023, [Online]. Available: https://journalinstal.cattleyadf.org/index.php/Instal/index

[8] N. W. Oktha Pratiwi, N. Widya Utami, and I. Gede Juliana Eka Putra, “Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar, Bali,” *J. Inform. Teknol. dan Sains*, vol. 4, no. 3, pp. 101–107, 2022, doi: 10.51401/jinteks.v4i3.1667.

[9] R. A. Islahudin, S. Rahmatullah, A. Afandi, and S. Safitri, “Algoritma C4.5 Untuk Memprediksi Kelayakan Penerima Bantuan Pangan Non Tunai,” *J. Inform.*, vol. 22, no. 2, pp. 147–159, 2022, doi: 10.30873/ji.v22i2.3367.

[10] M. Lasena, R. Aminuddin, and Z. Azhar, “Pembentukan Pola Peminjaman Buku Pada Perpustakaan Dengan Menerapkan Metode CART dan Normalisasi Z-Score,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 314–324, 2024, doi: 10.47065/bits.v6i1.5238.

[11] M. L. Radhitya and G. I. Sudipa, “Pendekatan Z-Score Dan Fuzzy Dalam Pengujian Akurasi Peramalan Curah Hujan,” *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 149–156, 2020, doi: 10.31598/sintechjournal.v3i2.567.

[12] M. A. Barata, Edi Noersasongko, Purwanto, and Moch Arief Soeleman, “Improving the Accuracy of C4.5 Algorithm with Chi-Square Method on Pure Tea Classification Using Electronic Nose,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 2, pp. 226–235, 2023, doi: 10.29207/resti.v7i2.4687.

[13] I. P. Fatricia and H. Irawan, “Penerapan Algoritme C4. 5 Terhadap Kelayakan Bantuan Pangan Non Tunai (Bpnt) Pada Kantor Kelurahan Kedaung,” *Pros. Semin. Nas. …*, vol. 2, no. September, pp. 1049–1058, 2023, [Online]. Available: http://senafti.budiluhur.ac.id/index.php/senafti/article/view/1007%0Ahttp://senafti.budiluhur.ac.id/index.php/senafti/article/download/1007/583

[14] N. Safitri, D. Kusnandar, and S. Martha, “Implementasi Algoritma K-Nearest Neighbor Dengan Normalisasi Z-Score Dalam Klasifikasi Penerima Bantuan Sosial Desa Serunai,” *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 1, pp. 99–106, 2024.

[15] R. H. Pambudi, B. D. Setiawan, and Indriati, “Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal,” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2637–2643, 2020, [Online]. Available: http://j-ptiik.ub.ac.id

[16] M. Fahmi, “Penerapan Algoritma K-Means Data Mining Pada Clustering Kelayakan Penerima UKT Dengan Normalisasi Data Model Z-Score,” vol. 6, no. 3, 2024, doi: 10.47065/bits.v6i3.6475.

[17] I. Permana and F. N. S. Salisah, “Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation,” *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022, doi: 10.57152/ijirse.v2i1.311.

[18] R. N. Amalda, N. Millah, and I. Fitria, “Implementasi Algoritma C5.0 Dalam Menganalisa Kelayakan Penerima Keringanan Ukt Mahasiswa Itk,” *Teorema Teor. dan Ris. Mat.*, vol. 7, no. 1, p. 101, 2022, doi: 10.25157/teorema.v7i1.6692.

[19] F. R. Valerian *et al.*, “Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix,” vol. 9, no. 2, pp. 2242–2249, 2025.

[20] B. Maulana and D. Febrian, “Comparison of Support Vector Machine , Random Forest , and C4 . 5 Algorithms for Customer Loss Prediction,” vol. 2, no. February, pp. 1–6, 2025.