Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Socio-Economic Status Classification of Neighborhood Residents Using the Decision Tree Algorithm

Elta Putri Setia Nengsi[1], Dia Komalla[2*], Ardeya Wulandari[3], Cintia Novita Lorensya[4], Mufid Faruq Aziz[5]

[1,2,3,4,5]*Teknik Informatika, Teknik, Universitas Muhammadiyah Bengkulu*
tatadorable21@gmail.com[1], diakumalla2@gmail.com[2], ardeyawulandari4@gmail.com[3], cintialorensya@gmail.com[4], mufidfaruq.a@gmail.com[5]*

Abstract

This study aims to analyze the socioeconomic classification of RT residents using the Decision Tree algorithm. The analysis was carried out using data that includes attributes such as the number of family members, education, and occupation. The results of the study show that the Decision Tree algorithm is capable of producing a clear and structured classification model, with the number of family members being the dominant factor in class distribution. Most residents were classified into the Middle socioeconomic category (68.3%), followed by the Low category (26.8%), and the High category (4.9%). These results reflect that the majority of residents have relatively stable socioeconomic conditions, although there are still groups that require special attention. This classification model provides important insights for policymakers to design more focused assistance and economic empowerment programs. This study also recommends further development by adding more diverse attributes and comparing the Decision Tree algorithm with other classification methods to improve the model's accuracy and validity.

Keywords: *Decision Tree, Number of Family Members, Occupation, Socio-Economic Classification*

1. Introduction

This The distribution of social assistance is one of the government's programs aimed at improving community welfare, especially for underprivileged families. However, in practice, the distribution process still faces various obstacles, one of which is the inaccuracy in targeting aid recipients. This often results in assistance being received by those who do not truly need it, while residents who genuinely require support fail to obtain it.

[1] stated that in practice, the distribution of basic food assistance is often mistargeted, thus requiring more valid data collection regarding families eligible or ineligible to receive aid. Therefore, a system is needed that can objectively, transparently, and data-based classify the eligibility of social assistance recipients.

Along with technological advancements, the use of algorithms in the field of data science has become a potential solution to support decision-making processes. One algorithm that can be utilized is the Decision Tree. Decision Tree is a machine learning method that forms a predictive model in the form of a decision tree. This algorithm can determine aid recipients based on several parameters such as income, occupation, and the number of dependents in the family. In previous studies, Decision Tree algorithms, including C5.0 and CART, have been widely used for social welfare classification due to their ability to efficiently and accurately process large datasets [2]. Decision Tree algorithms have been widely applied in various fields, including data processing to determine the eligibility of social assistance recipients. By using residents' socio-economic characteristic data, this algorithm can help predict recipients' eligibility more accurately and systematically.

Several previous studies related to the issue of social assistance eligibility have been conducted. One of them is a study titled "Classification of Social Assistance Recipients Using the C4.5 Algorithm", which utilized the C4.5 algorithm to determine the eligibility of social assistance recipients in Tangerang City. The results of this study indicated that the C4.5 classification method could provide objective and accurate decisions in determining eligibility based on criteria such as income, number of dependents, housing status, and number of vehicles, with income being the most dominant factor.

Furthermore, a study by Ni Wayan Oktha Pratiwi and Nengah Widya Utami (2022) entitled "Classification for Determining Cash Social Assistance (BST) Recipients Using the C4.5 Algorithm" in Keramas Village, Gianyar, Bali applied the C4.5 algorithm, achieving an accuracy of 97.83%, higher than both K-Nearest Neighbor and Naïve Bayes. The most influential root node was the attribute "previously received other assistance", followed by occupation and number of dependents[3].

Another study by Aldiyansyah et al. (2024) in "Comparison of Decision Tree and Random Forest Algorithm Accuracy Rates in Classifying BPNT Social Assistance Recipients in Slangit Village" showed that both algorithms achieved very high accuracy of 97.86%, with variables such as income, number of children, age, and vehicle ownership as important factors in determining the eligibility status of recipients [4].

Additionally, a study by Rosanti (2022) in "Implementation of Machine Learning Models to Determine the Classification of Social Assistance Types" compared three methods: Certainty Factor, Naïve Bayes, and Decision Tree to classify types of social assistance received by residents in the DKI Jakarta area. The study results showed that Certainty Factor delivered the best accuracy at 98.4%, while Naïve Bayes and Decision Tree achieved accuracies of 93.3% each. This study also developed a web-based application to facilitate the determination of assistance types based on applicable criteria [5]

Lastly, a study by Reni Pratiwi in "Comparison of C5.0 and CART Algorithm Classifications on Social Data of Household Heads in Teluk Baru Village" compared the accuracy of C5.0 and CART algorithms on community social data. The results showed that CART achieved a higher average accuracy of 84.63%, compared to C5.0 at 79.17% [2]. This proves that classification methods using different Decision Tree algorithms provide varying performance depending on the data characteristics used.

This current study aims to develop a classification system for social assistance eligibility for residents of RT 5 by utilizing the Decision Tree algorithm. This system will use residents' socio-economic characteristics data, such as income, occupation, number of dependents, and other relevant variables as parameters to determine eligibility for assistance. It is expected that the results of this research can provide a more effective solution for distributing social assistance so that it reaches those who truly need it.
.

2. Literature Review

An Socio-economic classification has become an important topic in various fields of research, such as sociology, economics, and computer science. Socio-economic class is generally used to describe the position of individuals or groups within the social structure based on several factors, such as income, occupation, education, and other social status indicators. In this study, the main focus is on the classification of socio-economic class based on three main factors: education, occupation, and number of family members. In this study, the main focus is on the classification of socio-economic class based on three main factors: education, occupation, and number of family members, which have been shown to support accurate classification between upper-middle and lower-middle socio-economic classes through data segmentation and pattern recognition [6].

2.1. Education as an Indicator of Socio-Economic Class

Education is often considered one of the main determinants in classifying socio-economic status. Several studies show that a person's education level has a strong correlation with income and social status. Education enhances an individual's ability to secure better jobs, which in turn can improve their socio-economic status. Individuals with higher education levels tend to have access to better-paying jobs and broader career opportunities.

2.2. Occupation and Socio-Economic Class

Occupation is an important indicator in determining socio-economic class. Work reflects the social and economic capital possessed by the individual. Certain professions, such as civil servants or entrepreneurs, are often associated with higher socio-economic classes compared to jobs like daily laborers or farmers. Occupation can affect the level of economic stability and access to various social resources that support well-being.

2.3. Number of Family Members and Economic Welfare

The number of family members is an important factor in determining the economic well-being of a household. Households with more family members tend to have higher expenses, which can reduce the family's financial capacity. Conversely, households with fewer family members tend to be more financially stable. Therefore, the number of family members can serve as an important indicator in classifying a family's socio-economic class.

2.4. Socio-Economic Class Classification Using Machine Learning Algorithms

In recent decades, the development of machine learning technology has opened up new opportunities to classify socio-economic classes automatically. One commonly used algorithm in classification is the Decision Tree. Decision tree classification technique is one of the most popular data mining techniques [7] .The Decision Tree is a method that builds predictive models in the form of a tree structure, where each branch represents a decision based on a specific feature, and each leaf represents a class category. Decision trees can handle both categorical and numerical data whereas other techniques are specialized for only one type of variable [8], making them particularly suitable for socio-economic classification that involves mixed data types such as education levels, occupation categories, and numerical family size data.

The Decision Tree algorithm has advantages in handling data with categorical attributes (such as occupation, education, and socio-economic class) and can produce models that are easy to understand and interpret. Decision Trees can be applied to various classification problems, including socio-economic class prediction, by identifying rules that connect input variables with output classes.
Several previous studies have also employed Decision Trees for socio-economic class classification.

3. Methodology

This study employed a quantitative approach using an experimental method to analyze and predict socio-economic class based on variables such as education, occupation, and number of family members. The research process was divided into several stages: data collection, data analysis, and evaluation of the predictive model. The following provides a more detailed explanation of each stage of this study:

3.1. Research Design

This research was designed with the objective of building and testing a Decision Tree model to classify the socio-economic class of individuals. The classes to be predicted are socio-economic classes categorized into three levels: Low, Middle, and High. The features used for prediction are education, occupation, and number of family members. Decision trees are effective tools for simplifying complex decisions by providing a transparent, rule-based structure that enhances interpretability in classification tasks [9]. This characteristic makes them particularly suitable for socio-economic analysis where explainability of the decision process is essential.

3.2. Data Source

The data used in this study was collected through field surveys, which included information about individuals or households in their social and economic aspects. The data includes the following variables:

a. Education: The education level category possessed by the head of the household or main family member (e.g., completed elementary school, junior high school, senior high school, diploma, and others).
b. Occupation: The type of job performed by the head of the household or main family member, such as laborer, entrepreneur, trader, civil servant, and others.
c. Number of Family Members: The total number of family members in the household.
d. Socio-Economic Class: The socio-economic class category to be predicted, namely Low, Middle, and High.

A total of 42 individual records were collected, containing information about the attributes mentioned above.

Table 1: Respondent Data

Respondent Code	Education	Occupation	Number of Family Members	Socio-Economic Class
R1	Junior High School Equivalent	Daily Laborer	5	Low
R2	Junior High School Equivalent	Retired	4	Middle
R3	Diploma	Entrepreneur	4	Middle
…	…	…	…	…
R42	Senior High School Equivalent	Retired	1	Middle

3.3. Socio-Economic Class Classification

Socio-economic classification (SEC) is generally defined as a descriptive tool used to illustrate social stratification and inequalities within society. It does not inherently carry theoretical or analytical weight but is widely applied to reflect how societies are organized and differentiated based on socio-economic structures[10]. The classification of residents' socio-economic classes in this study was determined automatically by the Decision Tree model based on patterns formed from the analyzed data. The model used the Number of Family Members attribute as the main factor in determining the socio-economic class categories, with certain values set as split points.

The classification decisions were made without any predefined manual rules, but rather followed the results generated by the Decision Tree algorithm based on the combination of attribute values in the dataset.
The socio-economic classes of the residents were classified into three categories as the final result of the classification process:

- Low
- Middle
- High

These categories represent the final classification outcomes generated automatically by the Decision Tree model according to the structure of the decision tree formed.

3.4. Data Preprocessing

Before the data could be used for further analysis, several preprocessing steps were carried out, including:

a. Data Cleaning: Incomplete or invalid data were removed or imputed to prevent affecting the analysis results.
b. Data Transformation: Categorical variables, such as education and occupation, were encoded into numerical form using label encoding or one-hot encoding methods.
c. Normalization: Numerical data such as the number of family members were normalized so that all features had a comparable scale, which is important for some machine learning algorithms.

3.5. Data Partitioning

The obtained data was divided into two sets for model training and testing:

- Training Data: 80% of the data was used to train the model. In this stage, the Decision Tree model learned the pattern of relationships between the features (education, occupation, number of family members) and socio-economic class.
- Testing Data: 20% of the data was used to test the accuracy of the model after training.

3.6. Implementation with RapidMiner

This study was conducted using RapidMiner software, which provides a graphical interface for importing, processing, and analyzing data. The processes performed in RapidMiner included:

a. Importing the cleaned and processed dataset.
b. Applying data preprocessing methods using the available operators.
c. Building the Decision Tree model using the Decision Tree operator.
d. Evaluating the model using Cross Validation and Performance Evaluation operators.

4. Results and Discussion

4.1. Decision Tree Structure

a. The decision tree was built using Number of Family Members as the main attribute.
b. The root node is the Number of Family Members attribute, with a split value of > 5.500 or ≤ 5.500.
c. If the Number of Family Members > 5.500, the classification is directly assigned as High.
d. If the Number of Family Members ≤ 5.500, the decision tree continues splitting based on the next values of Number of Family Members, namely > 3.500, > 2.500, and > 1.500, until it reaches the leaf node with a final result of Middle or Low.

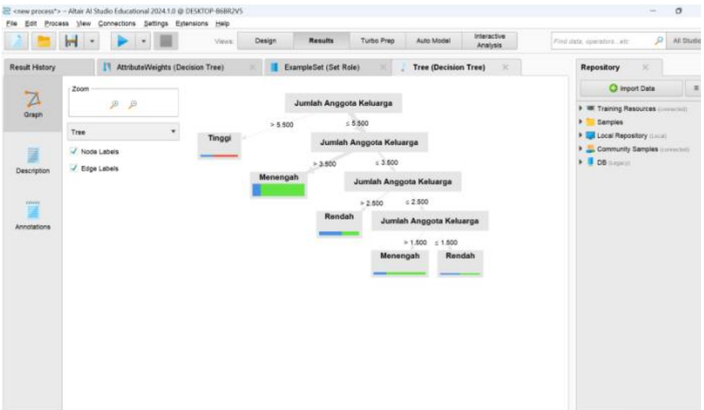

Fig. 1: Decision Tree Based on Number of Family Members

4.2. Classification Distribution

The results of the socio-economic classification of residents were divided into three categories:
a. Middle: 28 residents (68.3%)
b. Low: 11 residents (26.8%)
c. High: 2 residents (4.9%)

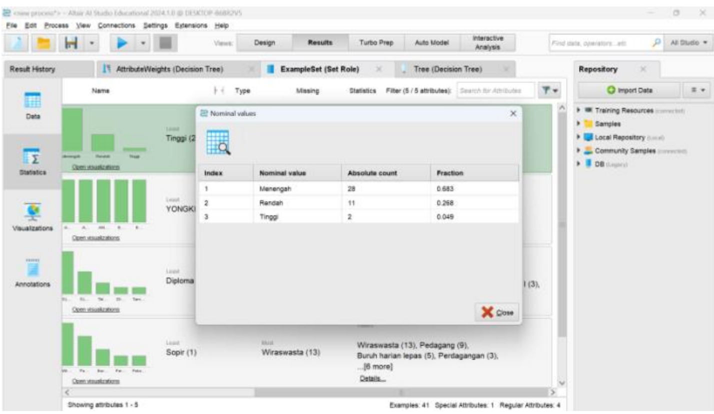

Fig. 2: Class Value Distribution in the Dataset

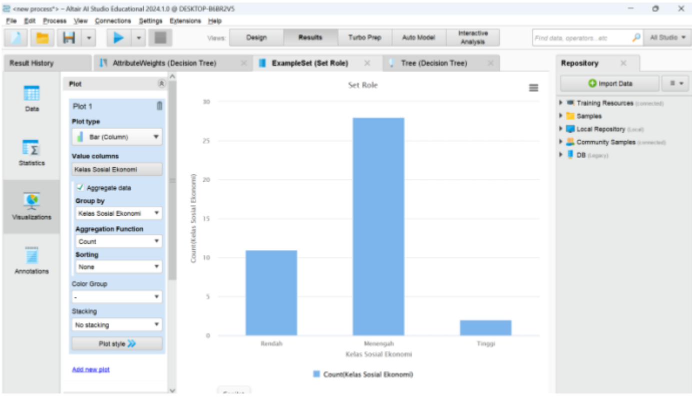

Fig. 3: Visualization of Socioeconomic Class Distribution in the Dataset

4.3. Dominant Attribute

The Number of Family Members attribute became the primary factor influencing the initial classification. This attribute was also used for further splitting of the data that did not meet the criteria at the root node, until reaching the leaf nodes.

5. Conclusion

This study demonstrated that the Decision Tree algorithm was successfully used to classify the socio-economic conditions of RT residents in a clear, structured, and easily interpretable manner. The main attribute, such as the Number of Family Members, proved to be significant in determining the classification, with the majority of residents falling into the Middle category (68.3%), followed by the Low category (26.8%), and the High category (4.9%). This reflects that most residents have a relatively stable socio-economic level, although there are still groups that require special attention to improve their welfare. These classification results have important implications for local policymakers, such as designing assistance programs focused on the Low category and economic empowerment programs to maintain the stability of the Middle category and encourage improvement toward the High category.

This study can be further developed by adding more relevant variables, using a larger dataset, or comparing the Decision Tree algorithm with other methods to obtain more optimal results. Thus, this research provides valuable insights into understanding the socio-economic conditions of residents and supports data-driven decision-making.

References

[1] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.

[2] R. Pratiwi, M. N. Hayati, and S. Prangga, “Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification and Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019),” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 2, pp. 273–284, 2020, doi: 10.30598/barekengvol14iss2pp273-284.

[3] I. G. J. E. P. Ni Wayan Oktha Pratiwi, Nengah Widya Utami, “KLASIFIKASI PENENTUAN PENERIMA BANTUAN SOSIAL TUNAI (BST) MENGGUNAKAN ALGORITMA C4.5 DI DESA KERAMAS, GIANYAR BALI,” *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 4, no. 3, pp. 101–107, 2022.

[4] A. Aldiyansyah, A. Irma Purnamasari, and I. Ali, “Perbandingan Tingkat Akurasi Algoritma Decision Tree Dan Random Forest Dalam Mengklasifikasi Penerima Bantuan Sosial Bpnt Di Desa Slangit,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 127–132, 2024, doi: 10.36040/jati.v8i1.8290.

[5] N. Rosanti, “Jurnal Teknologi Terpadu PENERAPAN MODEL MACHINE LEARNING UNTUK MENENTUKAN,” vol. 8, no. 2, pp. 127–135, 2022.

[6] J. Brmanda, “Klasifikasi Masyarakat Penerima Bantuan Sosial dari Pemerintah dengan Metode Algoritma C4.5,” *J. Komput. Antart.*, vol. 3, no. c, pp. 34–41, 2025.

[7] H. Sharma and S. Kumar, “A Survey on Decision Tree Algorithms of Classification in Data Mining,” *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.

[8] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, “Analysis of Various Decision Tree Algorithms for Classification in Data Mining,” *Int. J. Comput. Appl.*, vol. 163, no. 8, pp. 15–19, 2017, doi: 10.5120/ijca2017913660.

[9] T. C. Kershaw, S. Bhowmick, C. C. Seepersad, and K. Hölttä-Otto, “A decision tree based methodology for evaluating creativity in engineering design,” *Front. Psychol.*, vol. 10, no. JAN, pp. 1–19, 2019, doi: 10.3389/fpsyg.2019.00032.

[10] D. Rose and E. Harrison, “The European socio-economic classification: A new social class schema for comparative European research,” *Eur. Soc.*, vol. 9, no. 3, pp. 459–490, 2007, doi: 10.1080/14616690701336518.