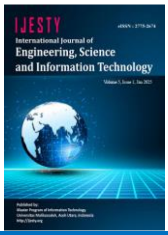
International Journal of Engineering, Science and Information Technology
Volume 5 No. 3 (2025) pp. 122-131
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i3.896
Research Paper, Short Communication, Review, Technical Paper



Analysis of Boarding House Feasibility and Satisfaction Using Data Mining with the C4.5 Algorithm Based on Service Quality and Facilities

Aktina Manik*, Dahlan Abdullah, Ar Razi

[1]*Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia*

**Corresponding author Email: aktina.210170046@mhs.unimal.ac.id*

The manuscript was received on 26 November 2024, revised on 10 January 2025, and accepted on 2 April 2025, date of publication 25 May 2025

Abstract

This research develops a boarding house eligibility classification system using the C4.5 algorithm based on service quality and available facilities. The system evaluates boarding house eligibility by considering various factors such as management services, cleanliness, security, room facilities, public facilities, internet access, comfort, and price. Each of these factors is given a specific weight based on its importance to the tenants, and they are used to classify boarding houses as luxury, standard, and economical. The classification results show that 43% of luxury boarding houses were deemed eligible, while 57% were not. In the standard boarding house category, 21% were classified as eligible, and 79% as ineligible, while in the economical category, 23% were eligible and 77% were ineligible. Using the Confusion Matrix and Classification Report, model evaluation revealed precision ranging from 0.4 to 1.0, recall from 0.67 to 1.0, and F1-scores from 0.5 to 0.91, demonstrating a reasonably high overall accuracy. Additionally, feature importance analysis revealed that price, water and electricity availability, and room facilities are the most influential factors in determining boarding house eligibility. The system's performance was tested against a dataset of real-world boarding houses, and the results suggest that it can accurately classify boarding houses based on key factors that affect tenant satisfaction. The system has the potential to serve as a valuable decision-making tool for boarding house owners, helping them improve service quality and for prospective tenants, enabling them to make more informed housing choices based on their preferences and needs.

Keywords: *Boarding House, C4.5, Classification, Data Mining, Machine Learning,*

1. Introduction

According to data from Statistics Indonesia (Badan Pusat Statistik/BPS), in 2023, intercity migration—referring to individuals relocating to other cities—reached 139.4 million people out of Indonesia's total population of 278.7 million (Badan Pusat Statistik, 2024). In other words, approximately 50% of Indonesians have engaged in intercity migration. One of the primary destinations for such migration is Aceh Province, which ranks among the top eight provinces with the highest number of incoming migrants [1]. The capital city of Aceh Province attracts a significant number of migrants, including both the general public and students. Aceh's economic growth, supported by trade, services, tourism, and real estate sectors, creates numerous opportunities for job seekers and students pursuing higher education. The presence of multiple educational institutions, particularly universities and colleges, is a significant factor driving the increase in student migration to the province. As a result, Aceh Province has become a center of growth in both economic and educational aspects. Blang Pulo Boarding House is strategically located among several essential buildings, making it ideal for migrants and students to carry out their activities and daily routines. The boarding house offers 1,000 rental units with varying levels of feasibility and facilities. However, some of the Blang Pulo Boarding House facilities are damaged, and the overall feasibility is low, prompting the owner to plan repairs. The influx of migrants into Aceh Province also drives the owner to expand the business by adding more rental units.
The increasing number of students opting to study in major cities has contributed to the rise in boarding houses or rental units. Various boarding houses are offered, supported by facilities that aim to enhance students' comfort and well-being, making them more inclined to choose these accommodations as their residence. In addition to facilities, the surrounding environment and the type of boarding house

also need to be considered. Parents whose children are pursuing education in another city often give them the autonomy to select a suitable boarding house to foster their children's independence, maturity, and comfort [2].

A boarding house is a building used as temporary accommodation for students or employees living far from their campus, office, or workplace. The primary purpose of boarding houses is to save time commuting to campus, office, or work. Additionally, having a boarding house helps reduce fatigue from lengthy journeys to reach the campus, office, or workplace [3].

In the boarding house business world, the quality of service and facilities offered are key factors influencing consumer decisions, especially in campus areas bustling with migrants, such as in Blang Pulo Village. At Blang Pulo Boarding House, located in the Muara Satu subdistrict, the quality of service and facilities plays a vital role in attracting tenants, particularly with the increasing number of migrants coming to the city. Boarding house tenants, consisting of the general public and students, consider various aspects when choosing a place to stay, such as cleanliness, security, comfort, and additional facilities such as internet access and en-suite bathrooms.

However, with the issues regarding some facilities at Blang Pulo Boarding House and the increasing competition in the boarding house industry, the owner needs to understand the factors determining tenant feasibility and satisfaction. One way to achieve this is by analyzing tenant data using Data Mining methods. The C4.5 algorithm is one of the practical algorithms for building decision models based on the available data. This algorithm can generate decision trees that help classify tenant feasibility and satisfaction based on various variables, such as service quality and available facilities.

Applying the C4.5 algorithm will help boarding house owners understand tenant preferences and make targeted service improvements or enhancements. Through deeper data analysis, owners can determine strategic steps to improve the feasibility of the boarding house while increasing tenant or customer satisfaction [4]. The C4.5 algorithm is one of the classification techniques in machine learning used in the Data Mining process by constructing a decision tree represented in rules, generating new knowledge [5].

The study by [6] focuses on predicting the faculty selection of high school students at SMAN XXX Tangerang Selatan using the C4.5 algorithm. By analyzing 203 student data, including their national exam scores, the research applies the decision tree model of the C4.5 algorithm, achieving a prediction accuracy of 77.16%. The results from the decision tree are used as a basis for determining the appropriate faculty choice for students, offering a decision support system for the school to aid in faculty selection after graduation.

And the study by Ronna Putri Fadhillah, Raisya Rahma, Arni Sepharni, Ratna Mufidah, Betha Nurina Sari, and Agung Pangestu (2022) [7] focuses on classifying diabetes mellitus based on its causative factors using the C4.5 algorithm. The research uses the Pima Indians Diabetes dataset from Kaggle, which includes 769 samples with nine attributes. After feature selection using a heatmap, the study identifies key features such as pregnancies, glucose, BMI, age, and outcome. The classification process achieved an accuracy rate of 76%, outperforming previous research using the SVM algorithm, which had a lower accuracy of 70%.

Another study by Annas Prasetio, Muhammad Hari Hasibuan, and Primatua Sitompul (2021) [8] focuses on applying the C4.5 decision tree algorithm in determining the nutritional status of toddlers. Using a dataset of 123 entries, the research applies the decision tree method to classify nutritional status based on weight and height. The results indicate that the C4.5 algorithm provides high accuracy in classifying toddler nutritional status, with the classification process producing a perfect accuracy rate of 100%. This demonstrates the effectiveness of the C4.5 algorithm in health-related classifications.

Therefore, the author provides a solution to classify boarding houses' feasibility and satisfaction levels using the C4.5 algorithm. This algorithm was chosen because it is one of the methods in the Decision Tree that is widely used for making predictions on a case [9]. Thus, boarding house owners can understand their tenants' feasibility and satisfaction levels based on the quality of service and facilities provided. This study uses a decision tree based on the C4.5 algorithm to improve the accuracy of boarding house feasibility and satisfaction assessments.

2. Literature Review

2.1. Boarding House

In the context of accommodation, a boarding house refers to a building or residence that provides rooms for rent and is occupied by tenants for a specified period. Boarding houses are often chosen by individuals who need temporary housing, such as students, workers, or migrants from other areas. The facilities offered by boarding houses typically include bedrooms, bathrooms (shared or private), a shared kitchen, and additional amenities such as internet access, cleaning services, and a shared living room [10].

In assessing boarding houses, aspects of service quality and facilities play a crucial role in determining the feasibility and satisfaction of the residents. Service quality includes the reliability of the services provided by the owner or manager, responsiveness to complaints, and the ability to maintain the comfort and security of the residents. Meanwhile, the available facilities include room furnishings, sanitation conditions, availability of clean water, and access to supporting facilities such as parking areas and environmental security. Boarding houses in various regions, including in Blang Pulo, are essential for affordable and comfortable housing. Residents' choice to stay in a boarding house is based on a combination of these various factors. Therefore, a study on the feasibility evaluation of boarding houses based on service quality and facilities using Data Mining methods, such as the C4.5 algorithm, can help boarding house managers improve service standards and make it easier for prospective tenants to make informed decisions [11].

2.2. Data Mining

Data mining is a technique used to extract information from data, enabling the acquisition of knowledge that can be used as a basis for decision-making. Companies worldwide utilize data mining techniques to support their business decisions, which can result in significant profits [12]. Descriptive data mining is used to uncover patterns that are understandable to humans and explain the characteristics of the data. On the other hand, predictive data mining is employed to create a knowledge model that can be used to make predictions. Based on its functionalities, data mining is divided into seven categories: clustering, which involves grouping data based on similarities; classification, which assigns data to predefined categories; forecasting, which predicts future trends or behaviors; regression, which models the relationship between variables; association, which identifies relationships between data items; sequencing, which analyzes sequences or patterns over time; and descriptive, which summarizes and interprets the characteristics of data. By combining disciplines such as statistics, artificial intelligence, and machine learning, data mining becomes a valuable tool for analyzing data and extracting useful information from large datasets [13].

2.3. Classification

Classification is the process of finding a set of models or functions that describe and distinguish the categories of data to predict the category of objects whose category is unknown. The model is obtained through data analysis with known class labels [14]. This classification technique is suitable for describing datasets with data types such as binary or nominal. However, the limitation of this technique is that it is not ideal for ordinal data sets due to the implicit order in the categories of data [15].

2.4. C4.5 Algorithm

The C4.5 algorithm is used for classification or clustering in a dataset [16]. Developed by Ross Quinlan, C4.5 is an enhancement of the ID3 algorithm and was designed to improve classification performance more efficiently and accurately. This algorithm is highly popular in data analysis and is used to predict the class or category of data based on existing attributes. It is also one of the most commonly used decision tree techniques and generates a decision tree that is easier to understand [17].

The steps for performing calculations with the C4.5 algorithm include preparing the training data, determining the tree's root by calculating entropy, then calculating the gain, and selecting the tuple to partition. To calculate entropy, the formula is used as follows [18]:

$$Entropy(s) = -\sum_{i=1}^{n} pi \times \log(pi) \quad (2)$$

Where:

s : Set of cases

n : Number of Samples

pi : Proportion of the class

After calculating the entropy, the next step is to calculate the gain using the following equation:

$$Gain(s, A) = Entropy(s) - \sum_{i=1}^{n} \frac{|si|}{|s|} \times \log(pi) \quad (3)$$

Where:

s : Set of Cases

A : Attribute

n : Number of partitions of attribute A

|si| : Proportion of Si relative to S

|s| : Set of Cases

After obtaining the gain and entropy values to choose the attribute as the root, select the attribute with the highest gain ratio from the available attributes. Then, calculate the gain ratio using the following equation:

$$gain\ ratio(s, A) = \frac{Gain(s,A)}{Splitinfo(s,A)} \quad (4)$$

Where:

s : Set of Cases

A : Attribute

$Gain(s, A)$: Gain Information of Attribute A

$Splitinfo(s, A)$: Split Information of Attribute A

$$Split\ info(s, A) = \sum_{i=1}^{n} \left(\frac{|si|}{|s|} log_2 \frac{|si|}{|s|}\right) \quad (5)$$

A : Attribute

n : Number of Partitions of Attribute

|s| : Data Number of Attribute

|si| : Data of Partition of Attribute I

2.5. Python

Python is a high-level programming language developed by Guido van Rossum and first released in 1991. Python features an easy-to-write syntax, and it also has a comprehensive library and strong community support because it is open source. This language is well-known for its easy-to-understand and straightforward syntax, making it very suitable for beginners [19].

Python is easier to understand due to its simple code writing; it is available for free and open source, flexible because it can run on almost all operating systems, and versatile because it can be implemented in web development, mobile apps, and desktop apps. Python is one of the high-level programming languages that is interpreted, interactive, and object-oriented. This language can operate on almost all platforms, including UNIX-based systems, Mac, Windows, and others [20].

3. Research Method

3.1. Place and Time of Research

This research was conducted in Blang Pulo Village to collect data for the boarding house's feasibility test and tenant satisfaction. This activity was carried out because the aspects that support the system's needs to be built must be considered so that the research can proceed well. Data was collected during the odd semester of 2024 until the study was completed. The data collected comes from the tenants who have stayed at the Blang Pulo Boarding House and provided assessments based on their experience regarding the services and facilities available.

3.2. Research Flow

This study will be conducted through several stages of the research method, as shown in the image below:

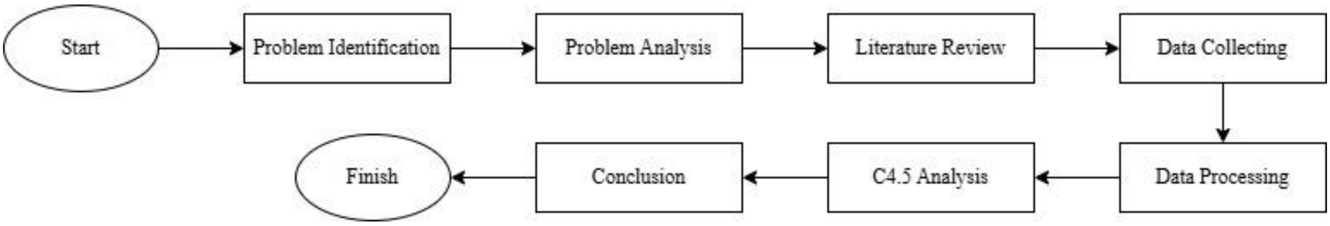


Fig 1. Research Flow Diagram

This diagram illustrates the systematic flow of the research process, beginning with identifying the problem. The next stage involves analyzing the situation, followed by a literature review to gather relevant background information. Afterward, data is collected to gather the information required for the research. The collected data is then processed in the Data Processing phase, refined, and prepared for analysis. Subsequently, the C4.5 analysis is performed on the processed data to draw meaningful conclusions. Finally, the research concludes with a summary of the findings in the Conclusion section, completing the entire research process.

3.3. System Schema

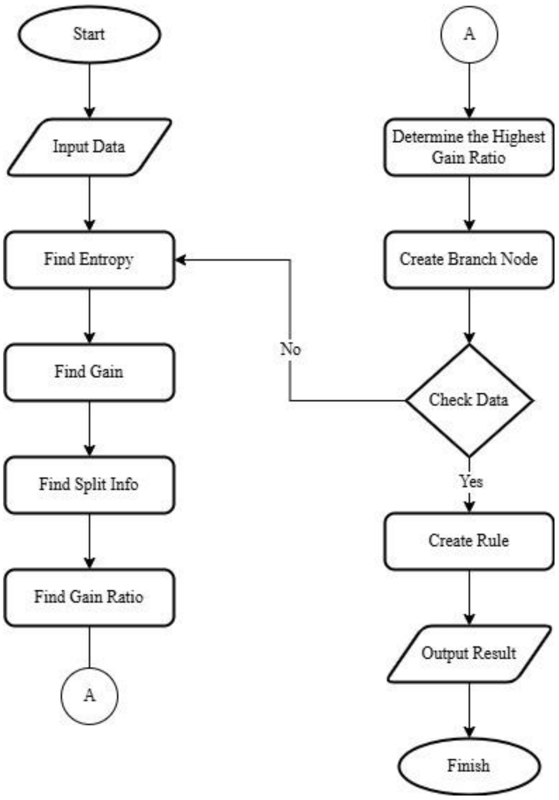


Fig 2. System Schema

3.3. Data Collecting

The data in this study is classified into three categories: low, medium, and luxury, based on various aspects such as service quality, cleanliness, security, facilities, and comfort. Boarding houses in the low category have limited facilities and services, and several factors, such as cleanliness, security, and access to public facilities, are not optimal. Meanwhile, the medium category shows improvements in service quality and maintenance, although there is still variation in the completeness of the provided facilities.

The luxury category offers better service quality, with well-maintained cleanliness, security, and comfort. Boarding houses in this category also provide more complete facilities, including stable internet access and better-equipped rooms. The classification system used in this study offers more structured information about the quality of boarding houses based on the assessed aspects, helping prospective tenants choose suitable accommodations and offering insights for boarding house owners to improve their services.

3.4. Problem Analysis

Determining the feasibility and satisfaction levels of boarding house residents is a complex challenge, considering the numerous variables that must be considered, such as available facilities, cleanliness, security, rental prices, and the quality of service provided by the owner or manager. An accurate understanding of these factors is crucial for prospective tenants when choosing a boarding house that suits their needs and preferences.

Traditional approaches to assess feasibility and resident satisfaction are often done through direct testimonials, simple surveys, or word-of-mouth recommendations. However, these methods have drawbacks as they are subjective, unstructured, and difficult to analyze on a larger scale. Moreover, prospective tenants often face challenges when comparing the available options without a clear classification system. At the same time, boarding house owners lack a solid basis for improving their services based on accurate data.

Applying the C4.5 algorithm in Data Mining offers a relevant solution to address these issues. The C4.5 algorithm can analyze historical data and classify boarding houses based on feasibility and resident satisfaction levels. This algorithm works by building a decision tree based on attributes that influence the quality of the boarding house, such as available facilities, environmental conditions, rental prices, and cleanliness levels. With this approach, this study aims to produce a classification model that can assist prospective tenants in selecting suitable boarding houses and provide insights for owners to improve their services.

With the presence of a data mining-based system, the decision-making process for choosing a boarding house is expected to be carried out more objectively, accurately, and efficiently. Furthermore, the classification model developed from this study can also serve as a foundation for creating smarter, data-driven boarding house recommendation systems in the future.

4. Result and Discussion

This study aims to apply the C4.5 algorithm to determine boarding houses' feasibility and satisfaction levels based on the quality of services and available facilities. The developed system utilizes data related to boarding house amenities and services, including cleanliness, security, availability of facilities (e.g., WiFi, air conditioning, en-suite bathrooms), and rental prices. This data was collected through questionnaires and direct surveys with tenants and boarding house owners. The primary objective is to design a classification model using the C4.5 algorithm to categorize boarding houses into specific groups: feasible or not feasible and satisfactory or unsatisfactory. By implementing the C4.5 algorithm, this research seeks to produce a decision tree that serves as a reference for assessing boarding houses' feasibility and tenant satisfaction. The results are expected to provide a solution for prospective tenants in selecting accommodations that match their preferences, assist owners in improving service quality, and contribute to developing a more structured and optimal data mining-based recommendation system.

4.1. Problem Analysis

Determining the feasibility and satisfaction levels of boarding house tenants is a complex challenge due to the many variables that must be considered, such as available facilities, cleanliness, security, rental price, and the quality of service the owner or manager provides. An accurate understanding of these factors is crucial for prospective tenants when selecting accommodations that align with their needs and preferences. Traditional approaches to assessing boarding house feasibility and tenant satisfaction often rely on direct testimonials, simple surveys, or word-of-mouth recommendations. However, these methods are subjective, unstructured, and difficult to analyze on a larger scale. Without a clear classification system, prospective tenants frequently struggle to compare available options, and owners lack a solid foundation for improving service quality based on accurate data. Applying the C4.5 algorithm in Data Mining offers a relevant solution to address this issue. C4.5 can analyze historical data and classify boarding houses based on feasibility and tenant satisfaction levels by constructing a decision tree from attributes influencing boarding house quality, such as available facilities, environmental conditions, rental price, and cleanliness. Through this approach, the study aims to develop a classification model that assists tenants in making informed choices and provides insights for owners to enhance their services. A Data Mining-based system is expected to make the decision-making process more objective, accurate, and efficient while laying the groundwork for developing a more intelligent, data-driven boarding house recommendation system.

4.2. Data Descriptive

The data in this study is classified into three categories—low, medium, and premium—based on various aspects such as service quality, cleanliness, security, facilities, and overall comfort. Limited facilities and services, with suboptimal conditions in areas such as cleanliness, safety, and access to public amenities, characterize boarding houses as being in the low category. Meanwhile, the medium category indicates improved service and maintenance quality, although variations in the completeness of facilities still exist. Premium boarding houses offer higher service quality, with better-maintained cleanliness, security, and comfort. This category typically includes more comprehensive facilities, such as stable internet access and well-equipped rooms. The classification system used in this study provides a more structured overview of boarding house quality based on the assessed aspects, enabling prospective tenants to make more informed housing choices and offering valuable insights for property owners to improve their service standards.

4.3. System Implementation

In this subsection, the researcher discusses the implementation of the C4.5 algorithm within a boarding house classification system based on feasibility and tenant satisfaction. This implementation follows the system design phase, where users can input boarding house data, including the property name, available facilities, cleanliness, security, management service quality, and overall comfort. The system processes the input data using the C4.5 algorithm to generate a decision tree that classifies boarding houses into specific categories, such as low, medium, or premium.

The classification results are presented in reports and data visualizations, making it easier for users to understand the feasibility and satisfaction levels based on predefined parameters. The system interface is designed to be user-friendly, enabling administrators or property owners to quickly and efficiently access classification outcomes. Additionally, the system includes features for reviewing training history and classification analysis, allowing users to evaluate the quality of the input data and the resulting classifications.

This system implementation is crucial in providing objective and accurate information for prospective tenants when selecting suitable accommodations. It supports property owners in improving their services based on analyzed data. Furthermore, it opens opportunities for future development, such as integration with web- or mobile-based applications, to enhance accessibility and efficiency in managing and monitoring boarding house quality.

4.3.1. Dashboard Page

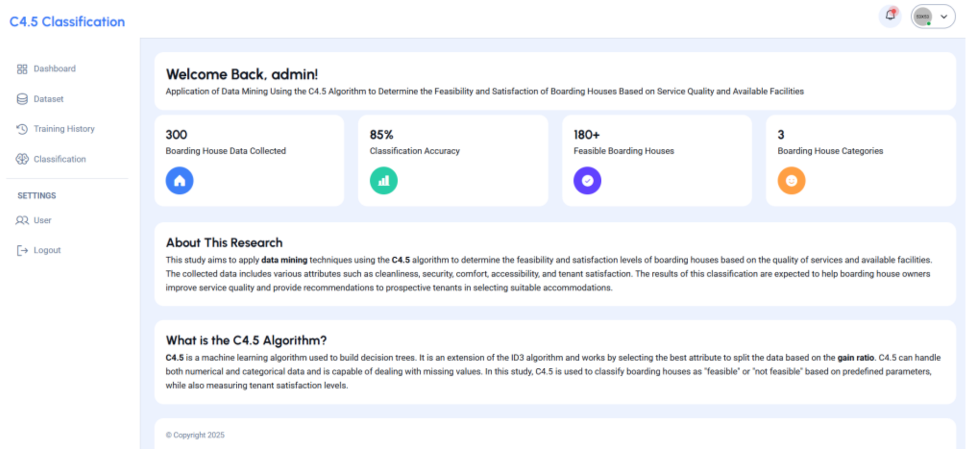


Fig 3. Dashboard Page Display

The Dashboard page in the boarding house classification system based on the C4.5 algorithm is designed to provide a general overview of the data analysis results and offer user-friendly navigation. At the top of the page, a welcome message—"Welcome Back, admin!"—is displayed along with a brief description of the system's objective: determining the feasibility and satisfaction levels of boarding houses based on service quality and available facilities.

Below this, four key statistic boxes present essential information, including the number of collected boarding house data (300), classification accuracy (85%), the number of boarding houses classified as feasible (180+), and the number of boarding house categories (3). This information helps administrators understand the system's performance and the effectiveness of its classification process.

Further down, two additional sections provide informative content: "About This Research," which explains the objective behind the application of the C4.5 algorithm in boarding house classification, and "What is the C4.5 Algorithm?", which gives an overview of how the algorithm builds decision trees to categorize boarding houses based on predefined parameters. With a clean and informative layout, this dashboard ensures that administrators can easily access key data and comprehend the classification process within the system.

4.3.2. Dataset Page

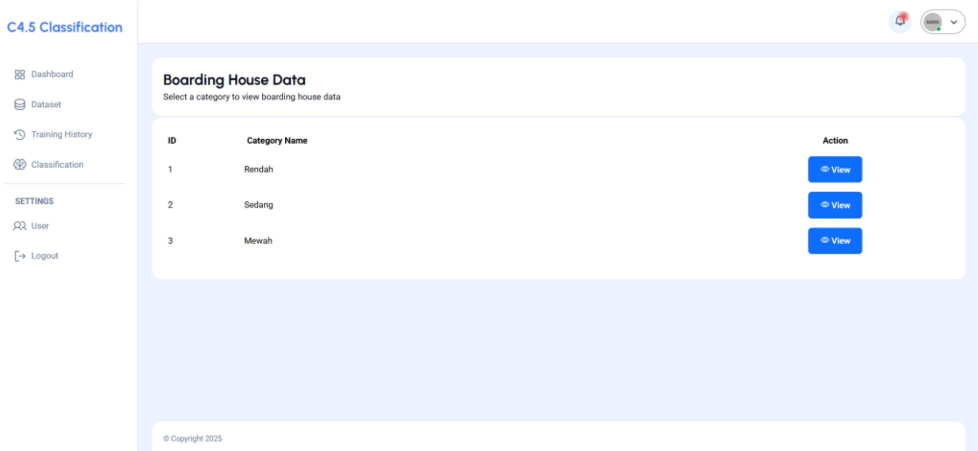


Fig 4. Dataset Page Display

4.3.3. C4.5 Implementation

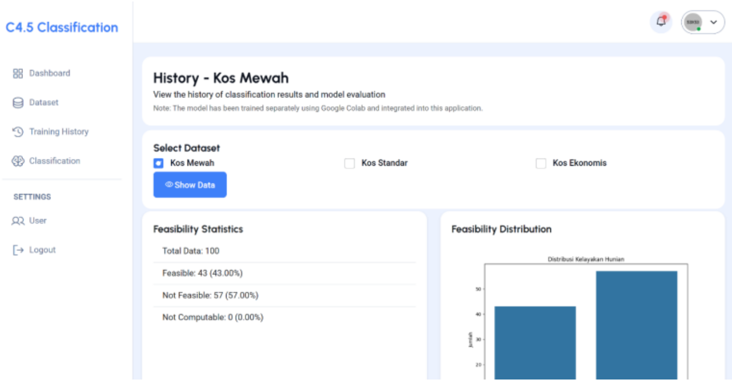


Fig 5. Training History Kos Mewah

The "History - Luxury Boarding Houses" page displays the classification results and model evaluation in the boarding house classification system based on the C4.5 algorithm. At the top is a description explaining that the model has been separately trained using Google Colab and integrated into this application, ensuring that the classification process is carried out with an optimal machine-learning approach.

Below that, there is a dataset selection feature where users can choose the category of boarding house they want to analyze, such as Luxury Boarding House, Standard Boarding House, or Economic Boarding House. After selecting a category, the "Show Data" button allows users to process and display the classification results. In the Feasibility Statistics section, the classification information is presented as percentages, where 43% of the 100 data points analyzed are categorized as Feasible, while 57% are classified as Not Feasible. The system also presents the Feasibility Distribution as a bar chart, providing a more precise visualization of the comparison between feasible and non-feasible boarding houses. With this display, the admin can evaluate the classification quality performed and further analyze the model's results for more accurate decision-making.

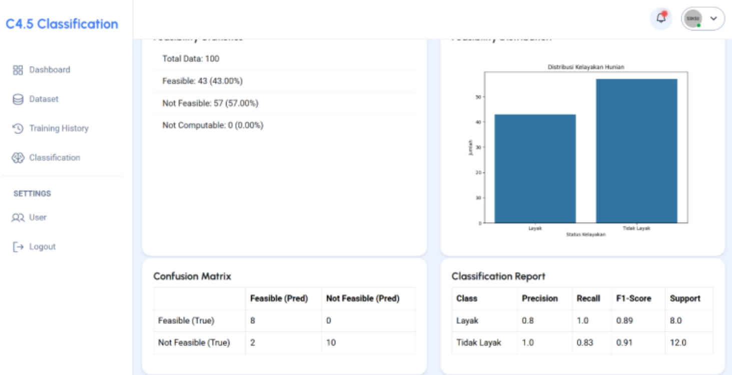


Fig 6. Result Analysis Classification Kos Mewah

The "History - Luxury Boarding Houses" page displays the model evaluation results with various metrics that provide an overview of the performance of the C4.5 algorithm in determining the feasibility of boarding houses. In the Confusion Matrix section, the table compares the predicted results and the actual values. The results displayed show eight boarding houses that are truly feasible (True Positive), 10 boarding houses that are genuinely not feasible (True Negative), and two boarding houses that are feasible but predicted as not feasible (False Negative). There are no False Positives, meaning the system did not incorrectly classify any boarding house as possible.

The Classification Report section provides evaluation metrics such as Precision, Recall, F1-Score, and Support. For the Feasible category, precision is at 0.8, recall is 1.0, and the F1-score is 0.89, indicating that the system is quite good at recognizing truly feasible boarding houses. Meanwhile, the Not Feasible category has a precision of 1.0, a recall of 0.83, and an F1-score of 0.91, which means the system accurately identifies non-feasible boarding houses. However, there are still some errors in classifying feasible boarding houses.

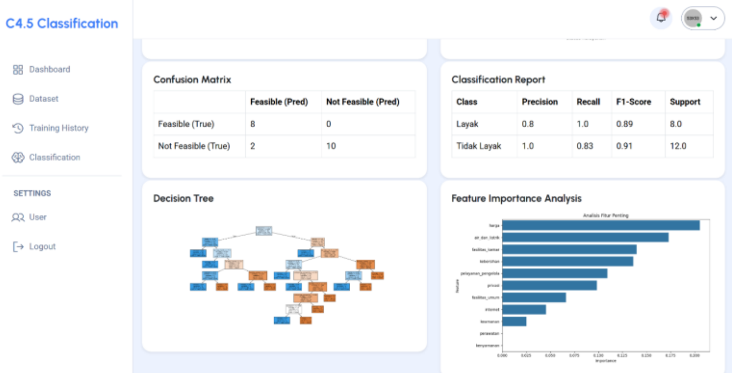


Fig 7. Result Analysis Classification Kos Mewah (Continue)

In the Decision Tree section, the system displays the learning results from the C4.5 algorithm in the form of a Decision Tree, illustrating how variables such as price, water and electricity, room facilities, and cleanliness play a role in determining the classification of a boarding house. This decision tree shows the relationships between factors and the final decision, making it easier to understand how the system determines whether a boarding house is classified as feasible or not feasible. With this visual representation, the admin can see the decision paths used by the system in analyzing the boarding house data.

Furthermore, further analysis can be performed through the Feature Importance Analysis chart, which displays the key factors influencing the classification of boarding houses. From these results, it is evident that price and access to water and electricity are dominant factors in determining the feasibility of a boarding house, followed by room facilities, cleanliness, and management services. With this visualization, the admin can better understand the most influential factors in the classification, which can be used as a basis for decision-making or improving the quality of boarding house services.

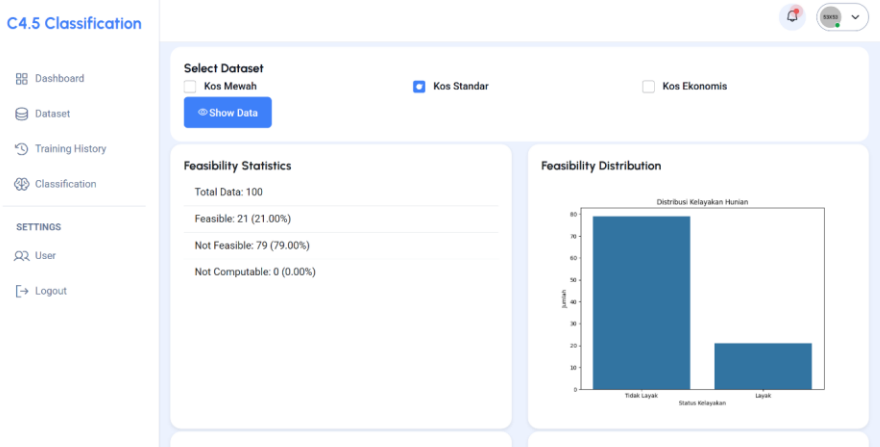


Fig 8. Result Analysis Classification Kos Standard

The "History - Standard Boarding Houses" page displays the classification results and model evaluation for the standard boarding house category using the C4.5 algorithm. Based on the Feasibility Statistics, of the total 100 boarding houses analyzed, 21% are classified as Feasible, while 79% fall into the Not Feasible category. This distribution is visualized in a bar chart, which shows that most boarding houses in the standard category do not meet the feasibility criteria based on the parameters used in the system.

In the Confusion Matrix section, the table illustrates how the model predicts the boarding house categories. The model correctly classified two feasible boarding houses (True Positive), but one boarding house was feasible but was classified as not feasible (False Negative). Meanwhile, 14 boarding houses that are indeed not feasible were correctly classified (True Negative), while three not feasible boarding houses were incorrectly predicted as feasible (False Positive).

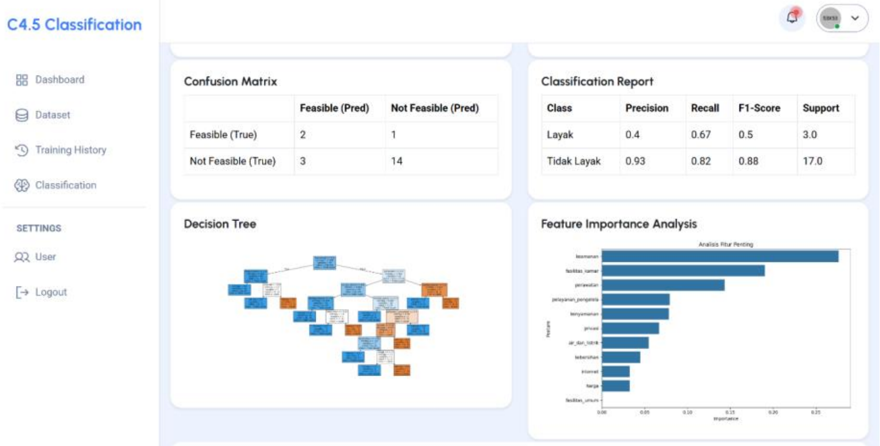


Fig 9. Result Analysis Classification Kos Standard (Continue)

The Classification Report section further evaluates each category's Precision, Recall, and F1-Score. For the feasible boarding houses, precision only reaches 0.4, indicating that there are quite a few errors in classifying boarding houses into this category. Meanwhile, recall is 0.67, showing that the model is somewhat capable of recognizing feasible boarding houses, although it still has limitations in overall accuracy. On the other hand, precision is 0.93, and recall is 0.82 for the not feasible category, indicating that the model is more accurate in identifying not feasible boarding houses than the feasible ones.

The Decision Tree section displays the system's learning results as a Decision Tree, showing how factors such as security, room facilities, maintenance, and management services are key variables in determining the feasibility of standard boarding houses. Additionally, the Feature Importance Analysis chart shows that security is the most dominant factor, followed by room facilities, maintenance, and management services. This visualization helps the admin understand the factors influencing the classification results and can serve as a basis for further evaluating the boarding houses in the standard category.

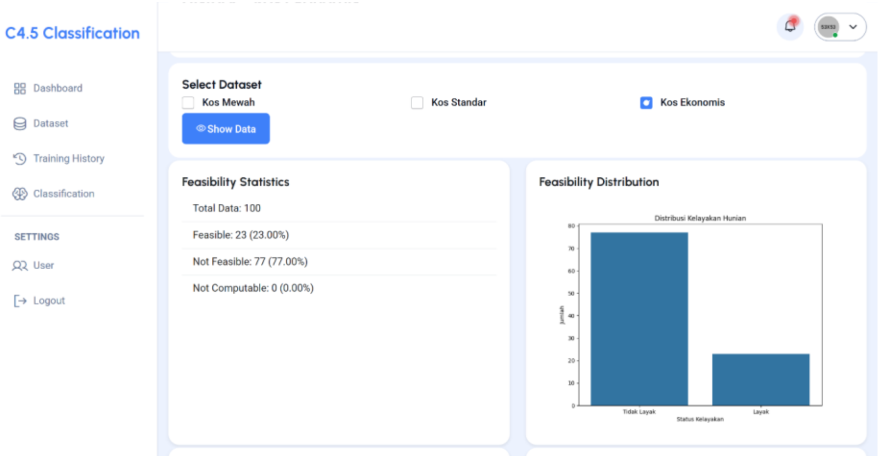


Fig 10. Result Analysis Classification Kos Ekonomis

The "History - Economic Boarding Houses" page displays the classification results and model evaluation for the economic boarding house category using the C4.5 algorithm. Based on the Feasibility Statistics, of the total 100 boarding houses analyzed, 23% are categorized as Feasible, while 77% are classified as Not Feasible. This distribution is visualized in a bar chart, which shows that most boarding houses in this category do not meet the feasibility criteria set based on the system's parameters.

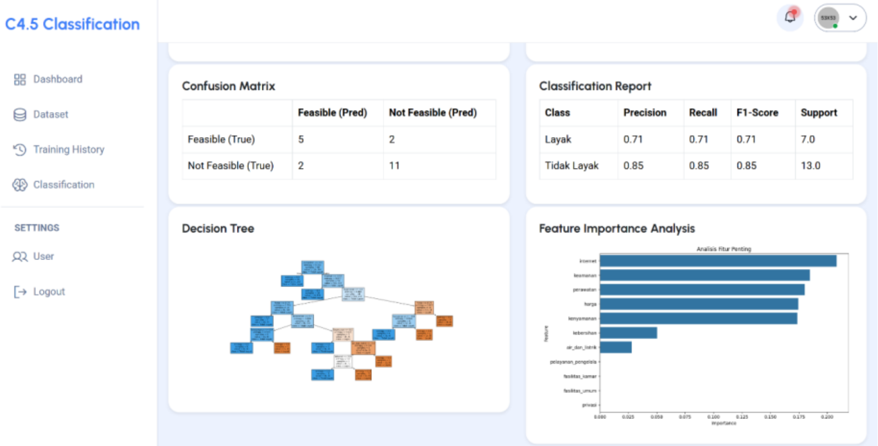


Fig 11. Result Analysis Classification Kos Ekonomis

In the Confusion Matrix section, the model correctly classified five feasible boarding houses (True Positive), but two boarding houses were feasible but classified as not feasible (False Negative). Meanwhile, 11 boarding houses that are genuinely not feasible were correctly classified (True Negative), while two not feasible boarding houses were incorrectly predicted as feasible (False Positive). The Classification Report shows Precision, Recall, and F1-Score for each category. For the feasible boarding houses, precision is at 0.71, recall is 0.71, and F1-score is 0.71, indicating that the model accurately recognizes feasible boarding houses but still has a relatively high error rate. For the not feasible category, precision is 0.85, and recall is 0.85, showing that the model is better at identifying not feasible boarding houses than feasible ones.

The Decision Tree section shows how the system uses the C4.5 algorithm to build the Decision Tree, illustrating the relationship between the main factors influencing the classification. The Feature Importance Analysis shows that internet access, security, maintenance, price, and comfort are dominant factors in determining the feasibility of boarding houses in the economic category. With this information, the admin can further understand how the model determines classifications and use it as a basis for analysis or decision-making to improve the quality of boarding houses.

Based on the classification results of the three categories of boarding houses (Luxury, Standard, and Economic) using the C4.5 algorithm, it is evident that the feasibility of boarding houses is heavily influenced by factors such as facilities, security, cleanliness, and other supporting aspects like price and service availability. From the analyzed data, the Luxury Boarding House category has the highest feasibility rate compared to the other two categories, with 43% classified as feasible. In contrast, the Standard Boarding House and Economic Boarding House categories show lower percentages of feasible boarding houses, with 21% and 23%, respectively, meaning that most of the accommodations in these categories still have limitations regarding services and available facilities.

Model evaluation based on the Confusion Matrix and Classification Report shows that the system is more accurate in identifying non-feasible boarding houses than feasible ones. This is reflected in the higher Precision and Recall values for the not feasible category, especially in the Standard and Economic Boarding House categories. Additionally, the Feature Importance Analysis reveals that security, room facilities, cleanliness, and price are the dominant factors in determining the feasibility of boarding houses. Based on these findings, boarding house owners can use these factors as references to improve service quality and meet feasibility standards better. In contrast, prospective tenants can use this classification as a reference when choosing a boarding house that fits their needs.

5. Conclusion

This study demonstrates the effectiveness of the C4.5 algorithm in developing a boarding house feasibility classification system that categorizes accommodations into Luxury, Standard, and Economic classes based on service and facility parameters. The research provides a comprehensive approach to evaluating boarding house quality through advanced data mining techniques, offering significant insights into the multifaceted nature of accommodation assessment.

The model's performance evaluation revealed nuanced classification results across different boarding house categories. The Luxury category exhibited the highest precision, with a 0.93 score for the unfeasible class, a recall of 0.83, and an F1 score of 0.91. Standard category accommodations showed more varied performance, with precision ranging from 0.71 for feasible classes to 0.85 for unfeasible classes, and F1-scores fluctuating between 0.71 and 0.85. The Economic category presented the most complex classification challenge, demonstrating a high precision of 0.93 for unfeasible classes but a lower precision of 0.40 for feasible courses, highlighting the complexity of evaluating lower-cost accommodations.

Critical factors influencing boarding house feasibility emerged as price, room facilities, and access to essential utilities such as water and electricity. The research underscores the variable impact of these factors across different accommodation categories, revealing that the significance of each parameter differs depending on the boarding house's classification. This nuanced understanding provides valuable guidance for boarding house owners and managers seeking to improve their service quality and facility offerings.

The study's findings have broader implications for the accommodation sector, offering a data-driven approach to assessing and improving boarding house quality. Leveraging the C4.5 algorithm allows stakeholders to gain deeper insights into the critical attributes determining accommodation feasibility and customer satisfaction. Future research could expand on this methodology by exploring additional factors, refining the classification model, and developing more comprehensive predictive frameworks for accommodation assessment.

In conclusion, this research contributes a sophisticated analytical approach to boarding house evaluation, demonstrating the potential of data mining techniques in providing actionable insights for accommodation providers. The study offers a robust classification methodology and highlights the importance of comprehensive, data-driven approaches in understanding and improving accommodation services.

References

[1] J. A. Terapan and S. K. Bisnis, “ANALYSIS OF INVESTMENT FEASIBILITY FOR BUSINESS,” vol. 2, pp. 171–185, 2024.

[2] M. S. Ummah, “Analisis Struktur Kovarians pada Indikator Kesehatan yang Berfokus pada Persepsi Kesehatan Subjektif pada Lansia yang Tinggal di Rumah” *Sustainability (Switzerland)*, vol. 11, no. 1, pp. 1–14, 2019.

[3] L. Susilowati, “Kualitas Pelayanan, Harga Dan Fasilitas Rumah Kos Di Desa Plosokandang Dalam Memberikan Kepuasan Kepada Penghuninya,” *An-Nisbah: Jurnal Ekonomi Syariah*, vol. 7, no. 2, pp. 561–595, 2020, doi: 10.21274/an.v7i02.3605.

[4] Reza Fauzy, Riki Winanjaya, and Susiani, “Analisis Tingkat Kepuasan Pelanggan dengan Menerapkan Algoritma C4.5,” *Bulletin of Computer Science Research*, vol. 2, no. 2, pp. 41–46, 2022, doi: 10.47065/bulletincsr.v2i2.162.

[5] N. L. Mauliddiyah, “Analisis Struktur Kovarians pada Indikator Kesehatan yang Berfokus pada Persepsi Kesehatan Subjektif pada Lansia yang Tinggal di Rumah” p. 6, 2021.

[6] C. Tristianto and D. A. Zein, “Prediksi Pemilihan Fakultas Di Universitas Pada Sekolah Menengah Atas Negeri XXX Tangerang Selatan Dengan Menggunakan Algoritma C4.5,” *Jurnal Penelitian dan Pengkajian Sains dan Teknologi*, vol. 32, 2022, doi: 10.37277/stch.v32i4.

[7] R. Putri Fadhillah, R. Rahma, A. Sepharni, R. Mufidah, B. Nurina Sari, and A. Pangestu, “KLASIFIKASI PENYAKIT DIABETES MELLITUS BERDASARKAN FAKTOR-FAKTOR PENYEBAB DIABETES MENGGUNAKAN ALGORITMA C4.5,” 2022. [Online]. Available: www.kaggle.com

[8] A. Prasetio, M. H. Hasibuan, and P. Sitompul, “Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 4, no. 3, 2021.

[9] P. V.M, I. Gunawan, B. E. Damanik, I. Parlin, and A. W. Saputra, “Penerapan Data Mining Menggunakan Algoritma C4.5,” *Sostech*, vol. 1, no. 5, pp. 2774–5147, 2021.

[10] R. Aminah, Replita, I. Dianto, and N. Hasibuan, “Dampak Keberadaan Rumah Kost Mahasiswa Terhadap Kehidupan Ekonomi Masyarakat,” *Jurnal at-taghyir*, vol. 6, pp. 239–252, 2024.

[11] A. Rahmatillah, T. Saputra, and W. B. Hartiningsih, “Sistem Informasi Rumah Kost Berbasis Web Studi Kasus pada Paviliun Sejahtera,” *Explore:Jurnal Sistem informasi dan telematika*, vol. 11, no. 2, p. 102, 2020, doi: 10.36448/jsit.v11i2.1495.

[12] Ar Razi, “Klasifikasi Penerima Beasiswa Aceh Carong (Aceh Pintar) Di Universitas Malikussaleh Menggunakan Algoritma Knn (K-Nearest Neighbors),” *Jurnal Tika*, vol. 7, no. 1, pp. 79–84, 2022.

[13] C. Zai, “IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA,” 2022.

[14] Asrianda, A. Razi, and Zulfadli, “Tingkat Kepuasan Peserta Pelatihan Pengabdian Dosen Universitas Malikussaleh Menggunakan Fuzzy Ahp,” *INFOTECH journal*, vol. 9, no. 1, pp. 20–24, 2023, doi: 10.31949/infotech.v9i1.4339.

[15] Fauziah, Dedy Hartama, and Irfan Sudahri Damanik, “Analisa Kepuasan Pelanggan Menggunakan Klasifikasi Data Mining,” *Jurnal Penerapan Kecerdasan Buatan*, 2020.

[16] D. A. C, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes,” vol. 2, no. 1, pp. 39–46.

[17] M. Adriansa, L. Yulianti, L. Elfianty, U. Dehasen Bengkulu, and J. Meranti Raya, “Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5,” 2022.

[18] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, “Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prosiding Seminar Nasional Matematika*, vol. 3, pp. 64–71, 2020, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[19] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning,” *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.

[20] M. Abdul muthalib, I. Irfan, K. Kartika, and S. M. Selamat Meliala, “Pengiraan Pose Model Manusia Pada Repetisi Kebugaran Ai Pemograman Python Berbasis Komputerisasi,” *INFOTECH journal*, vol. 9, no. 1, pp. 11–19, 2023, doi: 10.31949/infotech.v9i1.4233.