

SISTEM PENDUKUNG KEPUTUSAN PENENTUAN JEMAAH PRIORITAS PADA PT HAFIZ WISATA DARUSSALAM BERBASIS ALGORITMA C4.5 DAN PROFILE MATCHING

DECISION SUPPORT SYSTEM FOR DETERMINING PRIORITY PILGRIMS AT PT HAFIZ WISATA DARUSSALAM BASED ON THE C4.5 ALGORITHM AND PROFILE MATCHING

Herlina Dianaria¹, Ermatita^{*2}, Anggina Primanita³

***Email: 09012682226014@studen.unsri.ac.id**

^{1,2,3}Master Of Computer Science, Sriwijaya University, Palembang, Indonesia

Abstrak - Persaingan bisnis penyelenggara perjalanan ibadah umrah menuntut perusahaan untuk memanfaatkan data historis secara optimal dalam mendukung pengambilan keputusan. PT Hafiz Wisata Darussalam memiliki sekitar 500 data historis alumni jemaah periode 2022–2025, namun data tersebut belum dimanfaatkan secara maksimal sehingga penentuan jemaah prioritas masih dilakukan secara subjektif. Penelitian ini bertujuan membangun Sistem Pendukung Keputusan untuk menentukan jemaah prioritas menggunakan kombinasi algoritma C4.5 dan metode *Profile Matching*. Penelitian menggunakan 500 data historis dengan atribut mantan jemaah, pekerjaan, aktivitas media sosial, paket umrah, umur, dan keberangkatan bersama keluarga. Algoritma C4.5 digunakan untuk mengklasifikasikan jemaah prioritas berdasarkan atribut yang paling berpengaruh melalui pembentukan *decision tree*, sedangkan metode *Profile Matching* digunakan untuk menentukan peringkat jemaah berdasarkan tingkat kesesuaian terhadap profil ideal perusahaan. Sistem diimplementasikan menggunakan PHP dan MySQL. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan data dengan tingkat akurasi sebesar 82,6%, *precision* 81,5%, *recall* 81,2%, dan *F1-score* 81,4%. Dari 500 data historis, sistem mengidentifikasi 233 jemaah prioritas yang selanjutnya dirangking menggunakan metode *Profile Matching* sehingga menghasilkan rekomendasi jemaah prioritas secara objektif. Sistem yang dikembangkan dapat membantu perusahaan menentukan sasaran pemasaran yang lebih tepat, meningkatkan loyalitas alumni jemaah, serta mendukung proses pengambilan keputusan yang lebih efektif dan berbasis data.

Kata Kunci: Algoritma C4.5; Data Mining; Jemaah Prioritas; Profile Matching; Sistem Pendukung Keputusan; Travel Umroh

Abstract– The increasing competition among Umrah travel service providers requires companies to optimize the use of historical data to support decision-making. PT Hafiz Wisata Darussalam possesses approximately 500 historical records of Umrah pilgrims from the 2022–2025 period; however, these data have not been fully utilized, resulting in a subjective process for determining priority pilgrims. This study aims to develop a Decision Support System (DSS) for determining priority pilgrims by integrating the C4.5 algorithm and the Profile Matching method. The study employed 500 historical records consisting of the following attributes: previous pilgrim status, occupation, social media activity, Umrah package, age, and family group departure. The C4.5 algorithm was applied to classify priority pilgrims by identifying the most influential attributes through a decision tree, while the Profile Matching method was used to rank pilgrims based on their similarity to the company's ideal profile. The system was implemented using PHP and MySQL. The experimental results show that the C4.5 algorithm achieved an accuracy of 82.6%, a precision of 81.5%, a recall of 81.2%, and an F1-score of 81.4%. Out of 500 historical records, the system identified 233 priority pilgrims, who were subsequently ranked using the Profile Matching method to generate objective priority recommendations. The proposed system assists the company in identifying appropriate marketing targets, improving alumni pilgrim loyalty, and supporting more effective and data-driven decision-making.

Keywords: C4.5 Algorithm; Data Mining; Decision Support System; Priority Pilgrims; Profile Matching; Umrah Travel.

I. INTRODUCTION

The rapid growth of the Umrah travel industry in Indonesia has intensified competition among Umrah travel agencies. This situation requires travel companies not only to attract new pilgrims but also to maintain the loyalty of former pilgrims as part of their business development strategy [1]. PT Hafiz Wisata Darussalam manages approximately 500 historical records of pilgrims from the 2022–2025 period, including information on pilgrimage history, occupation, social media activity, Umrah package selection, age, and family participation. However, these historical data have not been optimally utilized, as the determination of priority pilgrims is still largely based on subjective considerations.

Decision Support Systems (DSS) provide a systematic approach to assisting decision-making processes by utilizing data and computational methods to produce more objective recommendations [2]. In this study, a combination of the C4.5 algorithm and the Profile Matching method is proposed. The C4.5 algorithm is employed to identify the most influential attributes for determining priority pilgrims by constructing a decision tree [2], [3]. Subsequently, the Profile Matching method is used to measure the suitability of each pilgrim against the company's ideal profile and to generate a ranked recommendation of priority pilgrims [4], [5].

Previous studies have demonstrated that the C4.5 algorithm achieves satisfactory classification performance in various application domains, including the identification of Hajj departure patterns [6] and disease diagnosis classification [7]. Meanwhile, the Profile Matching method has been successfully applied to generate ranking-based recommendations for selecting prospective Hajj pilgrims [8]. Furthermore, the integration of the C4.5 algorithm and the Profile Matching method has shown promising results in decision support systems by improving decision accuracy and effectiveness [9], [10]. Nevertheless, the application of this combined approach for determining priority pilgrims in the context of Umrah travel agencies remains limited.

Therefore, this study aims to develop a Decision Support System by integrating the C4.5 algorithm and the Profile Matching method to provide objective recommendations for priority pilgrims. The proposed system is expected to assist PT Hafiz Wisata Darussalam in supporting its marketing strategy, enhancing customer loyalty among former pilgrims, and improving the effectiveness of data-driven decision-making.

II. LITERATURE REVIEW

The literature review in this study, entitled "Sistem Pendukung Keputusan Penentuan Jemaah Prioritas pada PT Hafiz Wisata Darussalam Berbasis Algoritma C4.5 dan Profile Matching," examines previous studies related to decision support systems for priority determination, particularly in the context of Umrah services and data-driven decision-making. It also reviews recent developments in the application of the C4.5 algorithm and the Profile Matching method based on studies published within the last five years. This review aims to establish the theoretical foundation of the research, identify the strengths and limitations of previous studies, and highlight the research gap and novelty of the proposed approach.

A. Theoretical Foundation

1. Jemaah

According to the Kamus Besar Bahasa Indonesia (KBBI), Jemaah refers to a group of people who perform a religious worship, while Jemaah Umrah refers to a group of pilgrims who perform the Umrah pilgrimage in Makkah [1]. At PT Hafiz Wisata Darussalam, Jemaah are categorized into three groups based on the selected Umrah package, namely Jemaah Gold, Jemaah Regular, and Jemaah Hemat.

In this study, Jemaah Prioritas refers to all former pilgrims who have previously performed Umrah with PT Hafiz Wisata Darussalam through the Gold, Regular, or Hemat packages and have the potential to contribute to the company's business development. This potential is reflected in factors such as loyalty, preference for specific Umrah packages, and the ability to recommend or bring new prospective pilgrims. Pilgrims identified as Jemaah Prioritas receive marketing support, business development guidance, incentives for successfully referring new pilgrims, and additional benefits when participating in future Umrah trips organized by PT Hafiz Wisata Darussalam.

2. Decision Support System

A Decision Support System (DSS) is a computer-based system designed to assist decision-makers in solving semi-structured and unstructured problems by utilizing relevant data, models, and information, thereby enabling faster, more objective, and more accurate decision-making [11]. In this study, the DSS is employed to support the determination of priority pilgrims based on the analysis of historical data using the C4.5 algorithm and the Profile Matching method.

3. Definition of the C4.5 Algorithm

The C4.5 algorithm is one of the most widely used classification algorithms for constructing a decision tree based on the calculation of entropy and information gain values. It is employed to classify data into predefined classes by selecting the attribute that provides the highest discriminatory power among the available attributes [2], [3] [12]. In addition to handling both categorical and numerical data, the C4.5 algorithm generates decision rules that are easy to interpret, making it widely applicable to various classification problems and decision support systems [13], [14].

4. Profile Matching

The Profile Matching method is a decision-making approach within a Decision Support System (DSS) that is used to measure the level of compatibility between an alternative's profile and a predefined ideal profile. The compatibility level is determined by calculating the gap between the actual value and the standard value for each criterion. The resulting gap values are then converted into weighted scores to produce a final score, which serves as the basis for the ranking process [4], [5]. In this study, the Profile Matching method is applied to determine the priority ranking of pilgrims based on their level of conformity with the ideal profile established by PT Hafiz Wisata Darussalam [8].

III. RESEARCH METHODS

The research methodology consists of a series of systematic stages that serve as a guideline for conducting the study, starting from problem identification to the generation of priority pilgrim recommendations. This study integrates the C4.5 algorithm and the Profile Matching method in the decision-making process. The C4.5 algorithm is employed to analyze historical data and identify the most influential attributes in determining priority pilgrims through the construction of a decision tree [2]. Subsequently, the Profile Matching method is applied to evaluate the compatibility of each pilgrim with the ideal profile established by the company, resulting in a ranked list of recommendations based on the level of conformity [5]. The integration of these two methods is expected to provide a more objective and accurate approach to determining priority pilgrims [9].

The research methodology consists of the following stages: problem identification, data collection, criteria determination, data analysis using the C4.5 algorithm, priority determination using the Profile Matching method, system implementation, and testing and result analysis. These stages are systematically organized to ensure that each phase of the research is conducted in accordance with the research objectives. Figure 1 illustrates the research methodology for determining priority pilgrims in an Umrah travel agency.

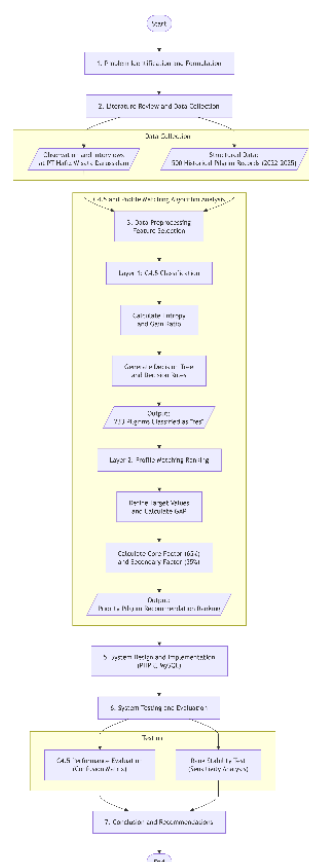


Figure 1 illustrates the research methodology

Figure 1 illustrates the proposed research methodology for developing a Decision Support System (DSS) to determine priority Umrah pilgrims. The research consists of six main stages: problem identification, literature review and data collection, data analysis, system implementation, system evaluation, and conclusion.

During the data collection stage, both primary and secondary data were utilized. The primary data were collected through observations and interviews with the management of PT Hafiz Wisata Darussalam to identify the evaluation criteria, define the ideal profile, and determine the weighting scheme used in

the Profile Matching method. Meanwhile, the secondary data were obtained through documentation of the company's customer database, consisting of 500 historical Umrah pilgrim records from the 2022–2025 period. The dataset includes six evaluation attributes: previous pilgrim status, occupation, social media activity, Umrah package, age, and family group departure, while the priority class (Yes/No) was used as the target variable for classification. The details of the dataset are presented in Table 1.

After the data collection stage, a data preprocessing step was performed to prepare the dataset before analysis using a two-layer filtering approach. In the first layer, the C4.5 algorithm was applied to classify pilgrims into the "Yes" and "No" categories by calculating entropy, information gain, and gain ratio, which were then used to generate decision rules in the form of a decision tree. Only pilgrims classified as "Yes" proceeded to the second layer, where the Profile Matching method was applied. At this stage, the GAP values, Core Factor (65%), Secondary Factor (35%), and final scores were calculated to produce a ranked list of priority pilgrim recommendations. Finally, the system was implemented using PHP and MySQL and evaluated using Confusion Matrix metrics and sensitivity analysis to assess the effectiveness and robustness of the proposed method.

A) Data Processing Using the C4.5 Algorithm

The process begins by inputting the pilgrim dataset into the system. The system then automatically calculates the overall entropy and the entropy of each attribute using Equation (1). Subsequently, the Information Gain, Split Information, and Gain Ratio values are computed for all available attributes. The attribute with the highest Gain Ratio is selected as the root node of the decision tree.

The decision tree is then constructed recursively by generating new branches based on the remaining attributes until all data in the leaf nodes are classified into homogeneous classes ("Yes" or "No"). Finally, each path from the root node to a leaf node is transformed into a set of decision rules, which are used as the basis for classifying priority pilgrims.

Figure 2 presents the flowchart of the data processing procedure using the C4.5 algorithm. This process is performed recursively to generate new branches until all records in the leaf nodes are classified into homogeneous classes ("Yes" or "No").

Each path from the root node to a leaf node is then extracted as a decision rule, which serves as the basis for the classification process.

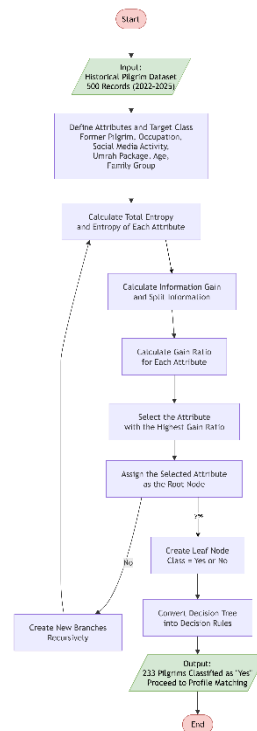


Figure 2 presents the flowchart of the data processing procedure using the C4.5 algorithm.

B) Profile Matching Calculation

After the classification process using the C4.5 algorithm is completed, the next stage is to determine the priority of the pilgrims using the Profile Matching method. This method is employed to measure the level of compatibility between each pilgrim's profile and the ideal profile established according to the requirements of PT Hafiz Wisata Darussalam.

In this study, the Profile Matching method is used to evaluate the suitability of each pilgrim based on the predefined attributes, resulting in a final score that serves as the basis for ranking the priority pilgrims. The smaller the GAP between a pilgrim's profile and the ideal profile, the higher the compatibility score and the greater the likelihood that the pilgrim will be prioritized for marketing purposes.

The logical workflow and the systematic calculation stages of the ranking process using the Profile Matching method are illustrated in the flowchart presented in Figure 3. The Profile Matching flowchart shown in Figure 3 illustrates the ranking process for determining priority pilgrims. The process

begins by receiving the input data of pilgrims who have been classified as "Yes" by the C4.5 algorithm. The system then compares each pilgrim's actual profile with the predefined ideal profile established by the company to calculate the GAP value. The GAP values are subsequently converted into weighted scores and grouped into Core Factors (CF) and Secondary Factors (SF).

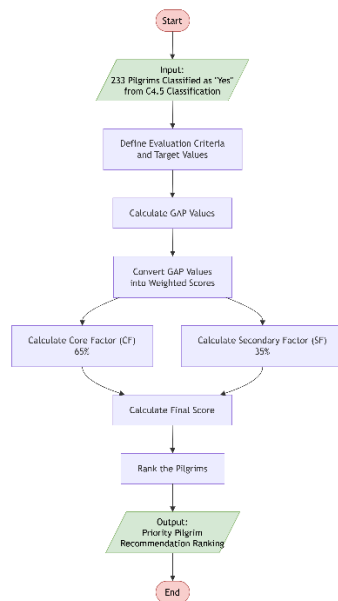


Figure 3. Flowchart of the Profile Matching Method

After calculating the average Core Factor (NCF) and Secondary Factor (NSF) scores, the system computes the final score (N) by applying the predetermined weighting scheme, consisting of 65% for the Core Factor and 35% for the Secondary Factor. Finally, the pilgrims are ranked by sorting the final scores in descending order, producing a priority list from the highest to the lowest score.

C) System Design

The system design phase aims to model the software architecture, data flow, and user interactions of the Decision Support System (DSS) for determining priority pilgrims. This stage provides a comprehensive overview of the system's operational workflow before implementation. The system is designed using the Unified Modeling Language (UML), which serves as a standard modeling language for representing the system's functional requirements and the relationships among its components in a structured manner. Through this design, the system integrates the classification process using the C4.5 algorithm with the weighting and ranking process using the Profile Matching

method, thereby enabling the generation of effective and well-structured recommendations for priority pilgrims [9], [11].

D) Use Case Diagram

The Use Case Diagram presented in Figure 4 illustrates the functional requirements of the proposed system and defines the interactions between the user (actor) and the Decision Support System for Determining Priority Pilgrims. The system is designed with a single primary actor, namely the Administrator/Management, who has full authority to manage all system functions. The proposed system consists of five main use cases, each representing a key functionality required to support the decision-making process. The proposed Use Case Diagram of the system is presented in Figure 4.

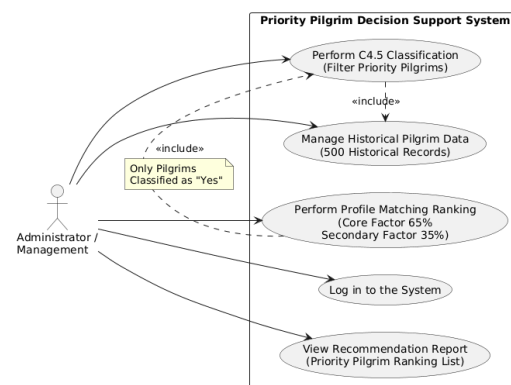


Figure 4 Use Case Diagram

E) C4.5 Algorithm Accuracy Evaluation

The accuracy evaluation was conducted to assess the ability of the C4.5 algorithm to classify pilgrim data into two recommendation classes, namely "Yes" and "No." The purpose of this evaluation is to determine the classification performance of the proposed model by comparing the predicted results with the actual data [14] [15]

In this study, the performance of the C4.5 algorithm was evaluated using a Confusion Matrix, which is a widely used evaluation method for measuring classification performance based on the number of correct and incorrect predictions for each class [14], [3]. The Confusion Matrix produces four performance metrics: Accuracy, Precision, Recall, and F1-Score. Accuracy measures the overall correctness of the classification model, Precision indicates the proportion of correctly predicted

positive instances, Recall evaluates the model's ability to identify all actual positive instances, while the F1-Score represents the harmonic mean of Precision and Recall [3], [7]. Higher values of these four metrics indicate better classification performance of the C4.5 algorithm in identifying priority pilgrims.

F) Ranking Stability Analysis

After obtaining the ranking results using the Profile Matching method, a ranking stability analysis was conducted to evaluate the consistency of the ranking outcomes when different weighting schemes were applied to the Core Factor (CF) and Secondary Factor (SF).

In this study, the primary weighting scheme consisted of 65% for the Core Factor and 35% for the Secondary Factor. Two additional weighting scenarios, 50%:50% and 70%:30%, were then applied as comparison scenarios. For each scenario, the final scores were recalculated to generate a new ranking of the priority pilgrims.

The rankings obtained from each comparison scenario were subsequently compared with the baseline ranking generated using the 65%:35% weighting scheme. The stability level was determined based on the number of pilgrims who either retained the same ranking position or experienced a change of no more than one rank relative to the baseline ranking.

$$\text{Ranking Stability} = \frac{\text{Number of stable records}}{\text{Total number of records}} \times 100\% \quad (1)$$

Where:

- Number of stable records = the number of pilgrims whose ranking remains unchanged or shifts by no more than one position compared with the baseline ranking.
- Total number of records = the total number of pilgrims processed using the Profile Matching method.

The results of this analysis are used to evaluate the consistency and robustness of the Profile Matching method when different weighting schemes are applied to the Core Factor (CF) and Secondary Factor (SF).

IV. RESULTS AND DISCUSSION

A) Data Processing Using the C4.5 Algorithm

After the preprocessing stage was completed, the next step was to calculate the entropy and information gain values using the C4.5 algorithm. These

calculations were performed to measure the uncertainty of the dataset and to identify the attribute with the highest capability to distinguish between the Priority and Non-Priority classes. The attribute with the highest information gain was selected as the root node for constructing the decision tree.

Table 1. results of the entropy, information gain, split information

Node	Attribute	Attribute Value	Number of Cases	Yes	No	Entropy	Gain	Split Info	Gain Ratio
1.	TOTAL		500	234	266	0.997043			
	FORMER PILGRIM						0.006336	1.884225	0.003999
		More than 2 times	166	68	98	0.97631			
		Once	172	90	82	0.988435			
		More than 1 time	162	76	86	0.99725			
2.	OCCUPATION						0.004274	0.998956	0.004274
		Employed / Worker	263	128	135	0.998999			
		Unemployed / Not	247	106	141	0.985467			
3.	ACTIVE ON SOCIAL MEDIA						0.001054	1.998461	0.000498
		Active	130	61	69	0.997267			
		Does Not Have Set	76	30	46	0.981233			
		Very Active	130	62	68	0.988463			
		Inactive	124	61	63	0.998932			
4.	WHICH UMRAH PACKAGE						0.002905	1.984889	0.00796
		Regular	171	73	98	0.988026			
		Economy (Premium)	170	86	84	0.9999			
		Gold	159	75	84	0.997488			
5.	AGE						0.001704	1.883697	0.001076
		20-30	172	84	88	0.99961			
		>30	173	83	90	0.998876			
		<20	155	67	88	0.986718			
6.	GOTO FOR UMRAH WITH FAMILY						0.001386	0.999938	0.001386
		Not (Branded)	251	112	139	0.997637			
		Yes	249	122	127	0.999705			

Table 1 presents the results of the entropy, information gain, split information, and gain ratio calculations obtained using the C4.5 algorithm on 500 historical records of pilgrims from PT Hafiz Wisata Darussalam. The dataset consists of 234 Priority records and 266 Non-Priority records, with an overall entropy value of 0.997043. An entropy value close to 1 indicates a high level of uncertainty in the dataset, making attribute selection essential for constructing an effective decision tree.

The calculation results show that the Previous Pilgrim/Previously Performed Umrah with Alhafiz attribute has the highest Information Gain value of 0.006336, followed by Occupation (0.004274), Selected Umrah Package (0.003105), Age (0.001704), Social Media Activity (0.001394), and Performing Umrah with Family (0.001386). The Information Gain value reflects the ability of an attribute to reduce the uncertainty of the dataset. Therefore, a higher Information Gain indicates that the corresponding attribute is more effective in distinguishing between the Priority and Non-Priority classes.

In addition to having the highest Information Gain, the Previous Pilgrim/Previously Performed Umrah with Alhafiz attribute also achieves the highest Gain Ratio value of 0.003999, making it the root node of the decision tree. This result indicates that a pilgrim's previous Umrah experience with PT Hafiz Wisata Darussalam is the most influential factor in determining priority status compared with the other

attributes. This finding suggests that pilgrims who have previously used the company's services are more likely to be classified as priority pilgrims than those without prior travel experience with the company.

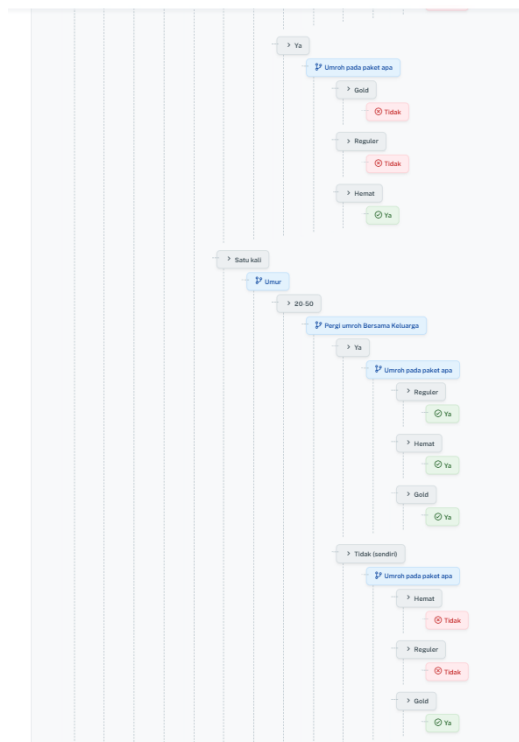


Figure 5. Decision Tree

Based on these findings, the C4.5 algorithm successfully identifies the most influential attribute for the classification process, providing the basis for constructing the decision tree. The classification results are subsequently used as input for the Profile Matching method to perform the ranking process and generate more objective recommendations for priority pilgrims.

B) Decision Tree Construction

After the Occupation attribute was selected as the root node based on the highest Gain Ratio value, the next step was to construct the decision tree. The decision tree was built by partitioning the dataset according to the values of the Occupation attribute, namely Employed and Unemployed, following the decision tree construction mechanism of the C4.5 algorithm [2], [3].

For each resulting branch, the entropy, information gain, split information, and gain ratio values were recalculated for the remaining attributes. These calculations were performed to identify the

most appropriate attribute as the next decision node based on the highest Gain Ratio value [2], [16].

The decision tree construction process was performed recursively until each branch resulted in a homogeneous class, namely "Yes" (Priority) or "No" (Non-Priority), or until no further attributes were available to partition the data [3][17]. The final output of this process was a decision tree representing the classification rules for determining priority pilgrims based on the selected attributes. The resulting decision tree is presented in Figure 5.

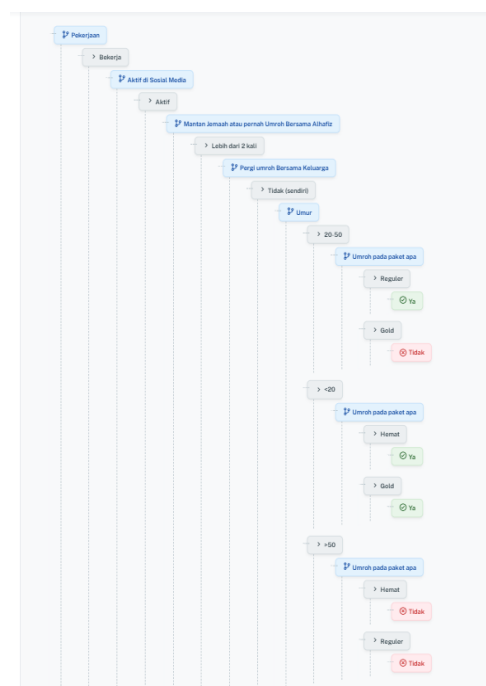


Figure 6. Decision Tree

Based on Figures 5 and 6, the Occupation attribute was selected as the root node because it achieved the highest Gain Ratio value. The dataset was subsequently divided into two branches, namely Employed and Unemployed. For each branch, the system selected the next attribute according to the highest Gain Ratio value, resulting in the construction of multiple decision nodes until reaching the leaf nodes, which represent the Priority or Non-Priority classes.

The resulting decision tree illustrates the hierarchical order of the most influential attributes in the classification process. In addition to serving as a classification model, the decision tree enhances interpretability, as each path from the root node to a leaf node can be transformed into a set of decision rules that clearly explain the classification outcomes.

C) Decision Rules

After the decision tree was constructed, each path from the root node to a leaf node was transformed into a decision rule. These rules were used by the system to classify pilgrim data into two categories: Priority ("Yes") and Non-Priority ("No").

Based on the resulting decision tree, a set of decision rules was generated to serve as the foundation of the classification process. Each rule represents a specific combination of attribute values that leads to a final classification outcome regarding the priority status of a pilgrim. These decision rules were subsequently implemented in the proposed system to automatically classify pilgrim data before proceeding to the ranking stage.

1. **IF** Pekerjaan = Bekerja **AND** Aktif di Media Sosial = Aktif **AND** Mantan Jemaah/Pernah Umrah Bersama Alhafiz = Lebih dari 2 Kali **AND** Pergi Umrah Bersama Keluarga = Tidak (Sendiri) **AND** Umur = 20–50 **AND** Paket Umrah = Reguler **THEN** Rekomendasi = **Ya**.
2. **IF** Pekerjaan = Bekerja **AND** Aktif di Media Sosial = Aktif **AND** Mantan Jemaah/Pernah Umrah Bersama Alhafiz = Lebih dari 2 Kali **AND** Pergi Umrah Bersama Keluarga = Tidak (Sendiri) **AND** Umur = 20–50 **AND** Paket Umrah = Gold **THEN** Rekomendasi = **Tidak**.
3. **IF** Pekerjaan = Bekerja **AND** Aktif di Media Sosial = Aktif **AND** Mantan Jemaah/Pernah Umrah Bersama Alhafiz = Lebih dari 2 Kali **AND** Pergi Umrah Bersama Keluarga = Tidak (Sendiri) **AND** Umur < 20 **AND** Paket Umrah = Hemat **THEN** Rekomendasi = **Ya**.

It can be seen that each rule is formed based on a combination of several attributes, namely Occupation, Social Media Activity, Previous Pilgrim or Has Performed Umrah with Alhafiz, Traveling for Umrah with Family, Age, and Umrah Package. The combination of these attributes produces the final decision in the form of Rekomendasi = Ya or Rekomendasi = Tidak.

For example, in Rule 1, pilgrims who have a certain occupation, are active on social media, have performed Umrah with Alhafiz more than twice, travel alone, are between 20 and 50 years old, and choose the Regular Package are classified as Rekomendasi = Ya. In contrast, in Rule 2, under almost identical conditions but with the Gold Package selected, the system produces Rekomendasi = Tidak. This indicates that a change in the value of a single

attribute can influence the classification result. Overall, the generated decision rules serve as the basis for the application to automatically classify pilgrim data before proceeding to the Profile Matching stage.

D) Algorithm Testing Results

The C4.5 algorithm was evaluated to assess its ability to classify pilgrim data into the Prioritas (Ya) and Tidak Prioritas (Tidak) categories. The evaluation was performed using a Confusion Matrix, which produced Accuracy, Precision, Recall, and F1-Score values. This evaluation aims to measure the model's classification performance before the classification results are used as input for the Profile Matching method.

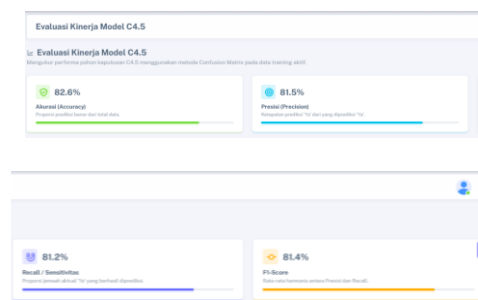


Figure 7 Results of the C4.5 Algorithm Testing

Based on the testing results of 500 pilgrim records (In Figure 7), the model achieved an Accuracy of 82.6%, Precision of 81.5%, Recall of 81.2%, and F1-Score of 81.4%. These results indicate that the C4.5 model has good performance in classifying pilgrim data based on the attributes used in this study.



Figure 8 Confusion Matrix Results

The Confusion Matrix in Figure 8 results show that the model correctly classified 190 records as True Positive (TP), representing pilgrims who actually belong to the Prioritas (Ya) category and were correctly predicted by the model. In addition, 223

records were classified as True Negative (TN), representing pilgrims who actually belong to the Tidak Prioritas (Tidak) category and were also correctly predicted. On the other hand, there were 43 False Positive (FP) records, representing pilgrims who actually belong to the Tidak Prioritas (Tidak) category but were predicted as Prioritas (Ya), and 44 False Negative (FN) records, representing pilgrims who actually belong to the Prioritas (Ya) category but were predicted as Tidak Prioritas (Tidak). Overall, the model correctly classified 413 out of 500 pilgrim records, while 87 records were misclassified. With an Accuracy of 82.6%, the C4.5 model demonstrates good performance and is suitable for use as the initial stage in the priority pilgrim selection process.

E) Pilgrim Data Classification Results

After testing the C4.5 algorithm, the classification results divided the pilgrim data into two categories, namely Prioritas (Ya) and Tidak Prioritas (Tidak). Based on the classification results, out of 500 pilgrim records, 233 were predicted as Prioritas (Ya), while 267 were predicted as Tidak Prioritas (Tidak).

In this study, the 233 records predicted as Prioritas (Ya) were used as input data for the Profile Matching method. This selection was made because the C4.5 algorithm serves as the initial filtering stage by identifying data based on classification patterns derived from historical data, whereas the Profile Matching method is used to calculate the degree of conformity of each alternative to the ideal profile, thereby producing a priority ranking [2], [5], [9].

Therefore, the Profile Matching method was not applied to all pilgrim records but only to those classified as Prioritas (Ya). This approach makes the ranking process more effective and efficient because only candidates considered eligible based on the C4.5 classification results are processed further in the Profile Matching stage. In addition to improving computational efficiency, this approach produces more focused priority recommendations since the assessment is performed only on pilgrims who have passed the initial filtering stage [9].

F) Data Processing Results Using the Profile Matching Method

After the classification process using the C4.5 algorithm was completed, the next stage was the implementation of the Profile Matching method [18]. This method was used to evaluate and determine the

priority ranking of pilgrims who had been classified as Prioritas (Ya) based on the C4.5 classification results. Therefore, Profile Matching was not used to reclassify the data but rather to rank the candidates who had passed the initial filtering stage [19].

Based on the previous classification results, 233 pilgrim records were predicted as Prioritas (Ya). All of these records were used as input for the Profile Matching process. Through this method, each pilgrim was evaluated based on the degree of conformity to the ideal profile using the predefined criteria, resulting in a final score that served as the basis for generating the priority recommendation ranking.

ID	Name	Core Factor	Secondary Factor	Final Score
P275	Hidayat	3.142	1.575	4.717
P366	Hidayat	3.142	1.575	4.717
P309	Hidayat	3.142	1.575	4.717
P60	Hidayat	3.142	1.575	4.717
P62	Hidayat	3.142	1.575	4.717
P140	Hidayat	3.142	1.575	4.717
P209	Hidayat	3.142	1.575	4.717

Table 2 Results of NCF, NSF, and Final Score Calculations

Based on Table 2, each pilgrim obtained different NCF and NSF values. These differences are influenced by the degree of conformity of each pilgrim to the ideal profile for each criterion. The higher the level of conformity to the ideal profile, the higher the NCF and NSF values obtained.

The calculation results show that P275 achieved an NCF value of 3.142 and an NSF value of 1.575, resulting in a final score of 4.717, which is the highest among the records presented. This result indicates that the characteristics of P275 have a high level of conformity with the ideal profile in both the Core Factor and Secondary Factor criteria.

Furthermore, P366 Hidayat obtained a final score of 4.659, while P309 obtained 4.600. Although both records have the same NCF value as P275 (3.142), the difference in the final score is caused by a lower NSF value. This finding indicates that, in addition to the Core Factor, the Secondary Factor also contributes to the final score, although its influence is smaller than that of the Core Factor.

In contrast, several pilgrims, namely P60, P62, P140, and P209, obtained the same final score of 4.550. This indicates that their levels of conformity with the ideal profile are relatively similar for both the Core Factor and Secondary Factor. Consequently, the system assigns the same final score, indicating that these pilgrims have an equivalent priority level based on the Profile Matching method.

Overall, the calculation results indicate that the Core Factor has a more dominant influence on the final score because it carries a weight of 65%, while the Secondary Factor serves as a supporting factor with a weight of 35%. Therefore, pilgrims who achieve higher scores on the main criteria tend to obtain higher final scores, even when there are slight differences in the supporting factors.

G) Ranking Results

Based on the final score calculation, the system ranked all 233 pilgrims who passed the classification stage using the C4.5 algorithm. The ranking process was performed by sorting the final scores in descending order, resulting in a priority ranking based on each pilgrim's degree of conformity to the ideal profile. This stage represents the primary function of the Profile Matching method, which identifies the best alternatives according to their level of conformity to the predefined profile [5].

The results indicate that P275 achieved the highest rank with a final score of 4.717, followed by P366 with 4.659 and P309 with 4.600. These pilgrims exhibited the highest degree of conformity to the ideal profile and therefore received higher priority recommendations than the other candidates. This finding demonstrates that the combination of Core Factor and Secondary Factor values effectively differentiates the priority level of each pilgrim based on the established criteria.

Overall, the results demonstrate that the Profile Matching method not only evaluates each pilgrim but also produces an objective priority ranking based on the weighted combination of Core Factor and Secondary Factor [5]. Furthermore, integrating the C4.5 algorithm with the Profile Matching method enables the selection process to be conducted in two sequential stages. The C4.5 algorithm performs the initial filtering by identifying pilgrims who should be prioritized, whereas the Profile Matching method determines their priority ranking. This approach is consistent with the findings of Gustiana and Sari, who reported that combining the two methods improves the effectiveness of decision-making through sequential classification and ranking processes [9]. Consequently, the resulting recommendations are more focused and can serve as a reliable basis for decision-making at PT Hafiz Wisata Darussalam.

Unlike the study by Sianturi *et al.* [6], which applied only the C4.5 algorithm to identify pilgrimage departure patterns, this study integrates the C4.5 algorithm as the initial filtering stage with the Profile Matching method as the ranking stage. This

integration enables the system not only to classify pilgrims who should be prioritized but also to generate recommendation rankings based on their degree of conformity to the company's ideal profile. Consequently, the proposed approach provides more comprehensive decision support than the use of the C4.5 algorithm alone.

H) Ranking Stability Analysis Results

After the Profile Matching process generated the final scores and priority rankings of the pilgrims, the next stage was to evaluate the stability of the ranking results. This evaluation was conducted to determine whether the recommendation results remained consistent when changes were made to the weights of the Core Factor (CF) and Secondary Factor (SF).

In this study, the default weighting scheme assigned 65% to the Core Factor and 35% to the Secondary Factor. Two additional weighting scenarios, 50%:50% and 70%:30%, were then evaluated for comparison. Under each scenario, the final scores were recalculated and a new ranking was generated. The rankings obtained from each scenario were subsequently compared with those produced using the default weighting scheme to assess the stability of the recommendation results. The results of the ranking stability evaluation are presented in Figure 4.5.

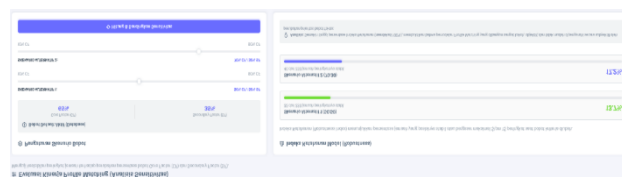


Figure 9 Ranking Stability Analysis Results

Based on Figure 9, the ranking stability under the 50%:50% weighting scenario was 13.7%, indicating that 32 out of 233 pilgrims maintained the same ranking or experienced a change of no more than one position compared with the ranking obtained using the default weighting scheme of 65%:35%.

Under the 70%:30% weighting scenario, the ranking stability increased to 17.2%, where 40 out of 233 pilgrims maintained their ranking positions or experienced a change of no more than one position. Among the two weighting scenarios, the 70%:30% weighting scheme produced a higher level of ranking stability than the 50%:50% weighting scheme. However, the overall stability remained relatively low, indicating that changes in the Core Factor and

Secondary Factor weights affect the ranking results. Therefore, determining the appropriate weighting scheme in the Profile Matching method is an important factor in producing priority recommendations that meet the requirements of PT Hafiz Wisata Darussalam.

V. CONCLUSION

Based on the research results, data processing, and system evaluation conducted on the Decision Support System for Determining Priority Pilgrims at PT Hafiz Wisata Darussalam, the following conclusions address the research objectives and research questions.

1. Implementation of the C4.5 Algorithm for Priority Pilgrim Classification
 - a. The C4.5 algorithm was successfully implemented to classify 500 historical pilgrim records (2022–2025) into "Ya" (Prioritas) and "Tidak" (Tidak Prioritas) categories.
 - b. Based on the Gain Ratio calculation, Pekerjaan was selected as the root node with the highest Gain Ratio value of 0.004274, followed by Mantan Jemaah/Pernah Umrah with a Gain Ratio value of 0.003999.
 - c. The C4.5 classification model generated a decision tree and decision rules (IF–THEN rules) that automatically filter the data. Based on the Confusion Matrix evaluation, the model achieved an Accuracy of 82.6%, Precision of 81.5%, Recall of 81.2%, and an F1-Score of 81.4%, indicating good classification performance.
2. Implementation of the Profile Matching Method for Priority Ranking
 - a. The Profile Matching method was successfully implemented as the ranking stage and applied exclusively to the 233 pilgrim records that had passed the filtering process and were classified as "Ya" (Prioritas) by the C4.5 algorithm.
 - b. By defining the ideal profile (the company's target standard), calculating GAP values, applying the weighting scheme, and grouping the evaluation criteria, the pilgrims were assessed objectively. The Core Factor (Mantan Jemaah, Pekerjaan, and Aktif Medsos) was assigned a weight of 65%, while the Secondary Factor (Umur, Pergi Bersama Keluarga, and Paket Umrah) was assigned a weight of 35%. As a result, the

system successfully generated a priority ranking, with Joko Pinem achieving the highest final score of 4.717.

- c. The ranking stability analysis showed that changes in the weighting scheme affected the ranking results. The 70%:30% weighting scenario achieved a higher stability level (17.2%) than the 50%:50% weighting scenario (13.7%).
3. Integration of the C4.5 Algorithm and the Profile Matching Method (Recommendation System)
 - a. The integration of the C4.5 algorithm and the Profile Matching method was successfully implemented as an adaptive and structured Decision Support System for PT Hafiz Wisata Darussalam.
 - b. The two methods operate sequentially in a two-stage process. The C4.5 algorithm performs the initial filtering by identifying potential pilgrims from the entire dataset, while the Profile Matching method performs the ranking of pilgrims who have already been identified as potential candidates.
 - c. The proposed integration resulted in a web-based application developed using PHP and MySQL, capable of providing objective priority recommendations for pilgrims. The system reduces subjective judgment in the decision-making process and supports PT Hafiz Wisata Darussalam in developing more effective marketing strategies and improving pilgrim loyalty.

REFERENCES

- [1] P. Deswara, "Isthita'ah Kesehatan Jemaah Haji," *Jurnal Persada Husada Indonesia*, vol. 10, no. 37, pp. 28–36, 2023, doi: 10.56014/jphi.v10i37.368.
- [2] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Ultimatics : Jurnal Teknik Informatika*, vol. 6, no. 1, pp. 15–20, 2014, doi: 10.31937/ti.v6i1.327.
- [3] D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System On Students Using C4.5 Algorithm," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 358–365, 2020, doi: 10.30812/matrik.v19i2.685.

- [4] H. Cahyono and Y. Yunita, "Sistem Penunjang Keputusan Promosi Jabatan Menggunakan Metode Profile Matching," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, pp. 31–35, 2020, doi: 10.31294/ijcit.v5i1.6467.
- [5] W. Nugroho Gultom, F. Prima Aditiawan, and R. Mumpuni, "Implementasi Metode Profile Matching Pada Aplikasi Sistem Pendukung Keputusan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3334–3340, 2024, doi: 10.36040/jati.v7i6.8194.
- [6] F. A. Sianturi, M. Kumari, and E. Laian, "Implementasi Algoritma C4.5 Menentukan Pola Berangkatan Jemaah Haji," *Jurnal Sains dan Teknologi*, vol. 3, no. 2, pp. 66–70, 2022, doi: 10.55338/saintek.v3i2.260.
- [7] A. Meiyanti and R. Komarudin, "Klasifikasi Diagnosa Penyakit Paru-Paru Pada Klinik Raditya Medical Center Dengan Metode Algoritma C4.5," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 12, no. 1, pp. 1894–1905, 2020, doi: 10.36706/jsi.v12i1.9456.
- [8] M. Fauzi, Y. Yudi, F. Tambunan, and ..., "Implementasi Profile matching Pada Pemilihan Calon Jemaah Haji," ... *Jurnal Penerapan Sistem ...*, vol. 2, no. 3, pp. 182–190, 2021, [Online]. Available: <http://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/80>
- [9] Z. Gustiana and A. Nia Sari, "Sistem Pendukung Keputusan Penentuan Kelulusan Mahasiswa Menggunakan Kombinasi Algoritma C 4.5 Dan Profile Matching," *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, vol. 6, no. 2, pp. 61–70, 2021, doi: 10.20527/jtiulm.v6i2.84.
- [10] E. Sutinah, "Kombinasi Algoritma C . 45 dan Profile Matching Pada Penilaian Kinerja Karyawan," *Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 4, no. 2, pp. 219–226, 2020.
- [11] Z. Muhyiddin, N. Khalifah, and M. Zidan, "Review Model of Decision Support System (DSS) Methods in The Contextual of Multimedia Platform Selection," vol. 5, no. 1, pp. 1–8, 2026.
- [12] M. A. U. Yosef Mulyanto Dawa, Abdul Aziz, "Optimasi Algoritma C4.5 Berbasis Particle Swarm Optimization (Pso)Untuk Menentukan Wholesales Penjualan Yosef," *Jurnal Riset Mahasiswa Bidang Teknologi Informasi*, vol. 6, pp. 21–26, 2023, [Online]. Available: <https://ejournal.unikama.ac.id/index.php/JFTI>
- [13] Y. Wang, "Prediction of rockburst risk in coal mines based on a locally weighted c4.5 algorithm," *IEEE Access*, vol. 9, pp. 15149–15155, 2021, doi: 10.1109/access.2021.3053001.
- [14] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [15] Rian Pratama, B. Huda, E. Novalia, and H. Kabir, "Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok," *Metik Jurnal*, vol. 6, no. 2, pp. 115–122, 2022, doi: 10.47002/metik.v6i2.379.
- [16] F. A. Lazuardy, A. Homaidi, and A. Lutfi, "Perbandingan Metode Data Mining Perbandingan Metode Klasifikasi Pengujian Kendaraan Bermotor Dengan Algoritma C4.5 Dan Naïve Bayes," *E-Link: Jurnal Teknik Elektro dan Informatika*, vol. 19, no. 2, p. 140, 2024, doi: 10.30587/e-link.v19i2.8113.
- [17] T. Gunantohadi and C. Crysdian, "Review Penerapan Metode Klasifikasi Pada Sistem Rekomendasi Sosial Kemasyarakatan," *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 3, no. 2, pp. 84–91, 2022, doi: 10.31102/jatim.v3i2.1578.
- [18] Saifulloh, "Decision Support System for Election of Madiunambassadors Using the Profile Matching Method," *Senatik*, vol. 1, no. 2, pp. 1–11, 2020, [Online]. Available: <http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1648>
- [19] F. Liu, Y. Ma, J. Wang, S. Wu, Y. Zhao, and Z. Wang, "Research on Multi-feature Constrained Adaptive Visual SLAM Localization Method Based on Decision Tree," *IEEE Access*, vol. 14, no. November 2024, pp. 31228–31240, 2025, doi: 10.1109/ACCESS.2025.3526669.