Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Analysis of the Most Popular Study Programs at Haji University of North Sumatra Using the Decision Tree Algorithm

Dewi Sartika[1], Muhammad Iqbal[2]
Master of Information Technology, University Pembangunan Panca Budi, Medan, Indonesia

Article Info
Keywords:
Decision Tree,
Study Program Clustering,
Applicants,
Gender,
Classification.

ABSTRACT
This study aims to analyze the clustering of the most popular study programs at Universitas Haji Sumatera Utara using the Decision Tree algorithm. This algorithm successfully grouped the study programs based on the applicants' interests, considering gender as the primary variable. The analysis results show that the most popular study programs among women are the Bachelor of Midwifery and the Bachelor of Nursing programs, which each have a very high number of female applicants. On the other hand, programs such as the Regular Bachelor of Law and Management show a more balanced interest between women and men, with Management having almost equal gender proportions. This classification model performed very well in detecting female applicants, with a high recall (95.51%) and good precision (79.84%). However, the model struggles to identify male applicants, with low recall (18.40%) and suboptimal precision (54.76%). This indicates that the model is more sensitive to predicting female applicants. Therefore, it is recommended that Universitas Haji Sumatera Utara enhance more inclusive and balanced marketing strategies, as well as optimize both regular and non-regular registration pathways to attract a more even interest from both genders, in order to achieve gender equality across various study programs and improve the efficiency of student admissions.

Corresponding Author:
Dewi Sartika
University Pembanguan Panca Budi, Medan, Indonesia
dewie.syartycka@gmail.com

INTRODUCTION

The rapid development of technology has brought significant changes across various sectors, including education, requiring higher education institutions to continuously adapt and enhance the quality of learning and services for students. In this context, one crucial aspect is a deep understanding of students' preferences in choosing academic programs [1]. The selection of a study program by prospective students is a key factor influencing their success throughout the educational process. Preferred study programs typically offer high-quality education with curricula aligned with contemporary demands, supported by modern facilities and competent faculty members. Moreover, these programs promise bright career prospects, making them a strategic choice for students aiming for professional success. Therefore, understanding and meeting students' preferences not only contribute to their academic success but also support the overall development of higher education institutions [2], [3], [4] .

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Haji Sumatera Utara University is one of the most sought-after higher education institutions in the field of health sciences in North Sumatra. This is due to the university's strong reputation for offering high-quality academic programs and promising career opportunities for its graduates. Consequently, this study focuses on analyzing the clustering of the most preferred academic programs at Haji Sumatera Utara University to understand student preferences and support the future development of better academic programs.

This study aims to analyze and identify the most favored academic programs among students by considering several key indicators. These indicators include the employability rate of graduates in the job market, the interest of prospective students, and recommendations from relevant institutions as a benchmark for program quality. To conduct this analysis, a classification method using the decision tree algorithm is employed, which is known for its effectiveness in categorizing data based on predefined criteria. This algorithm provides well-structured and easily interpretable results, facilitating the decision-making process for developing academic programs that align with students' needs and preferences [5]. Data mining is the process of exploring, analyzing, and extracting valuable patterns or information from large and complex datasets.

This process involves the use of statistical techniques, machine learning algorithms, artificial intelligence, and other methods to uncover hidden relationships, patterns, or trends in data that may not be immediately apparent [6], [7], [8], [9], [10], [11]. The decision tree algorithm is one of the popular and effective machine learning methods for data classification analysis. It offers advantages such as high classification accuracy and a simple model structure, making it a powerful tool for data-driven decision-making [12], [13].

The study conducted by Asmaul Husnah Nasrullah in 2021, titled "Implementation of the Decision Tree Algorithm for Best-Selling Product Classification," stated that the C4.5 decision tree algorithm proved to be accurate in classifying best-selling products, achieving an accuracy of 90% and an AUC of 0.709, indicating good performance on private data [14] A subsequent study by M. Riski Qisthiano et al. in 2023, titled "Application of the Decision Tree Algorithm in Predicting Student Graduation Classification," demonstrated that the decision tree model achieved an accuracy of 87.93% [15].

Other studies conducted by Ifani Hariyanti et al., Laila Qadrini et al., and Nanda Tri Hastuti et al. showed that the decision tree algorithm is more effective than other methods in various application fields. In their research, the algorithm was used to analyze product sales, classify poverty levels, and predict the impact of social media and sleep patterns on students' academic performance. The results indicated that the decision tree outperformed other methods, such as Adaboost and Naïve Bayes, in these tasks [16], [17], [18].

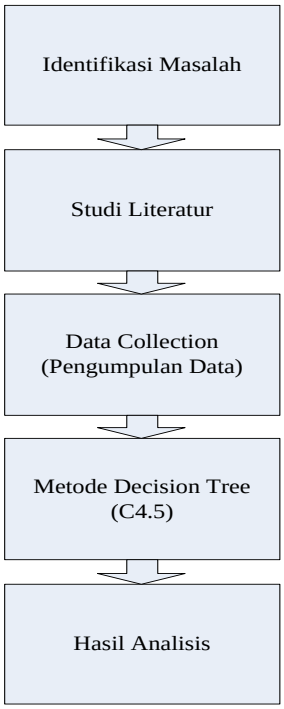
This study aims to analyze and identify the most preferred academic programs among students using the decision tree algorithm while considering gender factors. The findings are expected to assist the management of Haji Sumatera Utara University in formulating recruitment strategies and providing appropriate recommendations for prospective students in selecting study programs that align with their interests and abilities.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

METHODS

Research Stages

The following are the research stages conducted:



Fifure 1. Research Stages

1. Problem Identification
   The issue in grouping the most preferred academic programs at Haji Sumatera Utara University based on student enrollment across different admission waves arises due to several factors. First, there is an imbalance in the number of applicants per wave and differences in faculty preferences between waves. Additionally, external factors such as cost, gender, and scholarships influence study program choices. Variations in the selection process across waves, as well as inconsistencies or incomplete enrollment data, further complicate accurate classification. Therefore, an in-depth data analysis is required to identify enrollment patterns and formulate a more effective admission strategy.
2. Literature Review
   The researcher will gather theoretical references from various academic sources and relevant journals to support this study, focusing on the decision tree algorithm, classification, and clustering. The selected literature will help the researcher understand the fundamental theories of data analysis, particularly in identifying classification patterns in student data, such as study program and faculty selection trends. By studying these theories, the researcher can strengthen the theoretical foundation of the study, build an appropriate analytical framework, and compare previous research findings to gain deeper insights. This approach is expected to provide effective and

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

applicable solutions for Haji Sumatera Utara University, such as improving promotional strategies and supporting more data-driven decision-making.

3. Data Collection
   Student enrollment data from Wave 1 (February) to Wave 6 (October) 2024 will be collected as the primary sample for analysis. Data collection will be conducted through direct interviews with the new student admission committee and by extracting data from available databases. Subsequently, data preprocessing will be performed, including removing duplicates, handling missing values, and normalization, to ensure the dataset is clean and ready for analysis, ultimately generating valid patterns
4. Data Analysis
   The data processing steps for grouping the most preferred faculties using the decision tree method begin with collecting new student data from Wave 1 (February) to Wave 6 (October) 2024, which includes information such as admission wave, faculty, gender, and study program. The data is then cleaned to remove duplicates, address missing values, and ensure consistency. Relevant features, such as admission wave, faculty, gender, and study program, are identified, with faculty serving as the target variable for classification. The dataset is then split into a training set (80%) and a test set (20%) to train and evaluate the model. The decision tree model is trained using the training set, adjusting parameters such as tree depth and splitting criteria, and evaluated using metrics like accuracy and mean squared error (MSE) on the test set. The decision tree results are analyzed to identify patterns in faculty selection, and strategic recommendations, such as marketing approaches or resource planning, are formulated based on the findings. The model is then applied for faculty prediction in future admission waves and is periodically monitored to update the model with new data.

Decision Tree

A decision tree is a simple yet effective classification algorithm. One of its main advantages is that it provides classification rules that are easily interpretable by humans. However, decision trees also have some drawbacks, one of which is the need to sort all numerical attributes to determine the optimal split point for a node. This can be computationally expensive in terms of processing time and memory usage, especially when training a decision tree on large datasets.

Steps of the C4.5 Algorithm:

1. Select an Attribute as the Root
   Choose an attribute as the root of the decision tree by calculating the gain value for each attribute. The attribute with the highest gain value will be selected as the root node.
2. Create Branches for Each Attribute Value
   Once the root attribute is determined, create branches for each possible value of that attribute.
3. Partition Data into Branches Based on Attribute Values
   Divide the dataset into branches based on the corresponding attribute values.
4. Repeat the Process for Each Branch

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Repeat the previous steps for each branch until all data within a branch belong to the same class.

Steps in Building a Decision Tree Using the C4.5 Algorithm:

1. Prepare Training Data
   Prepare the training dataset derived from historical data. This dataset should already be classified into specific categories.
2. Determine the Root Node
   Select the attribute that will serve as the root node by calculating the gain value of each attribute. The attribute with the highest gain value will be chosen as the root.
   Before calculating the gain, the entropy value must first be computed using the following formula:

$$H(S) = -\sum_{i=1}^{n} p_i \log_2 p_i \quad (1)$$

   Information:
   S : is the dataset
   n : is the number of classes
   pi : is the proportion of each class in the dataset
3. Calculate the Gain Value
   After obtaining the entropy value, calculate the gain for each attribute using the following formula:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (2)$$

   Where:
   S : is the entire dataset
   A : is the attribute being evaluated
   Values(A) : represents all possible values of attribute A
   Sv : is the subset of S where attribute A has value v
   |Sv| : is the number of instances in subset Sv
   |S| : is the total number of instances in S
   Entropy(Sv) : is the entropy of subset Sv
4. Repartitioning
   Repeat the second step for each branch until all records are perfectly partitioned.
5. Stopping Criteria
   a. The partitioning process stops if one of the following conditions is met:
   b. All records in a node belong to the same class.
   c. There are no more attributes available for partitioning.
   d. No records remain in the branch.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Table 1. Preprocessed New Student Data for 2024

No	Period	Name	Gender	Registration Path	Chosen Study Program	Faculty
1	Gelombang 1	Lia Simahara	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
2	Gelombang 1	Lutfiah bunga sakila br lubis	Perempuan	Reguler	Manajemen	Fakultas Soshumdik
3	Gelombang 1	ahmad	Laki-laki	Reguler	Ilmu Keperawatan Program Diploma Tiga	Fakultas Ilmu Kesehatan
4	Gelombang 1	Lukmanl Hakim	Laki-laki	Reguler	Kebidanan Program Diploma Tiga	Fakultas Ilmu Kesehatan
5	Gelombang 1	Atik cibro	Perempuan	Reguler	Kebidanan Program Diploma Tiga	Fakultas Ilmu Kesehatan
6	Gelombang 1	Mulia jayanti	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
7	Gelombang 1	Anisah hasanah	Perempuan	Reguler	S1 Farmasi	Fakultas Ilmu Kesehatan
8	Gelombang 1	Marlina ariga	Perempuan	Reguler	S1 Farmasi	Fakultas Ilmu Kesehatan
9	Gelombang 1	Kayla Nazwa	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
10	Gelombang 1	kaidah nurul risti	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
11	Gelombang 1	Lili	Perempuan	Reguler	S1 Hukum	Fakultas Soshumdik
12	Gelombang 1	DAHMAN SIREGAR	Laki-laki	Non Reguler	Ilmu Kesehatan Masyarakat	Fakultas Ilmu Kesehatan
13	Gelombang 1	Nadila adzhani	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
14	Gelombang 1	Della Sulistiani	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
15	Gelombang 1	Inka saamita nasution	Perempuan	Reguler	S1 Farmasi	Fakultas Ilmu Kesehatan
16	Gelombang 1	Suci aurelia ananta	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
17	Gelombang 1	Anggi mutia	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
18	Gelombang 1	Tri Priono	Laki-laki	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
19	Gelombang 1	Fikra adelia	Perempuan	Reguler	S1 Farmasi	Fakultas Ilmu Kesehatan
20	Gelombang 1	Raisa Restyapani	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
21	Gelombang 1	annisa melani rambe	Perempuan	Reguler	Kebidanan Program Diploma Tiga	Fakultas Ilmu Kesehatan

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

No	Period	Name	Gender	Registration Path	Chosen Study Program	Faculty
22	Gelombang 1	dimas sanjaya pasaribu	Laki-laki	Reguler	S1 Hukum	Fakultas Soshumdik
23	Gelombang 1	Rudzi ramdani	Laki-laki	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
24	Gelombang 1	Naila fazira nasution	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
25	Gelombang 1	Nadia sapira ritonga	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
26	Gelombang 1	Adis aulia nainggolan	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
27	Gelombang 1	Kholila Indah sari harapah	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
28	Gelombang 1	Abdah Haswani Panggabean	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
29	Gelombang 1	Andini Permata Sari	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
....	....	....	....	....	....	....
539	Gelombang 6	Nikmah Pasaribu	Perempuan	Non Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
540	Gelombang 6	Annisa Rahman Hasibuan	Perempuan	Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
541	Gelombang 6	Nora Pradana Sukma	Perempuan	Non Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
542	Gelombang 6	Nur halimah Hasibuan	Perempuan	Non Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
543	Gelombang 6	Dian Puspita Sari	Perempuan	Reguler	S1 PGSD	Fakultas Soshumdik
544	Gelombang 6	Risma Ramadani	Perempuan	Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan
545	Gelombang 6	Rafly al farezy hasibuan	Laki-laki	Reguler	Manajemen	Fakultas Soshumdik
546	Gelombang 6	Bowo Satrio	Laki-laki	Reguler	S1 Hukum	Fakultas Soshumdik
547	Gelombang 6	Rubiat	Perempuan	Non Reguler	Sarjana Kebidanan	Fakultas Ilmu Kesehatan
548	Gelombang 6	Hendra Junaidi	Laki-laki	Non Reguler	Ilmu Keperawatan Program Sarjana	Fakultas Ilmu Kesehatan

Table 1 presents the data of new student applicants at Haji Sumatera Utara University from February to October 2024, with a total of 548 applicants after data cleaning. This dataset has undergone preprocessing and includes attributes such as registration wave, name, gender, registration path, chosen study program, and faculty. The registration process was conducted in six waves. This data will be analyzed using the decision tree algorithm to determine the most popular faculty in the 2024 new student admissions at Haji Sumatera Utara University.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

RESULTS AND DISCUSSION

This study utilizes new student registration data from Wave 1 to Wave 6, conducted from February to October 2024, with a total of 548 new student records in .xlsx (Excel) format. The dataset includes information such as the applicant's full name, registration number, ID card number, phone number, gender, place and date of birth, religion, nationality, registered address, and current address. Additionally, applicants are required to select a primary and optional study program along with the relevant faculty. Additional data includes the applicant's school of origin, student identification number (NISN), registration path, registration administration status, and parental information such as name, occupation, education, and income. Supporting documents required in the registration process include a birth certificate, photograph, ID card, family card, diploma, academic transcript, and other necessary documents. In this study, the attributes used include the registration period, name, gender, registration path, chosen study program, and faculty. The preprocessing stage is carried out to refine raw data by eliminating problematic or erroneous records, ensuring high-quality data ready for further processing. From the total 548 collected records, a thorough check was conducted to ensure there were no errors or inconsistencies in the data.

Decision Tree Modeling Using RapidMiner

RapidMiner is an open-source software used for data analysis, including classification with various descriptive and predictive techniques, to generate information and knowledge that assist in accurate decision-making. In RapidMiner version 2024 1.0, the C4.5 algorithm can be easily implemented to build a classification model based on a decision tree, allowing users to efficiently process data and generate reliable predictions. The C4.5 algorithm is one of the methods used for classification or predictive segmentation by constructing a decision tree, which is a powerful and easy-to-understand classification and prediction method because it can transform complex data into simple rules. According to Kusrini, the process of building a decision tree using the C4.5 algorithm generally involves several stages: selecting an attribute as the root, creating branches for each attribute value, splitting the data into these branches, and repeating this process for each branch until all data in a branch belong to the same class.

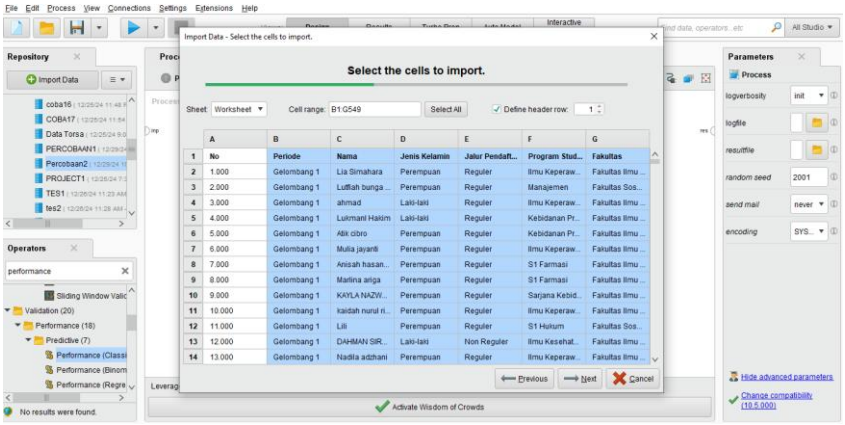


Figure 2. Data Transformation.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

In Figure 2, the data transformation and data reduction process is explained, aiming to improve efficiency and accuracy in data processing. During the transformation stage, each attribute is initialized with an appropriate data type for use by the C4.5 algorithm, where the polynomial data type is used, and the gender attribute is labeled as a classification determinant according to the needs of the data mining process. Next, data reduction is performed by removing irrelevant attributes, such as the number attribute, ensuring that only the necessary attributes are included, namely the period, name, gender, enrollment path, chosen study program, and faculty. This process aims to simplify the data by eliminating unimportant variables without reducing the quality of the obtained information. After data reduction, the modeling stage is carried out to select the appropriate data mining technique and algorithm. In this study, the C4.5 algorithm is used, with model implementation performed using RapidMiner software.

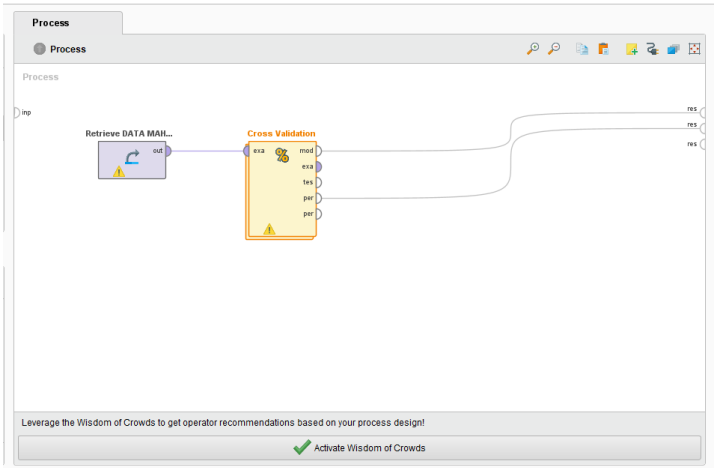


Figure 3. Model Cross Validation

Figure 3 illustrates the Cross-validation model, an evaluation technique that divides the dataset into multiple folds to ensure that each data point is alternately used as part of the training and testing data. This technique helps reduce bias in model performance evaluation and provides a more objective estimation of the model's performance on previously unseen data.

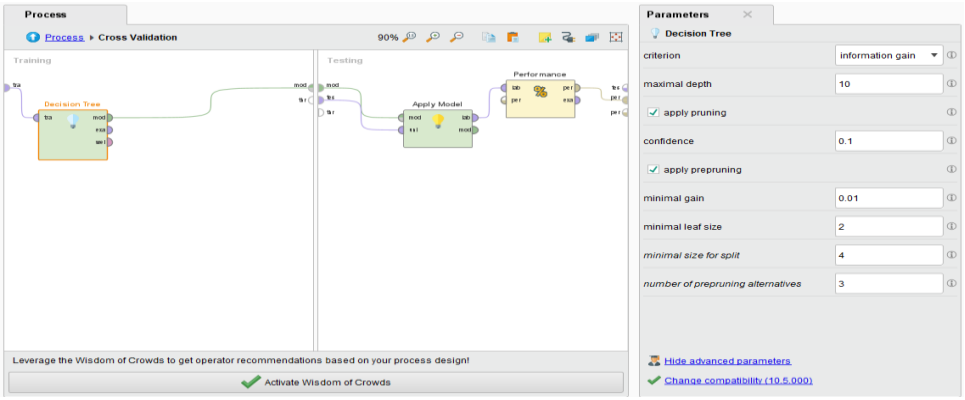


Figure 4. Model Decision Tree

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Figure 4 illustrates the operator design in the data processing workflow for new student admissions at Haji Sumatera Utara University in 2024 using RapidMiner. Each operator in the design is interconnected, forming an integrated workflow. The process begins with reading an .xlsx (Excel) data file using the Retrieve operator. This operator is connected to Cross Validation, which functions to split the dataset into multiple folds, ensuring that each data portion is alternately used as training data and testing data. In Cross Validation, the dataset is divided into two parts: 80% for training data and 20% for testing data. In this study, the training data is used to build a predictive model using a machine learning algorithm, while the testing data is used to validate the model, assess its accuracy, and evaluate its performance. Next, the Decision Tree algorithm operator is included as required for data classification. The Apply Model operator is then connected to the Performance operator, which evaluates the model's performance based on metrics such as accuracy, precision, and recall. Once all operators are configured, the model can be executed by clicking the Run button.

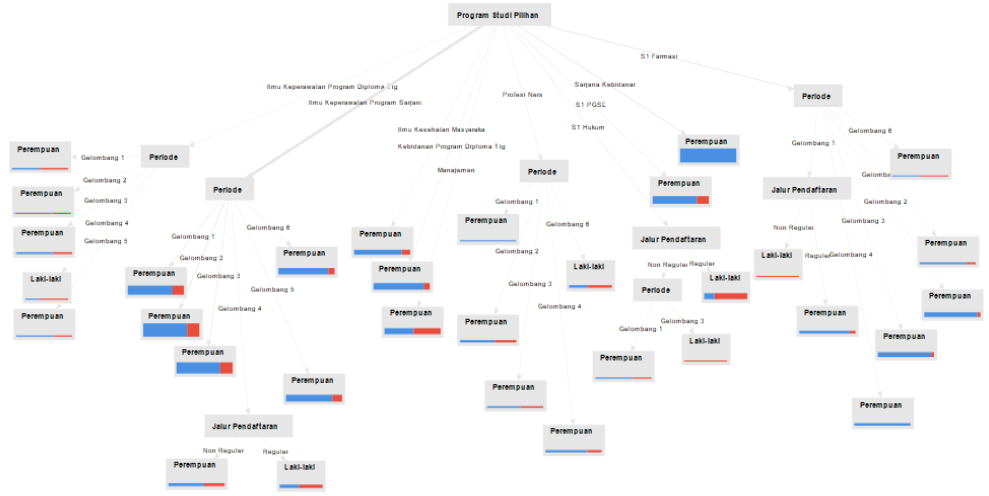


Figure 5. Decision Tree Result

Tree

```
Program Studi Pilihan = Ilmu Keperawatan Program Diploma Tiga
|   Periode = Gelombang 1: Perempuan {Perempuan=2, Laki-laki=2}
|   Periode = Gelombang 2: Perempuan {Perempuan=2, Laki-laki=1}
|   Periode = Gelombang 3: Perempuan {Perempuan=4, Laki-laki=2}
|   Periode = Gelombang 4: Laki-laki {Perempuan=1, Laki-laki=2}
|   Periode = Gelombang 5: Perempuan {Perempuan=2, Laki-laki=1}
Program Studi Pilihan = Ilmu Keperawatan Program Sarjana
|   Periode = Gelombang 1: Perempuan {Perempuan=32, Laki-laki=8}
|   Periode = Gelombang 2: Perempuan {Perempuan=47, Laki-laki=13}
|   Periode = Gelombang 3: Perempuan {Perempuan=39, Laki-laki=11}
|   Periode = Gelombang 4
|   |   Jalur Pendaftaran = Non Reguler: Perempuan {Perempuan=7, Laki-laki=4}
|   |   Jalur Pendaftaran = Reguler: Laki-laki {Perempuan=6, Laki-laki=7}
|   Periode = Gelombang 5: Perempuan {Perempuan=24, Laki-laki=5}
|   Periode = Gelombang 6: Perempuan {Perempuan=24, Laki-laki=3}
Program Studi Pilihan = Ilmu Kesehatan Masyarakat: Perempuan {Perempuan=17, Laki-laki=3}
Program Studi Pilihan = Kebidanan Program Diploma Tiga: Perempuan {Perempuan=25, Laki-laki=3}
Program Studi Pilihan = Manajemen: Perempuan {Perempuan=13, Laki-laki=12}
Program Studi Pilihan = Profesi Ners
|   Periode = Gelombang 1: Perempuan {Perempuan=2, Laki-laki=0}
|   Periode = Gelombang 2: Perempuan {Perempuan=5, Laki-laki=3}
|   Periode = Gelombang 3: Perempuan {Perempuan=3, Laki-laki=2}
|   Periode = Gelombang 4: Perempuan {Perempuan=6, Laki-laki=2}
|   Periode = Gelombang 6: Laki-laki {Perempuan=4, Laki-laki=5}
Program Studi Pilihan = S1 Farmasi
|   Periode = Gelombang 1
|   |   Jalur Pendaftaran = Non Reguler: Laki-laki {Perempuan=0, Laki-laki=2}
|   |   Jalur Pendaftaran = Reguler: Perempuan {Perempuan=9, Laki-laki=1}
|   Periode = Gelombang 2: Perempuan {Perempuan=18, Laki-laki=1}
```

Figure 6. Decision Tree Description Result

From the decision tree data, it is evident that applicants for study programs are predominantly female, especially in health-related fields such as the Bachelor's in Midwifery, Diploma Three in Midwifery, and Nursing Science. The Bachelor's in Midwifery recorded the

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

highest number of female applicants, with 63 individuals, while the Bachelor's in Nursing Science had a significant number of applicants in the second admission wave, totaling 60 (47 females and 13 males). On the other hand, programs such as Management and Law (Bachelor's) show a more balanced gender distribution, with the Bachelor's in Law (Regular) recording 19 male applicants, the highest among all programs. Male applicants tend to show greater interest in non-health programs like Law and Management. Additionally, the regular admission pathway is more preferred than the non-regular one, as seen in programs like Pharmacy (Bachelor's) and Law (Bachelor's). Overall, the programs with the highest number of applicants are the Bachelor's in Midwifery and the Bachelor's in Nursing Science, indicating a strong appeal for health-related fields. This data illustrates gender preference trends and the popularity of certain study programs, which can serve as a reference for academic planning and student admission strategies in the future.

accuracy: 77.91% +/- 3.92% (micro average: 77.92%)

	true Perempuan	true Laki-laki	class precision
pred. Perempuan	404	102	79.84%
pred. Laki-laki	19	23	54.76%
class recall	95.51%	18.40%	

Figure 7. Accuracy Results of the Decision Tree Method.

The accuracy results of the Decision Tree method using the C4.5 algorithm show that the Precision for the Female category is 79.84%, meaning that approximately 79.84% of all predictions classified as female are indeed correct. Meanwhile, the Precision for the Male category is 54.76%, indicating that only about 54.76% of male predictions are actually correct. For Recall, the model achieved 95.51% for the Female category, demonstrating that the model successfully identified 95.51% of all actual females in the data. However, for the Male category, the Recall is only 18.40%, meaning the model was able to identify only 18.40% of the actual males in the dataset.

CONCLUSION

Based on the research findings, it can be concluded that the Decision Tree algorithm successfully categorized study programs based on applicant interest, considering gender as the main variable. The study programs most favored by female applicants are the Bachelor's in Midwifery and the Bachelor's in Nursing Science, both of which have a significantly high number of female applicants. On the other hand, programs such as the Bachelor's in Law (Regular) and Management show a more balanced interest between males and females, with Management being the program with the most nearly equal gender proportion. This classification model performs exceptionally well in detecting female applicants, with a very high recall (95.51%) and a reasonably good precision (79.84%). However, the model struggles with identifying male applicants, as indicated by a very low recall (18.40%) and suboptimal precision (54.76%). This suggests that the model is more sensitive to predicting females and tends to have difficulty accurately classifying males. Overall, although the model effectively identifies female applicants, improvements are still needed to enhance its performance in classifying male applicants. Therefore, it is recommended that Haji Sumatera

Utara University consider implementing more inclusive and balanced marketing and promotional strategies, as well as optimizing the management of both regular and non-regular admission pathways to attract interest from both genders more evenly. With these improvements, the university is expected to enhance the efficiency of student admissions and achieve gender balance across various study programs.

ACKNOWLEDGEMENT

Acknowledgments are extended to the lecturers of the Master's Program in Information Technology at Universitas Pembangunan Panca Budi, friends, and the academic community of Universitas Haji Sumatera Utara for granting permission to use data for this research. Special thanks are also addressed to our families for their unwavering support, motivation, prayers, and both moral and material assistance. The author also expresses gratitude to everyone who has contributed to the completion of this scientific work, even though they cannot be mentioned individually. Hopefully, this work will be beneficial to readers and contribute to improving the quality of future research..

REFERENCE

[1] R. Ghaniy and F. Indriyaningsih, "Penerapan Metode Fuzzy C-Means dalam Pemilihan Program Studi Mahasiswa Baru di Perguruan Tinggi," *TeknoIS : Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 10, no. 2, pp. 19–30, Nov. 2020, doi: 10.36350/JBS.V10I2.84.

[2] N. Nursobah, "Penerapan Metode Preference Selection Index (PSI) Dalam Pemilihan Perguruan Tinggi Swasta Program Studi IT di Provinsi Kalimantan Timur," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 1045, Jul. 2021, doi: 10.30865/MIB.V5I3.3081.

[3] L. Indana and R. Hidayati, "Keterserapan Kerja Lulusan Program Studi Sistem Informasi Universitas Merdeka Malang di Era Pandemi," *Briliant: Jurnal Riset dan Konseptual*, vol. 7, no. 4, pp. 896–904, Nov. 2022, doi: 10.28926/BRILIANT.V7I4.1024.

[4] M. H. Mohd Anuar, M. Z. Mohd Nawi, R. Khalid, and A. J. Mohd Hatta @ Azman, "Persepsi Pelajar Sijil Operasi Perhotelan Terhadap Perlaksanaan Program Alumni di Kolej Komuniti," *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, vol. 7, no. 10, p. e001786, Oct. 2022, doi: 10.47405/MJSSH.V7I10.1786.

[5] J. J. Purnama, H. M. Nawawi, S. Rosyida, R. Ridwansyah, and R. Risnandar, "Klasifikasi Mahasiswa HER Berbasis Algoritma SVM dan Decision Tree," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 6, pp. 1253–1260, Dec. 2020, doi: 10.25126/JTIIK.0813080.

[6] N. Q. Rahmawati, "IMPLEMENTASI DATA MINING TERHADAP DATA MINAT MAHASISWA MENGGUNAKAN METODE APRIORI," *Jurnal Teknologi Kimia Unimal*, vol. 13, no. 2, pp. 215–228, Nov. 2024, doi: 10.29103/JTKU.V13I2.19563.

[7] Saefudin and D. Fernando, "Application of Book Recommendation Data Mining Using the Apriori Algorithm," *JSil (Jurnal Sistem Informasi)*, vol. 7, no. 1, pp. 50–56, 2020.

[8] T. D. Saribu, M. S. Wahyuni, and S. F. Rezky, "Data Mining Dalam Menentukan Pola Penjualan Barang Menggunakan Algoritma Apriori," *Jurnal Sistem Informasi Triguna*

Dharma (JURSI TGD), vol. 3, no. 2, pp. 183–190, Mar. 2024, doi: 10.53513/JURSI.V3I2.6145.

[9] F. Harahap, W. Fahrozi, R. Adawiyah, E. T. Siregar, and A. Y. N. Harahap, "Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes," *JURNAL UNITEK*, vol. 16, no. 1, pp. 41–51, Jun. 2023, doi: 10.52072/unitek.v16i1.541.

[10] A. I. Zalukhu, D. Sartika, and S. Wahyuni, "Penerapan Algoritma Apriori untuk Optimasi Strategi Penjualan Berdasarkan Analisis Pola Pembelian di Torsa Cafe," *Bulletin of Information Technology (BIT)*, vol. 5, no. 4, pp. 295–304, Dec. 2024, doi: 10.47065/BIT.V5I4.1715.

[11] Hamsiah, "Implementasi Data Mining Dalam Penerapan Clustering Algoritma K-Medoid Sebaran Mahasiswa Baru Pada STIE-SAK," *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, vol. 8, no. 1, pp. 16–24, Jan. 2025, doi: 10.47324/ILKOMINFO.V8I1.311.

[12] Y. Feng, H. Li, X. Wang, J. Tian, and Y. Qi, "Application of Machine Learning Decision Tree Algorithm Based on Big Data in Intelligent Procurement," *Journal of Social Science and Humanities*, vol. Vol.7, no. Issue 2 (Apr), pp. 108–115, Apr. 2024, doi: 10.53469/wjimt.2024.07(02).14.

[13] E. Rosida and A. Firmansyah, "Comparative Analysis of Classification of K-Nearest Neighbor (KNN) Algorithm and Decision Tree in Breast Cancer Using Rapidminer," *International Journal of Applied Research and Sustainable Sciences*, vol. 2, no. 12, pp. 1039–1060, Dec. 2024, doi: 10.59890/IJARSS.V2I12.48.

[14] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, Sep. 2021, doi: 10.35329/JIIK.V7I2.203.

[15] M. Riski, Q. 1∗, P. Armilia Prayesy, I. Ruswita, and R. Artikel, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/GTECH.V7I1.1850.

[16] I. Hariyanti, M. Al-Husaini, and A. R. Raharja, "Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa," *Technologia : Jurnal Ilmiah*, vol. 15, no. 2, pp. 332–240, Apr. 2024, doi: 10.31602/TJI.V15I2.14381.

[17] O. Laila Qadrini, A. Seppewali, A. Aina, P. Studi Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and U. Sulawesi Barat, "Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial," *Jurnal Inovasi Penelitian*, vol. 2, no. 7, pp. 1959–1966, Dec. 2021, doi: 10.47492/JIP.V2I7.1046.

[18] N. T. Hastuti and F. Budiman, "Optimasi Klasifikasi Stunting Balita dengan Teknik Boosting pada Decision Tree," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 655–664, Dec. 2024, doi: 10.29408/EDUMATIC.V8I2.27913.