Journal of Information Technology, computer science and Electrical Engineering (JITCSE)
Vol. 1, No. 3, October 2024, pp. 517~521
ISSN: 3046-4900, DOI: 10.30596/jitcse

Student Data Analysis for Decision Making Using Decision Tree

Risca Sri Mentari, Muhammad Iqbal
[1,2]Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

ABSTRACT

In today's digital era, data has become a valuable asset that supports effective decision-making in various fields, including education. In Indonesia, the education sector needs a technology-based strategy to utilize data optimally in improving the quality of education services. Data mining, as an information technology approach, plays an important role in extracting valuable information from big and complex data. Classification algorithms such as the Decision Tree, specifically the C4.5 algorithm, are widely used in data mining to build accurate decision models. This study aims to apply the C4.5 algorithm to student data to support evidence-based decision-making in education. Using RapidMiner software, this research focuses on classifying and analyzing student data to build a model that can simplify the decision-making process, making it easier to understand and implement. The results of this study show that there is a pattern of gender distribution in various classes, with some classes dominated by female students and others dominated by male students. The preprocessing stage successfully simplifies the data, so that relevant information can be analyzed more easily. These results underscore the importance of data mining technology in education data analysis for better decision-making, as well as provide new insights in designing data-driven education policies.

Keyword : Data mining, C4.5; Decision Tree; RapidMiner.

Corresponding Author:
Risca Sri Mentari,
Master of Information Technology
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing 20122, Medan, Indonesia.
Email : risca.mentari18@gmail.com

Article history:
Received Nov 28, 2024
Revised Nov 29, 2024
Accepted Nov 30, 2024

1. INTRODUCTION

In today's digital era, data has become one of the most valuable assets in supporting effective decision-making in various fields, including education. Education in Indonesia requires a technology-based strategy to optimally utilize data in improving the quality of education services (Sitanggang & Nabila, 2024). Data mining, as one of the information technology approaches, is used to extract valuable information from big and complex data (Risqi Ananda et al., 2023). In this process, classification algorithms such as Decision Tree play a major role.

The C4.5 algorithm is one of the classification techniques in machine learning used in data mining to form a decision tree with a high level of accuracy (Risnawati et al., 2023). Decision Trees turn data into decision rules, allowing decision-makers to understand problems in a simpler and more systematic way (Imam Nawawi & Zaehol Fatah, 2024). The Decision Tree model consists of nodes and branches, where each node represents the feature to be classified, while the branch defines the value that the node can retrieve (Firmansyach et al., 2023).

Classification is the process of identifying and grouping objects into the same category based on specific attributes (Adhitya et al., 2023). In this context, RapidMiner, as one of the software for data analysis, makes it easier for users to carry out data mining, text mining, and predictive analysis processes (Nahjan et al., 2023) (Sholeh et al., 2023).

This study aims to apply the Decision Tree algorithm to student data to support evidence-based decision making. By utilizing the C4.5 algorithm, this study focuses on the process of classifying and analyzing student data to produce a model that can make it easier to determine education policies. The main benefit of this approach is its ability to simplify complex decision-making processes to be easier to understand and implement (Hariyant et al., 2024).

2. RESEARCH METHOD

A. Data Collection

At this stage, the researcher collected data from students in grade XII at SMK Negeri 1 Stabat for the 2023-2024 academic year. Data collection is one of the most important stages in research (Yasin et al., 2024).

B. Preprocessing

Once the data is collected, the next step is preprocessing, Preprocessing is the stage of data cleaning and improvement. At the time of collecting data, the data obtained is usually unstructured and contains many irrelevant characters. The purpose of preprocessing is to remove the noise (Insan et al., 2023).

C. Modeling and Analysis

Next, the researcher modeled the decision tree using rapidminer, after being modeled, the researcher conducted an analysis. Modeling is one of the most important steps of effective data mining (Leni et al., 2023). Analysis is an effort to observe in detail a thing or object by describing its constituent components or constituents for further study (Alfayed et al., 2023).

3. RESULTS AND DISCUSSION

A. Data Collection

	A	B	C	D
1	NPSN	NAME	GENDER	CLASS
2	0072736864	Adela Safita	F	XII DPIB 1
3	0058680735	Adlina Zahra	F	XII DPIB 1
4	0067819915	Agung Syahputra	M	XII DPIB 1
5	0068525816	Ahmad Irzi Izham	M	XII DPIB 1
6	0057213235	Alindia Puri	F	XII DPIB 1
7	0051915880	Angga Syahputra	M	XII DPIB 1
8	0068796774	Angga Wiranata Kaban	M	XII DPIB 1
9	0056897839	Daffa Faatin Mtd	M	XII DPIB 1
10	0068256171	Danu Pratama	M	XII DPIB 1
11	0063459906	Desta Febriyansyah	M	XII DPIB 1
12	0064487937	Eka Rahma Salsabila	F	XII DPIB 1
13	3062698124	Felycya Adelya	F	XII DPIB 1
14	0067577359	Habi Septia Ardinata	M	XII DPIB 1
15	0062915580	Imelda Risma Sari	F	XII DPIB 1
16	0069540965	Juwandi	M	XII DPIB 1
17	0062371804	M. Amiza Rizky	M	XII DPIB 1
18	0051123558	M. Rizky Pratama	M	XII DPIB 1
19	0062554441	Milda Padela Putri	F	XII DPIB 1
20	3076232664	Muhammad Fahrozi	M	XII DPIB 1
21	0068670044	Nabilla Ramadhani	F	XII DPIB 1

Data

Fig 1. Grade XII Student Data

In Figure 1, data on grade XII students is presented consisting of several main attributes, namely NPSN as a unique identity of the school, student names that show the identity of each individual, gender as gender information, and classes that display the division of students based on learning level.

	A	B	C	D
1	GENDER	CLASS		
2	F	DPIB 1		
3	F	DPIB 1		
4	M	DPIB 1		
5	M	DPIB 1		
6	F	DPIB 1		
7	M	DPIB 1		
8	M	DPIB 1		
9	M	DPIB 1		
10	M	DPIB 1		
11	M	DPIB 1		
12	F	DPIB 1		
13	F	DPIB 1		
14	M	DPIB 1		
15	F	DPIB 1		
16	M	DPIB 1		
17	M	DPIB 1		
18	M	DPIB 1		
19	F	DPIB 1		
20	M	DPIB 1		
21	F	DPIB 1		

Preprocessing

Fig 2. Grade XII Student Data after Preprocessing

In Figure 2, the data of grade XII students after going through the processing process is displayed, which shows information that has been simplified into two main attributes, namely gender as the identification of the student's gender and class as the category of their learning level.

B. Pemodelan dan Analisis

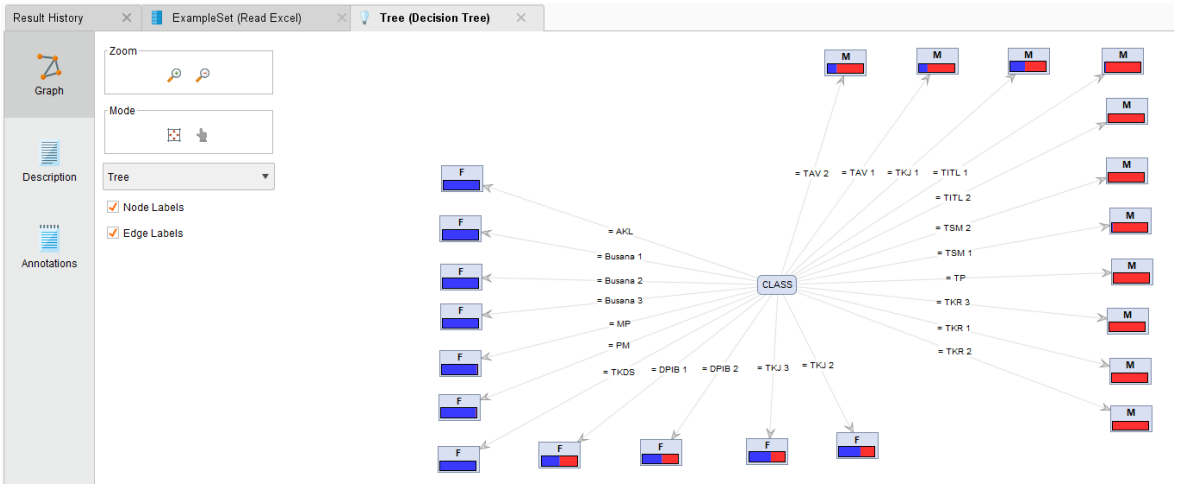

Fig 3. Model Decision Tree

In Figure 3, it can be seen that the decision tree model consists of 11 branches on the left side (7 branches are bright blue, and the other 4 branches are a combination of blue and red), and 11 branches on the right side (8 branches are bright red, and the other 3 branches are a combination of red and blue). The female gender dominated more in the classes (AKL, Busana 1, Busana 2, Busana 3, MP, PM, TKDS, DPIB 1, DPIB 2, TKJ 3, and TKJ 2), while the male gender dominated more in the classes (TKR 2, TRR 1, TKR 3, TP, TSM 1, TSM 2, TITL 2, TITL 1, TKJ 1, TAV 1, and TAV 2).

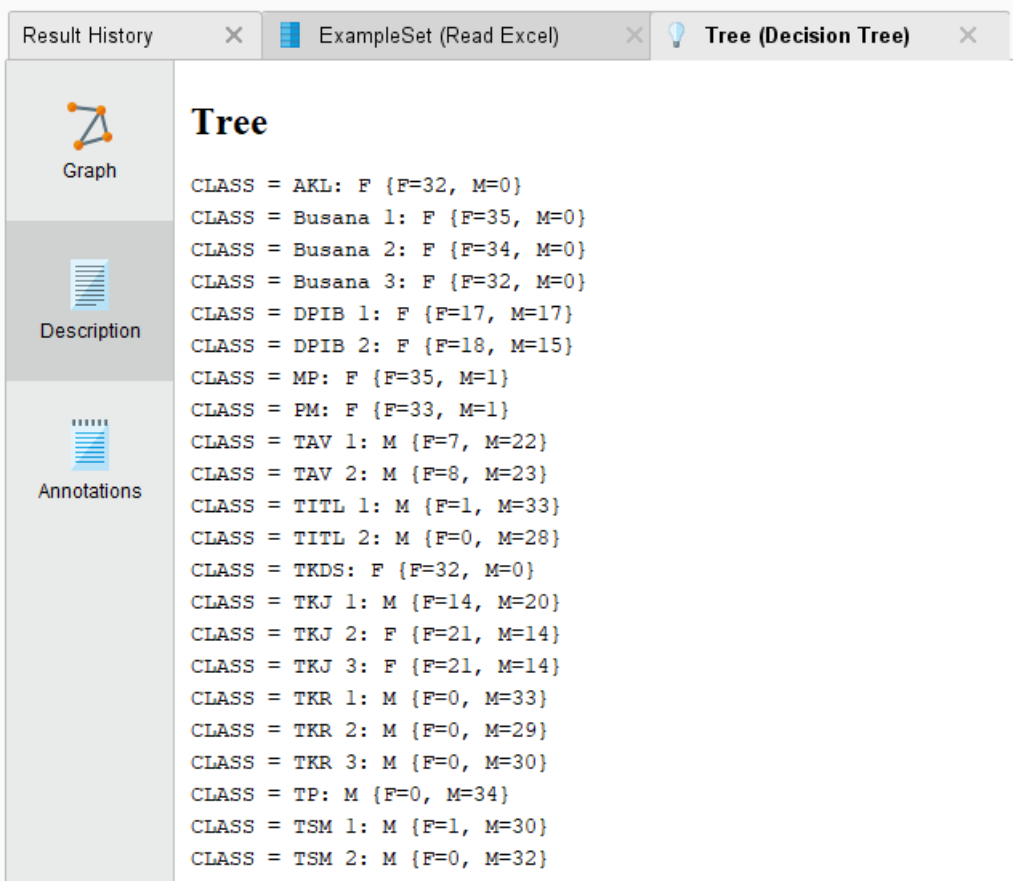

Fig 4. Description Decision Tree

In Figure 4, it can be seen that this data shows the distribution of female (F) and male (M) students across various classes. Classes with a majority of females (F) include "AKL" (32 F, 0 M), "Busana 1" (35 F, 0 M), "Busana 2" (34 F, 0 M), "Busana 3" (32 F, 0 M), "TKDS" (32 F, 0 M), "MP" (35 F, 1 M), "PM" (33 F, 1 M), "TKJ 2" (21 F, 14 M), and "TKJ 3" (21 F, 14 M). On the other hand, classes with a majority of males (M), such as "TAV 1" (7 F, 22 M), "TAV 2" (8 F, 23 M), "TITL 1" (1 F, 33 M), "TITL 2" (0 F, 28 M), "TKR 1" (0 F, 33 M), "TKR 2" (0 F, 29 M), "TKR 3" (0 F, 30 M), "TP" (0 F, 34 M), "TSM 1" (1 F, 30 M), and "TSM 2" (0 F, 32 M), show a dominance of males. Some classes, such as "DPIB 1" (17 F, 17 M) and "DPIB 2" (18 F, 15 M), have a nearly balanced distribution of female and male students.

4. CONCLUSION

This study shows that the application of the C4.5 algorithm in student data analysis using RapidMiner provides significant results in facilitating evidence-based decision-making in education. Based on the Decision Tree model that was built, it can be seen that there is a gender distribution pattern in various classes at SMK Negeri 1 Stabat, with certain classes dominated by female students and other classes dominated by male students. The preprocessing process that is carried out successfully simplifies the data, so that relevant information can be easily analyzed. These results illustrate how important data mining technology is in optimizing education data analysis for better decision-making, as well as providing new insights in designing data-based education policies.

REFERENCES

Adhitya, R. R., Wina Witanti, & Rezki Yuniarti. (2023). PERBANDINGAN METODE CART DAN NAÏVE BAYES UNTUK KLASIFIKASI CUSTOMER CHURN. *INFOTECH Journal*, *9*(2), 307–318. https://doi.org/10.31949/infotech.v9i2.5641

Alfayed, E., Ramadeli, L., Agnestasia, R., Amalina, V., Swid, Z. H. O., & Riofita, H. (2023). ANALISIS STRATEGI PEMASARAN DAN PENJUALAN E-COMMERCE PADA TIKTOK SHOP. *Jurnal Ekonomi Manajemen Dan Bisnis*, *1*(2), 195–201.

Firmansyach, W. A., Hayati, U., & Wijaya, Y. A. (2023). ANALISA TERJADINYA OVERFITTING DAN UNDERFITTING PADA ALGORITMA NAIVE BAYES DAN DECISION TREE DENGAN TEKNIK CROSS VALIDATION. *Jurnal Mahasiswa Teknik Informatika*, *7*(1), 262–269.

Hariyant, I., Al-Husaini, M., & Raharja, A. R. (2024). Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa. *Technologia : Jurnal Ilmiah*, *15*(2), 332. https://doi.org/10.31602/tji.v15i2.14381

Imam Nawawi, & Zaehol Fatah. (2024). Penerapan Decision Trees dalam Mendeteksi Pola Tidur Sehat Berdasarkan Kebiasaan Gaya Hidup. *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, *2*(4), 34–41. https://doi.org/10.59024/jiti.v2i4.969

Insan, M. K., Hayati, U., & Nurdiawan, O. (2023). ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 478–483.

Nahjan, M. R., Heryana, N., & Voutama, A. (2023). IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 101–104.

Risnawati, I., Rahma, S., Kusuma, F., Salsabila[4], N., Nourmansyah[5], A., Farhan[6], A., & Irfiani, E. (2023). Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5. *Jurnal INSAN (Journal of Information Systems Management Innovation*, *3*(2), 78–84. http://jurnal.bsi.ac.id/index.php/jinsan

Risqi Ananda, M., Maharani, N. S., Fadhila, E., Rahma, A., & Nurbaiti. (2023). Data Mining Dalam Perusahaan PT Indofood Lubuk Pakam. *Jurnal Ekonomi Manajemen Dan Bisnis (JEMB)*, *02*(1), 97–102. https://doi.org/10.47233/jemb.v2i1.1009

Sholeh, M., Kumalasari Nurnawati, E., & Lestari, U. (2023). Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner. *Jurnal Informatika Sunan Kalijaga)*, *8*(1), 10–21. https://archive.ics.uci.edu/ml/datasheets.php.

Sitanggang, A. S., & Nabila, A. (2024). Peran Teknologi Digital dalam Membangun Pendidikan Generasi Emas. *Jurnal Pendidikan Tambusai*, *8*(2), 30802–30809. https://www.researchgate.net/publication/382916341

Yasin, M., Garancang, S., & Hamzah, A. A. (2024). Metode dan Instrumen Pengumpulan Data (Kualitatif dan Kuantitatif). *Journal of International Multidisciplinary Research*, *2*(3), 161–173. https://journal.banjaresepacific.com/index.php/jimr