**Journal of Information Technology, computer science and Electrical Engineering (JITCSE)**
Vol. 1, No. 3, October 2024, pp. 160~163
ISSN: 3046-4900, DOI: 10.30596/jitcse


# Analysis of Student Graduation at SMK Negeri 1 Stabat Using the C4.5 Algorithm

**Sulis Sutiono[1], Darmeli Nasution[2]**
[1,2]Master of Information Technology, Universitas Pembangunan Panca Budi

**ABSTRACT**

This research aims to analyze the graduation rates of students at SMK Negeri 1 Stabat in the general subjects group using the C4.5 algorithm. Secondary data covering 717 students was collected and underwent a preprocessing stage to ensure accuracy. The modeling results indicate that the average score threshold for graduation is 73.823; of the 717 students, 709 are declared graduated and 8 did not pass. The use of the C4.5 algorithm has proven effective in providing insights into student graduation as well as generating decision tree visualizations that clarify the decision-making process. This study emphasizes the importance of applying data mining technology in education to enhance understanding of student learning outcomes.

**Keyword : C4.5 Algorithm; Data Mining; Student Graduation.**



***Corresponding Author:***
Sulis Sutiono,
Master of Information Technology
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing 20122, Medan, Indonesia.
Email : sulis671@guru.smk.belajar.id

***Article history:***
Received Oct 27, 2024
Revised Oct 29, 2024
Accepted Oct 31, 2024

## 1. INTRODUCTION

In this modern era, technological advancements have brought significant changes to various aspects of life, including education. According to Syarif & Astuti, technology offers numerous benefits that can enhance the quality of education and strengthen human resource development (Sitorus et al., 2024). Education is one of the main priorities in nation-building, where its quality greatly determines the future of the next generation (Noperia et al., 2023).

Information technology has become a driver of positive transformation, bringing innovations and efficiencies that benefit the global community (Priandika & Setiawansyah, 2023). One rapidly growing application of technology is data mining, which involves extracting valuable information from large datasets (Suhirman, 2022). In the context of education, data mining is used to analyze various aspects that can enhance learning effectiveness and student outcomes (Nofianti et al., 2023).

Classification, as one of the main techniques in data mining, plays a crucial role in grouping data based on specific labels or targets (Mukhsinin et al., 2024). One commonly used method in classification is the Decision Tree, which is an effective predictive modeling technique (Baiq Nurul Azmi et al., 2023). Among the various algorithms available, the C4.5 algorithm is well-known as a reliable tool for building decision trees in decision-making processes (Yulia Viska & Elisa, 2023).

Through this research, the author will analyze the graduation rates of students at SMK Negeri 1 Stabat in the general subjects group using the C4.5 algorithm. This study aims to provide deeper insights into student graduation as well as the application of data mining technology in the field of education.

## 2. RESEARCH METHOD

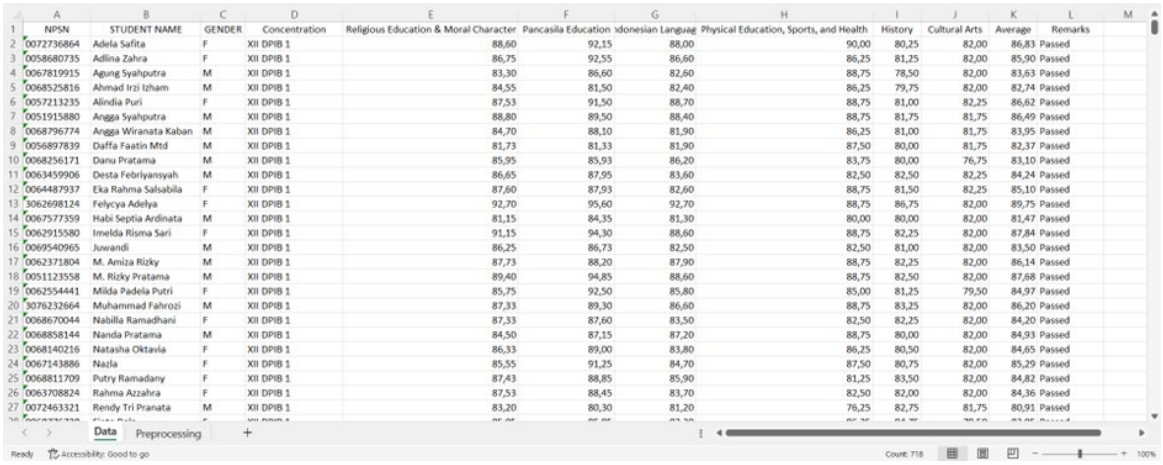
a. Graduation Classification: Based on the assessment results, if the average score is below 75, the student is declared not graduated; if the average score is 75 or above, the student is declared graduated.
b. Data Collection: The data collected for this study is secondary data. Rahmawati Wahyu Eka states that secondary data refers to information that is already available, sourced from literature or publications obtained through reading, studying, quoting, and summarizing information related to the topic being discussed (Putera et al., 2024). This secondary data was taken from the document

"List of Collected Diploma Scores for SMK Negeri 1 Stabat" for the 2023-2024 academic year, which includes a total of 717 students.

c. Preprocessing: The collected data underwent preprocessing. Data preprocessing is a crucial stage that prepares the data for further analysis (Sidiq et al., 2024). The main goal of preprocessing is to clean and prepare the data so that it is easier to analyze and yields more accurate results (Haq et al., 2024).
d. Modeling and Analysis: The processed data was modeled using RapidMiner, and the results of the modeling were subsequently analyzed. RapidMiner is one of the software tools for data mining (Indranu et al., 2024), and it also generates decision trees that can be viewed visually (Suriani et al., 2024).

## 3. RESULTS AND DISCUSSION

### A. Data Collection

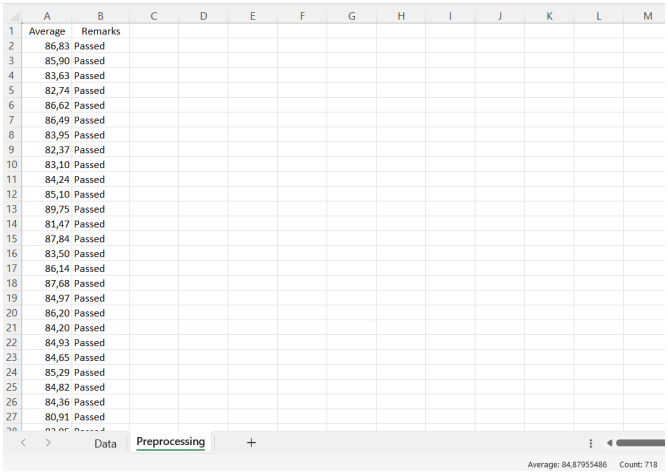


| NPSN | STUDENT NAME | GENDER | Concentration | Religious Education & Moral Character | Pancasila Education | Indonesian Language | Physical Education, Sports, and Health | History | Cultural Arts | Average | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0072736864 | Adela Safita | F | XII DPIB 1 | 88,60 | 92,15 | 88,00 | 90,00 | 80,25 | 82,00 | 86,83 | Passed |
| 0058680735 | Adlina Zahra | F | XII DPIB 1 | 86,75 | 92,55 | 86,60 | 86,25 | 81,25 | 82,00 | 85,90 | Passed |
| 0067819915 | Agung Syahputra | M | XII DPIB 1 | 83,30 | 86,60 | 82,60 | 88,75 | 78,50 | 82,00 | 83,63 | Passed |
| 0068525816 | Ahmad Irzi Izham | M | XII DPIB 1 | 84,55 | 81,50 | 82,40 | 86,25 | 79,75 | 82,00 | 82,74 | Passed |
| 0057213235 | Alindia Puri | F | XII DPIB 1 | 87,53 | 91,50 | 88,70 | 88,75 | 81,00 | 82,25 | 86,62 | Passed |
| 0051915880 | Angga Syahputra | M | XII DPIB 1 | 88,80 | 89,50 | 88,40 | 88,75 | 81,75 | 81,75 | 86,49 | Passed |
| 0068796774 | Angga Wiranata Kaban | M | XII DPIB 1 | 84,70 | 88,10 | 81,90 | 86,25 | 81,00 | 81,75 | 83,95 | Passed |
| 0056897839 | Daffa Faatin Mtd | M | XII DPIB 1 | 81,73 | 81,33 | 81,90 | 87,50 | 80,00 | 81,75 | 82,37 | Passed |
| 0068256171 | Danu Pratama | M | XII DPIB 1 | 85,95 | 85,93 | 86,20 | 83,75 | 80,00 | 76,75 | 83,10 | Passed |
| 0063459906 | Desta Febriyansyah | M | XII DPIB 1 | 86,65 | 87,95 | 83,60 | 82,50 | 82,50 | 82,25 | 84,24 | Passed |
| 0064487937 | Eka Rahma Salsabila | F | XII DPIB 1 | 87,60 | 87,93 | 82,60 | 88,75 | 81,50 | 82,25 | 85,10 | Passed |
| 3062698124 | Felycya Adelya | F | XII DPIB 1 | 92,70 | 95,60 | 92,70 | 88,75 | 86,75 | 82,00 | 89,75 | Passed |
| 0067577359 | Habi Septia Ardinata | M | XII DPIB 1 | 81,15 | 84,35 | 81,30 | 80,00 | 80,00 | 82,00 | 81,47 | Passed |
| 0062915580 | Imelda Risma Sari | F | XII DPIB 1 | 91,15 | 94,30 | 88,60 | 88,75 | 82,25 | 82,00 | 87,84 | Passed |
| 0069540965 | Juwandi | M | XII DPIB 1 | 86,25 | 86,73 | 82,50 | 82,50 | 81,00 | 82,00 | 83,50 | Passed |
| 0062371804 | M. Amiza Rizky | M | XII DPIB 1 | 87,73 | 88,20 | 87,90 | 88,75 | 82,25 | 82,00 | 86,14 | Passed |
| 0051123558 | M. Rizky Pratama | M | XII DPIB 1 | 89,40 | 94,85 | 88,60 | 88,75 | 82,50 | 82,00 | 87,68 | Passed |
| 0062554441 | Milda Padela Putri | F | XII DPIB 1 | 85,75 | 92,50 | 85,80 | 85,00 | 81,25 | 79,50 | 84,97 | Passed |
| 3076232664 | Muhammad Fahrozi | M | XII DPIB 1 | 87,33 | 89,30 | 86,60 | 88,75 | 83,25 | 82,00 | 86,20 | Passed |
| 0068670044 | Nabilla Ramadhani | F | XII DPIB 1 | 87,33 | 87,60 | 83,50 | 82,50 | 82,25 | 82,00 | 84,20 | Passed |
| 0068858144 | Nanda Pratama | M | XII DPIB 1 | 84,50 | 87,15 | 87,20 | 88,75 | 80,00 | 82,00 | 84,93 | Passed |
| 0068140216 | Natasha Oktavia | F | XII DPIB 1 | 86,33 | 89,00 | 83,80 | 86,25 | 80,50 | 82,00 | 84,65 | Passed |
| 0067143886 | Nazla | F | XII DPIB 1 | 85,55 | 91,25 | 84,70 | 87,50 | 80,75 | 82,00 | 85,29 | Passed |
| 0068811709 | Putry Ramadany | F | XII DPIB 1 | 87,43 | 88,85 | 85,90 | 81,25 | 83,50 | 82,00 | 84,82 | Passed |
| 0063708824 | Rahma Azzahra | F | XII DPIB 1 | 87,53 | 88,45 | 83,70 | 82,50 | 82,00 | 82,00 | 84,36 | Passed |
| 0072463321 | Rendy Tri Pranata | M | XII DPIB 1 | 83,20 | 80,30 | 81,20 | 76,25 | 82,75 | 81,75 | 80,91 | Passed |

Fig 1. Data Collection

In Figure 1, it can be seen that the data collection from the scores of students in the General Subjects group consists of 717 students, which includes NPSN, STUDENT NAME, GENDER, Concentration, Religious Education & Moral Character, Pancasila Education, Indonesian Language, Physical Education, Sports and Health, History, Cultural Arts, Average, and Remarks.

### B. Preprocessing



| Average | Remarks |
|---|---|
| 86,83 | Passed |
| 85,90 | Passed |
| 83,63 | Passed |
| 82,74 | Passed |
| 86,62 | Passed |
| 86,49 | Passed |
| 83,95 | Passed |
| 82,37 | Passed |
| 83,10 | Passed |
| 84,24 | Passed |
| 85,10 | Passed |
| 89,75 | Passed |
| 81,47 | Passed |
| 87,84 | Passed |
| 83,50 | Passed |
| 86,14 | Passed |
| 87,68 | Passed |
| 84,97 | Passed |
| 86,20 | Passed |
| 84,20 | Passed |
| 84,93 | Passed |
| 84,65 | Passed |
| 85,29 | Passed |
| 84,82 | Passed |
| 84,36 | Passed |
| 80,91 | Passed |

Fig 2. Preprocessing

In Figure 2, the data has undergone preprocessing, leaving the attributes of Average and Remarks.

### C. Modeling and Analysis

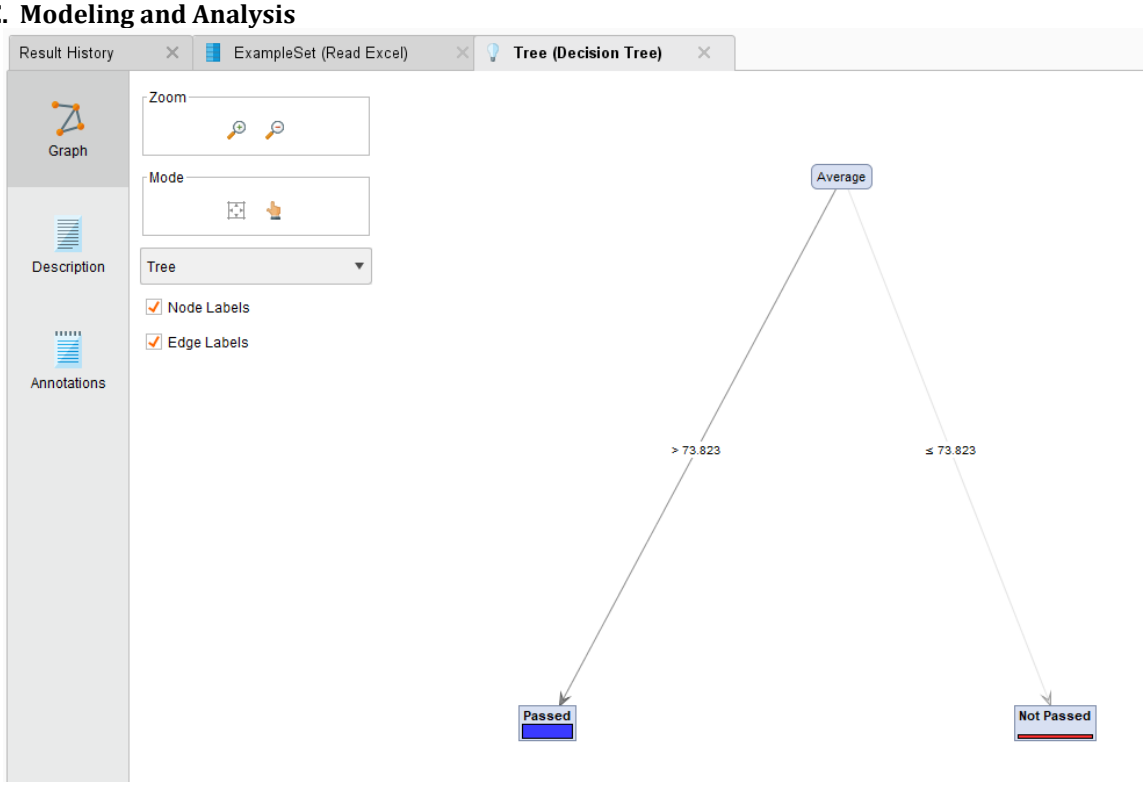



Fig 3. Modeling and Analysis

In this Tree modeling, the results of the evaluation based on the average scores of participants are illustrated. In the first branch, if the average score is greater than 73.823, the participant is considered "Passed." In the second branch, if the average score is less than or equal to 73.823, the participant is considered "Not Passed." Thus, this tree clearly shows the separation between the groups of participants who passed and those who did not, based on the established average criteria.

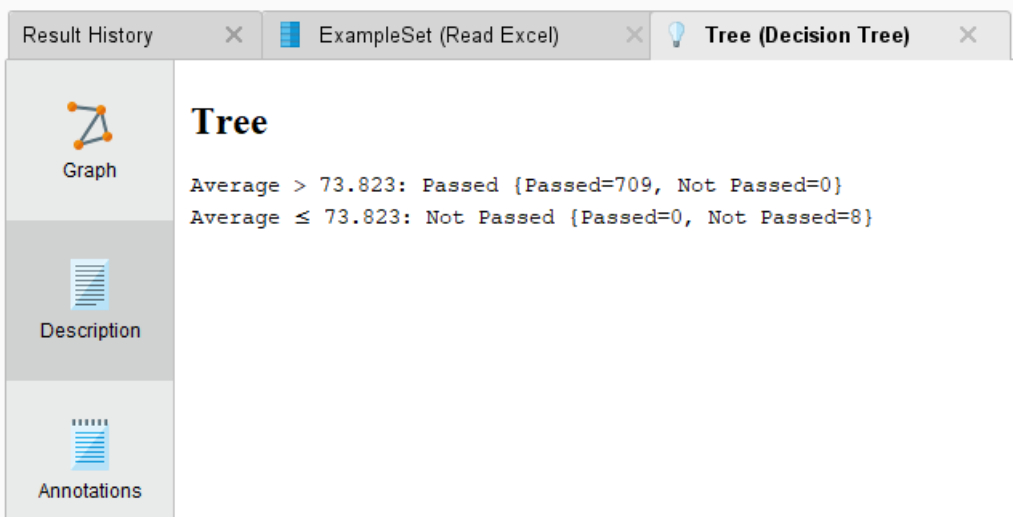



Fig 4. Description Tree

In this description of the tree, it explains that if the average is greater than 73.823, the status is 'Passed,' with a breakdown of 709 passing and 0 not passing. Conversely, if the average is less than or equal to 73.823, the status is 'Not Passed,' with a breakdown of 0 passing and 8 not passing.

## 4. CONCLUSION

This research analyzes the graduation rates of students at SMK Negeri 1 Stabat in the general subjects group using the C4.5 algorithm. Utilizing secondary data from 717 students, the researcher performed preprocessing to prepare the relevant data. The modeling results indicate that graduation is determined by the average score with a threshold of 73.823; of the 717 students, 709 are declared graduated, while 8 did not pass. The use of the C4.5 algorithm has proven effective in providing insights into student graduation, as well as generating decision tree visualizations that make the decision-making process more transparent and easier to understand. This study underscores the importance of applying data mining technology in education.

## REFERENCES


Baiq Nurul Azmi, Arief Hermawan, & Donny Avianto. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *4*(4), 281–290. https://doi.org/10.35746/jtim.v4i4.298

Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, *13*(1), 1079–1084. https://doi.org/10.33395/jmp.v13i1.13937

Indranu, F., Kesuma, M., & Kalifia, A. D. (2024). PENGARUH MEDIA SOSIAL TERHADAP TINGKAT ANXIETY PADA REMAJA : SEBUAH ANALISIS DENGAN RAPIDMINER. *Jurnal Ilmiah Multidisiplin*, *1*(3), 177–181. https://doi.org/10.62017/merdeka

Mukhsinin, D. A., Rafliansyah, M., Ibrahim, S. A., Rahmaddeni, R., & Wulandari, D. (2024). Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 570–579. https://doi.org/10.57152/malcom.v4i2.1255

Nofianti, A., Yawan, M. Y., & Nazar, M. A. (2023). Implementasi Data Mining dalam Pengolahan Data Transaksi Toko Sembako Menggunakan Algoritma Apriori (Studi Kasus : Toko Devan Mart). *G-Tech: Jurnal Teknologi Terapan*, *7*(1), 165–173. https://doi.org/10.33379/gtech.v7i1.1962

Noperia, M., Winda Sari, V., Studi Sistem Informasi, P., Triguna Dharma, S., & Penulis Korespondensi, E. (2023). Implementasi Data Mining Pengelompokan Data Nilai Untuk Menentukan Minat Belajar Seni Budaya. *JURNAL SISTEM INFORMASI TGD*, *2*(1), 65–72. https://ojs.trigunadharma.ac.id/index.php/jsi

Priandika, A. T., & Setiawansyah. (2023). Digitalisasi Aplikasi Keuangan Untuk Koperasi pada Dinas UMKM Provinsi Lampung. *Jurnal Abdimas Teknologi Informasi Dan Digitalisasi (JATI-DIG)Jati-Dig*, *1*(1), 17–23. https://doi.org/10.58602/jati-dig.v1i1.21

Putera, A., Siahaan, U., Jabar, A. A., Parhusip, N., Indrayani, M., & Barutu, S. (2024). Analysis of User Age Predictions in Public Satisfaction Surveys at Public Service Malls Using Decision Tree C4.5. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(2), 129–133. https://doi.org/10.61306/jitcse.v1i2

Sidiq, C. M., Faqih, A., & Dwilestari, G. (2024). ALGORITMA DECISION TREE C4.5 DIGUNAKAN UNTUK MENGKLASIFIKASIKAN DATA STROKE. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2). www.kaggle.com,

Sitorus, Z., Jabar, A. A., Indrayani, M., Barutu, S., & Situkkir, M. (2024). Analysis of User Satisfaction in the Development of ICT Learning Media Using AR-Based Assemblr Edu with the PIECES Framework Method. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(2), 110–120. https://doi.org/10.61306/jitcse.v1i2

Suhirman. (2022). Data WareHouse dan Mining. *GAES - PACE Book Publisher*, 14–25. https://digitalpress.gaes-edu.com/index.php/gaespace/index

Suriani, U., Palingga Ninditama, I., Syaputra, W., Studi Teknologi Rekayasa Multimedia, P., & Prasetya Mandiri, P. (2024). Pemodelan Prediktif Keterlambatan Bicara pada Balita Terkait dengan Penggunaan Smartphone Menggunakan Data Mining. *Journal of Information Technology Ampera*, *5*(1), 2774–2121. https://doi.org/10.51519/journalita.v5i1.589

Yulia Viska, I., & Elisa, E. (2023). PENERAPAN ALGORITMA C.45 DALAM MENGANALISA KEPUASAN PELANGGAN PADA PENJUALAN FASHION TRIFT. *JURNAL COMASIE*.