

---

## Implementation Of the Decision Tree Algorithm in Predicting Student Graduation Rate Based on Academic and Absence Data at A Vocational High School in Bandung

---

Tijo Ramadhan Haryo Utomo<sup>1</sup>, Asep Ririh Riswaya<sup>2</sup>, Toni Kusnandar<sup>3</sup>, En Tay<sup>4</sup>

STMIK Mardira Indonesia, Bandung, Indonesia<sup>1,2,3,4</sup>

Email: 21110258@stmik-mi.ac.id<sup>1</sup>, asep\_ririh@stmik-mi.ac.id<sup>2</sup>, tonikusnandar@stmik-mi.ac.id<sup>3</sup>, entay@stmik-mi.ac.id<sup>4</sup>

### ABSTRACT

*Education is a fundamental component of human resource development, directly impacting a nation's advancement. Graduation serves as a benchmark for the efficacy of this educational process. A vocational high school in Bandung is experiencing difficulties with pupils failing to graduate on schedule, resulting in low graduation rates. The reactive management of this condition complicates early intervention. This study advocates for the application of a Decision Tree algorithm as a predictive mechanism to enhance student graduation rates. The research employs a quantitative methodology, utilizing a predictive model grounded in academic data from students at a vocational high school in Bandung. The applied methodology is Prototype, which includes steps such as needs analysis, design, prototype model development, testing, assessment, and refining. In the data preparation step, missing values are imputed, features are derived, irrelevant columns are removed, and categorical variables are encoded. The dataset is subsequently split into 80% for training and 20% for testing. The Decision Tree Classifier model is constructed and trained utilizing the processed training data. Model performance evaluation uses metrics such as accuracy, precision, recall, and F1-score, along with confusion matrix analysis and feature importance scores. Test findings demonstrate that in four tests conducted under diverse conditions, the model attains above 90% accuracy in experiments 1, 2, and 3, but accuracy declines to approximately 80% in experiment 4. This indicates that, across varying conditions, the model consistently yields high forecasts and is suitable for deployment as an early warning system for potential graduation delays. This research demonstrates that machine learning can enhance the optimization of student graduation rates.*

**Keywords:** *Decision Tree, Graduation Prediction, Academic Data, Semester 5 Grades, Din Warningg*

### INTRODUCTION

Education is the fundamental basis for cultivating human resources, and optimal graduation rates serve as a critical metric for assessing the effectiveness of an educational institution. (Al-Giffary & Martanto, 2024) A vocational-focused institution struggles to improve graduation rates. According to data from the past four years, 179 of 900 students experienced graduation delays, often due to insufficient academic performance or attendance that failed

to meet established criteria. The core issue lies in the reactive methodology for assessing graduation status, in which the institution becomes aware of at-risk students only after a comprehensive review and a whole assembly. The delay in identification complicates the implementation of early preventive measures, underscoring the necessity for a data-driven technological strategy capable of swiftly and reliably forecasting student graduation rates. (Rasendriya et al., 2025)

This study advocates for the application of a Decision Tree algorithm as a predictive methodology. This technique is selected for its ability to generate comprehensible, visually interpretable decision tree models (rule sets) that are proficient at classifying both numerical and categorical academic data. The prediction model is constructed using academic data from a vocational high school in Bandung, including examination scores, student attendance, and instructional engagement metrics. The data utilized is confined to the fifth semester of the 2021/2022 academic year to the 2024/2025 academic year. The system development methodology utilizes a Prototype approach that prioritizes efficiency, iteration, and model improvement based on evaluative feedback. This study does not investigate non-academic characteristics, such as socio-economic background or learning motivation. (Muflihatul Hasanah & Zaehol Fatah, 2025)

Investigations into forecasting student graduation using Data Mining and Machine Learning methodologies have gained prominence as a means to enhance early warning systems within educational institutions. (Charbuty & Abdulazeez, 2021) This method uses previous academic data to identify patterns and risk factors affecting graduation rates. Among the several categorization algorithms, the Decision Tree—specifically the C4.5 variant—distinguishes itself through its efficacy and interpretability. It operates by partitioning datasets based on features that yield the highest Information Gain, and subsequently refines nodes until they achieve purity or homogeneity. The fundamental principles of the C4.5 Decision Tree, encompassing metrics such as Entropy and Information Gain, facilitate the identification of optimal decision-node qualities, yielding logical rules comprehensible to non-technical users. (Al Faruq et al., 2024; Saed & Jaharadak, 2022)

Numerous studies have confirmed the efficacy of Decision Trees in educational settings, emphasizing their ability to make precise predictions. (Suriani, 2023) evaluated the C4.5 algorithm against ID3 for graduation prediction, concluding that C4.5 had a notable accuracy of 86%. Likewise, both (Zhao, 2022) and (Nailil Amani et al., 2024) employed the C4.5 algorithm in their graduation prediction models, achieving notable accuracies of 93%. (Malhotra & Prakash Singh, 2021) found an accuracy of 86.13% in predicting high school student graduation with the same system. These findings jointly demonstrate Decision Trees' capability to serve as dependable tools for improving educational outcomes through predictive analytics.

While prior studies have shown commendable accuracy, most have focused on general data or university students. This study focuses on the application of Decision Trees to vocational high school students, utilizing precise data (Semester 5 Scores and Attendance) and employing an iterative Prototype technique. Thus, the resultant system functions as a cohesive Early Warning System. The model's performance will be assessed using industry-

standard measures, including Accuracy, Precision, Recall, and F1-Score, with a specific focus on Recall to minimize the number of at-risk kids overlooked in the forecasts.

This research has two quantifiable primary aims, grounded in this backdrop. The primary objective is to evaluate the Decision Tree method's precision in forecasting student graduation rates at a vocational high school in Bandung, using academic data. Secondly, it aims to elucidate the phases of proficiently executing the Decision Tree method for forecasting student graduation. The findings of this study are anticipated to enhance the existing scientific literature and serve as an Early Warning System, offering data-driven insights for the school to devise more effective and focused preventive measures.

## **METHOD**

This study utilizes a descriptive quantitative methodology through a case study at a vocational high school in Bandung. This research is selected because the examined data comprise both numerical (academic scores and attendance) and categorical (graduation status) variables, presented in figures and percentages to delineate the data's properties and the predictive model's efficacy. The paper primarily concentrates on predictive modeling with the Decision Tree Classifier Algorithm.

The research subject consists of the academic records of pupils at a vocational high school in Bandung. The utilized data is historical, including exam results for each subject, attendance records, student engagement, and graduation status as the class label. The dataset encompasses semester 5 students from the 2021/2022 academic year to the 2024/2025 school year. The data undergo preprocessing, including handling missing values, feature engineering, eliminating irrelevant columns, and encoding categorical variables. The dataset is subsequently split into 80% for training and 20% for testing.

The adopted system development methodology is the Prototype Method. This method is selected for its ability to enable rapid, adaptable iterations for model enhancements, allowing the development of an initial predictive model that can be evaluated promptly. The executed phases encompass the articulation of requirements, which examines the existing manual graduation assessment method to pinpoint the fundamental issues causing delayed risk identification for students. The design step encompasses developing a data model, selecting features, and specifying classification targets. The Decision Tree model is implemented using Python libraries on the Google Colaboratory platform during the prototype construction phase. The assessment and feedback phase entails testing the model on test data and analyzing performance using metrics such as Accuracy, Precision, Recall, and F1-Score.

## **RESULTS AND DISCUSSION**

Model Implementation Results, the execution of the graduation prediction model occurs in phases, utilizing the Decision Tree Algorithm alongside a Prototyping development methodology. This implementation employs the Google Colaboratory computing

environment in conjunction with the Python machine learning package (Scikit-learn) for classification modeling.

The preliminary phases involve preparing academic and attendance data for students at SMK Widya Dirgantara. The historical dataset has been sanitized, addressed for absent values, and encoded for categorical variables. The dataset is split into 80% for training and 20% for testing. The Decision Tree model has been effectively trained using the Gini Impurity (or Entropy) criterion. The model's training process is assessed through four distinct experimental scenarios to measure its stability and predictive performance across varying data conditions. The resulting decision tree shows that the Average Semester 5 Score and Attendance Rate are the primary determinants of graduation.

Model Evaluation, the evaluation is performed on the testing data to assess the model's predicted performance under each experimental circumstance. The comparison outcomes from the four tests indicate consistent, exceptionally high model performance, as outlined in Table 1.

Table 1. Comparison Test Results

| Experiment   | Accuracy        | Information                                                                 |
|--------------|-----------------|-----------------------------------------------------------------------------|
| Experiment 1 | > 90% (Highest) | Best model performance (able to classify up to 98% of test data correctly). |
| Experiment 2 | > 90%           | Consistent results above 90%.                                               |
| Experiment 3 | > 90%           | Consistent results above 90%.                                               |
| Experiment 4 | ~ 80%           | Accuracy has decreased but is still at a fairly good figure.                |

The model's performance has been excellent, particularly in Experiment 1, indicating that the Decision Tree can accurately classify student graduation status. A comprehensive assessment using the Confusion Matrix and Classification Report for the optimal model corroborates these findings, indicating remarkable proficiency in identifying 'Not Graduated' instances.

=== RINGKASAN PERBANDINGAN 4 EKSPERIMEN ===

|                                         | Akurasi Total | F1-Score (belum_lulus) | Recall (belum_lulus) | Precision (belum_lulus) | Support (belum_lulus) |
|-----------------------------------------|---------------|------------------------|----------------------|-------------------------|-----------------------|
| Eksperimen                              |               |                        |                      |                         |                       |
| Random Split (Eks 1)                    | 0.9837        | 0.9565                 | 0.9706               | 0.9429                  | 34                    |
| Stratified Split (Eks 2)                | 0.9620        | 0.9014                 | 0.8889               | 0.9143                  | 36                    |
| Manual 50/50 (Eks 3)                    | 0.9244        | 0.8941                 | 0.8444               | 0.9500                  | 90                    |
| Belum Lulus Lebih Banyak di Uji (Eks 4) | 0.8014        | 0.7521                 | 0.6111               | 0.9778                  | 144                   |

Figure 1. Evaluation Results

The consistent accuracy of over 90% across most experiments demonstrates the efficacy of the Decision Tree implementation for academic data and attendance at a vocational high school in Bandung. Despite a decline in accuracy in Experiment 4 (about 80%), this metric still indicates commendable and consistent performance.

The primary discourse centers on the elevated Recall value, particularly in Experiment 1, for the "Not Graduated" category. This elevated Recall indicates the model's

remarkable ability to identify almost all students at risk of failing to graduate, thereby significantly reducing False Negatives. This skill directly pertains to the research objective of the early identification of students at risk of not graduating. The analysis of the resultant decision tree substantiates the conclusion that Semester 5 Scores and Attendance are the paramount determining criteria. This discovery establishes a robust foundation for the school to focus interventions on pupils lacking these two traits. The Decision Tree methodology delivers predictable outcomes and articulates comprehensible logical rules to reinforce academic policies.

Implementation, this research's execution pertains to the technical application of the model and the strategic outcomes attained. The highly accurate prediction model with robust Recall is ideally suited to serve as an Early Warning System for the school.

The subsequent steps for implementation are as follows: The institution consistently updates students' academic and attendance records. Subsequently, the Decision Tree model is executed to produce predictions regarding graduation status (At Risk or Safe). The list of students identified as "At Risk of Not Graduating" is employed by guidance counselors to devise proactive strategies, including intensive tutoring, grade enhancement programs, or disciplinary actions, prior to graduation. Consequently, implementing this model is both technical and operational, facilitating the school's transition from a reactive to a proactive strategy to improve student graduation rates.

## **CONCLUSION AND RECOMMENDATION**

This research's execution pertains to the technical application of the model and the strategic outcomes attained. The highly accurate prediction model with robust Recall is ideally suited to serve as an Early Warning System for the school.

The subsequent steps for implementation are as follows: The institution consistently updates students' academic and attendance records. Subsequently, the Decision Tree model is executed to produce predictions regarding graduation status (At Risk or Safe). The list of students identified as "At Risk of Not Graduating" is employed by guidance counselors to devise proactive strategies, including intensive tutoring, grade enhancement programs, or disciplinary actions, prior to graduation. Consequently, implementing this model is both technical and operational, facilitating the school's transition from a reactive to a proactive strategy to improve student graduation rates.

Suggestions, in light of the research findings and recognized limitations, the following recommendations are proposed for implementation at a vocational high school in Bandung and for subsequent research endeavors:

The school is urged to employ strategic forecasting to develop more proactive intervention techniques. Students identified as "At Risk of Not Graduating" should be prioritized for follow-up via guidance counseling, timely reminders, or academic enhancement initiatives. This methodology can aid guidance counselors by obviating the necessity to manually or arbitrarily identify at-risk students. Secondly, to improve the model's accuracy and generalizability, it is recommended to incorporate supplementary non-academic predictor variables, such as interest and aptitude test outcomes, learning motivation levels, and socio-

economic status. This will facilitate the evaluation of the influence of external factors on graduation rates. Third, it is advisable to investigate alternative machine learning algorithms, such as Random Forest or XGBoost, as they may yield more consistent and precise forecasts than a solitary Decision Tree model. Fourth, implementing time-based validation using longitudinal data (spanning many academic years) would verify that the model's performance remains consistent and relevant despite changes in curriculum or student behavior. Fifth, digitizing the data recording system is vital, as the existing academic data is susceptible to human error and may lack centralization. It is recommended to digitize data and integrate it into the school's academic information system, enhancing accessibility, facilitating updates, and enabling automated analysis. Ultimately, predictive outcomes should be included in an automated reporting system (dashboard) or in notifications (by email/app) dispatched directly to guidance counselors and homeroom instructors. This will expedite the intervention procedure in accordance with the school's response.

## REFERENCES

- Al Faruq, U., Fauzi, M. A. N., Fatayasya, I., Daniati, E., & Ristyawan, A. (2024). Prediksi Data Kelulusan Mahasiswa Dengan Metode Decision Tree menggunakan Rapidminer. N Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi), 131–138.
- Al-Giffary, F. R., & Martanto, M. (2024). KLASIFIKASI KELULUSAN SISWA TAHUN 2024 MENGGUNAKAN METODE DECISION TREE (STUDI KASUS SMA ISLAM ALAZHAR 5 CIREBON). Jurnal Manajemen Informatika Jayakarta, 4(2), 195. <https://doi.org/10.52362/jmijayakarta.v4i2.1408>
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20–28.
- Malhotra, K., & Prakash Singh, A. (2021). Implementation of decision tree algorithm on FPGA devices. IAES International Journal of Artificial Intelligence (IJ-AI), 10(1), 131. <https://doi.org/10.11591/ijai.v10.i1.pp131-138>
- Muflihatul Hasanah, & Zaehol Fatah. (2025). Penerapan Decision Tree Pada Klasifikasi Kelulusan Mahasiswa. Jurnal Mahasiswa Teknik Informatika, 4(2), 225–230. <https://doi.org/10.35473/jamastika.v4i2.4487>
- Nailil Amani, N., Martanto, M., & Hayati, U. (2024). PENGGUNAAN ALGORITMA DECISION TREE UNTUK PREDIKSI PRESTASI SISWA DI SEKOLAH DASAR NEGERI 3 BAYALANGU KIDUL. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 473–479. <https://doi.org/10.36040/jati.v8i1.8355>
- Rasendriya, R., Fahrian, Marundrury, A. O., Jumadi, Y. L., Sumanto, & Kuswanto, A. D. (2025). KOMPARASI DECISION TREE, RANDOM FOREST, DAN K-NN MEMPREDIKSI KELULUSAN SISWA MENGGUNAKAN ORANGE. Jurnal Komputer Dan Teknologi, 4(2), 60–71. <https://doi.org/10.64626/jukomtek.v4i2.414>

Saed, A. A. A., & Jaharadak, A. A. (2022). Implementation with performance evaluation of decision tree classifier for uncertain data: Literature review. *International Journal of Multidisciplinary Research and Publications*, 5(5), 125–132.

Suriani, U. (2023). Penerapan data mining untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma decision tree C4. 5. *Journal of Computer and Information Systems Ampera*, 4(2), 55–65.

Zhao, J. (2022). Application of Decision Tree Algorithm in Teaching Quality Analysis of Physical Education. *Wireless Communications and Mobile Computing*, 2022(1). <https://doi.org/10.1155/2022/3556136>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC Attribution-NonCommercial-ShareAlike 4.0) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).