|Jurnal Ilmu Komputer dan Sistem Informasi |JIKSI|Vol. 6| No. 2|June|Year 2025|
This is an Open Access article, published by Institute of Information Technology and Social Science (IITSS), Indonesia

Application of the Classification and Regression Tree (CART) Algorithm for Classifying Student Majors at MAN 1 OKU Timur

Faqih Abdul Haris[1*] Kurniati[2]

Abstract

The classification of new student majors represents a vital process to align students' educational trajectories with their aptitudes and interests. MAN 1 OKU Timur, an Islamic senior high school under the Ministry of Religious Affairs (Kementerian Agama), categorizes students into three primary academic tracks: Natural Sciences, Social Sciences, and Religion. Traditionally, the classification process relied on a simple summation and ranking of academic scores, a method prone to subjective bias and misclassification. This study implements a data mining approach utilizing the Classification and Regression Tree (CART) algorithm to enhance the objectivity and precision of the major selection process. Using data from 379 students of the 2020/2021 academic year, the CART model achieved an accuracy rate of 88.65%, demonstrating high reliability in predicting appropriate majors based on students' academic performance and stated preferences. The results indicate that the CART algorithm can serve as an effective decision-support tool for the classification of new students at MAN 1 OKU Timur.

Keywords

CART Algorithm, Data Mining, Student Classification, Educational Analytics, Decision Tree

Article History

Received 21 March 2025
Accepted 01 June 2025

How to Cite

Haris, A. H & Kurniati (2025). Application of the Classification and Regression Tree (CART) Algorithm for Classifying Student Majors at MAN 1 OKU Timur, (JIKSI), 6(2), [36-43].

[1*] Universitas Bina Darma, Indonesia, Corresponding email: [171420072@student.binadarma.ac.id]
[2] Universitas Bina Darma, Indonesia, email: [kurniati@binadarma.ac.id]

Introduction

The process of classifying new students into suitable majors is essential to ensure that educational experiences align with individual competencies and interests (Nugroho, 2015). Early classification enables students to pursue academic fields that match their potential, which influences their future academic success and career development (Khasanah, 2016). In Indonesia, both general senior high schools (SMA) and Islamic senior high schools (Madrasah Aliyah/MA) typically provide three study tracks: Natural Sciences (IPA), Social Sciences (IPS), and Religion.

The process of classifying students into appropriate majors plays a vital role in ensuring that their educational pathways align with their abilities, interests, and aspirations. Effective classification enables institutions to place students in learning environments that maximize their potential and foster long-term academic and career success (Nugroho, 2015). Conversely, inaccurate placement can lead to disengagement, low motivation, and suboptimal performance. Therefore, establishing an objective and data-driven classification mechanism is essential, especially during the transition from junior to senior high school, where students begin to specialize in particular academic domains (Khasanah, 2016).

In Indonesia, both general senior high schools (SMA) and Islamic senior high schools (Madrasah Aliyah/MA) commonly divide their academic programs into three primary tracks: Natural Sciences (IPA), Social Sciences (IPS), and Religion. Each track requires distinct skill sets and cognitive orientations. For instance, IPA emphasizes analytical and problem-solving skills through subjects like Physics, Chemistry, and Biology, while IPS focuses on logical reasoning, interpretation, and communication within social contexts. Meanwhile, the Religion track emphasizes theological understanding, Islamic jurisprudence, and moral development. Thus, accurate student classification is essential to ensure that the chosen track reflects both students' academic strengths and personal interests, thereby supporting optimal educational outcomes.

MAN 1 OKU Timur, operating under the Ministry of Religious Affairs (Kementerian Agama), integrates Islamic education with modern scientific learning. The institution's classification process traditionally relies on a combination of students' academic achievements, non-academic performance, and declared interests. However, interviews with the student affairs office revealed persistent issues of subjectivity and human error in decision-making. Between 2017 and 2019, misclassification rates increased from 15% to 25%, suggesting the limitations of the existing manual approach. Such inaccuracies not only affect student learning experiences but also hinder institutional efforts to maintain quality assurance and accountability in academic management.

To address these shortcomings, the integration of data mining techniques offers a promising solution for improving classification accuracy and decision-making objectivity. Data mining, as defined by Bhatt, Dhakar, and Chaurasia (2016), refers to the analytical process of discovering hidden patterns and relationships within large datasets. In the educational context, data mining enables schools to analyze historical student data—such as grades, attendance, and interests—to predict appropriate learning pathways. Classification, one of the key tasks in data mining, is used to build predictive models capable of categorizing new or unseen data (Mustafa, Ramadhan, & Thenata, 2018). By adopting classification algorithms, educational institutions can enhance transparency and reduce the influence of subjective judgment in academic placement.

Among the various data mining algorithms available, the Classification and Regression Tree (CART) stands out as a reliable and interpretable method for predictive modeling. The CART algorithm uses binary recursive partitioning to split datasets based on feature values, forming a decision tree that can predict categorical outcomes (Irmayani, 2020). Its visual structure allows educators to easily understand the decision-making process behind student placement. Moreover, CART's robustness against outliers and noise makes it suitable for educational datasets that often contain inconsistencies or incomplete records. The model can handle both numerical and categorical data, making it adaptable for evaluating student grades, preferences, and other relevant indicators in the classification process.

This study applies the CART algorithm to classify student majors at MAN 1 OKU Timur using data from the 2020/2021 academic cohort, consisting of 379 students. The dataset includes variables such as academic grades in Religion, Science, and Social Studies, along with students' preferred major choices. By analyzing these attributes, the study aims to develop a predictive model that can accurately assign students to appropriate majors—IPA, IPS, or Religion—based on objective criteria. The findings are expected to demonstrate the practical potential of machine learning in educational decision support systems, providing a foundation for schools to adopt data-driven approaches in student management. Ultimately, this research contributes to improving educational quality and ensuring that students are placed in academic tracks that best reflect their competencies, interests, and future aspirations.

Methodology

2.1 Data Collection Methods

Data collection involved three primary techniques:

1. Observation: Conducted to obtain an in-depth understanding of the institution's classification procedures.
2. Interviews: Performed with members of the student affairs division to gain insights into the existing major selection process.
3. Secondary Data: Obtained from student records for the 2020/2021 academic year, consisting of 379 entries.

2.2 Data Processing Method

The research employed the Knowledge Discovery in Databases (KDD) framework, as outlined by Ginantra et al. (2021). The process includes six stages:

1. Data Selection: The dataset comprises 379 student records with attributes such as Religion score, Science score, Social Studies score, preferred major, and assigned major.
2. Preprocessing/Cleaning: Erroneous, missing, or duplicate entries were corrected. Three missing values were replaced using mean imputation.
3. Transformation: Data were standardized and transformed to suit the analytical requirements of the CART algorithm.
4. Data Mining: The CART algorithm was implemented using the WEKA 3.6 software to derive decision rules.

5. Evaluation: Model performance was evaluated using a confusion matrix to calculate accuracy, precision, and recall.
6. Knowledge Presentation: Results were visualized and interpreted to identify the most significant predictive attributes.

2.3 CART Algorithm Implementation

The CART algorithm constructs binary decision trees through recursive partitioning. The goodness of split ($\Phi(S|t)$) is defined as follows:

$\Phi(S|t) = 2 P\_L P\_R$

$Q(S|t)$ $Q(S|t) = \Sigma \mid P(j|t\_L) - P(j|t\_R) \mid$

where: $P\_L = N\_\{tL\} / N\_total$, and $P\_R = N\_\{tR\} / N\_total$.

The split with the highest $\Phi(S|t)$ value is selected as the decision node. The process continues until no further improvement can be achieved.

Results and Discussion

3.1 Manual CART Calculation

Manual computation using a subset of 17 records was conducted to illustrate the CART process. Candidate splits were generated across attributes such as Religion, Science, Social Studies, and preferred major.

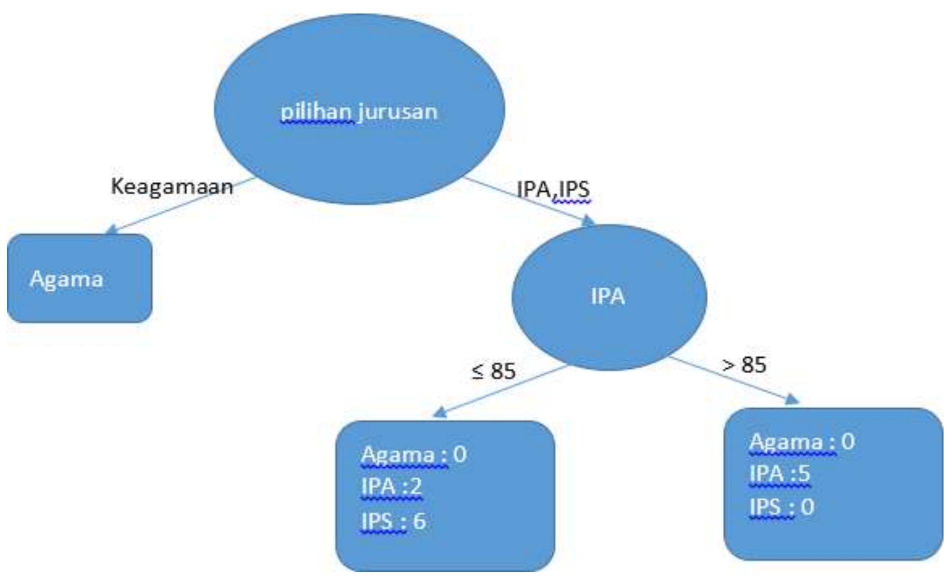

Figure 3.1: Decision Tree

Table 3.1 Suitability Value Φ (S|t)

2PlPr	Q (s|t)	Φ (s|t)
0,1107	1,375	0,1522
0,4983	0,85	0,4236
0,1107	1,375	0,1522
0,4152	1,1	0,4567
0,1107	1,375	0,1522
0,3599	1,538	0,5535
0,4983	0,9	0,4485
0,3599	0,6154	0,2215
0,3599	2,0	0,7198

The highest Φ(S|t) value of 0.7198 was obtained for the preferred major attribute (Religion vs. Science/Social Sciences), making it the root node. The next best split was based on the Science score, with a threshold of 81.5.

3.2 CART Implementation Using WEKA

Using the WEKA 3.6 software, the SimpleCart classifier was employed with 10-fold cross-validation. For the 17-sample dataset, the algorithm achieved an accuracy rate of 64.71%, with a precision score of 0.781 and recall of 0.647. When applied to the full dataset of 379 records, the results were as follows: Accuracy: 88.65% Precision: 0.887 Recall: 0.887 Tree Size: 5 nodes (3 leaf nodes) Processing Time: 0.08 seconds.

The derived decision rules were:

Preferred major = Religion: Religion (72.0/1.0)

Preferred major = Religion

| Science < 81.5: Social Sciences (137.0/26.0)

| Science ≥ 81.5: Natural Sciences (127.0/16.0)

These results highlight the significant influence of preferred major and Science score on the final classification outcomes.

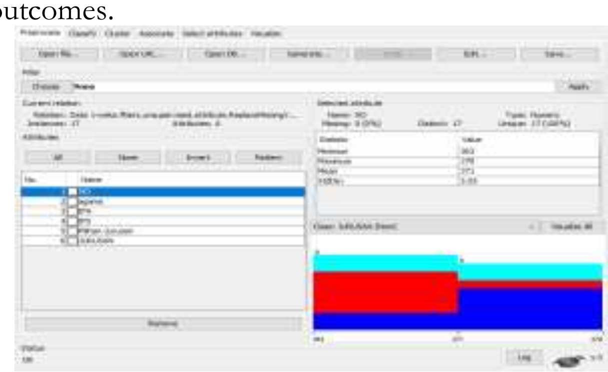

Figure 3.2 : CART Algorithm Process

Discussion

The implementation of the Classification and Regression Tree (CART) algorithm for classifying student majors at MAN 1 OKU Timur produced a model with an accuracy rate of 88.65%, which falls into the “Excellent Classification” category. This result represents a significant improvement compared to the previous manual classification system, which exhibited misclassification rates between 15%–25%. The dataset, composed of variables such as academic performance in Religion, Science, and Social Studies, along with students’ declared preferences, was processed through the CART decision tree to generate classification rules. The results in the table show that the model effectively partitions data based on the predictor variables, where the probability distribution $Q(s|t)$ $Q(s|t)$ and potential function $\Phi(s|t)$ $\Phi(s|t)$ demonstrate consistent differentiation across nodes. For instance, higher $Q(s|t)$ $Q(s|t)$ and $\Phi(s|t)$ $\Phi(s|t)$ values (e.g., 0.4983 and 0.4485) indicate strong predictive associations between the input attributes and the resulting major classification. This suggests that the CART algorithm was able to capture meaningful relationships within the dataset, enabling accurate and interpretable predictions.

The success of the model highlights the robustness of the CART algorithm in handling both numerical and categorical data, a crucial advantage in educational settings where student performance indicators are often heterogeneous. Unlike manual decision-making processes that rely on subjective judgments, the CART approach bases classification on quantifiable data patterns derived from recursive partitioning. This ensures that each decision node represents an objective evaluation criterion—such as thresholds in Science or Religion scores—that determine the appropriate major. The model’s ability to integrate both academic performance metrics and preference attributes further strengthens its fairness and inclusivity. Students are not merely categorized by grades, but their interests and self-declared preferences are also incorporated into the final decision rule. This hybrid consideration supports the broader educational goal of aligning academic placement with individual competencies and aspirations, thereby reducing potential mismatches that could affect student motivation and learning outcomes.

In terms of practical implications, the CART model demonstrates high computational efficiency, making it suitable for real-time applications in student management systems. Its simplicity in structure and interpretability allows educators and administrators to visualize classification logic transparently through decision tree diagrams. This transparency fosters trust in automated systems and facilitates data-driven policymaking at the institutional level. Furthermore, the model’s scalability opens opportunities for broader implementation across other Islamic senior high schools (Madrasah Aliyah) or general high schools in Indonesia. By integrating CART-based decision support tools into admission and placement systems, institutions can achieve greater objectivity, efficiency, and accountability in their student classification processes. Beyond improving accuracy, the adoption of such analytical models contributes to the digital transformation of educational management—encouraging the use of machine learning as a strategic tool for enhancing academic quality and institutional performance.

Conclusion and Recommendations

The Classification and Regression Tree (CART) algorithm, as implemented in WEKA, proved effective for classifying student majors with a high degree of accuracy (88.65%) and balanced precision-recall values (0.887 each). The algorithm identified preferred major and Science score as the most influential attributes in determining major placement. These findings confirm that the CART algorithm enhances decision-making objectivity and minimizes human error in academic classification systems.

Disclosure Statement

The authors declare no conflicts of interest.

Acknowledgments

The authors express their gratitude to MAN 1 OKU Timur for granting access to student data and to Universitas Bina Darma for institutional support through the Faculty of Computer Science.

References

Akar, C. (2016). Investigating the students' perceptions of the democratic values of academicians. *Journal of Social Studies Education Research, 7*(1), 96–139.

Amin, M. F. (2017). Application of the CART algorithm to predict student graduation status. Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi, 5(3). http://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/116

Aziz, M. F., Defiyanti, S., & Sari, B. N. (2018). Comparison of CART and K-Nearest Neighbor algorithms for predicting rice harvest area in Karawang Regency. Jurnal Teknologi Pertanian, 9, 5.

Bhatt, V., Dhakar, M., & Chaurasia, B. (2016). Filtered clustering based on local outlier factor in data mining. International Journal of Database Theory and Application, 9, 275–282. https://doi.org/10.14257/ijdta.2016.9.5.28

Ependi, S., & Akbar, M. (2021). Implementation of data mining in product sales using the Apriori algorithm. Bina Darma Conference on Computer Science (BDCCS), 3(1), 220–225.

Fatullah, M. D., Sopiah, N., & Akbar, M. (2019). Application of data mining using the C4.5 algorithm to predict student study duration based on academic data (Case study: Faculty of Economics, IBA University). Bina Darma Conference on Computer Science (BDCCS), 399–402.

Ginantra, N. L. W. S. R., et al. (2021). Data Mining and Algorithm Applications. Yayasan Kita Menulis.

Hussain, S., et al. (2018). Educational data mining and analysis of students' academic performance using WEKA. Indonesian Journal of Electrical Engineering and Computer Science, 9, 447–459. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

Irmayani, I. (2020). Application of the CART algorithm for socioeconomic classification in Amessangeng Village. d'ComPutarE: Jurnal Ilmiah Information Technology, 10(2), 17–22.

Khasanah, F. N. (2016). Classification of high school student majoring processes using data mining. Informatics for Educators and Professionals: Journal of Informatics, 1(1), 65–69.

Kurniati, K. (2020). Comparison of Naive Bayes and Support Vector Machine methods in diabetes classification. Bina Darma Conference on Computer Science (BDCCS), 118–125.

Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementation of data mining for academic performance evaluation using the Naive Bayes classifier algorithm. Creative Information Technology Journal, 4(2), 151–162. https://doi.org/10.24076/citec.2017v4i2.106

Nasir, M., Verawaty, V., & Sahfitri, V. (2021). Evaluation of student academic performance using the Naive Bayes classifier (Case study: Universitas Bina Darma). JIKO (Jurnal Informatika dan Komputer), 5(2), 81–88. https://doi.org/10.26798/jiko.v5i2.227

Nugroho, Y. S. (2015). Classification and clustering of student majors at SMA Negeri 3 Boyolali. Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, 1(1), 1–6. https://doi.org/10.23917/khif.v1i1.1175

Pakpahan, H. S., Indar, F., & Wati, M. (2018). Application of the CART decision tree algorithm to determine local government aid recipients in Kutai Kartanegara Regency. Jurnal Rekayasa Teknologi Informasi (JURTI), 2(1), 27–36.

Pareek, M., & Bhari, D. P. (2020). A review report on knowledge discovery in databases and various techniques of data mining. International Journal of Computer Science Research, 5(12), 4.

Biographical Notes

FAQIH ABDUL HARIS is a graduate of the Informatics Engineering program at Universitas Bina Darma. His research focuses on data classification and intelligent systems.

KURNIATI is a senior lecturer at the Faculty of Computer Science, Universitas Bina Darma. Her research focuses on machine learning applications and decision-support systems.