WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis of Student Academic Performance Prediction Using Decision Tree Algorithms

Yin Nyein Aye

University of Computer Studies (Taunggyi), Taunggyi, Myanmar
yinnyeinaye@ucstgi.edu.mm,

A B S T R A C T

Student performance prediction in the classroom is a valuable tool that helps educators identify learners who may need additional support. Using data such as attendance, homework completion, quiz scores, participation, and behavioral patterns, machine learning models can analyze trends to predict future academic outcomes. Predicting student performance in the classroom has become increasingly important in the modern education system. With advancements in educational data mining and machine learning, educators and institutions can now use real-time data to forecast how well students are likely to perform in upcoming assessments or overall academic progress. The purpose of this study is to compare analysis of predictive model for student academic performance using machine learning decision tree algorithms (ID3, C4.5 and CART). This approach not only improves academic achievement but also supports holistic student development across academic, behavioral, and social dimensions.

Keywords: Student Performance Prediction, Decision Tree, ID3C4.5,CART

1. Introduction

In a traditional classroom setting, performance is often measured through test scores, assignments, and final grades. However, these metrics are reactive and only highlight issues after they've occurred. By contrast, a student performance prediction system aims to be proactive, giving educators insights into student behavior, learning trends, and potential academic struggles before they manifest as poor results. To build an effective prediction model, there are multiple factors that influence student performance. These include are academic records, attendance, involvement, behavioral data, socioeconomic factors, learning style and pace and use of LMS or Digital Tools. . These predictions help in early identification of at-risk students, enabling timely interventions and support. This paper aims to many educators know how many students are need for help such as struggling students for timely support, tailoring instruction to individual needs and more enhancing academic achievement and reduces dropout rates.

Predicting student performance in classroom management is a growing area of interest in educational research and data science. It combines insights from pedagogy, psychology, and technology to help educators create more effective learning environments. Student performance prediction needs for identifying struggling students early allows for timely support, tailoring instruction to individual needs boosts engagement and outcomes, helping administrators and enhancing academic achievement and decreases student failure rates. Table1 shows the multiple factors of student performance in the classroom.

1.1 Importance of Academic Performance Prediction

In recent years, educational institutions have increasingly adopted data-driven approaches to improve student outcomes. Predictive models offer a proactive way to identify students at risk of underperforming, enabling timely interventions and personalized support. This paper presents a Decision Tree-based model using the ID3 ,C4.5 and CART algorithms to classify student performance and explore key influencing factors.

1.2 Challenges in Traditional Evaluation of Student Performance

Traditional methods of evaluating student performance—such as standardized tests, report cards, and teacher assessments—have long been the cornerstone of academic measurement. However, these approaches face several limitations that hinder their effectiveness in identifying and supporting students at risk. These are limited scope, delayed feedback, subjective and bias, one-size fits all testing, lack of predictive insight and data fragmentation.

1.3 Predictive Analytics Addresses Traditional Evaluation Challenges

Predictive analytics leverages data science and machine learning to forecast student outcomes based on historical and real-time data. By identifying patterns and correlations across diverse variables, it offers a proactive and personalized approach to academic support.

2. Comparative Analysis of ID3, C4.5 and CART

ID3 (Iterative Dichotomiser 3) is a decision tree algorithm that uses information gain to select the best feature to split the data at each node. It builds the tree recursively until all data is classified or no further splits are possible. Comparing ID3, C4.5, and CART helps clarify which decision tree algorithm best suits your data and goals. ID3 uses Information Gain, which can favor features with many values.C4.5 improves on this by using Gain Ratio, which normalizes Information Gain to reduce bias. CART uses the Gini Index for classification and Mean Squared Error (MSE) for regression. C4.5 generally offers better accuracy and flexibility than ID3.CART is preferred for binary classification and regression tasks. The choice depends on the dataset characteristics and the need for interpretability vs. predictive power. A dataset includes continuous values or missing entries, C4.5 or CART would be more suitable than ID3. Comparative analysis of ID3, C4.5 and CART decision tree algorithms are shown in Table1.

Table 1–Comparative analysis of ID3, C4.5 and CART

Algorithm	Splitting Criterion	Output Type	Handles Missing Values	Pruning
ID3 (Iterative Dichotomiser 3)	Information Gain	Categorical	✗ No	✗ No
C4.5(Successor to ID3)	Gain Ratio	Categorical	☑ Yes	☑ Yes
CART (Classification and Regression Trees)	Gini Index (Classification) / MSE (Regression)	Categorical & Numerical	☑ Yes	☑ Yes

Decision tree models will help create adaptive learning paths tailored to each student's strengths and weaknesses. Governments and institutions may use predictive insights to shape curriculum, allocate resources, and design support programs. Integration with digital learning platforms could allow continuous monitoring and dynamic adjustment of teaching strategies.

3. Appropriateness for Student Performance Prediction Model

Decision trees offer clear, rule-based insights that educators can easily understand and act upon. They can be applied to large datasets across institutions, making them suitable for national or regional education systems. Future systems may combine decision trees with other machine learning techniques (e.g., ensemble methods or neural networks) for improved accuracy and adaptability. Incomplete or biased data can affect prediction accuracy. Ensuring clean, representative datasets is crucial. Without proper pruning or validation, decision trees may overfit training data, reducing generalizability. Predictive models must be used responsibly to avoid labeling or limiting students based on algorithmic predictions.

4. Conclusion

ID3 is simple and good for small, clean datasets with categorical features.C4.5 is more robust and versatile, especially with pruning and continuous data. CART is powerful for both classification and regression, and handles binary splits well. The best algorithm for specific needs- based on classifying students' performance categories. The paper aims to compare the analysis of ID3,C4.5 and CART algorithms for student academic performance prediction. The goal is to identify which algorithm provides better predictive accuracy and reliability for the type of dataset

References

1.Jain, L., Behera, H., Mandal, J., Mohapatra, D. (eds) (2015) Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest Computationa Intelligence in Data Mining - Volume 1. Smart Innovation, Systems and Technologies, vol 31. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2205-7_51

2.Songul Cinaroglu (2016) Comparison of Performance of Decision Tree Algorithms and Random Forest: An Application on OECD Countries Health Expenditures International Journal of Computer Applications (0975 – 8887) Volume 138 – No.1

3.Dede Kurniadi (2023) Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification