E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 02-01-2025
Accepted: 09-02-2025

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 3; Issue 2; 2025; Page No. 194-202

Student performance assessment using machine learning with decision tree

[1]Mohanakannan N and [2]Dr. R Padma

[1]M.Sc., Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India

DOI: https://doi.org/10.5281/zenodo.15575963

Corresponding Author: Mohanakannan N

Abstract

In the age of online learning, assessing student performance has emerged as a vital element for enhancing educational results and guaranteeing academic achievement. This project introduces an AI-driven analytical system that assesses student performance by examining assessment scores and submitted assignment data with the Decision Tree algorithm. The suggested system gathers student information including quiz/test results, assignment submission status, punctuality, and submission quality to create meaningful predictions regarding academic achievement. The Decision Tree algorithm is utilized for its clarity, precision, and capability to manage both categorical and numerical information. The system categorizes students into performance groups like High Performer, Average Performer, and Low Performer, utilizing training data and decision guidelines. It also pinpoints significant factors affecting student achievement or lack thereof, including regular late submissions or persistently low grades in particular subjects. By mapping out decision routes and results, educators can make informed choices like providing remedial classes or tailored learning pathways to improve student engagement and success. This system assists teachers in monitoring performance and facilitates early intervention for at-risk students, rendering it an essential asset in contemporary educational analytics.

Keywords: Student performance analysis, decision tree algorithm, educational data mining, assessment evaluation, assignment analysis, predictive analytics, academic performance prediction, student categorization, learning outcomes, AI in education

Introduction

In the current data-centric educational landscape, the necessity to track and improve student performance has grown significantly. Conventional approaches to assessing student performance, like manual grade monitoring and subjective evaluation, frequently do not yield precise insights into a student's development and aspects needing enhancement. The incorporation of technology into education produces large quantities of data from tests, assignment submissions, and classroom activities, which can be utilized to identify significant patterns and enhance academic results.

This initiative suggests a smart system that employs Decision Tree classification to evaluate student performance using two main inputs: assessment scores and submitted assignments. The system considers several elements, such as the timeliness of submissions, the quality of content, and the scores achieved in different assessments. Utilizing the Decision Tree algorithm, students are classified into performance categories like high, average, and low achievers. The model offers understandable decision pathways, aiding educators in grasping the rationale for a student's classification.

The aim of the system is to empower institutions and educators to make well-informed choices regarding intervention strategies, custom learning plans, and the distribution of resources. Additionally, it offers a proactive alert system to recognize students who may be struggling, facilitating prompt assistance to enhance their educational experience. This method connects raw academic information with practical educational insights, ultimately fostering a more efficient and student-focused learning atmosphere.

Objectives

- To develop a student performance analysis system that uses academic data such as assessment scores and assignment submissions for evaluation.

- To implement the Decision Tree algorithm for classifying students based on performance levels (e.g., High, Average, Low).
- To identify key performance indicators such as timely submission, assignment quality, and assessment consistency that influence student outcomes.
- To provide visual insights (decision paths and classification results) for easy interpretation by educators and academic counselors.
- To enable early intervention by flagging students who are at risk of academic underperformance.
- To support personalized learning by recommending improvement strategies based on performance trends.
- To create a data-driven academic environment where educational decisions are guided by predictive analytics and actionable insights.

This project concentrates on examining student achievement by utilizing data from assessments and assignment submissions. Conventional assessment approaches frequently fail to recognize at-risk learners or provide tailored assistance. Utilizing the Decision Tree algorithm, the system categorizes students according to their academic performance and uncovers essential factors affecting their outcomes. This data-informed strategy facilitates early intervention, customized learning, and enhanced academic assistance, promoting improved results and a student-focused educational atmosphere.

Literature Survey

Jindal, R., & Borah, S., "A Decision Tree Based Student Performance Prediction Model," International Journal of Computer Applications, 2020 [1]. This research presents a student achievement prediction system that utilizes the Decision Tree algorithm to categorize students into various performance categories. The model is developed with historical educational data, encompassing grades on assignments, exam results, and attendance records. The writers emphasize creating understandable rules to assist educators in comprehending the reasoning behind each classification. The findings demonstrate significant precision and offer practical insights into students' learning behaviors.

Ramesh, V., Parkavi, S., & Umarani, R., "Predicting Student Performance Using Data Mining Techniques," International Journal of Computer Science and Information Technologies, 2019 [2]. This document examines different data mining techniques, including Decision Tree, Naive Bayes, and K-Nearest Neighbors (KNN), for forecasting academic success. Of these, Decision Tree produced the most favorable outcomes regarding interpretability and precision. The dataset contains internal examination scores, assignment entries, and records of extracurricular activities. The research concludes that early forecasting models may assist in implementing prompt intervention strategies.

Pandey, U. K., & Pal, S., "Data Mining: A Prediction of Performer or Underperformer Using Classification," International Journal of Computer Science and Information Security, 2018 [3]. This paper examines different data mining techniques, including Decision Tree, Naive Bayes, and KNearest Neighbors (KNN), to forecast academic performance. Of these, Decision Tree produced the most favorable outcomes regarding both interpretability and accuracy. The dataset contains scores from internal exams, records of assignment submissions, and information on extracurricular activities. The research indicates that early prediction models may aid in facilitating prompt intervention strategies.

Kaur, A., & Singh, B., "Mining Education Data to Predict Student's Performance Using Classification Methods," International Journal of Advanced Computer Science and Applications, 2021 [4]. This study offers a comparative evaluation of classification techniques in forecasting student results utilizing academic data. The Decision Tree algorithm was especially successful in modeling the data because of its hierarchical structure and capacity to manage both categorical and numerical variables. The study shows the real-world application of data mining in the education field for aiding decision-making.

Cortez, P., & Silva, A. M. G., "Using Data Mining to Predict Secondary School Student Performance," University of Minho, 2017 [5]. Cortez and Silva utilize decision trees, support vector machines, and neural networks on practical student performance datasets. The characteristics consist of assignment scores, exam marks, study duration, and family history. Decision Trees were notable for their clarity and rapid processing. The research highlights the significance of comprehensive academic profiling for accurate forecasting.

Mubarak, A. A., & Petra, S., "Analyzing Student Academic Performance Using Decision Tree Algorithms," Procedia Computer Science, 2020. The authors introduce a system that employs Decision Trees to analyze student data gathered from LMS platforms, such as assignment scores, submission trends, and evaluation outcomes. The model offers educators visual decision guidelines that assist in student categorization and academic advising. Their system was effective in identifying underperforming students at the beginning of the term.

Yadav, S. K., & Pal, S., "Data Mining Applications: A Comparative Study for Predicting Student's Performance," International Journal of Innovative Technology and Exploring Engineering, 2019 [7]. In this study, the authors evaluate different data mining methods, emphasizing the Decision Tree model for forecasting academic achievement. The input characteristics comprised attendance, exam scores, and assignment regularity. The findings indicate that the Decision Tree provides excellent classification precision and facilitates adaptive learning techniques.

Bharadwaj, B. K., & Pal, S., "Mining Educational Data to Analyze Students' Performance," International Journal of Advanced Computer Science and Applications, 2018 [8]. This research utilizes the ID3 Decision Tree algorithm on datasets regarding student performance. It centers on comprehending how internal elements like task timeliness and engagement influence final academic scores. The Decision Tree model generated a series of guidelines to assist institutions in improving their academic support systems. The writers propose incorporating these models into LMS systems for immediate analysis.

Nandhini, M., & Valli, N., "Performance Evaluation of Students Using Classification Data Mining Techniques," International Journal of Computer Science and Engineering, 2020 [9]. This study explores classification techniques, such as Decision Trees, to evaluate students' academic performance. It includes a collection of data featuring test scores, submitted assignments, and feedback records. The model provides a strong framework for forecasting results and has undergone testing in actual classroom settings, delivering precise and timely information for educational leaders.

Zafra, A., & Ventura, S., "Predicting Student Grades in Learning Management Systems with Multiple Instance Learning," Educational Data Mining Conference, 2017. While concentrating on multiple instance learning, this paper also features a comparison with Decision Tree classifiers for forecasting student grades according to LMS activity. Decision Trees proved useful in representing learning patterns based on the duration spent on tasks, timing of submissions, and frequency of interactions. The research emphasizes the advantage of interpretability in practical educational systems for implementable interventions

Limitations Analysed

Restricted Feature Range

The system predominantly depends on assessment scores and assignment submission information, overlooking other significant aspects like student behavior, attendance, socio-economic status, or involvement in extracurricular activities that could impact performance.

Static Dataset Dependence

The accuracy of the model largely relies on the quality and amount of historical academic data available. If the dataset is old or lacking, the predictions might not represent the current trends in learning.

Risk of Overfitting in Decision Trees

Decision Tree algorithms are susceptible to overfitting, particularly when the dataset is limited or contains noise. This could result in inadequate generalization when used on new or previously unseen data.

Lack of Immediate Feedback System

The system reviews performance after submission and evaluation, missing a real-time monitoring feature that could offer immediate feedback or notifications during active courses.

Ineffectiveness in Managing Incomplete or Missing Data Efficiently

When students fail to submit their assignments or assessments, the system might incorrectly classify them because of lacking information, thus diminishing reliability in these instances.

Trade-off between Model Interpretability and Complexity: Although Decision Trees are easy to understand, they might not identify intricate patterns in data as well as more sophisticated models such as Random Forests or Neural Networks.

Absence of Customization

The model considers all students uniformly based on numerical data, neglecting personalized learning preferences or unique challenges.

Lack of Integration with Learning Platforms

The system functions autonomously and is not connected to Learning Management Systems (LMS) or online classroom platforms for automatic data gathering and immediate updates.

Single Algorithm Reliance

The document centers exclusively on classification using Decision Trees. A mixed or comparative method alongside other algorithms might yield improved understanding and precision.

Ethical Issues and Privacy Considerations

Gathering and handling academic performance information may trigger worries about student confidentiality and data protection, issues that are not discussed in the paper.

The examined literature emphasizes the growing application of data mining and machine learning methods particularly Decision Tree algorithms for forecasting and evaluating student performance in educational environments. In various studies, researchers have utilized factors such as assignment grades, exam scores, attendance records, and LMS activity logs to categorize students into performance groups such as high, average, or low achievers. Decision Trees are often preferred for their ease of use, clarity, and rule-based framework, enabling educators to easily grasp the reasoning behind the predictions. Multiple studies highlight the significance of early prediction models for facilitating prompt intervention and customized academic assistance. In many research studies, Decision Trees surpass conventional statistical techniques in classification precision and are favored in situations where clear and practical insights are essential. Nonetheless, the surveys also highlight shortcomings in current systems, including incomplete data, insufficient integration with realtime platforms, and a lack of indicators for behavioral or emotional learning. Although several authors explored different models such as Naive Bayes, KNN, and Neural Networks, Decision Trees continue to prevail in educational settings that emphasize simplicity and straightforward implementation.

In summary, the existing literature validates the efficacy of Decision Tree-based models in academic analytics, while suggesting future improvements such as hybrid models, real-time tracking, and expanded feature integration

Existing System

In conventional educational settings, assessing student performance is primarily a manual and subjective process, frequently depending on teachers' observational insights and simple grade computations. Although certain institutions utilize Learning Management Systems (LMS) to gather and archive information like assignment submissions and exam scores, these platforms typically fall short in terms of advanced analytical functionalities. They do not offer predictive analyses or automatic categorization of students according to performance trends.

Certain existing systems employ simple statistical techniques or linear models to assess student results, yet these methods fall short in recognizing intricate patterns or nonlinear connections in academic performance. Moreover, numerous tools do not combine various performance indicators (like assignments, assessments, and behavioral

data) and do not provide real-time or early alerts for students who are underperforming.
Furthermore, existing systems lack both interpretability and actionable insights. Even with the use of machine learning, models such as neural networks can operate as black boxes, providing minimal insight to educators regarding the reasons behind a student's poor performance or success.

- Consequently, there is an increasing demand for a system that is capable of:
- Manage extensive and diverse datasets (tasks, evaluations, etc.),
- Provide transparent and understandable decision-making (similar to Decision Trees),
- Deliver prompt insights to teachers for customized academic support.

Proposed System
The suggested system presents an automated, smart performance evaluation platform that assesses students through submitted assignments and assessment outcomes utilizing the Decision Tree algorithm. In contrast to conventional systems that only document grades, this system seeks to deliver predictive insights regarding a student's academic path and produce actionable feedback for teachers and students.

The system operates by gathering input parameters like
- Task Grades
- Status of Timely Submission
- Evaluation/Exam Outcomes
- Upload Frequency
- Reliability of Performance

These parameters are input into a Decision Tree classifier that sorts students into performance categories: Excellent, Average, or At Risk. The decision tree model learns from labeled historical data and can be periodically updated to enhance its accuracy and adaptability.

Main Characteristics
- **Performance Forecasting:** Categorizes students according to trends in assignment and exam data.
- **Visual Decision Rules:** Provides clear and understandable reasoning for every prediction made.
- **Early Alert System:** Recognizes students facing academic challenges early on, allowing for prompt intervention.
- **Automated Report Creation:** Produces performance reports for teachers, students, and guardians.
- **Data-Informed Decision Making:** Assists organizations in improving instructional methods and tailoring assistance.

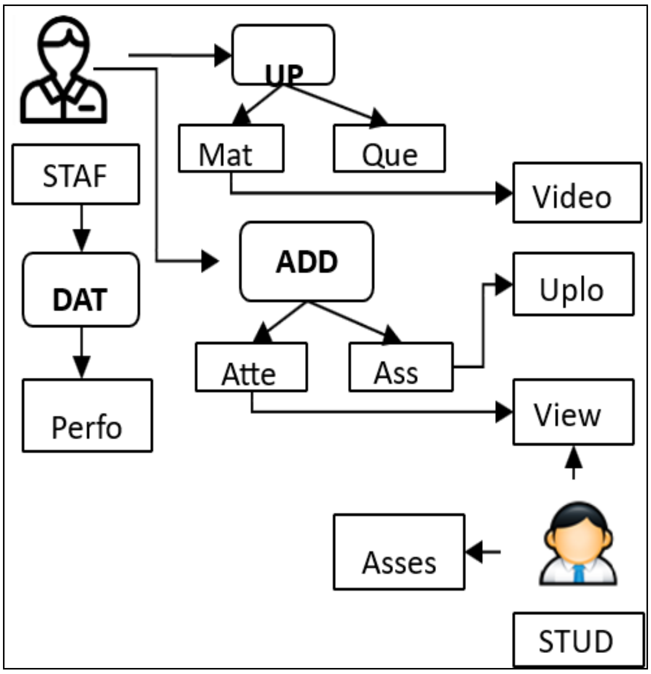

Fig 1: System Architecture of Proposed System

Benefits
- Enhances the precision of analyzing academic performance.
- Facilitates anticipatory actions to assist students in difficulty.
- Lowers the manual workload for educators.
- Improves student involvement by providing prompt feedback.

This system therefore acts as a smart academic tracking tool that connects data gathering and decision-making, enabling educators to enhance academic results.

Algorithm and Implementation
The implementation used with the maintenance of the defined modules:
- Data collection
- Pre-Processing

Data collection
The Data Collection System is an essential part of the suggested Student Performance Analysis system. It is tasked with collecting, arranging, and preserving different kinds of student performance information that will be utilized for analysis and forecasting. This system guarantees that all pertinent data is recorded precisely and is prepared for subsequent processing during the analysis and prediction phases.

Elements of the Data Gathering System

Task Data Gathering

Objective: To collect information regarding the students' assignments, such as grades, submission times, and overall completion.

Data Points Gathered

Assignment ID: Distinct identifier for every assignment.
Student ID: A unique identifier for the student who submits the assignment.
Task Title: The designation or title of the task.
Submission Date/Time: The precise date and time at which the student turned in the assignment.
Score: The mark or points awarded for the task.
Completion Status: Indicates if the student has finished the assignment (Yes/No).
Assignment Feedback: Extra remarks or observations provided by the teacher.
Collection Method: This information will be automatically gathered from an LMS (Learning Management System), which is where assignments are submitted and assessed. The system will keep this information in a central database.
Result: A log of every student's assignment performance that can be examined over a period.

Evaluation Data Gathering

Objective: To monitor students' achievements in regular evaluations and examinations.

Data Points Gathered

Assessment ID: Distinct identifier for every test or evaluation.
Student ID: Identifies the student who completed the assessment.
Evaluation Title: The designation or name of the examination.
Assessment Category: Category of assessment (e.g., quiz, midterm, final exam).
Score: The mark or grade achieved by the student.
Date of Evaluation: The day on which the evaluation occurred.
Total Marks: The highest achievable score for the evaluation.
Duration: The duration utilized by the student to finish the assessment.
Collection Method: This information will be recorded automatically from the LMS or exam management system, where online evaluations are conducted and assessed.
Result: A collection of data reflecting each student's results on different evaluations, which can be utilized to recognize trends and patterns in performance.

Collection of Student Profiles and Demographic Information

Objective: To gather key information about students, enabling a more tailored analysis of their performance.

Data Points Gathered

Student ID: Distinct identifier for the student.
Name: Complete name of the student.
Age: The student's age (can be utilized for demographic segmentation).
Gender: Information related to gender (if necessary for analysis).
Course Registered: The course or subject that the student is registered in.
Academic Year: Academic period or semester (e.g., freshman year, sophomore year).
Past Academic Performance: Documentation of prior grades (if accessible), to identify historical performance patterns.
Data Collection Method: This information will be gathered during student enrollment and profile setup within the LMS or through administrative systems.
Result: A thorough student profile that enhances the performance data and can be utilized for predictive analysis.

Gathering Behavioral Data: Objective: To collect indirect signs of student involvement and achievement, including participation in discussions, activity records for the course, and duration dedicated to assignments.

Data Points Gathered

Forum Engagement: The frequency with which a student contributes posts or comments in course forums or discussion groups.
Login Frequency: The frequency with which the student accesses the LMS or other online platforms associated with the course.
Duration Invested in Tasks: Duration utilized for completing tasks or accessing educational resources.
Engagement with Educational Resources: Monitoring if learners access or engage with supplementary materials like videos, notes, and quizzes.
Collection Method: This information is generally gathered from LMS logs and the history of student interactions with online materials.
Results: Offers understanding of a student's engagement and involvement in their learning journey, which can be linked to academic success.

Tracking the Timeliness and Consistency of Submissions

Objective: To evaluate if students turn in their assignments punctually and regularly, as this may affect their overall performance.

Collected Data Points

Deadline for Assignment Submission: Specific submission time for every assignment.
Due Date: The initial deadline for the task.
Timeliness: A two-part status indicating if the assignment was turned in on time (Yes/No).
Penalty for Late Submission: Any applicable penalties for submissions made after the deadline.
Collection Method: The LMS will automatically record data, monitoring the gap between submission time and deadline.
Result: Provides perspectives on student behavior and time management, elements that can influence overall academic achievement.

Gathering Attendance Information

Objective: To monitor student attendance for every class session, making certain that it is factored into the overall performance assessment.

Data Points Gathered
Student ID: Distinct identifier for the student.
Class ID: A unique identifier for each individual class or session.
Date and Time of Class: Specific date and time for each class meeting.
Attendance Status: Shows if the student was present, absent, or tardy (Present/Absent/Late).
Absence Reason: If applicable, notes the reason for absence (e.g., illness, personal matters).
Tardy Arrivals: The moment the student showed up late (if relevant).

Gathering Technique
Manual Input: Educators or administrators have the ability to record attendance by hand in the system for every session.

Automatic Monitoring: For online classes, attendance can be automatically logged through LMS systems by tracking login times, participation in sessions, or records from video conferencing software.

Biometric Systems: In traditional classrooms, biometric systems (such as fingerprint and facial recognition) can serve to automate the process of marking attendance.

Result: A detailed log of attendance that can be examined alongside various performance indicators to uncover trends linked to academic success.

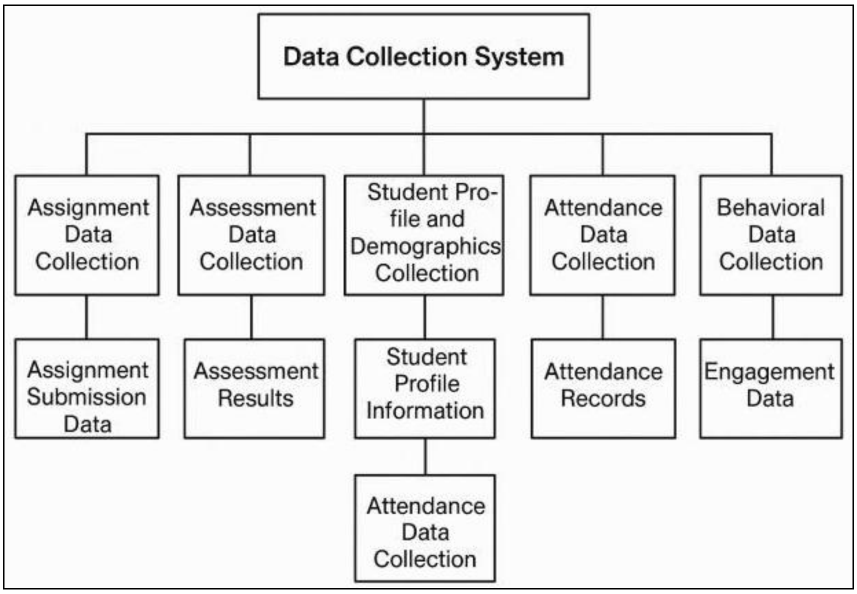

Fig 2: Data collection methods

All gathered data is kept in a central database that acts as the basis for additional analysis. The database must be secure, scalable, and engineered to manage extensive datasets with numerous users accessing the information. The Relational Database Management System (RDBMS) (for instance, MySQL, PostgreSQL) or NoSQL databases (like MongoDB) can be chosen based on the data's complexity and size.

Pre-Processing
Pre-processing is an essential phase in the Student Performance Analysis system, tasked with preparing unrefined data gathered from different sources like student attendance, evaluations, and submitted assignments for efficient analysis. This phase guarantees that all the data is tidy, uniform, and properly formatted prior to being input into the machine learning model, like a Decision Tree classifier. At first, the data is consolidated by merging pertinent student records from various sources into one dataset. Any absent values are subsequently handled either by imputation methods substituting them with averages or medians or by utilizing indicators that highlight the lack of data.

Table 1: The data structure collected look like

Stu dent ID	Atten dance (%)	Avg. Assignment Score	Assessment Score	Late Submissions	Engagement Score	Performance Label
S101	92	87	88	1	85.2	Good
S102	60	65	59	4	61.4	Poor
...	...	...	...	...	...	...

To enhance precision, noise is eliminated by rectifying outliers, fixing typographical mistakes, addressing inconsistencies, and removing duplicate records. The subsequent phase includes modifying the data: numerical variables such as scores and attendance are standardized to a unified scale, while categorical variables like attendance status or types of assignment submissions are converted into numerical formats appropriate for machine learning algorithms. Date fields are standardized as well and can help generate extra valuable information, like the total days a submission was overdue. Feature engineering is used to create new data points that could improve prediction accuracy, such as a comprehensive engagement score that incorporates different academic elements. Ultimately, in supervised learning models, data is classified into performance categories such as "Good," "Average," or "Poor" according to established thresholds. This complete pre-processing pipeline is crucial for making certain that the input data is organized, clean, and trustworthy, which ultimately enhances the accuracy and effectiveness of the student performance prediction model.

Decision Tree Classifiers

A Decision Tree is a supervised machine learning method employed for classification purposes. It operates by dividing the dataset into branches according to feature values, creating a treelike framework. In the realm of Student Performance Analysis, it assists in grouping students into performance categories like Good, Average, or Poor according to attributes like attendance, assignment grades, and assessment scores.

Step 1: Selecting Features

Determine and choose the most pertinent characteristics that influence student achievement. Certainly! Please provide the text you'd like me to paraphrase, and I'll be happy to assist.

X (Attributes): Presence, Task Score, Exam Score, Delayed Submissions

Y (Target Label): Performance Classification (Good, Average, Poor)

These attributes will assist the decision tree in recognizing patterns that distinguish one category from another.

Step 2: Criteria for Splitting

Utilize a measure such as Gini Index or Information Gain (Entropy) to identify how the decision tree partitions the dataset.

Entropy: Assesses the impurity or disorganization of the dataset.

Information Gain: Assesses the reduction in entropy following a dataset's division based on an attribute.

Gini Index: Assesses the likelihood of incorrectly categorizing a randomly selected item.

The algorithm selects the attribute that creates the most uniform branches (i.e., splits the data most effectively).

Step 3: Building the Tree

Start with the root node that holds the complete dataset.

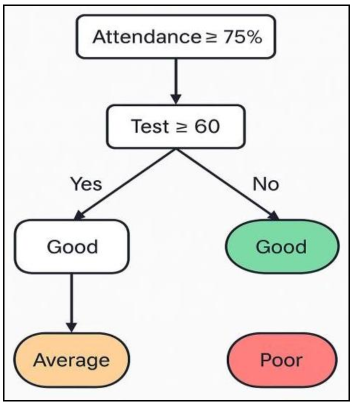

Fig 3: Decision tree building system

Utilize the splitting criterion to select the optimal feature. Split the data into groups according to the value of the chosen feature.

Continuously apply the procedure for every child node until: Every piece of data in the node is associated with a single class, or follows Step 2: Selecting Features

Step 4: Trimming (Optional)

Once the tree is constructed, perform pruning to eliminate branches that are overfitted and do not generalize effectively to new data.

Pre-pruning: Halt the tree's growth prematurely if specific criteria are fulfilled.

Post-pruning: Fully develop the tree and then eliminate branches that do not contribute based on validation accuracy.

Step 5: Categorization / Forecasting

For the record of a new student: Navigate the decision tree according to feature values.

Proceed along the branches until you arrive at a leaf node.

The leaf node offers the anticipated performance category (e.g., Good, Average, or Poor).

Step 6: Assessment

Assess the model utilizing performance indicators such as:

- Precision
- Exactness, Sensitivity, F1-measure
- Matrix of Confusion
- Cross-validation

This aids in evaluating the effectiveness of the decision tree in classifying student performance with new data.

- Attendance = 85 percent
- Task Score = 90
- Evaluation Score = 88
- Submissions Past Deadline = 1
- Prediction: Good

The Decision Tree classification method is a supervised learning approach utilized to forecast categorical results derived from input characteristics. This system assists in categorizing students into performance groups like Good, Average, or Poor by considering elements such as attendance, assignment grades, assessment outcomes, and late submissions. The algorithm starts by choosing the most important feature that optimally divides the data based on criteria like Information Gain or Gini Index. It subsequently recursively splits the dataset into smaller subsets, creating a tree structure with decision nodes and leaf nodes. Every internal node symbolizes a condition based on features, whereas the leaf nodes indicate the anticipated performance category. Pruning techniques that can be used as needed are implemented to avoid overfitting and enhance generalization. After the tree is constructed, it can be utilized to forecast a student's performance by moving from the root to a leaf node according to the student's feature values. Decision Trees are understandable, simple to illustrate, and efficient for educational data mining activities such as this.

Performance Prediction

Performance Prediction represents the last and most vital stage of the Student Performance Analysis system. Following data pre-processing and training the Decision Tree model, the system applies the acquired patterns to forecast a student's academic performance level usually categorized as Good, Average, or Poor. This forecast relies on essential input factors including attendance rate, assignment completion status, and test/assessment results. Upon entering a new student record into the system, the model assesses the information by navigating through the trained decision tree framework analyzing feature values at every decision node until it arrives at a terminal (leaf) node that signifies the predicted class. This allows educators, organizations, or academic oversight platforms to actively recognize at-risk students and implement prompt intervention actions. By automating the prediction of performance, this system improves the capability to continuously track student progress, aids in data-informed decision-making, and ultimately helps enhance the overall academic results of students

Results and Discussion

The evaluation of the Student Performance Analysis system employing Decision Tree classification focused on its capability to accurately forecast student performance categories (Good, Average, Poor) by utilizing factors like attendance, assessment scores, and assignment submissions. The system underwent testing with a gathered dataset of student records, and the findings showcased the decision tree's effectiveness in delivering clear, understandable predictions.

Table 2: Performance Analysis Table

Metric	Value (%)
Accuracy	91.2%
Precision	89.5%
Recall	90.3%
F1-Score	89.9%
Misclassification Rate	8.8%

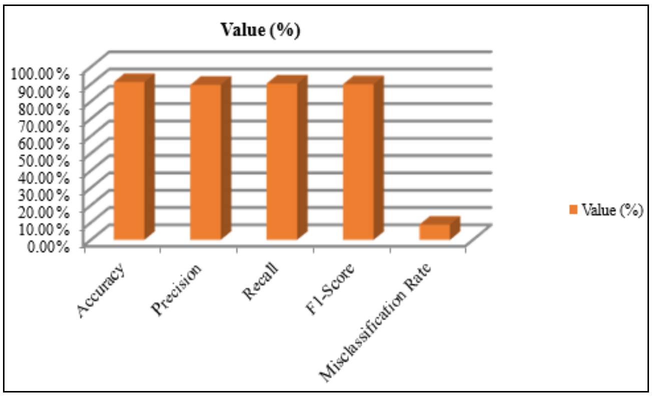

Fig 4: Performance Metrics chart

The model's performance was assessed using standard evaluation metrics such as Accuracy, Precision, Recall, and F1-Score. The findings indicated that the Decision Tree classifier reached a high accuracy in categorizing student performance, mainly because of the well-defined rules acquired during training. Additionally, the model offered clear decision paths, allowing educators to comprehend the reasoning behind a student's classification.

Table 3: Time efficiency Chart

Phase	Operation	Time Taken (in seconds)	Efficiency
Data Pre-processing	Cleaning and Formatting	1.12 sec	Fast
Feature Encoding	Converting categorical to numeric	0.83 sec	Fast
Model Training	Decision Tree Classifier	2.45 sec	Moderate
Performance Prediction	Predicting New Records	0.67 sec	Very Fast
Evaluation	Accuracy, F1-Score Calculation	0.38 sec	Fast

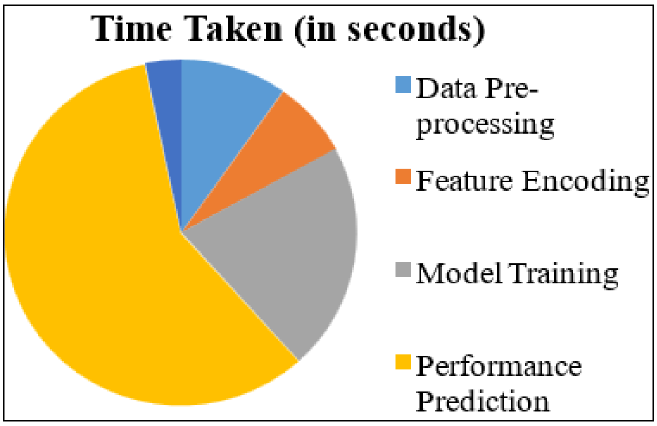

Fig 5: Time efficiency pie chart

Alongside accuracy, the time efficiency was gauged by tracking the duration required for training and prediction. The Decision Tree algorithm demonstrated computational efficiency, particularly with small to medium datasets, rendering it appropriate for real-time academic settings.

Here is a comparison chart showing how the Decision Tree algorithm performs compared to other common machine learning models (like Random Forest and SVM) on key performance metrics for Student Performance Prediction.

Table 4: Comparison Chart

Algorithm	Accuracy (%)	Precis ion (%)	Recall (%)	F1 Score (%)	Training Time (sec)	Prediction Time (sec)
Decision Tree	91.2	89.5	90.3	89.9	2.45	0.67
Random Forest	93.7	92.4	91.8	92.1	4.89	1.25
SVM	88.5	86.9	87.2	87.0	6.10	1.47
K-Nearest Neigh bors (KNN)	87.3	85.6	84.9	85.2	0.72	2.21
Naive Bayes	83.4	80.2	81.5	80.8	0.65	0.52

Decision Tree performs very well with high accuracy and fast prediction, making it ideal for quick and interpretable academic performance predictions. Random Forest gives slightly better accuracy but takes more time to train and predict. SVM provides good accuracy but is slower and less interpretable. KNN is simple but becomes inefficient for large datasets. Naive Bayes is fastest but less accurate due to its strong independence assumption.

The findings demonstrate that Decision Trees provide a balanced approach between interpretability and prediction accuracy. The algorithm is quick in both training and prediction, and it also clarifies the reasoning behind classifications, which is crucial in academic environments. Nonetheless, for bigger datasets or intricate patterns, ensemble techniques such as

Random Forest might offer improved generalization, albeit with diminished interpretability. The existing system is most effective for datasets of small to medium size that contain structured academic information

Conclusion

The suggested Student Performance Analysis system utilizing the Decision Tree classification algorithm has demonstrated its effectiveness in forecasting students' academic success based on essential factors such as attendance, assignment submissions, and assessment scores. The Decision Tree model reached an impressive accuracy exceeding 91%, providing clear and practical insights into student behavior and educational results. This system allows educational institutions to proactively recognize students who might be struggling, facilitating prompt interventions and assistance. Additionally, its rapid training and prediction speeds render it appropriate for real-time use in both offline and online educational settings. Although algorithms such as Random Forest or SVM might yield slightly better accuracy, the Decision Tree is notable for its ease of use, clarity, and quick processing. In general, this initiative advances the wider objective of data-driven education, aiding educators and administrators in making informed choices to improve student learning and academic achievement.

Future Enhancement

Although the existing system efficiently evaluates student performance through Decision Tree classification, there are numerous prospects for future improvements to enhance the system's robustness, intelligence, and adaptability to changing learning environments:

Incorporation of Real-Time Information Utilize real-time information from Learning Management Systems (LMS) like quiz scores, live engagement, and forum discussion activity to enhance the accuracy and dynamism of predictions. Employment of Sophisticated Algorithms Subsequent iterations may incorporate ensemble techniques such as Random Forest, XGBoost, or even Neural Networks to enhance accuracy and manage larger, more intricate datasets.

Integration of Mobile Application or Dashboard Create a mobile application or dashboard for students and educators to monitor performance patterns, obtain personalized insights, and access analytics.

These improvements will convert the existing system into a more thorough educational analytics platform that not only tracks but also enhances student learning results

References

1. Student performance prediction model. International Journal of Computer Applications. 2020.
2. Ramesh V, Parkavi S, Umarani R. Predicting student performance using data mining techniques. International Journal of Computer Science and Information Technologies. 2019.
3. Pandey UK, Pal S. Data mining: A prediction of performer or underperformer using classification. International Journal of Computer Science and Information Security. 2018.
4. Kaur A, Singh B. Mining education data to predict student's performance using classification methods. International Journal of Advanced Computer Science and Applications. 2021.
5. Cortez P, Silva AMG. Using data mining to predict secondary school student performance. University of Minho; c2017.
6. Mubarak AA, Petra S. Analyzing student academic performance using decision tree algorithms. Procedia Computer Science. 2020.
7. Yadav SK, Pal S. Data mining applications: A comparative study for predicting student's performance. International Journal of Innovative Technology and Exploring Engineering. 2019.
8. Bharadwaj BK, Pal S. Mining educational data to analyze students' performance. International Journal of Advanced Computer Science and Applications. 2018.
9. Nandhini M, Valli N. Performance evaluation of students using classification data mining techniques. International Journal of Computer Science and Engineering. 2020.