

**A
Project On**

“Decision Tree–Based Student Performance Prediction System”

By

Vijayalakshmi Madduru

Abstract

Student performance prediction is an important area of research in the field of educational data mining. Educational institutions generate a large amount of data related to students' academic activities, such as attendance, internal assessments, assignments, and examination results. Analyzing this data using machine learning techniques can help predict student performance and identify students who may need additional academic support.

This project focuses on predicting student performance using a Decision Tree classification algorithm. The system uses academic attributes such as attendance percentage, daily study hours, internal marks, and assignment scores to predict whether a student will Pass or Fail. The Decision Tree model is trained on historical student data and tested for accuracy. The proposed system is simple, interpretable, and effective, making it suitable for educational institutions to improve academic planning and decision-making.

Introduction

In the modern education system, evaluating student performance plays a vital role in maintaining academic quality. Traditionally, student performance evaluation is done after examinations, which is a reactive approach. By the time poor performance is identified, it is often too late to take corrective measures.

With the advancement of machine learning and data analytics, it has become possible to predict student performance at an early stage using historical data. Machine learning models can analyze patterns in student behavior and academic records to predict future outcomes.

Among various machine learning algorithms, the Decision Tree algorithm is widely used due to its simplicity and transparency. It represents decision-making in the form of a tree structure, which is easy to understand even for non-technical users. This project demonstrates how a Decision Tree model can be used to predict student performance effectively.

Problem Statement

Educational institutions face challenges in identifying students who are at risk of poor academic performance. Manual monitoring of each student is difficult, time-consuming, and inefficient. There is a need for an automated system that can analyze student data and predict academic outcomes in advance.

Problem Statement:

To design and implement a machine learning-based system using a Decision Tree algorithm that predicts student performance (Pass/Fail) based on academic attributes such as attendance, study hours, internal marks, and assignment scores.

Objectives of the Project

The main objectives of this project are:

1. To study and analyze student academic data
2. To apply the Decision Tree algorithm for classification
3. To predict student performance accurately
4. To identify important factors influencing student results
5. To provide an interactive system that accepts user input
6. To demonstrate the practical application of machine learning in education

Scope of the Project:

The scope of this project is limited to predicting student performance based on selected academic attributes. The system is designed for small to medium-sized datasets and is suitable for academic demonstration purposes.

The project does not include:

- i. Psychological or behavioral analysis
- ii. Real-time institutional databases
- iii. Online deployment

However, it provides a strong foundation for future enhancements.

Literature Review

Educational Data Mining (EDM) has gained significant attention in recent years. Several studies have applied machine learning algorithms such as Decision Trees, Naive Bayes, Support Vector Machines, and Random Forests to predict student performance.

Research shows that Decision Trees are effective due to:

- i. Easy interpretability
- ii. Ability to handle both numerical and categorical data
- iii. Minimal preprocessing requirements

Previous studies indicate that factors like attendance, internal assessments, and assignment performance significantly impact final academic results. This project builds on these findings by implementing a Decision Tree-based prediction system.

Proposed System

The proposed system is a machine learning-based student performance prediction system. It uses a Decision Tree classifier trained on historical student data to predict whether a student will pass or fail.

Features of the Proposed System

- i. Automated prediction of student performance
- ii. Accepts real-time user input
- iii. Easy-to-understand prediction results
- iv. Visual representation of decision rules
- v. Efficient and user-friendly

System Architecture

The system architecture consists of the following components:

1. Input Module

- i. Student academic data (attendance, study hours, marks)

2. Data Processing Module

- i. Feature selection
- ii. Data organization

3. Model Training Module

- i. Decision Tree classifier training

4. Evaluation Module

- i. Accuracy calculation Confusion matrix

5. Prediction Module

- i. User input-based prediction

6. Output Module

- i. Pass / Fail result

Methodology

Data Collection

A sample dataset containing student academic details was created. The dataset includes:

- Attendance percentage
- Daily study hours
- Internal marks
- Assignment score
- Final result (Pass/Fail)

Data Preprocessing

- Data is organized in tabular format using Pandas DataFrame
- Relevant features are selected
- Target variable is identified

Feature Selection

Input features:

- Attendance
- StudyHours
- InternalMarks
- AssignmentScore
- Target variable: Result (Pass / Fail)

Train–Test Split

The dataset is divided into:

- 70% training data
- 30% testing data

Model Training

The Decision Tree classifier is trained using the entropy criterion to calculate information gain.

Prediction

The trained model predicts:

- Test dataset results
- User-entered student data

Model Evaluation

The model is evaluated using:

- Accuracy score
- Confusion matrix
- Classification report

Decision Tree Algorithm

A Decision Tree is a supervised learning algorithm used for classification and regression tasks. It splits the dataset into subsets based on the feature that provides the highest information gain.

Working of Decision Tree

1. Select the best feature using entropy
2. Split the dataset
3. Repeat recursively for child nodes
4. Stop when a leaf node is reached

Advantages

- Simple and interpretable
- Easy to visualize
- Handles non-linear relationships

Disadvantages

- Overfitting on small datasets
- Sensitive to data changes

Tools and Technologies Used

- **Programming Language:** Python
- **Libraries:**
 - Pandas
 - NumPy
 - Scikit-learn
 - Matplotlib
- **IDE:** Jupyter Notebook

Results and Discussion

The Decision Tree model successfully predicts student performance based on academic features. The model achieves good accuracy on the test dataset. The decision tree visualization clearly shows how decisions are made based on different attributes.

The system also allows real-time prediction using user input, making it interactive and practical for demonstration.

Applications

- Academic performance analysis
- Early identification of weak students
- Student counselling support
- Educational research
- Academic decision-making

Limitations

- Small dataset size
- Limited features
- Not deployed as a real-time system

Future Scope

- Use real-world datasets
- Apply ensemble algorithms (Random Forest)
- Add more features (behavioral, psychological)
- Develop a web or mobile application
- Improve model accuracy

Conclusion

This project successfully demonstrates the application of a Decision Tree algorithm for predicting student performance. By analyzing academic attributes, the system can predict whether a student will pass or fail. The model is simple, interpretable, and effective, making it suitable for educational institutions. The project highlights the importance of machine learning in improving academic planning and student success.