

Student Marks Prediction Using Machine Learning

1. Introduction

With the growth of data-driven technologies, machine learning is increasingly used to analyze and predict real-world outcomes. In the education domain, predicting student performance can help identify learning patterns and provide early academic support.

This mini-project focuses on predicting student marks using supervised learning algorithms based on study-related features.

2. Problem Statement

To predict the marks obtained by a student based on the number of courses enrolled and the time spent studying, using machine learning regression models.

3. Dataset Description

The dataset used in this project is the Student Marks Dataset obtained from Kaggle.

Attributes:

Attribute Name	Description
number_courses	Number of courses taken by the student
time_study	Time spent studying (hours)
Marks	Final marks obtained

4. Methodology

1.Data Collection:

The dataset was loaded using Pandas.

2.Data Preprocessing:

Relevant features were selected, and the target variable was separated.

3.Train-Test Split:

The data was split into training (80%) and testing (20%) sets.

4.Model Training:

- Linear Regression
- Decision Tree Regression

5.Evaluation:

Models were evaluated using:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)

6. Prediction:

User inputs were taken to predict marks using the trained model.

5. Algorithms Used

5.1 Linear Regression

Linear Regression models the relationship between input features and the target variable by fitting a linear equation.

5.2 Decision Tree Regression

Decision Tree Regression predicts output by learning decision rules inferred from the data features.

6. Results and Discussion

- Linear Regression showed stable performance for continuous data.
- Decision Tree Regression captured complex patterns but showed higher variance.
- Linear Regression produced more consistent predictions.

7. Conclusion

The project successfully demonstrates how machine learning can be used to predict student marks using basic academic parameters. The comparison of models highlights the strengths and limitations of different regression techniques.

8. Future Scope

- Include more student-related features
- Apply advanced regression models
- Deploy the model as a web application
- Use cross-validation for better evaluation

9. References

- Kaggle Dataset: Student Marks Dataset
- Scikit-learn Documentation
- Python Official Documentation