# Predicting Student Academic Performance Using Machine Learning Techniques

**Authors:** Shreyash Vachkal
**Affiliation:** Department of Artificial Intelligence and Data Science, XYZ Institute of Technology, Pune, India
**Publication Year:** 2025
**Journal/Conference:** International Journal of Machine Learning and Data Analytics

## Abstract

Predicting student academic performance has become an important research area in educational data mining and machine learning. Accurate prediction models can help educators identify at-risk students early and design personalized learning interventions. This research paper presents a comparative study of multiple machine learning algorithms to predict student academic outcomes based on demographic, academic, and behavioral features. Algorithms such as Linear Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) are evaluated on a publicly available student performance dataset. Experimental results show that ensemble-based models outperform traditional regression techniques in terms of accuracy and robustness. The findings highlight the potential of machine learning in improving educational decision-making.

**Keywords:** Machine Learning, Student Performance Prediction, Educational Data Mining, Random Forest, Classification

---

## 1. Introduction

Education systems generate large volumes of data related to student demographics, attendance, assessments, and learning behavior. Analyzing this data manually is time-consuming and often ineffective. Machine learning (ML) techniques provide automated methods to discover hidden patterns and make accurate predictions from such data.

Student performance prediction is a critical task that enables early identification of students who may require additional academic support. Traditional statistical approaches often fail to capture complex, non-linear relationships present in educational data. In contrast, machine learning models can adapt to these complexities and deliver better predictive performance.

This paper focuses on building and comparing different ML models to predict student academic performance. The main objectives of this research are:

- To analyze the effectiveness of various machine learning algorithms for performance prediction.
- To compare model performance using standard evaluation metrics.
- To identify the most influential features affecting student outcomes.

---

# 3. Methodology

## 3.1 Dataset Description

The dataset used in this study contains records of secondary school students, including attributes such as age, gender, study time, attendance, parental education, and previous grades. The target variable represents the final academic performance of the student, categorized into performance levels.

## 3.2 Data Preprocessing

Data preprocessing steps included:

- Handling missing values using mean and mode imputation
- Encoding categorical variables using one-hot encoding
- Feature scaling using standardization
- Splitting the dataset into training (80%) and testing (20%) sets

## 3.3 Machine Learning Models

The following machine learning models were implemented:

- **Linear Regression**: Used as a baseline regression model
- **Decision Tree**: Captures non-linear relationships
- **Random Forest**: An ensemble method that reduces overfitting
- **Support Vector Machine (SVM)**: Effective for high-dimensional data

## 3.4 Evaluation Metrics

Model performance was evaluated using:

- Accuracy
- Precision
- Recall
- F1-score

---

# 4. Experimental Results

The experimental results indicate that Random Forest achieved the highest accuracy among all models. Linear Regression showed lower performance due to its inability to model complex patterns.