Student Performance Prediction Using Machine Learning

Abstract

This project focuses on predicting student academic performance using supervised machine learning techniques. Regression and classification models are implemented to analyze student exam performance based on demographic and academic features. Linear Regression is used to predict math scores, while Logistic Regression and Decision Tree classifiers are used to classify students as pass or fail.

1. Introduction

Educational institutions generate large volumes of student data. Analyzing this data using machine learning helps in predicting student performance and identifying students at risk. This project applies supervised learning techniques to predict academic outcomes using exam data.

2. Dataset Description

The dataset contains 1000 student records with the following attributes:

- Gender
- Race/Ethnicity
- Parental Level of Education
- Lunch Type
- Test Preparation Course
- Math Score
- Reading Score
- Writing Score

The dataset is used for both regression and classification tasks.

3. Data Preprocessing

- No missing values were found
- Categorical features were encoded using Label Encoding
- Dataset split into 80% training and 20% testing data

---

4. Methodology

4.1 Regression

- Algorithm: Linear Regression
- Target Variable: Math Score
- Evaluation Metrics: Mean Squared Error (MSE), R² Score

4.2 Classification

- Algorithms: Logistic Regression, Decision Tree
- Target Variable: Pass/Fail (Math Score $\geq$ 40)
- Evaluation Metric: Accuracy

---

5. Results

Model	Performance
Linear Regression	$R^2 \approx 0.88$
Logistic Regression	Accuracy $\approx 97.5\%$
Decision Tree	Accuracy $\approx 97.5\%$

---

6. Discussion

The regression model shows strong predictive performance. Classification models accurately distinguish between passing and failing students. Reading and writing scores are strong predictors of math performance, while parental education and test preparation positively impact outcomes.

---

7. Conclusion

The project demonstrates that supervised machine learning models can effectively predict student performance. These models can assist educators in early intervention and academic planning.

---

8. Future Enhancements

- Apply Random Forest or XGBoost
- Perform hyperparameter tuning
- Deploy the model using Streamlit or Flask