GradeGenius: Student Performance Predictor

Md. Talha Mahmud*, Md. Siam Hossain†, Asif Manowar‡
Department of CSE, R.P. Shaha University, Narayanganj
Batch: 26
*23100069, †23100084, ‡23100016

Course Code: 0613-CSE-3207
Course Title: Artificial Intelligence
Course Instructor: Sayma Akter Rupa, Lecturer, CSE, RPSU
Semester: Fall 2025

***Abstract*—This project aims to develop a machine learning-based solution to predict student performance using various attributes such as attendance, study time, previous grades, and parental background. Our goal is to classify students into pass/fail categories or predict final exam scores. We plan to use models like Linear Regression, Decision Tree, and Random Forest to analyze the data and uncover key factors influencing student success.**

***Index Terms*—student performance, machine learning, prediction, regression, classification**

I. Introduction

Student performance prediction is a critical educational challenge that can help identify at-risk students early and offer necessary academic support. Our project, GradeGenius, leverages supervised learning techniques to analyze multiple academic and personal features of students. By training models such as Linear Regression and Random Forest on real-world datasets, we aim to forecast final exam scores or pass/fail status with improved accuracy. This model can assist teachers and institutions in making data-driven decisions.

II. Literature Review

Leave it for Phase 02.

III. Methodology

Visit the following website and choose any algorithm that you prefer: https://scikit-learn.org/stable/index.html
(You don't have to describe the methodology right now; just mention the algorithms.)

Algorithms selected:

- Linear Regression
- Decision Tree
- Random Forest

IV. Data Set

We are using the following datasets for this project:

- https://www.kaggle.com/datasets/satayjit/student-performance-bd
- https://archive.ics.uci.edu/dataset/320/student+performance

The datasets contain features such as student demographics, past performance, attendance records, study hours, and parental education. These features will help us build robust models for performance prediction. We will also apply preprocessing techniques such as missing value imputation and feature scaling.

V. Results and Analysis

Leave it for Phase 02/03.

VI. Conclusion

Leave it for Phase 02/03.

VII. References

Please add the references here!