

# STUDENT GRADE ANALYSIS PREDICTION

Tikam (2022542)

Kartikeya Malik (2022243)

## Different Models Tried

- 1) Linear Regression
- 2) Random Forest
- 3) Gradient Boosting
- 4) Support Vector Regression (SVM)

## Different Datasets Used

- 1) [Dataset1.csv](#)
- 2) [Dataset2.csv](#)
- 3) [Dataset3.csv](#)

Contains information about students, including their Student ID, Gender, Age, Family Size, Parent's Education, Student's Education, Subject, Grades for four periods, and various social and college-related factors.

## Problem Being Addressed

- The project aims to predict student's grades in the fourth period based on various demographic and academic factors.
- The objective is to assist educators in identifying students who may need additional support or interventions to improve their academic performance.

## Relevant Literature

- Reviewing existing literature on student performance prediction, machine learning in education, and educational psychology can provide insights into effective methodologies and potential challenges in this domain.
- This review guides the selection of appropriate models and features for the prediction task.

## Methodology

**Data Preprocessing:** This involves cleaning the dataset, handling missing values, and encoding categorical variables.

**Feature Selection:** Identifying relevant features that may influence student's grades and preparing the data for model training.

**Model Selection:** Exploring various regression models such as Linear Regression, Random Forest, Gradient Boosting, and Support Vector Regression to determine the best-performing model for the task.

**Model Evaluation:** Assessing model performance using metrics like Mean Squared Error (MSE).

## Experimental Settings

**Feature Scaling:** Standardization is applied to scale the features to a standard range, ensuring that all features contribute equally to the model.

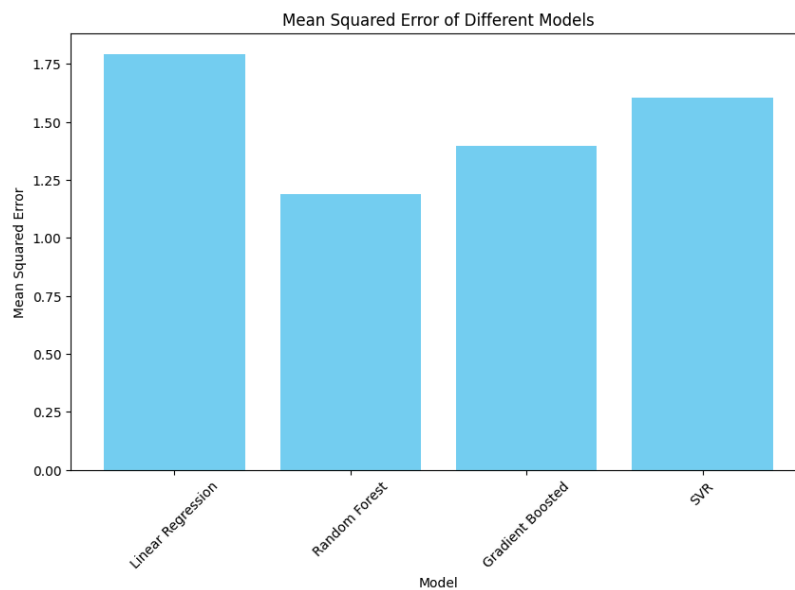
**Model Training:** Each selected model is trained on the training data.

**Model Evaluation:** The trained models are evaluated on the testing data, and their performance is compared based on MSE.

## Results

→ Mean Squared Error (MSE) is used as the evaluation metric for the trained models.

→ Results show the performance of each model in predicting student grades.



## Comparisons:

→ Comparison of the performance of different models in terms of MSE and other relevant metrics.

→ Insights into the strengths and weaknesses of each model and potential areas for improvement.

## Analysis Of Results

→ The analysis of results involves comparing the MSE of different models and identifying the best-performing model for predicting student grades.

```
Analysis of Results:  
Linear Regression: Mean Squared Error = 1.7932320794550862  
Random Forest: Mean Squared Error = 1.17676784138968  
Gradient Boosted: Mean Squared Error = 1.396983568112189  
SVR: Mean Squared Error = 1.6029581780590252
```

## Visualization:

→ Visualizations such as KDE plots, box plots, histograms, and count plots are used to analyze the dataset and gain insights into the relationship between different attributes and the final grade.

