**Student Exam Score Prediction – Report**

# Abstract

This project focuses on predicting student exam scores based on a variety of academic, socio-economic, and behavioral factors. Using machine learning models, the aim is to identify the most influential predictors of exam performance and provide a reliable predictive system. The dataset consists of 6,607 records and 20 features, including study habits, parental involvement, access to resources, attendance, and socioeconomic indicators.

# Introduction

Education is a key driver of personal and professional growth. However, students’ academic performance is influenced by multiple external and internal factors. The objective of this study is to build a machine learning model capable of predicting exam scores with reasonable accuracy, thereby assisting institutions in identifying students at risk and improving targeted interventions.

The dataset was obtained from Kaggle and includes features such as study hours, parental involvement, peer influence, school type, family income, and access to resources.

# Methodology

## Data Preprocessing

The dataset was cleaned and preprocessed to handle categorical features and ensure compatibility with regression models.

Correlation analysis was conducted to identify key predictors.

## Models Applied

Linear Regression – Baseline model to establish a simple linear relationship between features and exam scores.

Polynomial Regression (Degree 2 & 3) – To capture non-linear relationships between predictors and outcomes.

## Evaluation Metrics

Root Mean Squared Error (RMSE) – To measure prediction error.

R² Score – To assess goodness-of-fit and model accuracy.

# Results

**Linear Regression:** Provided a baseline accuracy but with relatively higher error.

**Polynomial Regression:** Improved results compared to linear regression, reducing RMSE and increasing R².

Overall, Polynomial Regression was identified as the best-performing model for this dataset.