Problem and Research Question

Problem: Predicting student exam scores based on various factors to identify key predictors of academic performance.

Research Question: What factors most significantly predict student exam performance, and how accurately can we predict exam scores using suitable machine learning models?

Significance

- Understanding factors that affect academic achievement.
- Obtaining early indicators of student underperformance.
- Improving resource allocation for the affected students.
- Obtaining personalized learning pathways for certain students.

Dataset Description

Source: Pulled from Kaggle.com

Number of Observations: 6,608

Features (19 predictors, 1 response variable)

Quantitative Features

- Hours_Studied
- Attendance
- Sleep_Hours
- Previous_Scores
- Tutoring_Sessions

- Physical_Activity

Qualitative Features

- Parental_Involvement
- Access_to_Resources
- Extracurricular_Activities
- Motivation_Level
- Internet_Access
- Family_Income
- Teacher_Quality
- School_Type
- Peer_Influence
- Learning_Disabilities
- Parental_Education_Level
- Distance_From_Home
- Gender

Response Variable

- Exam_score

Methods

- Linear Regression
- Regularization
- Cross validation
- Model Evaluation

Hypothesis: Factors like Previous_Scores, Hours_Studied, and Attendance will be strong predictors

Evaluations

- Implementing cross validation
- Comparing the performance of test sets of different models
- Metrics for evaluation include: MSE, RMSE,