

**machine learning report (Regression project STUDENTS’ EXAM SCORES)**

**By:**

|  |  |
| --- | --- |
| Name | ID |
| **Ahmed Sofi** | 42010635 |

**Under the Supervision of:**

Eng. Ahmed Nousir

Contents

1. [Introduction 3](#_bookmark0)
2. [The Problem understanding: 3](#_bookmark1)
   1. [Data Description 3](#_bookmark2)
   2. [Problem Statement 3](#_bookmark3)
3. [Data Exploration 3](#_bookmark4)
   1. [Descriptive Statistics 3](#_bookmark5)
   2. [Correlation Analysis 4](#_bookmark6)
4. [Data Preprocessing 5](#_bookmark7)
   1. [Handling Missing Values 5](#_bookmark8)
   2. [Encoding Categorical Variables 5](#_bookmark9)
   3. [Feature Scaling 5](#_bookmark10)
5. [Model Building 5](#_bookmark11)
   1. [Splitting Data 5](#_bookmark12)
   2. [Model Selection 5](#_bookmark13)
   3. [Training the Model 5](#_bookmark14)
   4. [Model Evaluation 5](#_bookmark15)
6. [Conclusion 6](#_bookmark16)

Figure 1 Statistics for data 4

Figure 2 Correlation 4

Figure 3 plots 4

Figure 5 Accuracy 6

# Introduction

This dataset includes scores from three test scores of students at a (fictional) public school and a variety of personal and socio-economic factors that may have interaction effects upon them.

# The Problem understanding:

The data set is about scores of students this data contain 14 columns and 30641 rows

## Data Description

Explain the importance of each column in the data set. Highlighting the role of factors such as the type of school lunch and the number of siblings a student has on the student’s overall grades.

## Problem Statement

Define the problem statement clearly. For example, you could state that the goal is to predict based on various factors in the dataset.

# Data Exploration

This data from Kaggle is about 30641 rows and 14 columns in his part we need to get more information about data and another statistics.

## Descriptive Statistics

Provide summary statistics for each column. This could include mean, median, standard deviation, and other relevant metrics. Identify any trends or patterns.

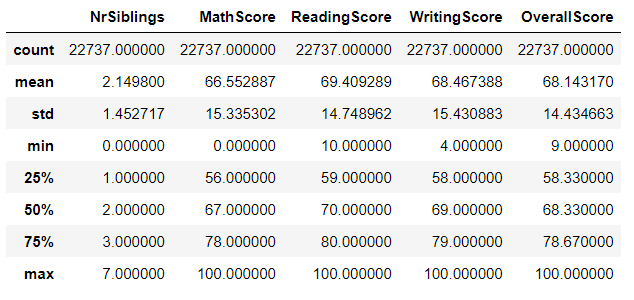


Figure 1 Statistics for data

## Correlation Analysis

Explore the correlation between all columns.

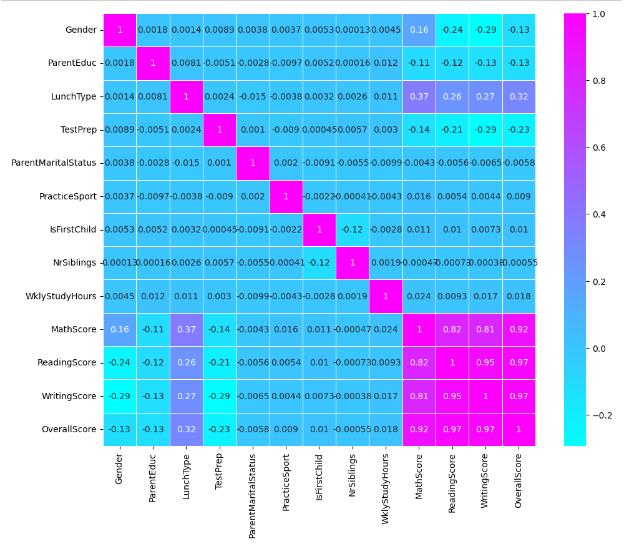


Figure 2 Correlation

**3.3 Visualizations**

P a g e

# Data Preprocessing

This part for clean data from nulls and duplicates and check outliers to make data ready for deploy model.

## Handling Missing Values

Address any missing values in the dataset through imputation or removal.

## Encoding Categorical Variables

If there are categorical variables, encode them into numerical format for model compatibility.

## Feature Scaling

Normalize or standardize numerical features if necessary to ensure fair treatment by the model.

# Model Building

## Splitting Data

Divide the dataset into training and testing sets.

## Model Selection

Choose a suitable machine learning model for the task (e.g., Random Forest, Linear Regreesion).

## Training the Model

Train the chosen model on the training dataset.

## Model Evaluation

Evaluate the model's performance on the testing dataset. Mention metrics such as accuracy, precision, recall, and F1 score.

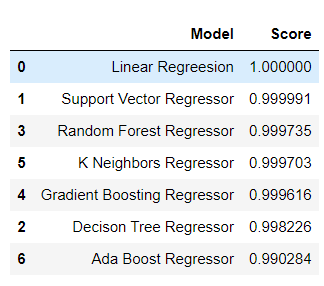


Figure 4 Accuracy

# Conclusion

Summarize the key findings and insights from the analysis. Discuss the limitations and potential areas for improvement. Offer recommendations for further research or model enhancements.