AI PROJECT

Students GPA Estimation

OVERVIEW

- Introduction
- Problem Statement
- Related Work
- Dataset Overview
- Data Preprocessing
- Visualizations
- Prepare for Models
- Applied Models
- Compare the Models
- Result
- Conclusion
- Thank You

INTRODUCATION

Objective :

Develop an AI-powered model to predict students' GPA using relevant factors. This tool aims to enhance early identification of at-risk students.

Impact:

Empowers educators to provide timely interventions, improving student success and well-being through targeted support.

PROBLEM STATEMENT

Students face a mix of academic and non-academic challenges (e.g., socioeconomic status, mental health, family support). Traditional methods may fail to address these complex issues effectively

Goal

Empower educators with data-driven insights to foster an inclusive and supportive environment, addressing the holistic needs of students.

Solution

Develop an AI-powered predictive model to estimate GPA using diverse factors (demographic, academic, behavioral, and non-academic data). Leverage machine learning to proactively identify at-risk students and provide timely, personalized interventions.

RELATED WORK

1. Does the Education Level of a Parent Affect a Child's Achievement in School?
2. Parental Involvement is Key to Student Success
3. The influence of sports participation on academic performance among students in higher education
4. School absenteeism and academic achievement: Does the timing of the absence matter?

DATASET OVERVIWE

- **Format** CSV file
- **Size** 2,393 rows and 15 features.
- **key Features**

Absence: Tracks student attendance.

Demographic Information: Age, gender, socioeconomic status, etc.

Academic Performance: study time, test scores, and study habits.

Non-Academic Factors: Family support, mental health indicators, etc.

PREPROCESSING

Missing Data:

We have no missing data

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2392 entries, 0 to 2391
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Age                2392 non-null   int64  
 1   Gender             2392 non-null   int64  
 2   Ethnicity          2392 non-null   int64  
 3   ParentalEducation  2392 non-null   int64  
 4   StudyTimeWeekly    2392 non-null   float64
 5   Absences           2392 non-null   int64  
 6   Tutoring           2392 non-null   int64  
 7   ParentalSupport    2392 non-null   int64  
 8   Extracurricular    2392 non-null   int64  
 9   Sports             2392 non-null   int64  
 10  Music              2392 non-null   int64  
 11  Volunteering       2392 non-null   int64  
 12  GPA                2392 non-null   float64
dtypes: float64(2), int64(11)
memory usage: 243.1 KB
```

PREPROCESSING

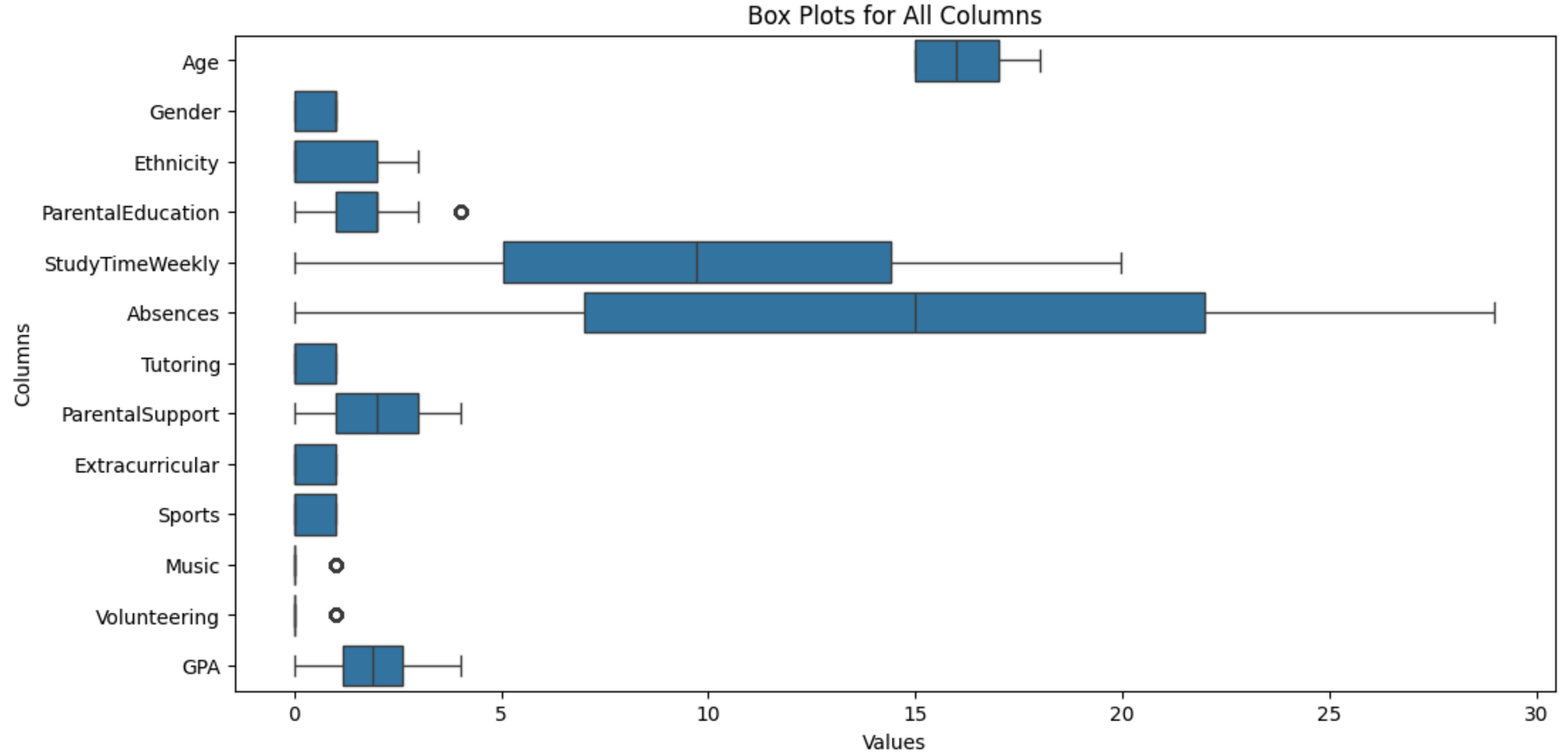
Duplicated Values:

And no duplicated values

```
df.duplicated().sum()
```

```
0
```

PREPROCESSING

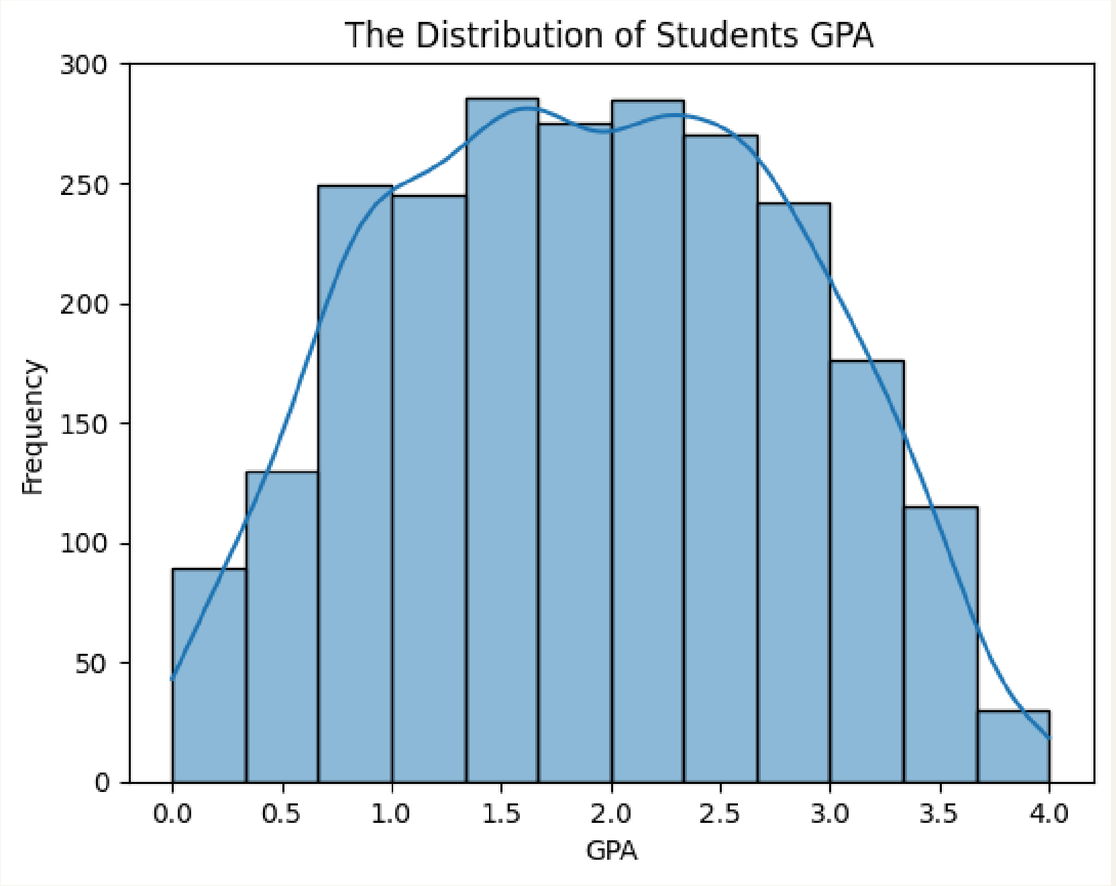


Check for outliers:

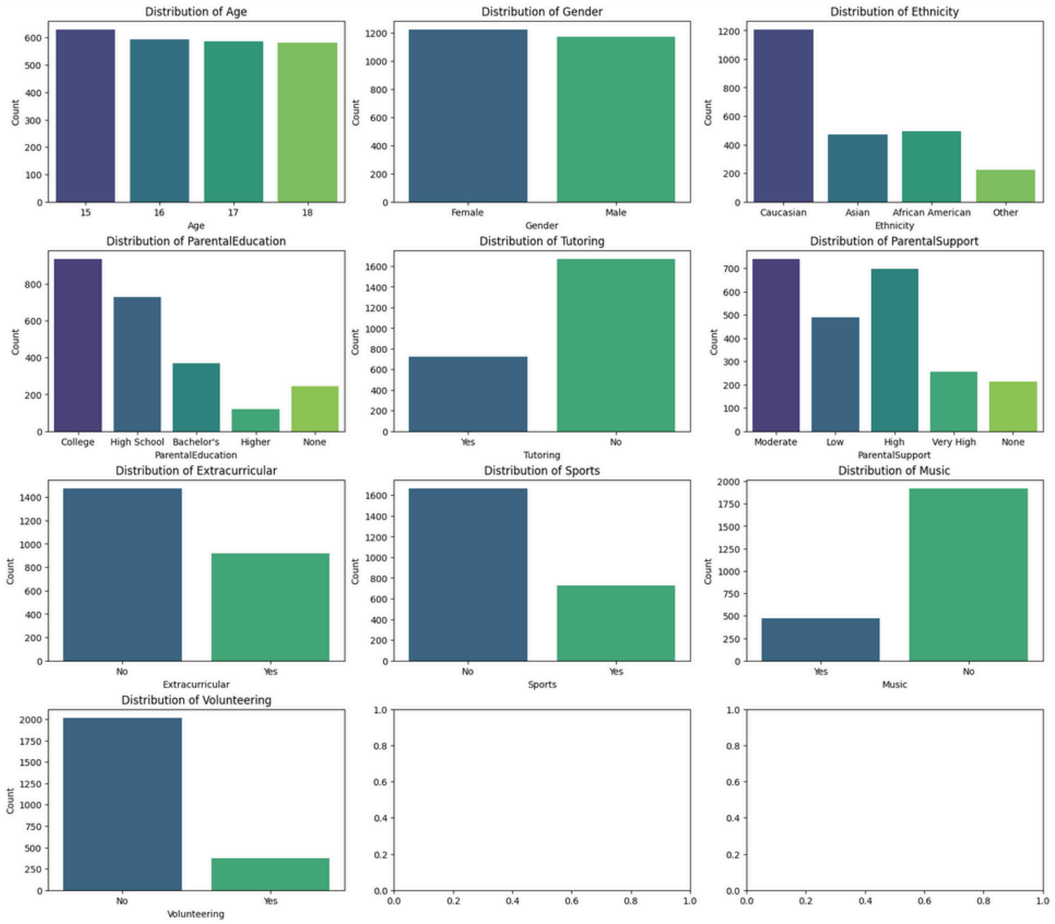
There is a small amount of outliers

VISUALIZATIONS

The Distribution of the Target Feature



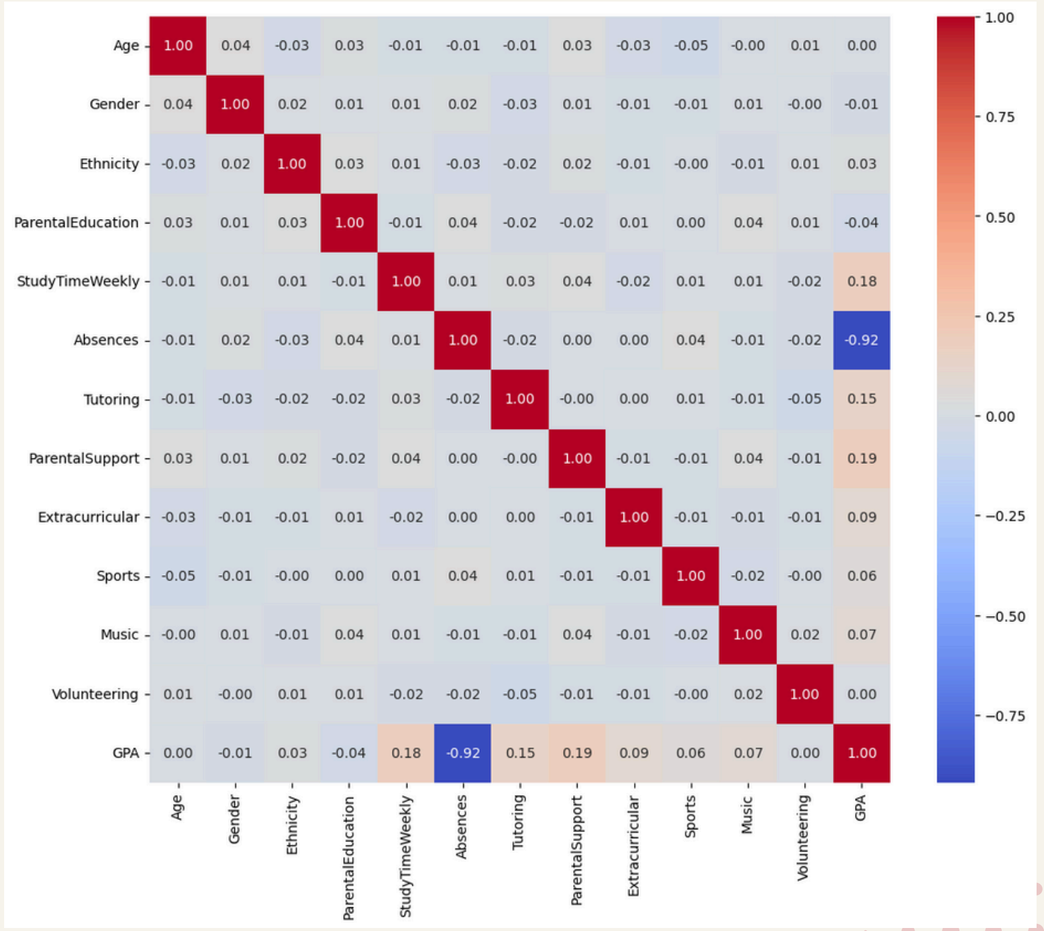
VISUALIZATIONS



The Count of Each Categorical Feature

VISUALIZATIONS

Heatmap for Detecting the Most Important Features



PREPARE FOR MODELS

```
from sklearn.model_selection import train_test_split
x = df.drop(['GPA'],axis=1)
y = df['GPA']
```

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
```

Splitting the data for train & test data

MODELS

APPLIED MODELS

- **DECISION TREE**

  The R2 Score : 85.5
  Mean Squared Error : 0.11

- **SUPPORT VECTOR MACHINE (SVM)**

  The R2 Score : 95.0
  Mean Squared Error : 0.04

- **RANDOM FOREST**

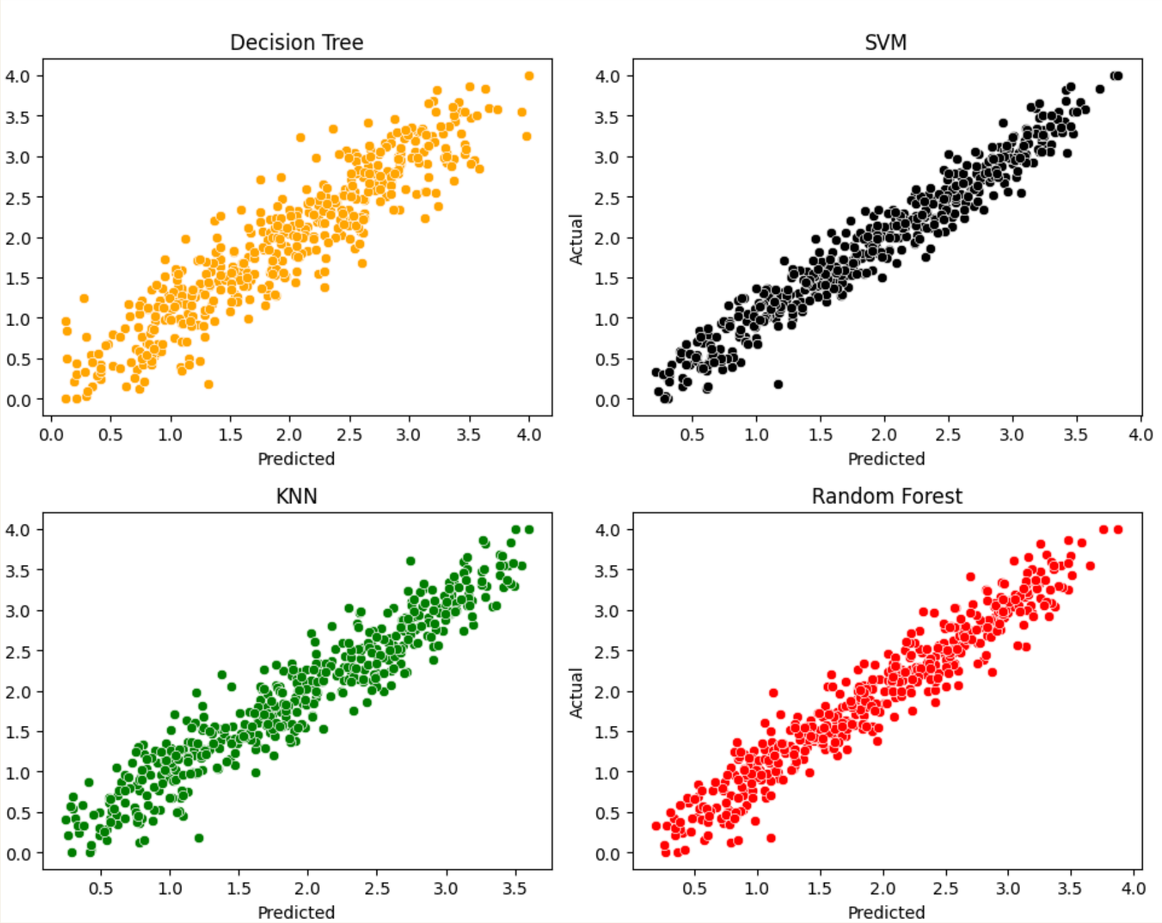
  The R2 Score : 93.0
  Mean Squared Error : 0.05

- **K-NEAREST NEIGHBORS (KNN)**

  The R2 Score : 90.8
  Mean Squared Error : 0.07

COMPARING MODELS

SCATTER PLOT



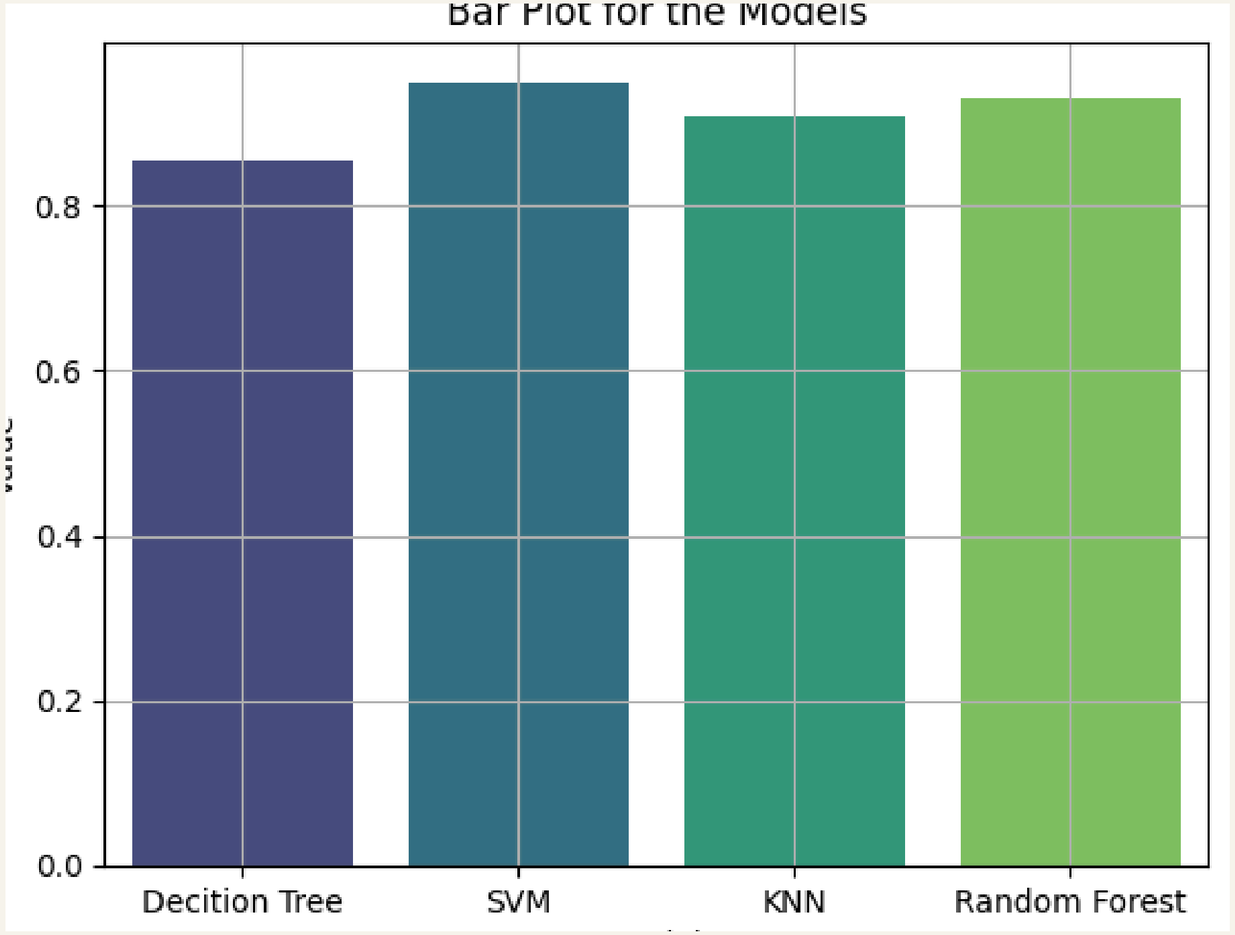
The Distribution of the Predicted and Actual data points

LOOKS LIKE SVM IS THE BEST MODEL!

Lets Make Sure

RESULT

The SVM Has the Best Accuracy



CONCLUSION

This project demonstrates the power of AI in predicting GPA and identifying at-risk students early. By leveraging data-driven insights, educators can provide targeted support, improve student outcomes, and enhance overall academic success. Ultimately, this model aims to create a more supportive and proactive learning environment.

PRESENTED BY:

1.Ziad Wael El-Gharib

2.Mahmoud Rafaat El-Alawy

3.Ahmed Abd-Elfatah Mohammed

4.Islam Mostafa Abd-Elaziz

5.Mohammed Mahmoud Fathey

THANK YOU