GPA PREDICTION PROJECT REPORT

1. Introduction

The GPA Prediction Project aims to leverage various study habits and lifestyle factors to predict the GPA of college students. By analyzing data such as class hours, attendance, and involvement in extracurricular activities, the project provides insights into which factors most significantly impact academic performance. This predictive model can be used by educators and students alike to optimize study strategies and improve academic outcomes.

2. Data Overview

The dataset used in this project includes the following key variables:

- **Attendance**: The percentage of classes attended by a student.

- **Class Hours per Week**: The number of hours a student spends in classes each week.

- **College Course**: The specific college or major the student is enrolled in (e.g., Arts, Business, Engineering, Science).

- **Online Learning**: The percentage of learning done online.

- **Extracurricular Activities**: The number of extracurricular activities a student participates in weekly.

- **Library Hours**: The number of hours a student spends in the library each week.

- **Sleep Hours**: The average number of hours a student sleeps per night.

- **Jobs**: Whether the student has a part-time job.

These variables were analyzed to predict the college GPA using Power BI's modeling capabilities.

3. Methodology

Metrics Used

The model performance was assessed using the following metrics:

- **Mean Squared Error (MSE)**: 0.02

- **Root Mean Squared Error (RMSE)**: 0.16

- **Percent RMSE**: 3.94%

- **R^2 Score**: 0.84

These metrics indicate a strong predictive capability, with an R^2 score of 0.84 suggesting that 84% of the variance in GPA can be explained by the model.

Coefficient Determination

Coefficients for each variable were calculated to understand their impact on GPA:

- **Arts**: 0.15

- **Business**: 0.14

- **Engineering**: 0.17

- **Science**: 0.15

- **Attendance**: 0.01

- **Class Hours per Week**: 0.02

- **Online Learning**: 0.07

- **Extracurricular Activities**: 0.00

- **Library Hours**: 0.00

- **Sleep Hours**: 0.00

- **Jobs**: -0.01

The constant (intercept) value was 0.61.

4. Analysis and Findings

Impact of Variables on GPA

- **Engineering and Science majors** showed a slightly higher positive impact on GPA compared to other majors like Arts and Business.

- **Online Learning** also had a positive but moderate impact on GPA, with a coefficient of 0.07.

- **Attendance** and **Class Hours per Week** had minimal but positive effects, indicating that while these factors contribute, they are not the sole determinants of GPA.

- **Extracurricular Activities**, **Library Hours**, and **Sleep Hours** showed no significant impact on GPA, while holding a part-time job had a slightly negative effect.

Visualization

The project utilized various visualizations in Power BI, including pie charts, to represent the distribution of coefficients and the predictive metrics. The pie chart showed that the Engineering coefficient contributed the highest percentage (12.61%) towards predicting GPA, followed by Science (11.65%), and Arts (11.15%).

5. Conclusion

The analysis indicates that while major choice and online learning contribute significantly to predicting GPA, other factors such as attendance and class hours also play a role, albeit a smaller one. The model successfully predicts GPA with a high degree of accuracy, as evidenced by the R^2 score of 0.84.

6. Recommendations

- **Focus on High-Impact Areas**: Students in majors like Engineering and Science should continue to leverage their course rigor as it positively impacts GPA.

- **Enhance Online Learning**: Given the positive influence of online learning, students should consider supplementing traditional study methods with online resources.

- **Optimize Study Habits**: While attendance and class hours are important, they should be complemented with effective study techniques tailored to individual needs.

For further research, it is recommended to explore additional variables such as psychological factors, study techniques, and time management to enhance the model's predictive power