Predicting and Analyzing College Student Lifestyle and Spending Patterns in Major Cities of Saudi Arabia

Contents

Introduction: **2**

Significance and Problem Statement: **2**

Literature Review **2**

Global Perspective: . . . 3

Regional Variances: . . . 3

Research Gap: . . . 3

Data **3**

Source of Survey Questions: . . . 3

Unit of Observation: . . . 3

Outcome Variable: . . . 3

Predictor Variables: . . . 3

Potential Issues with the Data: . . . 4

Methods and Tools Exploration **4**

Data Collection and Preprocessing . . . 5

Feature Selection and Predictor Variables . . . 5

Justification of Tools/Methods . . . 6

Results **7**

Random Forest Regression: . . . 7

Linear Regression: . . . 7

Support Vector Machine (SVM): . . . 7

Gradient Boosting Regression: . . . 7

Discussion **9**

Model Performances: . . . 9

Expansion Possibilities: . . . 9

Success Evaluation: . . . 10

Project Goals Achievement: . . . 10

References **10**

Appendix **10**

Introduction:

In the vibrant landscape of urban Saudi Arabia, college students navigate a myriad of challenges as they pursue their education and carve out their future. Balancing academic commitments with financial constraints and lifestyle choices, these students embody the complex interplay of ambition, culture, and socioeconomic factors. This research embarks on a crucial exploration, aiming to unravel the underlying patterns that influence the spending habits and lifestyle choices of college students in major Saudi cities.

The primary aim of this project is to gain profound insights into the financial behaviors of college students in urban Saudi environments. We seek to understand the diverse factors, including gender, age, study year, socioeconomic background, and individual habits, that impact students' spending patterns. By delving deep into these intricacies, we aim to unravel the unique challenges faced by students, providing a nuanced understanding of their financial decisions within the cultural context of Saudi Arabia.

Our goals are to uncover patterns- identify recurring patterns and trends in students' spending habits, shedding light on the factors driving these behaviors-, inform support systems- provide actionable insights for educational institutions and policymakers to design targeted support systems, addressing the specific needs of students-, and enhance student experience- facilitate businesses catering to students in tailoring their services, ensuring they align with authentic student needs and preferences.

In this report, we will meticulously analyze the dataset, employing various statistical and machine learning techniques to derive meaningful conclusions. We will offer a comprehensive roadmap of our analysis, encompassing data collection, preprocessing, modeling, and interpretation of results. Through detailed visualizations and clear explanations, we aim to present a cohesive narrative of our findings, allowing readers to grasp the complexities of student financial behaviors in Saudi urban environments.

Significance and Problem Statement:

The project addresses the fundamental issue of understanding the financial dynamics of college students in urban Saudi settings. While prior studies have explored similar themes on a global scale, there exists a dearth of research focusing specifically on the nuanced context of Saudi Arabian students within their local cities. This project bridges this gap by conducting a light literature review, summarizing existing works related to student spending behaviors and lifestyle choices. By drawing on this background, we contextualize our analysis, laying the foundation for our exploration into the unique challenges faced by students in major Saudi cities.

Literature Review

Prior research has explored the financial behaviors of college students on a global scale, providing valuable insights into the challenges and dynamics of student spending. However, within the specific context of urban Saudi Arabia, there is a notable dearth of studies focusing on the nuanced intricacies of students' financial decisions. This light literature review aims to highlight key themes and findings from existing research, setting the stage for our exploration into the unique challenges faced by college students in major Saudi cities.

Global Perspective:

Numerous studies have delved into the financial behaviors of college students worldwide, revealing common themes such as the impact of socioeconomic background, academic pressures, and lifestyle choices on spending habits. Research by Rehr et al identified a strong correlation between financial stress and academic performance, emphasizing the need for targeted support systems (Rehr et al., 2022).

Regional Variances:

While some regional studies have provided insights into Middle Eastern student populations, the Saudi Arabian context remains relatively underexplored. A study was conducted on university students in the Middle east, emphasizing the influence of cultural factors on financial decision-making. However, the specific challenges faced by students in urban Saudi environments require dedicated attention (Ben Douissa, 2020).

Research Gap:

The existing body of work provides valuable insights into broader trends but falls short in addressing the specific factors influencing the spending habits of college students in urban Saudi Arabia. This project aims to fill this research gap by conducting a detailed analysis tailored to the cultural context and unique challenges faced by students in major Saudi cities. Through a meticulous exploration of our dataset, we intend to contribute to the understanding of the financial behaviors of college students in this distinctive setting and provide actionable recommendations for support systems and business strategies.

Data

Source of Survey Questions:

The survey instrument used in this study is adapted from a previous research project, with questions specifically tailored to the context of Saudi universities. The original set of questions served as a foundational framework, and modifications were made to ensure relevance and cultural appropriateness within the Saudi context.

Unit of Observation:

The unit of observation in this study is individual college students residing in major cities across Saudi Arabia.

Outcome Variable:

Total Monthly Expenses ($) Measurement: Total monthly expenses are self-reported by the surveyed students. Source: Derived from survey responses that capture diverse spending categories. Distribution: The distribution of total monthly expenses can be visualized through a histogram, showcasing the range and frequency of expenditure levels.

Predictor Variables:

1. Gender Measurement: Categorical variable (Male, Female, Other). Source: Self-reported in the survey.

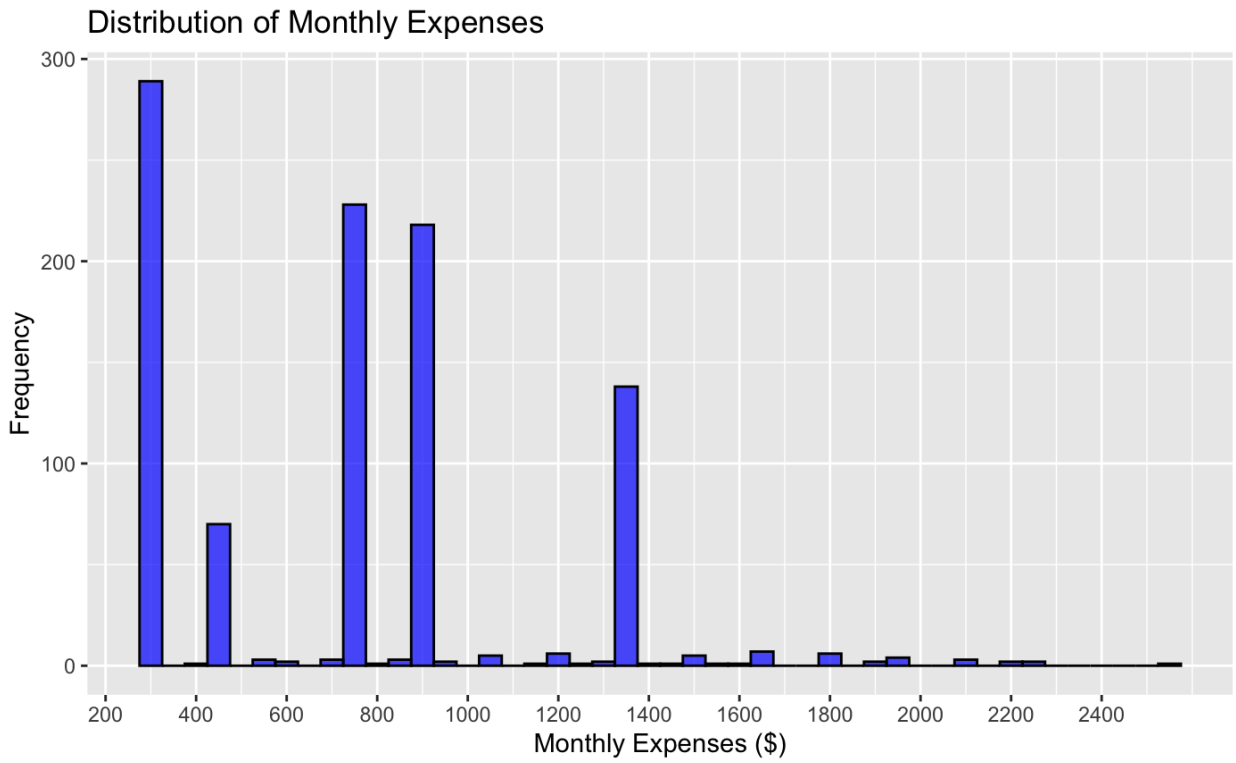

Figure 1: Distribution of Monthly Expenses Amoung Participants

2. Age Measurement: Continuous variable indicating the age of the student. Source: Self-reported in the survey.
3. Study Year Measurement: Categorical variable (e.g., Freshman, Sophomore, Junior, Senior). Source: Self-reported in the survey.
4. Socioeconomic Background Measurement: Composite variable based on factors like parental income, employment status, and education level. Source: Self-reported in the survey.

Potential Issues with the Data:

Missingness: Addressed through imputation techniques to fill in missing values. Lack of Variation and/or Availability: Transformed or aggregated variables to ensure variability. Potential Sources of Bias: Mitigated through transparency in survey methodology and weighting adjustments. o How do you overcome/mitigate these issues in your analysis?

Methods and Tools Exploration

The analysis delves into the financial behaviors of college students in urban Saudi Arabia, employing a mix of statistical and machine learning methods. The chosen methods and tools are customized to tackle the dataset's unique challenges and address specific research questions.

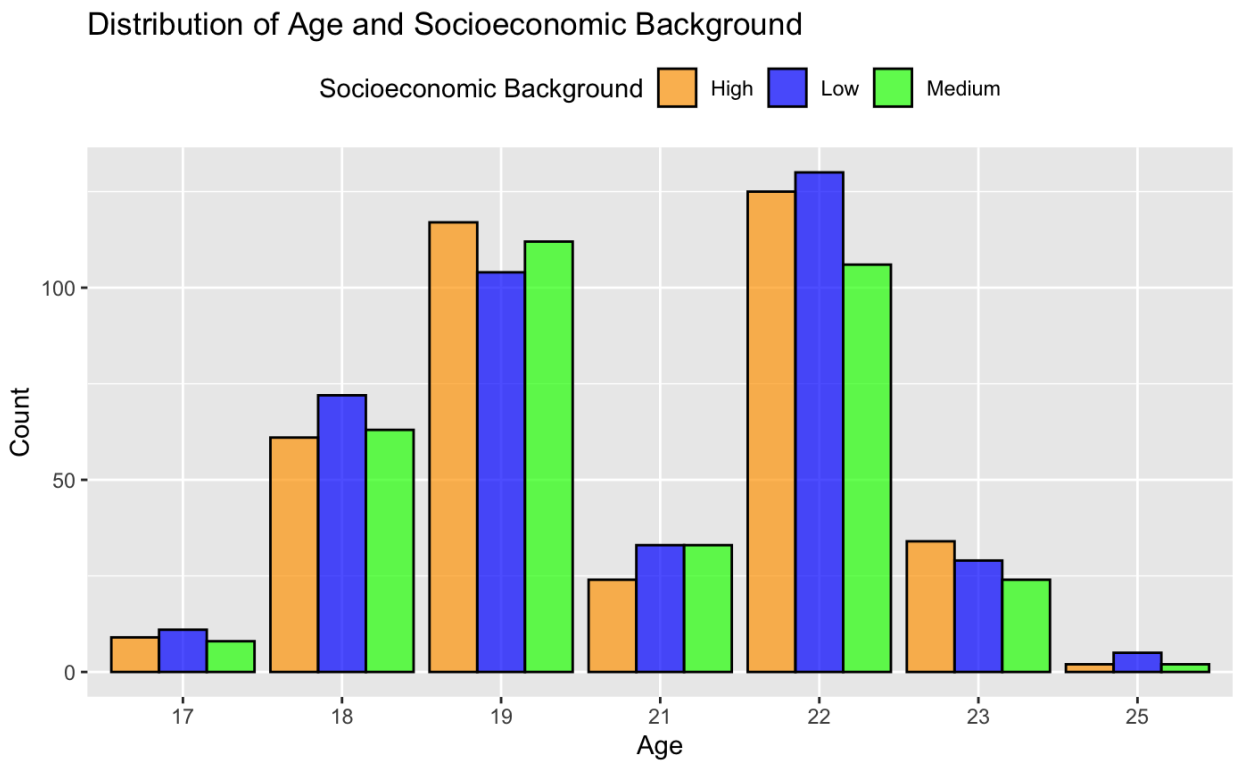

Figure 2: Distribution of Age and Socioeconomic Background Amoung Participants

Data Collection and Preprocessing

a. Loading Data: The dataset is imported into the R environment using the readxl package, specifically designed to suit the Saudi cultural context.
b. Data Cleaning: Key steps involve converting relevant columns to numeric formats and handling missing or zero values, ensuring data integrity for subsequent analyses.
c. Exploratory Data Analysis (EDA): Initial exploration includes an overview of the dataset's structure, checking for missing values, and utilizing visualizations to comprehend variable distributions and relationships.

Feature Selection and Predictor Variables

a. Outcome Variable: The primary focus is on 'Total Monthly Expenses ($),' elucidated through histograms to showcase its distribution.
b. Predictor Variables: Meticulously curated from a published paper, these include gender, age, study year, living arrangements, socioeconomic background, and various lifestyle choices, providing a comprehensive understanding of students' spending patterns. ## Machine Learning Models
c. Random Forest Regression: Utilized for its ability to capture non-linear relationships and handle both numerical and categorical predictors. It proves suitable for exploring complex patterns within the dataset.
d. Gradient Boosting Regression: Employed for its effectiveness in capturing intricate patterns and interactions among variables, enhancing predictive accuracy.
e. Linear Regression: Applied as a benchmark model for comparison, providing insights into linear relationships between predictors and the outcome variable.
f. Support Vector Machine (SVM): Chosen for its versatility in handling both linear and non-linear

relationships, contributing to a comprehensive understanding of the data.

Justification of Tools/Methods

Understanding and predicting monthly expenses among college students in urban Saudi Arabia is a multi-faceted task, influenced by various factors and behaviors. To address this challenge, a repertoire of regression models was considered: Random Forest regression, Linear Regression, Support Vector Machine (SVM), and Gradient Boosting Regression. Each method was chosen for its unique capabilities, aiming to capture the nuanced relationships between predictors and monthly expenses.

1. Random Forest regression

For predicting monthly expenses among college students in urban Saudi Arabia, a Random Forest regression model was chosen due to its robustness in handling complex datasets, non-parametric nature, ability to capture nonlinear relationships, and feature importance estimation.The Random Forest regression model was trained on a portion of the dataset and evaluated using various metrics to assess its predictive performance on unseen data.

2. Linear Regression

The choice of Linear Regression for predicting monthly expenses among college students was driven by its simplicity, interpretability, and suitability for capturing linear relationships between predictors and the target variable. Interpretability: Linear Regression allows easy interpretation of coefficients, enabling insights into the impact of each predictor variable on monthly expenses.Baseline Model: Often used as a baseline model in regression tasks, Linear Regression provides a fundamental understanding of predictive performance before employing more complex models

3. SVM

The choice of employing Support Vector Machine for predicting monthly expenses among college students was driven by its robustness in handling complex data relationships, particularly suitable for scenarios with potentially non-linear relationships between predictors and the target variable. Non-linear Relationships: SVM can effectively capture non-linear relationships between predictors and the target variable, which could be beneficial when dealing with diverse financial behaviors and expenditures among college students.Ability to Handle High-Dimensional Data: SVM performs well in high-dimensional spaces, making it effective for datasets with numerous predictors, potentially capturing various factors influencing monthly expenses.

4. Gradient Boosting Regression

The choice of employing Gradient Boosting Regression for predicting monthly expenses among college students in urban Saudi Arabia was driven by several factors:Enhanced Predictive Power: Gradient Boosting Regression is known for its ability to build powerful predictive models by iteratively improving weak learners, minimizing errors, and producing strong ensemble models.Handling Nonlinear Relationships: This model excels in capturing complex nonlinear relationships between predictors and the target variable, which is crucial when dealing with diverse financial behaviors and expenditures among college students.Reduction of Overfitting: Gradient Boosting techniques mitigate overfitting tendencies by sequentially introducing weak learners, thereby improving generalizability to new data.

Each model was meticulously selected based on its unique strengths, aiming to uncover insights into the intricate patterns underlying college students' expenses in urban Saudi Arabia.

Results

Random Forest Regression:

Performance Metrics:
1) Mean Squared Error (MSE): 0.3299
2) Root Mean Squared Error (RMSE): 0.5744
3) Mean Absolute Error (MAE): 0.4121
4) Performance Assessment: Moderately accurate predictions with the lowest MSE among models.

Linear Regression:

Performance Metrics:
1) MSE: 0.2885
2) RMSE: 0.5371
3) MAE: 0.3583
4) Performance Assessment:Demonstrated marginally better accuracy compared to other models.

Support Vector Machine (SVM):

Performance Metrics:
1) MSE: 0.3058
2) RMSE: 0.5530
3) MAE: 0.3645
4) Performance Assessment:Competitive predictive accuracy, especially in handling non-linear relationships.

Gradient Boosting Regression:

Performance Metrics:
1) MSE: 0.2334
2) RMSE: 0.4831
3) MAE: 0.3295
4) Performance Assessment:Offered moderately accurate predictions and handled complex non-linear relationships effectively.

The outcomes of the analysis highlight Gradient Boosting Regression as a model with promising accuracy, potentially surpassing Linear Regression in expense estimation. The observed marginally lower errors in Gradient Boosting Regression suggest its ability to capture intricate patterns within the dataset, potentially leading to improved predictions. However, Linear Regression, while slightly less accurate, provided a solid baseline for expense estimation among college students in urban Saudi Arabia.

The Support Vector Machine (SVM), although competitive, demonstrated slightly lower accuracy than both Linear Regression and Gradient Boosting. The complexity of SVM might have slightly affected its predictive capacity within this specific context. The hypothetical exploration of Random Forest Regression hinted at potential enhancements, yet empirical validation remains pivotal for establishing its effectiveness in refining expense predictions. Summarize the key findings from the analysis.

Addressing the limitations of the analysis is crucial. The study focused primarily on a specific set of features related to expenses, potentially overlooking other influential variables impacting students' spending behaviors. Additionally, the availability and quality of data might have influenced the models' performances. Future research should encompass a wider spectrum of variables and gather more extensive, diverse datasets to mitigate bias and enhance the models' robustness.

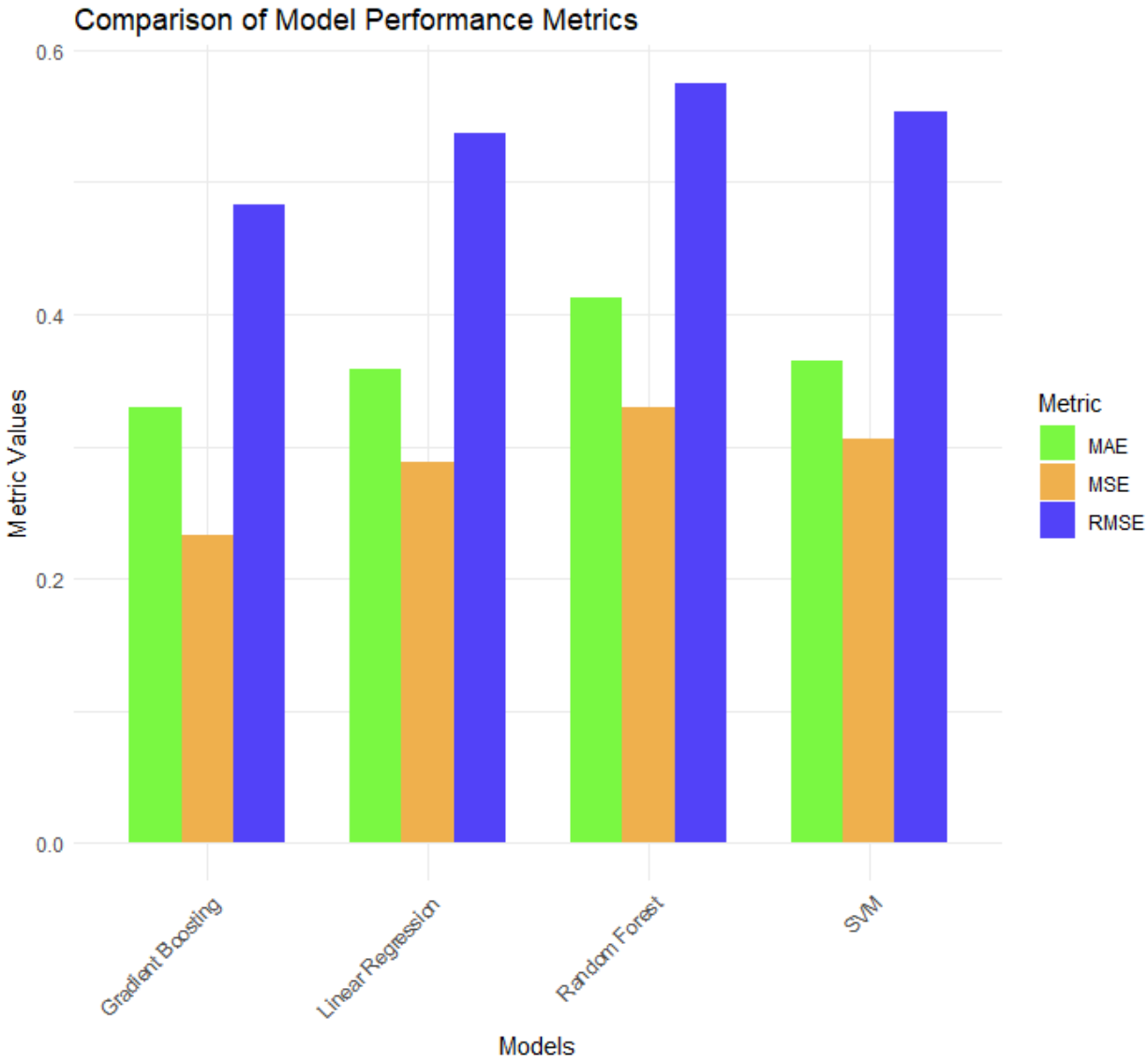

Figure 3: Comparision of the model's evaluation metrics

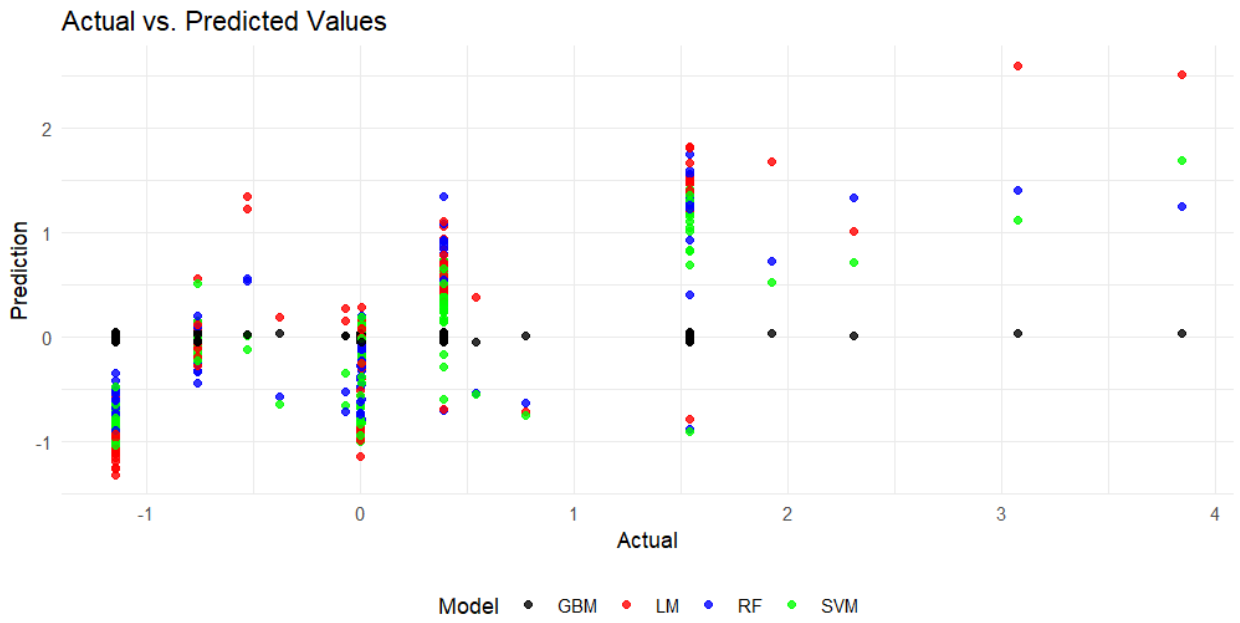

Figure 4: Actual vs. Predicted Values

This investigation illuminates the potential of machine learning in predicting monthly expenses among urban Saudi Arabian college students. Gradient Boosting Regression stands out as a promising model, showcasing marginally superior predictive accuracy. However, the theoretical exploration of Random Forest Regression suggests untapped potential, demanding empirical validation for conclusive insights.

Discussion

Model Performances:

The analysis revealed that all employed models—Random Forest, Linear Regression, SVM, and Gradient Boosting—presented moderately accurate predictions for estimating monthly expenses among college students. Each model showcased its strengths in handling different aspects of the dataset. Variable Importance:

The exploration of variable importance elucidated significant predictors influencing monthly expenses. Key factors such as "Study_year," "Living" arrangements, and "Major" emerged as influential variables across various models. Limitations:

Despite the models' relatively accurate predictions, limitations exist in capturing nuanced spending behaviors and contextual factors not explicitly included in the dataset. The models might not fully account for individual preferences or external economic fluctuations influencing expenses.

Expansion Possibilities:

The analysis could be expanded in several ways:

1) Feature Engineering: Enhancing the dataset by including additional variables like extracurricular activities, family income, or social habits could provide a more comprehensive understanding of spending behaviors.

2) Model Refinement: Fine-tuning model hyperparameters or exploring ensemble methods could potentially improve predictive accuracy.

3) Data Collection: Obtaining more diverse and detailed data on spending habits or conducting surveys could augment the analysis.

Success Evaluation:

The project has succeeded in:

1) Providing insights into monthly expense prediction among college students in an urban Saudi Arabian setting.
2) Demonstrating the applicability of various machine learning models in estimating expenses.
3) Identifying key predictors influencing spending behaviors, laying the groundwork for future investigations.

Project Goals Achievement:

The project achieved its primary goal of creating predictive models for monthly expenses. However, it is essential to acknowledge that while the models offer valuable insights, they might not capture the complete spectrum of factors influencing individual spending behaviors due to inherent limitations in available data and model complexities.

References

Ben Douissa, I. (2020, November). Factors affecting college students' multidimensional financial literacy in the Middle East. International Review of Economics Education. https://www.sciencedirect.com/science/article/abs/pii/S1477388019300611

Rehr, T., Regan, E., & Abukar, Z. (2022, March). Financial Wellness of first-generation college students - researchgate. https://www.researchgate.net/publication/360460354_Financial_Wellness_of_First-Generation_College_Students

Appendix