CHRIST
(DEEMED TO BE UNIVERSITY)
PUNE LAVASA CAMPUS
The Hub of Analytics

HACKVERSE'23

Overall Project Documentation

Team Name:
Blue Horns

Team Members:
Annet Jacob 22122011
Aleena Joby 22122303
Abhidev SP 23122001
Gokul Manoj 23122117

Email:
annet.jacob@msds.christuniversity.in
aleena.joby@msea.christuniversity.in
abhidev.sp@msds.christuniversity.in
gokul.manoj@msds.christuniversity.in

Department:
Department of Data Science

Date & Month & Year:
03RD December 2023

Index

4.1 Task Allocation Among Team Members

4.2 Collaborative Efforts and Synergy

5. Presentation of Results

5.1 Model Performance Assessment User

5.2 Interface Demonstration

6. Documenting

6.1 Documentation Quality

6.2 Work Execution and Organization

6.3 Presentation and Delivery

Project Documentation: Predictive Analysis for Quality Education

1. Introduction

1.1 Initiative Overview

The Department of Education, Government of India, stands at the forefront of a crucial investigation into the underlying factors shaping student performance—a mission deeply rooted in advancing UN Sustainable Development Goal 4: Quality Education. With meticulousness, the government has undertaken an extensive survey, amassing comprehensive data from schools that delves into the intricate web of socio-economic factors within families. This dataset, a treasure trove of insights, holds the keys to understanding the nuanced elements significantly impacting educational outcomes. By dissecting these socio-economic influences, the research initiative aims to democratize accessible insights, empowering policymakers and educators with targeted interventions. Such initiatives foster inclusive, equitable education, aligning directly with the principles of UN Sustainable Development Goal 4 and catalyzing societal advancement.

1.2 University's Role

Your esteemed university occupies a pivotal role in this initiative. Tasked with leveraging advanced analytics and machine learning, your mission is to untangle intricate patterns within this rich dataset. However, your mandate extends beyond technical prowess. Your responsibility lies in bridging the gap between complex insights and public accessibility. The goal isn't just to provide technical solutions but to render these invaluable insights easily understandable and usable by the general public. This approach contributes significantly to the democratization of education data, aligning seamlessly with the principles of SDG 4.

1.3 Objective and Evaluation Criteria

A prototype demonstrating diverse insights and the utilization of predictive models is anticipated by the government. The evaluation criteria encompass a holistic assessment, focusing on the approach to solve the problem, analysis from available data, model selection and performance, presentation of findings through a user-friendly interface, and the overall professionalism and standard of work.

2. Approach to Solve the Problem

2.1 Understanding the Dataset

Overview of the dataset attributes and their descriptions.

Initial data exploration techniques used to comprehend the dataset's characteristics.

2.2 Feature Analysis

Examination of the features' potential impact on educational outcomes.

Statistical methods or visualizations employed to identify key influencing factors.

2.3 Model Building and Comparison

List of models experimented with (Linear Regression, Bayesian Ridge, Random Forest Regressor, GradientBoosting Regressor, Decision Tree Regressor, Ridge Model, Lasso Model).

Evaluation metrics used for model comparison (Mean Absolute Error, R-squared value).

Selection of the best-performing model (GradientBoosting Regressor) based on performance evaluation.

2.4 User Interface Development

Role of UI in making insights accessible to the wider public.

Description of the website developed using HTML, CSS, and Flask.

2. Findings and Analysis

3.1 Impact of Features on Education Quality

- Identification of Key Predictive Variables

Your analysis scrutinized the dataset to discern the features significantly affecting the quality of education. Through statistical measures and data visualization techniques, certain variables emerged as pivotal. Factors like parental education (Medu and Fedu), study time, failures, and absenteeism showcased strong correlations with students' academic performance.

- Influence of Socio-Economic Factors

Examining socio-economic variables like parental occupations (Mjob, Fjob), family size (famsize), and home location (address) revealed nuanced influences on educational outcomes. For instance, students with parents in teaching or healthcare-related professions tend to exhibit higher academic success rates. Moreover, an interesting observation emerged regarding the impact of urban/rural residency on academic performance.

3.2 Social and Parental Influence

- Family Dynamics and Student Performance

The exploration extended to understanding the role of family dynamics and social life on education quality. Quality of family relationships (famrel), free time (freetime), and the frequency of going out (goout) exhibited correlations with academic performance. Moreover, the presence of a romantic relationship (romantic) and its implications on educational outcomes was a noteworthy discovery.

- Parental Impact on Education

The analysis unveiled how parents' jobs (Mjob, Fjob) correlate with a student's academic performance. Additionally, the education levels of parents (Medu, Fedu) displayed a substantial influence on their children's educational attainment, underscoring the intergenerational impact of education.

3.3 Insights on External Factors

- Alcohol Consumption and Health

The examination extended to unconventional factors. Surprisingly, a correlation was observed between alcohol consumption (Dalc, Walc) and student performance, highlighting the potential impact of lifestyle choices on educational outcomes. Furthermore, students' health status (health) showcased a relationship with academic success, emphasizing the holistic nature of student well-being.

- Attendance and Grades

An intriguing finding emerged concerning the relationship between absences and grades. Higher rates of absenteeism correlated with lower academic performance, emphasizing the significance of regular attendance in educational success.

4. Responsibilities and Collaborative Efforts

4.1 Task Allocation Among Team Members

a) Aleena and Annet's Focus: Data Analysis and Model Building

Aleena and Annet spearheaded the data analysis phase, meticulously exploring the dataset's intricacies. Their responsibilities encompassed:

- Conducting comprehensive exploratory data analysis (EDA) to understand the dataset's nuances.
- Identifying key features and variables significantly impacting educational outcomes through statistical methods and visualization techniques.
- Experimenting with various machine learning models, including Linear Regression, Bayesian Ridge, Random Forest Regressor, Decision Tree Regressor, Ridge Model, and Lasso Model, to derive the most accurate predictive model.
- Comparing model performances based on evaluation metrics such as Mean Absolute Error (MAE), R-squared value, and selecting the GradientBoosting Regressor due to its superior performance.

b) Abhidev and Gokul's Focus: User Interface Development

Abhidev and Gokul spearheaded the development of a user-friendly interface, ensuring the democratization of insights for the wider public. Their responsibilities included:

- Creating a website utilizing HTML, CSS, and Flask to present the findings and predictions derived from the selected predictive model.
- Designing an intuitive interface that enables easy navigation and comprehension of complex educational insights for diverse audiences.
- Integrating the model into the website, facilitating user interaction and providing a seamless experience for accessing educational predictions based on various factors.

4.2 Collaborative Efforts and Synergy

- Team Synergy and Coordination

While distinct roles were assigned, the team operated in synergy, fostering seamless communication and collaboration. Regular meetings and progress updates ensured that insights from data analysis seamlessly integrated into the UI development phase. This cohesive approach allowed for a comprehensive solution that harmoniously combined technical accuracy with user accessibility.

- Feedback Loops and Iterative Improvements

Constant feedback loops were established between the data analysis and UI development teams. Insights derived from data analysis guided the design of the user interface, ensuring the presentation of meaningful and impactful information. This iterative process allowed for continuous improvements and refinements, resulting in an integrated and cohesive final product.

5. Presentation of Results

5.1 Model Performance Assessment

- Evaluation Metrics

The selected GradientBoosting Regressor model underwent rigorous evaluation based on established metrics. These metrics include Mean Absolute Error (MAE) and R-squared value. The MAE of 0.65 signifies the average absolute difference between predicted and actual values, showcasing the model's accuracy. The high R-squared value of 0.91 indicates the model's excellent fit to the data, portraying its predictive capability.

- Justification for Model Selection

The rationale behind choosing the GradientBoosting Regressor among the array of models experimented with was its superior performance in accurately predicting educational outcomes. This model demonstrated the lowest error rate and the highest explained variance, making it the optimal choice for reliable predictions.

5.2 User Interface Demonstration

- Accessibility of Insights

The user interface developed by the team provides a seamless platform for disseminating insights derived from the predictive model. Users can interactively explore and understand how various socio-economic and lifestyle factors impact educational outcomes. The intuitive design ensures that complex insights are presented in an easily understandable manner.

- Functionality Overview

The website created using HTML, CSS, and Flask houses the predictive model, enabling users to input different parameters and receive predictions regarding student performance. It offers an interactive experience, allowing users to explore the potential effects of altering different variables on educational outcomes.

- Demonstration of Use Case

A walkthrough of the user interface illustrates how users, including policymakers, educators, and the general public, can access and utilize the predictive model's insights. This demonstration showcases the practicality and user-friendliness of the interface, aligning with the goal of democratizing education data.

6. Documenting

6.1 Documentation Quality

- Clarity and Coherence

The documentation maintains a coherent structure, ensuring that each section flows logically and is interconnected. Clear headings and subheadings facilitate easy navigation and comprehension. The language used is concise, avoiding unnecessary jargon to enhance accessibility.

- Completeness and Detail

Every aspect of the project, from problem statement to solution, is comprehensively covered. Detailed descriptions of methodologies, findings, and analyses provide a comprehensive understanding of the project's depth and breadth. Additionally, references and attributions are included where applicable to uphold academic and professional integrity.

6.2 Work Execution and Organization

- Systematic Approach

The project follows a systematic approach from data analysis to model building and user interface development. Clear delineation of tasks, as outlined in the responsibilities section, ensures efficient progress and avoids redundancy.

- Collaboration and Coordination

Efficient communication channels and regular meetings among team members ensure seamless coordination. Feedback loops are established, fostering an environment of continuous improvement and collaborative problem-solving.

6.3 Presentation and Delivery

- Professionalism in Presentation

The final documentation is presented professionally, incorporating suitable visual aids such as graphs, tables, and diagrams to enhance understanding. Consistent formatting, proper grammar, and accurate data representation contribute to the overall professionalism.

- Adherence to Deadlines and Guidelines

The project adheres strictly to deadlines, ensuring timely completion of tasks and milestones. It also complies with all specified guidelines, meeting the expectations set forth in the project's requirements and evaluation criteria.