Next-Gen Diabetes Management:

Pioneering Solution

Overview

- This project aims to investigate and analyse various factors related to diabetes management and health outcomes. The dataset consists of several important variables related to diabetes, lifestyle choices, health conditions, and socioeconomic factors. By conducting a thorough examination of these variables, the project aims to identify potential patterns, correlations, and insights that can contribute to enhanced diabetes care and overall health improvements.

Business Understanding

- A health insurance company wants to identify individuals at risk of developing diabetes or prediabetes based on the Behavioral Risk Factor Surveillance System (BRFSS) survey data for the year 2015. The company aims to use this information to design targeted prevention and intervention programs to reduce the risk of diabetes and improve the overall health of their policyholders.

Business Objectives

1. To improve diabetes care, the project aims to analyse the dataset and identify key factors that can lead to better diabetes outcomes and overall health for patients.

2. To determine personalized treatment plans for diabetic patients based on identified patterns and correlations in the dataset.\

3. To develop a robust predictive model that accurately estimates the probability of individuals contracting diabetes.

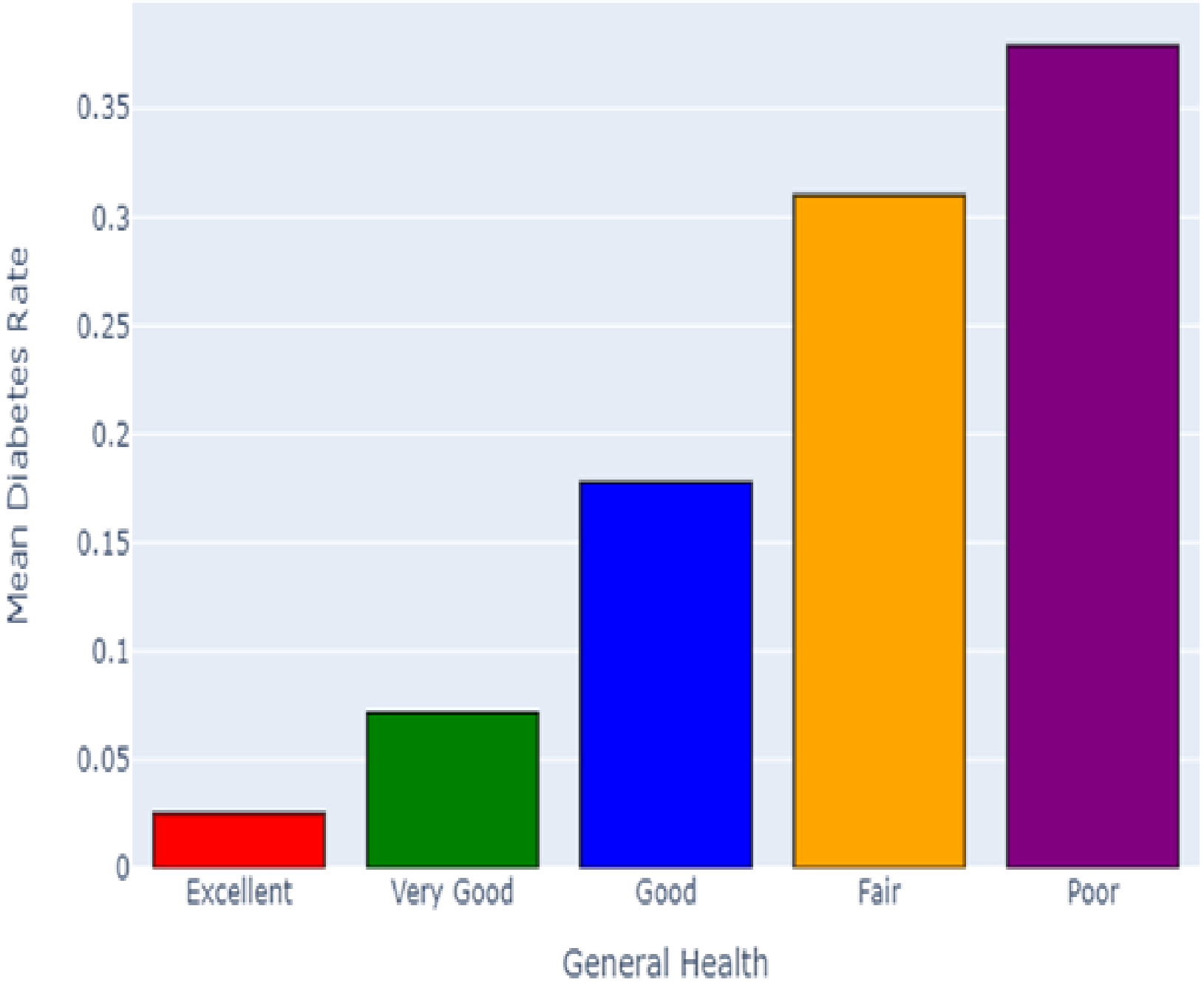
Data Understanding

- “diabetes _ binary _ health _ indicators _ BRFSS2015.csv” is a clean dataset of 253,680 survey responses to the CDC's BRFSS2015; obtained from Kaggle.
- The target variable “Diabetes binary” has 2 classes.
- 0 represents no diabetes, and 1 represents prediabetes or diabetes.
- This dataset has 21 feature variables and is not balanced.

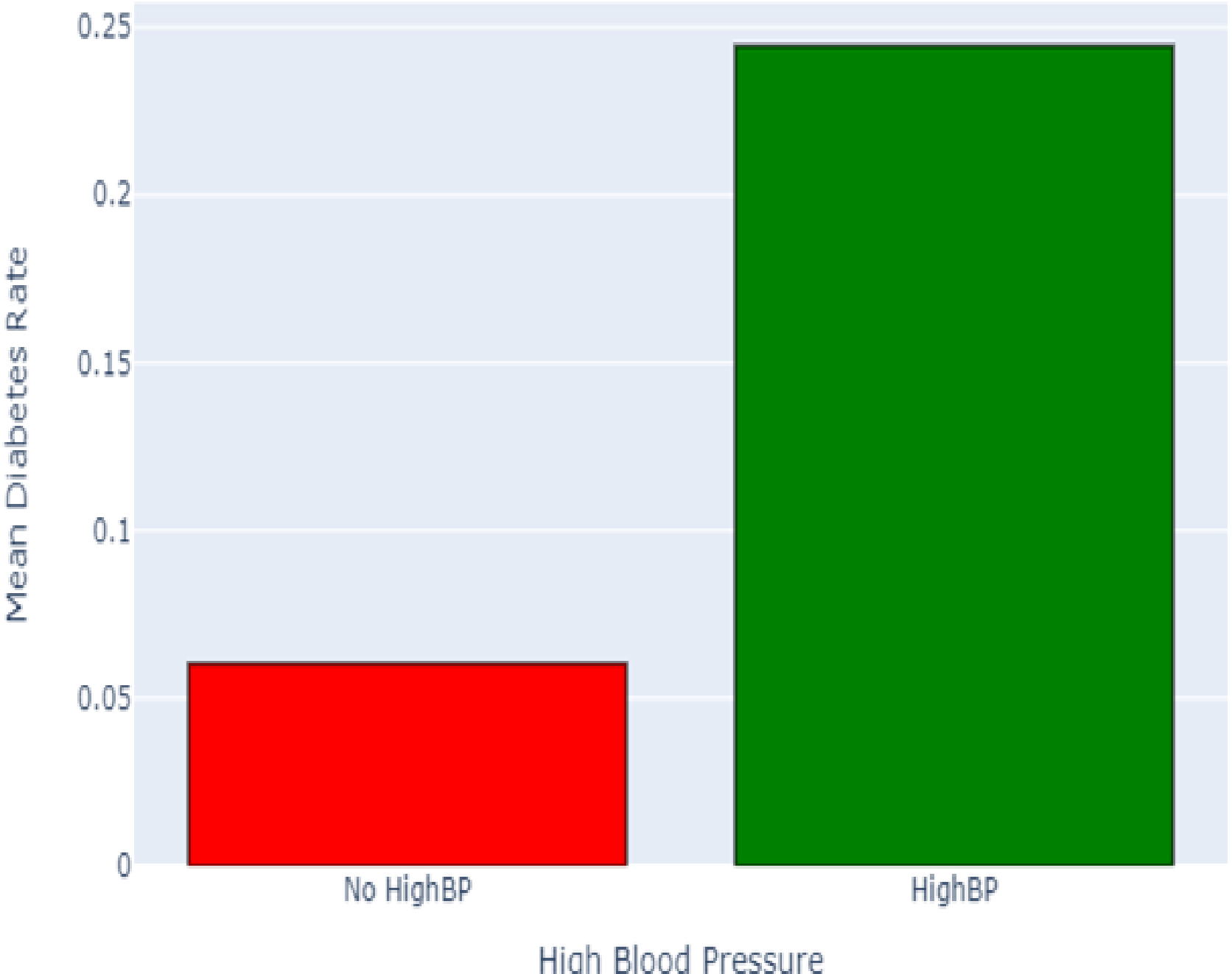
Exploratory Data Analysis

• Features that were most important in predicting whether someone has diabetes or not included:

- General Health
- High Blood Pressure
- Age
- BMI
- Income
- High Cholesterol

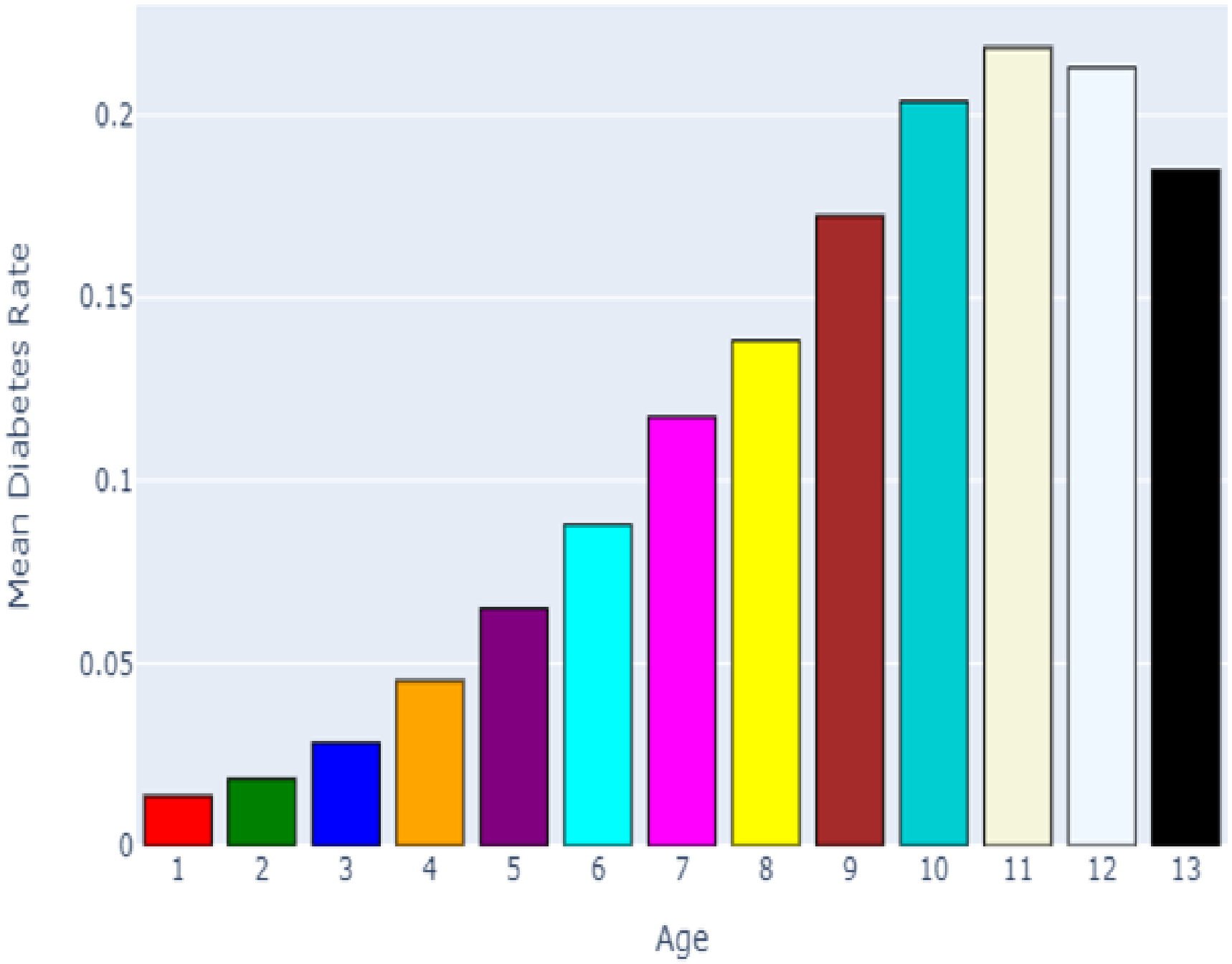


From this we can tell that someone with poor general health stands a higher chance of contracting diabetes compared to someone with excellent health



A person who has high blood pressure is at a higher risk of contracting diabetes compared to a person with normal blood pressure.

This graph reveals a notable trend: as age increases, the risk of developing diabetes also rises significantly. The data unmistakably indicates that older individuals are more vulnerable to contracting diabetes.



MODELING

We performed an iterative approach to modeling.

Models were evaluated based on their ROC-AUC values:

- Logistic Regression – 80%
- Decision Tree – 69%
- Random Forest – 78%
- XGBoost – 81%

Best Model:

- Our best performing model was XGBoost;
- With a discrimination ability score of 81%, our model effectively distinguishes between diabetes contraction and non-contraction based on input features.
- Important features provide insights on the key factors contributing to high diabetes contraction and improving model predictions.

Recommendations

1. Poor general health increases diabetes risk
2. Promote healthy lifestyles: exercise, balanced diet, stress management
3. Address high blood pressure through screening and management
4. Target diabetes screening in older age groups
5. Address socioeconomic disparities in healthcare and nutrition Manage cholesterol levels through diet and activity
6. Raise awareness through public campaigns and education

Next Steps

1. Refine the Model: Although the XGBoost model performed well, continuous refinement and fine-tuning of the model can lead to even better predictions and insights. Explore other advanced machine learning algorithms and ensemble methods to improve model accuracy and robustness.

2. Risk Stratification: Develop a risk stratification system to identify individuals at high risk of developing diabetes based on their demographic, health, and lifestyle factors. This will enable targeted interventions and personalized care plans for individuals at different risk levels.

Thank You!