CardioCare

A Personalised Heart Disease Risk Assessment and Lifestyle Improvement App

Ollie Leach

BrainStation Data Science Program

Project Pipeline

- Define the problem statement
- Source the data

- Clean the data.
- Perform EDA

- Model the data

- Create a minimum viable product

Problem Statement & Business Value

- Cardiovascular disease is the leading cause of death worldwide.
- In the USA alone, CVD is estimated to cost **$555 billion** per year.
- On average, early intervention has a **700% return** on investment in the long term.
- Early intervention could come in the form of an app.

Problem Statement & Business Value

- What are the Key Indicators of heart disease?
- How do these Key Indicators contribute to CVD risk?
- How can we use this to provide personalised, actionable advice?

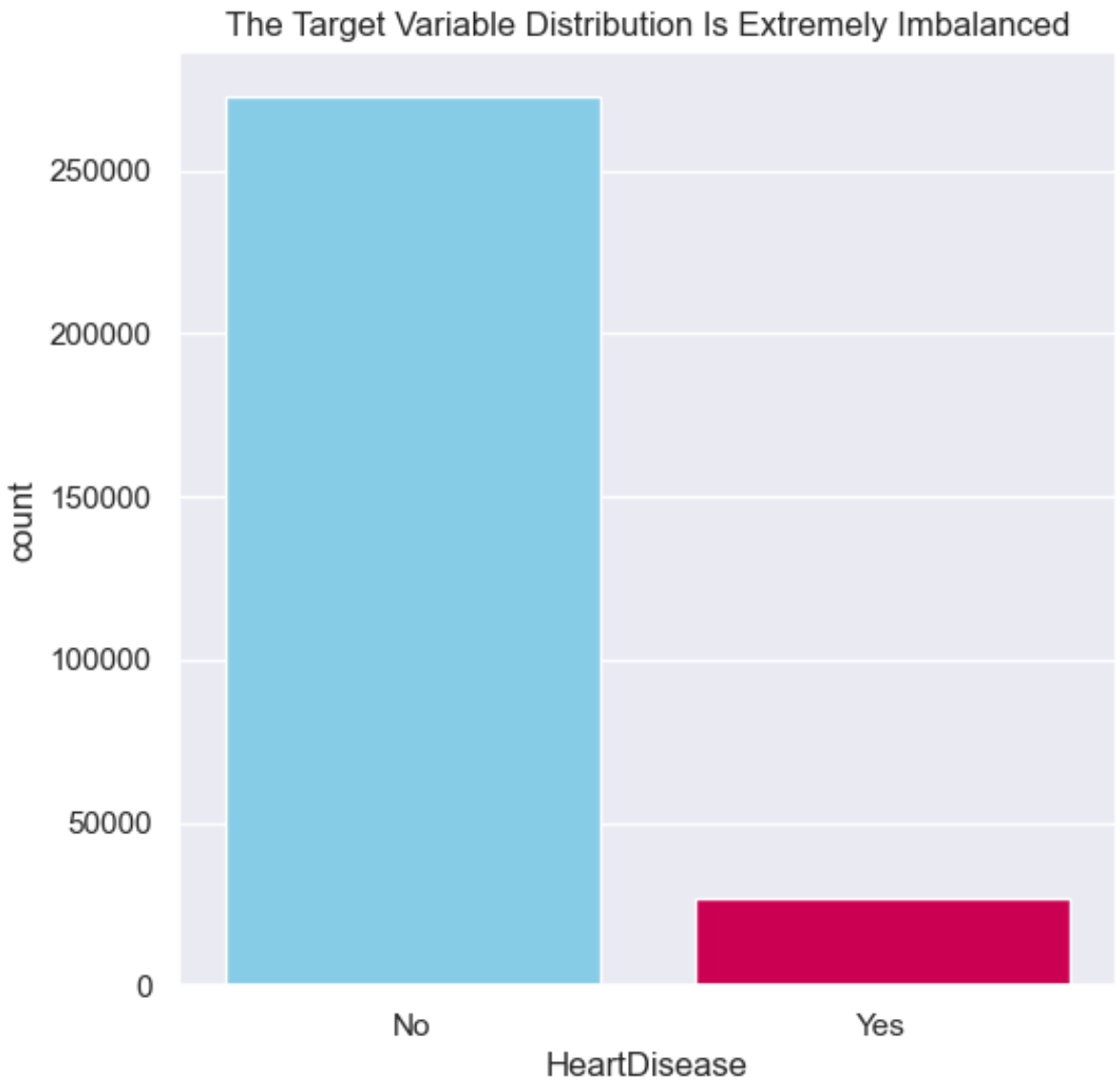
The Data Source

- 18 Columns
- 300,000 Rows
- 5,400,000 Data points

Exploratory Data Analysis Lifestyle Data

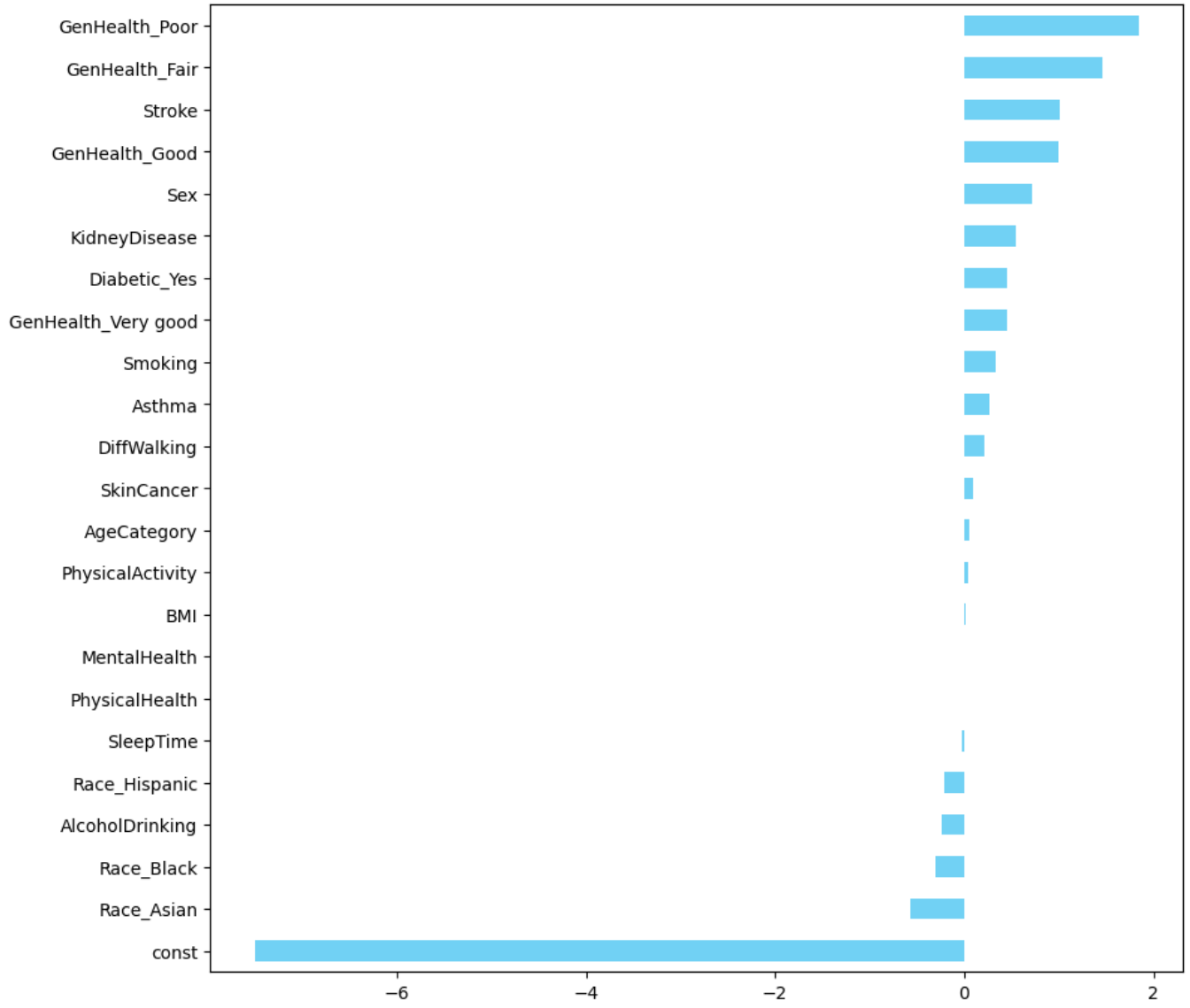
Key Takeaways:

- All EDA was congruent with medical literature.
- The dataset is very **imbalanced**.



Baseline Logistic Regression Model Results

- **Train Accuracy = 91.2%**
- **Test Accuracy = 91.2%**



…The dataset is very imbalanced

- Only 9% of our data contains rows with positive cases of CVD.

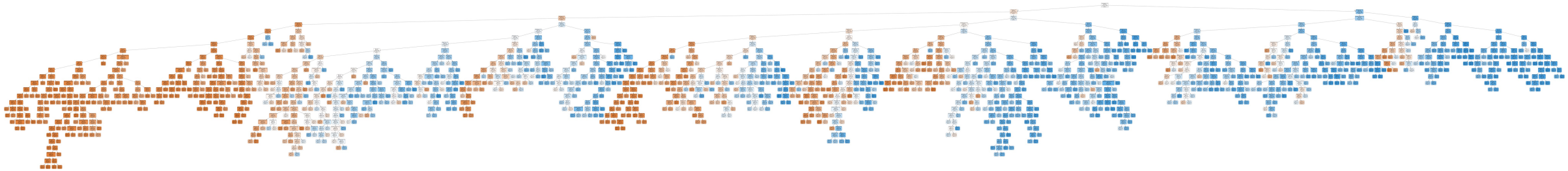
- Results:
- Precision Score for baseline model V4: **54.3%**
- Recall Score for baseline model V4: **10.4%**

Model Optimisation

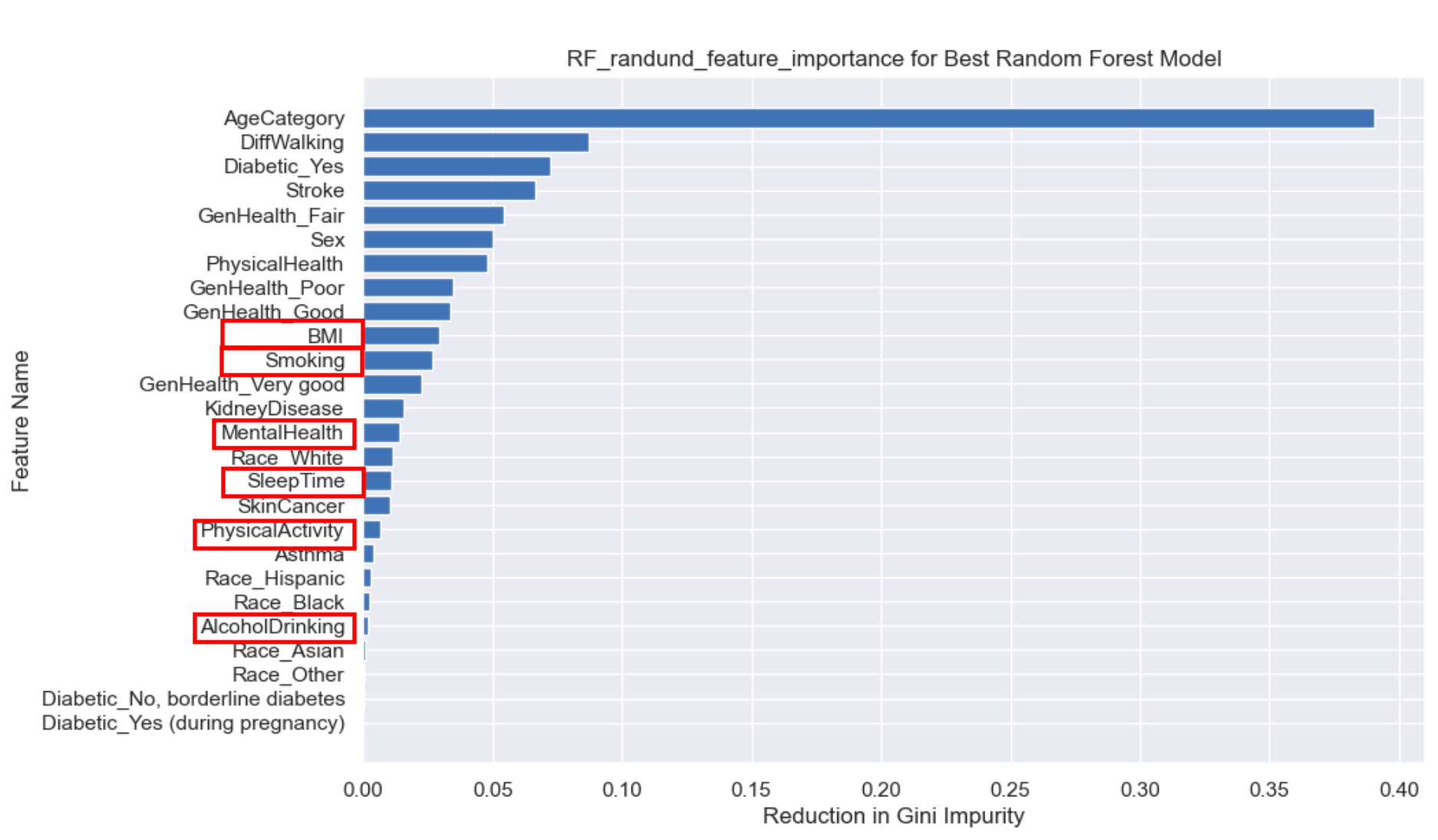
Select	Select interpretable algorithm, Logistic Regression, Decision Trees, Random Forest.
Try	Try different sampling techniques to address class imbalance
Optimise	Optimise by tuning hyperparameters.

Best Candidate Model

- Model Confidence: **84%** | Test Accuracy: **72.3%**
- Recall: **80.9%** | Precision: **21.9%**



Feature Importance



See the README file for a link to the demo.