Heart Disease Prediction

by Korolis Alexandros
Avramidis Theofilos

Introduction

About us

Part of Data Science department of National Health Association of USA

Purpose

Predict whether a person has heart disease or not based on specific features

Table of Contents

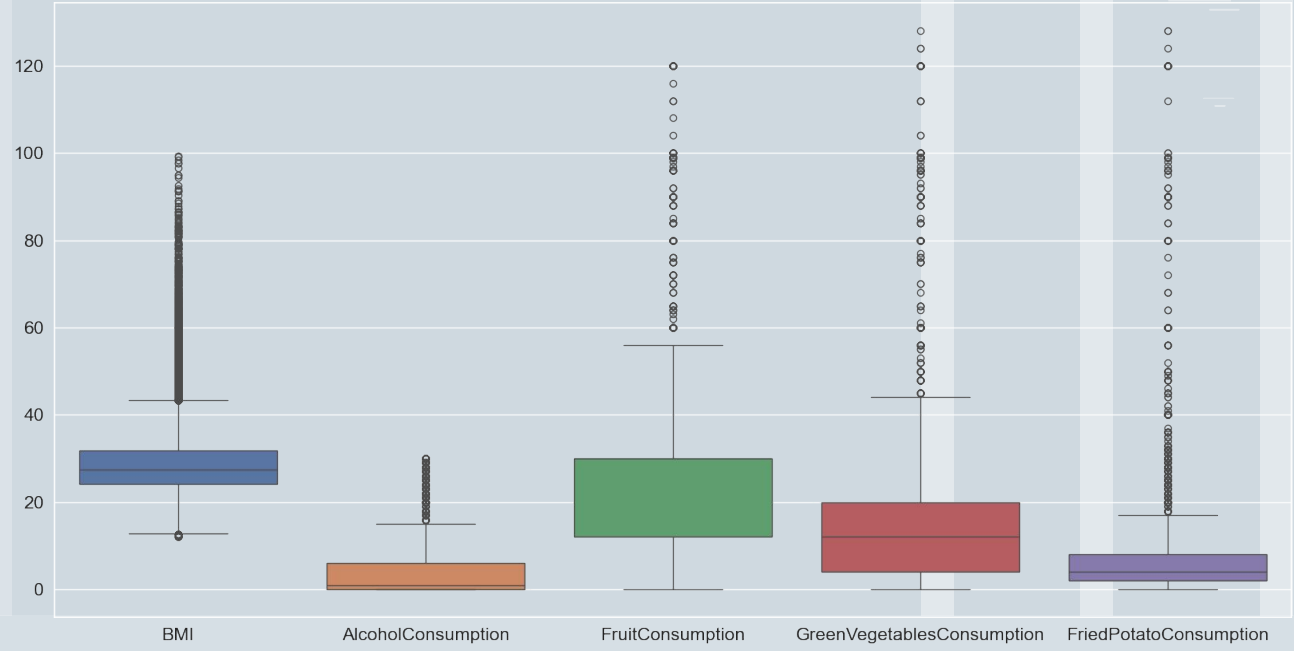
01 Exploratory data analysis (EDA)

02 Model Selection and Feature selection

03 Conclusion

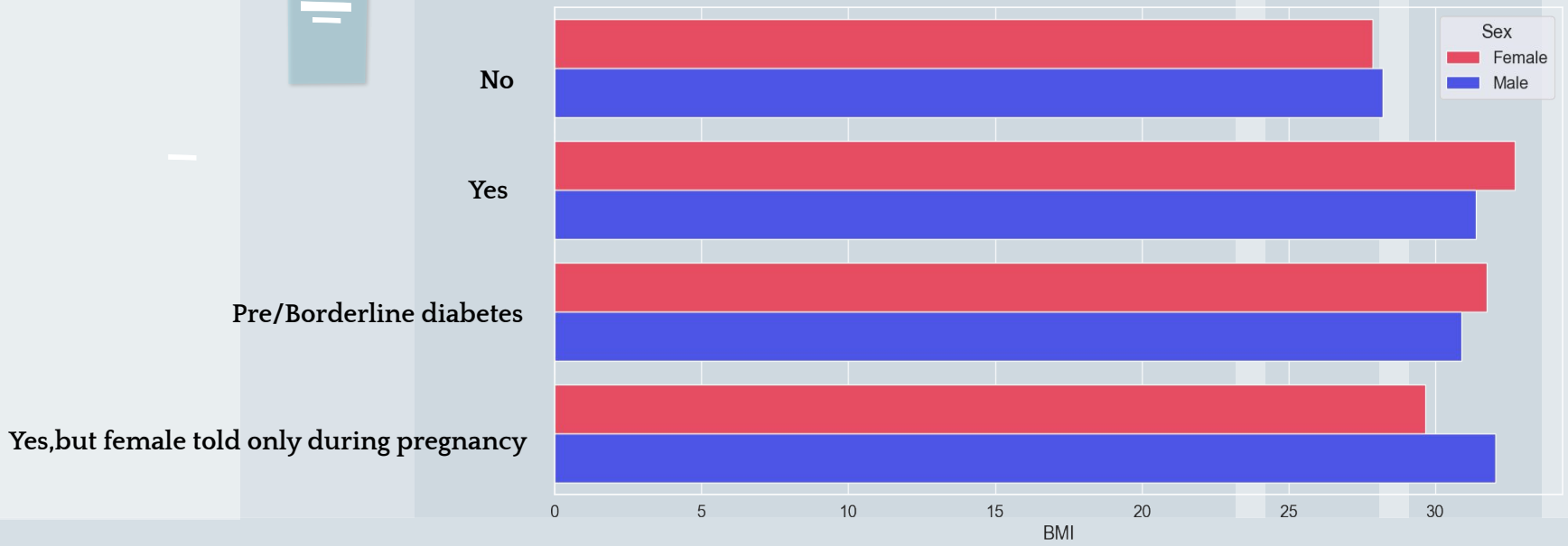
EDA

- Big dataset - **300.000 observations**
- **A lot of extreme values**



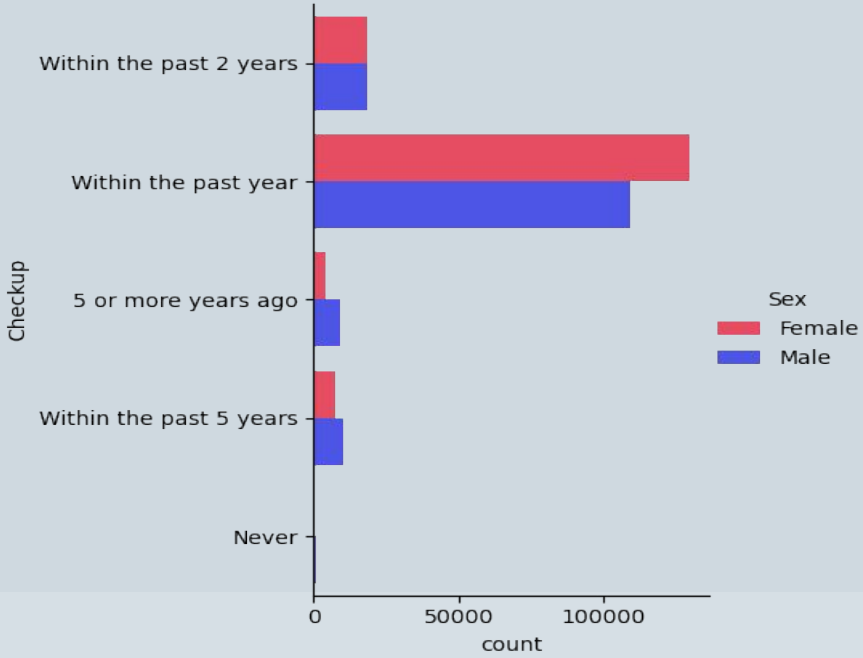
EDA

- **Females** with **Diabetes** had higher **BMI**

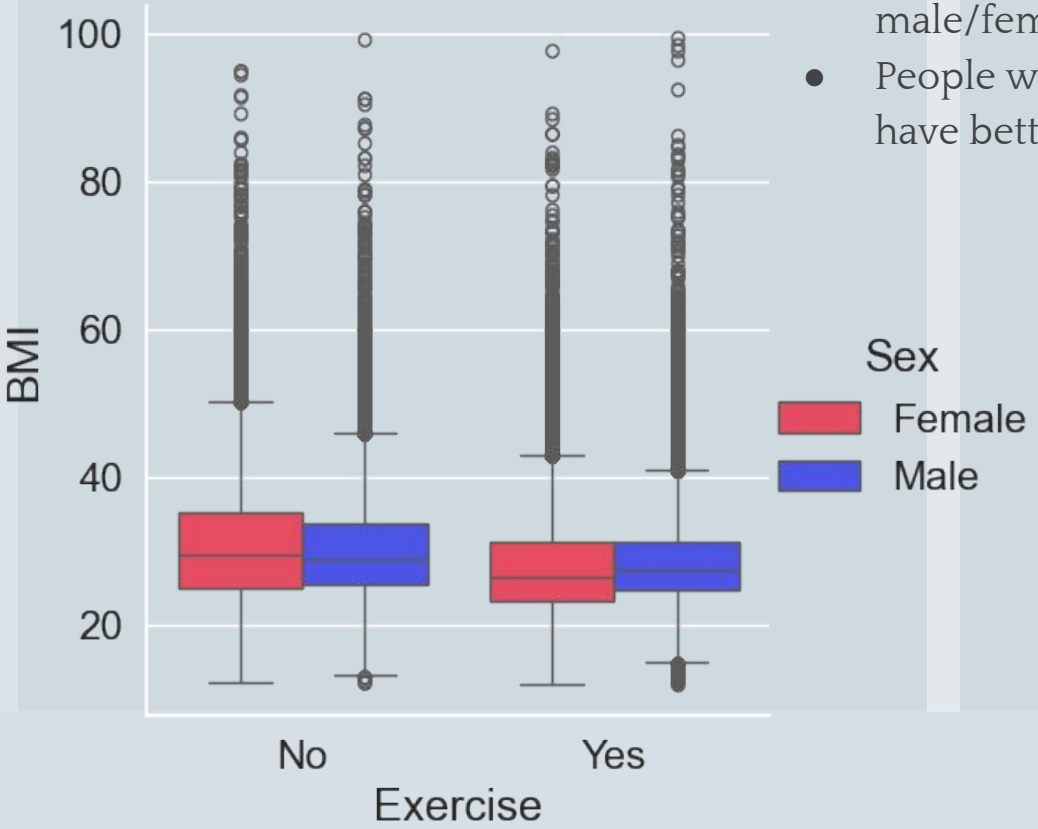


EDA

- **Checkups** based on sex
- Most checkups were **within the past year**
- **Females** do checkups more often



EDA



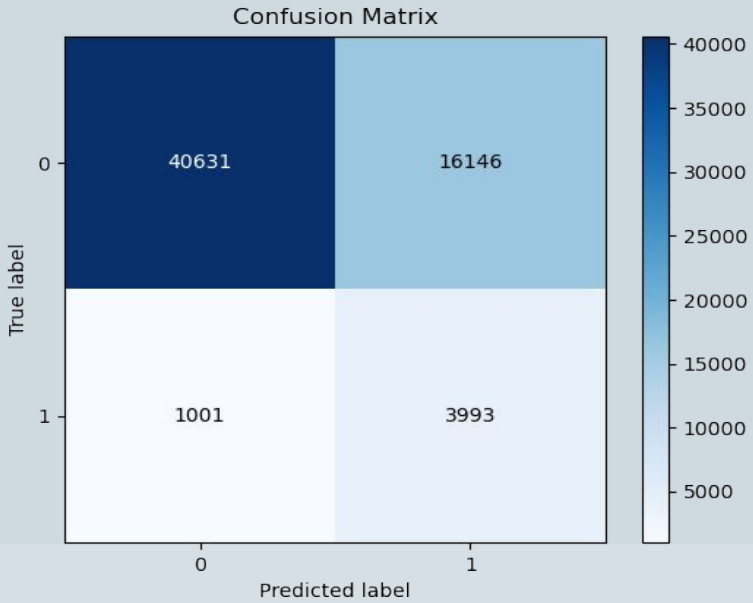
- **BMI** based on whether male/female exercise
- People who **exercise** tend to have better **(lower) BMI**

Model Selection

- Used **80%** of observations for training and **20%** to test the model.
- Selected the **Light Gradient Boosting Machine Classifier**.
- From the ones that had heart disease, we **predicted right 81%** on them.
- Prediction was made considering all of our initial features.

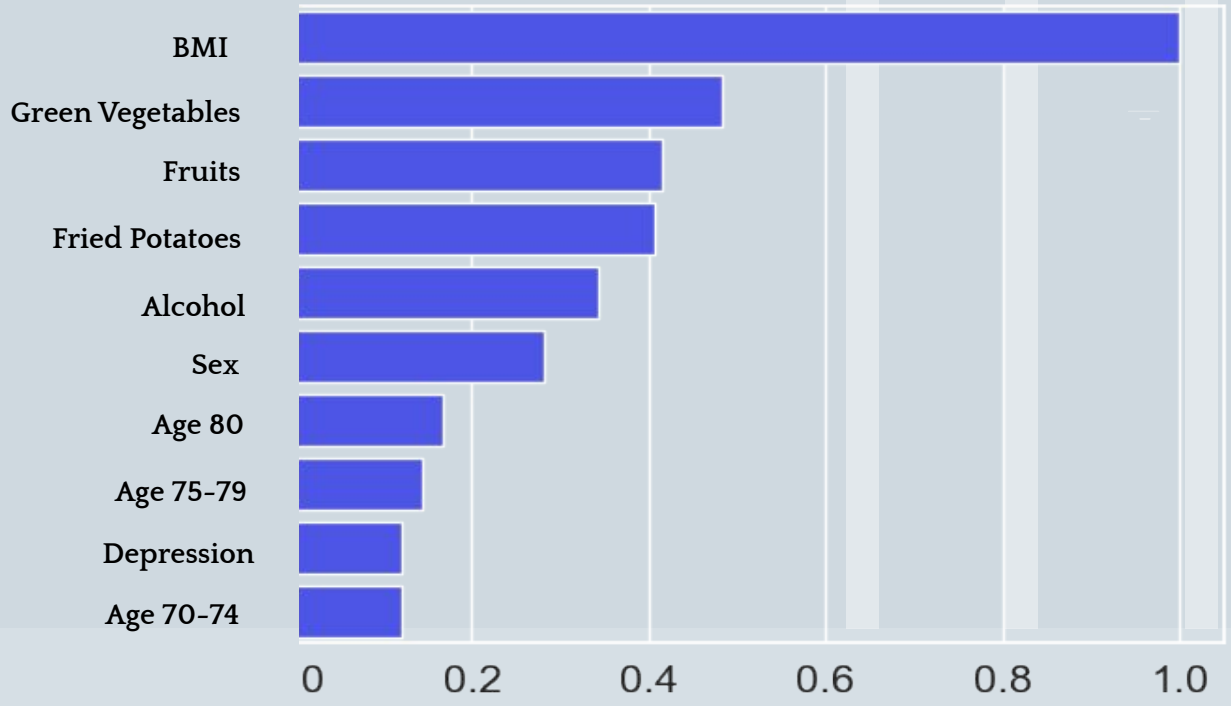
Model Selection

- **Model Selection** based on minimizing **false negative** cases (meaning we predicted wrong that someone did not have heart disease)



Feature Selection

- **Top 10 features** that are contributing the most to our predictions.



Conclusion

- From the ones that had heart disease, we **predicted right 81%** on them.
- Based on the top 10 features that our model generated we conclude :

1. Consumption of fried potatoes

2. Avoid big amounts of alcohol consumption

1. Exercise

2. Eat more green veggies

3. Check up once per year

THANKS