METIS®

Heart Disease Patients Classification

predictive model

METIS Data Science and Machine Learning Bootcamp

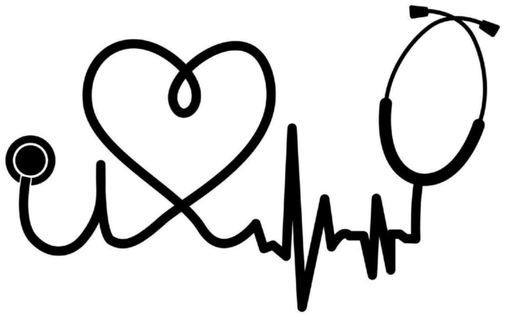
by Krystian Krystkowiak, 2022

Introduction

METIS®

- **Early heart disease detection** is important for **medical** institutions
- **Insurances**, medical **apps**, **fitness** or **nutrition** professionals and **individuals** conscious about their health

- GOAL: Develop a **predictive model** to aid in **early detection** of heart disease and raise red flags **during initial questionnaires**

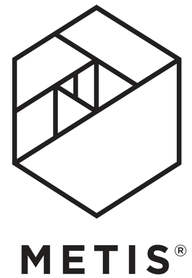


Methodology

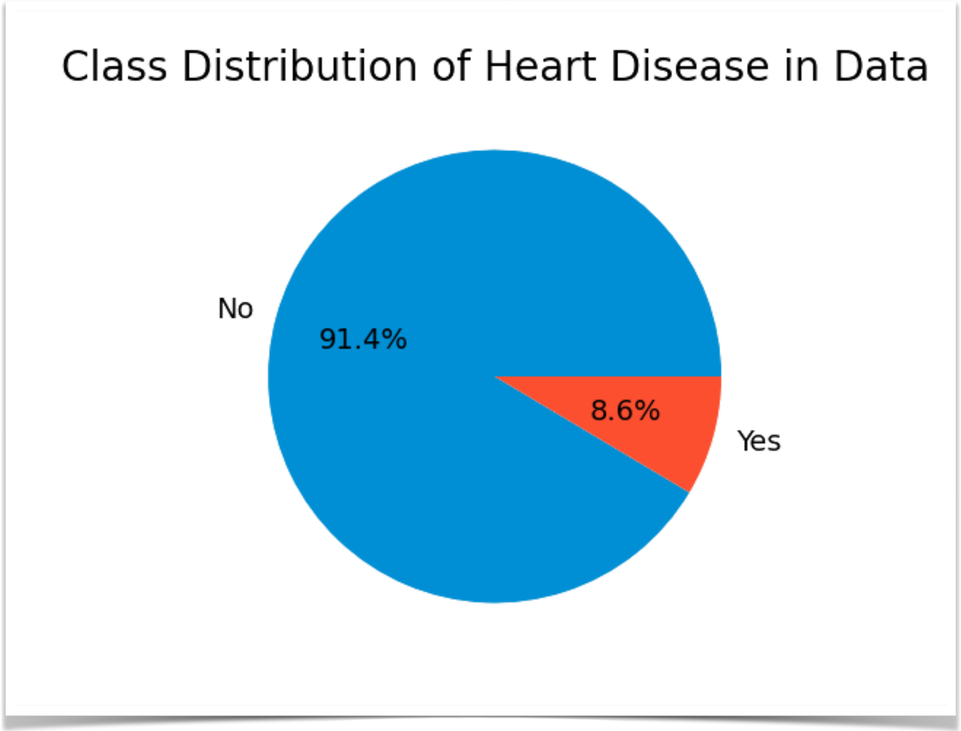
METIS®

- Data from the **Behavioral Risk Factor Surveillance System**, an annual survey that gathers information on the health status of U.S. residents
- Data from **319,000 individuals**, including **19 factors** (medical and lifestyle)
- **HeartDisease (target)**, BMI, Smoking, AlcoholDrinking, Stroke, PhysicalHealth, MentalHealth, DiffWalking, Sex, AgeCategory, Race, Diabetic, PhysicalActivity, GenHealth, SleepTime, Asthma, KidneyDisease, SkinCancer

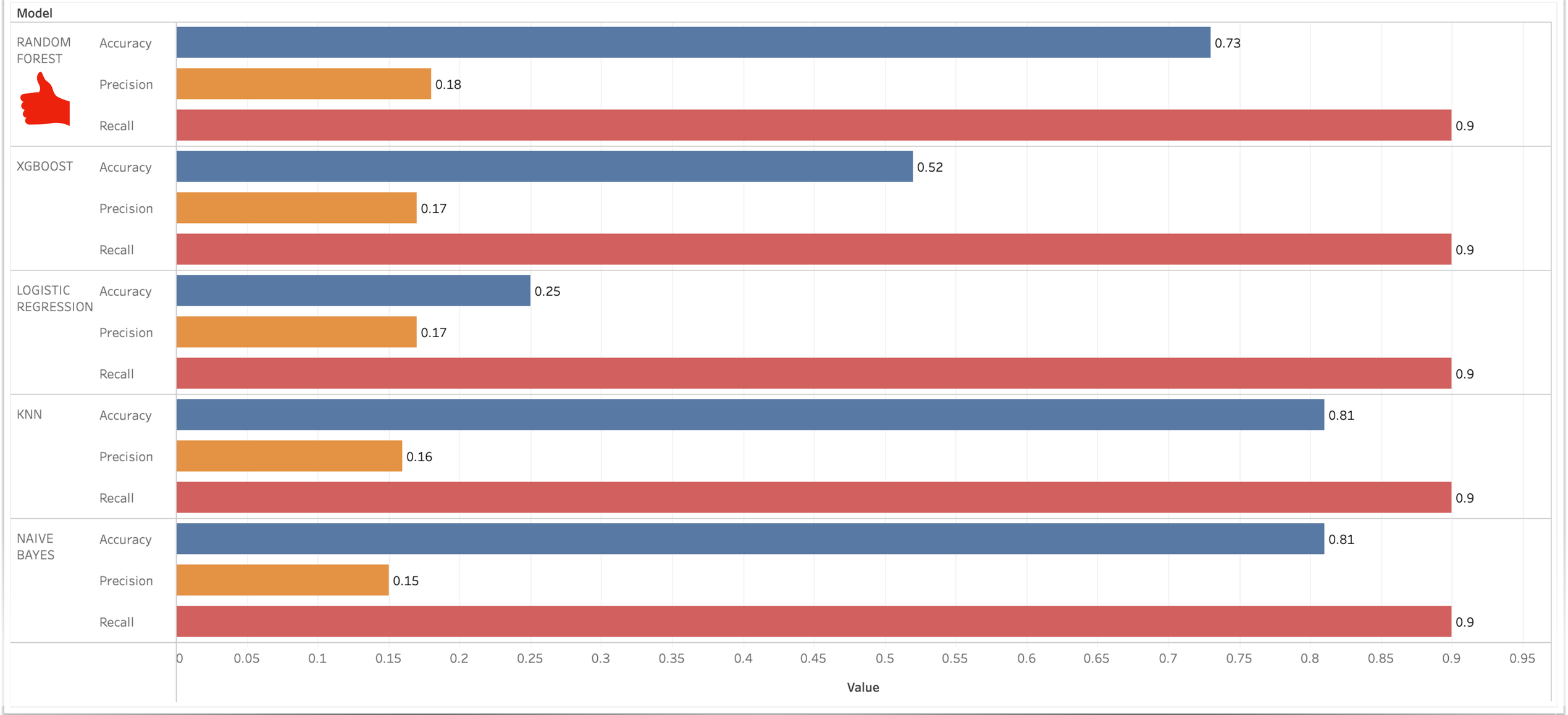
Methodology

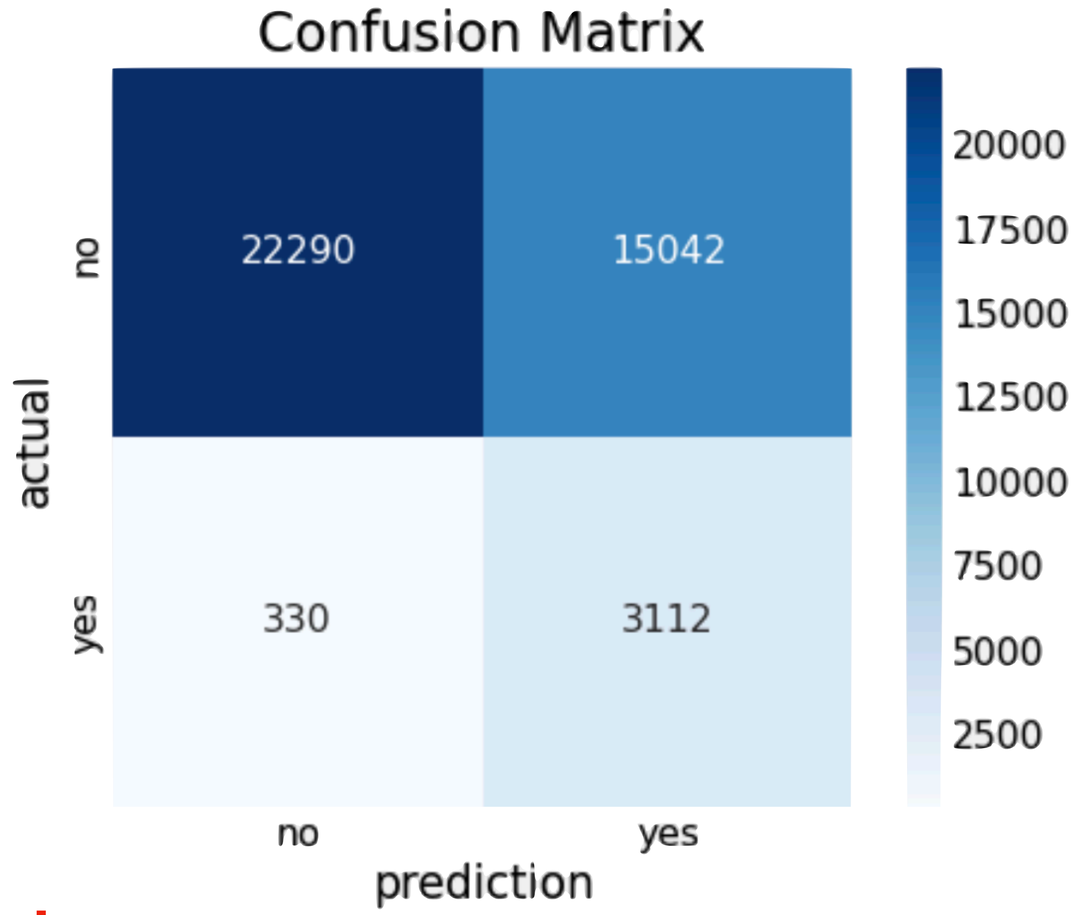


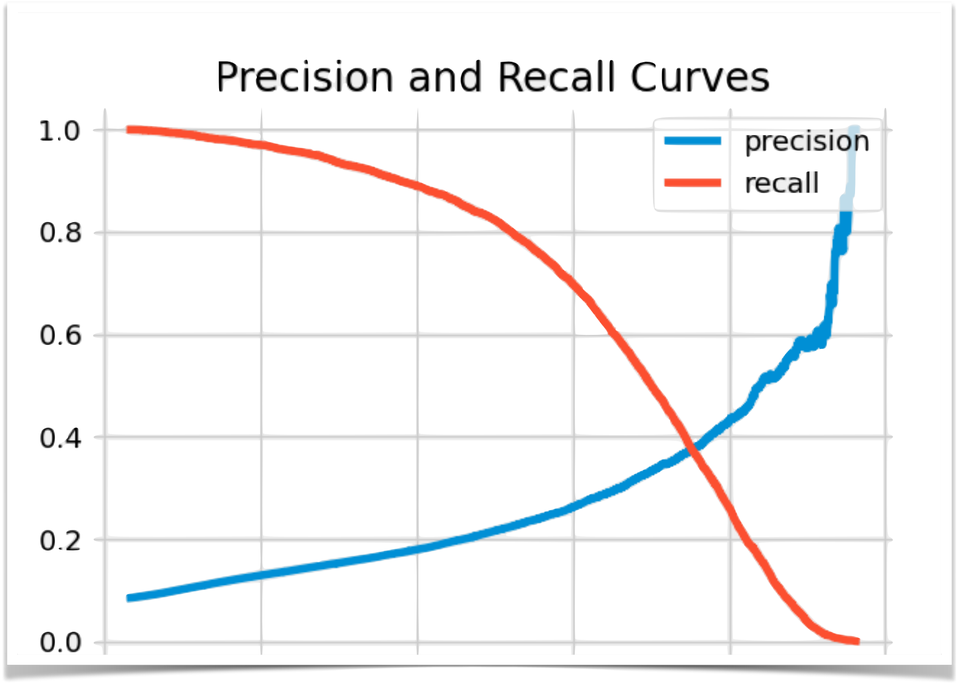
- Focus on achieving a high **recall score (90%)**
- To address **class imbalance**, techniques such as **undersampling**, class **weights**, and **threshold** adjustment were used
- Data **preprocessing** included converting categorical variables to dummy variables and dividing the data into train, validation, and test sets. Model tuning was done using **RandomizedSearchCV**

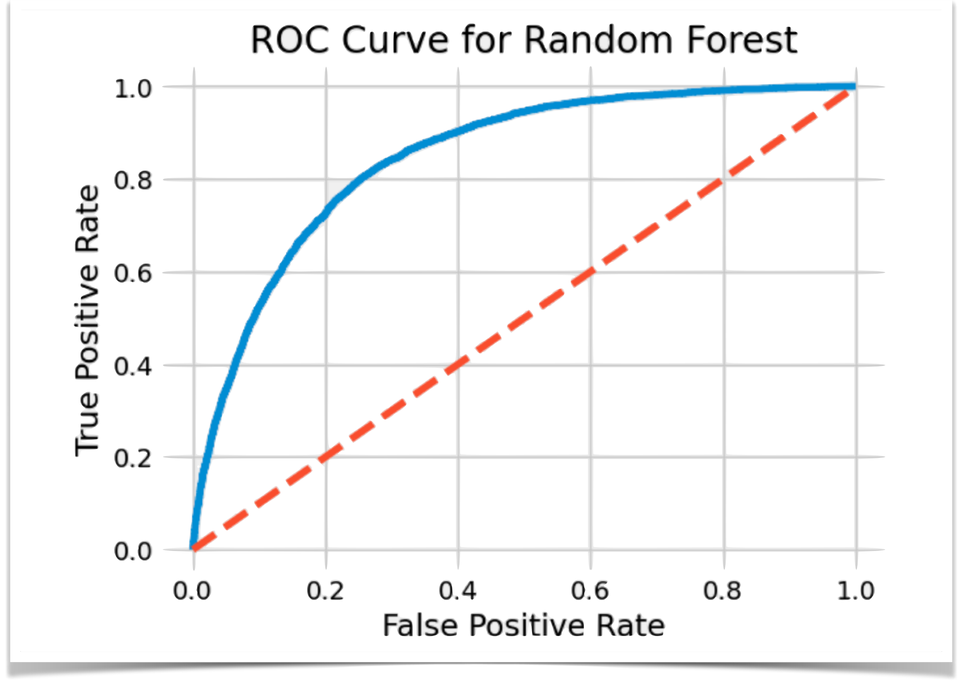


Results


Model
RANDOM FOREST
Accuracy
Precision
Recall
0.73
0.18
0.9
XGBOOST
Accuracy
Precision
Recall
0.52
0.17
0.9
LOGISTIC REGRESSION
Accuracy
Precision
Recall
0.25
0.17
0.9
KNN
Accuracy
Precision
Recall
0.81
0.16
0.9
NAIVE BAYES
Accuracy
Precision
Recall
0.81
0.15
0.9
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
Value







- **Random Forest**

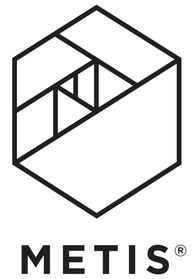
- Highest precision and accuracy while maintaining a recall rate of 0.90

Conclusions/Recommendations

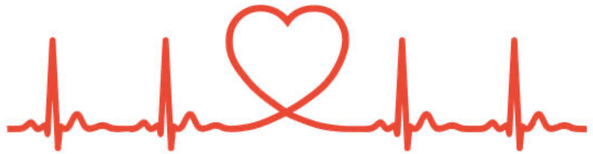
METIS®

- The **Random Forest** model developed in this project has the potential to aid in **early detection of heart disease**
- Other considered models were showing similar performance
- Systematic preventive medical **examination** to decrease chance of heart decease and **healthy** life style
- Constructed **models are ready** base for similar searches

Future Work



- Further exploration of feature relationships through **deeper EDA**
- **Fine-tuning of the Random Forest** model
- Exploring the use of **deep learning** models for improved performance
- Development of a **web app** for easy access
- **More data** (next years update)

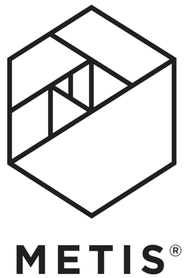


METIS®

Thank you!

Questions?

Appendix



- The data used for this project was sourced from Kaggle, from a dataset that was initially cleaned by Kamil Pytlak. The dataset can be found at: www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease