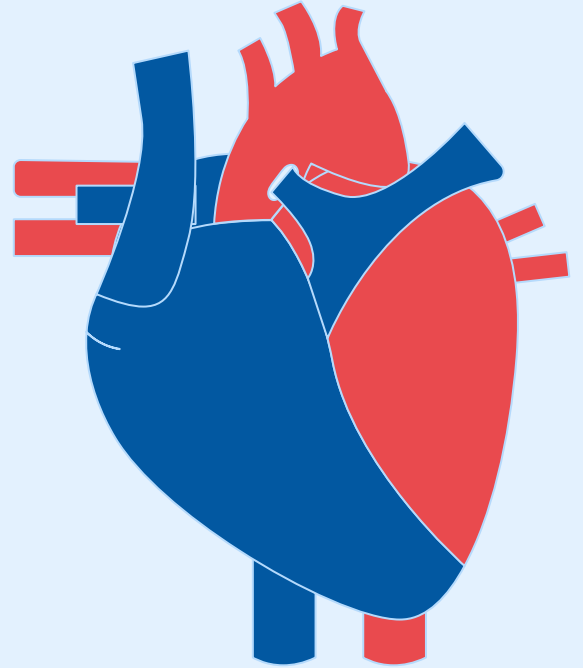


Cardiovascular Disease Prediction



OUTLINE

Introduction

Data
Description

Cleaning

EDA

Modeling

Tools



```
graph LR; A[Introduction] --> B[Data Description]; B --> C[Cleaning]; C --> D[EDA]; D --> E[Modeling]; E --> F[Tools];
```

INTRODUCTION :

- Cardiovascular disease is a class of diseases that involve the heart or blood vessels, And are the leading cause of death worldwide by representing.
- The goal of this project is to build machine learning models to classify or identify patient's condition on Cardiovascular Disease based medical examination for the patient.

Data Description :

DATA COUNT

It contained **70,000** observations
and **13** features.

After cleaning

it became **61,296** observations
and **24** features.

THE Target VARIABLE

Disease

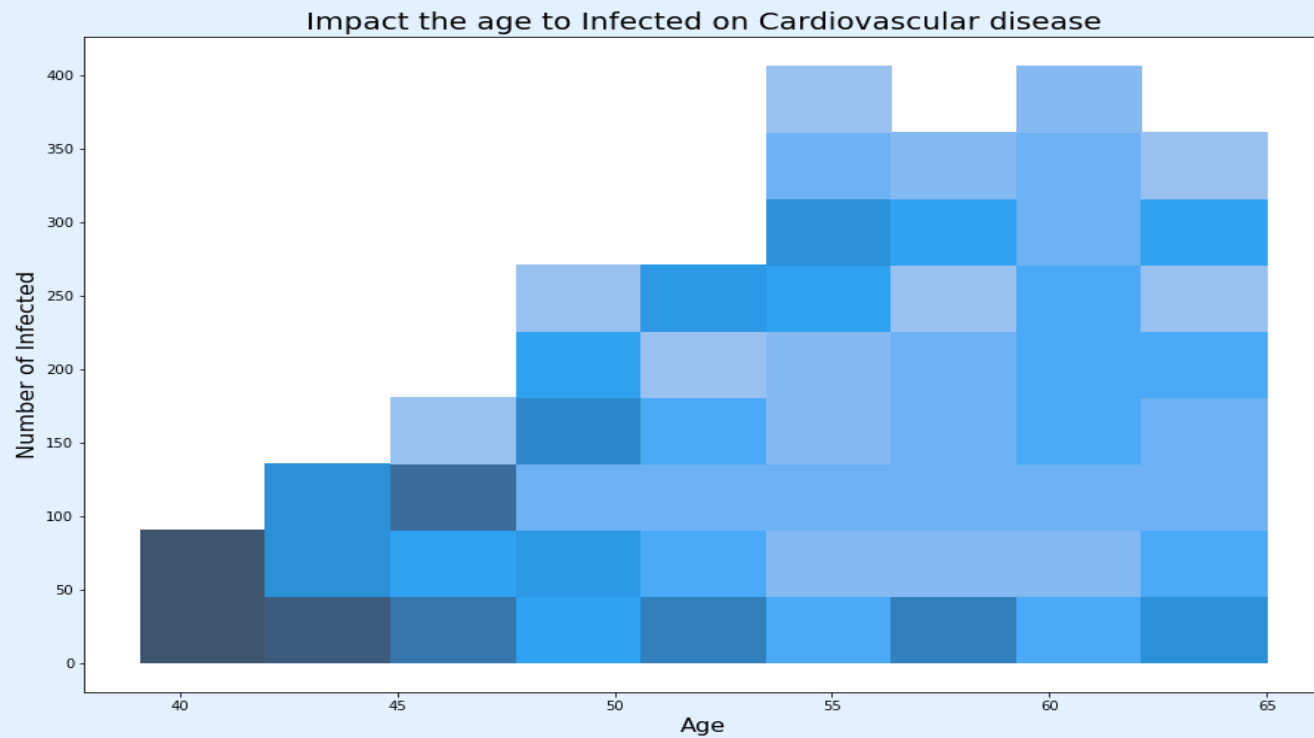
Source

from [Kaggle](#).

Data Cleaning :

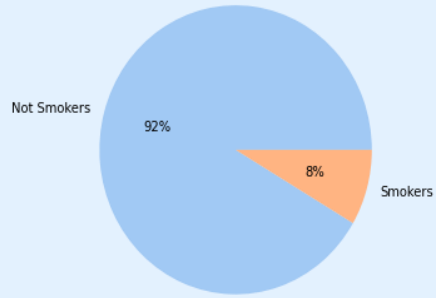
- ➔ Check null.
- ➔ Drop column.
- ➔ Rename columns.
- ➔ convert **[age]** from days to years.
- ➔ Remove outlier.
- ➔ Feature Engineering.
- ➔ Get Dummies.

EDA :

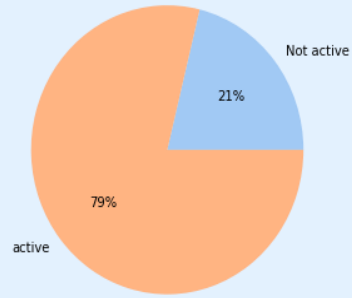


EDA :

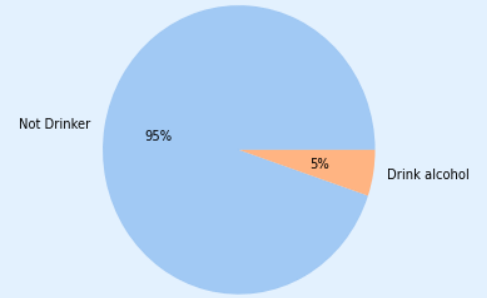
Ratio of smoking patient getting the Cardiovascular Disease



Ratio of active patient getting the Cardiovascular Disease



Ratio of alcoholic patient getting the Cardiovascular Disease



Modeling : Beasline model

Model	Training	validation
Logistic Regression	0.709	0.693

Improving models:

- Dealing with outliers
- Add features engineering
- Scaling dataset
- GridSearch
- Feature Selection

Models:

Model	Training	validation
Logistic Regression	0.730	0.708
MLP Classifier	0.737	0.718
Random forest	0.745	0.724
Extreme Gradient Boosting	0.746	0.723
Average Voting	0.744	0.721
Stacking Classifier	0.744	0.723

Note: these the best 5 models fitting with this dataset

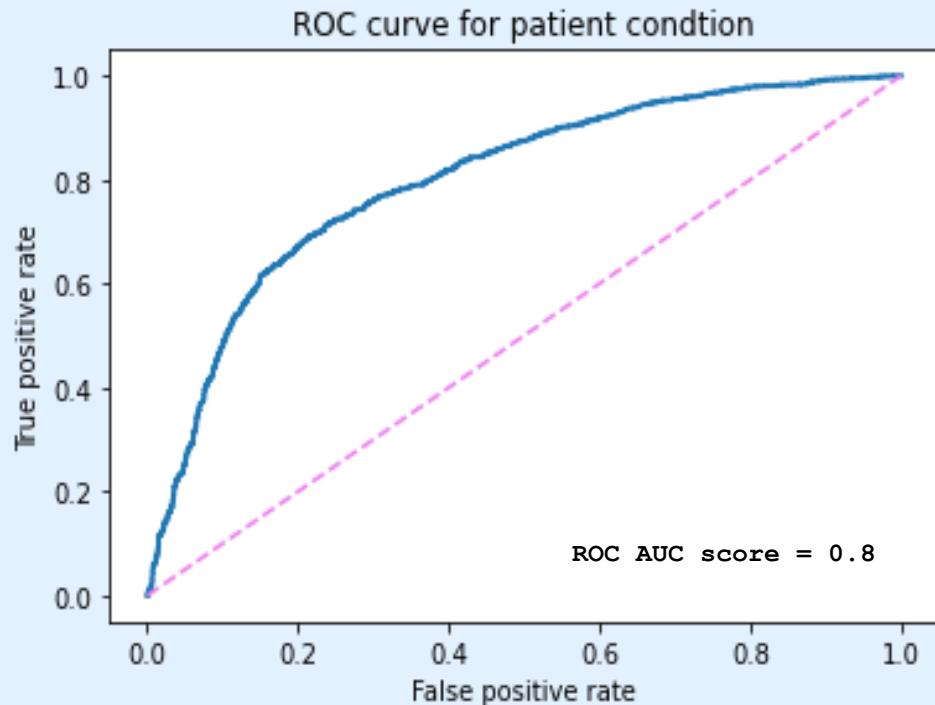
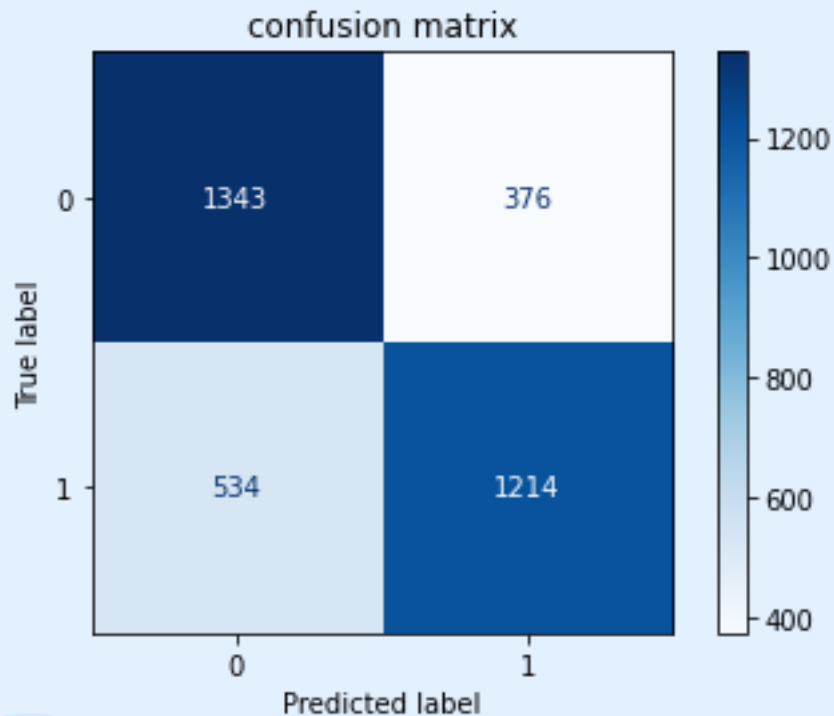
Models: Feature Selaction

Model	Training	validation
Logistic Regression	0.729	0.726
MLP Classifier	0.735	0.737
Random forest	0.745	0.732
Extreme Gradient Boosting	0.746	0.739
Average Voting	0.743	0.737
Stacking Classifier	0.743	0.739

Selected model :

Model	Training	Testing
Extreme Gradient Boosting	0.739	0.737

Confusion Matrix AND ROC :



TOOLS:



Thanks!

Does anyone have any questions?

Hala Alanzi

GitHub @ halaalanzi

Mashael Alfehaid

GitHub @ Mashael999