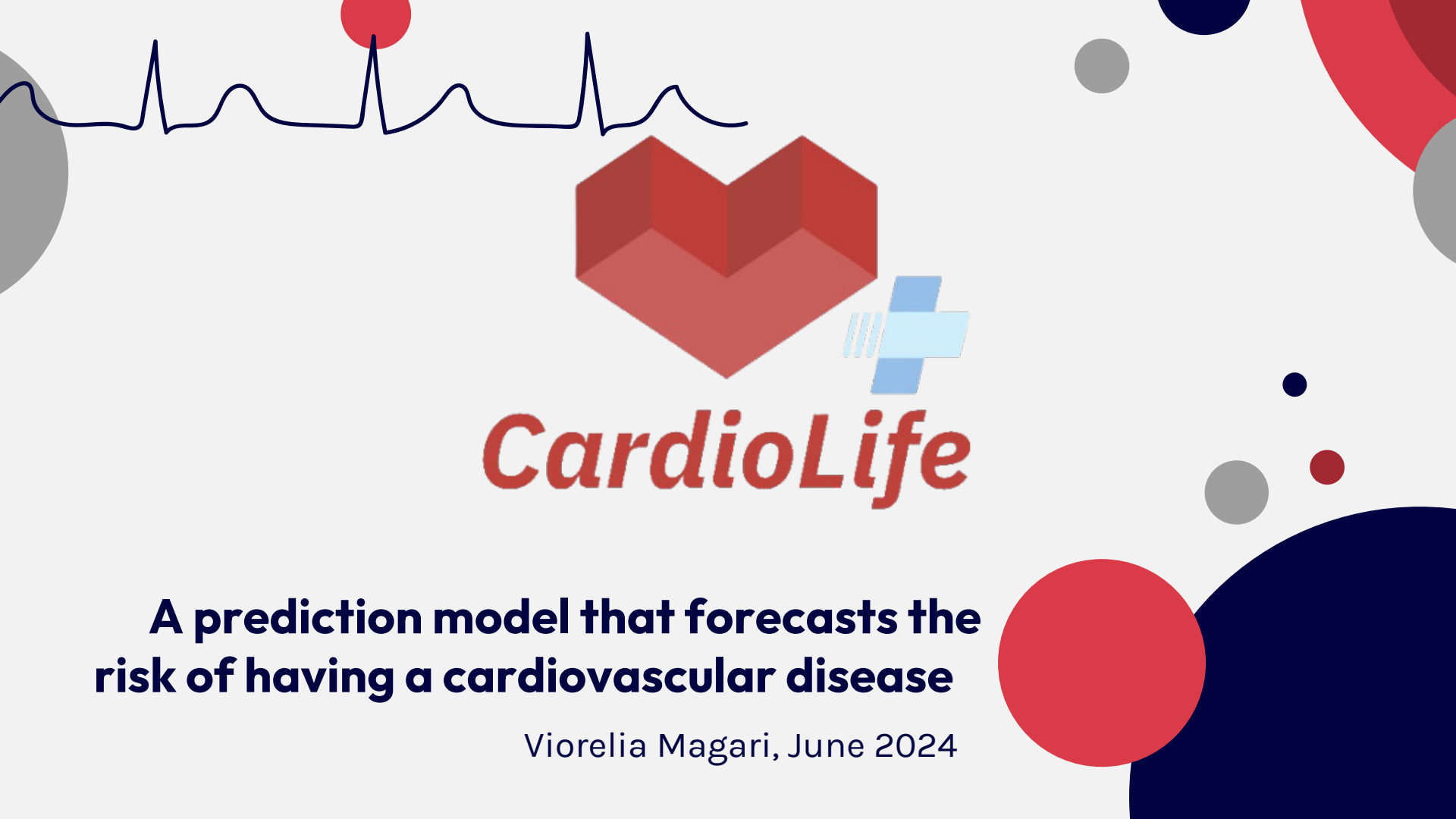


A prediction model that forecasts the risk of having a cardiovascular disease

Viorelia Magari, June 2024

Overview

Subject area:

- Detect the population with the higher risk of CVD
- Identify key risks that influence the apparition of CVD
- Make recommendations to minimize the risk

Proposed solution:

- Build the best model with ⬆ accuracy and ⬇ False Negative predictions
- Make a user-friendly web app, predict CVD and display recommendations

Potential impact:

⬆ The overall wellbeing of the population

⬇ Death levels

⬇ Cost of treating patients

⬆ Income of people, government

Dataset pre-processing

Overview:

70000 records of patients:

- ***Personal data***
- ***Medical data***
- ***Lifestyle data***

Data cleaning:

- Dropped 1,57% of invalid data (negative blood pressure, too short/tall patients, etc.)
- 68903 records remained
- Dropped id column

Data pre-processing:

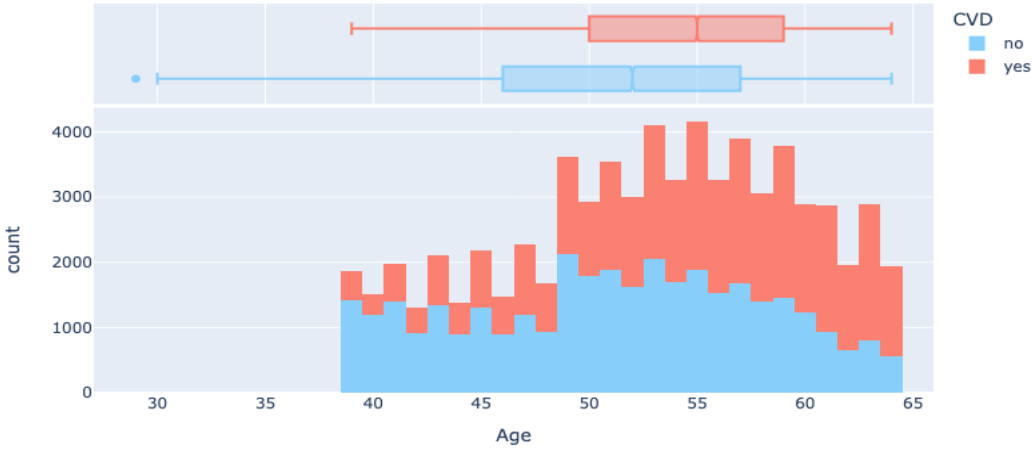
- *Mapped categorical columns*: gender, smoking, drinking, active, cardio
- *OneHotEncoder columns*: cholesterol, glucose

Findings EDA

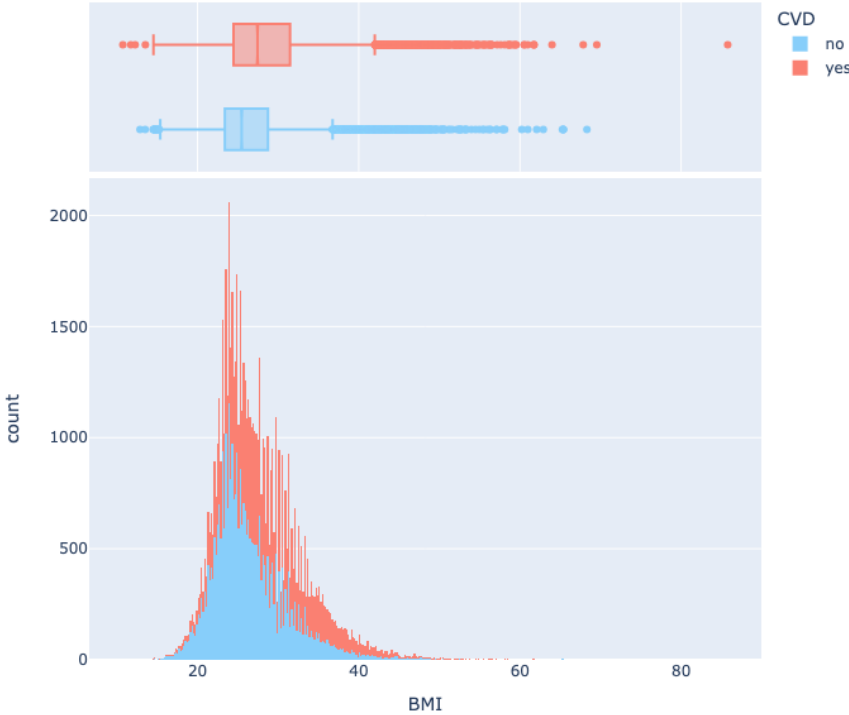
High risk:

- The oldest patients, 55+ years old
- Men, with 0.8%
- Overweight and obese patients, >25 BMI
- High blood pressure, >120mmHg
- Patients who have very high cholesterol and/or glucose
- Patients who have a sedentary lifestyle, with 7.7%

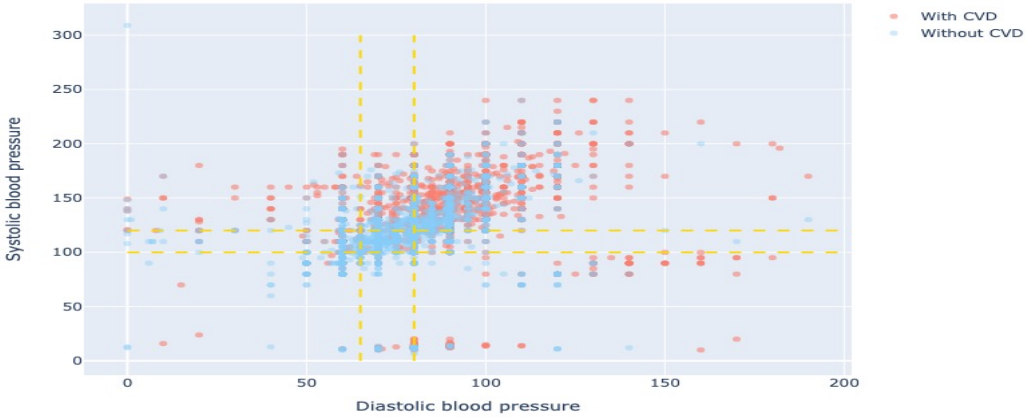
Age of patients, by having a CVD



The distribution of BMI, by having a CVD



The relationship between blood pressure and presence of a CVD

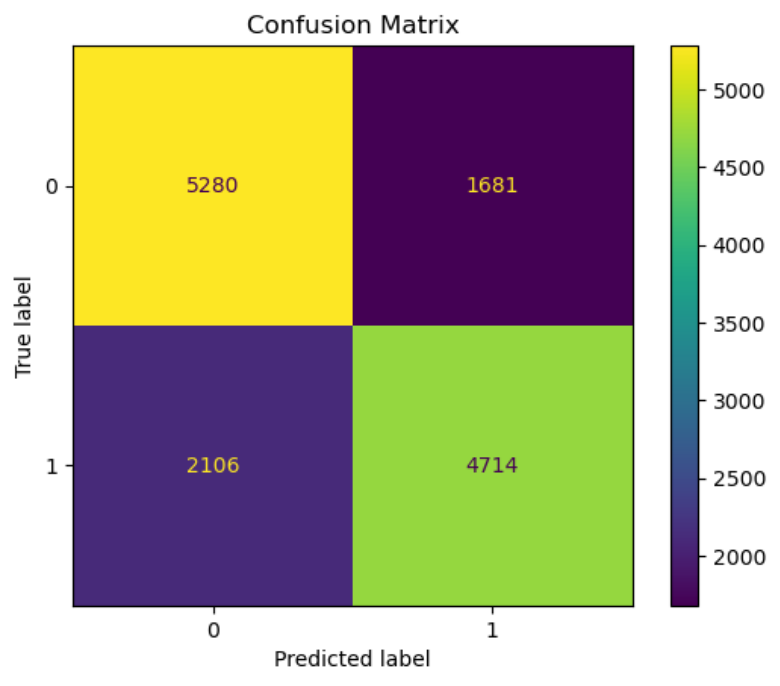


Baseline Modelling

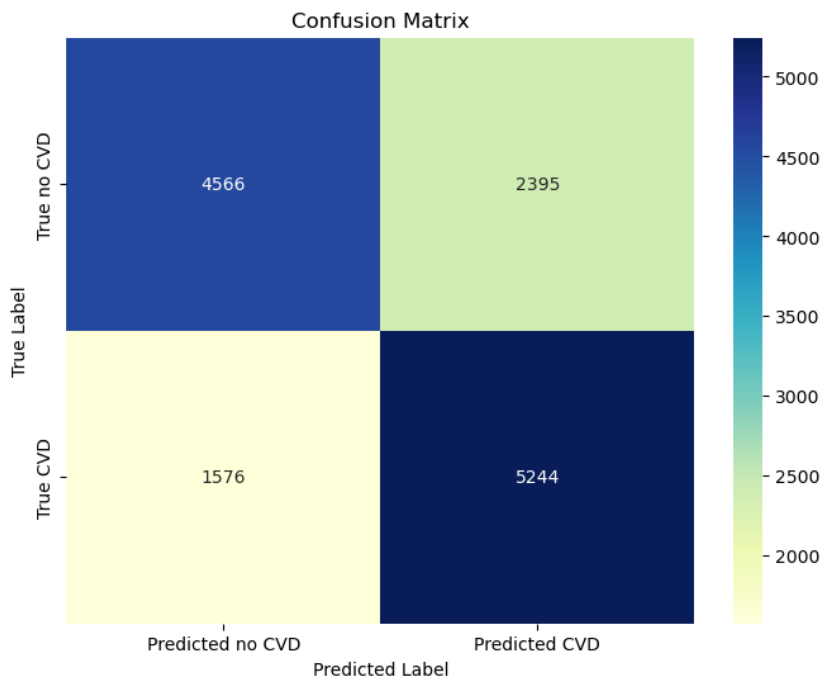
Model Evaluation	Logistic Regression	Decision Tree	K-Nearest Neighbors
Accuracy	0.73	0.73	0.73
Train score	0.73	0.73	**0.74**
Test score	0.73	0.73	0.73
Precision	0.75	0.76	0.74
Recall	0.68	0.68	**0.69**
F1-score	0.71	0.73	0.72

Confusion Matrix

K-nearest neighbors:



KNN with threshold 0.43:



Next steps

1

Advanced Machine Learning

Random Forest
Neural Network
Gradient Boosting
etc.

2

Choose the Model

Best accuracy
Lowest False Negative predictions

3

Optimize the model

And save it

4

Develop web app

Using Streamlit

5

Predict CVD

Make recommendations