Predicting Heart Disease

Diogo Cerveira	Diogo Mendes	João Amorim
201907427	202005393	202009074

Data Set:

For this project, there will be used data provided by the Cleveland Heart Disease Database. This database refers to data from 270 observations and includes 13 attributes (extracted from a larger set of 75):

1. Age	Real
2. Sex	Binary
3. Chest pain type (4 values)	Nominal
4. Resting blood age	Real
5. Serum cholestoral (in mg/dl)	Real
6. Fasting blood sugar > 120 mg/dl	Binary
7. Resting electrocardiographic results (values 0,1,2)	Nominal
8. Maximum heart rate achieved	Real
9. Exercise induced angina	Binary
10. Oldpeak = ST depression induced by exercise relative to rest	Real
11. The slope of the peak exercise ST segment	Ordered
12. Number of major vessels (0-3) colored by flourosopy	Real
13. Thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	Nominal

There is one adittional column for the **variable to be predicted**, with the values (1) for Absense and (2) for Heart Disease.

Project Idea:

Heart disease is the number one cause of death worldwide. Its prediction using a machine learning model should be a revelant starting point for a student in biomedical engineering.

The idea is to predict the presence/absence of heart disease in patients given the aforementioned medical information, containing 4 types of attributes (real, binary, nominal and ordered). The model output is the probability of the patient having heart disease.

Software:

There will be used Visual Studio Code with Python Programming Language to implement machine learning algorithms. According to the literature we found, the most used Machine Learning models applied to this type of data are: Linear regression, Decision tree, Support Vector Machine, K-nearest Neighbour.

Papers:

Heart disease prediction using machine learning algorithms, Jindal H, Agrawal S, […] Nagrath P, IOP Conference Series: Materials Science and Engineering (2021)