Heart Disease

Ömer Faruk Işık  
Computer Engineering  
Kültür UniversityIstanbul, Turkey  
1700003708@stu.iku.edu.tr

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords – heart disease, dataset, model, algorithm, preprocessing

# Introduction (*Heading 1*)

Heart diseases are one of the major problems of this era, due to junk food consumption and anything bought and ordered is processed. But most of the heart diseases can be cured to a percentage of %70 to %100 with early diagnosis. But number of cardiologists are falling behind in a rapid rate against the parabolic increase of heart diseases. To overcome this and many similar problems from life hacks to fatal problems, day by day the usage of machine learning is getting more settled in our lives. And machine learning method is advanced tremendoulsy in couple of years. So there are lots of algorithms, data and powerful computers to process that data with intended algorithms to train a machine learning model. In the end with the necessary amountof technical and theoretical knowledge, one can achieve almost anything via machine learning.

# Preliminaries

1. **Python-** It is a programming language that is high-level, object-oriented with built in data structures and semantics; It commonly used for mathematical calculations, complex data analysis, artificial intelligence, deep learning and machine learning. It has an increasing popularity against other programming languages due to the wide array of uses, ease of use,
2. **Preprocessing-** Manipulating the dataset to fit your needs (create data-frames, splitting, merging, deleting (rows or/and columns), adding, imputing, normalizing, etc.).
3. **Decision Tree Classifier-** Belongs to the sklearn.tree library; A decision tree is a tree-like model that acts as a decision support tool, visually displaying decisions and their potential outcomes, consequences, and costs. From there, the “branches” can easily be evaluated and compared in order to select the best courses of action.
4. **Random Forest Classifier-** Formed by n number of decision trees. It builds decision trees on different samples and takes their majority vote for classification.
5. **Multi-Perceptron Classifier-** A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses back-propagation for training the network. MLP is a deep learning method.
6. **K Neighbors Classifier-** It is a supervised machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The number of nearest neighbors to a new unknown variable that must be predicted or classified is denoted by the symbol 'K'. Commonly used for linearly non-separable data to overcome inefficiency of linear algorithms.
7. **Ridge Classifier-** It is a linear classification algorithm and a way to create a parsimonious model when the number of predictor variables in a set exceeds the number of observations, or when a data set has multicollinearity (correlations between predictor variables).

# User Attributes

Information is provided to select correct users for this model.

By this way only users that are qualified enough to feed the model with correct inputs will be assigned to the job.

## **Skill Level**- From minimum to desired qualification requeirements for users.

* Level 1- Beginner Python skills, basic knowledge about the logic of machine learning, fundamental knowledge of preprocessing and low-level academic certification of cardiology.
* Level 2- Intermediate Python skills, moderate knowledge about the logic of machine learning, common knowledge of preprocessing and medium-level academic certification of cardiology.
* Level 3- Advanced Python skills, good knowledge about the logic of machine learning, advanced knowledge of preprocessing and high-level academic certification of cardiology.
* Level 4- Mastered Python skills, expert knowledge about the logic of machine learning, expert knowledge of preprocessing and expert-level academic certification of cardiology.

Identify applicable funding agency here. If none, delete this text box.

## Some Common Mistakes

* Trying to train binary classification data with algorithms other than tree and/or forest algorithms.
* Trying to normalize an already normalized data
* Not knowing the difference between an accurate model and an overfitted model.
* Not plotting the data.
* Skipping the preprocessing part
* Skipping the confusion matrix.

# Decide the Algorithm/s

To decide which algorithm/s to use one must determine if the data is regression or classification. After the decision, it must be proven if the data is linear, linearly non-seperable or completeley inseperable.

1. **Linear-** Linear data can be trained with linear algorithms, but the exact algorithm depens if the data is regression or classification. Linearity of the data can be proven buy plotting the data. Only exception to this method is binary classification. It is not logical to plot binary classification data.
2. **Linearly Non-Seperable-** If plot shows that the data is not linear, then the algorithm must be decided on this founding, also KNN is proved to be useful with non-linear data.
3. **Inseperable-** If the data is completely inseperable, then common machine leraning algorithms will not be efficient, deep learning methods can be applied to the data to train an efficient model.