

# READ ME

## Title: HEART DISEASE PREDICTION SYSTEM

### 1. ALGORITHMS

#### 1.1 Regression.py

The file Regression.py takes in cleaveland.csv file directly as an input. This file is built in python 3.6.1 and uses skitc learn dependencies. Then converts the input into the data frame format. Then the data is split into two sets one has an attribute from zero to thirteen this is the set of independent variables. This is x input to the regression. The last column is the y input. For the training set this y input is used to calculate the error and updates tetha. The y input for the testing set is used to calculate the error of the model. The data set is split into testing data and training data. The regression model is fit using training data. And then predictions are done using testing data. Finally, the model is evaluated using precision, recall, f-score. The output of the file is the predicted values of the testing data. The confusion matrix and the matrix that has all the evaluation measures.

#### 1.2 Ann\_optimizer.py

The file Ann\_optimizer.py takes in cleaveland.txt file directly as an input. This file is built in python 3.6.1 and uses tensor flow dependencies. The data is split into three sets that are testing set, validation set and training set. Testing set has 70% of the data, training set has 10% of the data and the validation set has 20% of the data. The data is split randomly. Then the model is trained and optimized. The error of the model of three different sets is calculated and plotted. The plots show the error trend. The parameters that give the least error are consider as input to the Ann\_predict. The output is the error trend. In this case 3 hidden layers and 20 neurons gave the best results

#### 1.3 Ann\_prediction.py

The Ann\_prediction.py takes in cleaveland.csv file directly as an input. This file is built in python 3.6.1 and uses skitc learn dependencies. Then converts the input into the data frame format. Then the data is split into two sets one has an attribute from zero to thirteen this is the set of independent variables.

This is x input to the regression. The last column is the y input. For the training set this y input is used to calculate the error and updates the valued in the weight matrix. The y input for the testing set is used to calculate the error of the model. The data set is split into testing data and training data. Then the training of neural network is done according to the parameters that are obtained from the Ann\_optimizer.py, here training data is used to make the model run. And then predictions are done using testing data. Finally, the model is evaluated using precision, recall, f-score. The output of the file is the predicted values of the testing data. The confusion matrix and the matrix that has all the evaluation measures.

### 2. EXECUTION ENVIRONMENT

All the program files are written in python 3.6.1. To execute these file we need numpy, matplotlib, sklearn, pandas, tensorflow. The output of program is displayed in the console.

### 3. DATASET DETAILS

The cleaveland.txt is a data file in the text format. This is given as an input file to the Ann\_optimizer. This file has 14 attributes that are the risk factors that are used to predict weather a person has a heart disease or no.

The file cleaveland.csv has the same data in the csv format. This is done because it is easy to access in the code. Similarly, the procedure goes same with the other datasets namely Hungarian Dataset, Switzerland Dataset and Long Beach Dataset.

### 4. MAKEME

All the py files that are in the make file should be in the same folder and it should be executed with make command.