**Heart Disease Prediction using Data Mining Techniques**

**README**

**Submitted by,**

**Kanika Mittal (MT18075)**

**Nitindeep Singh (MT19069)**

**Akshaj Anil Patil (MT19111)**

**Dhawal Singh Pundir (MT19120)**

**INPUT:**

**Two Datasets containing 13 attributes**

The first data set is taken from the Cleveland Heart Disease database which consists of 303 records and the second dataset is taken from Statlog Heart Disease database. The second dataset contains 270 records.

**OUTPUT:**

* Accuracy metrics for all the classification techniques.
* Precision-Recall curve for each model
* ROC curve for each model
* F-score for each model

# Steps performed in Heart Disease Prediction using Data Mining Techniques

* Read two different data set (These datasets having same attributes one consists 303 records while other 270 records)
* These different datasets are merged and shuffled.
* If NULL value is present for any attribute in the dataset then it will replaced with the column mean.
* Now, Standard scalar is applied over datasets to scale values of all the columns.
* We have used Stratified K-fold cross validation for training and validation split.
* For generating features PCA have been used.
* For prediction various models have been used such as Artificial Neural Networks, Decision Trees, Naïve Bayes, Random Forest, SVM, LSTM, and Logistic Regression.
* For each model we have plotted PR curve, ROC curve, MCC Values.
* For all models accuracy comparison graph has been plotted.
* Various evaluation parameters have been used such as accuracy, specificity, sensitivity, precision, AUC curve, and ROC curve.
* Best accuracy was given by Decision Tree and Random Forest.

**Steps to execute the code**

* Download the py file to any directory.
* Add the two datasets to the same folder.
* Run the code using python BDMH\_18075\_19069\_19111\_19120.py

**Submitted python file:**

* BDMH\_18075\_19069\_19111\_19120.py