Heart Diseases Prediction

Submitted By: -
Parth Parashar (923928157)
Kirtan Patel (973582669)
Reeya Kambhojkar (900823980)
Manish Manchegowdar (956383070)
Tarun Kumar (939420464)

Objective

To compute and compare the accuracies of various Predictive algorithms on our data set namely: Heart-Diseases

Our objective in the computation of accuracies is to find out the value of the "target" column present in the dataset. This target column can take values as 0 and 1. These 0 and 1 values refer to whether the algorithm is able to predict correctly or not.

The Heart Diseases dataset chosen to compute our accuracies has been taken from Kaggle.com.

The dataset that we have used is attached below: -

heart.csv

Predictions' methodology

In our project, we will be using a dataset known as heart diseases dataset. This dataset is present in the form of a csv file given on the previous page.

There are multiple features present in the dataset. Some of those important features are listed below: -
1) Age
2) Sex (Gender)
3) Cp
4) Trestbps
5) Chol
6) Fbs
7) Restecg
8) Thalach
9) Exang

The first step of our prediction algorithms would be to split the dataset into testing and training datasets.
By using the training dataset, we would train our model to predict the diseases properly while the testing dataset would be used to test the model and to see whether the trained model is predicting the results correctly or not.

The second step would be enrichment of the datasets to avoid any run-time errors

The third step would be to use the algorithm for predicting the results.

The results would be stored in the form of targets (values 0 and 1). These values would provide an insight in our algorithms' predictive capabilities.

The various algorithms that we will be using for our predictions would be: -
1) Neural Network
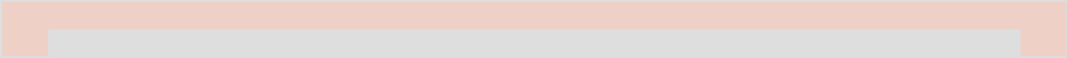2) SVM(s)
3) Naïve Bayes
4) Logistic Regression
5) Sklearn

The algorithms will also give us the following: -
1) Confusion Matrix
2) Probabilities (if applies)
3) Accuracy
4) Precision
5) Recall (if applies)

After we are done with the algorithms' predictive part, we shall also compare the results using the following parameters: -
1) Accuracies
2) Confusion Matrix

This comparison would help us identify the best algorithm used to predict the correct results.

Heart Diseases have been affecting the human society for decades now and with the help of these predictive algorithms, help can be provided to those who are the most vulnerable.