Project Name : Heart Attack Detection

Introduction:

The objective of this project is to analyse Heart Attack data and predict Heart Attack with the help of python libraries (like numpy, pandas, matplotlib, seaborn etc.). We took data from kaggle.

Columns : 'age', 'sex', 'cp','trestbps', 'chol', 'fbs','restecg', 'thalach', 'exang', 'oldpeak', 'slope','thal'.

Target column : 'num'

Steps :

Stored data in Dataframes(pandas library) as it is better in terms of I/O speed, consumed memory and disk space.

First we visualize the shape of the dataset and print all the columns of the dataset and preprocess the data to manage null entries in the dataset and drop the column if required.

For plotting graphs, we imported matplotlib.pyplot and for colors in it matplotlib.colors, as using colors we can visualize things efficiently and those who are not from technical background can easily relate to the data being represented.

Heatmap is a data visualization technique that shows magnitude of a phenomenon as color in two dimensions. The variation in color may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space.So we first draw heatmap.

Distplot plots a univariate distribution of observations so we plot distplot of each of the feature versus target variable.

Countplot is kind of like a histogram or a bar graph for some categorical area so we draw countplot for each of the features.

Running the tests :

First we split our dataset in to training and test set with testsize is 0.2

Principal component analysis (PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance. So we used PCA in our project.

SVC : The objective of a Linear SVC (Support Vector Classifier) is to fit to the data you provide, returning a "best fit" hyperplane that divides, or categorizes, your data.

GridSearchCV : we used this model to train our dataset. GridSearchCV lets you combine an estimator with a grid search preamble to tune hyper-parameters. The method picks the optimal parameter from the grid search and uses it with the estimator selected by the user. GridSearchCV inherits the methods from the classifier.

Result :

We evaluate this project using accuracy value, f1-score, Recall, support and this project provide the best score with the available dataset.