ABSTRACT

We are in a period of “Information Age”where the traditional industry can pressure the rapid shift to the industrial revolution for industrialization, based on the economy of information technology. Terabytes Of Data are produced and stored in day-to-day life because of fast growth in „InformationTechnology‟. Terabytes Of Data are produce stored day-today life because of fast growth in InformationTechnology‟.The data which is collected is converted into knowledge by data analysis by using various combinations of algorithms.For Example: the huge amount of the data regarding the patients is generated by the hospital such as x-ray results,lung results

,heart paining results, chest pain results , personal health records(PHRs) ., etc. There is no effective use of the data which is generated from the hospitals.Some Certain tools are used to extract the information from the database for the detection of heart diseases and other functions are not accepted. The main theme of the paper is the prediction of heart disease using machine learning techniques by summarizing the few current researches. In this paper the logistic regression algorithms are used and the health care data which classifies the patients whether they are having heart diseases or not according to the information in the record. Also I will try to use this data as a model which predicts the patient whether they are having heart disease or not.

## *INTRODUCTION*

To initiate the work we can use different types of techniques and algorithms. In this paper, machine learning techniques are used to increase the accuracy rate. In machine learning technique we can use the following algorithm

***1. Logistic regression***

Sklearn Logistic regression

The logistic regression is also known as sigmoid function which helps in the easy representation in graphs. It Also Provides high accuracy. In this algorithm first the data should be imported and then trained. By using equations the logistic regression algorithm is represented in the graphs showing the difference between the attributes.From the training data we have to estimate the best and approximate coefficient and represent it.

1. ***Comparing and confusion matrices***

The comparison is a confusing matrix, it is the summary for the prediction of the result which we classified. Based on the classification of attributes the correct and incorrect predictions are marked with count values.A confusion matrix,which is represented in table format will explain about the performance of the characterization model on the trained dataset.The most of the performance measures are calculated using this confusion matrix.

The most important organ of the human body is the heart. The function of the heart is to pump blood and circulate the entire body. It is protected by rib cage and it is surrounded by two layered tissue membrane called Pericardium.Is a four chambered organ which separates oxygenated and deoxygenated blood.Heart is having the five types of blood vessels:arteries,veins,capillaries, arterioles, venules and .The sizeof the human heart is about the size of the fist and weight approximately 300grams,the weight in females being about 25% lesser than males.Arteries and veins are present in heart which helps to collect the blood from all the parts and purifies it and the circulates to all the body parts. The nutrients and oxygen present in the body parts are provided by the blood and also will help in the removal of metabolic wastes. Nowadays the lifespan of human beings is reduced due to heart diseases. The factors which may lead to heart disease are obesity, high cholesterol,smoking, increase in blood pressure, diabetes and other factors.

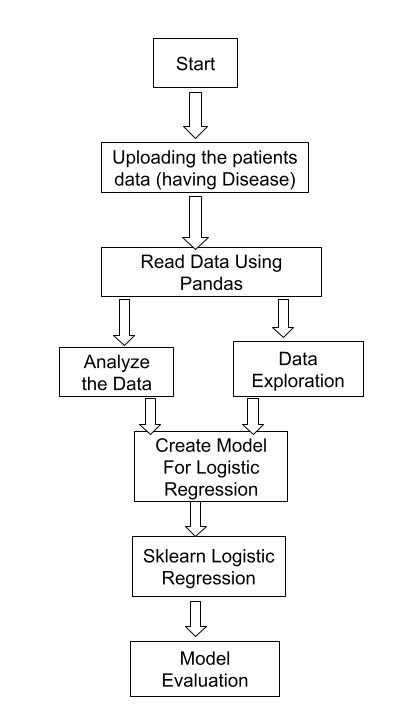
As per the WHO(World Health Organization) records, each and every year millions of people die with different types of cardio attacks such as heart stroke, chest pain,etc. We here proposed the collection of relevant data from the hospitals where the data is so enormous. Now we have to separate the data regarding the patients related to heart diseases. We train the data as per proposed algorithm of machine learning[3] by using logistic regression. For the purpose of detecting heart disease , we can enter patient medical details into the trained data.

## ***EXISTING SYSTEM:***

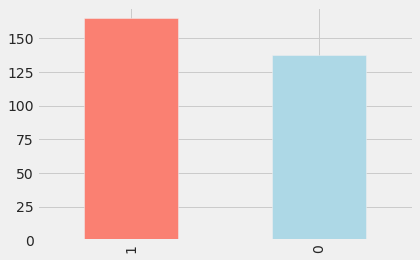
Before all existing systems work, the onset of both Deep learning and data mining[5]. The existing system modules generates comprehensive report by implementing the strong prediction algorithm The main aims of the existing system to compare and check the before patient whose having disease outputs and new patient disease and determine future possibilities of the heart disease to a particular patient By Implementing the above mentioned models will get the goal of developing a system with increased rate of accuracy of estimating the new patient getting heart attack percentage. The model which is proposed for the Heart Attack Prediction System is invented for using Deep learning algorithms and approaches. But By Using all the existing systems the accuracy is very less.

## *PROPOSED SYSTEM:*

This proposed system has data which classified if patients have heart disease or not according to features in it. This proposed system can try to use this data to create a model which tries to predict(reading data and data Exploration) if a patient has this disease or not.In This proposed system, use logistic regression (classification) algorithm. By using sklearn library to calculate score. Finally analysing the results by the help of Comparing Models and ConfusionMatrix.From the dataweave having,it should be classified into different structured data based on the features of the patient heart. From The availability of the data,we have to create a model which predicts the patient disease using a logistic regression algorithm. First,we havetoimport the datasets. Read the datasets, the datashouldcontain different variableslikeage,gender,sex,cp(chest pain),slope, target. The data should be explored so that the information is verified.Create a temporary variable and also build a model for logistic regression. Here,we use sigmoid function which helps in the graphical representation of theclassifieddata.By using logistic regression, naïve bayes the accuracy rate increases***.***



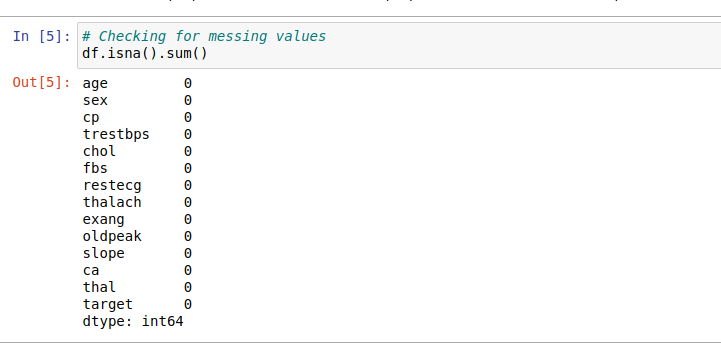
***Targets count of the Disease***

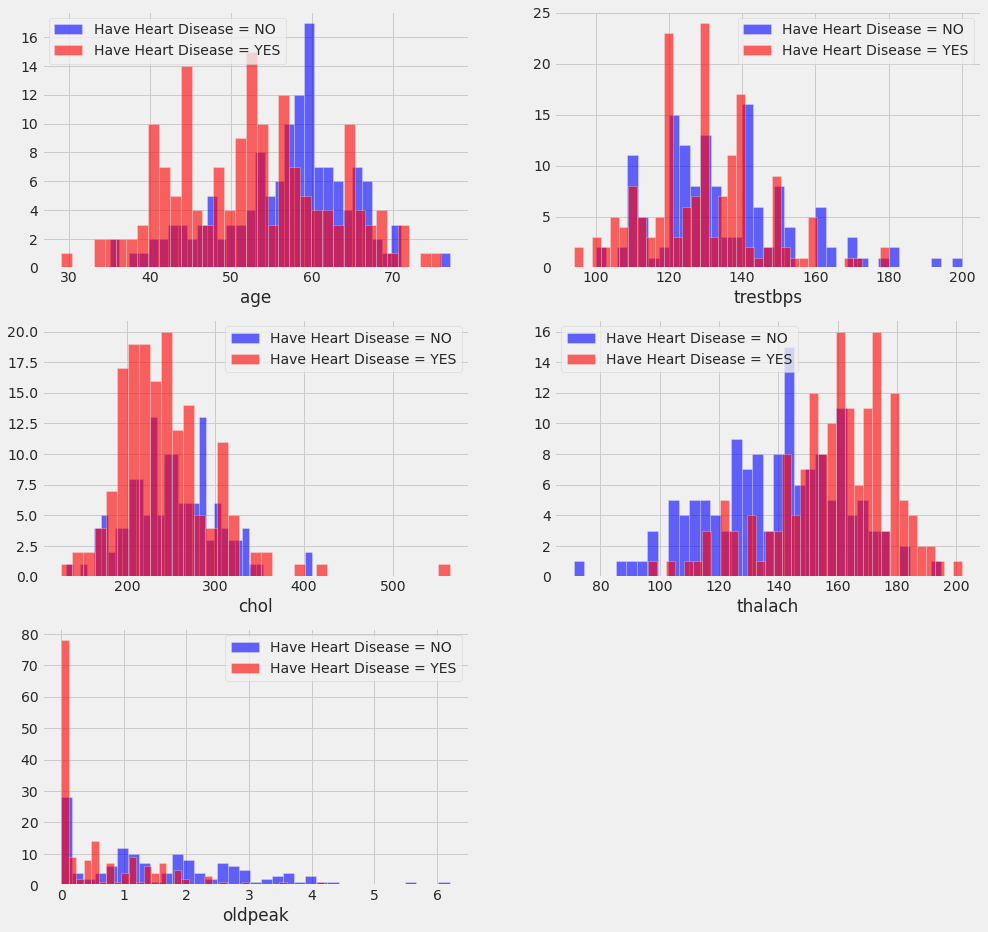


***We have 165 people with heart disease and 138 people without heart disease, so our problem is balance***

***Missing Value Analysis:***

***we have to remove null values from the dataset***

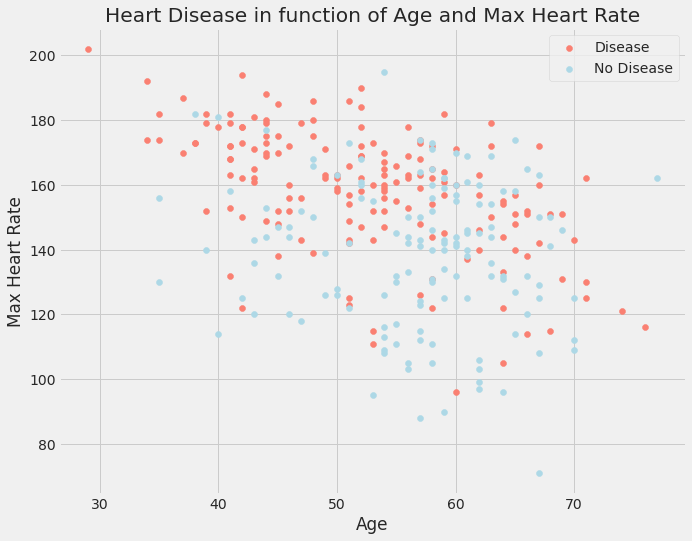


***Visualisation of the Dataset*** 

***conclusions:***

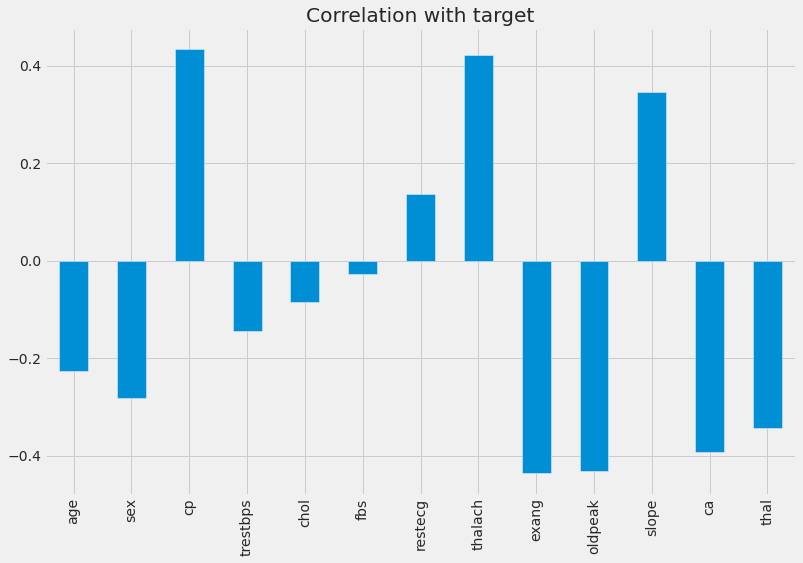
1. ***cp {Chest pain}: People with cp 1, 2, 3 are more likely to have heart disease than people with cp 0.***
2. ***restecg {resting EKG results}: People with a value of 1 (reporting an abnormal heart rhythm, which can range from mild symptoms to severe problems) are more likely to have heart disease.***
3. ***exang {exercise-induced angina}: people with a value of 0 (No ==> angina induced by exercise) have more heart disease than people with a value of 1 (Yes ==> angina induced by exercise)***
4. ***slope {the slope of the ST segment of peak exercise}: People with a slope value of 2 (Downsloping: signs of an unhealthy heart) are more likely to have heart disease than people with a slope value of 2 slope is 0 (Upsloping: best heart rate with exercise) or 1 (Flat Sloping: minimal change (typical healthy heart)).***
5. ***ca {number of major vessels (0-3) stained by fluoroscopy}: the more blood movement the better, so people with ca equal to 0 are more likely to have heart disease.***
6. ***thal {thallium stress result}: People with a thal value of 2 (defect corrected: once was a defect but ok now) are more likely to have heart disease.***

***Relation Between Heart Rate and Diseases***



***Observation : As we can see the There is High chance of having heart disease if person is detected with high heart rate***

***Correlation Analysis:***



* ***fbs and chol are the least correlated with the target variable.***
* ***All other variables have a significant correlation with the target variable.***

# ***Result and Accuracy:***

# 

# ***As We can see our model Achieved 86.81 % Accuracy .***

## *CONCLUSION*

***The amount of Heart diseases can exceed the control line and reach a maximum point. Heart disease are complicated and each and every year lots of people are dying with this disease By using this all systems one of the major drawbacks of these works is mainly focus only to the application of classify techniques and algorithms for heart disease prediction, by all these studying various data cleaning and mining techniques that prepare and build a dataset appropriate for data mining. So that I can use this***

***Machine Learning in logistic regression algorithms by predicting if a patient has heart disease or not. Any non-medical employee can use this software and predict the heart disease and reduce the time complexity of the doctors.***