**Heart Attack Predictions With The Help Of Data Science**

* Dr. Monica Mehrotra,

Director BBDEC

* Dr. Santosh Kumar Shukla

Head Of Department

* Mrs. Priyanka Gupta

Assistant Professor

* Sakshi Khare

[Sakhare18102002@gmail.com](mailto:Sakhare18102002@gmail.com)

* Annushree Upadhyay

[Annushreeupadhyay135@gmail.com](mailto:Annushreeupadhyay135@gmail.com)

* Yasharth Singh

[anjusingh7298@gmail.com](mailto:anjusingh7298@gmail.com)

**Abstract:**

We know that heart attack is the major cause of death worldwide. And now a days it is common and dangerous because every second person get affected by this heart Attack. So to give treatment for heart attack there are lot of advance technologies are used. Generally, heart attack problem has seen in old age person but now a days this problem has also seen in the younger age people. Increasing heart attack cases in the country have forced us to think about the reason behind the increases in such cases. According to National Crime Records Bureau (NCRB) in the last ten years, about 2.25 Lakh Indians have died from a heart attack. To overcome from these problems nowadays we are using big data analytics and machine learning to predict heart attack problems and this become very easy in the treatment of the patient. There are many algorithms in machine learning to predict heart attack like naïve bayes algorithm, k nearest neighbour (KNN) algorithm, etc these are some algorithms in machine learning to check heart attack problems.

**Keyword:**

Big Data Analytics ; Prediction ; Detection Tree ; Logistic Regression ; SVC ; K Neighbors Classifier; Non-Linear ML Algorithms; Decision Tree Classifier ; Random Forest classifier ; Gradient Boosting Classifier

**Introduction:**

On analysing DEATH RATE from 1990 to 2023 it was shown that death rate due to Heart Attack was increased from 12.1 million to 18.6 million (equally distributed between MALES & FEMALES) all over the world. It was predicted that those who has suffered from severe COVID-19 infections not to exert themselves too much while doing exercise and away from hard labour for some time. It is important to recall that a significant ratio of these deaths (46%) is reason for people under 70 years of age, in the most productive period of life. Furthermore, CVD is the main cause for in this age group almost 79%.

Our heart is a muscular organ that pumps blood containing the oxygen and nutrients you body needs. The main pumping chamber of our heart is left ventricle, when our left ventricle contracts it sends oxygen-rich blood to your body through a large artery called the aorta. Connected to our aorta our small arteries called coronary arteries. Blood flow from our aorta through the coronary arteries to supply our heart muscle with oxygen and nutrients. During the heart attack blood flow through one of our coronary arteries may have been severely reduced or completely blocked. Our reduced blood flow may cause by a buildup of a fatty substance called plaque in our coronary arteries. If this plaque become disrupted a blood clot might form and severely worsen the narrowing or lead to a sudden complete blockage stopping blood flow down the artery. A blockage in our coronary arteries prevented the oxygen and nutrients in our blood from reaching the part of our heard supplied by the artery as a result heart muscle in that area started to die. Damage to part of your heart muscle is called a heart attack (also known as a Myocardial Infarction or MI).

Our study for this problem is the part of Big Data Analytics applications, where we detect cardiac patients based on well-defined attributes such as age, sex, cholesterol, blood pressure, etc. With the help of patients, we can collect the data, which is very important to train the learning algorithms, where we use a data set collected from hospitals which certain a group of people are sick, and others are not. Before starting to present our study, we present a state of the art on the most recent research work in this field. This followed by pre-processing where we select the most relevant attributes that give the best results, using the correlation matrix. Finally, we apply learning algorithms on different sizes of the data set (600, 800, 1000, 1200 lines), to develop the most appropriate and stable prediction approach.

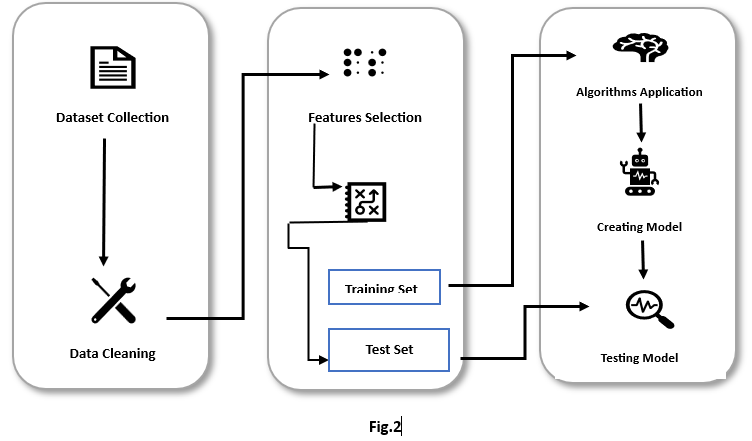
**Related Work:**

In this section, a brief literature has been employed from the latest research related to heart attack prediction. The detection and prediction of heart attack is an important issue. For this reason, a lot of work has been done in this area. The existing work is divided into two categories: The first presents approaches that select the most relevant patient by features selection, and the second is to explore the learning algorithms that offers high accuracy. Feature extraction, classification and predictions are the major steps involved in intelligence algorithms. The research utilized a variety of big data methods to detect cardiac illness, as well as hyperparameter to improve the accuracy of results. Heart attack is concertedly contributed by hypertension, diabetes, overweight and unhealthy lifestyles. Based on the performance of different machine learning models, the best fit model is chosen for the prediction of heart disease. The accuracy score achieved by different algorithms is compared and the satisfactory model is selected. The dataset used is taken from Kaggle propose application of knowledge discovering process on prediction of stroke patients based on different algorithms such as, Logistic Regression, Support Vector Machine, K Nearest Neighbour, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier. The accuracy score of 79% was achieved by Logistic Regression, 80% by SVM, 74% by KNN, 70% by Decision Tree Classifier, 84% by Random Forest Classifier, 80% by Gradient Boosting Classifier. Cardio Vascular Disease was predicted using different machine learning algorithms while on comparison of accuracy of various algorithms highest accuracy of 84% was given by implementing Random Forest Classifier. The realized methods which are shown have less operational sensitivity, specificity and accuracy. Wang et al. (2021b) evaluate classification algorithms using a Machine Learning technique to predict heart attack. This work demonstrated the bagging technique prediction with a good performance rate, as well as accuracy level. Superior prediction models other than past techniques are necessary. Martins et al. (2021) offers a genetic approach for predicting human heart attack through echocardiographic, the designed method is limited to huge unstructured data. The above discussions are providing information about the prediction models and its limitations. It is clear that many cardiac diagnosis models are facing various low-level issues under dynamic conditions. Random Forest Classifier dominated rest of the machine learning algorithms. According to the study, in this research paper Random Forest Classifier is chosen as the final model to make sure the balance between precision and transparency of the model used for predicting the cardiac arrest.

**Proposed Approach (Module):**

Our study is mainly based on following two phases are:-

* Pre-Processing
* Analyses the Dataset and Apply ML



**Pre-Processing:**

Data pre-processing is an important step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn; therefore, it is extremely important that we preprocess our data before feeding it into our model.

**Dataset Collection:**

We collects data related to our project in the form of structured, semi-structured and unstructured data so that we can mined the information according to the our project. W use a data set of people who have performed analyses and tests to detect heart attack. The data set is a matrix where the rows represent the patients and the columns represent the factors or attributes (features) to be tested.

**Data Cleaning:**

Data can be messy, with missing values or errors. Data scientists clean up this data to make sure it's accurate and complete. For instance, you might remove duplicate reviews or fill in missing product information.

**Features selection:**

In a Dataset where we have large set of data or feature, so in that case we have to select the attributes that we can detect the link between the attributes. The chosen attributes are highly dependent each other in order to apply Machine Learning algorithms and achieve better accuracy.

**Data Analysis:**

This step is very important in the development of Machine Learning algorithms. You use statistical and computational techniques to analyse the data. Because we analyse the data set, where we rank or label each person as they showing the symptoms of heart attack or not. To give the algorithms the training dataset, we can form the data set.

**Apply Algorithms & Testing :**

**Splitting Data set:**

To train the Machine Learning algorithm, we mention the target column in the data set, then we divide the data set into two small data sets. Training-set to train the algorithm is the Test-set to test it.

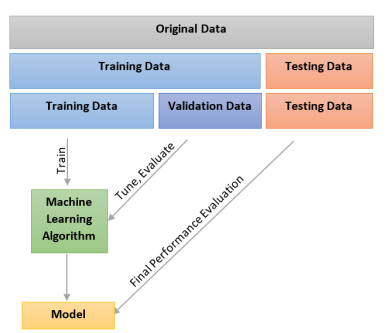


Fig. 1 Splitting the Dataset

**Machine Learning:**

Data science also involves machine learning (ML), it is a branch of AI that enables computers to learn from data and improve over time. ML algorithms can detect or find the hidden patterns in data and make predictions.

In our study we use different algorithms ( like KNN, SVM, etc ) and select the perfect algorithm to our project. Our main approach for finding perfect algo. are based on the following :-

* Accuracy
* Space & Time Complexicity

**Testing algorithms:**

e confusion matrix and the accuracy ratio to test the algorithms, on the test set versus manual exploration. A confusion matrix is a table that is often used to describe the performance of a classification model classifier on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing

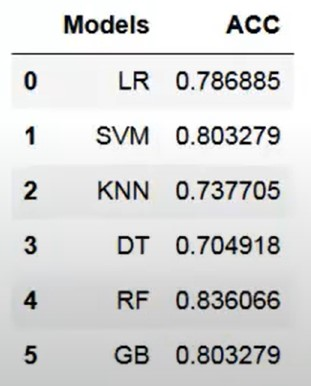
**Results and Discussion:**

The purpose of our study was to use machine learning algorithms for heart attack in healthcare. So for this we performed experiment by using different algorithms on heart attack patients. Through implementation we can know which classification algorithm is best for predicting heart disease.

We use more the 5 algorithms in our study are:

* Logical Regration (LR)
* SVC
* KNN
* Decision Tree (DT)
* Random Forest (RF)
* Gradient Boosting Classifier (GB)

We use all the Algo. and find which one give us best percentage accuracy among all. The following fig.3 table show the accuracy (ACC) of different algo. we used



**Fig.3**

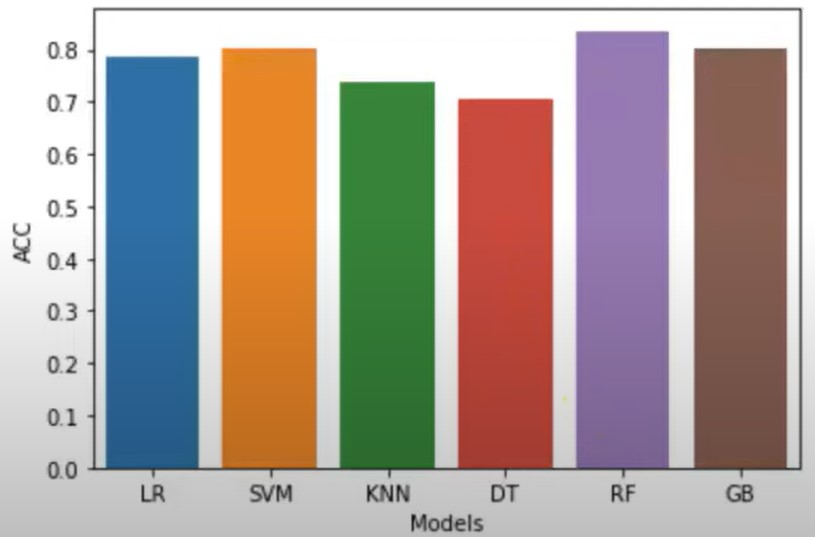
**Algorithm:**

* Import the python libraries
* Pandas
* NumPy
* Matplotlib
* Seaborn
* Import the CSV File **pd.read\_csc(“XYZ.csv”)**
* Check that if there is any Null Value or not in the Dataset.

**If val==True:**

**Drop the duplicate value**

* Drow the correlation matrix and find the following thing
* No. of death rate from Heart Attack,
* No. of Male & Female,
* Gender distribution acc. To the Target Value,
* Age distribution in the Dataset,
* Chest Pain Types and distribution as per Target Variable,
* Show Fasting Blood Suger acc. To the Target, much more.
* Plot the histogram chart with the continuous value
* Splitting the Data-set into the Training Set and Test Set.
* Now apply the different algo. such as KNN, SVC, LR, RF, DT, GB and select most suitable and best accuracy algo.



* Now Prediction on new data to check wither our module work properly or not. And save our model using Joblib

**References:**

* [**Using\_Machine\_Learning\_for\_Heart\_Disease\_Prediction4[1].pdf**](file:///C:\Users\hp\AppData\Local\Microsoft\Windows\INetCache\IE\O405HOIG\Using_Machine_Learning_for_Heart_Disease_Prediction4%5b1%5d.pdf)
* [**https://www.grammarly.com/?q=grammar&utm\_source=placement&utm\_term=plagiarismdetector\_net&utm\_medium=cpc&utm\_content=72890\_3**](https://www.grammarly.com/?q=grammar&utm_source=placement&utm_term=plagiarismdetector_net&utm_medium=cpc&utm_content=72890_3)
* **https://www.kaggle.com**
* **YouTube:**
* **Data Thinkers**
* **CarreerFoundry**
* **Max Healthcare**
* **hanmugasundaram, G., Selvam, V.M., Saravanan, R., Balaji, S.: An investigation of heart disease prediction techniques. In: 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA). pp. 1–6. IEEE (2018)**
* **Google Data-Set Sea**
* **Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology 3(3), 1873–7 (2014)**
* **Radhimeenakshi, S.: Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). pp. 3107–3111. IEEE (2016)**
* [**HeartDiseasePrediction[1].pdf**](file:///C:\Users\hp\AppData\Local\Microsoft\Windows\INetCache\IE\DLJXZV5L\HeartDiseasePrediction%5b1%5d.pdf)
* **Venkatalakshmi, B., Shivsankar, M.: Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology**