



## HEART ATTACK PREDICTION SYSTEM USING MACHINE LEARNING AND WEB APPLICATION

KUNTE YASWANTH RAO

Dr Reeja S R

Register No: 17MIS7152

Professor

VIT-AP University

VIT-AP University

Email Id:[yaswanth.kunte@vitap.ac.in](mailto:yaswanth.kunte@vitap.ac.in)[reeja.sr@vitap.ac.in](mailto:reeja.sr@vitap.ac.in)

### ABSTRACT

The health care industries and hospitals maintain a large amount of data that holds important information and it is very useful for data analysing and also to make very useful decisions. For making this data useful some machine learning algorithms are available and also that machine learning algorithms are very useful. In our project "Heart Attack Prediction System" was developed using these machine learning algorithms by finding the best accuracy. These machine learning algorithms will be explained soon. In this Dataset called "framingham.csv" which is the Heart Attack Dataset, the dataset has 16 (features) columns and 4241 rows that contained the data of the features. The 16 (features) columns are male, age, education, Current Smoker, Cigarettes per day, BP Meds, prevalent Stroke, diabetes, total Cholesterol, systolic blood pressure, diastolic blood pressure, BMI, heartrate, glucose and Ten-Year CHD and these 16 features have its own 4241 rows of data. For this Dataset, by using different ML algorithms to find the best accuracy and to find the Heart Attack chances of a respective person

**Keywords:**Machine Learning, Heart Attack Prediction, Selecting Dataset, Creating App, URL link.

### 1 INTRODUCTION

Heart Attack is one of the major causes of medical issues and it has a large number of deaths in this whole world. According to World Health Organization (WHO), there are 17.9 million people died due to Heart Attacks. If the Health centres and doctors can find a person who will die with a heart attack or not, then it will be a great help for our global society. So, Doctors and health centres are moved to Machine Learning techniques to develop a solution for this problem. And now there are many developed solutions available in health centres.

So, to develop such solution that will predict the heart attack chances of the respective person and helps society. The Heart Attack Prediction Project is about whether a person is going to attack with a heart attack or not in ten years by collecting his current body details. In this Dataset called "framingham.csv" which is the Heart Attack, the dataset has 16 (features) columns and 4241 rows that contained the data of the features. The 16 (features) columns are



male, age, education, current Smoker, Cigarettes per day, BP Meds, prevalent Stroke, diabetes, total Cholesterol, systolic blood pressure, diastolic blood pressure, BMI, heartrate, glucose and TenYearCHD[4]. And these 16 features have its own 4241 rows of data. For this Dataset, to perform operations using different ML algorithms to find the best accuracy and to find the Heart Attack chances of a respective person.

### 1.1 Objectives

The following are the objectives of this project:

The main objective of his Heart Attack Prediction Project is to find the respective person's correct heart attack chances are high or low within ten years by collecting his body details. To find the correct heart attack chances, we need the best accuracy by using a suitable ML algorithm. When we get the best accuracy, then we can build a website and connect it to the cloud, so the user can access the website anywhere by URL link and the user can predict the heart attack chances by giving his body details as we mentioned.

## 2. Background and Literature Survey

A similar project that was found and the project report which is based on Heart Attack Prediction but there is a huge difference between our project and their report. The name of the report is Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction and developed by Amayo Mordecial [3] describes that only selected the dataset of heart attack and cleaned the dataset and they did the clustering using K-means. After getting the K-means, then calculate the Significance Weightage Calculation which can predict the heart attack chances.

In [1] Shah, D., Patel and [2] Diwakar describes that same as in Amayo Mordecial [3] that only using one machine learning algorithm should be enough if it's getting a high accuracy.

Now a days in some Health centres there are various problems related to machines and devices to predict the Heart Attack chances of the respective persons. If we look at some health centres, they may give the correct body details which are cholesterol, systolic pressure, diastolic pressure, BMI, Heart Rate, Glucose but they can't predict the heart attack chances. But some may find the correct heart attack chances by collecting the body details, because of the accuracy rate[5]. By calculating both K-means and Significance Weightage Calculation which can predict the heart attack chances that can find the best accuracy, so that it can give prediction chances of heart attack chances which helps to maintain project standard[6]. Overall, of that report is based on clustering using K-means and Significance Weightage Calculation which can predict the heart attack chances. In this project, the operations different methods to find the best accuracy to find the best prediction of the heart attack chances of the user[7]. And also, to create a website and link it to the cloud, so the user can open that URL link and find his heart attack rate by entering his body details.

## 3. HEART ATTACK PREDICTION SYSTEM



This methodology describes, working methodology, implementation, design and software details.

### 3.1 Methodology

The following diagram (fig.1) shows the Methodology of this project. The Methodology process are Selecting Dataset, Exploratory Data Analysis, Feature Selection, Test Train – Split, Model Training, Model Selection, Pickling Model, Creating App, Designing Webpage, Deploying Website in Cloud Platform and URL link of the Website.

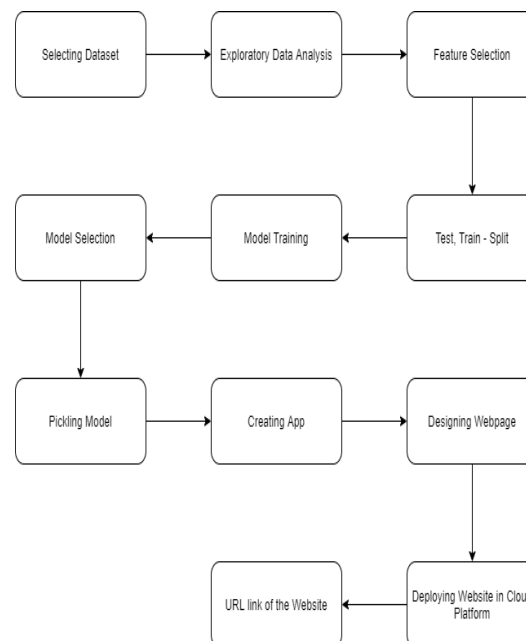


Fig.1: Methodology

### 3.2 Working Methodology

This Heart Attack Prediction System has eleven processes, and those eleven processes are going to do in Jupyter Notebook (Anaconda 3) software to execute our python code to find the best accuracy and also create a pickle file, Spyder (Anaconda 3) to create the python app using pickle file, Visual Studio Code to create a web application, Heroku Cloud to link the project to cloud. The whole process of the Working Methodology will be shown in 2.3 Implementation section with details explanation.

## 4. Implementation



Implementing a Heart Attack Prediction project means carrying out activities step by step. One of the most important ones is to implement our project in the right way and implementing our steps gives us a perfect result. The implementation of Heart Attack Prediction is shown below:

#### **4.1.1 Selecting a Dataset**

The Dataset that took for the analysis is the Framingham heart dataset. In this Dataset called "framingham.csv" which is the Heart Attack, the dataset has 16 (features) columns and 4241 rows that contained the data of the features. The 16 (features) columns are male, age, education, current Smoker, Cigarettes per day, BP Meds, prevalent Stroke, diabetes, total Cholesterol, systolic blood pressure, diastolic blood pressure, BMI, heartrate, glucose and Ten-year CHD. And these 16 features have its own 4241 rows of data. For this Dataset, we were using different ML algorithms to find that best accuracy and to find the Heart Attack chances of a respective person.

#### **4.1.2 Exploratory Data Analysis (EDA)**

EDA refers to the essential process of conducting original data exploration to identify patterns in order to spot irregularity, to test hypothesis and to check assumptions with the help of summary statistics. The step-by-step processes are:

#### **4.1.3 Importing the Dataset:**

In this shows importing this dataset in Jupyter Notebook (Anaconda 3), so it can read the selected dataset and check it.

#### **4.1.4 Pandas Profiling Report:**

In Pandas profiling report shows the report of Summarize the dataset, Generate the report statues and also Render HTML file of the report so that can checked whenever want to checked.

#### **4.1.5 Data Manipulation:**

In Data Manipulation, we are dropping the education column because the education column is not useful to our heart attack prediction and it has missing values. Instead of deleting those null records, we are replacing the null values with 'mean ()'.

#### **4.2 Feature Selection**

In Feature Selection process, we are using the Boruta package which can predict all relevant features and also shows the most important features and they are age and systolic blood pressure. Only with age and systolic blood pressure, we can't find the best accuracy, so we are selecting the top 6 features from the Boruta feature selection. They are age, systolic blood pressure, total Cholesterol, diastolic blood pressure, BMI, Heart Rate, Glucose. And also creating a new dataset in feature selection, so it can be used to do test, train – split process. The presence of irrelevant characteristics in your data can reduce the accuracy of your models. After selecting the top 6 features, now we can create a new balanced dataset and move on to the next process.

#### **4.3 Test, Train – Spilt**



In Test, Train – Split, First, we Test and Train our dataset, then we can split our dataset which is trained. Therefore, we use train data to match the model and test data to test it. The dataset is split into train and test set to check accuracies, by training and testing it on it. So, we can test 20% of our dataset and 80% of our dataset to train dataset. The split method involves taking a dataset and split it into two parts, then we can perform Model Training and selection.

#### 4.4 Model Training

Model training is the process of feeding an ML algorithm with train data and select the best ML algorithm which gives best accuracy. The following ML algorithms for Model Training are:

In Table.1, the accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. F1 Score is the weighted average of Precision and Recall. Area Under the Curve (AUC) provides an aggregate measure of performance across all possible classification thresholds.

	Accuracy	AUC	F1 score
<b>Logistic Regression</b>	0.655895	0.694543	0.625475
<b>K-nearest neighbours</b>	0.848035	0.883416	0.827723
<b>Decision Tree</b>	0.757205	0.771860	0.718623
<b>Support Vector Machine</b>	0.867249	0.932529	0.854406

Table1: Model Training

##### 4.4.1 Logistic Regression

In Logistic regression uses as the representation, very same like linear regression.

Logistic Regression equation:  $y = e^{(b_0 + b_1 * x)} / (1 + e^{(b_0 + b_1 * x)})$ . The accuracy in Logistic Regression we get is 66.2% and f1 score is 63.52%.

##### 4.4.2 K-nearest neighbours (KNN)

In The k-nearest neighbour (KNN) algorithm is a simple ML algorithm that can be used to solve both classification and regression problems. It's easy to apply and understand, but has a major drawback because it becomes slows as the size of that data in use grows.

The k-nearest neighbour equation:  $Pr(Y=j|X=x_0) = \frac{1}{k} \sum_{i \in N_0} I(y_i=j)$ . The accuracy in K-nearest neighbours we get is 83.67% and f1 score is 83.5%.

##### 4.4.3 Decision trees





In Decision Tree it is a tree-structured classifier, where nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. The accuracy in Decision Tree we get is 74.41% and f1 score is 70.55%.

#### 4.4.4 Support Vector Machine (SVM)

In SVM is a supervised ML algorithm that can be used for both classification or regression challenges. SVM is highly uses in classifications. The SVM equation is:  $L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i [y_i(w \cdot x + b) - 1]$ . The accuracy in Support Vector Machine (SVM) we get is 86.99% and f1 score is 85.85%.

#### 4.5 Model Selection

In the Model Selection process, we are selecting the Support vector machine (SVM) model from Model Training. We did Model Training on 4 ML algorithms and in all 4 ML algorithms, the Support vector machine (SVM) has the highest accuracy rate, f1 score, and AUC (Area under the Curve) compared to other ML algorithms. So, we our Model Selection is Support vector machine.

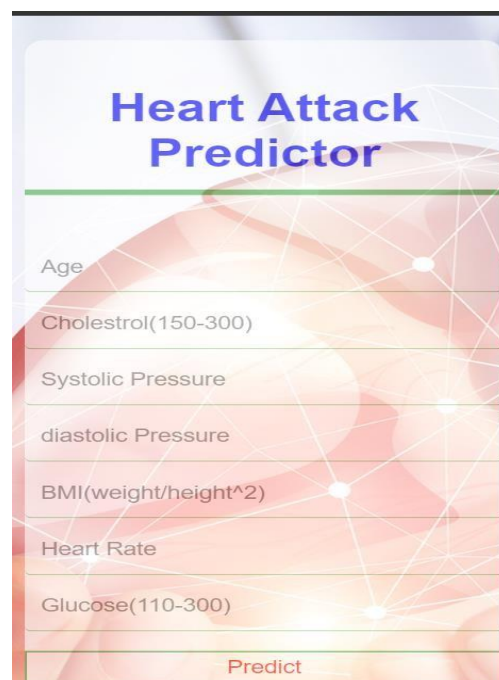


Fig.2: Opening the Website using URL link

#### 4.6 Pickling

In pickling process, the performing pickle operation to serialize the SVM ML algorithm and save the format as a pickle file.

#### 4.7 Creating App



In creating python app, when we creating an App, the data process will be done by import our pickle file and HTML file in app.py.

#### 4.8 Deploying our Website on Heroku Cloud:

In Deploying our website on Heroku Cloud, after creating a app.py we can deploy our webpage in Heroku Cloud platform, then we finally get URL link.

URLLink:<https://svmheartdidease.herokuapp.com/>

App Link: <http://127.0.0.1:5000/>

## 5. RESULTS AND DISCUSSIONS

### 5.1 Opening the Heart Attack Prediction System Website using URL link:

In Fig.2, Accessing Heart Attack Prediction System using URL link we can access the Heart Attack Prediction System website and the URL is: <https://svmheartdidease.herokuapp.com/>

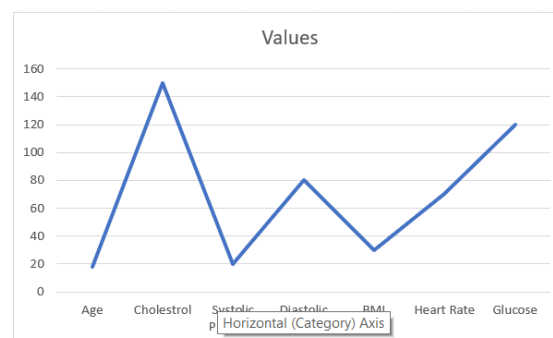
### 5.2 Enter Body Details of respective person's:

In this, Enter Body Details of respective person means if a person wants to know his/her heart attack chances after assessing with URL link, then he/her wants to enter their body details which were given in below.

### 5.3 Result of the Heart Attack Prediction Chances:

Body Details	Values
Age	18
Cholesterol	150
Systolic Pressure	120
Diastolic Pressure	80
BMI	30
Heart Rate	70
Glucose	120

Table.2: Result



**Fig.3: Result**

In Fig.3 and Table.2 Result, after entering body details of respective person, he/her should click predict button which was visible bottom of the webpage, then the heart attack predictor will predict the chances of the heart attack of respective person. In the result is Heart Probability is low. So that means details of the respective person's chances of the heart attack chances is low and he will not face any heart attack with 10 years.

**6. CONCLUSION AND FUTURE WORK**

The conclusion of this Heart Attack Prediction Project is to predict the heart attack chances whether low or high by body details of the person. By using the given URL link, we can open our Heart Attack Prediction application anywhere.

The Future work for this project is by applying all the Machine Learning algorithm, we can find even better accuracy and f1 score. And we also create a Android App, so people can easily open in their mobile at anywhere. We can do lot of better things to improve this project even better by using the even better features in dataset and so on. There are many possible chances to improvements can be done by adding even better machine learning algorithms, accuracy of the prediction system and also, we can develop best generalized system which is suitable for this project based on machine learning algorithm and so on.

**REFERENCES**

- [1] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, 1(6), 1-6.
- [2] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, 37, 3213-3218.
- [3] Amayo Mordecai (2020). A Medical Student and Software developer. And his Current status is at the intersect of biomedicine and info: 56-528.
- [4] S. R. Reeja and N. P. Kavya, "Real time video denoising," *2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA)*, 2012, pp. 1-5, doi: 10.1109/AICERA.2012.6306745
- [5] S.R. Reeja, Rino Cherian, Kiran Waghmare, Jothimani, Chapter 7 - EEG signal-based human emotion detection using an artificial neural network, Editor(s): Hemanth D. Jude, Handbook of Decision Support Systems for Neurological Disorders, Academic Press, 2021, Pages 107-124, ISBN 9780128222713
- [6] Dr Reeja S R, Mr. Ishfaq Yaseen, OBJECT DETECTION, TRACKING AND BEHAVIOURAL ANALYSIS FOR STATIC AND MOVING BACKGROUND,





Volume 7 Issue 1, January 2024

Paper 77

International Journal of Early Childhood Special Education (INT-JECSE)  
DOI:10.9756/INTJECSE/V14I5.79 ISSN: 1308-5581 Vol 14, Issue 05 2022

[7] . Bandi and R. S. R, "A Review On Thermal Infrared Semantic Distribution for Nightfall Drive," 2022 *Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, 2022, pp. 645-653, doi: 10.1109/ICICICT54557.2022.9917651