# HEART STROKE PREDICTION USING MACHINE LEARNING

## A UG PROJECT PHASE-1 REPORT

Submitted to

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In Partial fulfillment of the requirements for the award of the degree of

## BACHELOR OF TECHNOLOGY

## IN

## COMPUTER SCIENCE AND ENGINEERING

Submitted by

| | |
|---|---|
| **RUDROJU GOPICHAND** | **19UK1A05A5** |
| **NARLAPURAM SAI PRIYA** | **19UK1A05C9** |
| **SRIPATHI HARIKA** | **19UK1A0595** |
| **ALLAPURAM MANIKANTESH** | **19UK1A05C1** |

Under the esteemed guidance of

**Ms.T. SRAVANTHI**

(Assistant Professor)



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
## VAAGDEVI ENGINEERING COLLEGE

(Affilated to JNTUH,Hyderabad)
Bollikunta ,Warangal-506005
**2019-2023**

I

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## VAAGDEVI ENGINEERING COLLEGE
### BOLLIKUNTA,WARANGAL – 506005
### 2019-2023

## CERTIFICATE OF COMPLETION

## UG PROJECT PHASE-1

This is to certify that the UG Project Phase-1 entitled **"HEART STROKE PREDICTION USING MACHINE LEARNING"** is being submitted by *R.GOPICHAND* **(H.NO:19UK1A05A5) ,** *N.SAIPRIYA***(H.NO:19UK1A05C9),***S.HARIKA***(H.NO:19UK1A0595),***A.MANIKANTESH***(H.NO: 19UK1A05C1)** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** to **Jawaharlal Nehru Technological University Hyderabad** during the academic year **2022-23,**is a record of work carried out by them under the guidance and supervision.

**Project Guide**                                                    **Head of the Department**
**Ms.T.Sravanthi**                                                    **Dr.R.Naveen Kumar**
(Assistant Professor)                                                (Professor)

**External**

II

# ACKNOWLEDGEMENT

**R.GOPICHAND**       (19UK1A05A5)
**N.SAI PRIYA**       (19UK1A05C9)
**S.HARIKA**          (19UK1A0595)
**A.MANIKANTESH**   (19UK1A05C1)

# ABSTRACT

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. In recent times, Heart Stroke prediction is one of the most complicated tasks in the medical field. In the modern era, approximately one person dies per minute due to heart Stroke. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart stroke prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance.The project aims to predict the chances of Heart Stroke and classifies the patient's risk level by implementing different Machine Learning techniques such as KNN, Decision Tree and Random Forest. From these models the Best performing model is selected and saved. Here we will be building a flask application that uses a machine learning model to get the prediction of heart stroke. We will also train our model on IBM Cloud and deployment on IBM Cloud.

In recent times, Heart Stroke prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart Stroke. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart stroke prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. This paper makes use of heart stroke dataset. The proposed work predicts the chances of Heart Stroke and classifies patient's risk level by implementing different data mining techniques such as KNN, Decision Tree and Random Forest. Thus, this paper presents a comparative study by analyzing the performance of different machine learning algorithms. The trial results verify that Random Forest algorithm has achieved the highest accuracy of 99.17% compared to other ML algorithms implemented.

# TABLE OF CONTENTS :-

# LIST OF FIGURES                                    PAGE NO

# 1.INTRODUCTION

## 1.1. MOTIVATION:

Machine learning can be used to detect whether a person is suffering from a cardiovascular disease by considering certain attributes like chest pain,hypertension,smoking status,age of person and some other attributes.Amini conducted research to predict stroke incidence, collected 807 healthy and unhealthy subjects in their study categorized 50 risk factors for stroke, diabetes, cardiovascular disease, smoking, hyperlipidemia, and alcohol use. They used two techniques that had the best accuracy from c4.5 decision tree algorithm, and it was 95%, and for K-nearest neighbor, the accuracy was 94%.

## 1.2. PROBLEM DEFINITION:

Many researchers have already used machine learning based approached to predict heart strokes Cheng published a report on the estimation of the heart stroke prognosis. In their analysis, 82 stroke patient data were used, two ANN models were used to find precision, and 79% and 95% were used. Cheon performed a study to predict stroke patient mortality. In their study, they used 15099 patients to identify heart stroke occurrence. They used a deep neural network approach to detect heart strokes. The authors used PCA to extract medical record history and predict heart stroke.

## 1.3. PROJECT OBJECTIVE:

The main aim of this study is to effectively predict if the patient suffers from heart stroke.The health professionals enters the input values from the patient's health report.The data is fed into model which predicts the probability of having heart stroke.The diagnosis of heart disease is usually based on signs ,symptoms and physical examination of the patients.

## 1.4. LIMITATIONS OF PROJECT:

Individuals can obtain appropriate therapy more rapidly if they are observed and have their biosignals detected and precisely assessed in real-time. Stroke is a prominent cause of disability in adults and the elderly, resulting in a slew of social and financial issues. Stroke can be fatal if left untreated. we developed a stroke prediction system that detects stroke using real-time bio-signals with machine learning techniques. The purpose of this study was to analyse and diagnose electromyography data using machine learning.

## 1.4. ORGANIZATIONAL OF DOCUMENTATION:

Various classification methods are investigated, and the best accurate model for predicting the patient's heart condition is found. Numerous category methods are investigated, and the pleasant correct version for predicting the affected person's heart condition is discovered.The goal is to determine whether the patient has any type of cardiac disease. The input values from the patient's health report are entered by the health professional. The information is then incorporated into a machine learning algorithm.

# 2.PROBLEM STATEMENT

Heart disease can be managed effectively with a combination of lifestyle changes,medicine and in some cases surgery.The symptoms of heart disease can be reduced and the functionality of the heart is improved.The predicted results can be used to prevent and thus reduces the cost of surgical treatment sand other expensieve.To develop a heart disease prediction model with improved and enhanced accuracy.The specific objectives are to quickly identify new patients,reduce diagnostic time,reduce heart attacks and save lives.By using the Machine Learning algorithms and doing some performance analysis , it predicts the person from the heart stroke.It aims to predict heart stroke prediction using machine learning with the help of some health vitals of patients. The proposed work predicts the chances of Heart Stroke and classifies patient's risk level by implementing different ML Algorithms.

In a wide range of research projects, therefore, the employment of a verified subtype classification system that allows for comparisons of outcomes is vital. The method is used to select a subset of important features (variables, predictors) for use in the creation of a model in machine learning and statistics, also known as variable choices, attribute selections, or variable subset choices.In this research to predict whether a person will have a heart attack or not depending on numerous input characteristics such as age, gender, smoking status, employment type, and so on. The dataset is trained on a variety of machine learning methods, and the results are analyzed to see which one is the most successful in predicting heart stroke. To demonstrate the comparative and analysis study of each method, the accuracies acquired from each algorithm

# 3.LITERATURE SURVEY

## 3.1.INTRODUCTION:

Present day,Deaths due to heart strokes are increasing greatly day by day.Unfortunately,detectingsuch conditions in humans is a complex task.Handling such complex tasks can be done by using data sets. Heart strokes occurrence prediction can be done only by automation because we need to keep monitoring the heart rate. A machine learning-based approach has previously been utilized by many academics to predict cardiac attacks. Machine learning can be used to detect whether a person is suffering from a cardiovascular disease by considering certain attributes like chest pain, cholesterol level, age of the person and some other attributes.

## 3.2. EXISTING SYSTEM:

We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparision to the previously used classifier such as naïve bayes etc.

## 3.3. DISADVANTAGES OF EXISTING SYSTEM:

1. Prediction of cardiovascular disease results is not accurate.
2. Cannot handle enormous datasets for patient records.

## 3.4. PROPOSED SYSTEM:
The proposed work predicts heart stroke by exploring the above mentioned four classification algorithms and does performance analysis. We build a model by applying various machine learning algorithms to find the best accurate model.
We have used 4 different typesof ML algorithms to find the performance metrics for this project. They are

## 1.Decision Tree Algorithm:

Decision Tree algorithm comes under Managed Learning that can be used to solve both classification and regression problems. It is mainly based upon selecting the order of the nodes (attributes).
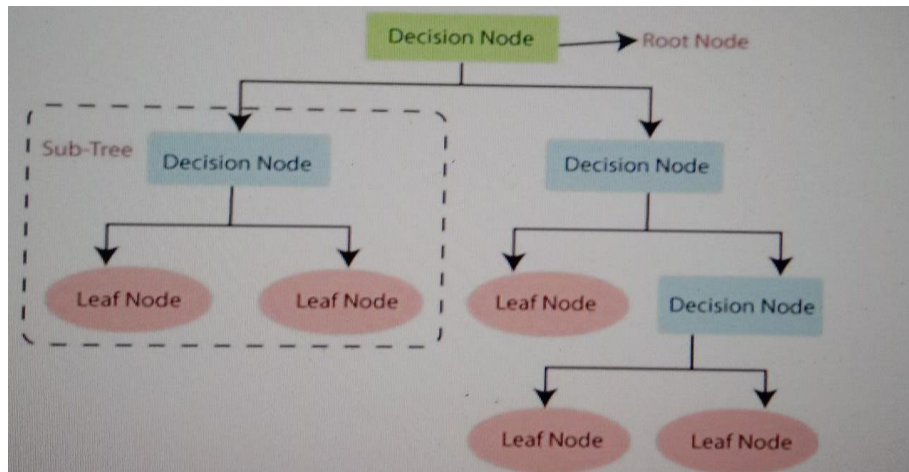


**Fig.1: Decision Tree Model**

## 2.NaïveBayesalgorithm:

The Naïve Bayes Classifier takes the reference of the bayes theorem, Bayesian classifier are the statistical classifiers. It contains evidence, hypothesis, likelihood.

A –HypothesisandB –EvidenceandP(B|A) –Likelihood
P(A|B) –it is a posterior probabilityandP(H) –Prior probability



$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

**Fig.2: Naive Bayes Model**

### 3.Random Forest algorithm:

A random forest algorithm is a classifier that combines a large no.of decision trees from different subsets of a dataset and averages the predictability of the dataset. It is a controlled machine learning algorithm commonly used to resolve classification and regression problems.
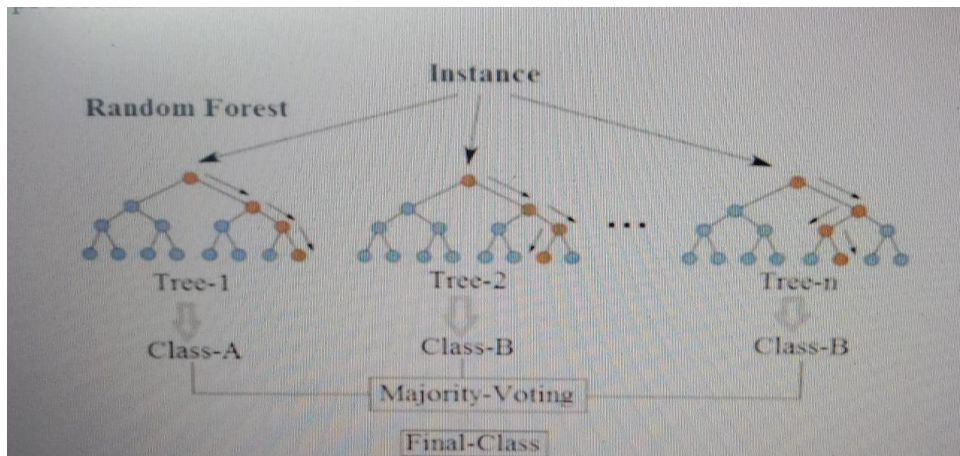


**Fig.3: Random Forest**

### 4. K-Nearest Neighbor algorithm:

The K-classifier nearest neighbor in the pattern searches for space training patterns and selects the pattern closest to the unknown pattern. The K-Nearest Neighbor method is a non-parametric algorithm that takes all existing data and organizes a new data point based on likeness.
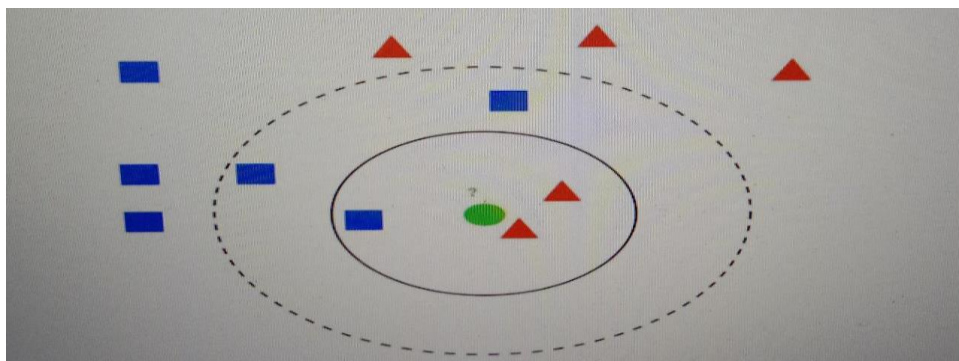


**Fig.4: K-Nearest Neighbor**

5

# 4.EXPERIMENTAL ANALYSIS

The health professional enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart stroke.The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes.This project gives us significant knowledge that can help us predict the patients with the heart disease.

**Data Collection**

Machine Learning heavily depends on the data, without data,a machine can't learn.In machine learning projects, we need a training dataset to train the model.We can collect datasets,from the different open sources like kaggle.com,data.gov etc.The dataset used for this project was obtained from kaggle.

**Data Pre-Processing**

It performs the following main tasks
- Importing the libraries
- Importing the dataset
- Analyse the data
- Taking care of missing Data
- Data Visualization
- Splitting Data into Train and Test

**Model Training**

The model building process involves setting up ways of collecting data.Model Building Includes:
- Import the model building libraries
- Initializing the model
- Training the model
- Model Evaluation
- Save the model

**Application Building**
- Create an HTML File
- Build python code
- Run the app in browser
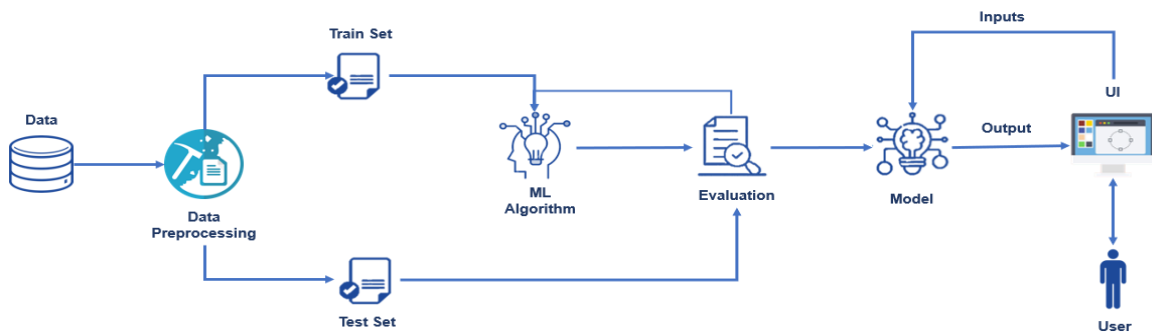- Showing the prediction on UI/Open CV window

# 4.1. PROJECT ARCHITECTURE



**Fig.5: Project Architecture**

# 4.2. SOFTWARE AND HARDWARE REQUIREMENTS:

.

The following are the software's that are required for the project:
- Anaconda Navigator
- Jupyter Notebook
- Flask
- And some other python packages that are need to be installed.

We must require the following Hardware to complete this project.
- Internet connection to download and activate.
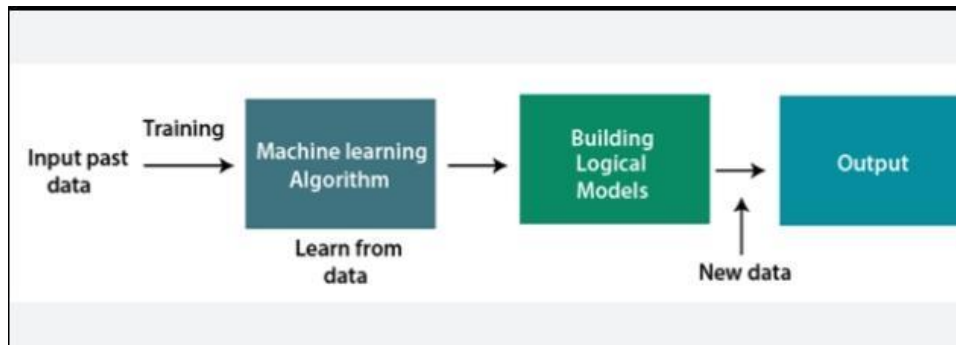- Administration access to install and run Anaconda Navigator

## 4.3. BLOCK DIAGRAM



**Fig.6: Block Diagram**

## 4.4. PROJECT FLOW

The dataset is collected from the patient's then heart stroke prediction is performed the use of a dataset inside the suggested model. primarily based on symptoms which include age, gender, average glucose degree, smoking popularity, body mass index, employment type, and residing type, the version forecasts the probability of a person having a stroke. It uses system mastering strategies which includes Random Forest, K Nearest Neighbor (KNN), Decision Tree, to classify someone's risk level. As a result, an assessment of the various algorithms is given, and the maximum is determined.
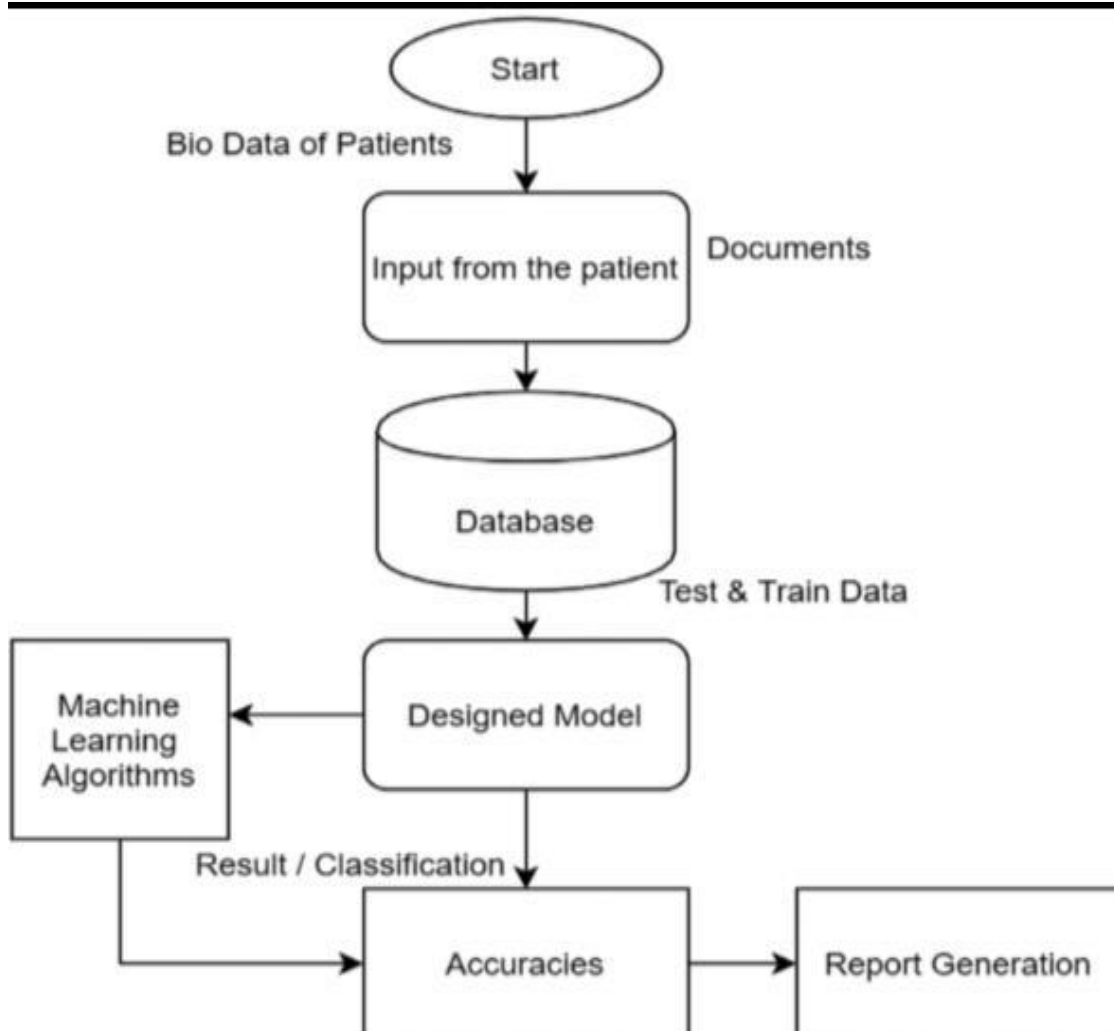
# 5.FLOWCHART



**Fig.7: Flowchart**

# 6. CONCLUSION

With the increasing number of deaths due to heart stroke's, it has become mandatory to develop a system to predict heart stroke effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart stroke. This study compares the accuracy score of Random Forest, Decision Tree and KNN algorithms for predicting heart stroke using kaggle dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 99.17% for prediction of heart stroke. In future the work can be enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the heart disease effectively and efficiently.

The Heart Stroke Prediction assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease.

# 7. FUTURE SCOPE

- Depending on the patient's cardiac forecast,their physician may begin appropriate medication to manage high blood pressure,diabetis,heart problems etc.
- System will be predict from heart stroke before being deployed in hospitals.