**HEART ATTACK ANALYSIS AND PREDICTION**

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

by

**D. SRINIDHI (2103A52079)**

**K. MADHU SRI (2103A52090)**

**V. SHIVANI (2103A52188)**

Under the guidance of

**Mr. D. RAMESH**

Assistant Professor, Department of CSE.

**

**Department of Computer Science and Artificial Intelligence**

**

**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that project entitled **“HEART ATTACK ANALYSIS AND PREDICTION**" is the bonafied work carried out by **D. SRINIDHI, K. MADHUSRI, V.SHIVANI** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

**Mr. D. RAMESH Dr. M. Sheshikala**

Asst. Professor, Assoc. Prof .& HOD (CSE)

S R University, S R University,

Ananthasagar ,Warangal Ananthasagar ,Warangal

**ACKNOWLEDGEMENT**

We express our thanks to Course co-coordinator **Mr.** **S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

**ABSTRACT**

Heart Attack Prediction using Machine Learning Technique in Big data analytics has started to play an important role in the healthcare practices and research. heart attack prediction will be found primarily on real-time processing, distributed and real-time classification and distribution, storage so; databases can be easily modified by the doctors. If you know all the attributes related to our health we can check easily how much chance to the Heart attack risk, using the system applications. It was recently used to train classification models. After that using extract the features that is condition to be find to be classified by Decision Tree (DT).Compared to existing; algorithms provides better performance. After classification, performance criteria including accuracy, precision, F-measure is to be calculated. If you are concern about the heart attack risks, you might be referred to a heart specialist. Some attributes are Heart Attack risk factors including which is the High blood pressure, high cholesterol and diabetes, increases your risk even more. Hence we are also checking your symptoms of heart attack and take about prevention.

**Table of Contents**

**Chapter No. Title Page No.**

1. Introduction
   1. Overview 1
   2. Problem Statement 3
   3. Existing system 7
   4. Proposed system 9
   5. Objectives 12
   6. Architecture 14
2. Literature survey
   * 1. Document the survey done by you 15
3. Data pre-processing

3.1. Dataset description 18

3.2. Data cleaning 19

3.3. Data augmentation 20

3.4. Data Visualization 21

1. Methodology
   1. Procedure to solve the given problem 22
   2. Model architecture 23

4.3. Software description 24

1. Results and discussion 25
2. Conclusion and future scope ………
3. References

**INTRODUCTION**

* 1. **OVERVIEW**

Heart disease is one of the biggest health risks for association today. According to the World Health Organization (WHO), stroke and heart attack are the most common cause of global death (85%). Therefore, the availability of data and data mining techniques, especially machine learning and early detection of Heart Attack, can help patients to anticipate a potential disease response. In the healthcare field, it is becoming more and more common nowadays to source large amounts of data (big data), streaming machines, advanced healthcare services, high throughput instruments, sensor networks, Internet of Things, mobile application applications, data archiving and processing, from many areas. The heart attack prediction is most significant and important duty in medical field which Requires more attention. However there are some techniques for data collection and analysis. Also huge set of medical data is required to correctly predict the heart attack.

**1.2. PROBLEM STATEMENT**

Heart Attack is one of the huge health risks for human’s healthy life. big data growth in medical and healthcare association today, early solution and accurate analysis of medical data benefits through patient care and community services. If the quality of medical data some data are incomplete, the accuracy of the analysis decreases. So we need huge medical data to predict the heart attack. The aim of heart attack analysis and prediction is to Design and Implement the Heart Attack analysis and Prediction System using machine learning techniques.

**1.3. EXISTING SYSTEM**

In this system, the input details are obtained from the patients. Then from the user inputs, and Using machine learning techniques heart attack is analyzed. Now the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of patient is collected from the UCI laboratory and used to discover the patterns with the existed machine learning models. The existing systems of predicting the heart attack are K nearest neighbours, support vector machines (SVM) and naive bayes. The results are compared for performance and accuracy with these machine learning algorithms.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

In this system, the input details are obtained

from the patient. Then from the user inputs,

using ML techniques heart disease is analyzed.

Now, the obtained results are compared with the

results of existing models within the same

domain and found to be improved. The data of

heart disease patients collected from the UCI

laboratory is used to discover patterns with NN,

DT, Support Vector machines SVM, and Naive

Bayes. The results are compared for

performance and accuracy with these

algorithms. The proposed hybrid method returns

results of 87% for F-measure, competing with

the other existing methods.

**1.4. PROPOSED SYSTEM**

The proposed work predicts heart disease by exploring the above mentioned four classification algorithms and does performance analysis. The objective of this study is to effectively predict if the patient suffers from heart disease. 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 disease. Figure shows the entire process involved.



Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which

requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge

set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure,

and blood sugar, it is possible to increase the possibility of heart diseases prediction



Data collection

classification



Testing data

Training data



Result

Test the model

Classification techniques



**1.5. OBJECTIVES**

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart Disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases. The prediction Provides the new approach to concealed patterns in the data. It Helps to avoid human biasness and also Reduce the cost of medical tests and gives accurate result whether the patient gets heart attack or not.

* 1. **ARCHITECTURE**

The architecture of the proposed system is as displayed in the figure below. The major components of the architecture are as follows: patient database, preprocessing, tokenization, training the model, test the model, design fitness function, application of genetic algorithm, results collection and prediction of heart disease.

Patient data base

Pre processing

tokenisation



Prediction of heart attack

Training model

Testing the model

Collection of result

Applying the algorithm

Design fit function of algorithm

**2.1.1 LITERATURE SURVEY**

There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centres.

K. Polaraju et al, [7] proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Marjia et al, [8] developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using k-fold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory.

S. Seema et al,[9] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

Ashok Kumar Dwivedi et al, [10] recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms

MeghaShahi et al, [11] suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other

data mining algorithms.

Jayami Patel et al, [14] suggested heart disease prediction using data mining and the machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques. The best algorithm J48 based on UCI data has the highest accuracy rate compared to LMT.

**3.DATA PRE-PROCESSING**

**3.1.1 DATASET DESCRIPTION**

|  |  |  |
| --- | --- | --- |
| **Sno** | **Attributes** | **Description** |
| 1.  2.  3.  4.  5.  6.  7.  8.  9.  10  11.  12.  13.  14. | Age  Gender  cp  chol  fbs  thalach  thal  ca  old peak  restecg  exng  sla  Out put | Age of the patient  Gender of the patient (male or female)  Chest pain type  Cholesterol  Fasting blood sugar  Maximum heart rate achieved in beats per minute  Displays the thalassemia  Number of major vessels  ST depression induced by exercise related to rest  Resting electro cardio graphic results  Exercise induced angina  Predicts whether patient gets heart attack or not |

**3.2 DATA CLEANING**

Data quality has become an important issue. This issue becomes more and more important in medicine area, where the need for effective decision making is high. In this context, the need for data cleaning to improve data quality is becoming crucial. Duplicate records elimination is a challenging data cleansing task. In this paper, we present a duplicate records elimination approach to improve the quality of data. We propose a deep learning-based approach for duplicate records detection using a sentence embeddings model. Also, we propose an algorithm for duplicated records correction. Then, we apply the proposed duplicate records elimination approach to analyze the effect of data cleaning on the quality of decisions.

Experiments show that the classification performance improves upon the application of the duplicate records elimination approach on datasets compared to that of datasets with duplicate records.

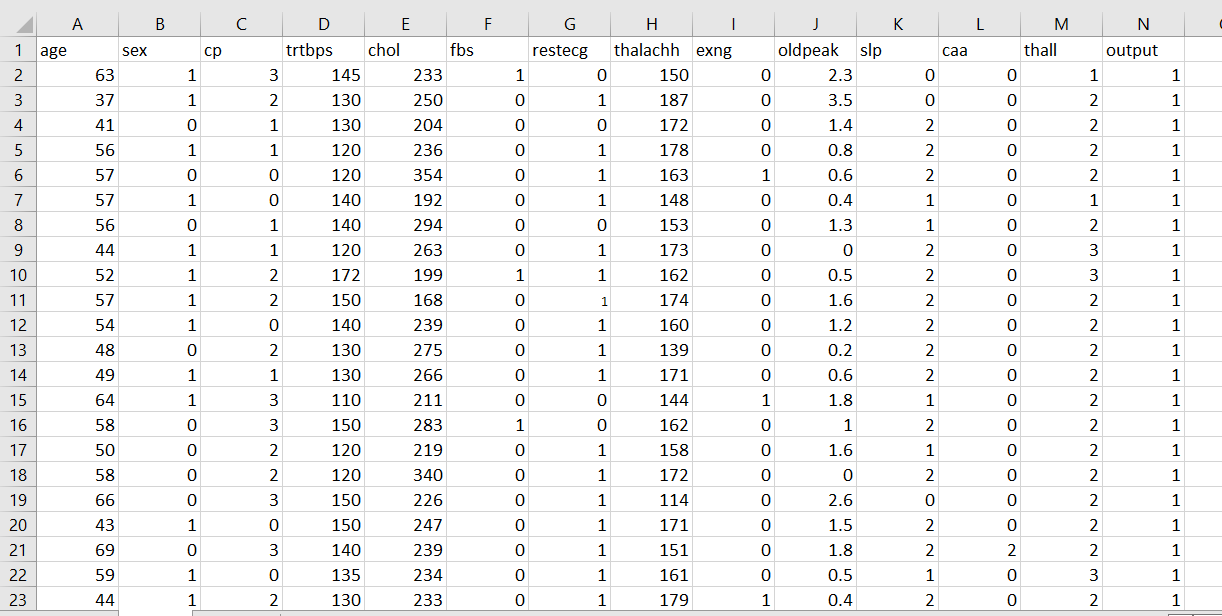
**3.3 DATA AUGUMENTATION**

Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points, Machine learning applications especially in the deep learning domain continue to diversify and increase rapidly. [Data-centric approaches to model development](https://research.aimultiple.com/data-centric-ai/) such as data augmentation techniques can be a good tool against challenges which the artificial intelligence world faces.

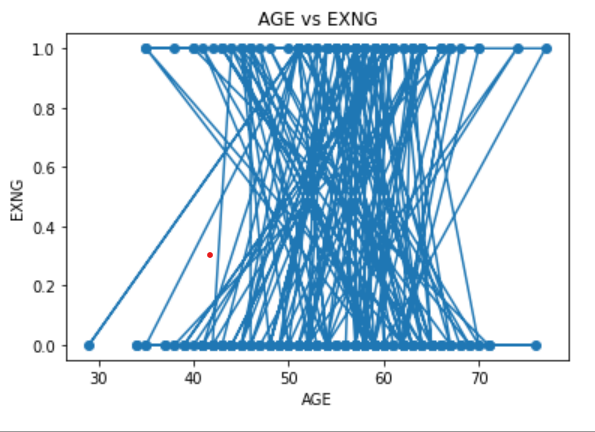
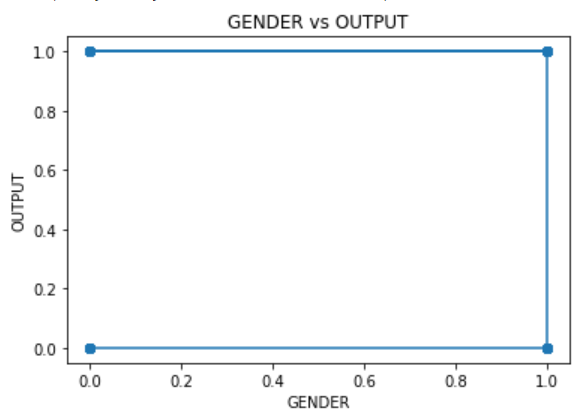
Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. If the dataset in a machine learning model is rich and sufficient, the model performs better and more accurately.

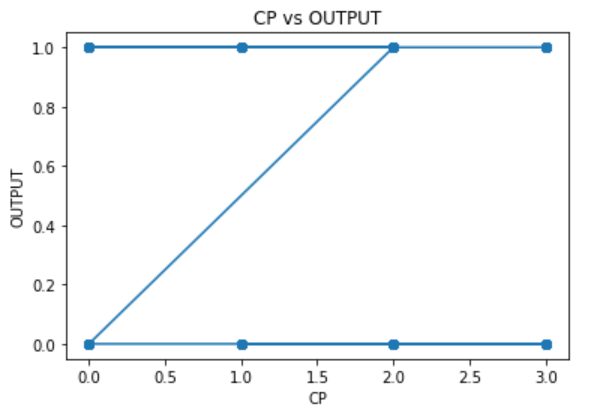
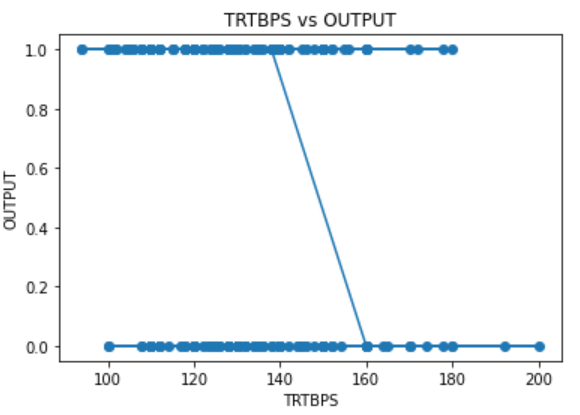
**3.4 DATA VISUALISATION**

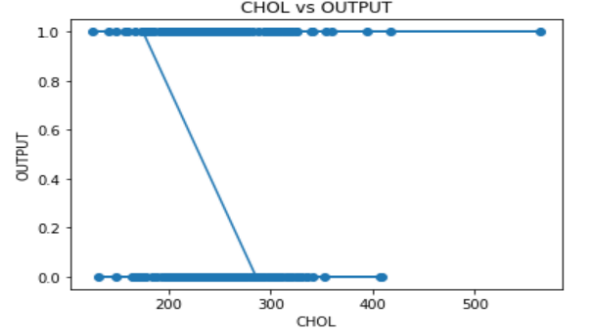
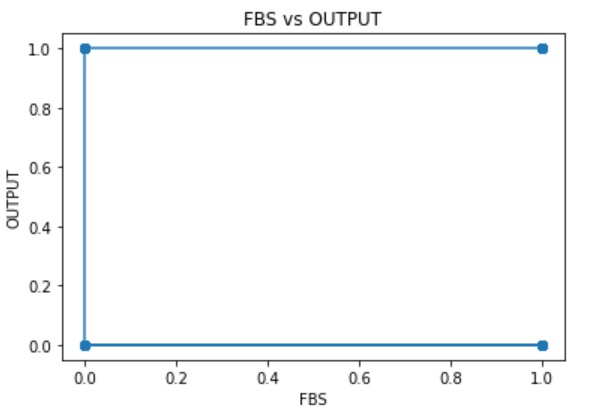
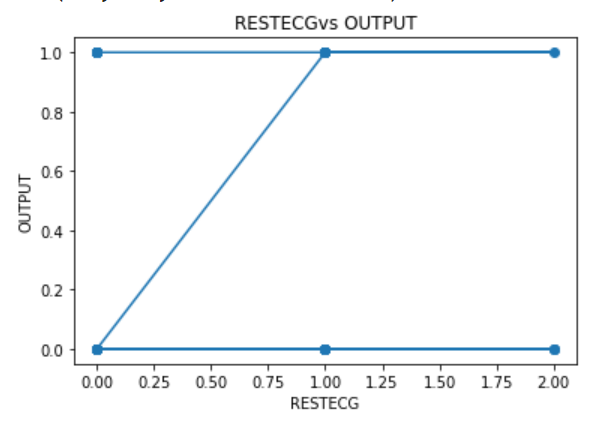
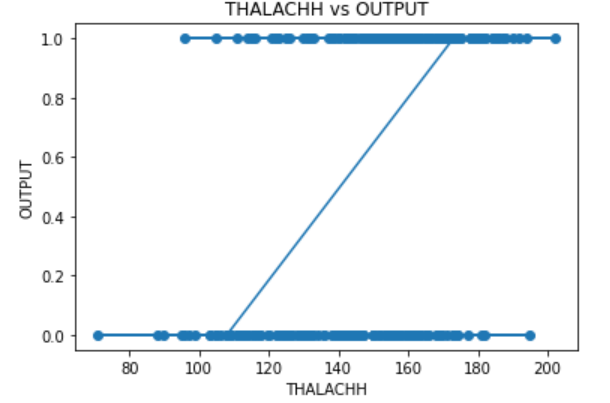
The heart attack analysis and prediction data set contains thirteen feature variables and one target variable output which contains symptoms of patients as features and whether the patient get heart attack or not as target

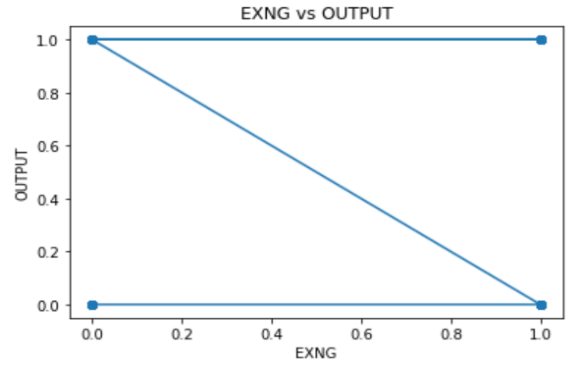
**DATASET**

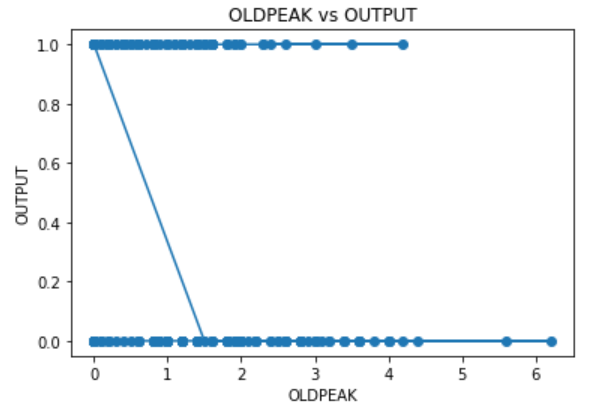
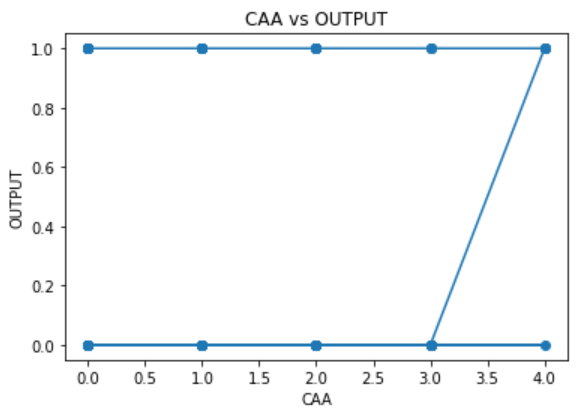
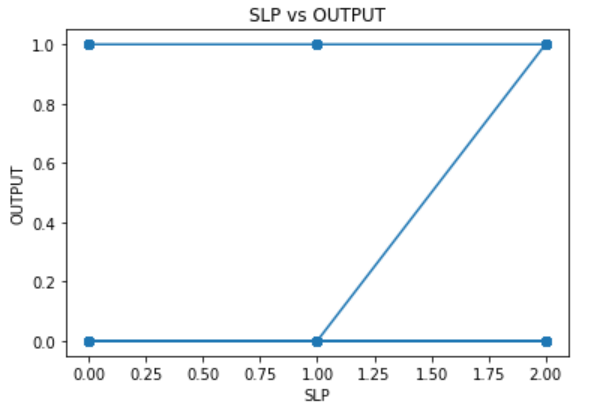
**GRAPHS PLOTTED BETWWEN FEATURE AND TARGET VARIABLES:**

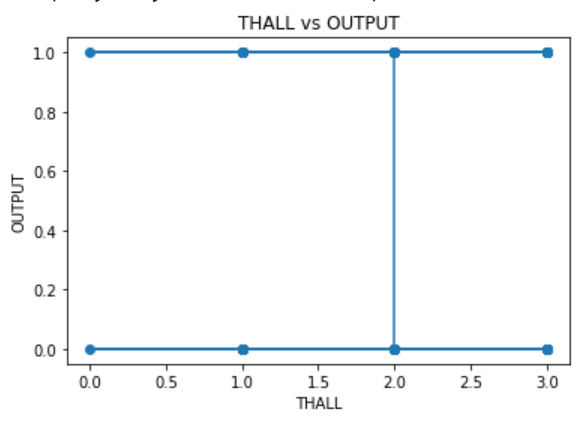
****



********





**4. METHODOLOGY**

**4.1 PROCEDURE TO SOLVE THE GIVEN PROBLEM**

In this project of heart attack analysis and prediction we use two approaches:

* Logistic regression
* KNN (K-Nearest neighbours)

**Logistic regression**

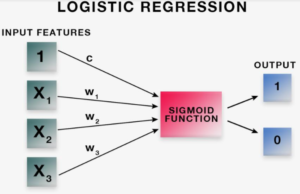
Logistic regression is a supervised machine learning algorithm.

This type of statistical model ,is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn’t vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure.

**Advantages of the Logistic Regression Algorithm:**

* Logistic regression performs better when the data is linearly separable
* It does not require too many computational resources as it’s highly interpretable
* There is no problem scaling the input features—It does not require tuning
* It is easy to implement and train a model using logistic regression
* It gives a measure of how relevant a predictor (coefficient size) is, and its direction of association (positive or negative)





**K-Nearest neighbours**

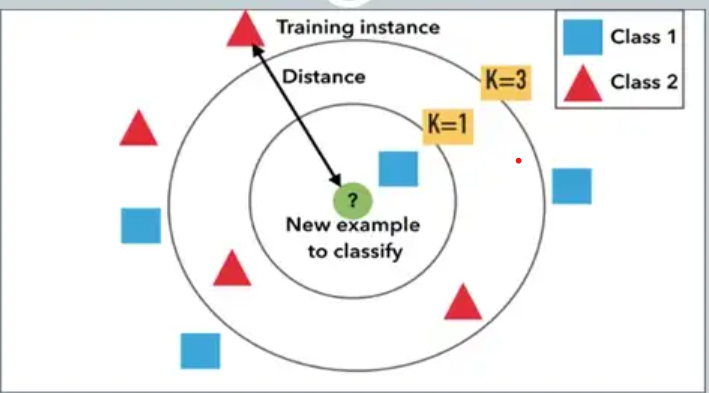
KNN is a non-parametric machine learning algorithm. The KNN algorithm is a supervised learning method. This means that all the data is labelled and the algorithm learns to predict the output from the input data. It performs well even if the training data is large and contains noisy values. The data is divided into training and test sets. The train set is used for model building and training. Ak- value is decided which is often the square root of the number of observations. Now the test data is predicted on the model built. There are different distance measures. For continuous variables, Euclidean distance, Manhattan distance and Minkowski distance measures can be used.

However, the commonly used measure is Euclidean distance. The formula for Euclidean distance is as follows:

**How does K-NN work?**

The K-NN working can be explained on the basis of the below algorithm:

* **Step-1:** Select the number K of the neighbours
* **Step-2:** Calculate the Euclidean distance of **K number of neighbours**
* **Step-3:** Take the K nearest neighbours as per the calculated Euclidean distance.
* **Step-4:** Among these k neighbours, count the number of the data points in each category.
* **Step-5:** Assign the new data points to that category for which the number of the neighbour is maximum.
* **Step-6:** Our model is ready.

****

**4.2 MODEL ARCHITECTURE**

HEART PATIENT DATA BASE COLLECTED FROM HOSPITALS



LOADING DATA SET



IDENTIFYING THE ATTRIBUTES PERTAING THE RISK OF HEART ATTACK



COLLECTION OF DATA AND PRE -PROCESSING

-



LOGISTIC REGRESSION OR K-NEAREST NEIGHBOURS



ARTIFICIAL NEURAL NETWORKS



OBTAIN RESULTS



CONCLUSION

**4.3 SOFTWARE DESCRIPTION**

**Software requirements:**

**Operating system:** Windows

**Platform**: google colab

**Programing language:** python

**5. RESULTS**

**CODE:**

**Logistic regression :**

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

data=pd.read\_csv('/content/stml data set.csv')

print(data)

x=data.iloc[:,0:8]

y=data.iloc[:,8:9]

from sklearn.preprocessing import StandardScaler

stsc=StandardScaler()

data=stsc.fit(x)

dd=stsc.transform(x)

print(data)

print(dd)

print(x)

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.25,random\_state=True)

print(x\_train.shape)

print(y\_train.shape)

print(x\_test.shape)

print(y\_test.shape)

lr= LogisticRegression(random\_state = 88)

mm=lr.fit(x\_train,y\_train)

print(mm.score(x\_train,y\_train))

print(mm.score(x\_test,y\_test))

yp=mm.predict(x\_test)

from sklearn.metrics import accuracy\_score

print(accuracy\_score(yp,y\_test))

from sklearn.metrics import classification\_report

print(classification\_report(yp,y\_test))

from sklearn import metrics

metrics.plot\_roc\_curve(mm,x\_test,y\_test)

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import ConfusionMatrixDisplay

cm=confusion\_matrix(yp,y\_test)

d=ConfusionMatrixDisplay(cm).plot()

**RESULT:**

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp \

0 63 1 3 145 233 1 0 150 0 2.3 0

1 37 1 2 130 250 0 1 187 0 3.5 0

2 41 0 1 130 204 0 0 172 0 1.4 2

3 56 1 1 120 236 0 1 178 0 0.8 2

4 57 0 0 120 354 0 1 163 1 0.6 2

.. ... ... .. ... ... ... ... ... ... ... ...

298 57 0 0 140 241 0 1 123 1 0.2 1

299 45 1 3 110 264 0 1 132 0 1.2 1

300 68 1 0 144 193 1 1 141 0 3.4 1

301 57 1 0 130 131 0 1 115 1 1.2 1

302 57 0 1 130 236 0 0 174 0 0.0 1

caa thall output

0 0 1 1

1 0 2 1

2 0 2 1

3 0 2 1

4 0 2 1

.. ... ... ...

298 0 3 0

299 0 3 0

300 2 3 0

301 1 3 0

302 1 2 0

StandardScaler()

[[ 0.9521966 0.68100522 1.97312292 ... 2.394438 -1.00583187

0.01544279]

[-1.91531289 0.68100522 1.00257707 ... -0.41763453 0.89896224

1.63347147]

[-1.47415758 -1.46841752 0.03203122 ... -0.41763453 -1.00583187

0.97751389]

...

[ 1.50364073 0.68100522 -0.93851463 ... 2.394438 0.89896224

-0.37813176]

[ 0.29046364 0.68100522 -0.93851463 ... -0.41763453 0.89896224

-1.51512489]

[ 0.29046364 -1.46841752 0.03203122 ... -0.41763453 -1.00583187

1.0649749 ]]

age sex cp trtbps chol fbs restecg thalachh

0 63 1 3 145 233 1 0 150

1 37 1 2 130 250 0 1 187

2 41 0 1 130 204 0 0 172

3 56 1 1 120 236 0 1 178

4 57 0 0 120 354 0 1 163

.. ... ... .. ... ... ... ... ...

298 57 0 0 140 241 0 1 123

299 45 1 3 110 264 0 1 132

300 68 1 0 144 193 1 1 141

301 57 1 0 130 131 0 1 115

302 57 0 1 130 236 0 0 174

[303 rows x 8 columns]

(227, 8)

(227, 1)

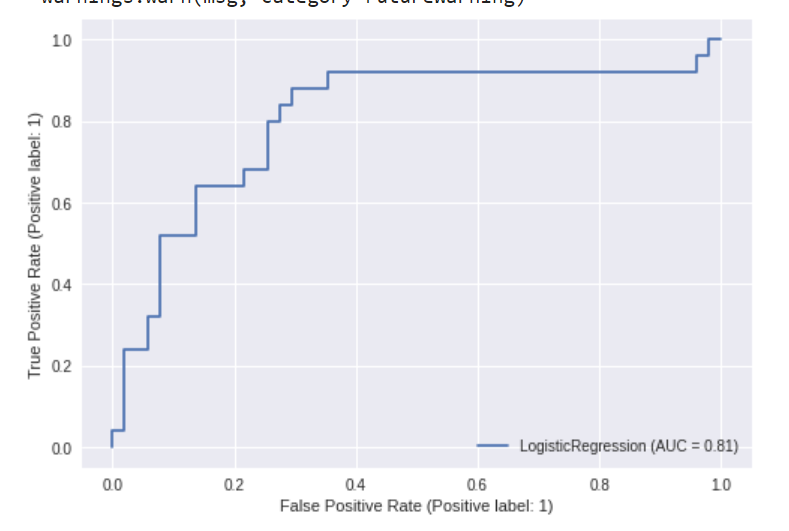
(76, 8)

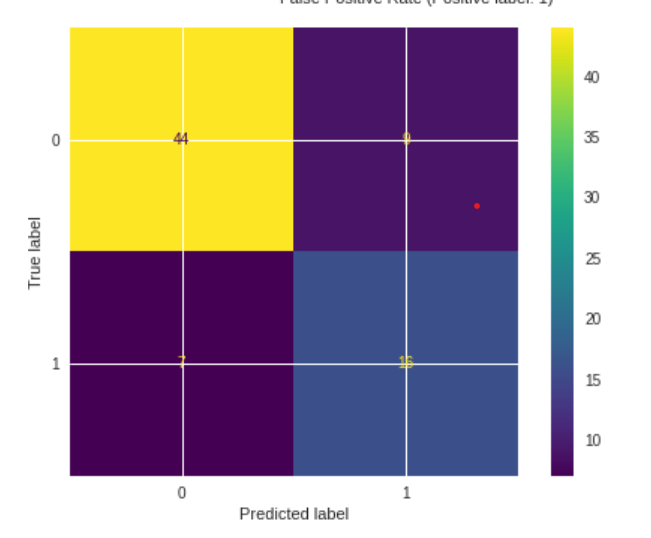
(76, 1)

0.801762114537445

0.7894736842105263

0.7894736842105263

****

****

**KNN:**

import pandas as pd

import matplotlib.pyplot as plt

from sklearn import numpy as np

.datasets import make\_blobs

from sklearn.neighbors import KNeighborsClassifier

from sklearn.model\_selection import train\_test\_split

data=pd.read\_csv('/content/stml data set.csv')

print(data)

X, y = make\_blobs(n\_samples = 500, n\_features = 2, centers = 4,cluster\_std = 1.5, random\_state = 4)

plt.style.use('seaborn')

plt.figure(figsize = (10,10))

plt.scatter(X[:,0], X[:,1], c=y, marker= '\*',s=100,edgecolors='black')

plt.show()

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, random\_state = 0)

knn1 = KNeighborsClassifier(n\_neighbors = 1)

knn3 = KNeighborsClassifier(n\_neighbors=3)

knn1.fit(X\_train, y\_train)

knn3.fit(X\_train, y\_train)

y\_pred\_1 = knn1.predict(X\_test)

y\_pred\_3 = knn3.predict(X\_test)

from sklearn.metrics import accuracy\_score

print("Accuracy with k=1", accuracy\_score(y\_test, y\_pred\_1)\*100)

print("Accuracy with k=3", accuracy\_score(y\_test, y\_pred\_3)\*100)

plt.figure(figsize = (15,5))

plt.subplot(1,2,1)

plt.scatter(X\_test[:,0], X\_test[:,1], c=y\_pred\_5, marker= '\*', s=100,edgecolors='black')

plt.title("Predicted values with k=1", fontsize=20)

plt.subplot(1,2,2)

plt.scatter(X\_test[:,0], X\_test[:,1], c=y\_pred\_1, marker= '\*', s=100,edgecolors='black')

plt.title("Predicted values with k=3", fontsize=20)

plt.show()

**RESULT:**

age sex cp trtbps chol fbs restecg thalachh exng oldpeak slp \

0 63 1 3 145 233 1 0 150 0 2.3 0

1 37 1 2 130 250 0 1 187 0 3.5 0

2 41 0 1 130 204 0 0 172 0 1.4 2

3 56 1 1 120 236 0 1 178 0 0.8 2

4 57 0 0 120 354 0 1 163 1 0.6 2

.. ... ... .. ... ... ... ... ... ... ... ...

298 57 0 0 140 241 0 1 123 1 0.2 1

299 45 1 3 110 264 0 1 132 0 1.2 1

300 68 1 0 144 193 1 1 141 0 3.4 1

301 57 1 0 130 131 0 1 115 1 1.2 1

302 57 0 1 130 236 0 0 174 0 0.0 1

caa thall output

0 0 1 1

1 0 2 1

2 0 2 1

3 0 2 1

4 0 2 1

.. ... ... ...

298 0 3 0

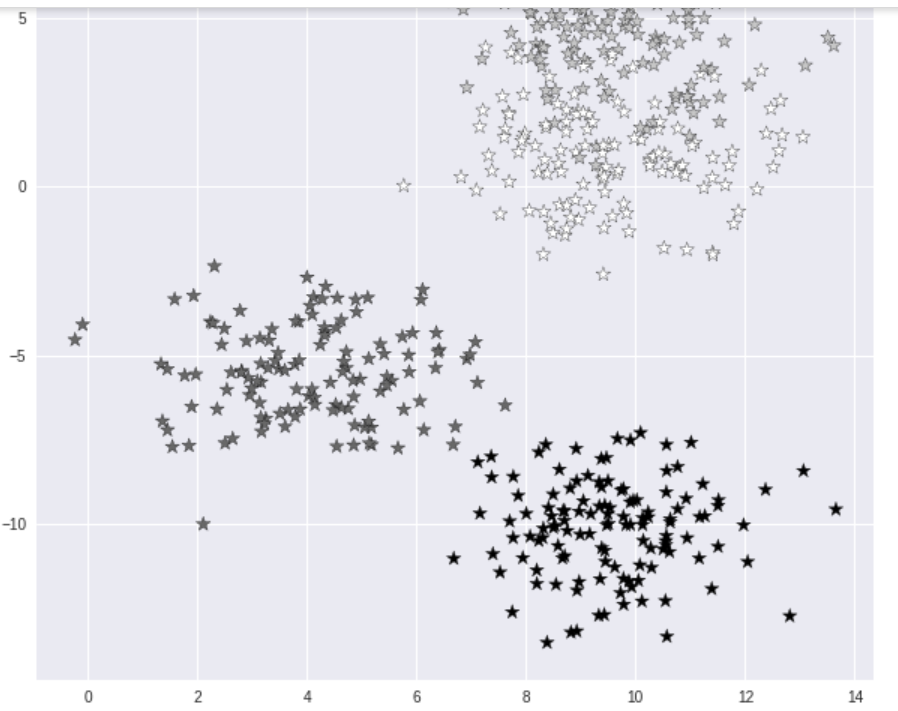
299 0 3 0

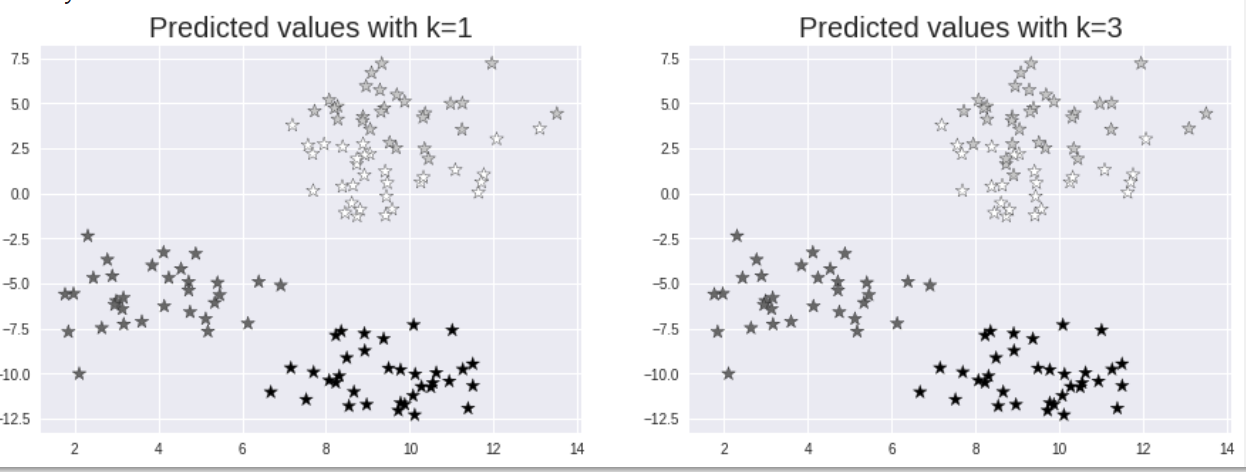
300 2 3 0

301 1 3 0

302 1 2 0

[303 rows x 14 columns]

****

****

Accuracy with k=1 90.4

Accuracy with k=3 92.80000000000001

**6. CONCLUSION AND FUTURE SCOPE**

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system.

The proposed working model can also help in reducing treatment costs by providing Initial

diagnostics in time. The model can also serve the purpose of training tool for medical students

and will be a soft diagnostic tool available for physician and cardiologist. General physicians can

utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements

that could be explored to improve the scalability and accuracy of this prediction system. As we

have developed a generalized system, in future we can use this system for the analysis of

different data sets. The performance of the health’s diagnosis can be improved significantly by

handling numerous class labels in the prediction process, and it can be another positive direction

of research. In DM warehouse, generally, the dimensionality of the heart database is high, so

identification and selection of significant attributes for better diagnosis of heart disease are very

challenging tasks for future research

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system.

The proposed working model can also help in reducing treatment costs by providing Initial

diagnostics in time. The model can also serve the purpose of training tool for medical students

and will be a soft diagnostic tool available for physician and cardiologist. General physicians can

utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements

that could be explored to improve the scalability and accuracy of this prediction system. As we

have developed a generalized system, in future we can use this system for the analysis of

different data sets. The performance of the health’s diagnosis can be improved significantly by

handling numerous class labels in the prediction process, and it can be another positive direction

of research. In DM warehouse, generally, the dimensionality of the heart database is high, so

identification and selection of significant attributes for better diagnosis of heart disease are very

challenging tasks for future research

This paper discusses the various machine learning such as logistic regression and k- nearest neighbour which were applied to the data set. It utilizes the data such as blood pressure, cholesterol, diabetes and then tries to predict the patient gets heart attack or not. Family history of heart disease can also be a reason for developing a heart disease as mentioned earlier. So, this data of the patient can also be included for further increasing the accuracy of the model. This work will be useful in identifying the possible patients who may suffer from heart disease in the next 10 years. This may help in taking preventive measures and hence try to avoid the possibility of heart disease for the patient. So when a patient is predicted as positive for heart disease, then the medical data for the patient can be closely analysed by the doctors. An example would be - suppose the patient has diabetes which may be the cause for heart disease in future and then the patient can be given treatment to have diabetes in control which in turn may prevent the heart disease.

**7.REFERENCES**

[1] Monika Gandhi, Shailendra Narayanan Singh Predictions in heart disease using techniques of data mining (2015)

[2] J Thomas, R Theresa Princy Human heart disease prediction system using data mining techniques (2016)

[3] Sana Bharti, Shailendra Narayan Singh, Amity university, Noida, India Analytical study of heart disease prediction comparing with different algorithms (May 2015)

[4] Purushottam, Kanak Saxena, Richa Sharma Efficient heart disease prediction system using Decision tree (2015)

[5] Sellappan Palaniyappan, Rafiah Awang Intelligent heart disease prediction using data mining techniques (August 2008)

[6] Himanshu Sharma,M A Rizvi Prediction of Heart Disease using Machine Learning Algorithms: A Survey (August 2017)

[7] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review (2017)

[8] V.Krishnaiah, G.Narsimha, N.Subhash Chandra Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review (February 2016)

[9] Ramandeep Kaur, 2Er. Prabhsharn Kaur A Review - Heart Disease Forecasting Pattern using Various Data Mining Techniques (June 2016)

[10] J.Vijayashree and N.Ch. SrimanNarayanaIyengar Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review (2016)

[11] Benjamin EJ et.al Heart Disease and Stroke Statistics 2018 At-a-Glance (2018)

[12] Abhay Kishore, Ajay Kumar, Karan Singh, Maninder Punia, Yogita Hambir Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra Professor, Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra Heart Attack Prediction Using Deep Learning (2018)

[13] M.Nikhil Kumar, K.V.S Koushik, K.Deepak Department of CSE, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India Prediction Heart Diseases using Data mining and machine learning algorithms and tools.(2018)

[14] Amandeep Kaur, Jyoti Arora Dept of CSC Desh Bhagat University, Punjab, India heart disease prediction using data mining techniques: a survey (2018)

[15] Stephen F Weng, Jenna Reps, Joe Kai, Jonathan M Garibaldi and Nadeem Qureshi Can machine learning improve cardio vascular risk prediction using routine clinical data?(2017)

[16] A.Sahaya Arthy, G. Murugeshwari A survey on heart disease prediction using data mining techniques (April 2018)

[17]https://www.kaggle.com/amanajmera1/framing ham-heart-study-dataset

[18] A.Sudha, P.Gayathri, N.Jaisankar Effective analysis and prediction model for stroke disease using classification methods (April 2012)