IR@AIKTC-KRRC

ir.aiktclibrary.org

IR@AIKTC-KRRC

A PROJECT REPORT

ON

"HEART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING"

Submitted to

UNIVERSITY OF MUMBAI

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
ELECTRONICS AND TELECOMMUNICATION ENGINEERING

BY

Mirkar Naif Shaukat Rahat	**17DET50**
Sain Naif Abdul Rashid Zubeda	**17DET59**
Mohammed Aquib Mohammed Sohel Asiya	**17DET52**

UNDER THE GUIDANCE OF
PROF. Geeta Desai

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 & 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206 **2019-2020**

AFFILIATED TO

UNIVERSITY OF MUMBAI

ir.aiktclibrary.org

IR@AIKTC-KRRC

A PROJECT II REPORT
ON

"HEART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING"

Submitted to
UNIVERSITY OF MUMBAI

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
ELECTRONICS AND TELECOMMUNICATION ENGINEERING

BY

Mirkar Naif Shaukat Rahat	17DET50
Sain Naif Abdul Rashid Zubeda	17DET59
Mohammed Aqib Mohammed Sohel Asiya	17DET52

UNDER THE GUIDANCE OF
PROF. GEETA DESAI

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 & 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206

2019-2020
AFFILIATED TO

UNIVERSITY OF MUMBAI

ir.aiktclibrary.org

IR@AIKTC-KRRC

Anjuman-i-Islam's Kalsekar Technical Campus

Department of Electronics and Telecommunication Engineering

SCHOOL OF ENGINEERING TECHNOLOGY

Plot No. 2 3, Sector - 16, Near Thana Naka,

Khandagaon, New Panvel - 410206

CERTIFICATE

This is certify that the project entitled

"Heart Disease Prediction System Using Machine Learning "

submitted by

Mirkar Naif Shaukat Rahat	**17DET50**
Sain Naif Abdul Rashid Zubeda	**17DET59**
Mohammed Aqib Mohammed Sohel Asiya	**17DET52**

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Electronics and Telecommunication Engineering) at ***Anjuman-I-Islam's Kalsekar Technical Campus, Navi Mumbai*** under the University of MUMBAI. This work is done during year 2019-2020, under our guidance.

Date: / /

(Prof. Geeta Desai)
Project guide

(Prof. Siraj Rashid Pathan)
Project Co-ordinator

(Prof. Afzal Shaikh)
HOD, EXTC Department

DR. ABDUL RAZAK HONNUTAGI
Director

External Examiner

ir.aiktclibrary.org

IR@AIKTC-KRRC

Acknowledgements

We would like to take the opportunity to express my sincere thanks to our guide **Geeta Desai**, Assistant Professor, Department of Electronics and Telecommunication Engineering, AIKTC, School of Engineering, Panvel for his invaluable support and guidance throughout our project research work. Without his kind guidance & support this was not possible.

We are grateful to him for his timely feedback which helped us track and schedule the process effectively. His time, ideas and encouragement that he gave is help us to complete our project efficiently.

We would like to express deepest appreciation towards **DR. Abdul Razak Honnutagi**, Director, AIKTC, Navi Mumbai, **Prof. Afzal Shaikh**, Head of Department of Electronics and Telecommunication Engineering and **Prof. Siraj Pathan**, Project Coordinator whose invaluable guidance supported us in completing this project.

At last we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped us directly or indirectly during this course of work.

Mirkar Naif Shaukat Rahat

Sain Naif Abdul Rashid Zubeda

Mohammed Aqib Mohammed Sohel Asiya

ir.aiktclibrary.org

IR@AIKTC-KRRC

Project I Approval for Bachelor of Engineering

This project entitled ***Ëheart Disease Prediction System Using Machine Learning¨*** by ***Mirkar Naif Shaukat Rahat, Sain Naif Abdul Rashid Zubeda and Mohammed Aqib Mohammed Sohel Asiya*** is approved for the degree of ***Bachelor of Engineering in Department of Electronics and Telecommunication Engineering.***

Examiners

1. .....
2. .....

Supervisors

1. .....
2. .....

Chairman

.....

ir.aiktclibrary.org

IR@AIKTC-KRRC

Declaration

We declare that this written submission represents my ideas in our own words and where others ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Mirkar Naif Shaukat Rahat
17DET50

Sain Naif Abdul Rashid Zubeda
17DET59

Mohammed Aqib Mohammed Sohel Asiya
17DET52

ir.aiktclibrary.org

IR@AIKTC-KRRC

ABSTRACT

Cardiovascular disease one of the lethal disease in the world.Around 17 million people lost their life because of Cardiovascular disease.Particularly myocardial infraction,also known as heart attack.Which lead to a necessity to stored the symptoms and bad health habits which has the end result of cardio vascular disease.In order to diagnosed the cardio vascular one has to go through various tests like auscultation,ECG,blood pressure,cholesterol and blood sugar.As the condition of the patient becomes more and more critical the significantly increase in tests with respect to time therefore prioritization of the tests became very important.Health habits lead to the cardiovascular disease need to be recognised and certified precaution must be taken.As data increasing day by day machine learning is an provocative and widely increasing field which lead to the output in a short time for the large amount of data.The objective of this project is to predict the heart disease using machine learning algorithm. From 100 percent of the data 70 percent will be used to trained and 30 percent will be tested.Classifiers like K-Nearest Neighbour and Support Vector Machine(SVM) are used to train data.

Keywords: Python, Machine learning, Support vector machine (SVM), K-Nearest Neighbors,Graphic User Interface(GUI).

ir.aiktclibrary.org

IR@AIKTC-KRRC

Contents

ir.aiktclibrary.org

IR@AIKTC-KRRC

ir.aiktclibrary.org

IR@AIKTC-KRRC

List of Figures

ir.aiktclibrary.org

IR@AIKTC-KRRC

List of Tables

ir.aiktclibrary.org

IR@AIKTC-KRRC

Chapter 1

Introduction

Provincially the rate of the heart disease is increasing day by day in the mankind.The ultimate cause for this is the poor and over busy lifestyle of the people lack of exercise and activeness has been a great cause for the cardiovascular disease.Over the years there is an drastic increase and which lead to the catastrophic like heart attack.In our country india there are high chances that 6 out of 10 people suffered from heart disease .Mortality rate of our country due to heart disease is 100,000 over the years and increasing significantly at an average of 5.6.Over an estimation of 18 million death occurs due to cardiovascular disease worldwide and this number is more in the middle income and low income countries.Stroke and heart attack are the reason for over 80 percent of salvation.In india as the clock ticking the number of cardiovascular cases are increasing which is over 30 million.Therefore for the prevention more 2.5 lakh bypass surgeries are performed. . A matter of growing concern is that the number of patients requiring coronary interventions has been rising at 20 to 30 percent for the past few years. People die because of heart disease because of not knowing the symptoms or laziness some time Our project objective is to predict the disease so that everybody knows their statues and little more cautious about there health.In this paper we are going to discuss how can predicting the heart disease using machine learning algorithm. for this we are using the various machine learning algorithm like SVM and KNN. Data set which will be use is in taken from kaggle. Kaggle consist of real data set which is used for research and development process. after the processing of data is done. finally the result of both algorithm is produce and according to the accuracy, sensitivity and efficiency it will predict whether the person has heart disease or not.At last we will compare the accuracy's and tell which algorithm is best for prediction.

ir.aiktclibrary.org

IR@AIKTC-KRRC

Chapter 2

Literature Survey

2.1 Prediction of Cardiovascular Disease Using Machine Learning Algorithm

2.1.1 Summary

- Author's: Dinesh Kumar G,Santhosh Kumar D,Arumugaraj K

  In this paper in order to predict whether a person having a heart disease they have suggested a prediction model which ultimately leads to awareness and even diagnosis. In todays world due to the advancement in the technology there is lots of data is generated lead to the effective and accurate prediction about disease.UCI machine learning repository is used to obtain the dataset , for the pre processing of the data technique like removal of the noise data,missing data removal,default values filling and at the every level classification of the attributes for decision making and prediction at various level.diagnosis of the algorithm is done by using "classification,accuracy,sensitivity and specificity analysis". SVM(Support Vector Machine), Gradient boosting,Random forest,Naive Bayes classifier and logistic regression.In this paper[1] authors have mentioned that they got best accuracy for Naive Bayes. They suggest the future scope that the hybrid solutions of the algorithms in order to have a best accuracy.

2.2 Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network

2.2.1 Summary

- Author: Bhuvaneswari Amma N.G In this paper author has proposed a medical system for the prediction of the cardiovascular disease .In this generic algorithm and neural network advantages had been combined.For the complex classification problem a feed forward multilayered neural network is applicable. Neural networks weights is been determined using generic algorithm because in less

ir.aiktclibrary.org

iteration it finds the weight of good set.In this dataset is taken from University of California .Repository of the machine learning is been used training and testing .In which 303 instance of data of cardiovascular disease with 14 attribute of class label.Prepossessing of the data is done 1st for suitable training purpose.For training generic based neural network is used .In the weight based the final weights are stored and used for predicting the risk of the disease.Final result propose by the author is with an ultimate accuracy of 94.17

2.3 Classification and Prediction of atherosclerosis disease using machine learning algorithm

2.3.1 Summary

- Author's:Oumaima Terrada,Bouchaib Cherradi, Abdelhadi Raihani,Omar Bouattane

  In this paper author's have used the patient clinical data for heart disease prediction.They explained different Machine learning techniques such as K-medoids and k-means clustering for classification, Artificial Neural Network (ANN) and K-Nearest Neighbor (KNN) for prediction the presence and the absence of Atherosclerosis disease.they used Cleveland Cardiovascular diseases database which has 13 attribute for each patient. this data is separated into two training and testing data which occurs many performance evaluation in sensitivities, specificity's, accuracy, and MCC of the testing set contains 83 patients. This gives an outcome of system reached accuracy around 96

2.4 Cardiovascular disease detection using a new ensemble classifier

2.4.1 Summary

- Author's:Hamidreza Ashrafi Esfahani,Morteza Ghazanfari

  In this paper,cardiovascular patient's data taken from the UCI which is used for discover pattern algorithms which includes Decision tree, Neural Networks, Rough Set, SVM, Naive Bayes, and analyze their accuracy and prediction.To increase the accuracy of these algorithms authors propose a hybrid algorithm. where three classifier fused together which in this case neural network,Rough Set and Naive Bayes have been combined. 303 patients datasets are evaluated for ensemble classifier.The result shows accuracy of this fusion algorithm improved upto 86.19

2.5 Reasons to used SVM and KNN

- From above research papers we have concluded that the Support vectors machine (SVM) is best the algorithm for the classification purpose which handles the dataset and separate them into various classes thus which lead to better accuracy as compared to others.
- KNN is very simple for implementation.Robust with regard to the search space;for instance,classes don't have to be linearly separable.
- SVM works relatively well when there is clear margin of separation between classes.It is more effective in high dimensional spaces.
- CLassifier can be updated online at very little cost as new instances with known classes are presented.

IR@AIKTC-KRRC

Chapter 3

Basic Concept

System is use to predict the heart disease so we are using machine learning algorithms like SVM and KNN to predict.We gathered the data to train the algorithm and studies the human heart system how it affect our body what symptoms it creates.

3.1 Classifier's

Classification may be defined as the process of predicting class or category from observed values or given data points. The categorized output can have the form such as "Black" or "White" or "spam" or "no spam". Mathematically, classification is the task of approximating a mapping function (f) from input variables (X) to output variables (Y). It is basically belongs to the supervised machine learning in which targets are also provided along with the input data set. In our project we will use three classifiers are following.

- Support vector machine (SVM).
- K-nearest neighbor(K-NN)

3.1.1 Support Vector Machine(SVM)

A SVM performs classification by finding the hyper plane that maximise the margin between two classes. The vectors that define the hyper plane are the support vectors. Steps for Calculation of Hyperplane.

- .Set up training data.
- Set up SVM parameter.
- Train the SVM.
- Region classified by the SVM.

ir.aiktclibrary.org

Usage of the SVM for data set classification has its own advantages and disadvantages. Medical data set can be nonlinear of high dimensional by observing properties. It is clear that SVM would be one of the favourite choices for classification. Some of the advantage to select the SVM for classification choice.

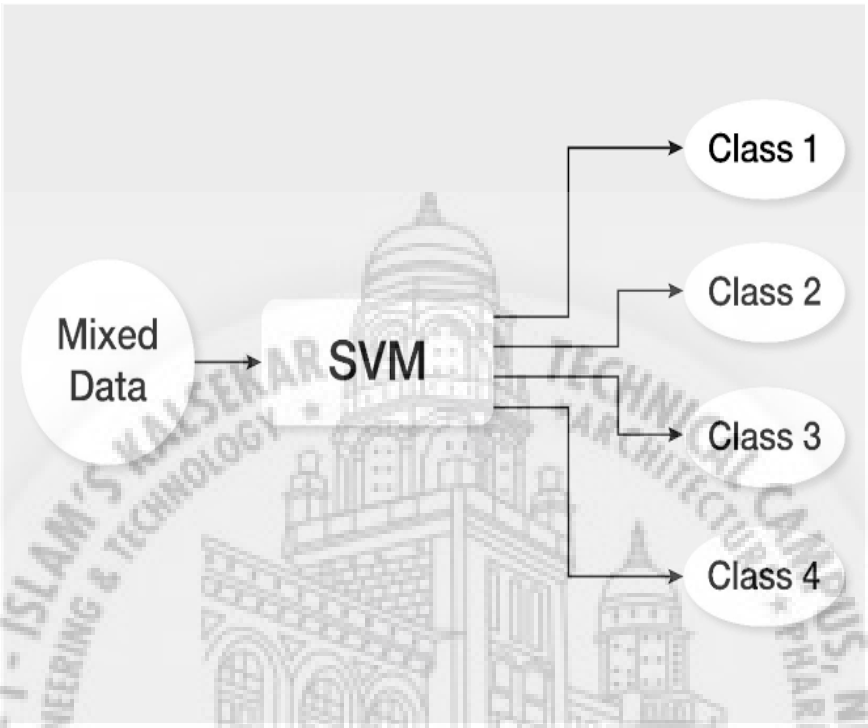

Figure 3.1: SVM Flowchart

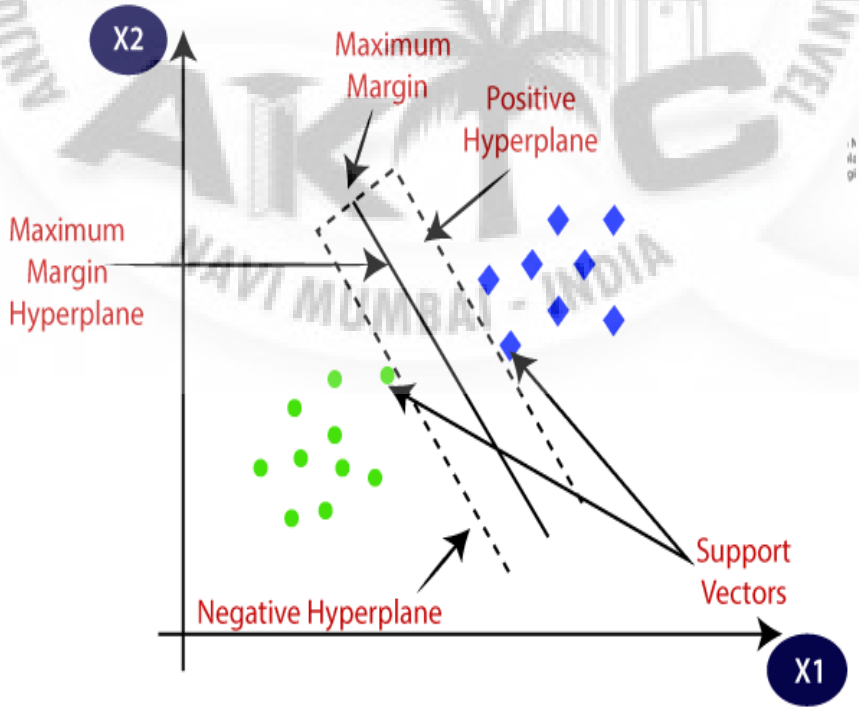

Figure 3.2: SVM Hyperplane

- Firstly regularization parameters which avoid problem of over fitting which one of the major challenges is in decision tree..
- Kernel tree is used to avoid the expert knowledge through the knowledge of kernel.
- SVM is an efficient method because it utilize convex optimisation problem (COP) which mean it has doesn't local minima.
- Error rated is tested which provide a greater support after miss classification of dataset.All the above features could be useful for medical diagnose dataset which is resulting in building more efficient predication system for the doctor. It doesn't mean it has all good side .coin has always two side on the other side it has best feature removal of over fitting problem is quite sensitive and it need optimizing parameter flaw in optimisation may result in error and may cause over fitting. .

3.1.2 K-Nearest Neighbor(KNN)

KNN is slow supervised learning algorithm, it take more time to get trained classification like other algorithm is divided into two step training from data and testing it on new instance . The K Nearest Neighbour working principle is based on assignment of weight to the each data point which is called as neighbour. in K Nearest Neighbour distance is calculate for training dataset for each of the K Nearest data points now classification is done on basis of majority of votes there are three types of distances need to be measured in KNN Euclidian, Manhattan, Minkowski distance in which Euclidian will be consider most one the following formula is used to calculate their distance.

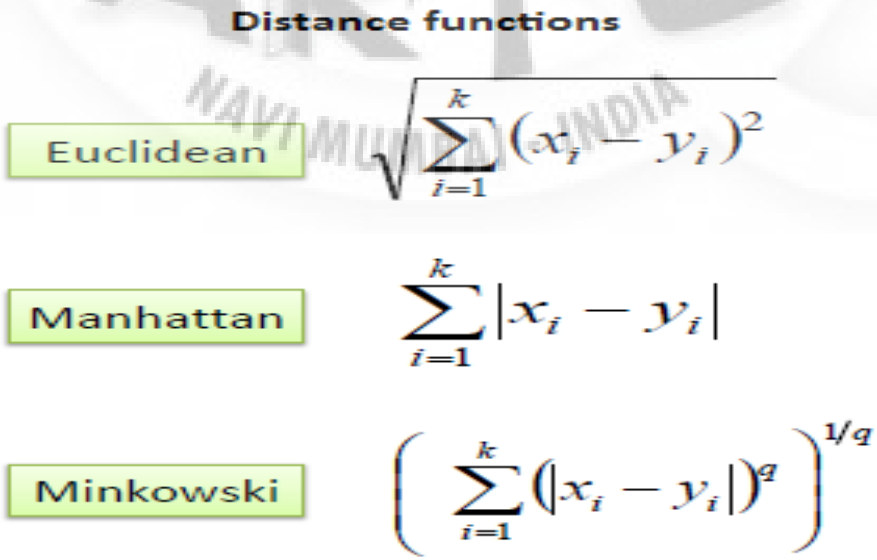

Figure 3.3: Distance Function

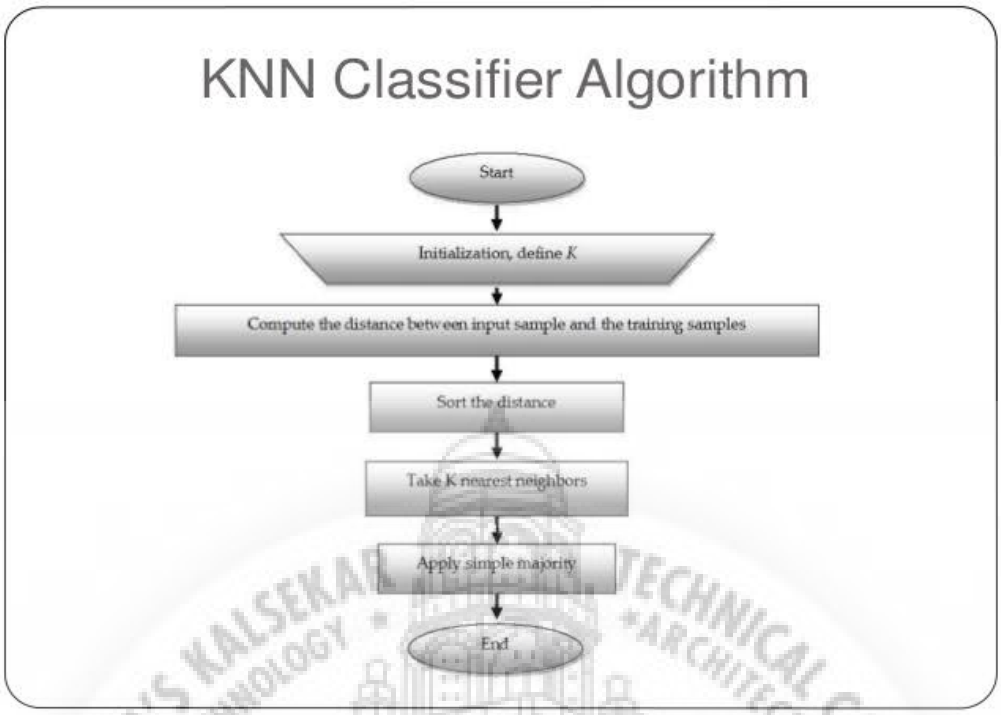

Figure 3.4: KNN Flow Chart

- Start
- Intialize the algorithm and define a suitable value of k
- Finding the distance from input data to the each sample value of training data
- Arranging the distance value in desending order
- Taking the top k defined value as k- nearest neighbour
- Finding which class has a majority in the that top k nearest neighbour
- Defining the input signal to that majority class.
- End

ir.aiktclibrary.org

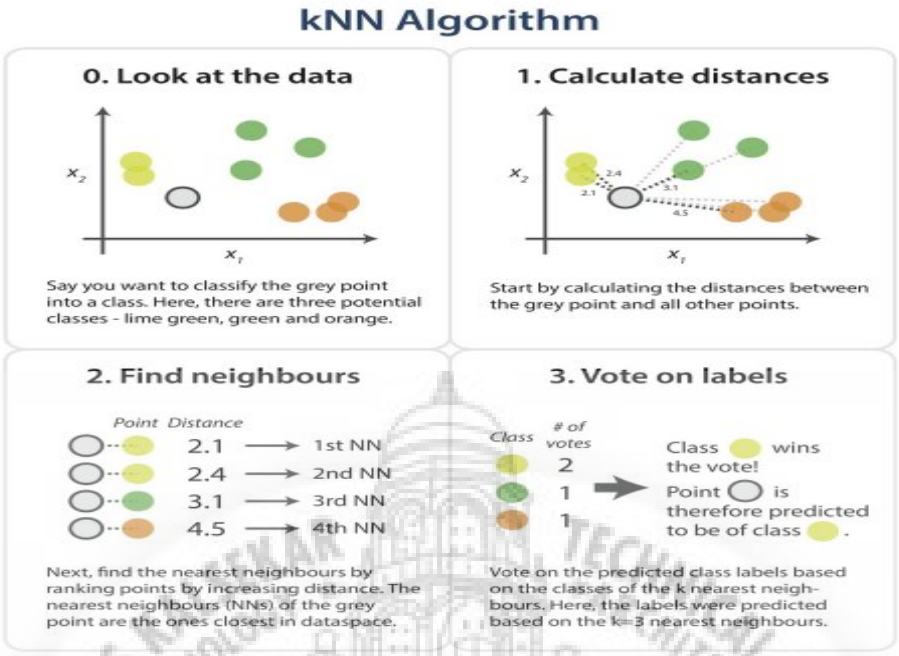

Figure 3.5: KNN Plot

Grouping of sample is based on super class in the KNN reduction of sample is the result of proper grouping which is used for further training. Selection of k value plays a pivotal role, if the k value is large then it precise and less noisy.The algorithm for KNN is defifined in the steps given below

- D represents the samples used in the training and k denotes the number of nearest neighbour.
- Create super class for each sample class
- Compute Euclidian distance for every training sample
- Based on majority of class in neighbour, classify the sample

3.2 Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics.Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

Reason to use python

- Pandas-Pandas is a library which is used to create a new type of data structure which is called as data frames.We can visualize it as a excel sheet with multiple

rows and ,multiple column.

- NumPy-NumPy is a library in which consist of multi-dimensional array and matrix data structure.It is used to perform a number of mathematical operations on arrays.Pandas is heavily rely on NumPy.
- Simple Syntax.
- Easily Understandable.
- Extensive Support Libraries.
- Integration Feature.

3.2.1 Jupyter

Jupyter is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks

Reason to use Jupyter

- The Jupyter can work for multifarious perceptual and language understanding tasks and to conduct complicated research on Machine Learning and Deep Neural Networks.
- It performs numerical computations through data flow graphs.
- Jupyter is highly parallel and designed to use various backends software .
- Scalability.

Chapter 4

Proposed Algorithm

4.1 System Flow Chart

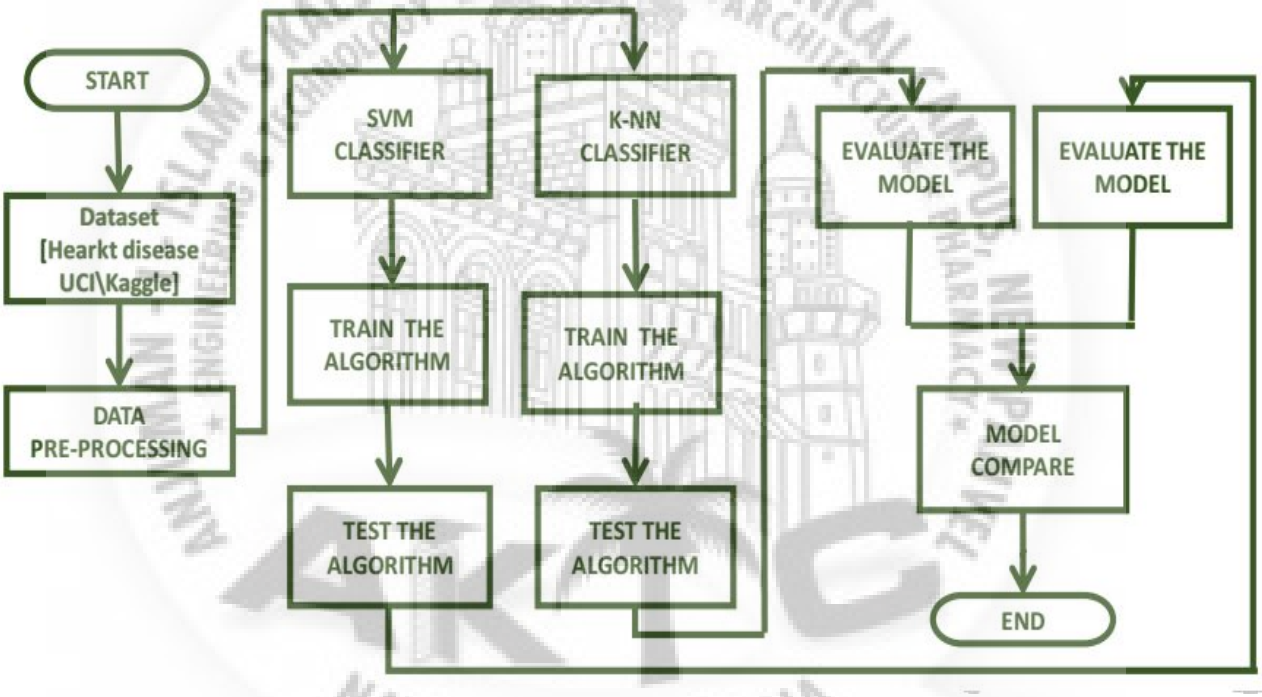

Figure 4.1: Flow Chart

- START.
- Dataset: The dataset contains the details of the patients.
- Data pre-processing :Converting the raw data into the suitable form that can be easily feed to the algorithm for prediction.
- Divide the data between training and testing in our system we have divided in 70 percent for training and 3 percent for testing.
- Used 2 different algorithm for heart disease prediction system. Support vector machine (SVM) K-Nearest Neighbour (KNN).

ir.aiktclibrary.org

- Train the algorithm:Train the both algorithm for the heart disease prediction system.
- Test the algorithm:Testing the result of both the algorithms.
- Evaluate the model: Evaluate both the algorithm and calculate the accuracy for both algorithms.
- Model comparing: Comparing accuracy's of both the algorithms and concluding which algorithm are accurate and gives best result.
- END.

4.2 System Architecture Design

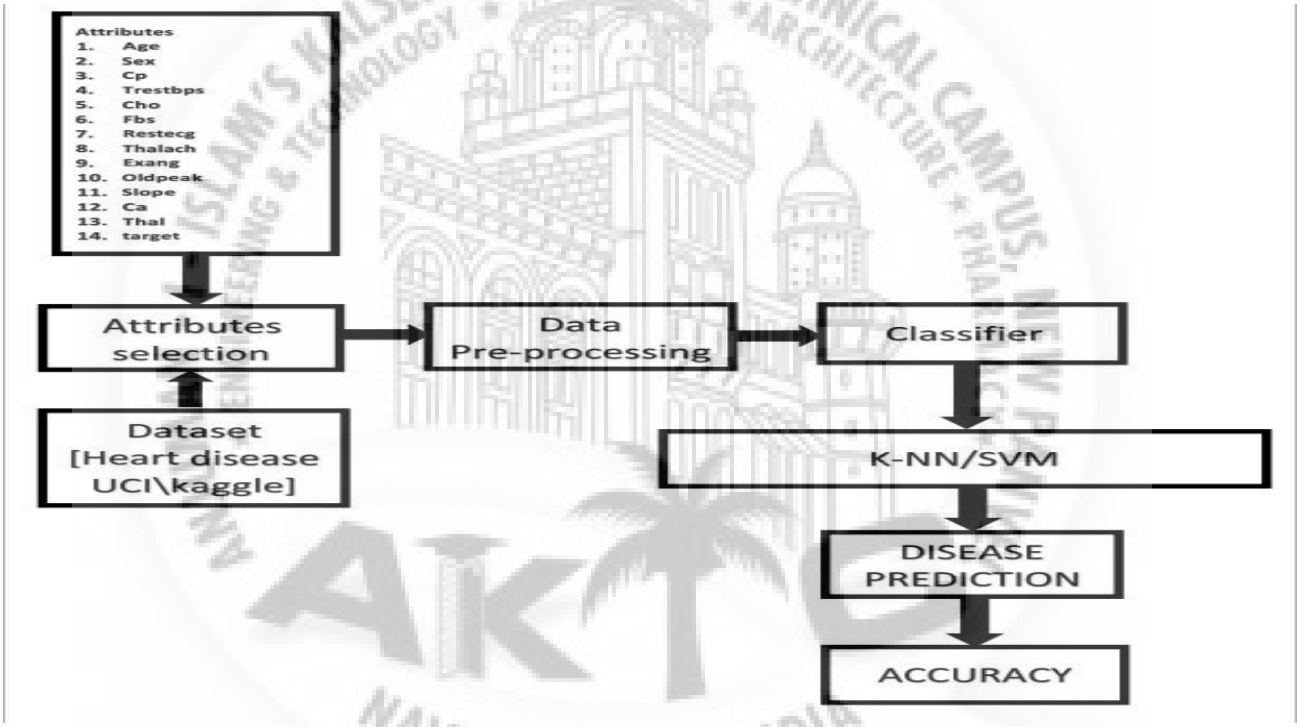

Figure 4.2: System Architecture

- Attributes will be taken as an input which will be pre-processed.Data pre-processing uses techniques like the removal of noisy data,removal of missing data,filling default values if applicable and classification of attributes for prediction and decision making.
- After pre-processing it will move towards heart disease Detection Engine. Here, it will extract the features from then dataset according to that heart disease will be predicted and provide an awareness or diagnosis on that.
- If heart disease is predicted then a system alert will be generated.

4.2.1 Data set

The dataset used here for predicting heart disease take online from the kaggle. Kaggle is an online community of data scientists and machine learners, owned by Google LLC. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, . The dataset used here is real dataset. The dataset consists of 214 instance of data with the appropriate 14 clinical parameters. The clinical parameter of dataset is about tests which are taken related to the heart disease as like blood pressure level, chest pain type, electrocardiographic result and etc.

4.2.2 Attributes

Selection Attribute selection even more important when the number of attributes are very large. You need not use every feature at your disposal for creating an algorithm. You can assist your algorithm by feeding in only those attributes that are really important. Importance of attribute selection are following:

- It enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the right subset is chosen.

As we take the dataset online from the kaggle. This dataset contains 76 attributes, but all publish experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The ”goal” field refers to the presence of heart disease in the patient.

Attributes Information

- Age.
- Sex.
- Chest pain type (4 values).
- Resting blood pressure.
- Serum cholesterol in mg/dl.
- Fasting blood sugar ¿ 120 mg/dl.
- Resting electrocardiography results (values 0,1,2).
- Maximum heart rate achieved.
- Exercise induced angina.

- Oldpeak = ST depression induced by exercise relative to rest.
- The slope of the peak exercise ST segment.
- Number of major vessels (0-3) colored by flourosopy thal: 3 = normal; 6 = fixed defect; 7 = reversable defect.The names and social security numbers of the patients were recently removed from the database,replaced with dummy values. One file has been ”processed”, that one containing the Cleveland database.

Data processing

After selecting the raw data for ML training, the most important task is data pre-processing. In broad sense, data pre-processing will convert the selected data into a form we can work with or can feed to ML algorithms. We always need to pre-process our data so that it can be as per the expectation of machine learning algorithm We have the following data pre-processing techniques that can be applied on data set to produce data for ML algorithms.

Data Visualisation

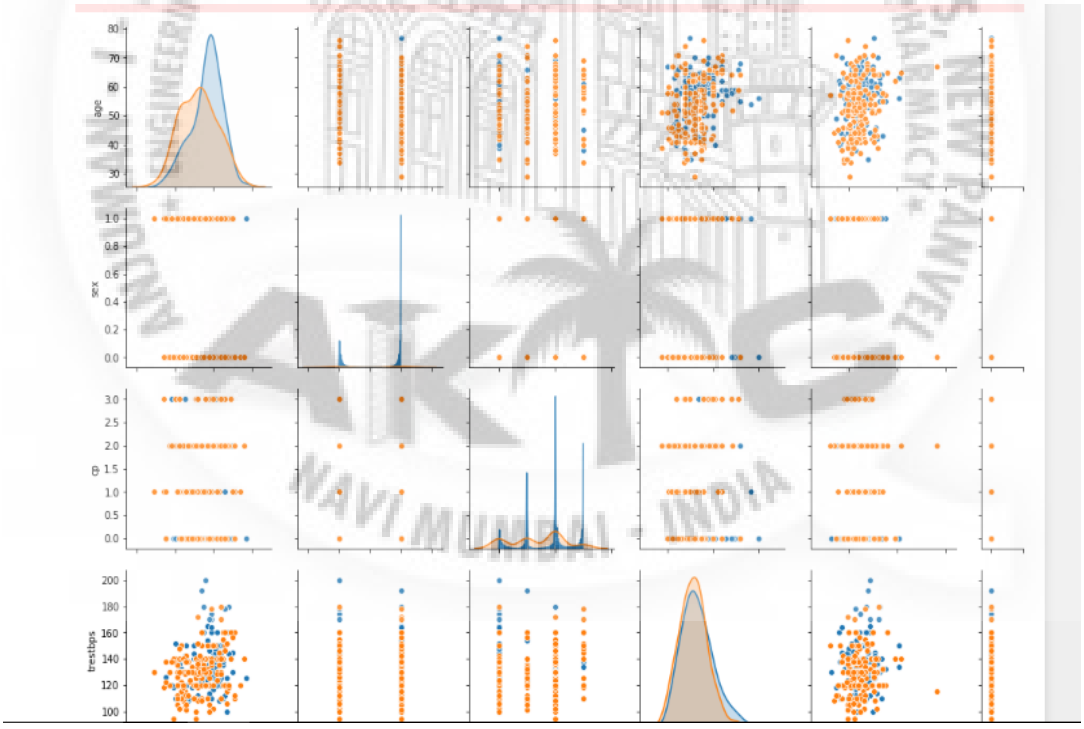

Figure 4.3: Data Visualization Based On Attributes

The data visualisation is the representation of data in graphical format.It represent the relationship among the different data so that it become easy for the user to observed.Above figure shows the scatter plot of different attribute’s of data set. Scatter plot is used to represent typical two variable in an graphical format.

4.3 Implementation

4.3.1 Importing all required librarires

```
# IMPORTS ALL LIBRARIES
from tkinter import *
from tkinter.ttk import Combobox
import tkinter.messagebox as tmsg
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
# % matplotlibinline
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import warnings
from sklearn.neighbors import K Neighbors Classifier
from sklearn.svm import SVC
```

4.4 Dataset Loading and Reading

We download the data set from the kaggle and save into project directory.python can not read the data in the tables form so first convert the data into CVS(comma separated value) form.after this data are loaded.

```
dataset = pd.read_csv('dataset.csv')
dataset.info()
```

4.5 Split data for training and testing

Next step are splitting data for training and testing.We are divide our data into 70 percent for training and 30 present for testing.

```
y = dataset['target']
 X = dataset.drop(['target'], axis=1)
 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
     random_state=120)
```

4.6 Fitting SVM Algorithm

```
svc_classifier = SVC(kernel="linear", gamma="auto")
svc_classifier.fit(X_train, y_train)
svc_classifier.score(X_test, y_test)
```

4.7 Evaluation of SVM

```
svc_classifier.score(X_test, y_test)
cd = svc_classifier.predict(X_test)
ab = svc_classifier.score(X_test, y_test) * 100
Print(cd)
print(ab, "is the total accuraccy of svm algorithm")
```

4.8 Fitting KNN Algorithm

```
knn_scores = []
knn_classifier = KNeighborsClassifier(n_neighbors=13)
knn_classifier.fit(X_train, y_train)
```

4.9 Evaluation of KNN

```
knn_classifier.score(X_test, y_test)
dd = knn_classifier.predict(X_test)
PP = knn_classifier.score(X_test, y_test) * 100
print(dd)
Print(pp, "is the total accuraccy of KNN algorithm")
```

- Full code with GUI designing

```
# IMPORTS ALL LIBRARIES
from tkinter import *
from tkinter.ttk import Combobox
import tkinter.messagebox as tmsg
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
# % matplotlibinline
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import warnings
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC



def sum():
    aa = agevalue.get()
    bb= gender.get()
    cc = chespain.get()
    ddd = restingbloodp.get()
    ff = sugar.get()
```

```
    gg = ecg.get()
    hh = maxheartrate.get()
    ii = angina.get()
    jj = slope.get()
    kk = ca.get()
    ll = thal.get()
    mm=oldpeak.get()
    nn=cholvalue.get()

    COU=0

    if bb=="MALE":
        bb=1
    elif bb=="FEMALE":
        bb=0
    else:
        COU=1+COU
        if COU==1:
            tmsg.showinfo(title="ERROR",message="YOU DOESNT ENTER ANY THING IN GENDER OR YOU ENTER SOMETHING WRONG "
                                                "IN THE GENDER. \n PLEASE CHECK IT AND REENTER THE CORRECT VALUE ")
    if cc=="TYPICAL ANGINA":
        cc=0
    elif cc=="ATYPICAL ANGINA":
        cc=1
    elif cc=="NON-ANGINA PAIN":
        cc=2
    elif cc=="ASYMPTOMATIC":
        cc=3
    else:
        COU=1+COU
        if COU==1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN CHEST PAIN TYPE OR YOU ENTER SOMETHING WRONG"
                                                 " IN THE CHEST PAIN TYPE. \n PLEASE CHECK IT AND REENTER THE CORRECT VALUE ")
    if ff=="GREATER THAN 120mg/dl":
        ff=1
    elif ff=="LESS THAN 120mg/dl":
        ff=0
    else:
        COU = 1 + COU
        if COU == 1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN FASTING BLOOD SUGAR OR YOU ENTER SOMETHING WRONG"
             " IN THE FASTING BLOOD SUGAR. \n PLEASE CHECK IT AND REENTER THE CORRECT VALUE ")
    if gg=="0=NORMAL":
        gg=0
    elif gg=="1=ST-T WAVE ABNORMALY":
        gg=1
    elif gg == "2=DEFINITE LEFT VENTRICULAR HYPERTROPHY BY ESTES":
        gg=2
    else:
        COU = 1 + COU
        if COU == 1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN ELECTROCARDIOGRAPHIC RESULT OR YOU ENTER SOMETHING WR"
```

```
                                                    "ONG IN THE
                                                        ELECTROCARDIOGRAPHIC RESULT
                                                        . \n PLEASE CHECK IT AND
                                                        REENTER THE CORRECT VALUE "
                                                        )
    if ii=="YES":
        ii=1
    elif ii=="NO":
        ii=0
    else:
        COU = 1 + COU
        if COU == 1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN
                EXERCIECE INDUCE ANGINA OR YOU ENTER SOMETHING WRONG"
         " IN THE EXERCIECE INDUCE ANGINA. \n PLEASE CHECK IT AND REENTER THE
            CORRECT VALUE ")
    if jj=="UPSLOPING":
        jj=0
    elif jj=="FLAT":
        jj=1
    elif jj=="DOWNSLOPING":
        jj=2
    else:
        COU=1+COU
        if COU==1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN
                SLOPE OR YOU ENTER SOMETHING WRONG "
                                                    "IN THE SLOPE. \n PLEASE CHECK
                                                        IT AND REENTER THE CORRECT
                                                        VALUE ")

    if kk== "0":
        kk=0
    elif kk== "1":
        kk =1
    elif kk== "2":
        kk=2
    elif kk== "3":
        kk=3
    else:
        COU = 1 + COU
        if COU == 1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN
                CA OR YOU ENTER SOMETHING WRONG "
                                                    "IN THE CA. \n PLEASE CHECK IT
                                                        AND REENTER THE CORRECT
                                                        VALUE ")

    if ll=="NORMAL":
        ll =1
    elif ll=="FIXED DEFECT":
        ll =2
    elif ll=="REVERSABLE DEFECT":
        ll=3
    else:
        COU = 1 + COU
        if COU == 1:
            tmsg.showinfo(title="ERROR", message="YOU DOESNT ENTER ANY THING IN
                THAL OR YOU ENTER SOMETHING WRONG "
                                                    "IN THE THAL. \n PLEASE CHECK
                                                        IT AND REENTER THE CORRECT
```

```
                                                    VALUE ")

    dataset = pd.read_csv('dataset.csv')
    dataset.info()
    new = [[aa,bb,cc,ddd,nn,ff,gg,hh,ii,mm,jj,kk,ll]]
    y = dataset['target']
    X = dataset.drop(['target'], axis=1)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
        random_state=120)

# SVM ALGORITHM

svc_classifier = SVC(kernel="linear", gamma="auto")
    svc_classifier.fit(X_train, y_train)
    svc_classifier.score(X_test, y_test)
    cd = svc_classifier.predict(X_test)
    ab = svc_classifier.score(X_test, y_test) * 100
    print(ab, "is the total accuraccy of svm algorithm")
    pre = svc_classifier.predict(new)
    print(pre)
    if pre == 1:
        SVMOP="YOU WILL HAVE HEART DISEASE"
    elif pre == 0:
        SVMOP="YOU WILL NOT HAVE HEART DISEASE"

    # KNN ALGORITHM

    knn_scores = []
    knn_classifier = KNeighborsClassifier(n_neighbors=13)
    knn_classifier.fit(X_train, y_train)
    knn_classifier.score(X_test, y_test)
    dd = knn_classifier.predict(X_test)
    pr1 = knn_classifier.predict(new)
    PP = knn_classifier.score(X_test, y_test) * 100
    print(pr1)
    if pr1 == 1:
        my="YOU WILL HAVE HEART DISEASE"
    elif pr1 == 0:
        my = "YOU WILL NOT HAVE HEART DISEASE"

    print(SVMOP,my)


    aaaa = tmsg.showinfo(title="resulty", message=f"HELLO  {NAMEL.get()}
        WELECOME TO THE HEART DISEASE PREDICTION SYSYTEM.\n WORKING OF THIS
        SYSTEMARE BASED ON TWO ALGORITHMS 1)SVM 2)KNN."
   f"\nTHIS TWO SYSTEM HAS DIFFRENT ACCURACY.\nSVM ACCURACCY IS{ab} and ACCURACY
        OF KNN IS {PP}.\nRESULT ARE BASED ON THIS TWO ALGORITHMS ARE FOLLOWING\n
       "
   f"ACCORDING TO SVM ALGORITHM{SVMOP}\nACCORDING TO KNN ALGORITHM {my} ")

    if aaaa == "ok":
        reset()

#RESET FUNCTION
def reset():
    maxheartrate.set("")
    cholvalue.set("")
    restingbloodp.set("")
    agevalue.set("")
    oldpeak.set("")
```

```
        gender.set("")
        chespain.set("")
        sugar.set("")
        ecg.set("")
        angina.set("")
        slope.set("")
        ca.set("")
        thal.set("")
        maxheartrate.set(0)
        cholvalue.set(0)
        restingbloodp.set(0)
        agevalue.set(0)
        oldpeak.set(0)

#GUI DESIGNING

from PIL import Image, ImageTk
root=Tk()
root.geometry("1350x900+0+0")
root.maxsize(1350,900)
root.minsize(300,300)
root.configure(background='RED')
meme=Frame(root,width=1350,height=50,bg="red")
Label(meme,text="HEART DISEASE PREDICTION SYSTEM",font=('arial',44,'bold'),bg="
    white",fg="black").grid(padx=100)
meme.grid(pady=(10,0),padx=15)

teme=Frame(root,width=1550,height=30,bg="red")
Label(teme,text="ENTER NAME",font=('arial',18,'bold'),bg="white",relief=SUNKEN,
    bd=5).grid(padx=(30,0))
NAMEL=StringVar()
Entry(teme,textvariable=NAMEL,bd=10,font=(50)).grid(row=0,column=1,padx=(30,0))

teme.grid(pady=10)


image=Image.open("y.jpg")
photo=ImageTk.PhotoImage(image)
ab=Label(root,image=photo)
ab.place(x=10,y=160,relwidth=1,relheight=1)
ab.image=photo
agel = Label(ab, text="AGE.                                                    ",font=("arial"
    ,18,"bold"),bd=5,relief=SUNKEN,bg='white').grid(row=0,column=0,padx=(100,0),
    pady=(10,0))
genderl= Label(ab, text="GENDER.                                                ",font=("arial
    ",18,"bold"),bd=5,relief=SUNKEN,bg='white').grid(row=0,column=3,padx=(40,0),
    pady=(10,0))
cpl = Label(ab, text="CHEST PAIN.                                          ",font=("arial",18,"bold
    "),bd=5,relief=SUNKEN,bg='white').grid(row=1,column=0,padx=(100,0),pady
    =(10,0))
rbpl = Label(ab, text="RASTING BLOOD PRESSURE.",font=("arial",18,"bold"),bd=5,
    relief=SUNKEN,bg='white').grid(row=1,column=3,padx=(40,0),pady=(10,0))
choll = Label(ab, text="CHOLESTROL.                                        ",bg='white',font=("arial
    ",18,"bold"),bd=5,relief=SUNKEN).grid(row=2,column=0,padx=(100,0),pady
    =(10,0))
fbsl = Label(ab, text="FASTING BLOOD SUGAR.       ",bg='white',font=("arial",18,"
    bold"),bd=5,relief=SUNKEN).grid(row=2,column=3,padx=(40,0),pady=(10,0))
ecgl = Label(ab, text="ELECTROCARDIOGRPHIC \n RESULT.",bg='white',font=("arial"
    ,18,"bold"),bd=5,relief=SUNKEN).grid(row=3,column=0,padx=(100,0),pady=(10,0)
    )
maxhrl = Label(ab, text="MAXIMUM HEART RATE.         ",bg='white',font=("arial"
```

```
    ,18,"bold"),bd=5,relief=SUNKEN).grid(row=3,column=3,padx=(40,0),pady=(10,0))
anginal = Label(ab, text="EXERCIECE INDUCED \n ANGINA.",bg='white',font=("
    arial",18,"bold"),bd=5,relief=SUNKEN).grid(row=4,column=0,padx=(100,0),pady
    =(10,0))
oldpeakl = Label(ab, text="OLD PEAK ",bg='white',
    font=("arial",18,"bold"),bd=5,relief=SUNKEN).grid(row=4,column=3,padx=(40,0)
    ,pady=(10,0))
slopel = Label(ab, text="SLOPE. ",bg='white',font
    =("arial",18,"bold"),bd=5,relief=SUNKEN).grid(row=5,column=0,padx=(100,0),
    pady=(10,0))
cal= Label(ab, text="CA ",bg='white',
    font=("arial",18,"bold"),bd=5,relief=SUNKEN).grid(row=5,column=3,padx=(40,0)
    ,pady=(10,0))
thall = Label(ab, text="THAL. ",bg='white',font
    =("arial",18,"bold"),bd=5,relief=SUNKEN).grid(row=6,column=0,padx=(100,0),
    pady=(10,0))

oldpeak=IntVar()
maxheartrate = IntVar()
cholvalue = IntVar()
restingbloodp = IntVar()
agevalue = StringVar()
age=Entry(ab,textvariable=agevalue,bd=2,font=('arial', 17, 'bold'),width=15).
    grid(row=0,column=1,pady=(10,0),padx=(10,0))
gendertype = ["MALE", "FEMALE"]
gender = Combobox(ab, values=gendertype, width=14, font=('arial', 17, 'bold'))
gender.grid(row=0,column=4,pady=(10,0),padx=(10,0))
cptype=["TYPICAL ANGINA","ATYPICAL ANGINA","NON-ANGINA PAIN","ASYMPTOMATIC"]
chespain=Combobox(ab,values=cptype, width=14,font=('arial',17,'bold'))
chespain.grid(row=1,column=1,pady=(10,0),padx=(10,0))
rbp=Entry(ab,textvariable=restingbloodp,bd=2,font=('arial', 17, 'bold'),width
    =15).grid(row=1,column=4,pady=(10,0),padx=(10,0))
chol=Entry(ab,textvariable=cholvalue,bd=2,font=('arial', 17, 'bold'),width=15).
    grid(row=2,column=1,pady=(10,0),padx=(10,0))
fastingsugartype = ["GREATER THAN 120mg/dl", "LESS THAN 120mg/dl"]
sugar = Combobox(ab, values=fastingsugartype, width=14, font=('arial', 17, 'bold
    '))
sugar.grid(row=2,column=4,pady=(10,0),padx=(10,0))
ecgtype = ["0=NORMAL", "1=ST-T WAVE ABNORMALY","2=DEFINITE LEFT VENTRICULAR
    HYPERTROPHY BY ESTES"]
ecg = Combobox(ab, values=ecgtype, width=14, font=('arial', 17, 'bold'))
ecg.grid(row=3,column=1,pady=(10,0),padx=(10,0))
maxheart=Entry(ab,textvariable=maxheartrate,bd=2,font=('arial', 17, 'bold'),
    width=15).grid(row=3,column=4,pady=(10,0),padx=(10,0))
anginatype = ["YES", "NO"]
angina = Combobox(ab, values=anginatype, width=14, font=('arial', 17, 'bold'))
angina.grid(row=4,column=1,pady=(10,0),padx=(10,0))
oldpeack=Entry(ab,textvariable=oldpeak,bd=2,font=('arial', 17, 'bold'),width=15)
    .grid(row=4,column=4,pady=(10,0),padx=(10,0))
slopetype = ["UPSLOPING", "FLAT","DOWNSLOPING"]
slope = Combobox(ab, values=slopetype, width=14, font=('arial', 17, 'bold'))
slope.grid(row=5,column=1,pady=(10,0),padx=(10,0))
catype = ["0", "1","2","3"]
ca = Combobox(ab, values=catype, width=14, font=('arial', 17, 'bold'))
ca.grid(row=5,column=4,pady=(10,0),padx=(10,0))
thaltype = ["NORMAL", "FIXED DEFECT","REVERSABLE DEFECT"]
thal = Combobox(ab, values=thaltype, width=14, font=('arial', 17, 'bold'))
thal.grid(row=6,column=1,pady=(10,0),padx=(10,0))
Button(ab, text="RESET",bd=10,font=('arial',15,'bold'),command=reset,bg="red").
    grid(row=6,column=3,pady=(10,0),padx=(10,0))

```


ir.aiktclibrary.org

```
Button(root, text="SUBMIT",command=sum, bd=20,font=('arial',20,'bold'),bg="red",
    fg="white").grid(row=30,column=0,pady=450)
maxheartrate.set(0)
cholvalue.set(0)
restingbloodp.set(0)
agevalue.set(0)
oldpeak.set(0)
root.mainloop()
```

IR@AIKTC-KRRC

Chapter 5

Result and Discussion

Everything which is devloped should get tested, because if developed system has some error that may cost user's lives.System testing is process of checking if fully developed system are working as per user requirement or not.As we already divided our data to 70percent for training and 30 percent testing process,now the 30 percent testing data are used here for the testing purpose.using that 30 percent data we checked our system accuracy for both SVM and KNN algorithm.The maximum accuracy generated by the algorithms are shows maximum chances of predicting a cardiovascular disease correctly.

5.1 Accuracy Measurement

The module predicts the accuracy by using machine learning algorithms. This module takes the maximum accuracy generated by the algorithms which predict the maximum chances of getting a cardiovascular disease. In this, each algorithm provides different accuracy rate for taken attributes which is the cause of the cardiovascular disease. You can calculate the accuracy of your model after generating the confusion matrix.

5.1.1 Confusion Matrix

By generating the confusion matrix we can find the accuracy of the system by applying the formula of accuracy. The formal of accuracy is given below:

ir.aiktclibrary.org

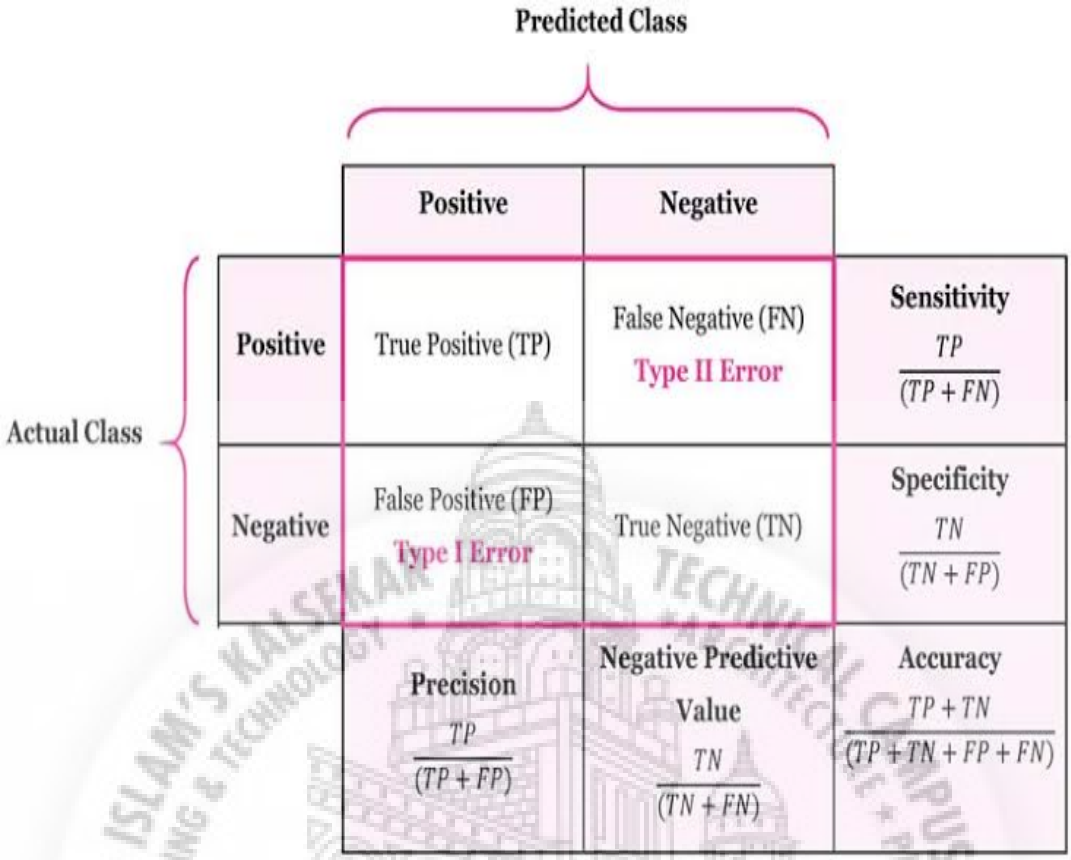

Figure 5.1: Confusion Matrix

5.2 Accuracy of SVM:

```
In [16]: from sklearn.metrics import confusion_matrix
         confusion_matrix(y_test,cd)

Out[16]: array([[33,  4],
                [ 6, 48]], dtype=int64)

In [17]: tt=confusion_matrix(y_test,cd)
         print("confusion matrix =",tt)
         dd=tt[0]
         tp=dd[0]
         fp=dd[1]
         dt=tt[1]
         fn=dt[0]
         tn=dt[1]

         print("TP(svm)=",tp)
         print("FP(svm)=",fp)
         print("FN(svm)=",fn)
         print("TN(svm)=",tn)

         confusion matrix = [[33  4]
          [ 6 48]]
         TP(svm)= 33
         FP(svm)= 4
         FN(svm)= 6
         TN(svm)= 48

In [18]: ACCURACY=(tp+tn)/(tp+tn+fp+fn)
         print("Accuracy of  SVM =",ACCURACY*100)

         Accuracy of  SVM = 89.01098901098901
```

Figure 5.2: SVM accuracy

As shown in abow figure we genrated the canfusion matrix of svm algorithm using confusion matrix librarry of sklearn.matrix,and after this we put all the values in the accuracy's formula. Using accuracy"s formula we get the accuracy of the svm algorthm which is 89.03 percent that means our svm algorithm gives 89.03 percent correct prediction.

SVM

Table 5.1: SVM Confusion Matrix

N=91	Prediction No	Prediction Yes
Actual No	33	4
Actual Yes	6	48

The Accuracy of Support Vector Machine is 89.01 Percent

5.3 Accuracy of KNN Algorithm:

```
In [21]: from sklearn.metrics import confusion_matrix
         confusion_matrix(y_test,dd)

Out[21]: array([[25, 12],
                [12, 42]], dtype=int64)

In [22]: ttl=confusion_matrix(y_test,dd)
         print("confusion matrix =",ttl)
         ddl=ttl[0]
         tpk=ddl[0]
         fpk=ddl[1]
         dtl=ttl[1]
         fnk=dtl[0]
         tnk=dtl[1]

         print("TP(knn)=",tpk)
         print("FP(knn)=",fpk)
         print("FN(knn)=",fnk)
         print("TN(knn)=",tnk)

         confusion matrix = [[25 12]
          [12 42]]
         TP(knn)= 25
         FP(knn)= 12
         FN(knn)= 12
         TN(knn)= 42

In [23]: ACCURAC=(tpk+tnk)/(tpk+tnk+fpk+fnk)
         print("Accuracy of knn",ACCURAC*100)

         Accuracy of knn 73.62637362637363
```

Figure 5.3: KNN accuracy

Similarly here we generated KNN confusion matrix using confusion matrix library of sklearn.matrix.after this we put all the values in the accuracy's formula. Using accuracy"s formula we get the accuracy of the KNN algorithm which is is 73.62 percent that means our SVM algorithm gives 73.62 percent correct prediction.

KNN

Table 5.2: KNN Confusion Matrix

N=91	Prediction No	Prediction Yes
Actual No	25	12
Actual Yes	12	42

The Accuracy of K-Nearest Neighbors is 73.62 Percent

5.4 Comparison of SVM and KNN Algorithm:

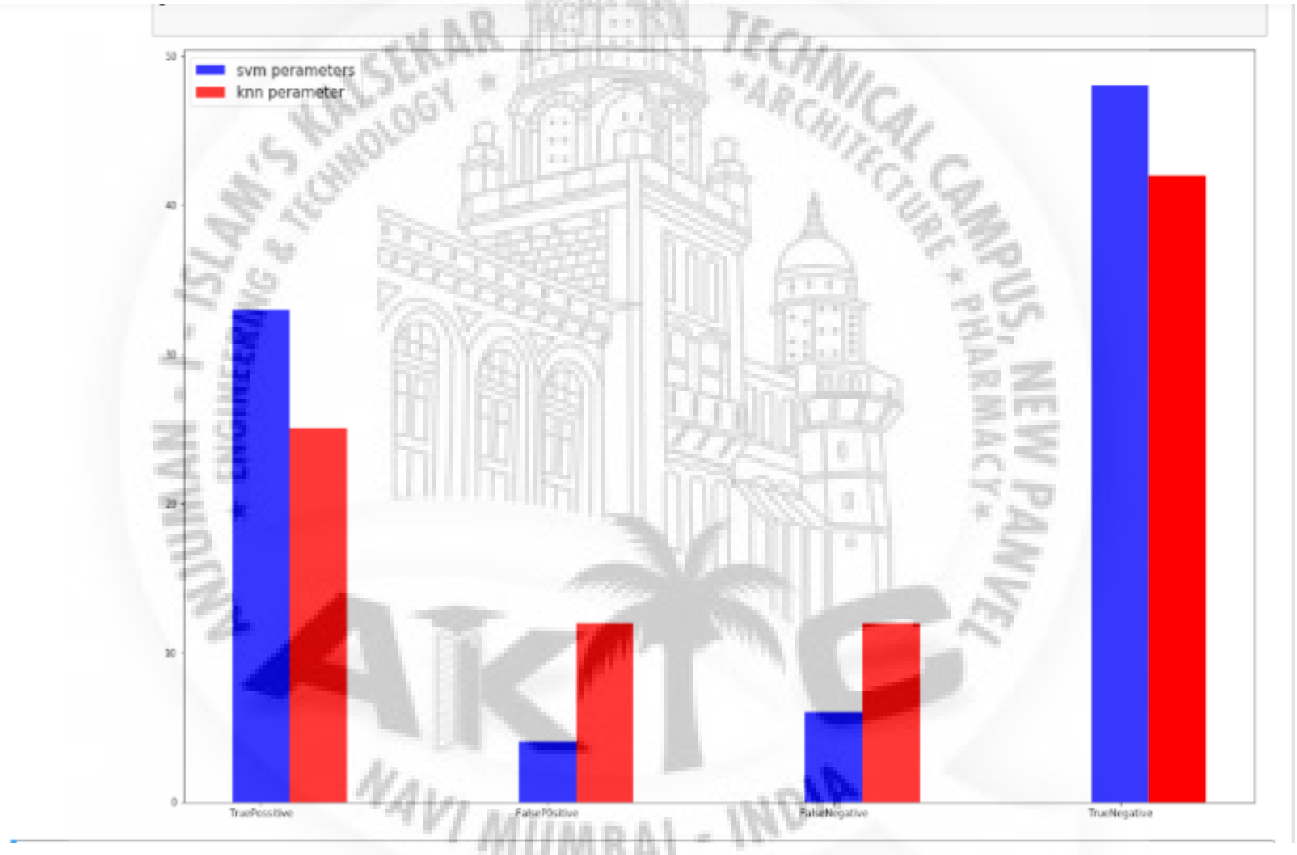

Figure 5.4: Bar plot of confusion matrix parameter

Another aim of our project is comparing the two algorithm SVM and KNN and finding which algorithm are better foe cardiovascular disease prediction.Above figure shows the bar plot of the different parameters of the confusion matrix.In this we compare the SVM KNN parameters. Blue bar represent the SVM and red bar represent KNN.

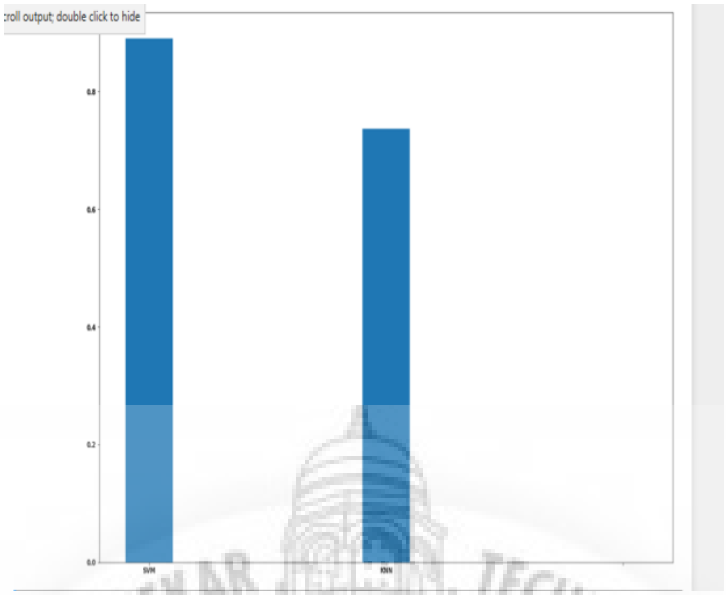

Figure 5.5: Bar plot of SVM and KNN accuracies

Above figure represent the bar plot of accurisies of SVM and KNN.In this value represent by bar of SVM is 0.88 and value represent by bar of KNN is 0.73.So, we can clearly see the accuracy of SVM is much higher than KNN.

5.5 GUI interface

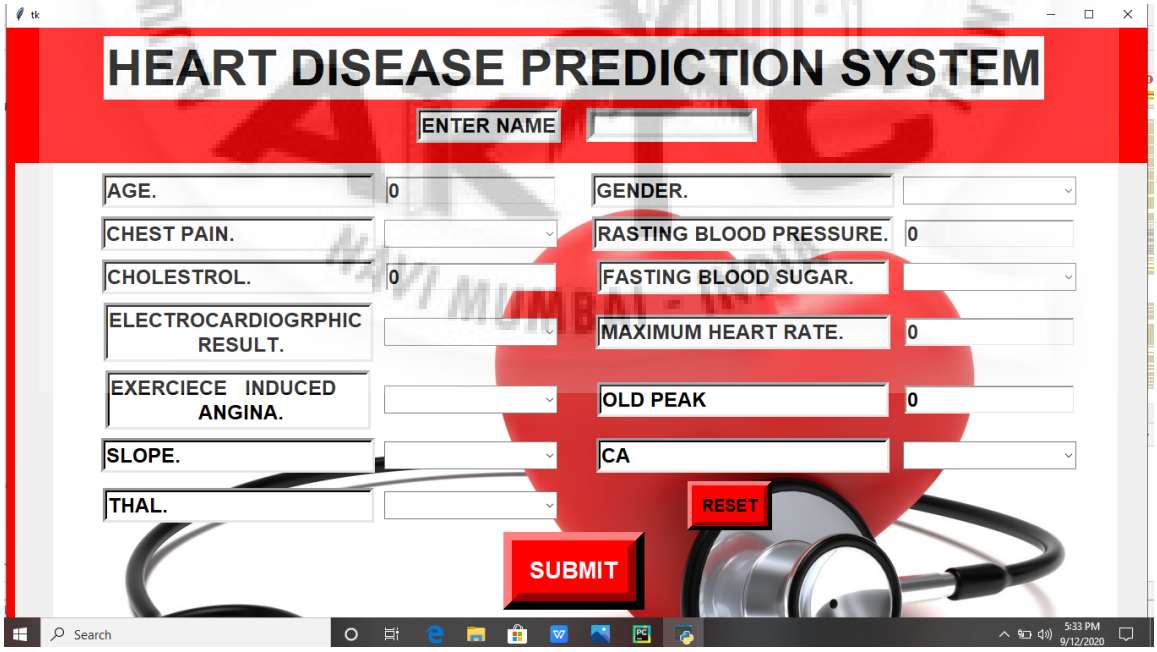

Figure 5.6: GUI interface

5.6 Attributes Entering

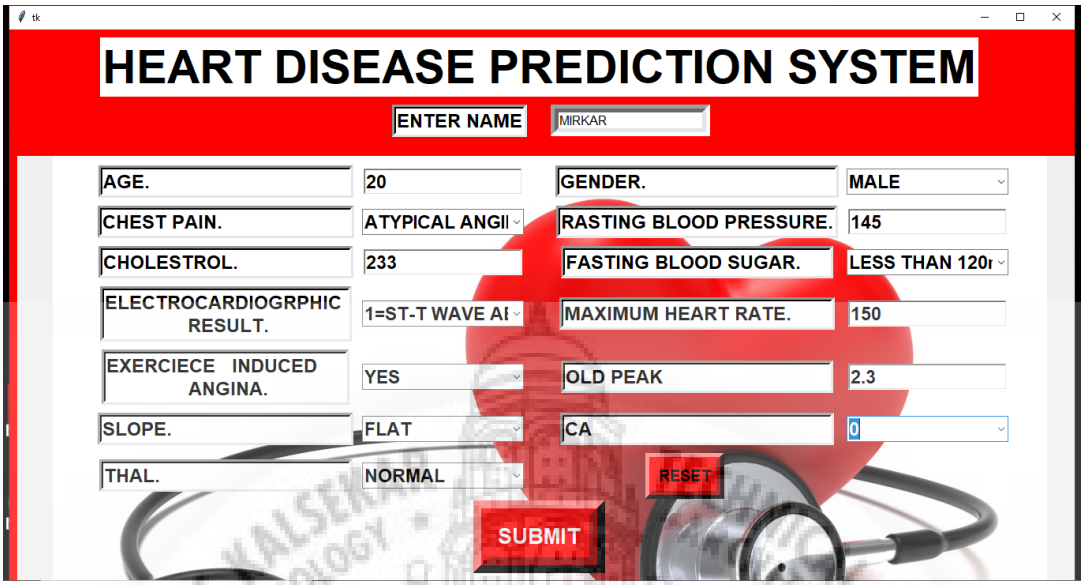

Figure 5.7: Attributes Entering

5.7 Predicted Result

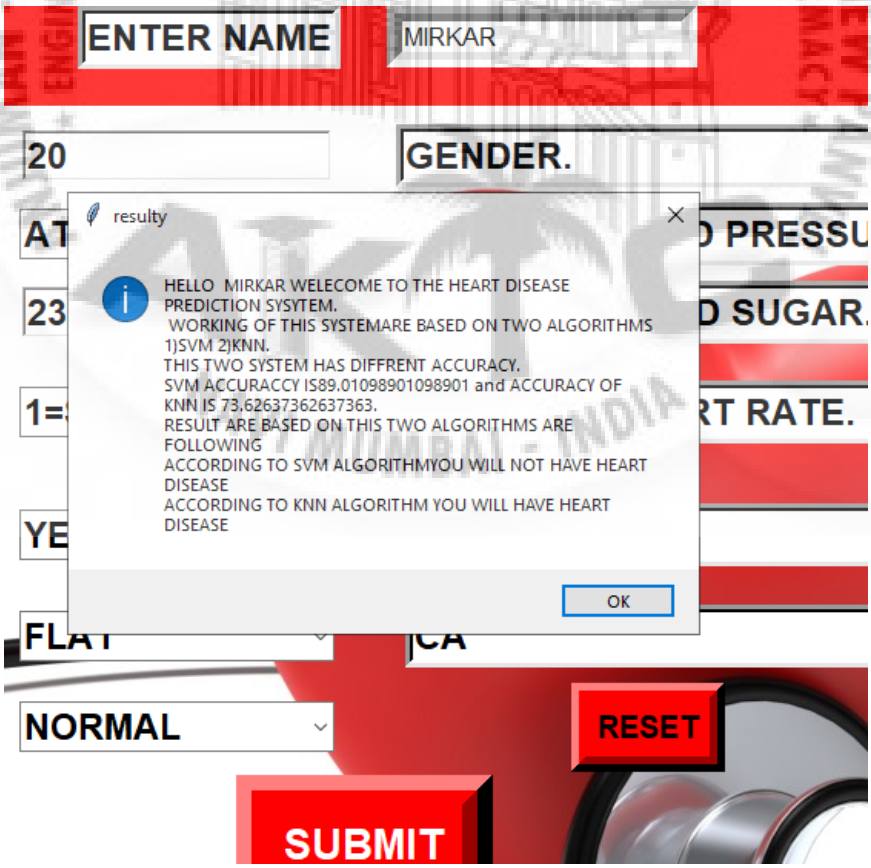

Figure 5.8: Predicted result

IR@AIKTC-KRRC

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

- Most of the data in today's era's are stored in computers and they are not properly used. Analysation of this data can lead to proper utilization of it .This area is been densely increase in the term of research and facilities.
- Basic aim of this project is to create a machine learning algorithm which can predicts the heart disease using software.
- It will be a user friendly and easy to used software that an non medical person can used easily and saving time for people

6.2 Future Plan

- Future work includes the hybridisation of different algorithm which can lead to a better performance and very good and wide selection of the features and attributes for the advancement in the medical field.
- Making more user friendly system so it will be more easy to operate by common person.

ir.aiktclibrary.org

References

[1] *Dinesh Kumar G, Arumugaraj K, Santhosh Kumar,Mareeswari V. "Prediction of cardio vascular disease using machine Learning Algorithm" IEEE International Conference on Current Trends 2018*

[2] *Bhuvaneswari Amma N.G. "Cardiovascular disease prediction system using Genetic Alogorithm and neural network". 2015: Department of CSE, Sudharsan Engineering College, Sathiyamangalam Pudukkottai, Tamilnadu, India.*

[3] *Oumaima Terrada, Bouchaib Cherradi, Abdelhadi Raihani,Omar Bouattane "Classification and prediction of artherocleroclersis disease using machine learning algorithm".978-1-7281-1482- 8/19 2019 IEEE.*

[4] *Hamidreza ashrafifi esfahani, Morteza ghazanfani. "Cardiovascular disease detection using a new ensemble classifier" 2017:IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI).*

[5] *Himanshu Sharma and M A Rizi"Prediction OF Heart Disease Using Machine Learning Algorithms:A Survey" International journal on recent and innovation trend in computing and communication.*