A

Minor Project

On

HEART DISEASE PREDICTION USING ML

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

ANANYA CHOWDARY BHEEMANENI (197R1A05J6)
SAI PRANAV GUNTHA (197R1A05L4)
SUMANTH CH (197R1A05K1)

Under the Guidance of

Najeema Afrin

(Asst. Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2019-2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"HEART DISEASE PREDICTION USING MACHINE LEARNING"** being submitted by **ANANYA CHOWDARY BHEEMANENI(197R1A05J6), SAI PRANAV GUNTHA (197R1A05L4) & CH. SUMANTH (197R1A05K1)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2022-23.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Najeema Afrin
(Asst. Professor)
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express my profound gratitude and deep regard to my guide **J.Narasimharao,** Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) **Dr. Punyaban Patel, Ms. Shilpa, Dr.M . Subha Mastan Rao & J. Narasimharao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

ANANYA CHOWDARY BHEEMANENI (197R1A05J6)
SAI PRANAV GUNTHA (197R1A05L4)
SUMANTH CH (197R1A05K1)

ABSTRACT

Heart related diseases or cardiovascular diseases (CVDs) are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life- threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. The health care industry produces a huge amount of data. This data is not always made use to the full extent and is often underutilized. Using this huge amount of data, a disease can be detected, predicted or even cured. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data.

Using machine learning techniques, the heart disease can be predicted. The medical data such as Blood pressure, hypertension, diabetes, cigarette smoked per day and so on is taken as input and then these features are modelled for prediction. The algorithms like K-nearest neighbour, Decision Trees, Random Forest are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the heart disease.

LIST OF FIGURES/TABLES

LIST OF SCREENSHOTS

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in rate of several heart related diseases. Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Thus, feasible and accurate prediction of heart related diseases is very important. Medical organisations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately

2. LITERATURE SURVEY

2. LITERATURE SURVEY

With growing development in the field of medical science alongside machine learning various experiments and researches has been carried out in these recent years releasing the relevant significant papers. The accuracy performance achieved by those algorithms are stillnot satisfactory. So that if the performance of accuracy is improved more to give batter decision to diagnosis disease. In a research conducted using Cleveland dataset for heart diseases which contains 303 instances and used 10-fold Cross Validation, considering 13 attributes, implementing 4 different algorithms, they concluded Gaussian Naïve Bayes and Random Forest gave the maximum accuracy of 91.2 percent. Using the similar dataset of Framingham, Massachusetts, the experiments were carried out using 4 models and were trained and tested with maximum accuracy K Neighbors Classifier: 87%, Support Vector Classifier: 83%, Decision Tree Classifier: 79% and Random Forest Classifier:84%

There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centers. K. Polaraju et al, 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, 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 kfold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease. S. Seema et al, 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, 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, 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 centers. 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. Chala Beyene et al, recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organization with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex is some of the attributes are included to identify if the person has heart disease or not. Analyses of dataset are computed using WEKAsoftware.

Sharmila et al, proposed to use non- linear classification algorithm for heart disease prediction. It is proposed to use bigdata tools such as Hadoop Distributed File System (HDFS), Mapreduce along with SVM for prediction of heart disease with optimized attribute set. This work made an investigation on the use of different data mining techniques for predicting heart diseases. It suggests to use HDFS for storing large data in different nodes andexecuting the prediction algorithm using SVM in more than one node simultaneously using SVM. SVM is used in parallel fashion which

yielded better computation time than sequential SVM. Jayami Patel et al suggested heart disease prediction using data mining and 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. Purushottam et al, proposed an efficient heart disease prediction system using data mining. This system helps medical practitioner to make effective decision making based on the certain parameter. By testing and training phase a certain parameter, it provides 86.3% accuracy in testing phase and 87.3% in training phase.

3.SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, “what must be done to solve the problem?” The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

3.1 PROBLEM DEFINITION

Cardiovascular disease(CVD) is a leading cause of deaths worldwide, accounting for approximately one third of all deaths. Prevention of CVD requires timely identification of people at risk. Over the last two decades, various prediction models have been developed. But these models had thier own set of issues and disadvantages. Hence, there was a absolute necessary to develop a reliable heart disease prediction model with high accuracy to help people to predict thier medical conditions.

3.2 EXISTING SYSTEM

Prediction using traditional methods and models involves various risk factors and it consists of various measures of algorithms such as datasets, programs and much more to add on. • High-risk and Low-risk patient classification is done on the basis of the tests that are done in group. But these models are only valuable in clinical situations and not in big industry sector. So, to include the disease predictions in various health related industries, we have used the concepts of machine learning and supervised learning methods to build the predictions system. Existing model gives less accuracy.

3.2.1 DISADVANTAGES OF EXISTING SYSTEM :

- There is a high probability of overfitting in Decision Tree.
- Generally, it gives low prediction accuracy for a dataset as compared to other machine learning algorithms.
- Information gain in a decision tree with categorical variables gives a biased response for attributes with greater number

3.3 PROPOSED SYSTEM

In the proposed system various techniques, algorithms and tools are used to create a system which predicts the disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references. Then the classification of that data is done with the help of Random Forest algorithm and the data goes in the recommendation model, there it shows the risk analysis that is involved in the system. The structured and unstructured form of data are combined for the overall risk analysis that is required for predicting the disease. Structured analysis is used to identify the chronic disease in a particular region and community. Unstructured analysis is used to select the features automatically with the help of algorithms and techniques.

3.3.1 ADVANTAGES OF PROPOSED SYSTEM

- Random algorithms avoids and prevents overfitting by using multiple trees.
- This gives accurate and precise results.

3.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis:

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

3.4.1 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on a project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require.

The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication that the system is economically possible for development.

3.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

3.4.3 BEHAVIORAL FEASIBILITY

This includes the following questions:

- Is there sufficient support for the users?
- Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developed and installed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible

3.5 HARDWARE & SOFTWARE REQUIREMENTS

3.5.1 HARDWARE REQUIREMENTS:

- RAM: 4 gb and above (8 or 16 recommended).
- Hard Disk: 15 gb and above.
- Processor: Core 2 duo and above.

3.5.2 SOFTWARE REQUIREMENTS:

- Windows 7 and above
- Python 3.7 and above
- Pandas,numpy,sklearn

4.ARCHITECTURE

4 . ARCHITECTURE

4.1 PROJECT ARCHITECTURE

This project architecture shows the procedure followed for classification, starting from input to final prediction.

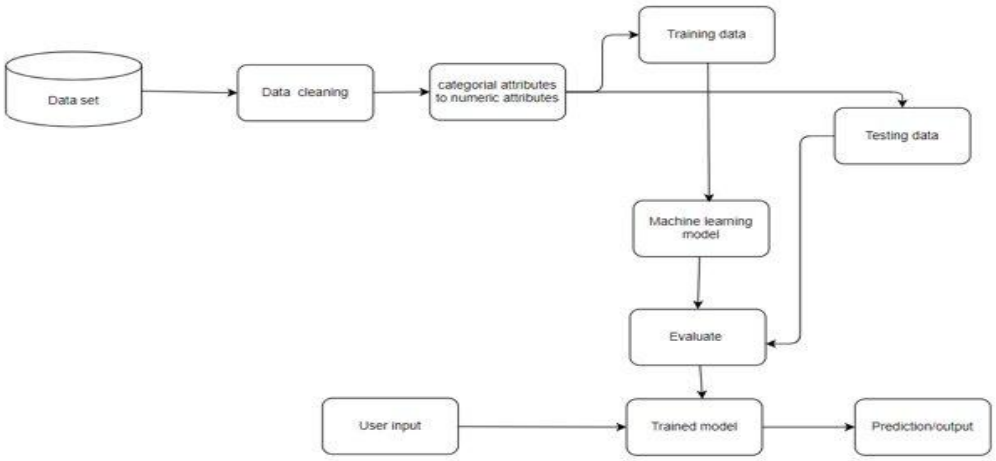


Figure 4.1: Project Architecture of Heart Disease Prediction using ML

4.2 DESCRIPTION

This project is totally based upon predicting heart disease of a patient using machine learning. This model is built to take the input data from the user and compare it with the training dataset to give the accurate prediction. This model is built using libraries pandas, NumPy, sklearn etc. Each library is used for a specific purpose for example pandas is used for data analysis. The library NumPy is used to work with arrays. Sklearn library is used for selecting efficient tools for machine learning and statistical model including classification, regression, clustering and dimensionality reduction via a consistence interface in python.

4.3 USE CASE DIAGRAM

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.

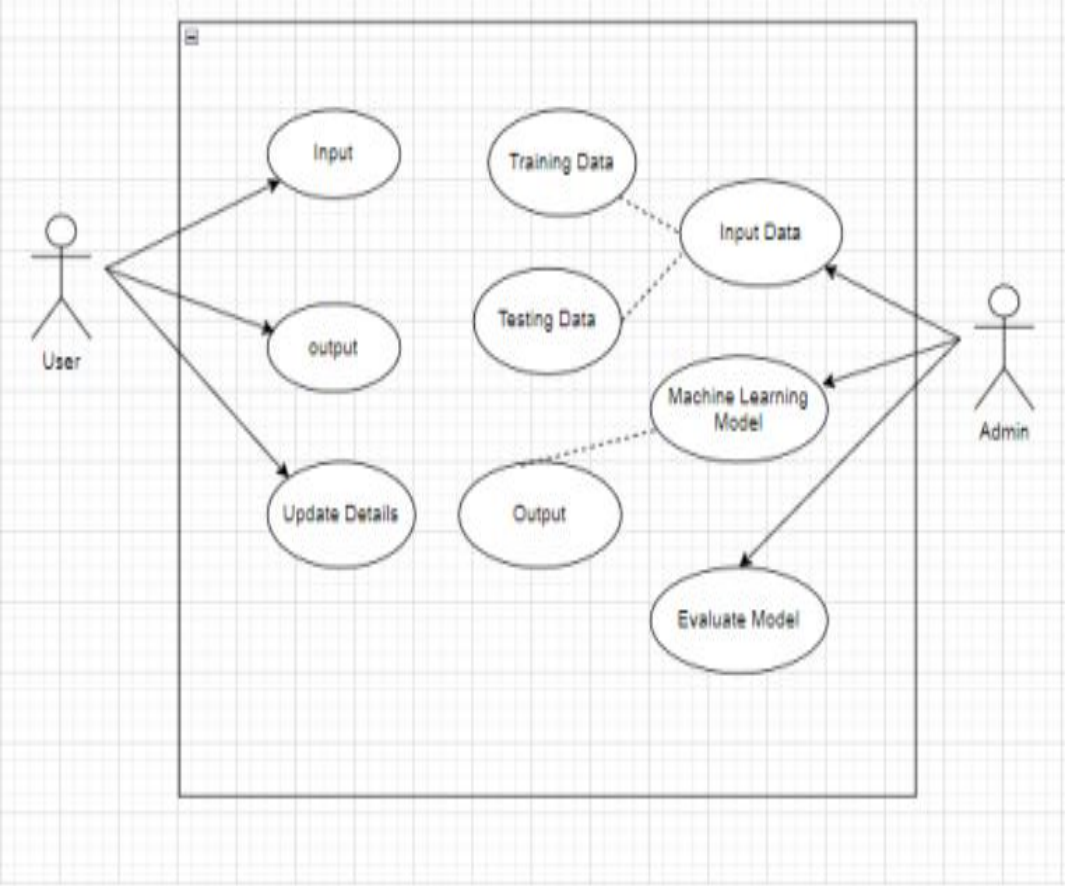


Figure 4.2: Use Case Diagram for Heart Disease Prediction using Machine Learning

4.4 CLASS DIAGRAM

Class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations(or methods), and the relationships among objects.

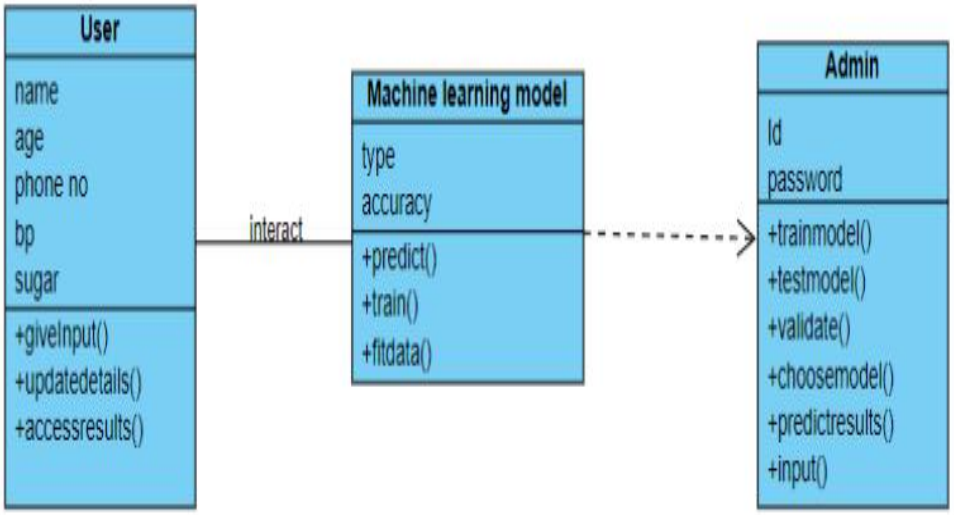


Figure 4.3: Class Diagram for Heart Disease Prediction using Machine Learning

4.5 SEQUENCE DIAGRAM

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development.

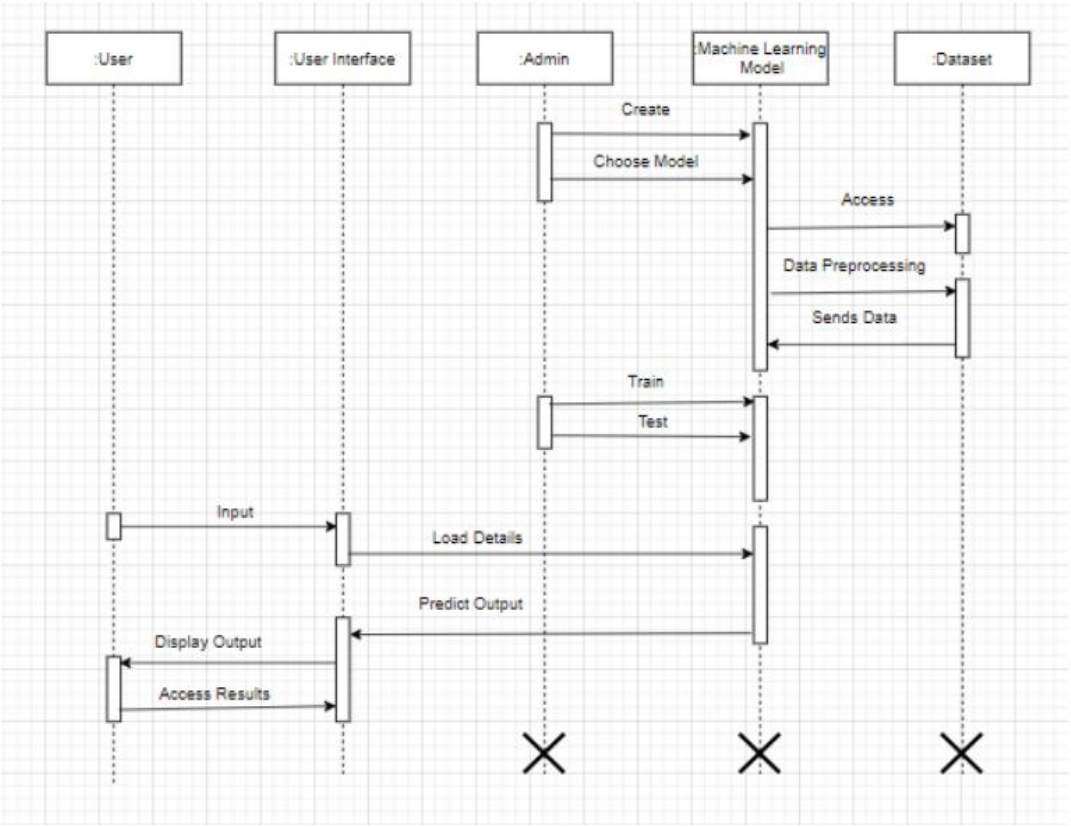


Figure 4.4: Sequence Diagram for Heart Disease Prediction using Machine Learning

4.6 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of work flows of stepwise activities and actions with support for choice, iteration and concurrency. They can also include elements showing the flow of data between activities through one or more data stores.

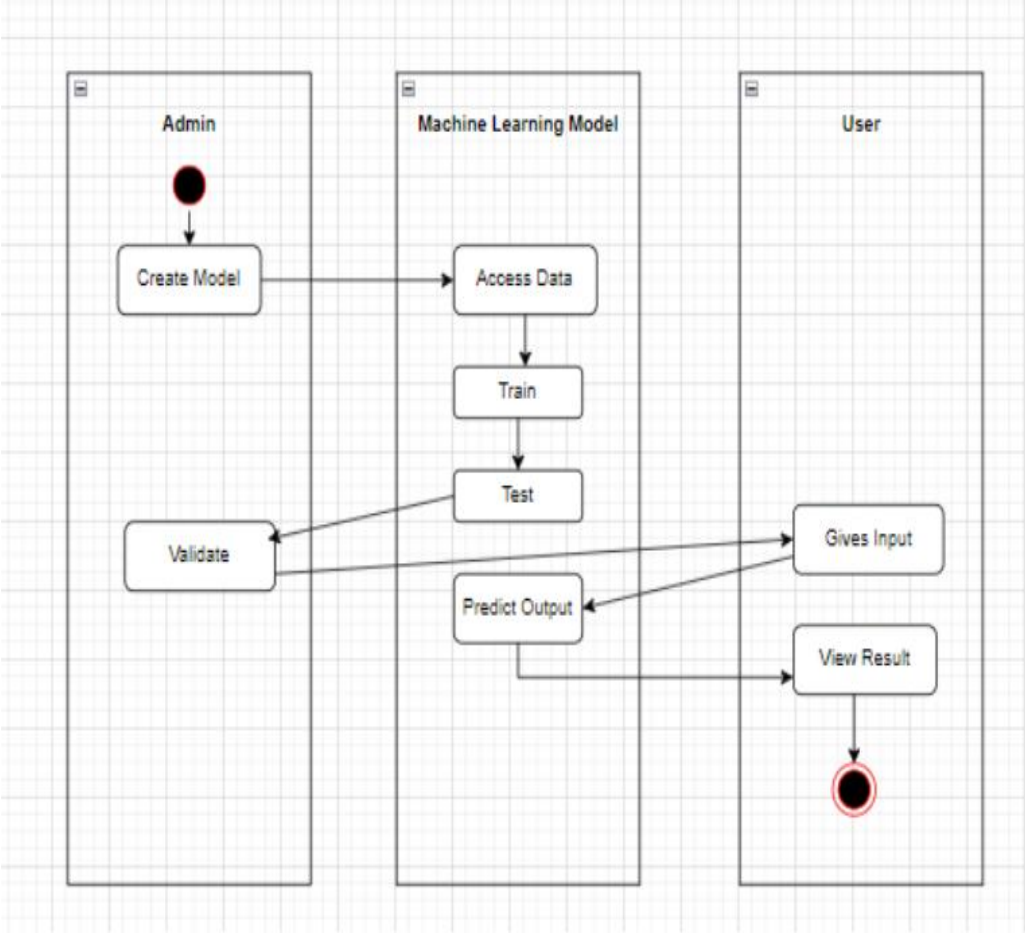


Figure 4.5: Activity Diagram for Heart Disease Prediction using Machine Learning

5. IMPEMENTATION

5.1 SAMPLE CODE

```
import pandas as pd
data = pd.read_csv('heart.csv')
data.isnull().sum()
data_dup = data.duplicated().any()
data_dup
data = data.drop_duplicates()
data_dup = data.duplicated().any()
data_dup
cate_val = []
cont_val = []
for column in data.columns:
    if data[column].nunique() <=10:
        cate_val.append(column)
    else:
        cont_val.append(column)
cate_val
cont_val
cate_val
data['cp'].unique()
cate_val.remove('sex')
cate_val.remove('target')
data = pd.get_dummies(data,columns = cate_val,drop_first=True)
data.head()
data.head()
from sklearn.preprocessing import StandardScaler
st = StandardScaler()
data[cont_val] = st.fit_transform(data[cont_val])
data.head()
X = data.drop('target',axis=1)
y = data['target']
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,
                                    random_state=42)
```

```
y_test
data.head()
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
log.fit(X_train,y_train)
y_pred1 = log.predict(X_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred1)

from sklearn import svm
svm = svm.SVC()
svm.fit(X_train,y_train)
y_pred2 = svm.predict(X_test)
accuracy_score(y_test,y_pred2)
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(X_train,y_train)
y_pred3=knn.predict(X_test)
accuracy_score(y_test,y_pred3)
score = []

for k in range(1,40):
    knn=KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train,y_train)
    y_pred=knn.predict(X_test)
    score.append(accuracy_score(y_test,y_pred))
    score
    import matplotlib.pyplot as plt
    plt.plot(score)
    plt.xlabel("K Value")
    plt.ylabel("Acc")
  plt.show()
  knn=KNeighborsClassifier(n_neighbors=2)
```

```
knn.fit(X_train,y_train)
y_pred=knn.predict(X_test)
accuracy_score(y_test,y_pred)
data = pd.read_csv('heart.csv')
data = data.drop_duplicates()
X = data.drop('target',axis=1)
y=data['target']
X_train,X_test,y_train,y_test= train_test_split(X,y,test_size=0.2,
                                    random_state=42)
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(X_train,y_train)
y_pred4= dt.predict(X_test)

accuracy_score(y_test,y_pred4)
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(X_train,y_train)
y_pred5= rf.predict(X_test)
accuracy_score(y_test,y_pred5)
from sklearn.ensemble import GradientBoostingClassifier
gbc = GradientBoostingClassifier()
gbc.fit(X_train,y_train)
y_pred6 = gbc.predict(X_test)
accuracy_score(y_test,y_pred6)
final_data = pd.DataFrame({'Models':['LR','SVM','KNN','DT','RF','GB'],
                    'ACC':[accuracy_score(y_test,y_pred1)*100,
                        accuracy_score(y_test,y_pred2)*100,
                        accuracy_score(y_test,y_pred3)*100,
                        accuracy_score(y_test,y_pred4)*100,
                        accuracy_score(y_test,y_pred5)*100,
                        accuracy_score(y_test,y_pred6)*100]})
```

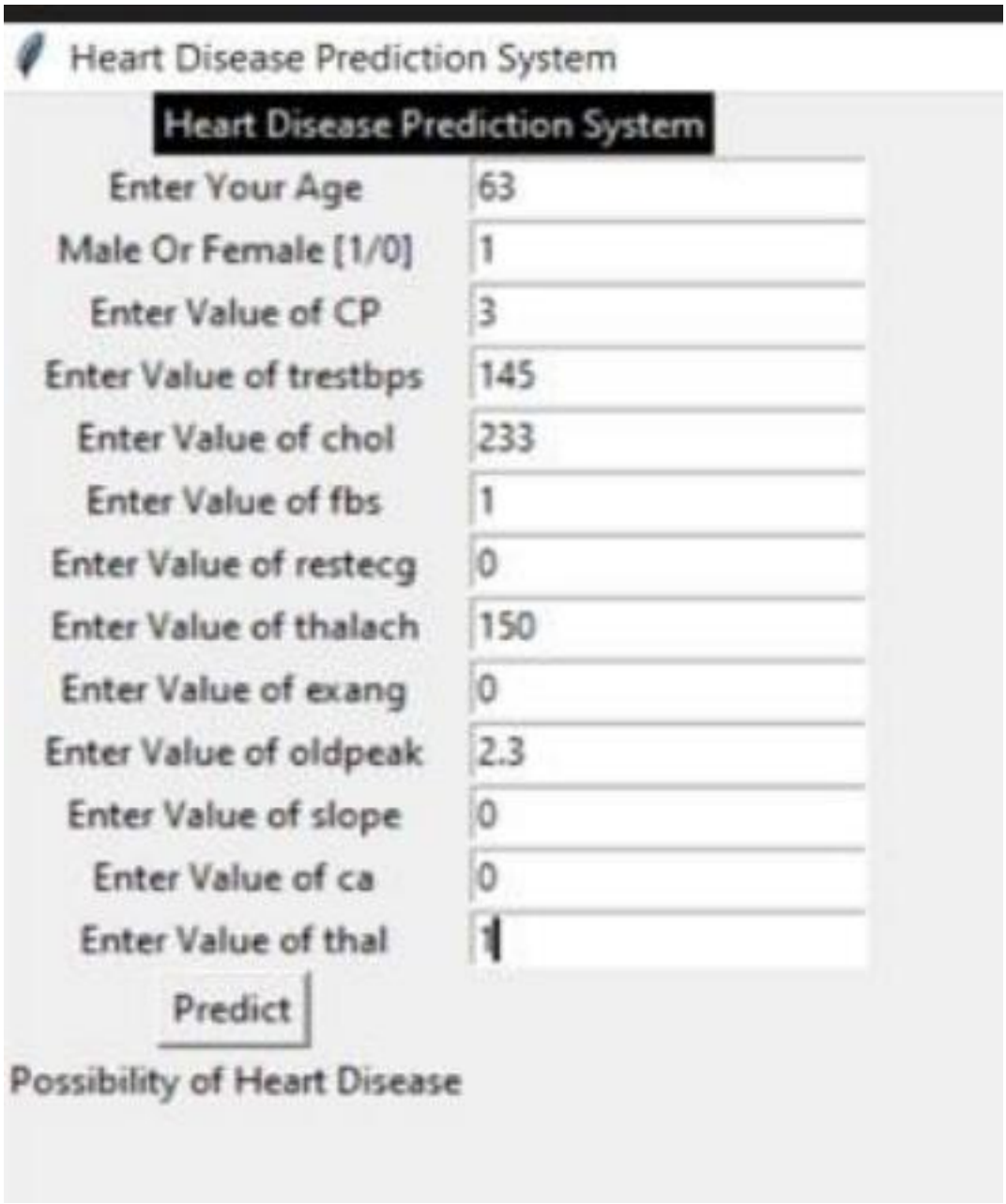
```
 final_data
import seaborn as sns
sns.barplot(final_data['Models'],final_data['ACC'])
X=data.drop('target',axis=1)
y=data['target']
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(X,y)
import pandas as pd
new_data = pd.DataFrame({
    'age':52,
    'sex':1,
    'cp':0,
    'trestbps':125,
    'chol':212,
    'fbs':0,
    'restecg':1,
    'thalach':168,
    'exang':0,
    'oldpeak':1.0,
     'slope':2,
    'ca':2,
    'thal':3,
},index=[0])
new_data
p = rf.predict(new_data)
if p[0]==0:
    print("No Disease")
else:
    print("Disease")
import joblib
joblib.dump(rf,'model_joblib_heart')
model = joblib.load('model_joblib_heart')
model.predict(new_data)
data.tail()
```

```
from tkinter import *
import joblib
def show_entry_fields():
   p1=int(e1.get())
   p2=int(e2.get())
   p3=int(e3.get())
   p4=int(e4.get())
   p5=int(e5.get())
   p6=int(e6.get())
   p7=int(e7.get())
   p8=int(e8.get())
   p9=int(e9.get())
   p10=float(e10.get())
   p11=int(e11.get())
   p12=int(e12.get())
   p13=int(e13.get())
   model = joblib.load('model_joblib_heart')
   result=model.predict([[p1,p2,p3,p4,p5,p6,p7,p8,p8,p10,p11,p12,p13]])

   if result == 0:
      Label(master, text="No Heart Disease").grid(row=31)
   else:
      Label(master, text="Possibility of Heart Disease").grid(row=31)


master = Tk()
master.title("Heart Disease Prediction System")


label = Label(master, text = "Heart Disease Prediction System"
                      , bg = "black", fg = "white"). \
                          grid(row=0,columnspan=2)
```

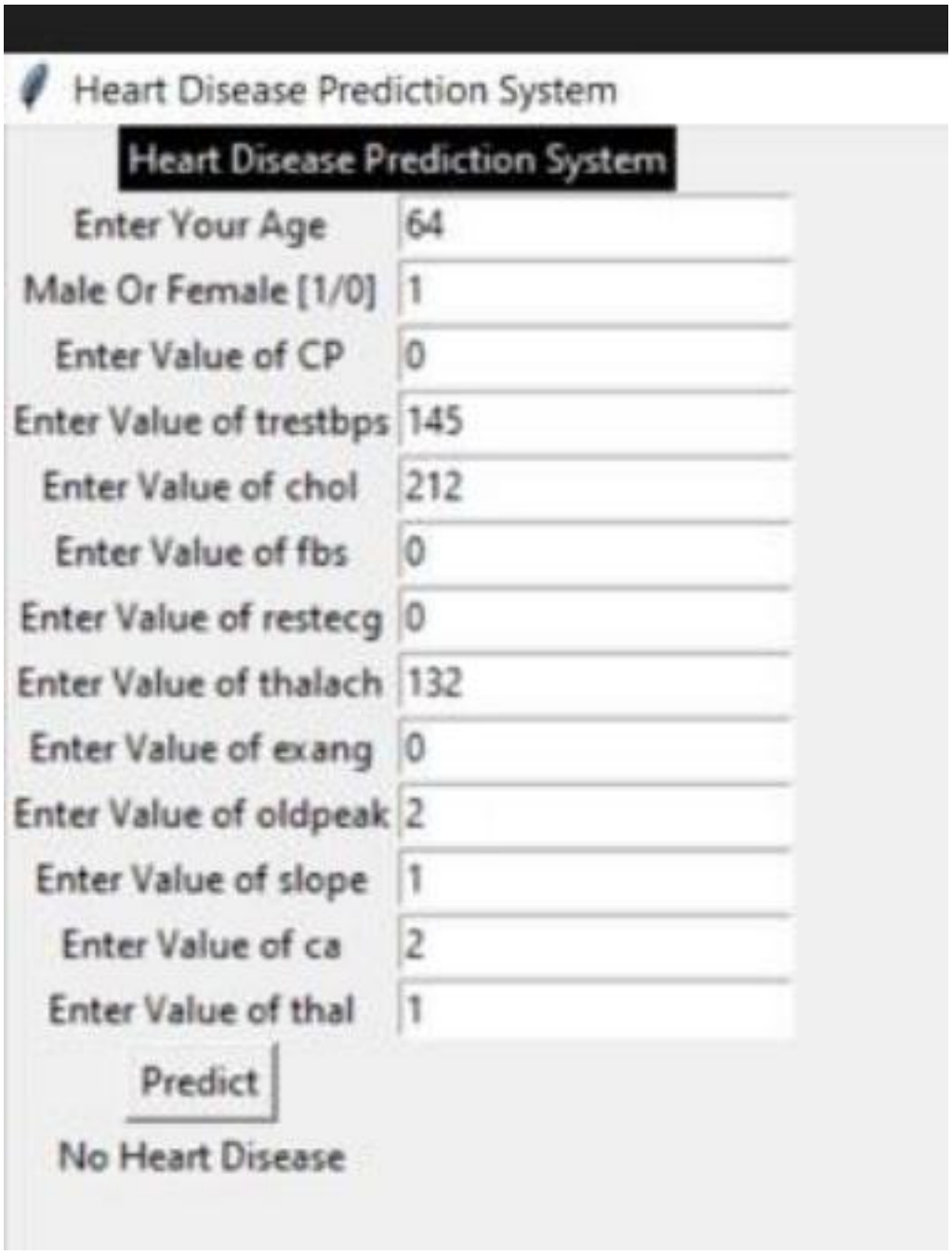
```
Label(master, text="Enter Your Age").grid(row=1)
Label(master, text="Male Or Female [1/0]").grid(row=2)
Label(master, text="Enter Value of CP").grid(row=3)
Label(master, text="Enter Value of trestbps").grid(row=4)
Label(master, text="Enter Value of chol").grid(row=5)
Label(master, text="Enter Value of fbs").grid(row=6)
Label(master, text="Enter Value of restecg").grid(row=7)
Label(master, text="Enter Value of thalach").grid(row=8)
Label(master, text="Enter Value of exang").grid(row=9)
Label(master, text="Enter Value of oldpeak").grid(row=10)
Label(master, text="Enter Value of slope").grid(row=11)
Label(master, text="Enter Value of ca").grid(row=12)
Label(master, text="Enter Value of thal").grid(row=13)


e1 = Entry(master)
e2 = Entry(master)
e3 = Entry(master)
e4 = Entry(master)
e5 = Entry(master)
e6 = Entry(master)
e7 = Entry(master)
e8 = Entry(master)
e9 = Entry(master)
e10 = Entry(master)
e11 = Entry(master)
e12 = Entry(master)
e13 = Entry(master)


e1.grid(row=1, column=1)
e2.grid(row=2, column=1)
e3.grid(row=3, column=1)
```

```
e4.grid(row=4, column=1)
e5.grid(row=5, column=1)
e6.grid(row=6, column=1)
e7.grid(row=7, column=1)
e8.grid(row=8, column=1)
e9.grid(row=9, column=1)
e10.grid(row=10, column=1)
e11.grid(row=11, column=1)
e12.grid(row=12, column=1)
e13.grid(row=13, column=1)



Button(master, text='Predict', command=show_entry_fields).grid()

mainloop()
```

6. SCREENSHOTS



Screenshot 6.1: Final Prediction of a patient with Heart Disease



Screenshot 6.2: Final Prediction of a patient without Heart Disease

7. TESTING

TESTING

6.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

6.2 TYPES OF TESTING

6.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .It is done after the completion of an individual unit before integration. This is a structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

7.2.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

7.2.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.
Invalid Input : identified classes of invalid input must be rejected.
Functions :. identified functions must be exercised.
Output : identified classes of application outputs must be exercised

Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases.

7.3 TEST CASES

7.3.1 CLASSIFICATION

Test case ID	Test case name	Purpose	Input	Output
1.	Disease Prediction	To detect if a person has heart disease	The user who has the disease gives the input of his health parameters	The output is the prediction if the person has heart disease
2.	Disease Prediction	To detect if a person has heart disease	The user who do not have the disease gives the input of his health parameters	The output is the prediction if the person has heart disease

8. CONCLUSION

8.CONCLUSION & FUTURE SCOPE

8.1 PROJECT CONCLUSION

With the increasing number of deaths due to heart diseases, it has become mandatory to develop a system to predict heart diseases effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart disease using UCI machine learning repository dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 90.16% for prediction of heart disease. 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.

8.2 FUTURE SCOPE

The accuracy of out project is around 90 percent, but with time and advanced hardware we can try to further improve the accuracy of our project .The dataset we used is available for public access. But the datasets that are available to public are very old data and is very less, so we would like to further improve the training process by providing bigger datasets. With availability of advanced hardware, we can further reduce the time taken to predict the results. We also have plans to further improve the GUI(graphical user interface).

9. BIBILIOGRAPHY

9. BIBILIOGRAPHY

9.1 REFERENCES

1. Guizhou Hu, Martin M.root, "Building prediction models for coronary disease", European science of cardiology,2005.

2. V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.

3. https://www.medicalnewstoday.com/articles/257484.php

9.2 GITHUB LINK