Heart Disease Prediction System

T.E. mini-project report submitted in partial

fulfilment of the requirements of the degree of

Information Technology

by

Swarali More (Roll no.11) [TEIT2]

Ghanshyam Patel (Roll no.22) [TEIT2]

Maitrey Phatak (Roll no.26) [TEIT2]

Hrutvik Rane (Roll no.33) [TEIT2]

Under the guidance of

Prof. Preeti Mishra

Department of Information Technology
Atharva College of Engineering, Malad(W)

University of Mumbai

2021–2022

CERTIFICATE

This is to certify that the T.E. mini-project entitled **"Heart Disease Prediction System"** is a bonafide work of **Swarali More** submitted to University of Mumbai in partial fulfilment of the requirement for the award of the degree of **"Information Technology"** during the academic year 2021–2022.

Prof. Preeti Mishra
Guide Name

Prof. Deepali Maste	**Dr. S.P. Kallurkar**
Head of Department	Principal

T.E. Mini-Project Report Approval

This mini-project synopsis entitled **Heart Disease Prediction System** by **Swarali More, Ghanshyam Patel, Maitrey Phatak** and **Hrutvik Rane** is approved for the degree of **Information Technology** from **University of Mumbai**.

Examiners

1.

2.

Date:

Place:

Declaration

We declare that this written submission represents our 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 cause 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.

Signature

Swarali More (Roll no. 11)

Signature

GhanshyamPatel(Roll no.22)

Signature

Maitrey Phatak (Roll no.26)

Signature

Hrutvik Rane (Roll no.33)

Date:

a
l

Abstract

Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance.. The proposed work predicts the person has Heart Disease or not by using machine leaning techniques such as Naive Bayes, Decision Tree, Logistic Regression. Thus, this paper presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verify that Logistic Regression algorithm has achieved the highest accuracy compared to other ML algorithms implemented.

Keywords : Machine Learning Logistic Regression, , Heart Disease Prediction

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACE	**Atharva College of Engineering**
CP	Chest Pain
Chol	Cholesterol
Exang	Exercise induced angina
Fbs	Fasting blood sugar
Slope	exercise ST segment
Thalach	Maximum heart rate achieved
Trestbps	Resting blood pressure
Oldpeak	Exercise relative to rest

Chapter 1 Introduction

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

1.1Motivation

Machine learning techniques have been around us and has been compared and used for analysis for many kinds of data science applications. The major motivation behind this research-based project was to explore the feature selection methods, data preparation and processing behind the training models in the machine learning. With first hand models and libraries, the challenge we face today is data where beside their abundance, and our cooked models, the accuracy we see during training, testing and actual validation has a higher variance. Hence this project is carried out with the motivation to explore behind the models, and further implement Logistic Regression model to train the obtained data. Furthermore, as the whole machine learning is motivated to develop an appropriate computer-based system and decision support that can aid to early detection of heart disease, in this project we have developed a model which classifies if patient will have heart disease or not using logistic regression. Hence, the early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine

1.2 Problem statement

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data

1.3 Objectives

The main objective of developing this project are:

1. To develop machine learning model to predict possibility of heart disease by implementing Logistic Regression
2. To determine significant risk factors based on medical dataset which may lead to heart disease.
3. To design an interactive application, in which user can give a single input to arrive the prediction.

Chapter 2 Review of Literature

Name	Authors	Publication	Year
A literature survey of predicting heart disease	M. Preethi, Dr. J. Selvakumar	International Research Journal of Engineering and Technology (IRJET)	2020
A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach	M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar	International Journal of Computer Applications	2018

Chapter 3 Report on Present Investigation

3.1 Proposed System

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data

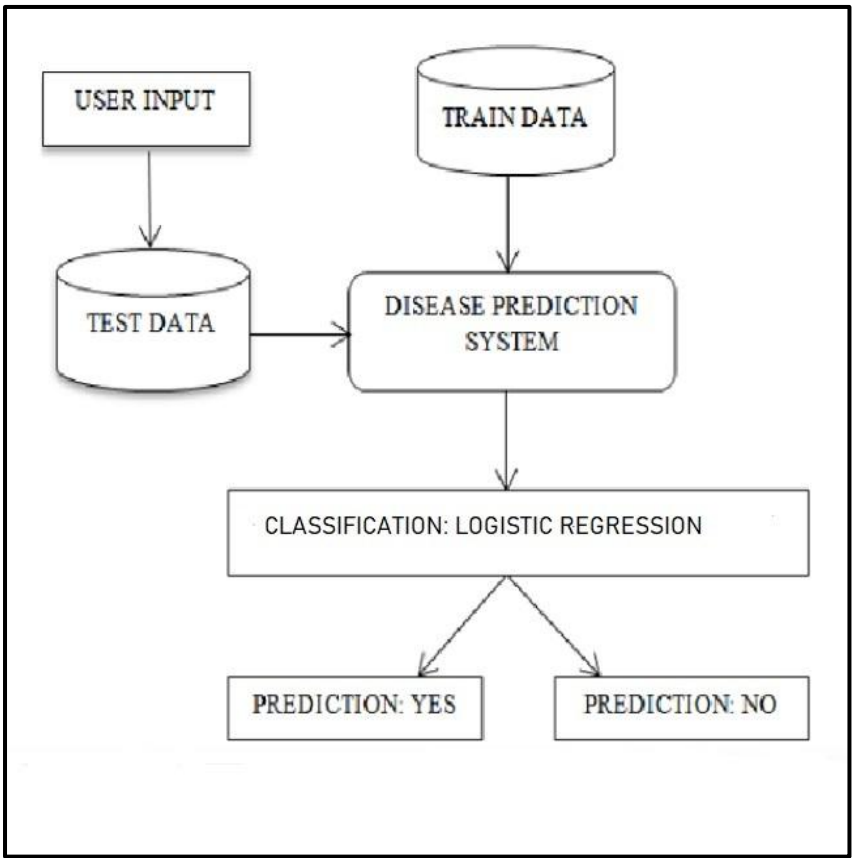

Figure 1: Architecture of Prediction System

3.1.1 Block diagram

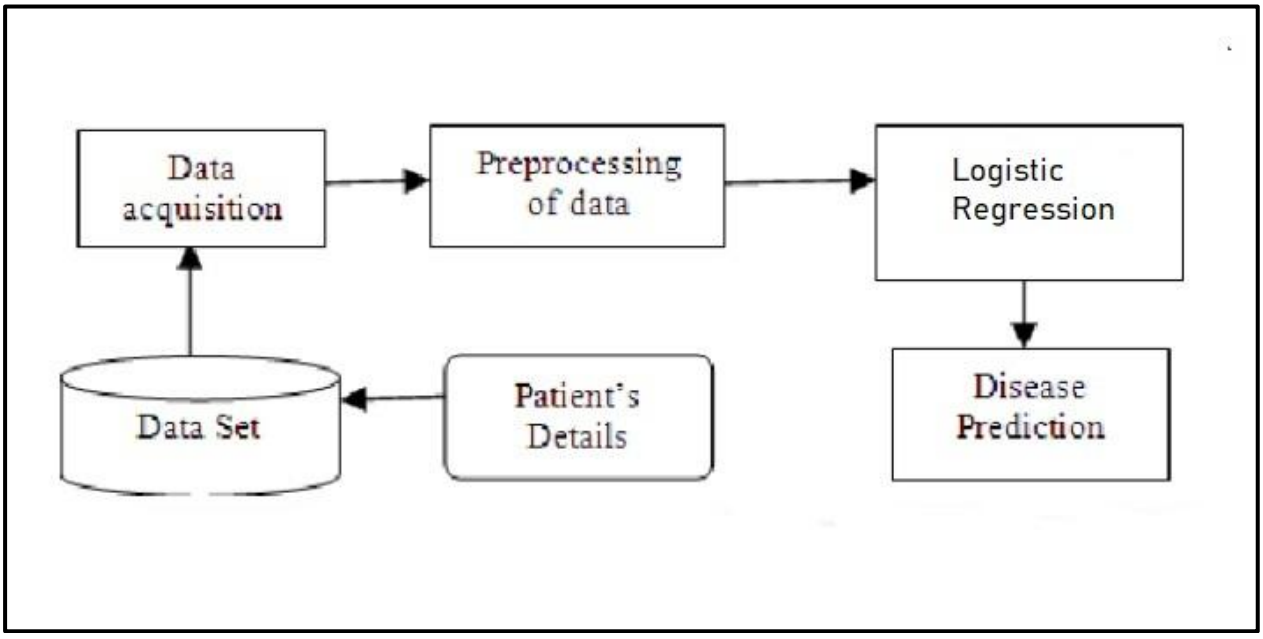

Figure 2 : Block diagram

3.2 Implementation

3.2.1 ML Algorithm

Logistic regression is a process of modelling the probability of a discrete outcome given an input variable. It is used in statistical software to understand the relationship between the dependent variable and one or more independent variables by estimating probabilities using a logistic regression equation.

3.2.2 Dataset description

The dataset is publicly available on the Kaggle Website which is from an on going cardiovascular study on residents of the town of Framingham, Massachusetts. It provides patient information which includes over 4000 records and 13 attributes. The attributes include: age, sex, chest pain (cp), resting blood pressure (trestbps), cholesterol(chol), fasting blood sugar(fbs), resting ECG(restecg), maximum rate rate acieved (thalach) ,exercise induced angina(exang), exercise relative to rest(oldpeak), exercise ST segment(slope), flourosopy(ca), thalassemia(thal) and target ranging from 0 and 1,where 1 is absence of heart disease. The data set is in csv (Comma Separated Value) format which is further prepared to data frame as supported by pandas library in python

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1

Figure 3: Original Dataset Snapshot

Code:

WEBUI Part

```
import numpy as np
import pickle
import streamlit as st


# loading the saved model
loaded_model = pickle.load(open('trained_modell.sav', 'rb'))


# creating a function for Prediction

def heartdisease_prediction(input_data):

    # changing the input_data to numpy array
    input_data_as_numpy_array = np.asarray(input_data)

    # reshape the array as we are predicting for one instance
    input_data_reshaped = input_data_as_numpy_array.reshape(1, -1)

    prediction = loaded_model.predict(input_data_reshaped)
    print(prediction)

    if (prediction[0] == 0):
      return 'The person does not have heart disease'
```

```
    else:
      return 'The person has heart disease'


def main():

    # giving a title
    st.title('Heart Disease Prediction Web App')

    # getting the input data from the user

    age = st.text_input('Age')
    sex = st.text_input('Sex(Male=1, Female=0)')
    cp = st.text_input('Chest pain type(cp)')
    trestbps = st.text_input('Resting blood pressure(trestbps)')
    chol = st.text_input('Cholesterol')
    fbs = st.text_input('Fasting blood sugar(fbs)')
    restecg = st.text_input('Resting electrocardiographic results(restecg)')
    thalach = st.text_input('Maximum heart rate achieved( thalach)')
    exang = st.text_input('Exercise induced angina(exang)')
    oldpeak = st.text_input('Exercise relative to rest(oldpeak)')
    slope = st.text_input('Slope of the peak exercise ST segments')
    ca = st.text_input('Number of major vessels(ca)')
    thal = st.text_input('Thalassemia(thal)')

    # code for Prediction
    diagnosis = ''

    # creating a button for Prediction

    if st.button('Disease Test Result'):
      diagnosis = heartdisease_prediction(
        [age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope,
ca, thal])

    st.success(diagnosis)


if __name__ == '__main__':
    main()
```

Rest of the Code:
https://colab.research.google.com/drive/1IBlEFSzAlcw2fI0YAZbJQHXmj7jbnXga?usp=sharing

Chapter 4 Model Implementation

4.1 Training of Model

Finally, this resulting data split into 80% train and 20% test data, which was further passed to the Logistic Regression model to fit, predict and score the model

4.2 Evaluation of Model

For the evaluation of our output from our training the data, the accuracy is calculated.

Chapter 5 Results and Discussion

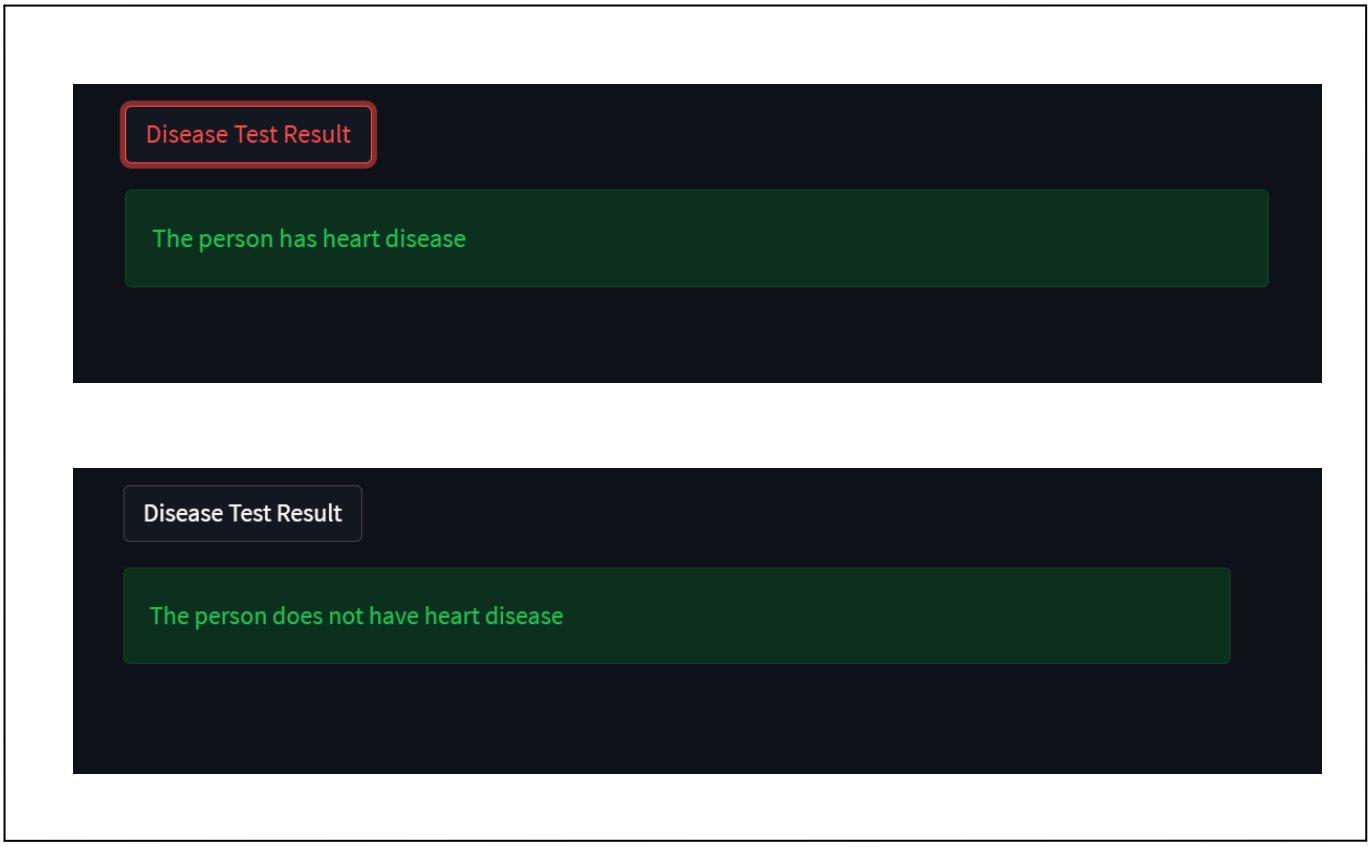

Figure 4: Output

Chapter 6 Model Deployment

We created an API using streamlit, which loads the trained ML model when the server first starts. We then enter patient's details, to determine healthiness of heart. The result is then predicted by the model.

Chapter 7 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 and Naive Bayes algorithms for predicting heart disease using machine learning repository dataset. The result of this study indicates that the Logistic Regression algorithm is the most efficient algorithm with accuracy score of 85% for prediction of heart disease.

Chapter 8 Future Scope

In future the work can be enhanced by developing a web application based on the Logistic Regression 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.
Further for its enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users, and also use more enhanced models

References:

- Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159.
- Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Network", The International Insitute of Science, Technology and Education, vol.5, No.4, 2015, pp.7-11.
- Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.
- 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.