

HEART DISEASE PREDICTION WEB APP USING MACHINE LEARNING

A Mini Project Submitted by

1. P. SRAVANI (19PA1A1297)
2. S.S. CHANDRIKA (19PA1A12B1)
3. T. S. N. MANIKANTA (19PA1A12B9)

Under the esteemed guidance of
Mr. D. UDAY KUMAR Sir



DEPARTMENT OF INFORMATION TECHNOLOGY

VISHNU INSTITUTE OF TECHNOLOGY
(Autonomous)

(Approved by AICTE, Accredited by NBA & NAAC and permanently affiliated to JNTU Kakinada)
BHIMAVARAM – 534 202
2022 – 2023

VISHNU INSTITUTE OF TECHNOLOGY

(Autonomous)

(Approved by AICTE, Accredited by NBA & NAAC and permanently affiliated to JNTU Kakinada)

BHIMAVARAM-534202

2022-2023

DEPARTMENT OF INFORMATION TECHNOLOGY



CERTIFICATE

This is to certify that the project entitled “HEART DISEASE PREDICTION WEB APP USING MACHINE LEARNING”, is being submitted by *P. Sravani, S.S. Chandrika and T. Manikanta*, bearing the **REGD.NOS: 19PA1A1297, 19PA1A12B1, 19PA1A12B9** submitted in fulfilment for the award of the degree of “**BACHELOR OF TECHNOLOGY**” in “**INFORMATION TECHNOLOGY**” is a record of bonafide work carried out by them under my guidance and supervision during the academic year 2022-2023 and it has been found worthy of acceptance according to the requirements of university.

Internal Guide

Mr. D. Uday Kumar

Head of the Department

Mrs. M. Sri Lakshmi

External Examiner

ACKNOWLEDGEMENT

It is nature and inevitable that the thoughts and ideas of other people tend to drift in to the subconscious due to various human parameters, where one feels acknowledge the help and guidance derived from others. We acknowledge each of those who have contributed for the fulfilment of this project.

We take the opportunity to express our sincere gratitude to **Dr.D.Suryanaryana**, director and principal, VIT, Bhimavaram whose guidance from time to time helped us to complete this project successfully.

We are very much thankful to **Mrs. M. Sri Lakshmi Madam**, Head of the Department, Department of Information Technology for his continuous and unrelenting support and guidance. We thank and acknowledge our gratitude to her for his valuable guidance and support expended to us right from the conception of the idea to the completion of this project.

We are very much thankful to **Mr. D. Uday Kumar Sir**, our internal guide whose guidance from time to time helped us to complete this project successfully.

Project Associates

P. Sravani	(19PA1A1297)
S.S. Chandrika	(19PA1A12B1)
T. Manikanta	(19PA1A12B9)

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 for a reliable, accurate, and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart-related diseases. Heart is the next major organ comparing to the brain which has more priority in the Human body. It pumps the blood and supplies it to all organs of the whole body.

Prediction of occurrences of heart diseases in the medical field is significant work. Data analytics is useful for prediction from more information and it helps the medical centre to predict various diseases. A huge amount of patient-related data is maintained on monthly basis. The stored data can be useful for the source of predicting the occurrence of future diseases. Some of the data mining and machine learning techniques are used to predict heart diseases, such as Logistic Regression, Artificial Neural Network (ANN), Random Forest and Support Vector Machine (SVM). Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. To reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop software with the help of machine learning algorithms which can help doctors and people to decide both prediction and diagnosing of heart disease. The main objective of this project is to predict the heart disease of a patient using machine learning algorithms...

TABLE OF CONTENTS

Sl.No	Contents	Page Numbers
1	Introduction	1
2	Dataset description	3
3	Data Pre-processing	5
	3.1 steps involved in data pre-processing	6
4	Technologies and Platforms	8
	4.1 Python	9
	4.2 Anaconda and Spyder	10
	4.3 GitHub and Heroku	10
5	Models	11
	5.1 Logistic Regression Model	12
6	Experimental setup	13
	6.1 How training data is given (Train Set, Test Set etc)	14
	6.2 Evolution metrics	14
	6.3 Developing App and Deploying	15
7	Results	16

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 the rate of several heart-related diseases. Heart diseases have emerged as one of the most prominent causes of death all around the world. According to World Health Organisation, heart related diseases are responsible for 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 diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15,2017. Heart-related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India has lost up to \$237 billion, from 2005-2015, due to heart-related or Cardiovascular diseases. Thus, feasible and accurate prediction of heart-related diseases is very important.

1.1 Hardware and Software Requirements:

1.1.1 Minimum Hardware Requirements:

- PROCESSOR: Intel I5
- RAM: 4GB
- HARD DISK: 256 GB

1.1.2 Minimum Software Requirements:

- OPERATING SYSTEM: Windows
- IDE: Google Colab, Anconda, Spyder
- PYTHON LIBRARIES: Pandas, Numpy, Streamlit, Pickle

DATASET DESCRIPTION

2. DATASET DESCRIPTION:

The dataset is available through the University of California, Irvine Machine learning repository. Here is the URL:

<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>

This dataset contains patient data concerning heart disease diagnosis that was collected at several locations around the world. There are 76 attributes, including age, sex, resting blood pressure, cholesterol levels, echocardiogram data, exercise habits, and many others. To date, all published studies using this data focus on a subset of 14 attributes - so we will do the same. More specifically, we will use the data collected at the Cleveland Clinic Foundation.

DATA PRE-PROCESSING

3. DATA PRE-PROCESSING:

Data pre-processing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance.

3.1 Steps involved in data pre-processing:

- Importing the required Libraries
- Importing the data set
- Handling the Missing Data.
- Encoding Categorical Data.
- Splitting the data set into test set and training set.
- Feature Scaling.

3.1.1 Libraries required for our project:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

3.1.2 Importing the dataset:

```
Heart_data = pd.read_csv('data.csv')
```

3.1.3 Handling the Missing Data: `isna()` is the method to check null values in dataset.

```
df=df.isna()
```

3.1.4 Splitting the data set into test set and training set:

Now we divide our data into two sets, one for training our model called the **training set** and the other for testing the performance of our model called the **test set**. The split is generally 80/20. To do this we import the “train_test_split” method of “sklearn.model_selection” library.

X_train (training part of the matrix of features),

X_test (test part of the matrix of features),

Y_train (training part of the dependent variables associated with the X train sets)

Y_test (test part of the dependent variables associated with the X test sets).

```
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, stratify=Y,  
random_state= 2)
```

TECHNOLOGIES AND PLATFORMS

4. TECHNOLOGIES

4.1 PYTHON

- Version Used – 3.10.8
- Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.
- Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it's relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

4.1.1 NUMPY

- NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

4.1.2 PANDAS

- Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools. Pandas module runs on top of NumPy and it is popularly used for data science and data analytics.

4.1.3 STREAMLIT

- Streamlit turns data scripts into shareable web apps in minutes All in pure Python. No front-end experience required.
- Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

4.1.4 PICKLE

- Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it “serializes” the object first before writing it to file. Pickling is a way to convert a python object (list, dict, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

PLATFORMS

4.2 ANACONDA AND SPYDER

- Conda is an open-source package and environment management system that runs on Windows, macOS, and Linux. Conda quickly installs, runs, and updates packages and their dependencies. It also easily creates, saves, loads, and switches between environments on your local computer. It was created for Python programs, but it can package and distribute software for any language.
- Spyder is short for Scientific PYthon Development EnviRonment Think of it as an IDE for scientific programming within Python.
- Spyder is a free and open source scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package

4.3 GITHUB AND HEROKU

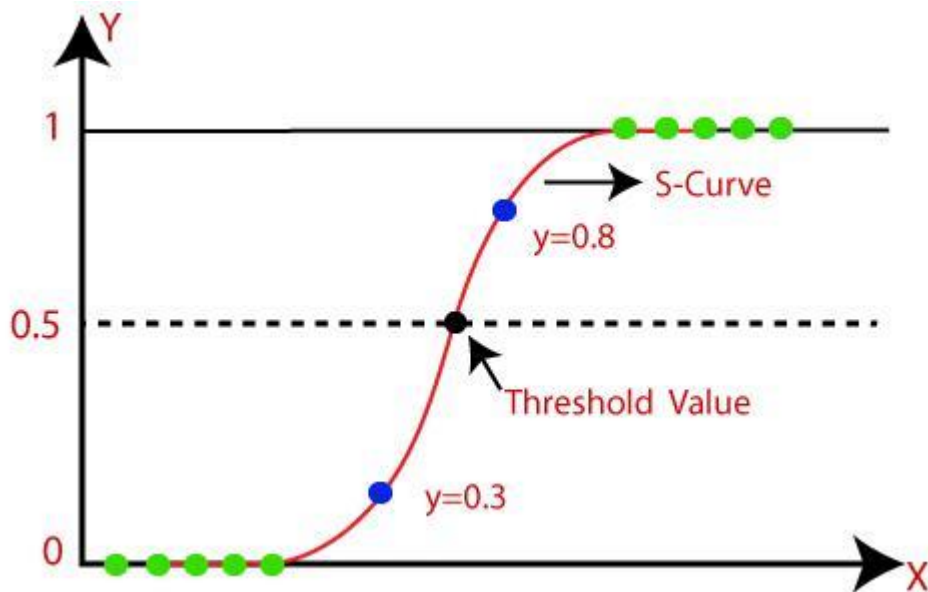
- GitHub is a code hosting platform for version control and collaboration. It lets you and others work together on projects from anywhere. This tutorial teaches you GitHub essentials like repositories, branches, commits, and pull requests.
- Heroku is a cloud platform as a service supporting several programming languages. One of the first cloud platforms, Heroku has been in development since June 2007, when it supported only the Ruby programming language, but now supports Java, Node.js, Scala, Clojure, Python, PHP, and Go.

MODELS

5. MODELS:

5.1 LOGISTIC REGRESSION:

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.



EXPERIMENTAL SETUP

6. Experimental Setup:

6.1 How training data is given:

We divide our data into two sets, one for training our model called the **training set** and the other for testing the performance of our model called the **test set**. The split is generally 80/20. To do this we import the “train_test_split” method of “sklearn.model_selection” library.

6.2 Evaluation Metrics Used:

6.2.1 Sklearn:

It is a free software machine learning library for Python programming language.

6.3 Developing and Deploying App:

Developing :

- In this phase we used the spyder ide for creating a web app in the local host.
- We used the Pickle module inorder to load the trained model which we have saved from the colab.
- We used a special module called STREAMLIT for creating the web app.
- We imported Streamlit as st and for displaying the title we used st.title() method.
- For taking the input from the user we used the st.text_input() method.
- So that we will be able to take the inputs from the user and send them to the function which computes the prediction whether the person have heart disease or not.
- Inorder to compute the prediction value we have created a button using st.button() method.

- It calls the heartdiseasepredict function and returns the result to the screen
- We then run the file using the command prompt of the spyder IDE.
- The web app will be displayed in the default browser.

Deploying:

- Till now we run the website in the local machine.
- Inorder to deploy the website we use the Heroku platform.
- First of all we upload the required files in the github by creating the repository.
- Then we have to create the PROCFILE, SETUP.sh file and requirement.txt file and upload them in the same repository.
- Now by logging into the Heroku we have to create the new app and then we have to connect to the github repository.
- Now we have click on the deploy button in the Heroku then after running all the modules the Heroku creates the app and provide a URL for the app.
- Link for the app : <https://heart-disease-predictionapp1.herokuapp.com/>.

RESULT:

Accuracy of Model: 82%

Code URL:

<https://colab.research.google.com/drive/1KdA9ad02n2gt5K5BedY4-9-5Dlm2GWfM?usp=sharing#scrollTo=eYMBypYXafda>

Web APP:

Heart Disease Prediction Web App

Age

Sex

CP of heart

Resting Blood Pressure

Cholestrol

Fasting Blood Sugar

Resting Cardiographic

Maximum Heart Rate Achieved

Exercise induced angina

Old Peak

Slope

Calcium present

Thalassemia

HeartDisease Test Result

heart-disease-predictionapp1.herokuapp.com

1

Resting Cardiographic

0

Maximum Heart Rate Achieved

150

Exercise induced angina

0

Old Peak

2.3

Slope

0

Calcium present

0

Thalassemia

1

HeartDisease Test Result

The Person has Heart Disease

heart-disease-predictionapp1.herokuapp.com

0

Resting Cardiographic

1

Maximum Heart Rate Achieved

123

Exercise induced angina

1

Old Peak

0.2

Slope

1

Calcium present

0

Thalassemia

3

HeartDisease Test Result

The Person does not have a Heart Disease