**HEART DISEASE PREDICTION**

 A

Report submitted in partial fulfilment

Of

the requirement for the degree

of

B.Tech.

In

###### Computer Science & Engineering

Under the Supervision of

##### Garima Verma (Assistant Professor)

By

##### Mohd Irfan (2001641530035)



Pranveer Singh Institute of Technology, Kanpur

Dr A P J A K Technical University

Lucknow

**Certificate**

*This is to certify that Mini Project Report entitled “****HEART DISEASE PREDICTION****” which is submitted by* ***Mohd Irfan*** *in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science and Engineering* ***of*** *Pranveer Singh Institute of Technology, affiliated to**Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my/our supervision. The project embodies result of original work and studies carried out by the students themselves and the contents of the project do not form the basis for the award of any other degree to the candidate or to anybody else****.***

|  |  |  |
| --- | --- | --- |
| *Signature:*  *Dr. Vishal Nagar*  *Head of Department*  *CSE Department,*  *PSIT, Kanpur* |  | *Signature:*  *Garima Verma*  *Assistant Professor*  *CSE Department,*  *PSIT, Kanpur* |

Date:

**DECLARATION**

*I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.*

|  |
| --- |
| *Signature*  *Name:*  *Roll No.:*  *Date:* |

***ACKNOWLEDGEMENT***

*It gives me a great sense of pleasure to present the report of the B.Tech. Project undertaken during B.Tech. second Year. I owe special debt of gratitude to my project supervisor* ***Garima Verma****, Department of Computer Science and Engineering, Pranveer Singh Institute of Technology, Kanpur for her constant support and guidance throughout the course of my work. Her sincerely, thoroughness and perseverance have been a constant source of inspiration for me. It is only his cognizant efforts that our endeavours have seen light of the day.*

*We also take the opportunity to acknowledge the contribution of Professor Dr. Vishal Nagar, Head, Department of Computer Science & Engineering, Pranveer Singh Institute of Technology, Kanpur for his full support and assistance during the development of the project.*

*We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.*

|  |  |
| --- | --- |
| *Signature*  *Name:*  *Roll No.:* |  |

**ABSTRACT**

**Heart Disease Prediction** is the application of science and technology to predict the person’s heart is defective or not for given their data set.  The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time. In this project different machine learning algorithms is applied to compare the results and analysis of the UCI Machine Learning Heart Disease dataset. The dataset consists of 14 main attributes used for performing the analysis. Various promising results are achieved and are validated using accuracy and confusion matrix. The dataset consists of some irrelevant features which are handled using Isolation Forest, and data are also normalized for getting better results. And how this study can be combined with some multimedia technology like mobile devices.

**TABLE OF CONTENT**

**CHAPTER NO. TITLE PAGE NO.**

TITLE …………………………………………………………………………………………………. 1

CERTIFICATE …………………………………………………………………………….………. 2

DECLARATION …………………………………………………………………………………… 3

ACKNOWLEDGEMENTS ……………………………………………………………...……. 4

ABSTRACT ………………………………………………………………………………….……… 5

LIST OF TABLES ………………………………………………………………………………… 6

Main Content Indexing Sample:

* 1. Introduction 7 - 8
     1. Overview 7
     2. Motivation 7
     3. Scope of Work 8
     4. Objective 8
  2. System Requirements 9-9
     1. Hardware Requirements 9
     2. Software Requirements 9
  3. Design / Implementation 10 - 19
     1. Flowchart 10
     2. Code 11 - 19
  4. Testing / Result analysis 20
  5. Conclusion and Future enhancement 21
  6. References 22

**Chapter 1: Introduction**

1. **Overview:**

## What Is Heart Disease?

The term “heart disease” refers to several types of heart conditions. The most common type of heart disease in the United States is coronary artery disease (CAD), which **affects the blood flow to the heart**. Decreased blood flow can cause a heart attack.

Using Machine Learning Model, we can predict the human’s disease is defective or not by training model with the help of their dataset with major accuracy.

1. **Motivation:**

The number of deaths due to cardiovascular diseases increased by 41% between 1990 and 2013, climbing from 12.3 million deaths to 17.3 million deaths in globally. In addition to that, half of the deaths in many developed countries are due to this issue. Therefore, early detection of heart diseases is required to reduce the health complications. Machine learning has been widely used in modern healthcare sector for diagnosing and predicting the presence of diseases using data models. Logistic regression is one such relatively used machine learning algorithms for complex diseases. Thus, identify the most significant model for cardiovascular diseases is predicting the overall risk by using logistic regression.

**3. Scope of Work:**

* The system uses 14 medical parameters such as **age, sex, blood pressure, cholesterol, and obesity** for prediction.
* The Heart Disease Prediction System predicts the likelihood of patients getting heart disease.
* It enables significant knowledge, relationships between medical factors related to heart disease and patterns, to be established.

**4. Objective:**

The objective of this study is **to effectively predict if the patient suffers from heart disease**. The health professional enters the input values from the patient's health report. The data is feed into model which predicts the probability of having heart disease.

**Chapter 2: System Requirements**

1. **Hardware Requirements**

* Intel Processor (3th Gen or Above) And AMD’s 2nd Generation Ryzen 5’s and 7’s Or Higher.
* SSD Storage for Faster Review but HDD Will Also Work.
* Ram (Min 4 Gb)
* A Stable Internet Connection for Faster Accessing.

1. **Software Requirements**

* Windows 10 Operating System (Recommended).
* **A Development Environment IDE** (Example: PyCharm, VSCode) or running code in any notebook like: Google Collab or Jupyter Notebook.
* Python Language
* Python Libraries like NumPy, Scikit-learn, logistic regression.
* Data Set from Kaggle

**Chapter 3: Design / Implementation**

**Flow chart for demonstrating working of Model:**

Diagram

Description automatically generated

**Understanding CODE:**

**Importing the Dependencies :)**

1. Importing following modules: numpy, pandas, and some sklearn modules.Graphical user interface, text, application, Word

   Description automatically generated

**Data Collection and Processing :)**

1. This is how to upload your data using Pandas module. Graphical user interface, text

   Description automatically generated
2. This will show you first 5 rows data from your data set. A screenshot of a computer

   Description automatically generated with medium confidence
3. This will show you last 5 rows data from your data set. Table

   Description automatically generated with medium confidence
4. How to know total number of rows and columns in your dataset

Graphical user interface, text, application

Description automatically generated

1. Getting some general information about the data. (303 non-null means we don’t have any missing values and Dtype tells their data type.)

Text

Description automatically generated with medium confidence

1. Another method for Checking for missing values in our data set. Table

   Description automatically generated with medium confidence
2. Statistical measures about the data. Graphical user interface, application

   Description automatically generated
3. Checking the distribution of Target Variable. 1 Represents the Defective Heart and 0 represents Healthy Heart.Graphical user interface, text, application, chat or text message

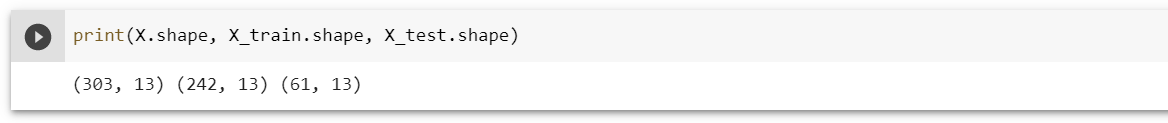
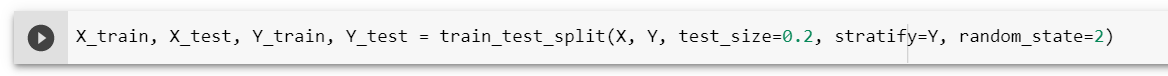
   Description automatically generated

**Splitting the features and targets :)**

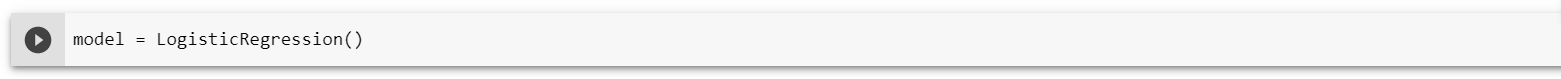
1. Splitting the features and targets. A picture containing graphical user interface

   Description automatically generated
2. Printing X. Table

   Description automatically generated
3. Printing Y. Graphical user interface, text, application, email

   Description automatically generated
4. Splitting the data into training data and test data. 

**Model Training :)**

1. Logistic Regression.Graphical user interface, text, application, email

   Description automatically generated

**Model Evaluation :)**

**Accuracy Score:**

1. Accuracy on Training data set.Text

   Description automatically generatedA picture containing text

   Description automatically generated
2. Accuracy on Test data set. Text, letter

   Description automatically generatedTable

   Description automatically generated with medium confidence

**Building a Predictive System:**

1. Building a predictive system.Graphical user interface, text, application, email

   Description automatically generated

**Code For Heart Disease Prediction:**

""" Importing the Dependencies """

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

""" Data Collection and Processing """

# loading the csv data to a Pandas DataFrame

heart\_data = pd.read\_csv('/content/data.csv')

# print first 5 rows of the dataset

heart\_data.head()

# print last 5 rows of the dataset

heart\_data.tail()

# number of rows and columns in the dataset

heart\_data.shape()

# getting some info about the data

heart\_data.info()

# checking for missing values

heart\_data.isnull().sum()

# statistical measures about the data

heart\_data.describe()

# checking the distribution of Target Variable

heart\_data['target'].value\_counts()

"""

1 --> Defective Heart

0 --> Healthy Heart

"""

""" Splitting the Features and Target """

X = heart\_data.drop(columns='target', axis=1)

Y = heart\_data['target']

print(X)

print(Y)

""" Splitting the Data into Training data & Test Data """

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.2, stratify=Y, random\_state=2)

print(X.shape, X\_train.shape, X\_test.shape)

"""Model Training"""

# Logistic Regression

model = LogisticRegression()

# training the LogisticRegression model with Training data

model.fit(X\_train, Y\_train)

""" Model Evaluation """

"""Accuracy Score """

# accuracy on training data

X\_train\_prediction = model.predict(X\_train)

training\_data\_accuracy = accuracy\_score(X\_train\_prediction, Y\_train)

print('Accuracy on Training data : ', training\_data\_accuracy)

# accuracy on test data

X\_test\_prediction = model.predict(X\_test)

test\_data\_accuracy = accuracy\_score(X\_test\_prediction, Y\_test)

print('Accuracy on Test data : ', test\_data\_accuracy)

""" Building a Predictive System """

input\_data = (62, 0, 0, 140, 268, 0, 0, 160, 0, 3.6, 0, 2, 2)

# change the input data to a numpy array

input\_data\_as\_numpy\_array = np.asarray(input\_data)

# reshape the numpy array as we are predicting for only on instance

input\_data\_reshaped = input\_data\_as\_numpy\_array.reshape(1, -1)

prediction = model.predict(input\_data\_reshaped)

print(prediction)

if prediction[0] == 0:

    print('The Person does not have a Heart Disease')

else:

    print('The Person has Heart Disease')

**Chapter 4: Testing / Result and Analysis**

**Test Output 1 (Defective heart):**

**Graphical user interface, text, application, email

Description automatically generated**

**Test Output 2 (Healthy Heart):**

**Graphical user interface, text, application, email

Description automatically generated**

Hence, according to the results, if we give all data except 14th column (Target) as an input then it will give us the data in 0 or 1 which describes that person has **‘Heart Disease’** or not.

**Chapter 5: Conclusion and Future Enhancement**

**Conclusion:**

This project provides the overview into machine learning techniques for predicting heart diseases. The role of regression model is crucial in healthcare industry so that the results can be used for predicting the treatment which can be provided to patients. The existing techniques very good for finding the efficient and accurate systems.

Machine learning techniques significantly improves accuracy of cardiovascular risk prediction through which patients can be identified during an early stage of disease and can be benefitted by preventive treatment.

It can be concluded that there is a huge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases.

**Future Enhancement:**

**Naïve Bayes** was found to be the best algorithm, followed by neural networks and decision trees.

Artificial neural networks are also employed for the prediction of diseases. Supervised networks have been used for diagnosis and they can be trained using the Back Propagation Algorithm.

**Limitations:**

1. The accuracy can be more achieve by using **Naïve Bayes**.
2. Deep Learning ANN are better for the prediction of diseases. Supervised networks have been used for diagnosis.

**References**

* <https://www.kaggle.com/>
* <https://numpy.org/doc/stable/user/index.html#user>
* <https://scikit-learn.org/stable/supervised_learning.html#supervised-learning>
* <https://www.youtube.com/>

**Signature of Mini Project Co-ordinator:**

**Signature of H.O.D.:**