INTER DEPARTMENT PROJECT REPORT ON

**“Heart Disease Prediction System”**

**Submitted**

By

|  |  |
| --- | --- |
| Ms. A. Ramya Nandini  191FA04002 | Ms. R. V. A. Sathwika Sri  191FA04040 |
| Ms. V. Divya Sree  191FA04052 | Ms. E. Rani Prathyusha  191FA04376 |
| Ms. G.Jhansi  191FA04377 | Mr. B. Sai Gopi  191FA01016 |

**Under the guidance of**

|  |
| --- |
| **Dr. M. NirupumaBhat**  **professor**  **Department of CSE** |
|  |



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY**

**Vadlamudi, Guntur.**

**VIGNAN’S FOUNDATION FOR SCIENCE TECHNOLOGY AND RESEARCH**

**Deemed to be UNIVERSITY**

VADLAMUDI, GUNTUR DIST, ANDHRA PRADESH, INDIA, PIN-522 213



**CERTIFICATE**

This is to certify that the Inter Department Project Report entitled **“Heart Disease Prediction System”** that is being submitted by **A. Ramya Nandini(191FA04002),R.Sathwika(191FA04040),V.DivyaSree(191FA04052),E.RaniPrathyusha(191FA04376),G.Jhansi(191FA04377)** in association with **Bio Technology - B. Sai Gopi(191FA01016)** in partial fulfilment of Inter Department Project course work is a bonafide work carried under the supervision of Dr. Venkatesulu Dondeti(Head Of The Department CSE), Dr. M. NirupumaBhat(Professor, Department of CSE), Mr.N.Venkata Raju(Assistant professor ,BT).

**Dr. M.NirupumaBhat Mr.N.Venkata Raju Dr. VenkatesuluDondeti**  
 Professor Assistant professor HOD,CSE

Department of CSE Department of BT

For the External Inter Department Project Viva-voce scheduled on 02-02-2022.

Internal Examiner External Examiner

**VIGNAN’S FOUNDATION FOR SCIENCE TECHNOLOGY AND RESEARCH**

**Deemed to be UNIVERSITY**

VADLAMUDI, GUNTUR DIST, ANDHRA PRADESH, INDIA, PIN-522 213



**DECLARATION**

We hereby declare that the Inter Department Project entitled **“Heart Disease Prediction System”** is being submitted by **A. Ramya Nandini(191FA04002),R.Sathwika(191FA04040),V.DivyaSree(191FA04052), E. Rani Prathyusha(191FA04376), G. Jhansi(191FA04377)** in partial fulfilment of Inter Department Project course work. This dissertation is our original work, and the project has not formed the basis for the award of any degree. We have worked in association with **Bio Technology - B. Sai Gopi(191FA01016)** under the supervision of Dr.M.NirupumaBhat(Professor) from Department of Computer Science & Engineering and Mr.N.Venkata Raju(Assistant professor ,BT).

By

**191FA04002**

**191FA04040**

**191FA04052**

**191FA04376**

**191FA04377**

**191FA01016**

Date: 2-02-2022

**ACKNOWLEDGEMENT**

It gives us great sense of pleasure to report on “Heart Disease Prediction System” . For this we are sincerely thankful to Dr.M.Nirupuma Bhat , Mr.N.Venkata Raju for their help, invaluable guidance and elating encouragement through out the course of the present work. I would like to acknowledge internet for doubt clarification during the project. Finally, we are deeply thankful to our parents and teachers who helped and inspired us in completing the project.

Date: 2-02-2022

**TABLE OF CONTENTS**

[1. PROBLEM STATEMENT 6](#_Toc40173436)

[2. REQUIREMENTS SPECIFICATION 7](#_Toc40173442)

[3. TECHNOLOGIES ADOPTED 8](#_Toc40173453)

[4. DESIGN AND IMPLEMENTATION OF SOLUTION 9](#_Toc40173458)

[5. DESIGN OF TESTCASES & TESTING 17](#_Toc40173463)

[6. CONCLUSION 17](#_Toc40173475)

[BIBLIOGRAPHY 18](#_Toc40173479)

**1.PROBLEM STATEMENT:**

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced

and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive.

The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is a wealth of data available within the health care

healthcare. The healthcare environment is still „information rich‟ but „knowledge poor‟. There

systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.

care and reduce costs. According to (Wurz & Takala, 2006) ⁠the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. The successful application of data mining in highly visible fields like ebusiness, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is

healthcare. The healthcare environment is still „information rich‟ but „knowledge poor‟. Thereis a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.

**1a). Scope:**

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions

**2.REQUIREMENTS SPECIFICATION:**

**2a) Software Requirements:**

Operating System Any OS with clients to access the internet Network Wi-Fi Internet or cellular Network Visio Studio Create and design Data Flow and Context Diagram Github Versioning Control Google Chrome Medium to find reference to do system testing, display and run shinyApp.In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted result.

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted resul

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted resul

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted result.

Operating System : Any OS with clients to access the internet

Network : Wi-Fi Internet or cellular Network

Visio Studio : For implementation of code

Google Chrome : Medium to find reference to do system testing, display and run streamlit Webapp.

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted result

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted result

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted result

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted resul

In this system we are implementing effective heart attack prediction system using Naïve Bayes

algorithm. We can give the input as in CSV file or manual entry to the system. After taking input

the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is

performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight,

age and the priority levels are by consulting expertise doctors and the medical experts. The heart

attack prediction system designed to help the identify different risk levels of heart attack like

normal, low or high and also giving the prescription details with related to the predicted resul

**2b) Hardware Requirements:**

For webapplication development, the following Software Requirements are:

Processor : Intel or high

RAM : 1024 MB

Space on disk: minimum 100mb

For running the application:

Device : Any device

Minimum space to execute: 20 MB

The effectiveness of the proposal is evaluated by conducting experiments with identical setting, configured with an Intel CORE™ i7-4770 processor (3.40GHZ, 4 Cores, 8GB RAM, running Ubuntu 18.04 LTS with 64-bit Linux 4.31.0 kernel)

**3.TECHNOLOGIES ADOPTED:**

**3a)Streamlit**



Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers. Data scientists or machine learning engineers are not web developers and they’re not interested in spending weeks learning to use these frameworks to build web apps. Instead, they want a tool that is easier to learn and to use, as long as it can display data and collect needed parameters for modeling. Streamlit allows you to create a stunning-looking application with only a few lines of code.

**3b)Visual Studio Code**



Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code .

**3c)Libraries Used**

**i)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.



**ii)Pandas:**

Python Pandas Introduction Pandas is defined as anopen-source library that provides high-performance data manipulation in Python.

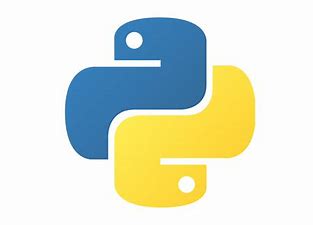


**iii)Scikit-Learn**



Scikit-learn (Sklearn) isthe most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machinelearning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

**3d)Language – Python**



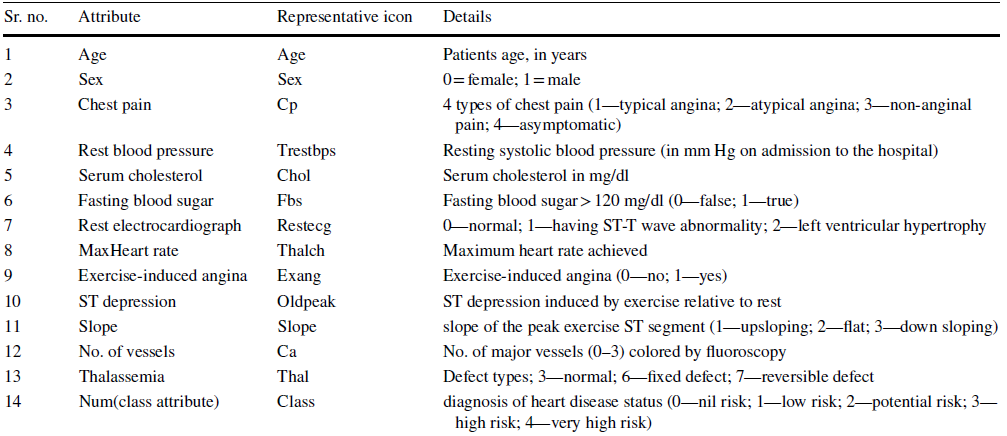
Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small & large-scale & projects.Python is dynamically-typed and garbage-collected.

**4.DESIGN AND IMPLEMENTATION OF SOLUTION:**

In this system we are implementing effective heart attack prediction system using Naïve Bayes, K-nearest neighbors, Decision tree, Random forest, Support Vector Machine (SVM), Logistic regression algorithms.

We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is the selected one. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

**4a)Attributes Info:**



**4b)Algorithms implemented:**

|  |  |
| --- | --- |
| **Algorithm Used** | **Accuracy** |
| Naïve Bayes | 84.71% |
| Decision Tree | 100% |
| K-NN | 78.10% |
| Random forest | 100% |
| SVM | 85.54% |
| Logistic regression | 85.12% |

**4c)PROJECT FLOW:**

Diagram

Description automatically generated

**4d) CODE:**

import streamlit as st

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import accuracy\_score

import warnings

warnings.filterwarnings('ignore')

heart\_data = pd.read\_csv("heart.csv")

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

df=pd.DataFrame(heart\_data)

Y = heart\_data['target']

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

st.title("Heart disease prediction system")

st.line\_chart(df)

st.sidebar.title("Enter input values ")

form = st.sidebar.form(key='my\_form')

age=form.number\_input(label="Age")

gender=form.radio("Gender",["Male","Female"])

cp=form.number\_input(label="Chest Pain")

bp=form.number\_input(label = "Enter Rest blood pressure")

chol=form.number\_input(label = "Enter Serum cholesterol")

fbs=form.number\_input(label = "Enter Fasting blood sugar")

restecg=form.number\_input(label = "Enter Rest electrocardiograph")

thalch=form.number\_input(label = "Enter MaxHeart rate")

exang=form.number\_input(label = "Enter Exercise-induced angina")

oldpeak=form.number\_input(label = "Enter ST depression")

slope=form.number\_input(label = "Enter slope")

ca=form.number\_input(label = "Enter No. of vessels ")

thal=form.number\_input(label = "Enter thalassemia")

Select=form.selectbox("Algorithm",["RANDOM FOREST","KNN","NAIVE BAYES","DECISION TREE","LOGISTIC REGRESSION","SVM"])

submit\_button = form.form\_submit\_button(label='Submit')

if gender=="Male":

gender=1

else:

gender=0

if submit\_button:

if Select=="RANDOM FOREST":

from sklearn.ensemble import RandomForestClassifier

model= RandomForestClassifier()

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

elif Select=="KNN":

from sklearn.neighbors import KNeighborsClassifier

model= KNeighborsClassifier()

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

elif Select=="NAIVE BAYES":

from sklearn.naive\_bayes import GaussianNB

model=GaussianNB()

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

elif Select=="DECISION TREE":

from sklearn.tree import DecisionTreeClassifier

model= DecisionTreeClassifier()

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

elif Select=="LOGISTIC REGRESSION":

from sklearn.linear\_model import LogisticRegression

model = LogisticRegression()

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

elif Select=='SVM':

from sklearn.svm import SVC

model = SVC(C=1,random\_state=1,kernel='linear')

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

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

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

input\_data = (int(age),int(gender),int(cp),int(bp),int(chol),int(fbs),int(restecg),int(thalch),int(exang),int(oldpeak),int(slope),int(ca),int(thal))

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

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

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

if prediction[0]==1:

st.subheader("Result: Positive")

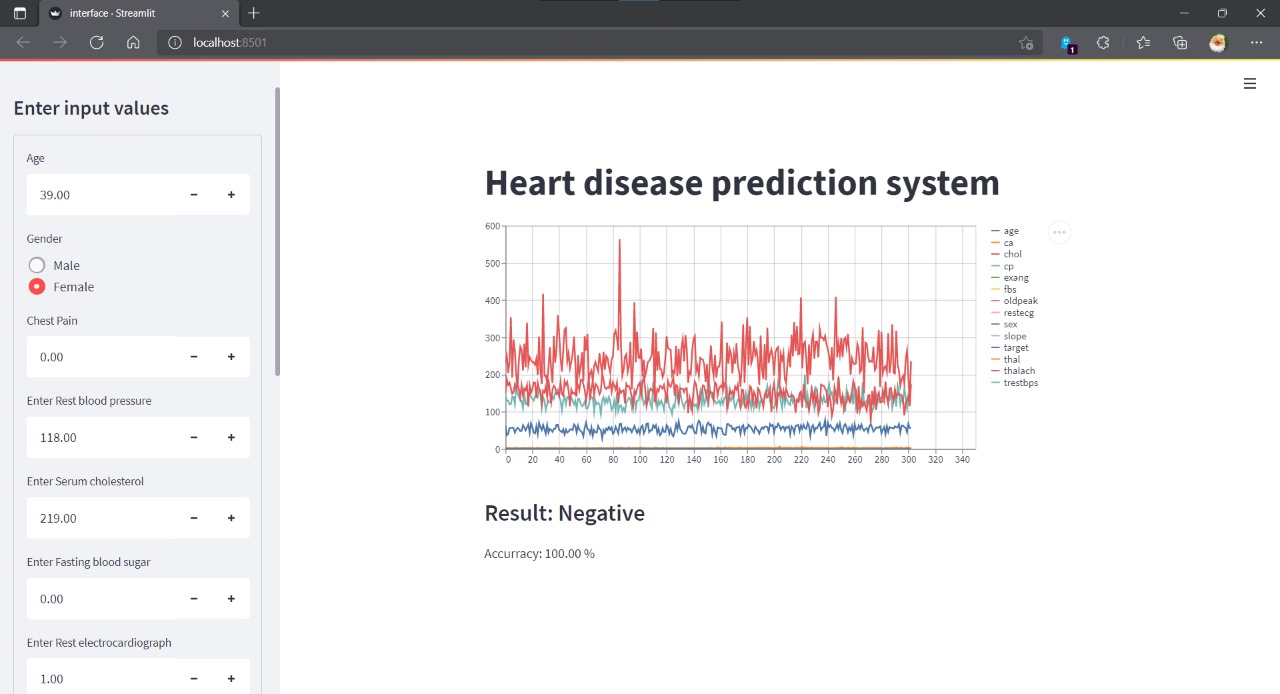
st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

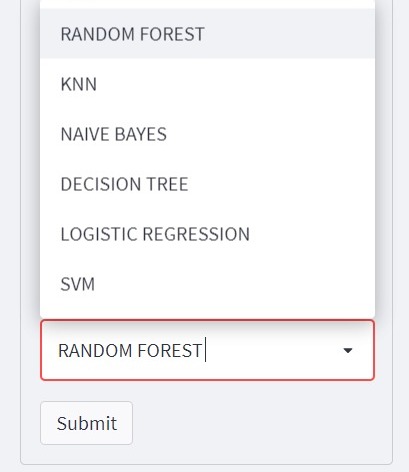
else:

st.subheader("Result: Negative")

st.write("Accurracy: ","%.2f"%(training\_data\_accuracy\*100),"%")

**5.DESIGN OF OUTPUT:**

****

****

**6.Conclusion:**

The overall aim is to define various data mining techniques useful in effective heart disease prediction. Efficient and accurate prediction with a lesser number of attributes and tests is our goal. In this study, We consider only 14 essential attributes. We applied 6 data mining classification techniques, K-nearest neighbor, Naive Bayes,

Decision tree, Logistic Regression, SVM, Random forest.

The data were pre-processed and then used in the model. Decision Tree and Random Forest are the algorithms showing the best results in this model.

**BIBLIOGRAPHY:**

A, A. S., & Naik, C. (2016). Different Data Mining Approaches for Predicting Heart Disease, 277– 281. <https://doi.org/10.15680/IJIRSET.2016.0505545>

Beyene, C., & Kamat, P. (2018). Survey on prediction and analysis the occurrence of heart disease using data mining techniques. International Journal of Pure and Applied Mathematics, 118(Special Issue 8), 165–173. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0- 85041895038&partnerID=40&md5=2f0b0c5191a82bc0c3f0daf67d73bc81 Kirmani, M. (2017). Cardiovascular Disease Prediction using Data Mining Techniques. Oriental Journal of Computer Science and Technology, 10(2), 520–528. <https://doi.org/10.13005/ojcst/10.02.38>

Science, C., & Faculty, G. M. (2009). Heart Disease Prediction Using Machine learning and Data Mining Technique. Ijcsc 0973-7391, 7, 1–9.

care and reduce costs. According to (Wurz & Takala, 2006) ⁠the opportunities to improve careand reduce costs concurrently could apply to as much as 30% of overall healthcare spending. The