IFET COLLEGE OF ENGINEERING

(AN AUTONOMOUS INSTITUTION)
VILLUPURAM

MINI PROJECT

Subject Code: 19UCSMP501

MACHINE LEARNING-BASED HEART PATIENT SCANNING, VISUALIZATION, AND MONITORING

By

HARI. K(201021028)

Department of CSE/Third Year

Under the guidance of

Mr. D. JAYAKUMAR /AP

Department of CSE

BONAFIDE CERTIFICATE

Certified that this mini project report on **"MACHINE LEARNING-BASED HEART PATIENT SCANNING, VISUALIZATION, AND MONITORING "** is a Bonafide work of **HARI. K (201021028)** who carried out the project work under my supervision.

SIGNATURE OF THE GUIDE **HEAD OF THE DEPARTMENT**

The mini project report submitted for the viva voce held on ….....

INTERNAL EXAMINER **EXTERNAL EXAMINER**

MACHINE LEARNING-BASED HEART PATIENT SCANNING, VISUALIZATION, AND MONITORING

ABSTRACT

Heart diseases leading most causes of death globally according to World Health Organization cardiovascular or all heart related disease are responsible for 17.9 million death every year. An early detection and diagnosis of the disease is very important and maybe its the key of cure. The major challenge is to predict the disease in early stages therefor most of scientists and researches focus on Machine learning techniques which have the capability of detection with accurate result for large and complex data and apply those techniques to help in health care. The purpose of this work is to detect heart diseases at early stage and avoid consequences by implementing different Machine Learning Algorithm , Logistic Regression.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	**4**
1	**INTRODUCTION**	**6**
	1.1 OBJECTIVES	
	1.2 SCOPE OF THE PROJECT	
2	**EXISTING SYSTEM**	**9**
	2.1 ADVANTAGES	
3	**METHODOLOGY**	**11**
	3.1 DATASET AND ATTRIBUTES	
4	**SYSTEM ARCHITECTURE**	**13**
5	**LANE DETECTION AND ASSISTANCE SYSTEM**	**14**
6	**SYSTEM REQUIREMENTS**	**17**
	6.1 SOWFTWARE REQUIREMENTS	
	6.2 HARDWARE REQUIREMENTS	
7	**RESULT**	**18**
8	**CONCLUSION**	**23**
	REFERENCE	**24**
	APPENDIX	

CHAPTER-1

INTRODUCTION

This paper focuses mainly on various Machine Learning techniques that are employed in heart disease prediction, cardiac (heart) is very important organ on the body which is responsible for regular blood flow throughout the body, therefore any irregularity to heart can cause distress in other parts of the body. Today, there is more reasons for heart diseases such as unhealthy lifestyle, smoking etc. Alcohols are major causes of heart disease. Good health style and early detection are the most way to prevent heart disease. Machine learning, a subfield of artificial intelligence, can learn from massive datasets and predict similarly previously unseen or new data based on its methods of learning or training. There are various kinds of cardiovascular diseases, within each variety of symptoms, such as: 1 - cardiovascular disease caused by an irregular heartbeat, low heartbeat, anxiety, and chest pain. 2- Blood vascular disease in the heart that causes chest discomfort and breathlessness. There are several causes of heart diseases, such as high blood pressure, hypertension, and drugs. Heart diseases include, heart infections, heart failure, cardiac arrest, hypertension, slow heartbeat, and stroke. Many factors for heart diseases are age, family history of coronary illness, blood pressure, and Cholesterol level.

1.1 OBJECTIVES

Machine learning methods potentially provide a highly accurate and detailed assessment of expected individual patient risk before elective cardiac surgery. Correct anticipation of this risk allows for the improved counselling of patients and avoidance of possible complications.

The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc.

The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest.

1.2 SCOPE OF THE SYSTEM

Conditions that fall within the scope of heart disease include cardiac arrhythmias, high blood pressure, heart failure, coronary artery disease, valve disorders, and congenital heart defects, among others.

Heart disease is a term used to describe a constellation of conditions that can affect the heart and/or its valves, vessels, structure, electrical system, or coronary arteries.

Conditions that fall within the scope of heart disease include cardiac arrhythmias, high blood pressure, heart failure, coronary artery disease, valve disorders, and congenital heart defects, among others.

Though each disease affects the heart differently, the ultimate problem with all varieties of heart disease is that, in one way or another, they can disrupt the vital pumping action of the heart.

CHAPTER 2

EXISTING SYSTEM

In this system, the input details are obtained from the patient. Then from the user inputs, using ML techniques heart disease is analyzed. Now, the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 87% for F-measure, competing with the other existing methods.

2.1 ADVANTAGES:

- Gives good results even if there is not enough information about the data. Also works well with unstructured data.
- Solves complex problems with a convenient kernel solution function.
- Relatively good scaling of high dimensional data.

CHAPTER 3

METHODOLOGY:

The below block diagram shows the basic steps which have been taken for all machine learning model, firstly data cleaning to convert the raw data into a manageable form because raw data cannot be used directly, then determine important features those steps are applied to each machine learning model predictions.

3.1 DATASET AND ATTRIBUTES:

- Age
- anaemia
- creatinine_phosphokinase
- diabetes
- ejection_fraction
- high_blood_pressure
- platelets
- serum_creatinine
- serum_sodium
- sex
- smoking
- time
- Death_event

CHAPTER 4

SYSTEM ARCHITECTURE

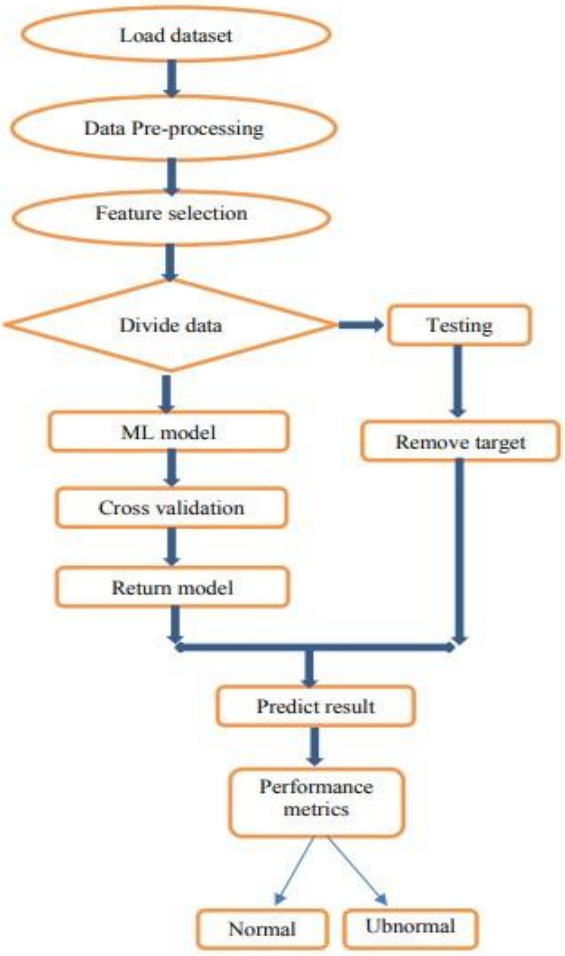

Figure 4.1 Prediction

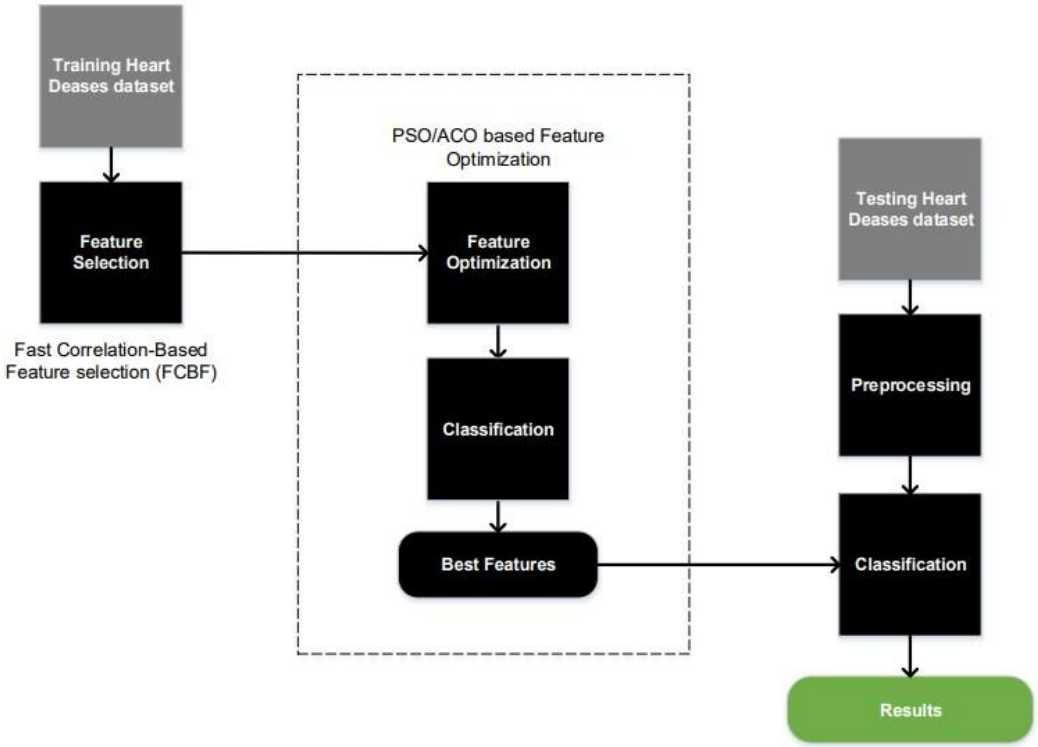

Figure 4.2 proposed architecture

CHAPTER 5

LANE DETECTION AND ASSISTANCE SYSTEM

5.1 HEART DISEASE PREDICTION

Machine Learning is used across many spheres around the world. The healthcare industry is no exception. Machine Learning can play an essential role in predicting presence/absence of Locomotor disorders, Heart diseases and more. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment per patient basis.

5.2 DATA COLLECTION

This synthetic dataset is generated using Matlab automotive driving toolbox to simulate a 77GHz FMCW millimeter-wave radar sensing in the road scenario. Especially for the Doppler ambiguity case, when the object vehicles move within or out of the unambiguous detecable velocity range. The dataset contains in total 20 recordings with the duration of 1 second each. Both time-division modulation (TDM) and binary phase modulation (BPM) data are provided. Each recording consists of complex CSV raw data and complex range-Doppler map, together with the ground-truth range and velocity.

5.3 PREPROCESSING

Heart disease data is pre-processed by using various collection of records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing.

5.3 ALGORITHM:

This dataset contains information about users from a company's database. It contains information about UserID, Gender, Age, Deathevents, and high_BP and etc. We are using this dataset for predicting whether a user will purchase the company's newly launched product or not.

CHAPTER 6

SYSTEM REQUIREMENTS

6.1 Software Requirements:

1. CSV
2. Python
3. Jupiter Notebook

6.1.1 CSV

A CSV file (Comma Separated Values file) is a type of plain text file that uses specific structuring to arrange tabular data. Because it's a plain text file, it can contain only actual text data—in other words, printable ASCII or Unicode characters. The structure of a CSV file is given away by its name.

6.1.2 Python

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.

6.1.3 Jupiter notebook

jupyterLab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality.

6.2 Package Requirements:

1. NumPy
2. Matplotlib
3. Pandas
4. Plotly
5. Seaborn

6.2.1 NumPy

The Python package NumPy is used to manipulate arrays. Additionally, it has matrices, Fourier transform, and functions for working in the area of linear algebra. In the year 2005, Travis Oliphant developed NumPy. You can use it for free because it is an open-source project.

6.2.2 Matplotlib

For the Python programming language and its NumPy numerical mathematics extension, Matplotlib is a graphing library. For integrating charts into programmes utilizing all-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK, it offers an object-oriented API.

6.2.3 Pandas

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

6.2.4 Plotly

The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

6.2.5 Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. For a brief introduction to the ideas behind the library, you can read the introductory notes or the paper.

6.3 Hardware Requirements:

1. Flow of the proposed methodology. After cleaning and analyzing all dataset, apply machine. learning models such as SVM Logistic Regression, Naïve.
2. Support Vector Machine. Support vectors are the closest data point set at the. Hyperplane and are very important points in the data.

CHAPTER 7

RESULT

Figure 7.1: sample output of Data entry page

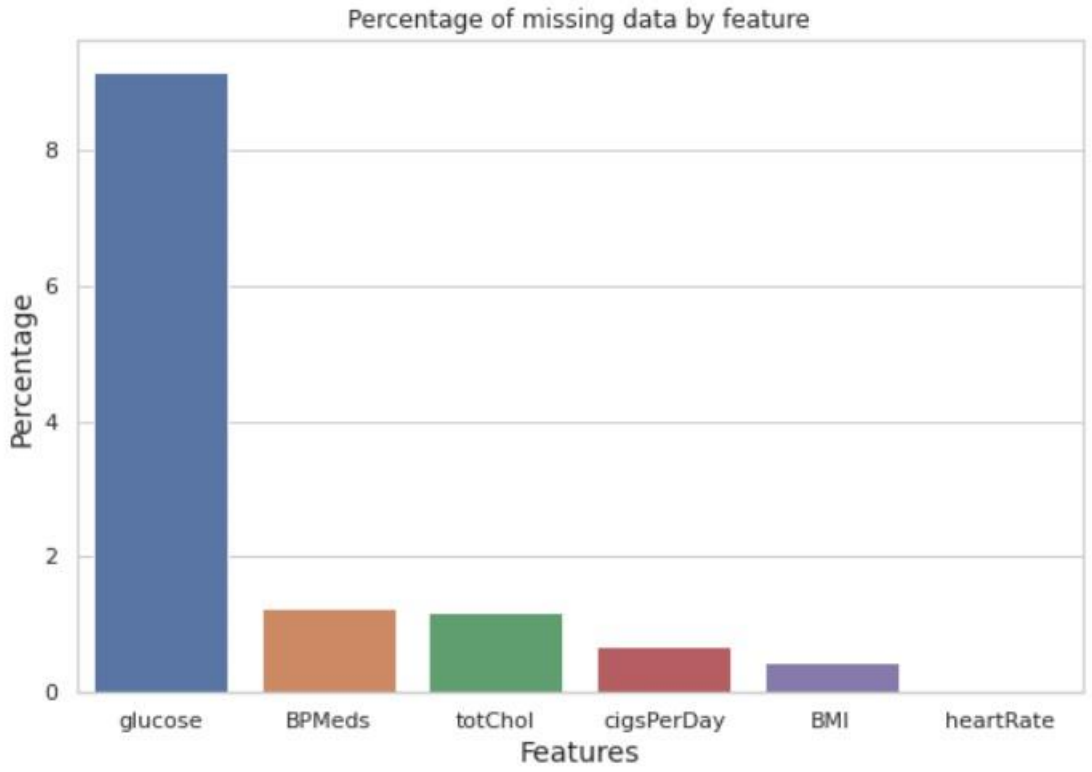

Figure 7.2: sample output of percent of detection

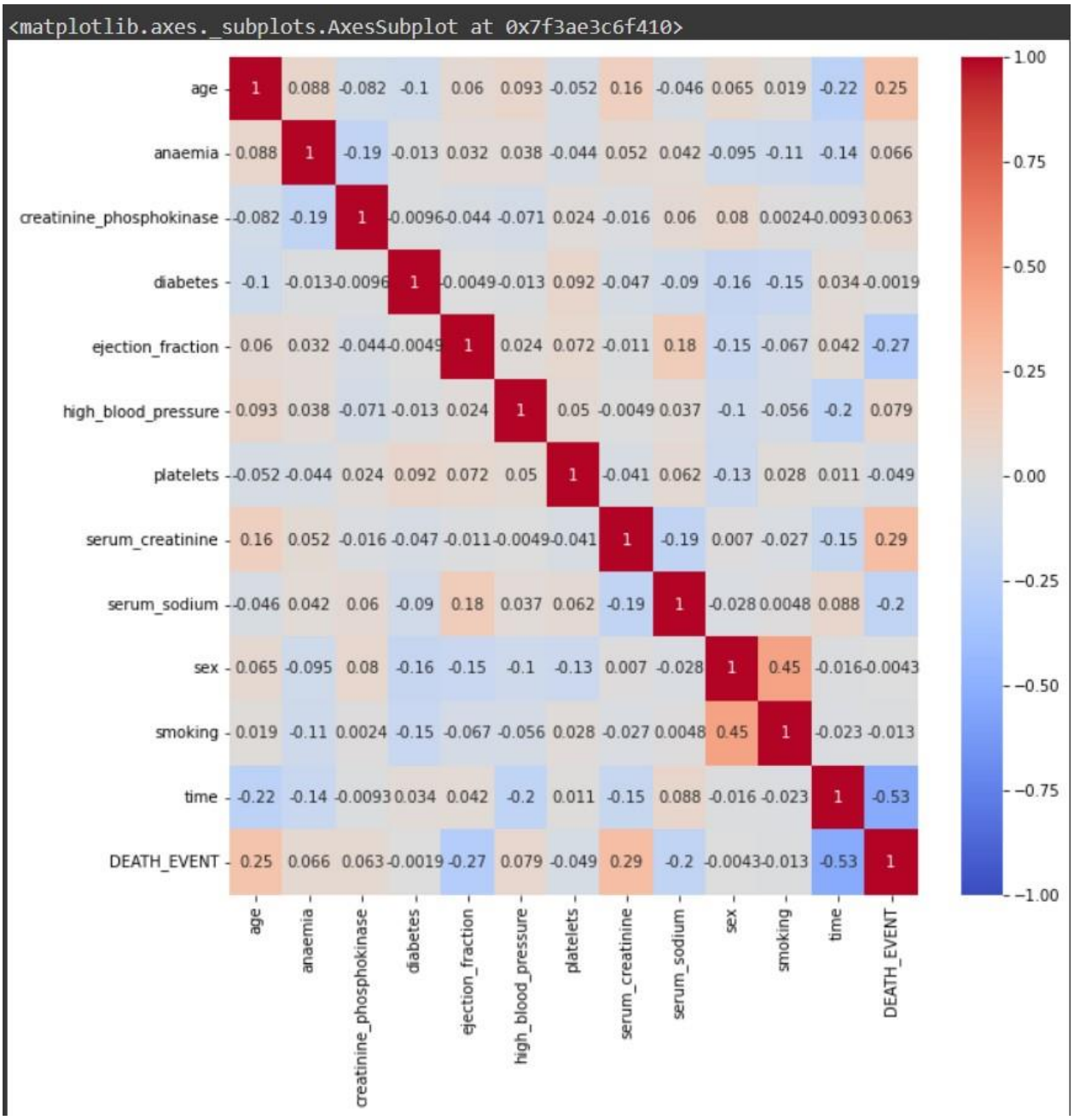

Figure 7.3: Sample output of matplotlib.axes

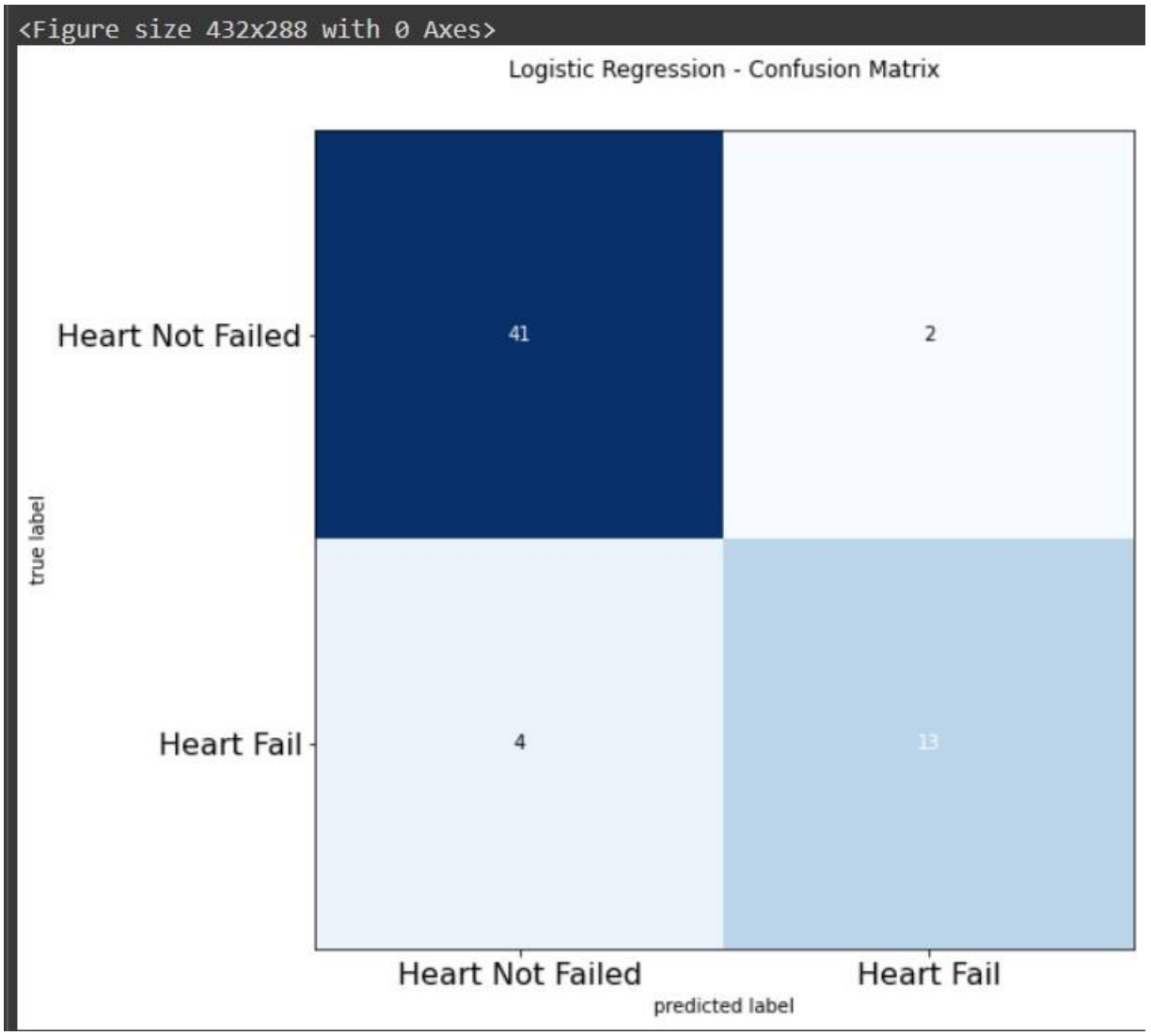

Figure 7.4: sample output of Heart disease predection

CHAPTER 8

CONCLUSION

In this work, a survey of Several Machine Learning techniques for predicting and detecting heart disease have been used., which generally is quite significant. Based on the above work, it is possible to conclude that machine learning algorithms have actually large potential for predicting and diagnosis cardiovascular illnesses or any heart-related diseases. With a large number of datasets, the Decision Tree method performs poorly. Logistic Regression scored extremely well as it addresses the issue of overfitting by combining numerous algorithms. (Many Decision Trees) The Nave Bayes classifier was highly quick and performed well in terms of computation. SVM improves efficiency in the vast majority of instances.

REFERENCE:

[1] “The Atlas of Heart Disease and Stroke”, [online]. http://www.who.int/cardiovascular_diseases/res ources/atlas/en/.

[2] Abhay Kishore1, Ajay Kumar2, Karan Singh3,Maninder Punia4, Yogita Hambir5,”Heart Attack Prediction Using Deep Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.

[3] J. S. Rumsfeld, K. E. Joynt, and T. M. Maddox, “Big data analytics to improve cardiovascular care: promise and challenges”, Nature Reviews Cardiology, Vol.13, No.6, pp.350, 2016.

[4] W. Dai, T. S. Brisimi, W. G. Adams, T. Mela, V. Saligrama, and I. C. Paschalidis, “Prediction of hospitalization due to heart diseases by supervised learning methods”, International Journal of Medical Informatics, Vol.84, No.3, pp.189–197, 2015.

[5] I. Kamkar, M. Akbarzadeh-T and M. Yaghoobi, "Intelligent water drops a new optimization algorithm for solving the Vehicle Routing Problem", In: Proc. of IEEE International Conference on Systems, Man and Cybernetics, pp.4142-4146, 2010.

[6] K. C. Tan, E. J. Teoh, Q. Yu, and K. C. Goh, "A hybrid evolutionary algorithm for attribute selection in data mining", Expert Systems with Applications, Vol.36, No.4, pp.8616-8630, 2009.

[7] K. Vembandasamy, R. Sasipriya, and E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm", IJISET-International Journal of Innovative Science, Engineering & Technology, Vol.2, pp.441-444, 2015.

[8] A. Malav, K. Kadam, and P. Kamat, "Prediction of heart disease using k-means and artificial neural network as a hybrid approach to improve accuracy", International Journal of Engineering and Technology, Vol.9, No.4, 2017.

APPENDIX

SOURCE CODE PY

```
import warnings
warnings.filterwarnings("ignore")
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.graph_objs as go
import plotly.express as px



labels=['No Diabetes','Diabetes']
diabetes_yes=heart_data[heart_data['diabetes']==1]
diabetes_no=heart_data[heart_data['diabetes']==0]
values=[len(diabetes_no),len(diabetes_yes)] fig =
go.fig(data=[go.pie(labels=labels,values=values,hole=.4)])
fig.update_layout(title_text="Analysis on Diabetes")

fig.show()

heart_data=pd.read_csv("heart_failure_clinical_records_dataset.csv")
heart_data.head()

heart_data.describe()

heart_data.isnull().sum()
fig=px.pie(heart_data,values='diabetes',names='DEATH_EVENT',title='Death
Analysis') fig.show plt.figure(figsize=(10,10))
sns.heatmap(heart_data.corr(),vmin=-1,cmap='coolwarm',annot=True) from
sklearn.model_selection

import train_test_split from sklearn.metrics

import confusion_matrix,accuracy_score
Feature=['time','ejection_fraction','serum_creatinine'] x=heart_data[Feature]
y=heart_data["DEATH_EVENT"]

xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_state=2) from
sklearn.linear_model import LogisticRegression log_re=LogisticRegression()
log_re.fit(xtrain,ytrain) log_re_pred=log_re.predict(xtest)
cm=confusion_matrix(ytest,log_re_pred) plt.figure()
plot_confusion_matrix(cm,figsize=(12,8),hide_ticks=True,cmap=plt.cm.Blues)
plt.title("Logistic Regression - Confusion Matrix") plt.xticks(range(2),["Heart Not
```

```
Failed","Heart Fail"],fontsize=16) plt.yticks(range(2),["Heart Not Failed","Heart
Fail"],fontsize=16) plt.show()
```