HEART DISEASE DETECTION USING MACHINE LEARNING ALGORITHM

Project report submitted in partial fulfillment of the requirement for the degree of
Bachelor of Technology
in
Computer Science and Engineering/Information Technology
By
Shivam Mittal(171276)

Under the supervision of
Mr.Praveen Modi
To

Department of Computer Science & Engineering and Information Technology
Jaypee University of Information Technology Waknaghat, Solan-173234,
Himachal Pradesh

DECLARATION

We hereby declare that the work reported in the B.Tech Project Report entitled **“<Heart Disease Detection Using Machine Learning Algortihm>”** submitted at **Jaypee University of Information Technology, Waknaghat, India** is an authentic record of our work carried out under the supervision of <**DR. Parveen Modi>.** We have not submitted this work elsewhere for any other degree or diploma.

29/6/21

<Signature>

Shivam Mittal

171276

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

18-06-2021

<Signature of the Supervisor>

Dr. Parveen Modi

Associate Professor

(Computer Science & Engineering and Information Technology)

Date:18-05-21

Head of the Department/Project Coordinator

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

&

INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the work in this Project title as "**Heart Disease Detection Using Machine Learning Algorithm**" is entirely written, successfully completed and demonstrated by the following students themselves as a fulfillment of requirement for Bachelor's of Engineeringin Computer Science.

Shivam Mittal(171276)

ACKNOWLEDGEMENTS

Thanks to Almighty GOD Who gave us courage and understanding to start and later on finish this work of project.Thanks to **HOD** Sir and our Supervisor **DR. PARVEEN MODI** who guided and helped us at the different phases of the project.

We would also like to thank my panel **Dr.Hari Singh** and **Mr. Ravindra Bhatt**, for giving me an opportunity to present my project and for judging my work and providing me feedback which would certainly help me in the future.

Last but not the least, we would like to thank my institution **Jaypee University of Information Technology** for giving me a platform to give me life and implementation, to the various fields I have studied till date.

ABSTRACT

Heart disease is one of the biggest cause of death in this era today. With Regular Physical work and decent eating habit can may be prevent it to some extent. In this paper we have tried to implement the important Machine Learning algorithm to predict the heart disease in a patent. In this paper we used five different algorithm of Machine Learning such as Logistic Regression, K-Nearest Neighbor, Naïve Bayes, Random Forest and decision tree but in which Logistic Regression gives the higher accuracy of approximately 91.80%.

TABLE OF CONTENTS

INTRODUCTION

9

1.1 INTRODUCTION

It is very difficult to predict the heart disease because of some risky factors such as diabetes, high blood pressure, high cholesterol and many other serous factors. Machine Learning is the most efficient technology in the today's era that provides systems the ability to automatically learn and improved from its own experiences.The heart disease is classified based on some methods such as- Logistic Regression, K-Nearest Neighbor (KNN), Naive Bayes (NB), Decision Tree [Abstract](DT) and Random Forest algorithm. After using these algorithm we find the accuracy mean of each algorithm that predicts which algorithm best measure the Heart disease efficiently. The data of heart disease patients collected from Kaggle.

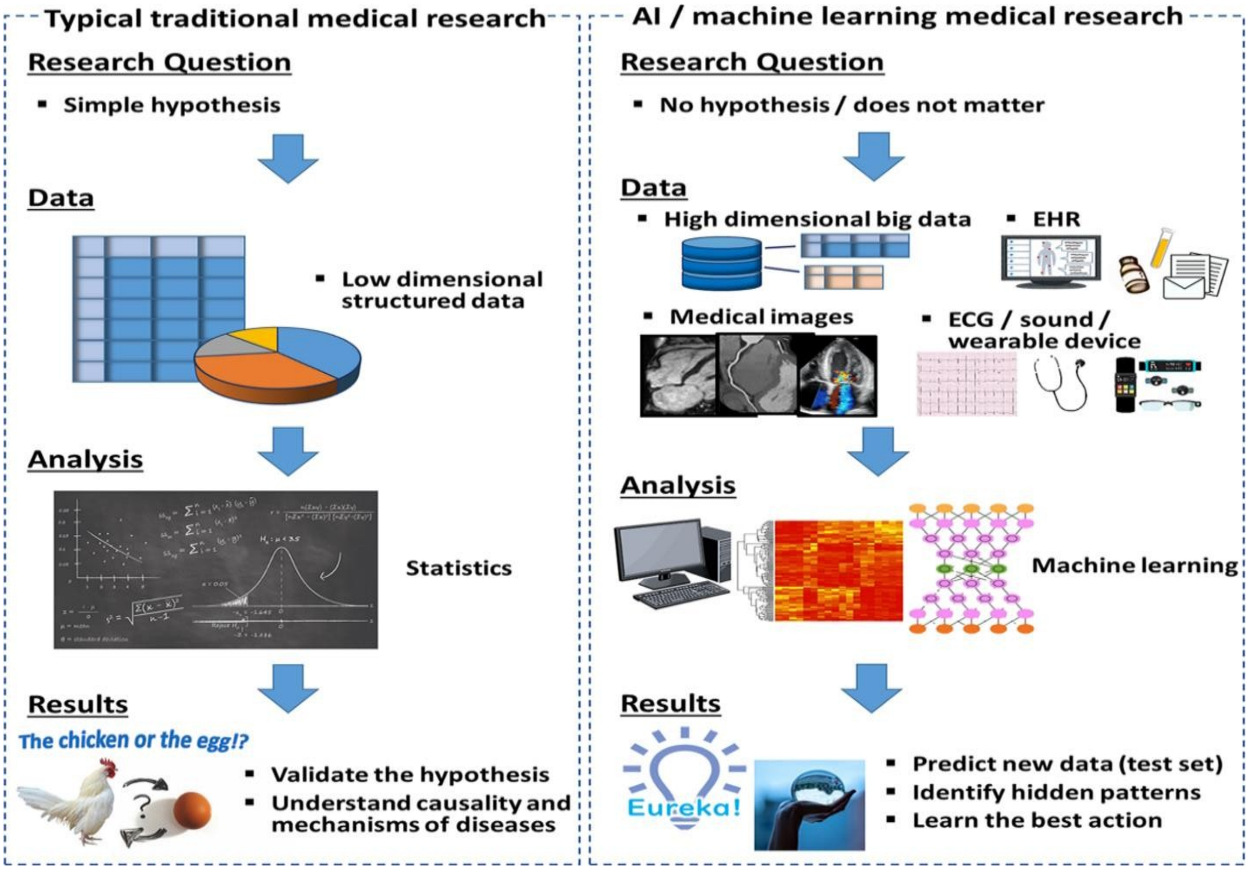

Typical traditional medical research
Research Question
Simple hypothesis
Data
Low dimensional structured data
Analysis
Statistics
Results
The chicken or the egg!?
Validate the hypothesis
Understand causality and mechanisms of diseases
AI / machine learning medical research
Research Question
No hypothesis / does not matter
Data
High dimensional big data
EHR
Medical images
ECG / sound / wearable device
Analysis
Machine learning
Results
Eureka!
Predict new data (test set)
Identify hidden patterns
Learn the best action

Fig: 1 (Research Overview)

1.2 OBJECTIVES

Advancement in the Medical Technologies.Detecting a person is suffering from any Heart disease or not?Its implementation as a desktop application

1.3 PROBLEM STATEMENT

Now understanding Heart disease so any kind of disorder in heart is refers as heart disease

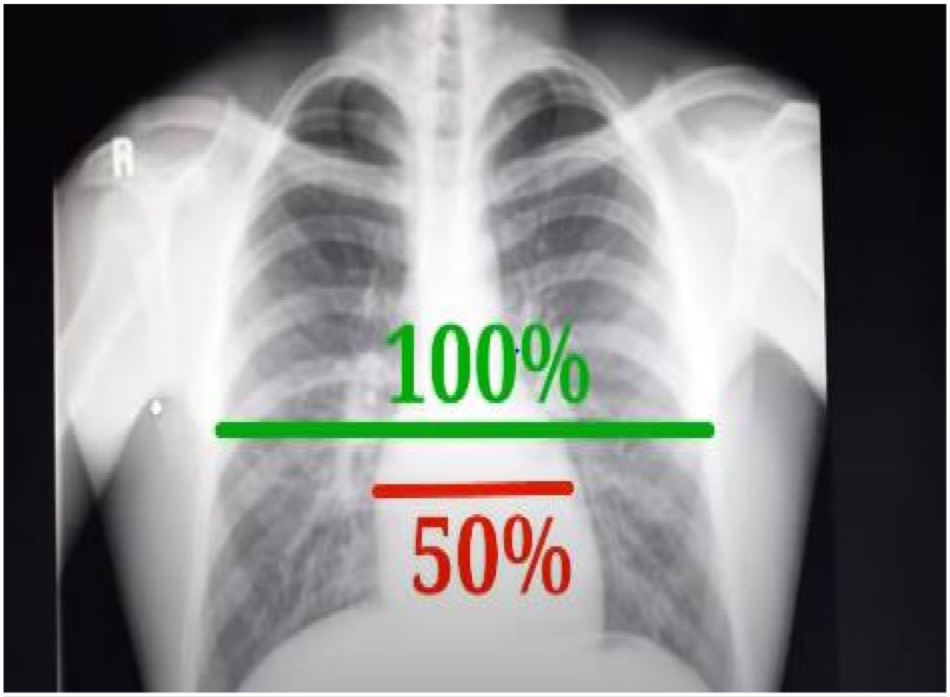


Fig: 2 (X-Ray Scan Image)

So if the shown area 50% increase while a little bit then there is a definitely chances of heart disease and the problem is to detect these heart disease so for this purpose we have a collected dataset which have records of 303 patients having 14 attributes so using these attributes and applying Machine learning model for detecting heart disease is our task.

1.4 METHODOLOGY

For this task as we defined earlier, we have record of 303 patients having 14 attributes so

- first task is testing and training for this task we use 80% of our data as training purpose and rest 20% is for testing purpose because this is an efficient way for study features.
- Now second task is data wrangling it means reduce the irrelevant components.

1.5 SCOPE

As the future of machine learning is very wide so scope of this project is very clear that this software is helpful for detecting a person is suffering from any kind of Heart disease or not ? and according to the latest study in the area of medical domain related to machine learning which is related to present scenario of COVID-19 epidemic Dr. Kamlesh Jain Professor in the department of Civil Engineering at Indian Institute of Technology, Roorkee develops a software that detects COVID-19 so in future we can also apply the similar task for Heart disease prediction.

LITERATURE REVIEW

2.1 HISTORY

Many researchers work in this field Dr. john moraque first develops the first machine learning model but after this many researchers works in that field and contributes in the medical domain.

2.2 ANALYSIS

Today medical services have come a long way to treat patients with various diseases. Among the most lethal one is the heart disease problem which cannot be seen with a naked eye and comes instantly when its limitations are reached. Today diagnosing patients correctly and administering effective treatments have become quite a challenge.

2.3 RECENT STUDY

The appreciable work of Dr. Deepak RanjanNayak(Currently working as an Assistant Professor in the Department of Computer Science & Engineering at SVNIT Surat) in the field of medical domain pathology and gives different-different ideas related to prediction of Medical disease through Computer vision and Machine learning Model.

2.4 APPLICATION

2.4.1 From Medical Perspective

SYSTEM DEVELOPMENT

3.1 OVERVIEW OF METHOD

Machine Learning process starts from preprocessing data phase followed by feature selection, classification.

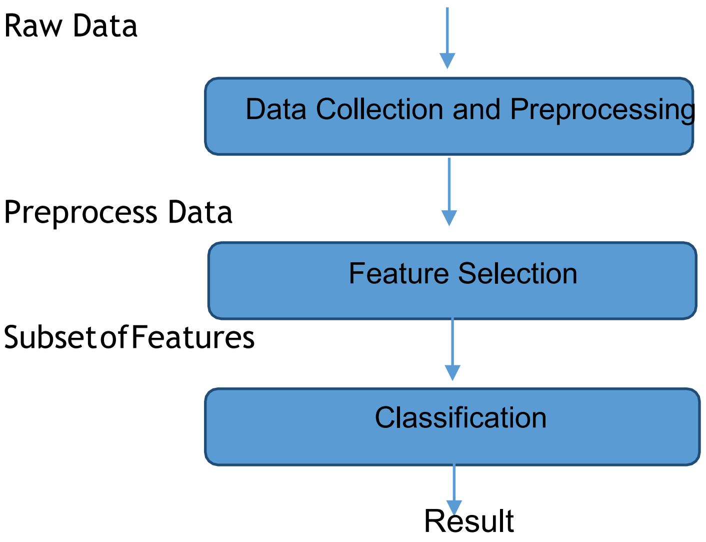


Fig: 3 (Overview Of Method)

3.2 DATA COLLECTION

Data is pre-processed after collection of various records. This dataset contains the record of 303 patients. The patient having heart disease or not is set to 1 or 0. There are 14 columns in our dataset which is described as:

ATTRIBUTE	DESCRIPTION
Age	Displays the age of the individual
Sex	Display the sex (0=female and 1=male)
Cp	Chest pain type (0=typical angina, 1=atypical angina, 2= non-angina, 3= asymptotic)
Trestbps	Level of blood pressure at resting mode (in mm/Hg at time of admitting in hospital)
Chol	Serum cholesterol (in mg/dl)
Fbs	Fasting blood sugar of individual with 120mg/dl (if fbs>120mg/dl then: 1(true) else: 0(false)
Restecg	Resting ECG (0=normal, 1=having ST-T wave abnormality,2=left ventricular hypertrophy
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina (1=yes and 0=no)
Oldpeak	ST depression induced by exercise relative to rest
Slope	Peak exercise ST segment (0=upsloping, 1=flat, 2= downsloping)
Ca	Fluoroscopy colored major vessels numbered from 0 to 4
Thal	Thalassemia (1=normal, 2=fixed defect, 3=reversible defect)
Target	Diagnosis of heart disease (0=absence, 1=present)

3.3 FEATURE SELECTION

among the 14 attributes there may be some columns which may be irrelevant can be dropped but for this there is need to train the data and the rest coming attributes is considered as an important factors for predicting the Heart disease.

Feature selection means after dividing our data into training and testing (here we reserve data for training for 80% and testing for 20% because this is an efficient way) then apply the tools of machine learning

The correlation between these attributes as shown below:

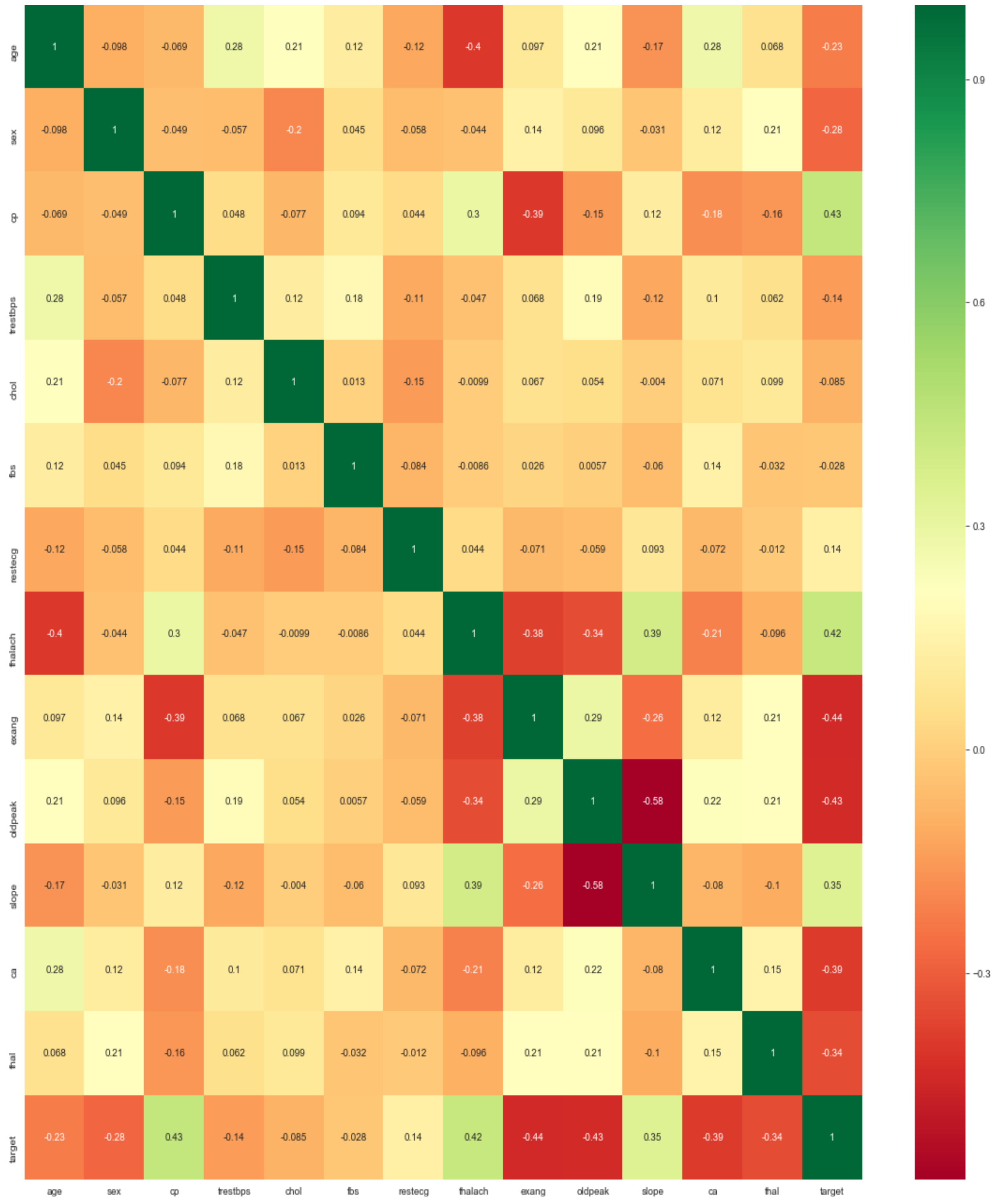
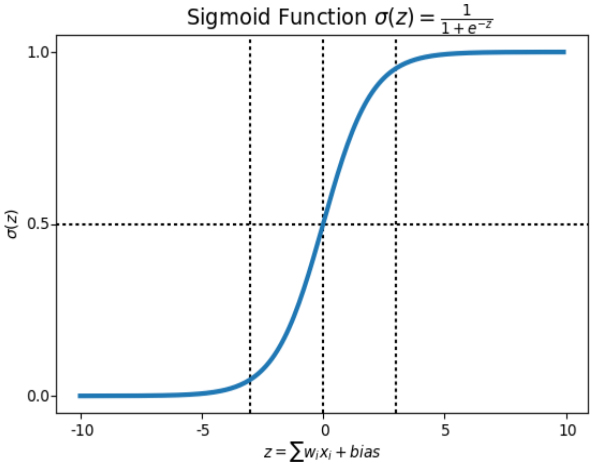


Fig: 4 (Correlation Matrix of Data)

3.4 CLASSIFICATION

Apply the different-different Machine Learning algorithm and Measurethe best one.

i) **LOGISTIC REGRESSION:** It is used for classification problem.It is a predictive analysis algorithm and based on the concept of probability. The hypothesis of logistic regression it to limit the cost function between and 1 and 0. In machine learning, we use sigmoid function to map predictions to probabilities which is represented as:



Graph: 1 (Logistic Regression)

It will first Analyzing the data followed by Collecting Data, Data Wrangling, Train & Test, Accuracy Check.

ii) **K-Nearest Neighbor:**It extract the knowledge based on the samples Euclidean distance function d(Xi,Yi) and the majority
of K-nearest neighbors:-
d(Xi,Yi)=[(X1-Y1)^2+(X2-Y2)^2+ ------------- +(Xi-Yi)^2]^0.5

iii) **NAIVE BAYES:** It is a probabilistic classifier. It is based on probability models that incorporate assumptions. It assumes that the value of a particular feature is independent of the value of any other feature, given the class variable.

iv) **DECISION TREE:** For training samples of data D, the trees are constructed based on high entropy inputs. These trees are simple and fast constructed in a top down recursive divide and conquer approach. Tree pruning is performed to remove the irrelevant samples on D.

v) **RANDOM FOREST:** In this technique, a set of decision trees are grown and each tree votes for the most popular class, then the votes of different trees are integrated and a class is predicted for each sample. This approach is designed to increase the accuracy of the decision tree, more trees are produced to vote for class prediction. This approach is ensemble classifier composed of some decision trees and the final result is the mean of individual trees results.

PERFORMANCE ANALYSIS

4.1 **System Requirement:**Python Editor as an essential such as Spyder (Anaconda). You can also used inPycharm,Python IDE 3.7.4 or many more editor for python development and you can also build this project using Matlab (Octave) depends on Software Developer but for this project we use Spyder (Anaconda) platform.

4.1.1 Minimum Requirements:

a. 200 MHz Processor

b. 64 MB of RAM

4.1.2 Best Requirements:

a. 1.6 GHz processor

b. 128 MB or more of RAM

4.2 Library Requirement:

4.2.1 NumPy :

NumPy stands for Numeric Python. It is used in adding support for large multi dimensional arrays and matrices along with large collection of high level mathematical functions to operate on arrays.

4.2.2 Pandas:

It is a software library written for python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.

4.2.3 Matplotlib:

It is a plotting library in python. It provides an object-orientedAPI for embedding plots into applications using general purpose GUI toolkits like Tkinter, rcParams, rainbow etc.

4.2.4 skLearn:

skLearn stands for skikit-learn. It is a open source library that has powerful tools for data analysis and data mining. It supports Python numerical and scientific libraries like NumPy and SciPy.

4.2.5 Seaborn:

It is a Python data visualization library based on matplotlib. It provides a high level interface for drawing attractive and informative statistical graphics.

4.3 System Development Approach :

Use Case Diagrams

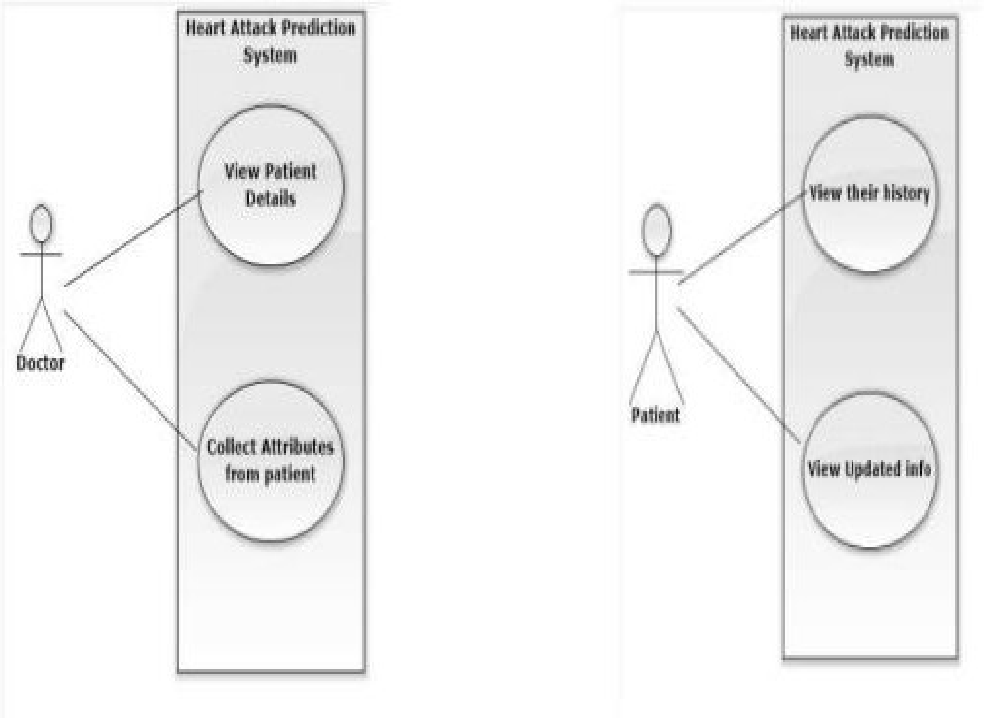
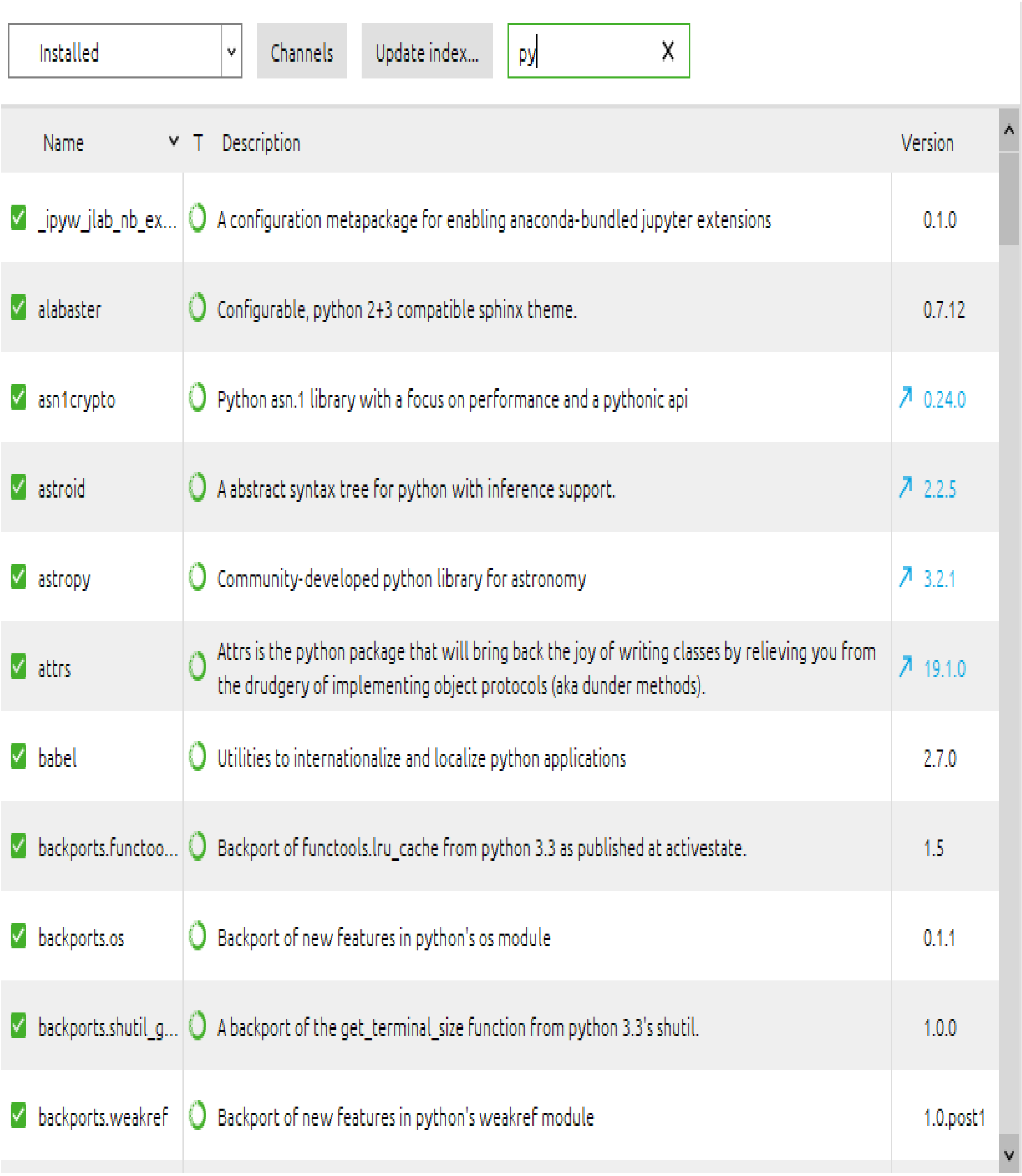

Heart Attack Prediction System
View Patient Details
Collect Attributes from patient
Doctor
Heart Attack Prediction System
View their history
View Updated info
Patient

Fig: 5 (Use Case Diagram)

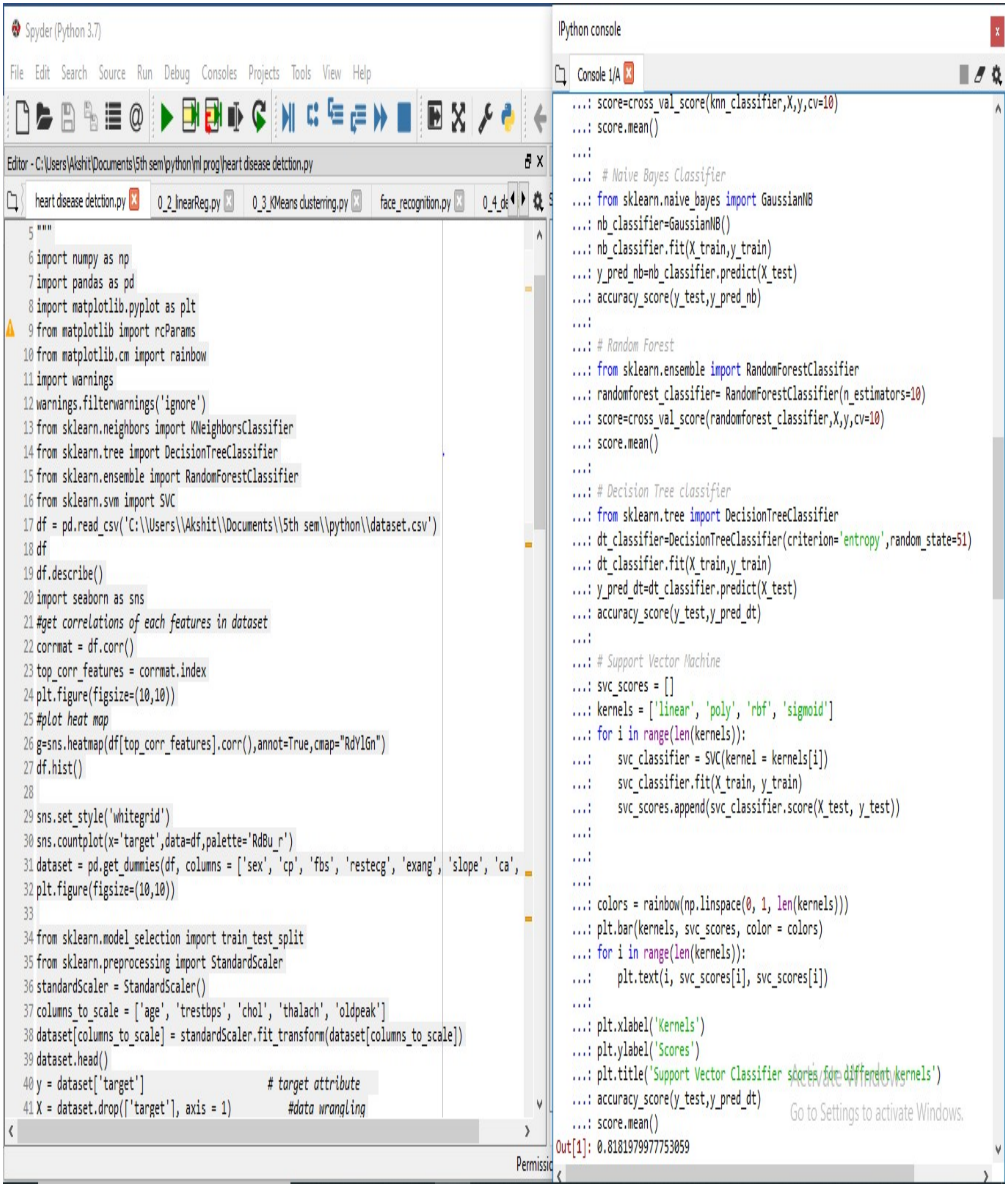
4.4 Installation of Packages:



	Name	T	Description	Version
☑	_ipyw_jlab_nb_ex...	○	A configuration metapackage for enabling anaconda-bundled jupyter extensions	0.1.0
☑	alabaster	○	Configurable, python 2+3 compatible sphinx theme.	0.7.12
☑	asn1crypto	○	Python asn.1 library with a focus on performance and a pythonic api	↗ 0.24.0
☑	astroid	○	A abstract syntax tree for python with inference support.	↗ 2.2.5
☑	astropy	○	Community-developed python library for astronomy	↗ 3.2.1
☑	attrs	○	Attrs is the python package that will bring back the joy of writing classes by relieving you from the drudgery of implementing object protocols (aka dunder methods).	↗ 19.1.0
☑	babel	○	Utilities to internationalize and localize python applications	2.7.0
☑	backports.functoo...	○	Backport of functools.lru_cache from python 3.3 as published at activestate.	1.5
☑	backports.os	○	Backport of new features in python's os module	0.1.1
☑	backports.shutil_g...	○	A backport of the get_terminal_size function from python 3.3's shutil.	1.0.0
☑	backports.weakref	○	Backport of new features in python's weakref module	1.0.post1

Fig: 6 (Installation of Packages)

4.5 Analysis:


Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\Akshit\Documents\5th sem\python\ml prog\heart disease detction.py
heart disease detction.py
0_2_linearReg.py
0_3_KMeans clusterring.py
face_recognition.py
0_4_de
Permissic
IPython console
Console 1/A

```
"""
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
import warnings
warnings.filterwarnings('ignore')
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
df = pd.read_csv('C:\\Users\\Akshit\\Documents\\5th sem\\python\\dataset.csv')
df
df.describe()
import seaborn as sns
#get correlations of each features in dataset
corrmat = df.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(10,10))
#plot heat map
g=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap="RdYlGn")
df.hist()

sns.set_style('whitegrid')
sns.countplot(x='target',data=df,palette='RdBu_r')
dataset = pd.get_dummies(df, columns = ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope', 'ca',
plt.figure(figsize=(10,10))

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
standardScaler = StandardScaler()
columns_to_scale = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
dataset[columns_to_scale] = standardScaler.fit_transform(dataset[columns_to_scale])
dataset.head()
y = dataset['target']                        # target attribute
X = dataset.drop(['target'], axis = 1)           #data wrangling
```

```
   ...: score=cross_val_score(knn_classifier,X,y,cv=10)
   ...: score.mean()
   ...:
   ...:  # Naive Bayes Classifier
   ...: from sklearn.naive_bayes import GaussianNB
   ...: nb_classifier=GaussianNB()
   ...: nb_classifier.fit(X_train,y_train)
   ...: y_pred_nb=nb_classifier.predict(X_test)
   ...: accuracy_score(y_test,y_pred_nb)
   ...:
   ...: # Random Forest
   ...: from sklearn.ensemble import RandomForestClassifier
   ...: randomforest_classifier= RandomForestClassifier(n_estimators=10)
   ...: score=cross_val_score(randomforest_classifier,X,y,cv=10)
   ...: score.mean()
   ...:
   ...: # Decision Tree classifier
   ...: from sklearn.tree import DecisionTreeClassifier
   ...: dt_classifier=DecisionTreeClassifier(criterion='entropy',random_state=51)
   ...: dt_classifier.fit(X_train,y_train)
   ...: y_pred_dt=dt_classifier.predict(X_test)
   ...: accuracy_score(y_test,y_pred_dt)
   ...:
   ...: # Support Vector Machine
   ...: svc_scores = []
   ...: kernels = ['linear', 'poly', 'rbf', 'sigmoid']
   ...: for i in range(len(kernels)):
   ...:     svc_classifier = SVC(kernel = kernels[i])
   ...:     svc_classifier.fit(X_train, y_train)
   ...:     svc_scores.append(svc_classifier.score(X_test, y_test))
   ...:
   ...:
   ...:
   ...: colors = rainbow(np.linspace(0, 1, len(kernels)))
   ...: plt.bar(kernels, svc_scores, color = colors)
   ...: for i in range(len(kernels)):
   ...:     plt.text(i, svc_scores[i], svc_scores[i])
   ...:
   ...: plt.xlabel('Kernels')
   ...: plt.ylabel('Scores')
   ...: plt.title('Support Vector Classifier scores for different kernels')
   ...: accuracy_score(y_test,y_pred_dt)
   ...: score.mean()
Out[1]: 0.8181979977753059
```

Fig: 7 (Result Analysis)

CONCLUSION

5.1 Advantages:

□User can search for doctor’s help at any point.

□User can talk about their heart disease and get instant Diagnosis

□It helps to make the treatment process fast and time consuming

5.2 Disadvantages:

□The system is not fully automated, it needs data from user for fully diagonis

5.3 Future Enhancement:

The Future of this project is very wide and at enhancement level we can also apply deep learning Neural Network and relate this concept with Image detection means through this model we can passes the images of our X-Ray scan which can detects the level of heart size if it increase or either decrease then it will definitely chances of heart disease and the related work in this field is also done by various eminent Professor of different NIT as well as IIT. 1. Dr. Deepak RanjanNayak (Currently working as an Assistant Professor in SVNIT Surat)2. Dr. Kamal Jain (Recently submit a report on COVID-19 Detection Software)

So, considering the appreciable work of above professor we can also apply the different model of neural networks.

5.4 Result:

Models	**Accuracy**
Logistic Regression	0.9180327868852459
K-Nearest Neighbor	0.8506637004078605
Naïve Bayes	0.8360655737704918
Decision Tree	0.8032786885245902
Random Forest	0.8049721913236929

Tab: 2 (Accuracy Result of Models) Since the logistic regression gives higher accuracy in this case so logistic regression algorithm in this dataset best depicts the person is suffering from heart disease or not basically there is an another way i.e cross validation through which we can also shows which algorithm is best.

5.5Conclusion:

Knowing the heart information will help in long term saving of human lives and early detection of abnormalities in heart conditions. Heart disease prediction is very important in the sector of medical. The Heart diseases can be easily controlled if it is detected at early stage.

REFERENCES

1. J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, 2016, pp. 1-5. doi: 10.1109/ICCPCT.2016.7530265

2. M. Gandhi and S. N. Singh, ''Predictions in heart disease using techniques of data mining,'' in Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manage. (ABLAZE), Feb. 2015, pp. 520–525

3. . M. Sultana, A. Haider and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, 2016, pp. 1-5. doi: 10.1109/CEEICT.2016.7873142

4. Praveen Kumar Reddy M, T Sunil Kumar Reddy, S.Balakrishnan, Syed MuzamilBasha Ravi Kumar Poluru, '' Heart Disease Prediction Using Machine Learning Algorithm,'' International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019

5. M. S. Amin, Y. K. Chiam, K. D. Varathan, ''Identification of significant features and data mining techniques in predicting heart disease,'' Telematics Inform., vol. 36, pp. 82–93, Mar. 2019. [Online]. Available: https:// linkinghub.elsevier.com/retrieve/pii/S0736585318308876

6. C. Sowmiya and P. Sumitra, ''Analytical study of heart disease diagnosis using classification techniques,'' in Proc. IEEE Int. Conf. Intell. Technology. Control, Optim. Signal Process. (INCOS), Mar. 2017, pp. 1–5

7. https://archive.ics.uci.edu/ml/dataset/Heart+Disease
8. G.Subbalakshmi et al. / Indian Journal of Computer Science and Engineering (IJCSE).Decision Support in Heart Disease Prediction System using naïve bayes.

9. International Journal of Computer Applications (0975 –888) Volume 47–No.10, June 2012 / Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques
10.https://www.google.com/url?sa=i&url=https%3A%2F%2Fpdfs.semanticscholar.org%2F2788

%2F86c15ffdfbf324ab17137d7df897183a9f79.pdf&psig=AOvVaw1oliTcA0AFrmgmAl78NPPs&ust=1589656166488000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCLDd_uzItukCFQAAAAAdAAAAABAD

11. Volume 4, Issue 1, January 2014 ISSN: 2277 128X International Journal of AdvancedResearch in Computer Science and Software Engineering /Study of Heart Disease Prediction using Data Mining

12. http://www.ijsrd.com/articles/IJSRDV8I10457.pdf

13. SH Wang, XX Zhou, JQ Yang, **DR Nayak**, C. Reyes, ZC Dong, R. Watson, C. Cattani, J. Young, W. Gray, LX Han, Y. Yao, LT Fang, P. Phillips, SD Du, YD Zhang, “ANTIDOTE: A Wechat Public Platform Provide Urgent Treatment for Patients with Tetrodotoxin Intoxication," ***Basic & Clinical Pharmacology & Toxicology, Wiley***, vol. 121, 2017

CODE

```
import numpy as npimport pandas as pdimport matplotlib.pyplot as pltfrom matplotlib import
rcParamsfrom matplotlib.cm import rainbowimport
warningswarnings.filterwarnings('ignore')from sklearn.neighbors import
KNeighborsClassifierfrom sklearn.tree import DecisionTreeClassifierfrom sklearn.ensemble
import RandomForestClassifierdf = pd.read_csv('C:\\Users\\Akshit\\Documents\\5th
sem\\python\\dataset.csv')dfdf.describe()
```

```
import seaborn as sns#get correlations of each features in datasetcorrmat =
df.corr()top_corr_features = corrmat.indexplt.figure(figsize=(20,20))#plot heat
mapg=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap="RdYlGn")df.hist()sns.set_st
yle('whitegrid')sns.countplot(x='target',data=df,palette='RdBu_r')dataset = pd.get_dummies(df,
columns = ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope', 'ca', 'thal'])plt.figure(figsize=(20,20))
```

```
from sklearn.model_selection import train_test_splitfrom sklearn.preprocessing import
StandardScalerstandardScaler = StandardScaler()columns_to_scale = ['age', 'trestbps', 'chol',
'thalach', 'oldpeak']dataset[columns_to_scale] =
standardScaler.fit_transform(dataset[columns_to_scale])dataset.head()y = dataset['target']X =
dataset.drop(['target'], axis = 1)          #data wrangling
```

```
from sklearn.model_selection import cross_val_scoreknn_scores = []for k in
range(1,21):knn_classifier = KNeighborsClassifier(n_neighbors =
k)score=cross_val_score(knn_classifier,X,y,cv=10)knn_scores.append(score.mean())plt.plot([k
for k in range(1, 21)], knn_scores, color = 'red')for i in range(1,21):plt.text(i, knn_scores[i-1], (i,
knn_scores[i-1]))plt.xticks([i for i in range(1, 21)])plt.xlabel('Number of Neighbors
(K)')plt.ylabel('Scores')plt.title('K Neighbors Classifier scores for different K
values')knn_classifier = KNeighborsClassifier(n_neighbors =
12)score=cross_val_score(knn_classifier,X,y,cv=10)score.mean()
```

```
from sklearn.ensemble import RandomForestClassifierrandomforest_classifier= RandomForestClassifier(n_estimators=10)score=cross_val_score(randomforest_classifier,X,y,cv=10)score.mean()
```

```
from sklearn.model_selection import train_test_splitX_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=5) X_trainy_trainy_test

# feature scaling

from sklearn.preprocessing import StandardScalersc=StandardScaler()X_tarin_sc=sc.fit_transform(X_train)X_test_sc=sc.transform(X_test)from sklearn.metrics import accuracy_score

###logistic regression

from sklearn.linear_model import LogisticRegressionlr_classifier=LogisticRegression(random_state=51,penalty='l1')lr_classifier.fit(X_train,y_train)y_pred_lr=lr_classifier.predict(X_test)accuracy_score(y_test,y_pred_lr)

# Naive Bayes Classifier

from sklearn.naive_bayes import GaussianNBnb_classifier=GaussianNB()nb_classifier.fit(X_train,y_train)y_pred_nb=nb_classifier.predict(X_test)accuracy_score(y_test,y_pred_nb)

# Decision Tree classifier

from sklearn.tree import DecisionTreeClassifierdt_classifier=DecisionTreeClassifier(criterion='entropy',random_state=51)dt_classifier.fit(X_train,y_train)y_pred_dt=dt_classifier.predict(X_test)accuracy_score(y_test,y_pred_dt)
```

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 29/06/2021

Type of Document : **B.Tech Project -YES**

Name: Shivam Mittal **Department: Computer Sci.** **Enrolment No 171276**

Contact No. 9468432555 **E-mail.shivammittal888@mail.com**

Name of the Supervisor: **Dr. Parveen Modi**

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): HEART DIESEASE DETECTION USING MACHINE LEARNING ALGORITHM

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages =28
- Total No. of Preliminary pages =6
- Total No. of pages accommodate bibliography/references =25

29/6/21

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at 16 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

Pravin

18-06-2021

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	**Excluded**	**Similarity Index (%)**	**Generated Plagiarism Report Details** (Title, Abstract & Chapters)	
	• All Preliminary Pages • Bibliography/Im a ges/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature **Librarian**

....

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com

ORIGINALITY REPORT

16%	19%	9%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Asian Institute of Technology Student Paper	7%
2	towardsdatascience.com Internet Source	6%
3	Submitted to Jaypee University of Information Technology Student Paper	5%
4	github.com Internet Source	3%
5	Satish Kumar Satti, K. Suganya Devi, Prasenjit Dhar, P. Srinivasan. "A machine learning approach for detecting and tracking road boundary lanes", ICT Express, 2020 Publication	1%
6	www.thushv.com Internet Source	1%
7	www.coursehero.com Internet Source	1%