TABLE OF CONTENT

CHAPTER1

1. INTRODUCTION

1.1 Project Overview

In this project , we will be keenly progressing with the heart disease prediction and for that, we will be looking into the dataset of the heart disease. From that dataset we will derive various insights that help us know the weightage of each feature and how they are interrelated to each other.our sole aim is to predict heartdiseases whether the person affected by heart diseass or not that will be affected by a savior heart problem or not and also visualize it using a dashboard.

Cardiovascular diseases (CVDs) or heart disease are the number one cause of death globally with 17.9 million death cases each year. CVDs are concertedly contributed by hypertension, diabetes, overweight and unhealthy lifestyles. You can read more on the heart disease statistics and causes for self-understanding. This project covers manual exploratory data analysis and using pandas profiling in Jupyter Notebook, on Google Colab.

Today, heart failure diseases affect more people worldwide than other autoimmune conditions. Cardiovascular Diseases (CVDs) affect the heart and obstruct blood flow through the blood vessels. Chronic ailments in CVD include heart disease (heart attack), cerebrovascular diseases (strokes), congestive heart failure, and many more pathologies. Worldwide, CVDs kill around 17 million a year, and death rates due to heart diseases have increased after the COVID-19 pandemic.

The datasets which we use contain huge data and sometimes we cannot view that data even in 3D, which is also called the curse of

dimensionality. So, when we perform operations on this data, we require a huge amount of memory, and sometimes the data can also grow exponentially and overfitting can happen.for visualizing the data we are using IBM cognos and IBM cloud for storage.

1.2 Purpose

The major purpose of visualising and predicting heart disease is to calculate the health check of affected ones by gathering reports and having a better research papers to predict the ones who's affected by heart disease and diagnose the further measures for the patient to recover from the disease quickly.

Aiding the best with appropriate treatment occurs only if we visualise and predict the heart disease properly. The EHDPS predicts the likelihood of patients getting heart disease.

The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time. The correct prediction helps to give them the proper treatment with quicker recovery.

CHAPTER2

2. LITERATURE SURVEY

2.1 Existing problem

Methaila et. al predicted heart disease using data mining Techniques. The main Methodology used for prediction is KNN Algorithms, Decision Trees like CART, C4.5, CHAID, J48, ID3 Algorithms, and Naive Bayes Techniques. This system uses 13 medical attributes as input and with that input, Data sets it to process the data mining techniques and shows the most accurate one.

The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise.

Disease prediction system provides only possible outcomes it does not guarantee that it will predict the disease correctly. But it has significantly higher accuracy for predicting possible diseases.

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.

Some risk factors for heart disease cannot be controlled, such as your age or family history. But you can take steps to lower your risk by changing

factors you can control.

Existing solutions are found to have,

1. 1.inefficient and inaccurate heart disease prediction system
2. No proper assistance of medical professionals in evaluating a patient's heart diseases based on the clinical data of the patient.

2.2 References

1. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. Int J Comput Sci Net Secur. 2008;8:343–350. [Google Scholar]
2. Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach. Int J Adv Sci Eng Technol. 2014;2:88–92. [Google Scholar]
3. Gudadhe M, Wankhade K, Dongre S. Decision support system for heart disease based on support vector machine and Artificial Neural Network. Computer and Communication Technology (ICCCT), 2010 International Conference on; 2010. pp. 741–745. [Google Scholar]
4. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating error. Nature. 1986;323:533–536. [Google Scholar]
5. Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.
6. Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013).
7. Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.
8. Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.

9. Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44-8.

2.3 Problem Statement Definition

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 our 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 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 healthcare environment is still „information rich“ but „knowledge poor“. There is 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.

Day by day the cases of heart diseases are increasing at a rapid rate and its very important and concerning to predict any such diseases beforhand. this diadnosis is a difficult task.it should be performed preisely and efficienty.

CHAPTER3

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

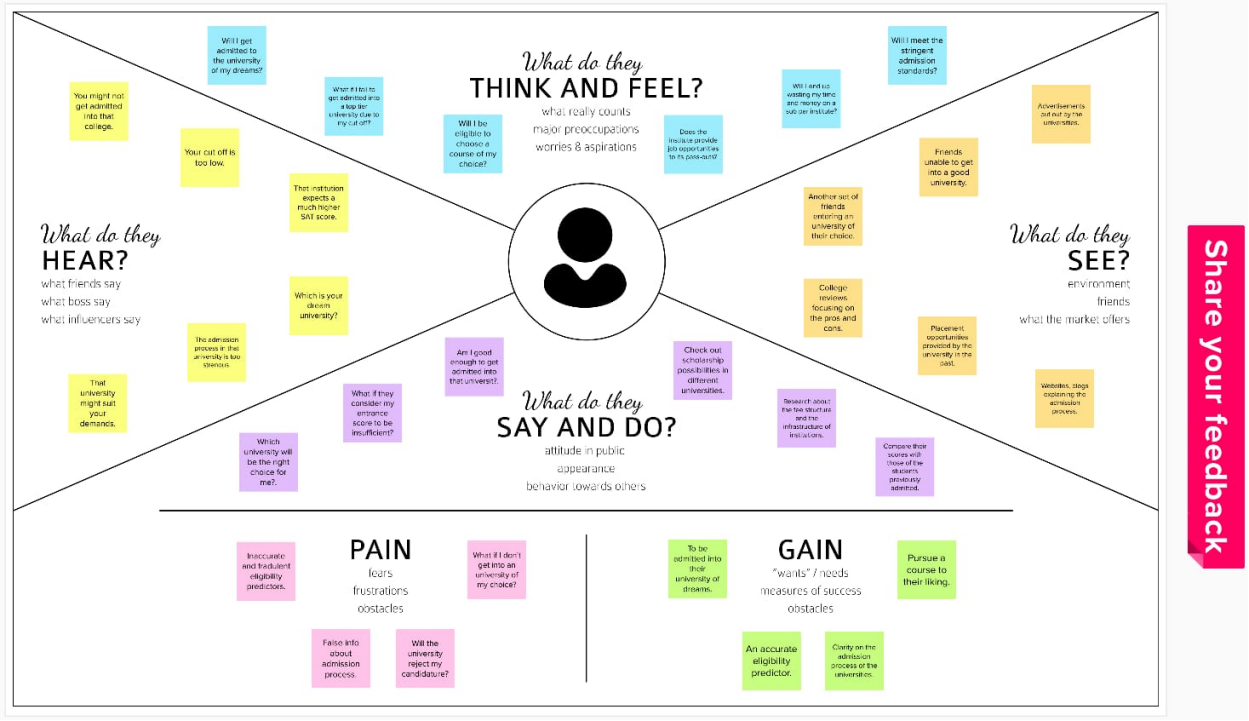
Edit this template
Right-click to unlock

Empathy Map Canvas

Gain insight and understanding on solving customer problems.

1

Build empathy and keep your focus on the user by putting yourself in their shoes.



Share your feedback

3.2 Ideation & Brainstorming

Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that ideation is

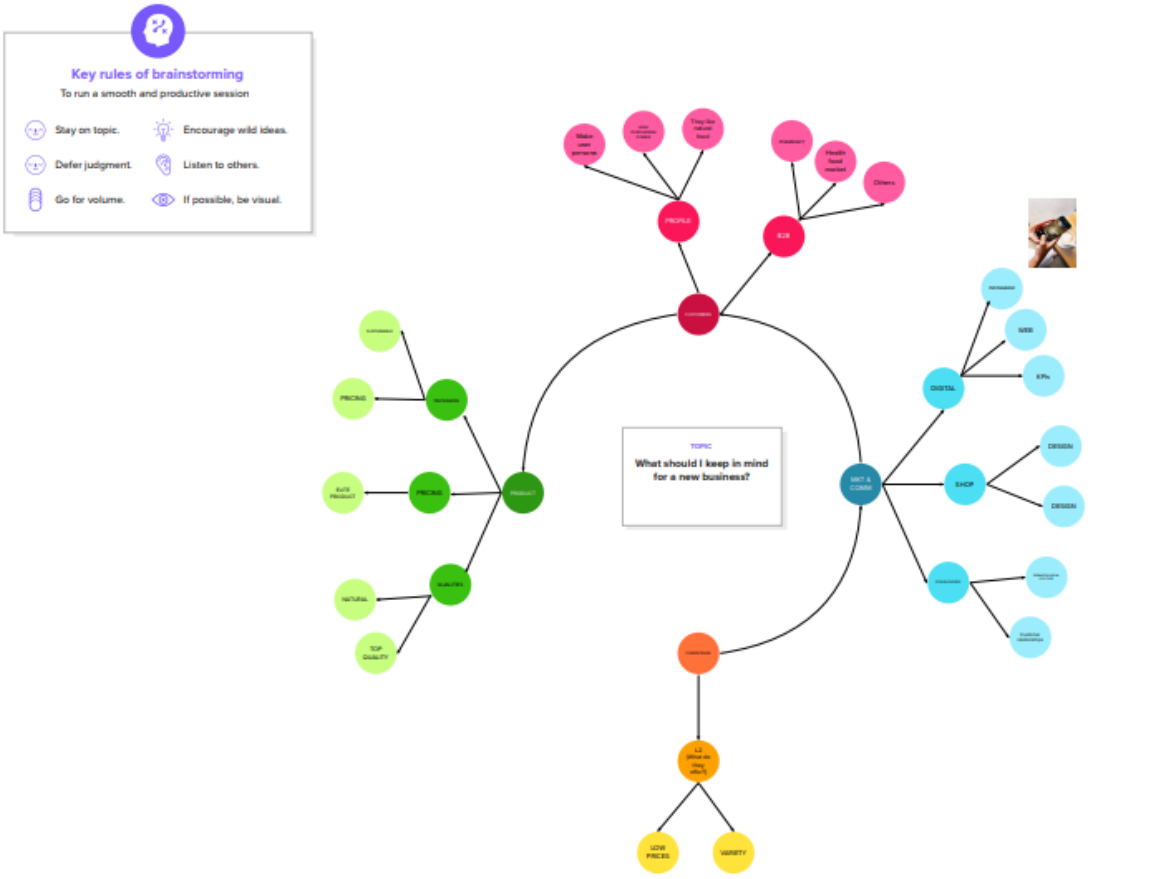
commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity.

Brainstorming is usually conducted by getting a group of people together to come up with either general new ideas or ideas for solving a specific problem or dealing with a specific situation.
For example, a major corporation that recently learned it is the object of a major lawsuit may want to gather together top executives for a brainstorming session on how to publicly respond to the lawsuit being filed. Participants in a brainstorming session are encouraged to freely toss out whatever ideas may occur to them. The thinking is that by generating a large number of ideas, the brainstorming group is likely to come up with a suitable solution for whatever issue they are addressing.

The lines between ideation and brainstorming have become a bit more blurred with the development of several brainstorming software programs, such as Brightidea and Ideawake. These software programs are designed to encourage employees of companies to generate new ideas for improving the companies' operations and, ultimately, bottom-line profitability.
The programs often combine the processes of ideation and brainstorming in that individual employees can use them, but companies may simulate brainstorming sessions by having several employees all utilize the software to generate new ideas intended to address a specific purpose.

INNOVATIX
CONSULTING

Brainstorming



3.3 Proposed Solution

3.3.1 Problem statement

Heart diseases Refers to several types of abnormalities in heart conditions.The leading cause of death is heart disease.It is Infeasible tests for ECG and so on.Hence,there needs a replacement for this,which must be handy and reliable

3.3.2 idea / solution description

The Idea Behind the proposed solution is to propose an interactive dashboard for visualizing and predicting heart Diseases in which user view his/her medical Report analysis and the predicted final result. The Dashboard will be generating using IBm cognos. the heart disease will be predicted Using naive bayes Algorithm.

3.3.3 Novely / Uniqueness

The novelty behind the prosed system is to provide suggestions to the user based on the preventive measures to take care of the user himself.

3.3.4 Social Impact /customer Satisfaction

The system helps the user as well as the doctor to make better decisions to predict heart disease.It is useful in predicting the disease in an eariler stage and makes the user alert about his current condition periodically.

3.3.5 Business Model

The interactive dashboard for the heart disease prediction can be deployed in health care centres and hospitals,so that It makes the analysis in a fast manner.

3.3.6 Scalability of the solution

The proposed solution will work efficiently in both smaller and larger datasets in a similar manner.In future,it can be changed to predict some other diseases with more accuracy

3.4 Problem solution fit

The problem solution-Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem.It helps entrepreneurs,Marjeters and corporate innovators identify behavioral patterns and recognize what would worka and why.

purpose:

1. solve complex problems in a way that fits the state of your customers.
2. succeed faster and increase your solution adoption by tapping into existing mediums and channels of behavior.

3. sharpen your communicstion and marketing strategy with the right triggers and messaging.

4. Increase touch points with your company by finding the right problem behavior fit and building trust by solving frequent annoyances or urgent or costly problems.

5. Understand the existing situation in order to improve it for your target group

CHAPTER4

4.1 FUNCTIONAL REQUIREMENT

- ☐ IBM cognos analytics account
- ☐ IBM cloud
- ☐ Preparation of Dataset
- ☐ Data exploration
- ☐ Creating Dashboard
- ☐ Generating report

4.2 NON FUNCTIONAL REQUIREMENT

- ☐ Usability
- ☐ Reliability
- ☐ Performance
- ☐ Availability

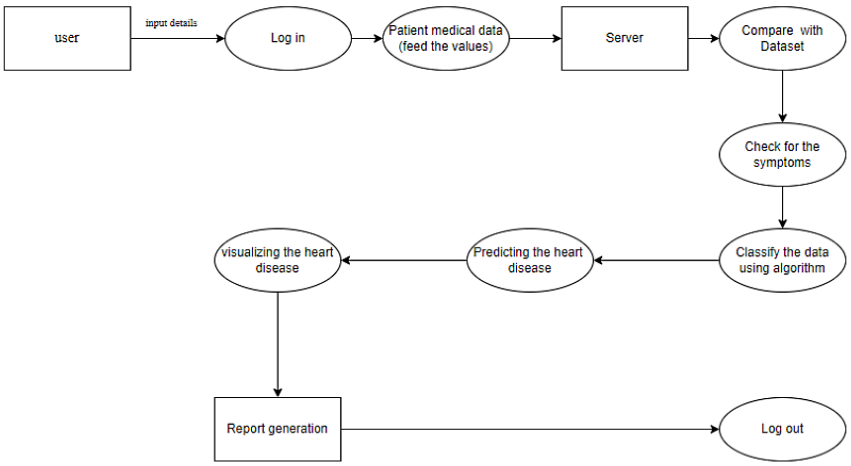
- [ ] Compatability
- [ ] Maintainability

CHAPTER5

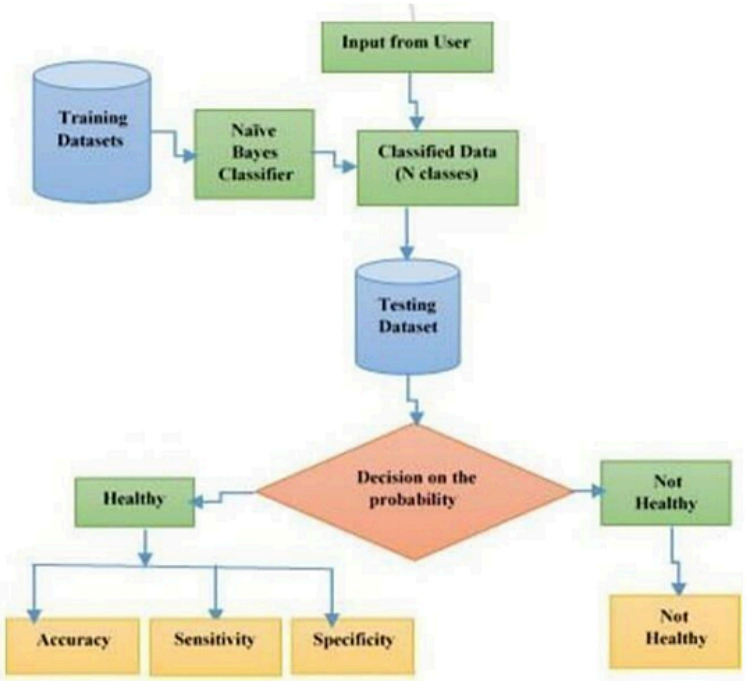
5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

Data Flow Diagrams:

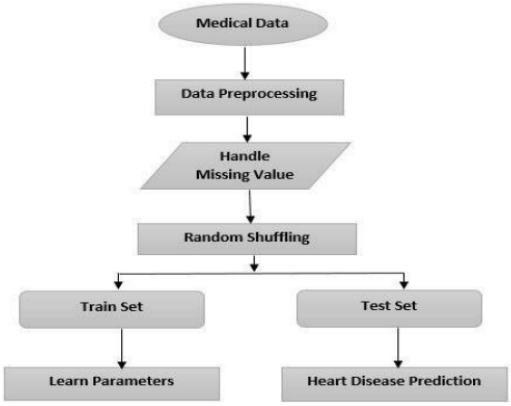


5.2 SOLUTION & TECHNICAL ARCHITECTURE


Input from User
Training Datasets
Naïve Bayes Classifier
Classified Data (N classes)
Testing Dataset
Decision on the probability
Healthy
Not Healthy
Accuracy
Sensitivity
Specificity
Not Healthy

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



5.3 USER STORIES

1.USER STORY NUMBER : USN1

USER TYPE : CUSTOMER (WEB USER)

FUNCTIONAL REQUIREMENT : IBM COGNOS

ACCEPTANCE CRITERIA : USER CAN SEE THE WEBPAGE

PRIORITY :LOW

RELEASE :SPRINT1

2.USER STORY NUMBER : USN2

USER TYPE : CUSTOMER

FUNCTIONAL REQUIREMENT : WEB USER

ACCEPTANCE CRITERIA : VISIT THE EXPLORATION

PRIORITY : LOW

RELEASE : SPRINT2

3.USER STORY NUMBER : USN 3

USER TYPE : CUSTOMER

FUNCTIONAL REQUIREMENT : BROWSER

ACCEPTANCE CRITERIA : USER CAN GIVE THE REQUIED DETAILS

PRIORITY : HIGH

RELEASE : SPRINT1

4.USER STORY NUMBER :USN 4

USER TYPE : CUSTOMER

FUNCTIONAL REQUIREMENT : IBM COGNOS

ACCEPTANCE CRITERIA :USER CAN SEE THE RESULT

PRIORITY :HIGH

RELEASE :SPRINT 1

5.USER STORY NUMBER :USN5

USER TYPE :CUSTOMER

FUNCTIONAL REQUIREMENT : REPRORT CAN BE VISUALIZED

ACCEPTANCE CRITERIA :WHEN USER CLICK DATA CAN BE VISUALIZED

PRIORITY : HIGH

RELEASE :SPRINT 4

CHAPTER6

6.1 SPRINT PLANNING AND ESTIMATION

SPRINT	FUNCTIONAL REQUIREMENT	USER STORY	USER STORY/TASK	STORY POINTS	PRIORITY	MEMBERS
SPRINT 1	DATA COLLECTION	USN 1	collecting required data from kaggle.com	1	low	**1**
SPRINT 2	DATA EXPLORATION	USN 2	data exploration can be viewed by the user	2	high	
SPRINT 1		USN 3	home data analysis	2	low	2
SPRINT 1	LINK	USN 4	the user will have to fill in the below 13 fields for predicting system -Age -chest	2	high	5

			pain type -fasting blood sugar -rest ecg -exang -slope -ca -thal -test blood pressure -thalach -oldpeak			
SPRINT 2	DATA MODULES	USN 5	data modules are prepared	3	high	1
SPRINT 1	system requirement	USN 6	hardware requirement: 1.laptop or pc 2.15 processor system or higher 3.4GB RAM or higher 4.128 GB ROM or higher	2	high	4
SPRINT 1		USN 7	software Requirement: 1.windows 10 or	2	high	4

			higher 2.browser 3.visual studio 4.IBM cognos 5.IBM cloud			
SPRINT 3	DASHBOARD CREATION	USN 8	dash doard can be created so that the user can visualize them	2	medium	2

6.2 sprint delivery schedule

sprint	**total story points**	**duration**	**sprint start date**	**sprint end date**	**story point compllted**	**sprint release date**
sprint1	20	6days	24 oct 2022	29 oct 2022	20	29 oct 2022
sprint2	20	6days	31 oct 2022	5 nov 2022	18	6 nov 2022
sprint3	20	6days	7 nov 2022	12 nov 2022	20	11 nov 2022

sprint4	20	6days				

CHAPTER 7

7.CODING AND SOLUTION

7.1FEATURE 1

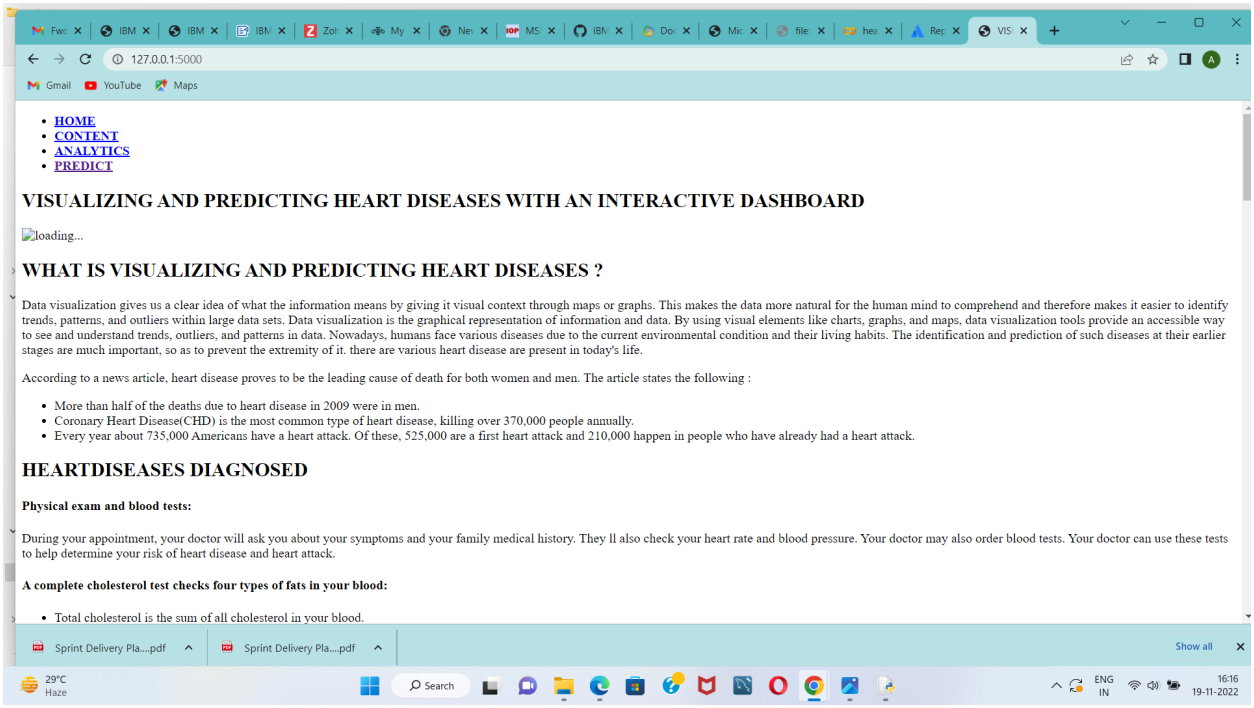


Figure 7.1 user can view the website

7.2 FEATURE 2

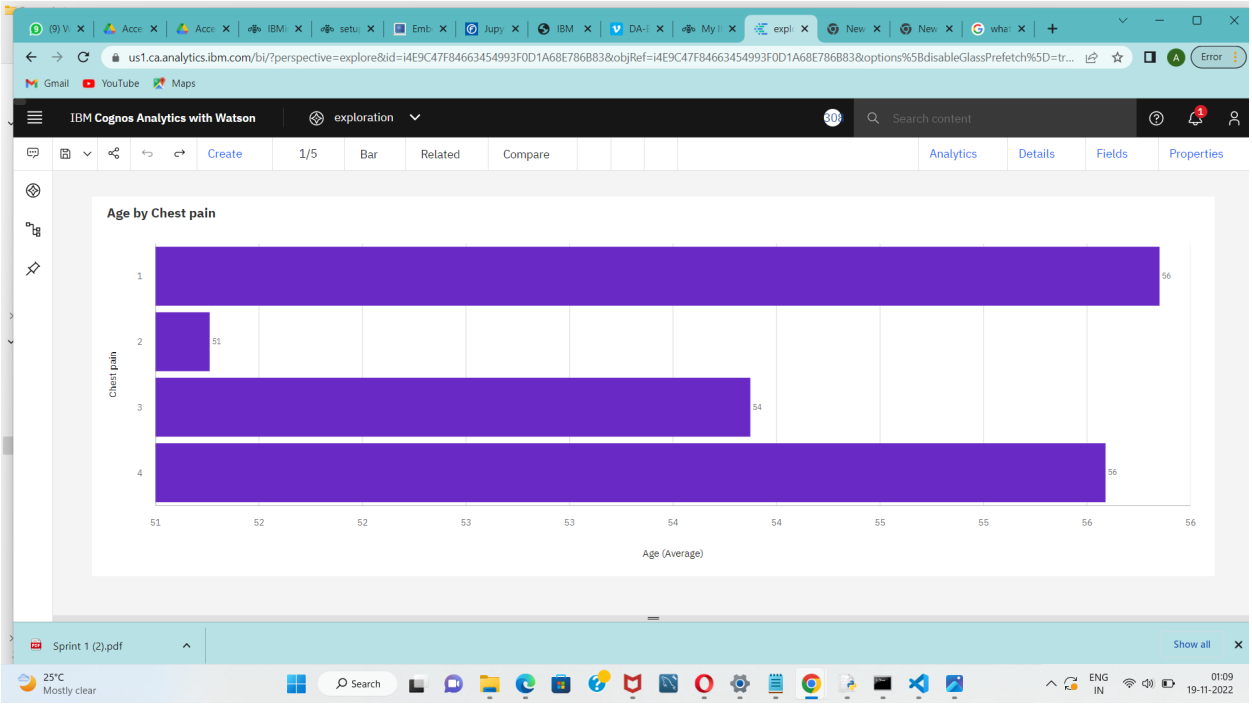


figure 7.2 user can visualize the data exploration by clicking the link

7.3 FEATURE 3

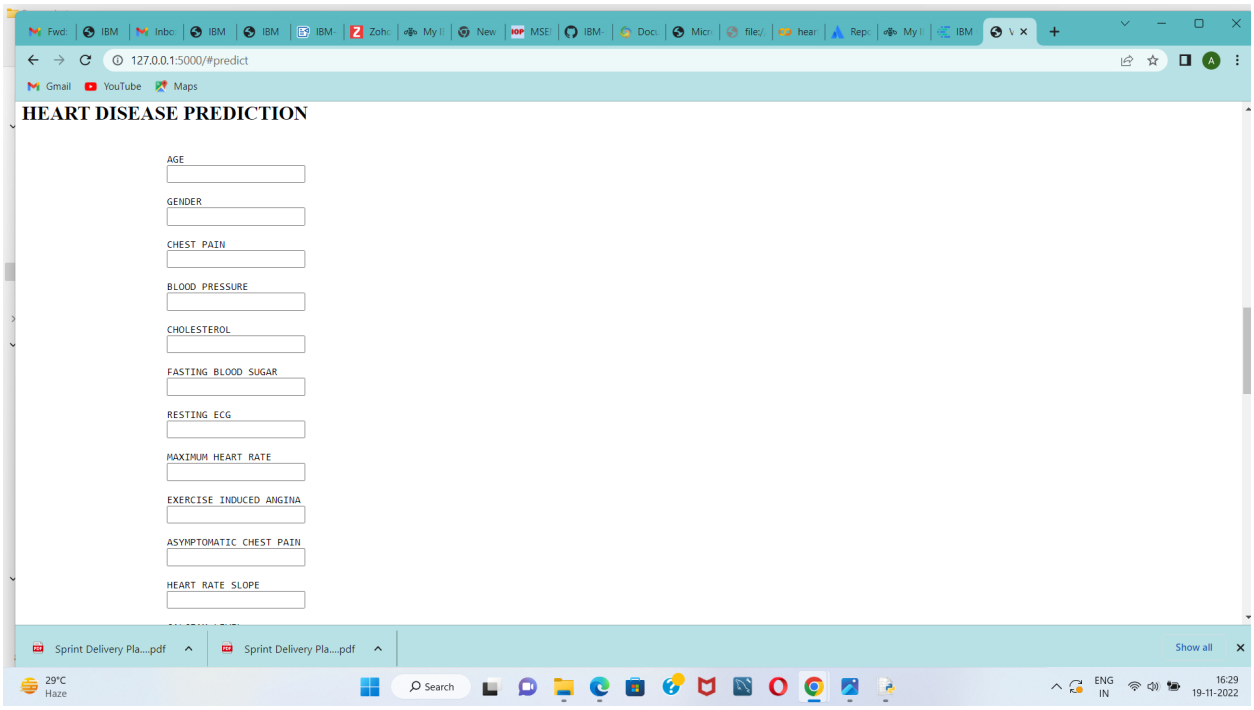
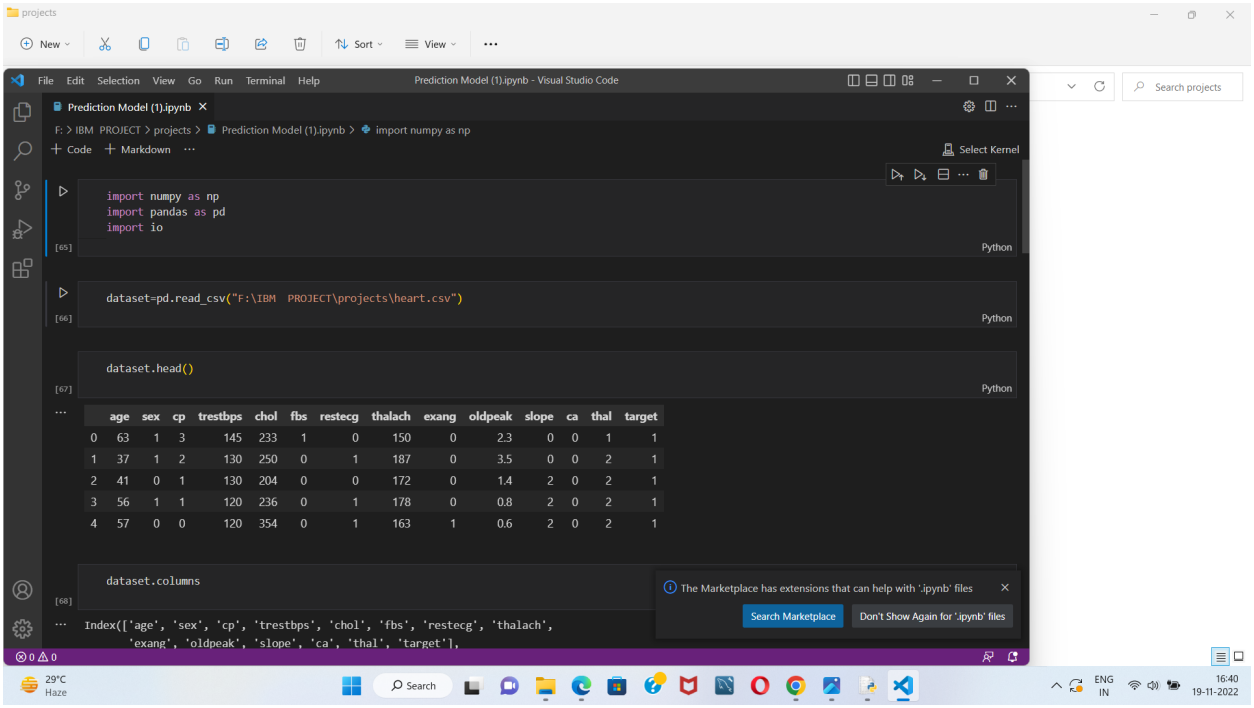


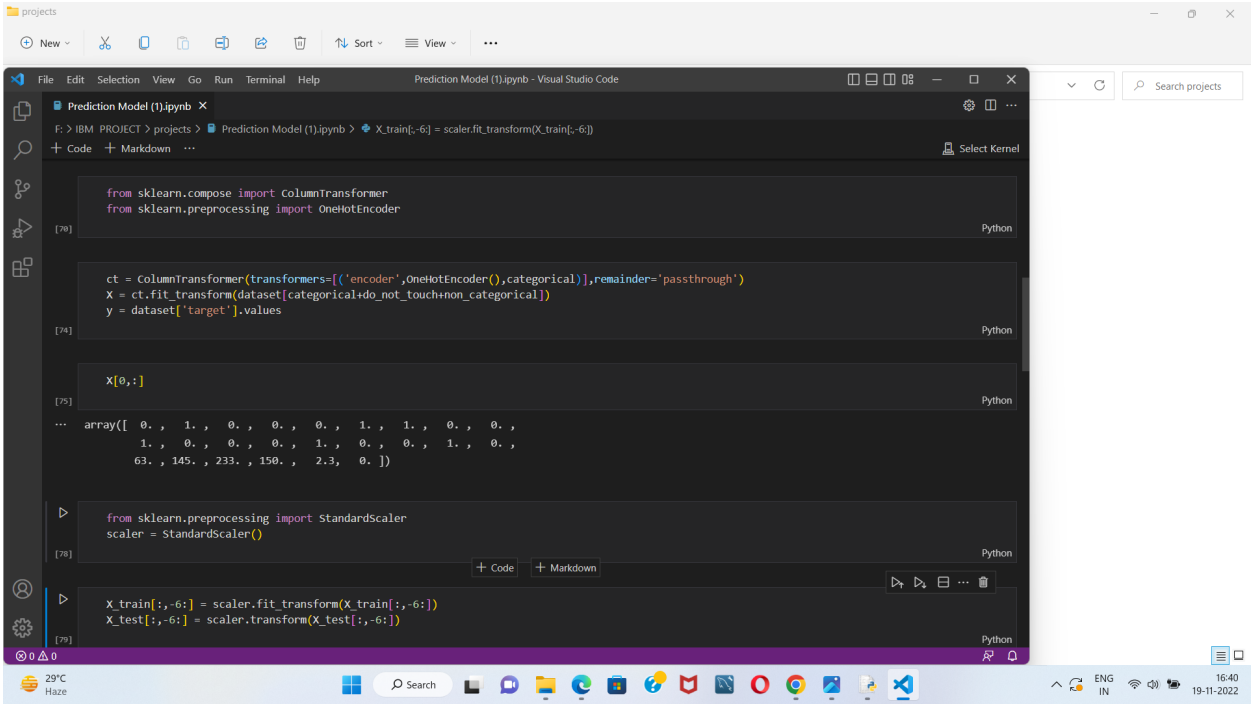
figure 7.3 Heart Disease prediction System

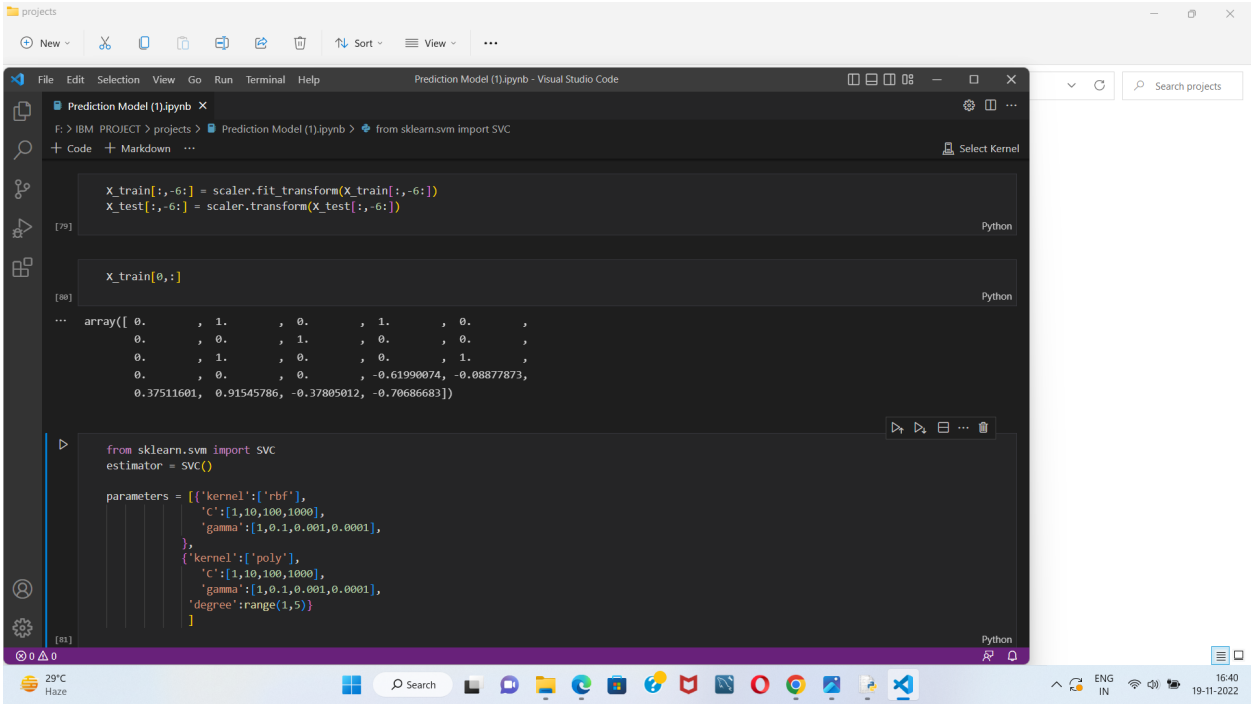
CHAPTER 8

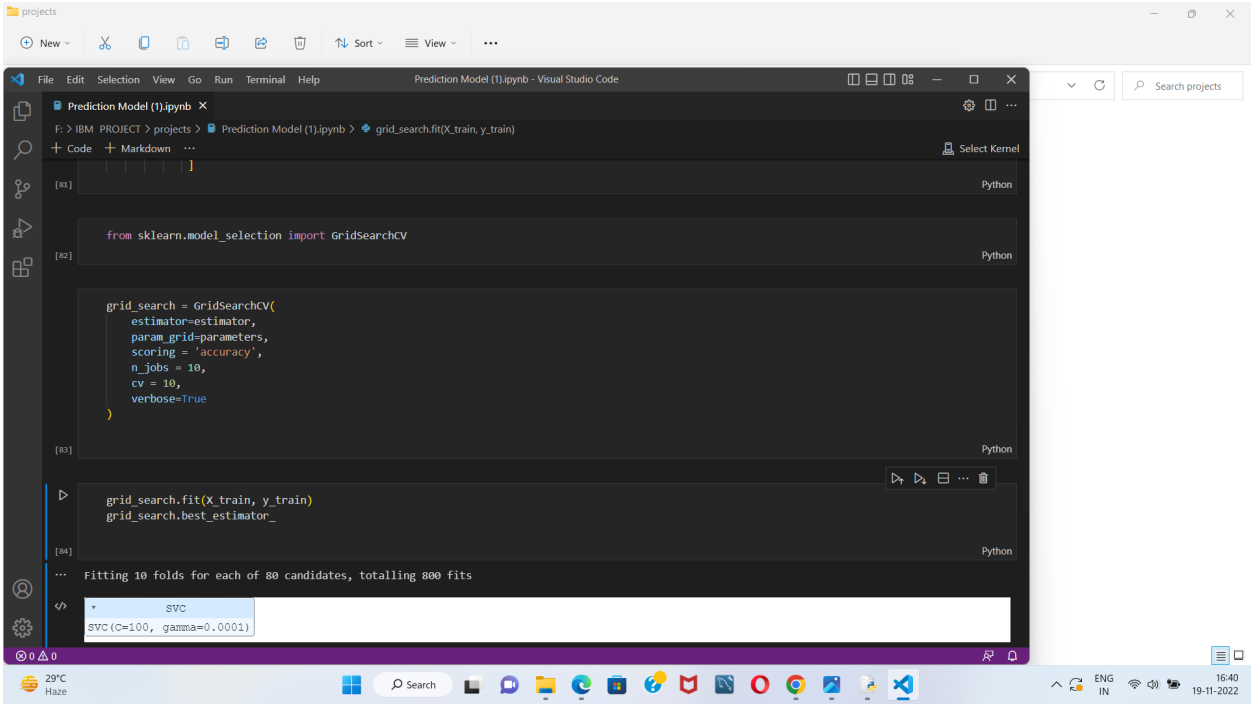
8.TESTING

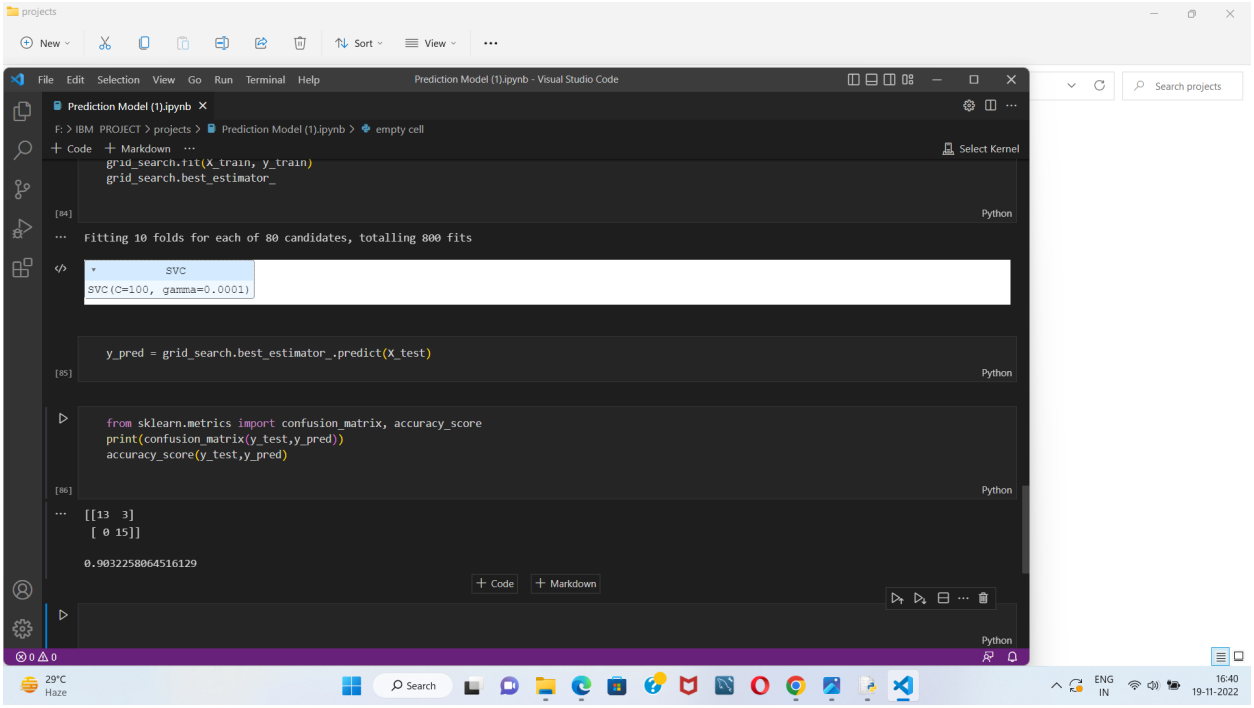
8.1 TEST CAESES




projects
New
Sort
View
Search projects
File Edit Selection View Go Run Terminal Help
Prediction Model (1).ipynb - Visual Studio Code
Prediction Model (1).ipynb
F: > IBM PROJECT > projects > Prediction Model (1).ipynb > X_train[:,-6:] = scaler.fit_transform(X_train[:,-6:])
Code
Markdown
Select Kernel
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
[70]
Python
ct = ColumnTransformer(transformers=[('encoder',OneHotEncoder(),categorical)],remainder='passthrough')
X = ct.fit_transform(dataset[categorical+do_not_touch+non_categorical])
y = dataset['target'].values
[74]
Python
X[0,:]
[75]
Python
array([0. , 1. , 0. , 0. , 0. , 1. , 1. , 0. , 0. ,
1. , 0. , 0. , 0. , 1. , 0. , 0. , 1. , 0. ,
63. , 145. , 233. , 150. , 2.3, 0.])
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
[78]
Python
Code
Markdown
X_train[:,-6:] = scaler.fit_transform(X_train[:,-6:])
X_test[:,-6:] = scaler.transform(X_test[:,-6:])
[79]
Python
0 0
29°C
Haze
Search
ENG
IN
16:40
19-11-2022


projects
New
Sort
View
Search projects
File Edit Selection View Go Run Terminal Help
Prediction Model (1).ipynb - Visual Studio Code
Prediction Model (1).ipynb
F: > IBM PROJECT > projects > Prediction Model (1).ipynb > from sklearn.svm import SVC
Code
Markdown
Select Kernel
X_train[:,-6:] = scaler.fit_transform(X_train[:,-6:])
X_test[:,-6:] = scaler.transform(X_test[:,-6:])
[79]
Python
X_train[0,:]
[80]
Python
array([0. , 1. , 0. , 1. , 0. ,
0. , 0. , 1. , 0. , 0. ,
0. , 1. , 0. , 0. , 1. ,
0. , 0. , 0. , -0.61990074, -0.08877873,
0.37511601, 0.91545786, -0.37805012, -0.70686683])
from sklearn.svm import SVC
estimator = SVC()
parameters = [{'kernel':['rbf'],
'C':[1,10,100,1000],
'gamma':[1,0.1,0.001,0.0001],
},
{'kernel':['poly'],
'C':[1,10,100,1000],
'gamma':[1,0.1,0.001,0.0001],
'degree':range(1,5)}
]
[81]
Python
29°C
Haze
Search
ENG
IN
16:40
19-11-2022


projects
New
Sort
View
Search projects
File Edit Selection View Go Run Terminal Help
Prediction Model (1).ipynb - Visual Studio Code
Prediction Model (1).ipynb
F: > IBM PROJECT > projects > Prediction Model (1).ipynb > grid_search.fit(X_train, y_train)
Code
Markdown
Select Kernel
]
[81]
Python
from sklearn.model_selection import GridSearchCV
[82]
Python
grid_search = GridSearchCV(
estimator=estimator,
param_grid=parameters,
scoring = 'accuracy',
n_jobs = 10,
cv = 10,
verbose=True
)
[83]
Python
grid_search.fit(X_train, y_train)
grid_search.best_estimator_
[84]
Python
Fitting 10 folds for each of 80 candidates, totalling 800 fits
SVC
SVC(C=100, gamma=0.0001)
29°C
Haze
Search
ENG
IN
16:40
19-11-2022

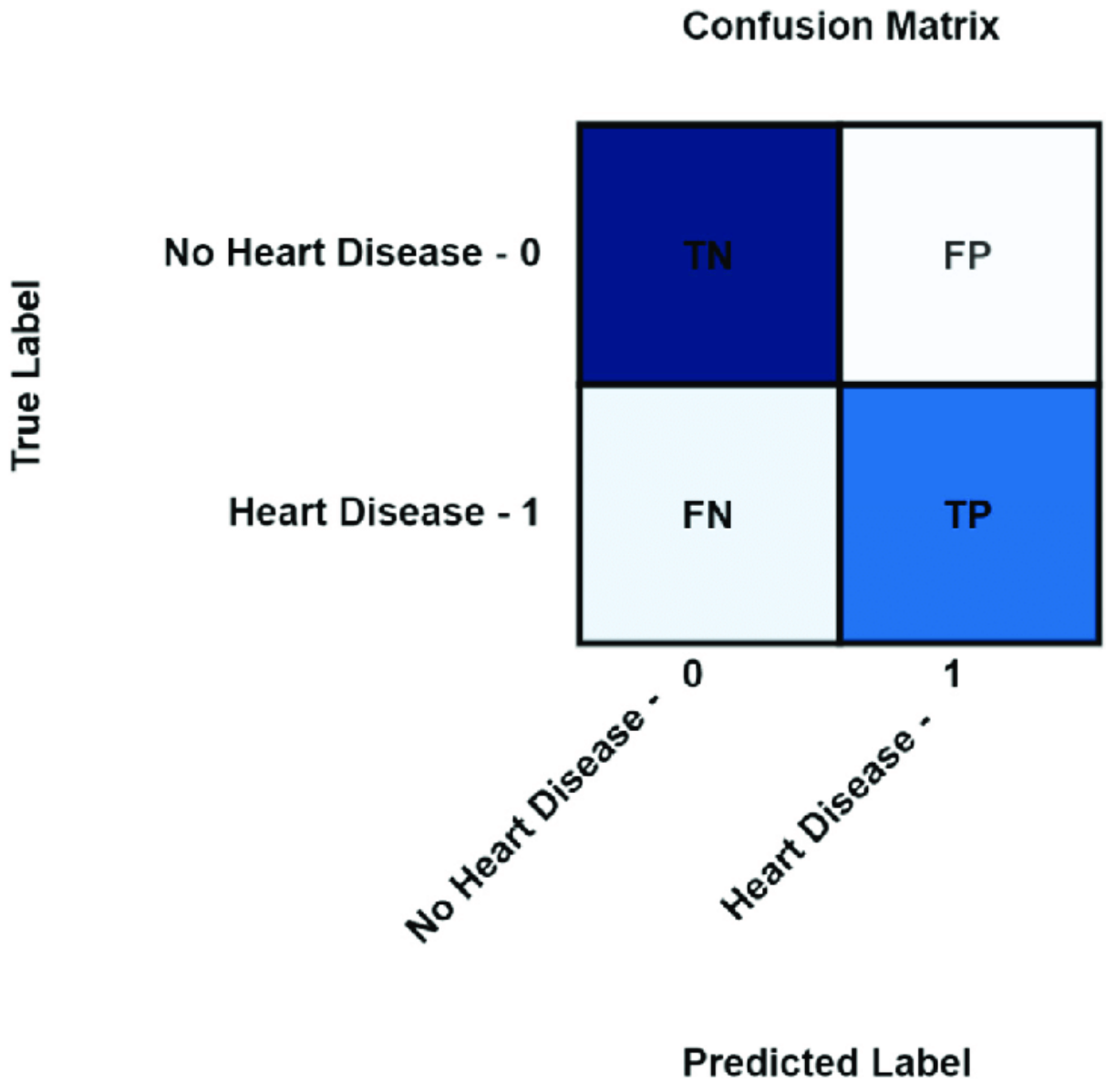

projects
New
Sort
View
Search projects
File Edit Selection View Go Run Terminal Help
Prediction Model (1).ipynb - Visual Studio Code
Prediction Model (1).ipynb
F: > IBM PROJECT > projects > Prediction Model (1).ipynb > empty cell
Code
Markdown
Select Kernel
grid_search.fit(X_train, y_train)
grid_search.best_estimator_
[84]
Python
Fitting 10 folds for each of 80 candidates, totalling 800 fits
SVC
SVC(C=100, gamma=0.0001)
y_pred = grid_search.best_estimator_.predict(X_test)
[85]
Python
from sklearn.metrics import confusion_matrix, accuracy_score
print(confusion_matrix(y_test,y_pred))
accuracy_score(y_test,y_pred)
[86]
Python
[[13 3]
[0 15]]
0.9032258064516129
Code
Markdown
Python
0 0
29°C
Haze
Search
ENG
IN
16:40
19-11-2022

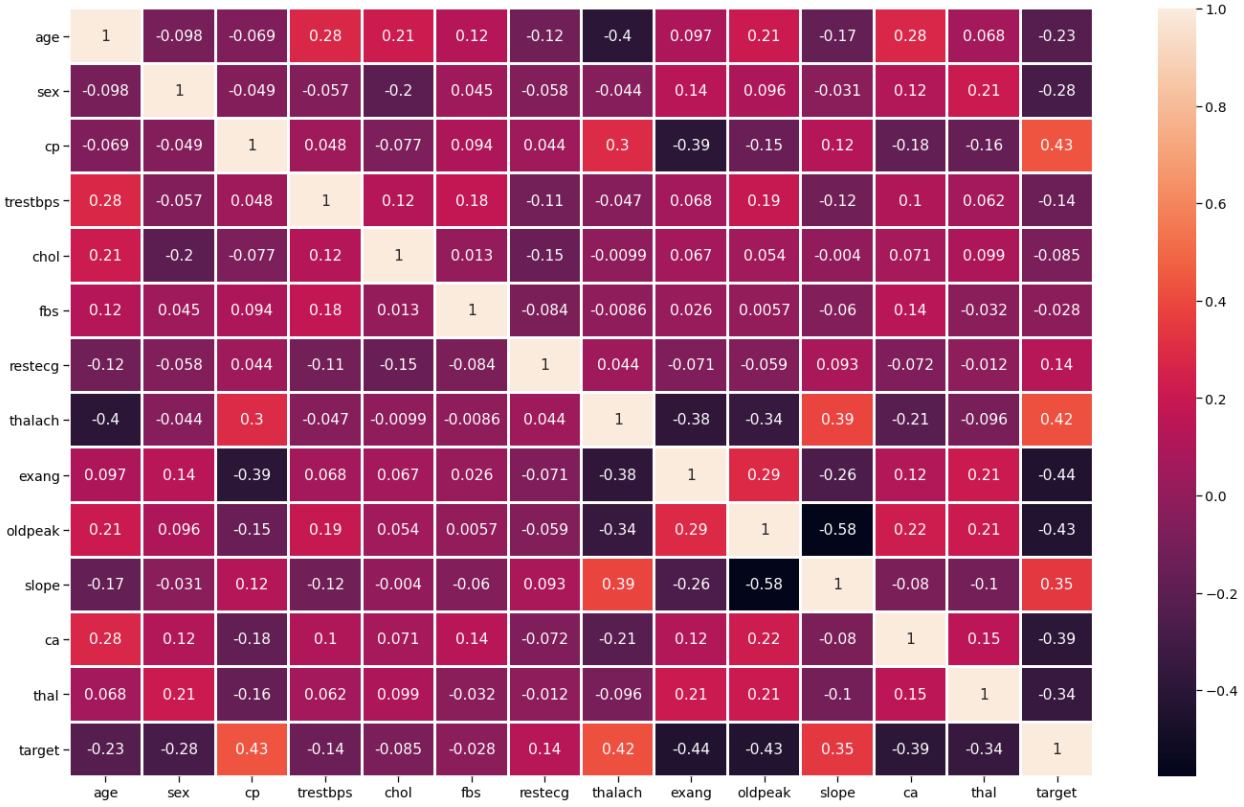
CHAPTER 9

9.RESULTS

9.1 PERFORMANCE METRICS

various performance metrices of the model is analyzed and documented as below




age 1 -0.098 -0.069 0.28 0.21 0.12 -0.12 -0.4 0.097 0.21 -0.17 0.28 0.068 -0.23
sex -0.098 1 -0.049 -0.057 -0.2 0.045 -0.058 -0.044 0.14 0.096 -0.031 0.12 0.21 -0.28
cp -0.069 -0.049 1 0.048 -0.077 0.094 0.044 0.3 -0.39 -0.15 0.12 -0.18 -0.16 0.43
trestbps 0.28 -0.057 0.048 1 0.12 0.18 -0.11 -0.047 0.068 0.19 -0.12 0.1 0.062 -0.14
chol 0.21 -0.2 -0.077 0.12 1 0.013 -0.15 -0.0099 0.067 0.054 -0.004 0.071 0.099 -0.085
fbs 0.12 0.045 0.094 0.18 0.013 1 -0.084 -0.0086 0.026 0.0057 -0.06 0.14 -0.032 -0.028
restecg -0.12 -0.058 0.044 -0.11 -0.15 -0.084 1 0.044 -0.071 -0.059 0.093 -0.072 -0.012 0.14
thalach -0.4 -0.044 0.3 -0.047 -0.0099 -0.0086 0.044 1 -0.38 -0.34 0.39 -0.21 -0.096 0.42
exang 0.097 0.14 -0.39 0.068 0.067 0.026 -0.071 -0.38 1 0.29 -0.26 0.12 0.21 -0.44
oldpeak 0.21 0.096 -0.15 0.19 0.054 0.0057 -0.059 -0.34 0.29 1 -0.58 0.22 0.21 -0.43
slope -0.17 -0.031 0.12 -0.12 -0.004 -0.06 0.093 0.39 -0.26 -0.58 1 -0.08 -0.1 0.35
ca 0.28 0.12 -0.18 0.1 0.071 0.14 -0.072 -0.21 0.12 0.22 -0.08 1 0.15 -0.39
thal 0.068 0.21 -0.16 0.062 0.099 -0.032 -0.012 -0.096 0.21 0.21 -0.1 0.15 1 -0.34
target -0.23 -0.28 0.43 -0.14 -0.085 -0.028 0.14 0.42 -0.44 -0.43 0.35 -0.39 -0.34 1
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

CHAPTER 10

10.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- Increased accuracy for effective heart diagnosis
- Reduce the time complexity of Doctors.
- Cost effective of patients.
- Helps in decreasing the difficulties both for the doctors and the patients.
- Medical Professionals can quickly recognize and respond to potential dangers with the right vizualization tools.
- Doctors may better define patient populations and allocate resources by displaying health data in real time.
- Visualization tools can significantly improve presentations and Reports.
- Making visualization may helpful for the easy understanding of patients

DISADVANTAGES:

- Inefficient data or poor structuring of data objects in the database may lead to inaccurate interpretation of visualized data.

CHAPTER 11

11.CONCLUSION

As we had worked through our project named " Visualising and predicting heart disease and analysis ", we had gathered to conclude the analysis through our entire progerss.The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes inhigh risk patients and in turn reduce the complications, which can be a great milestone in the fieldof medicine. We had almost tried to lessen complications for the disease prediction.We also hope that our website will be super useful for the forthcoming doctors and patients in the finest way.

This project of ours has resolved many things like complicating things lesser and make things easier,Manage to recover fast, trust in therapy and so on This project has also resolved the feature selection i.e. backward elimination ,Binary classifier etc.and successfully predict the heart disease, with 85% accuracy. The model used was Logistic Regression. Further for its enhancement, we can train on models and predict thetypes of cardiovascular diseases providing recommendations to the users and will also use more superior and enhanced models for future.

CHAPTER 12

12.FUTURE SCOPE

- The current project does for prediction and Visualization of Heart Diseases.It can be extended to add features related to hospital management.
- This project can be enhanced with adding some more diseases for prediction.
- include the deep learning for huge data set.
- AI features can be added in future for chatbot facilities.

CHAPTER 13

13.APPENDIX

REFERENCES

gitHub Repository LINK: **IBM-Project-46820-1660790953**

project Demo Link: http://127.0.0.1:5000