VisualizingandPredictingHeartDiseaseswithanInteractiveDashboard

NALAIYATHIRANPROJECTREPORT

2022

Submittedby

ABISHA M	**(822719104002)**
ABOORVA S	**(822719104003)**
DHANALAKSHMI K	**(822719104012)**
KARTHIKAYINI A	**(822719104026)**

TEAMID:PNT2022TMID47218

Content

VISUALIZINGANDPREDICTINGHEARTDISEASESWITHANINTERACTIVEDASHBOARD

1. Introduction

1.1 Project Overview

The leading cause of death in the developed world is heart disease. Therefore, there needs to beworkdonetohelppreventtherisksofhavingaheartattackorstroke.Thisprojectaimstocreatean interactive Dashboard using IBM Cognos Tool and dataset to predict which patients are mostlikelytosufferfromaheartdiseaseinthenearfutureusingthefeatures given.

1.2 Purpose

Heartdisease(HD) isamajorcauseofmortalityinmodernsociety.Medicaldiagnosisisanextremelyimportantbutcomplicatedtaskthatshouldbeperformedaccuratelyandefficiently.Cardiovasculardiseaseisdifficulttodetectduetoseveralriskfactors,includinghighbloodpressure, cholesterol, and an abnormal pulse rate. Based on the analytics we can analyze which patients are most likely to suffer from heart disease in the near future and based on the patient details we will make decisions to cure them.

2. Literature Survey

2.1 Existing Problem

Healthcare industries generate enormous amount of data, so called big data that accommodateshiddenknowledgeorpatternfordecisionmaking.Thehugevolumeofdataisusedtomakedecision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatoryvariables.Inthecontext,EDAisconsideredasanalyzingdatathatexcludesinferencesandstatistical modeling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Dataanalyticsisconsideredasacosteffectivetechnologyintherecentpast anditplaysanessentialroleinhealthcarewhichincludesnewresearchfindings,emergencysituations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive are and EDA is a vital step while analyzing data.

2.2 References

"Heart Disease Prediction via Evolutionary Rule Learning" by Aakash Chauhan and others (2018). In addition to assisting in directly extracting data from electronic records, this study eliminates the manual task. They used frequent pattern growth association mining on the patient dataset to produce strong association rules. For the prediction of cardiovascular disease, two

types of experiments were employed. In the first form, a random forest model is created, while in the second experiment, a random forest model based on the suggested Random Search Algorithm is created. In comparison to the traditional random forest model, this methodology is effective and simpler. It produces 3.3% greater accuracy when compared to traditional random forests. The suggested learning approach can aid medical professionals in better heart failure identification.

Senthil Kumar Mohan, ChandrasegarThirumalai and others "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" (2019) was an effective technique utilizing hybrid machine learning methodology. The hybrid method combines the random forest and linear approaches. The preprocessed cardiovascular disease dataset was used to choose a subset of specific attributes. Hybrid approaches were used to diagnose cardiovascular illness after preprocessing.

The paper on "Prediction and Diagnosis of Heart Disease Patients Using Data Mining Technique" was written by Mamatha Alex P and Shaicy P Shaji in 2019. The Artificial Neural Network, KNN, Random Forest, and Support Vector Machine techniques are used in this article

2.3 Problem Statement Definition

Who does the problem affect?

Peoplewithunhealthylifestyles,stress,depression,ageabove40andwhentheirancestorsgot Heart disease(since heart disease is hereditary).

When does the issue occur?

Theissueoccursforpeoplewithunhealthylifestylesandageabove40.Whereistheissue occurring? The issue is originating from an unhealthy life style. It mostly occurs in the blood values of the heart.

What would happen if we didn’t solve the problem?

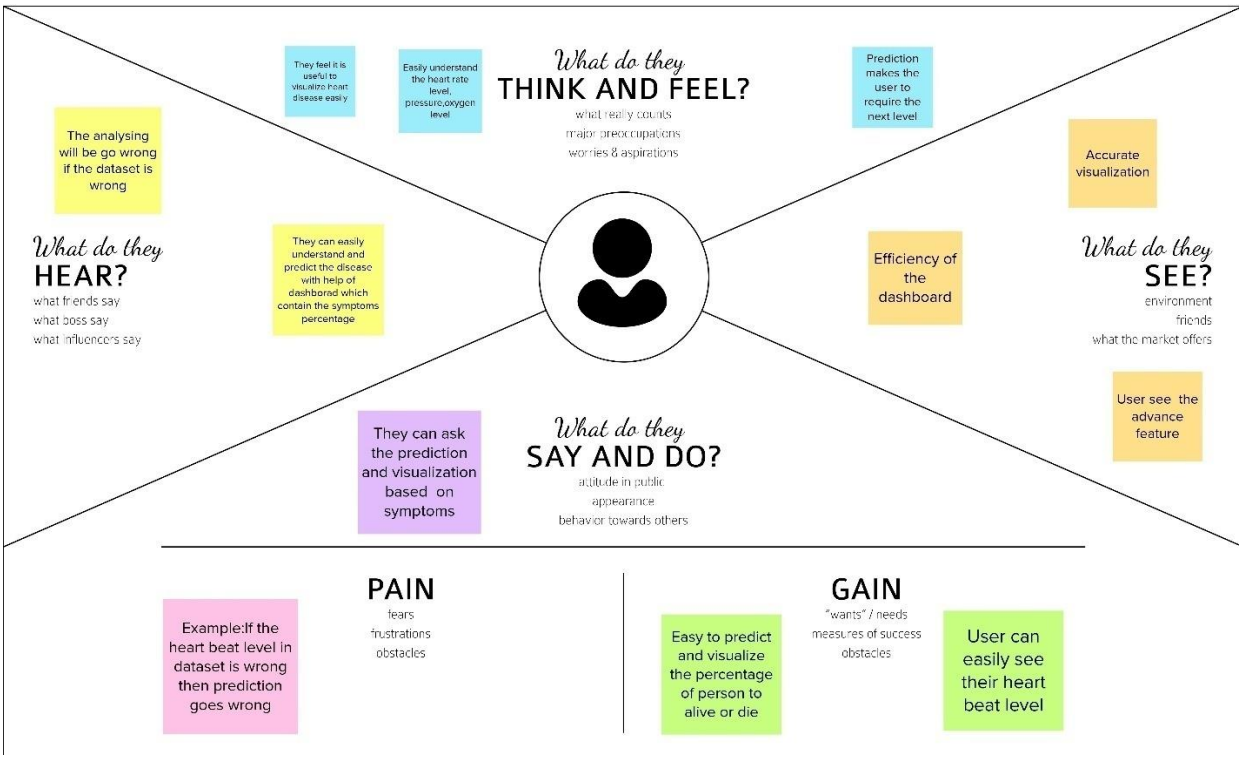
If we don’t solve the problem, many people will dieatayoung age. The death rate due to heart Disease will in crease rapidly.

Why is it important to fix the problem?

Weshouldpredicttheproblembeforegivingtreatmenttothepatients.Astheproblemis Predicted early, we can solve it easily and early.

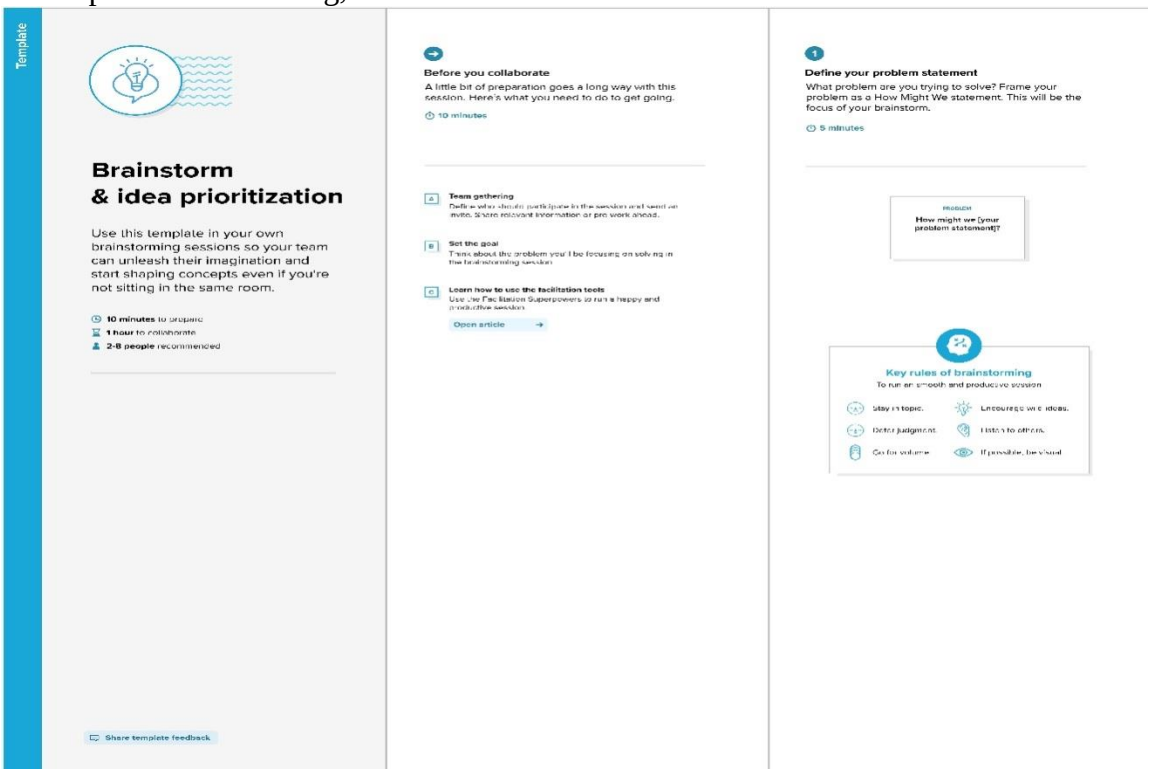
3. Ideation and Proposed Solution

3.1 Empathy Map Canvas

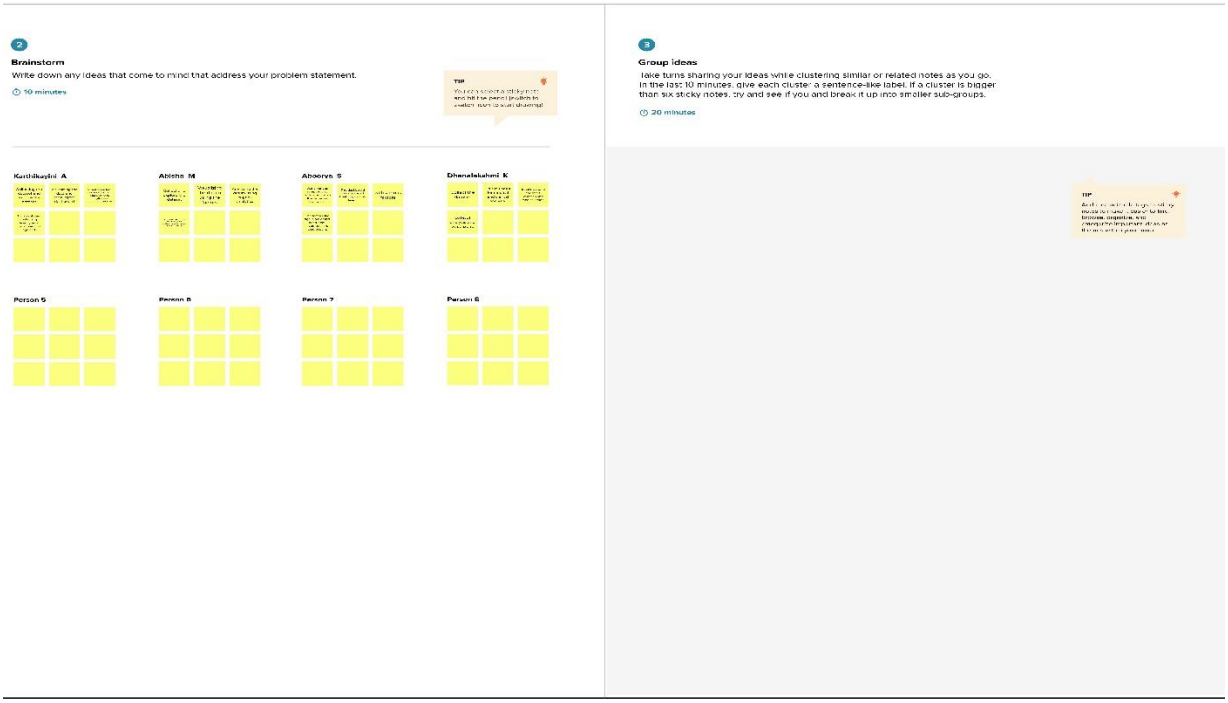


3.2 Ideation and Brainstorming

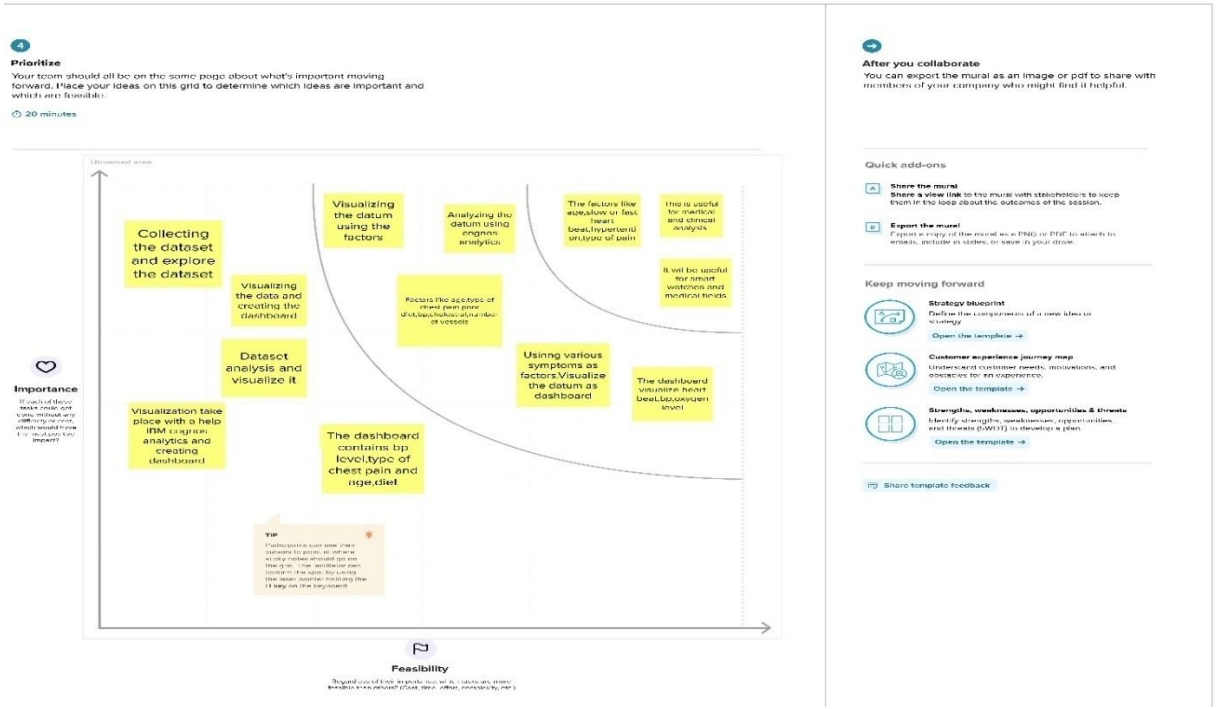
Step-1:TeamGathering,CollaborationandSelecttheProblemStatement



Step-2:Brainstorm,IdeaListingandGrouping



Step-3:IdeaPrioritization



3.3 Proposed Solution

S.NO.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The leading cause of death is heart disease. Heart disease refers to several types of abnormalities in heart conditions .It is inconvenient for a common man to take ECG tests periodically .Also, lack of proper diagnostic tools and accurate results affect the treatment of cardiac patients. Thus based on a patient's medical history, an expert's symptom analysis report, and physical laboratory results, invasive procedures are used to identify heart related problems .And so, there is a need for a replacement, which must be less complicated and reliable. The goal is to come up with a reliable prediction model so that the hospital can use this information to treat the patients at the starting state of the disease.
2.	Idea / Solution description	The idea is to provide an interactive dashboard for visualizing and predicting cardiac problems.IBM Cognos platform is used to visualize the given data. Machine learning techniques like Support Vector Machine, Decision tree, Naive Bayes, Random forest, K-Nearest Neighbour, and Neural networks are used to predict cardiac disease. To achieve greater accuracy, fusion of these algorithms is done. Exploratory Data Analysis (EDA) is a method to analyze data using advanced techniques to expose hidden structure, enhance the insight into a given dataset, identify the anomalies and build parsimonious models to test the underlying assumptions. The parameters provided in the data set help hospitals identify the patient's heart condition. An informative and creative dashboard can be created to present the data and utilize it for further medications.
3.	Novelty / Uniqueness	The prime uniqueness of the solution is the fusion of highly efficient algorithms, that eliminates the disadvantage of every

		algorithm when employed individually and also provides a higher level of accuracy in the prediction. Another innovation is employed in the dashboard by providing diet and fitness related suggestions to the user based on his/her medical reports and history. In addition to it, the patient is given a list of hospitals closer to the patient's locality and severity of the disease.
4.	Social Impact / Customer Satisfaction	It helps with disease prediction at an early stage and alerts the user about his/her current health status. Heart disease can be cured by a mix of medication, lifestyle modifications, and occasionally, surgery. The system helps the user as well as the doctor to make better decisions. Complex questions related to heart diseases can be answered by extracting hidden knowledge, i.e., patterns and relationships from the heart disease database.
5.	Business Model (Revenue Model)	Hospitals and healthcare facilities can install this interactive dashboard for heart disease prediction, allowing for speedy analysis. Predicted outcomes can be used to reduce the need for costly surgical operations by avoiding them altogether. Among all fatal disease, heart attacks disease reconsidered as the most prevalent. Medical practitioners conduct different surveys on heart diseases and gather information of heart patients, their symposiums and disease progression. It can be also used in educational institutions, industries and all types of workplaces to monitor the employee's health conditions and thereby helping them lead a healthier life.
6.	Scalability of the Solution	The proposed solution works efficiently in both smaller and larger datasets. This predictive model can be used to detect diseases in other internal organs too.

4. Requirement Analysis

4.1 Functional Requirement

Following are the functional requirements of the proposed solution.

FR. No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Web page	Preface of the Heart disease and the dashboard. Registration for new user and Sign in for registered user.
FR-2 U	User Registration	Registration through Gmail by entering the required details for registering
FR-3	User Confirmation	Confirmation via Email.
FR-4	User Login	The user will enter his/her email id and password used while in the time of registering.
FR-5	Upload other details	Upload the other clinical and non-clinical details that are asked for.
FR-6	Visualizing Data	Visualize the presence of heart diseases through dashboard created using IBM Cognos Analytics.
FR-7	Generation Report	View and generate the report.
FR-8	Result	POSITIVE-The result along with alert notification. NEGATIVE-The result along with precautionary measures

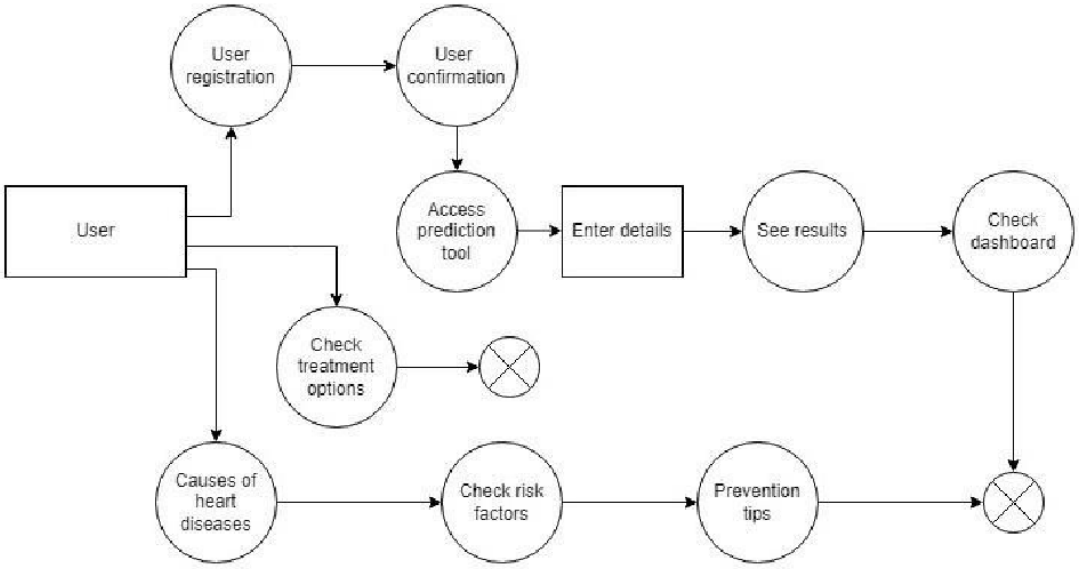
4.2 Non-FunctionalRequirement

Following are the non-functional requirements of the proposed solution.

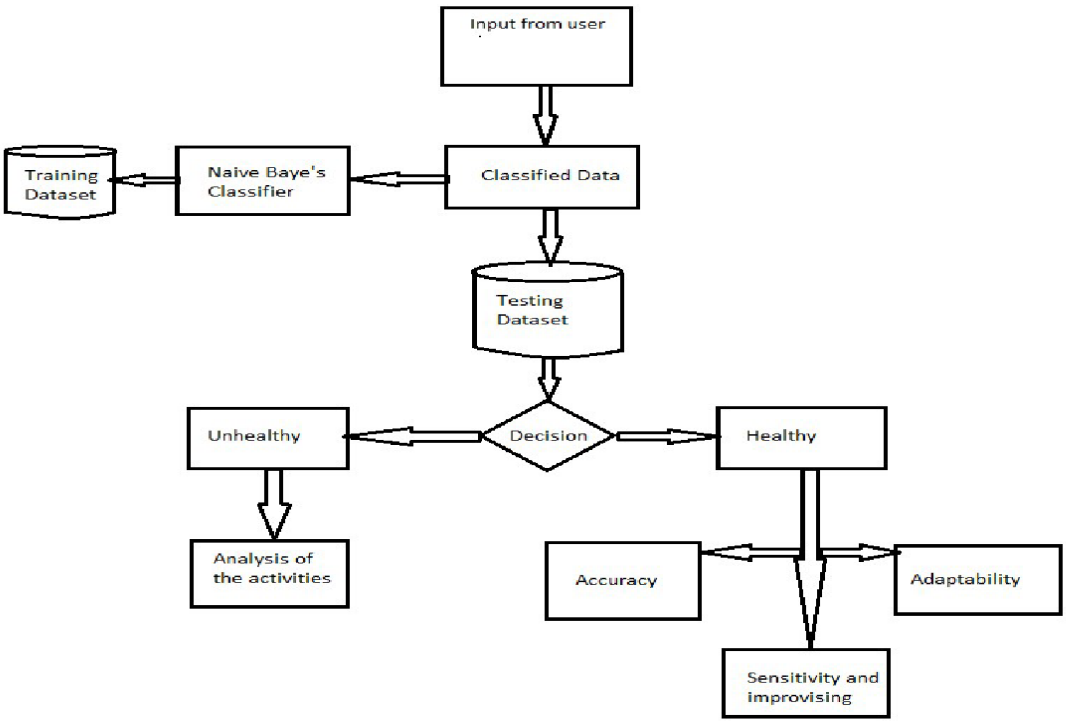
FR. NO.	Non-Functional Requirement	Description
NFR-1	Usability	Simple user interface with easy understanding. User friendly to access the dashboard.
NFR-2	Security	Maintain a secondary/backup dataset. User reports are kept safely and remove false dataset. Login details are shared via email on every login for confindentially.
NFR-3	Reliability	Must work without error or minimum error and able to work with various dataset. Highly reliable because predictions are made using the authorized clinical reports.
NFR-4	Performance	Depending on the error metrics we have to choose an algorithm with high response time. Low latency and high throughput.
NFR-5	Availability	It is based on the IBM Cognos Analytics so it should be available to all time. Exhibits good availability to all types of browsers
NFR-6	Scalability	Should handle a high number of request at the same time and a large datasets. It is highly scalable because it is an Interactive dashboard and the website is also maintained uninterrupted.

5. Project Design

5.1 Data Flow Diagram



5.2 Solution and Technical Architecture



6. Project Planning and Scheduling

6.1 Script Planning and Execution

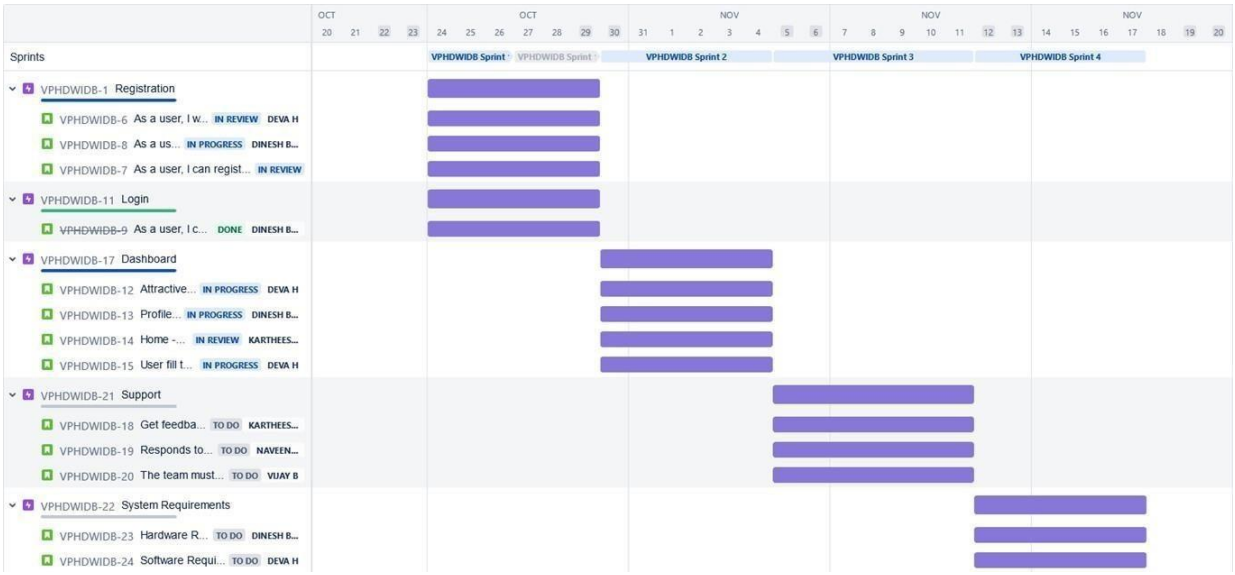
Sprint	Functionl Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Story points	Priority	Team Members
Sprint1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password	I can access my account / dashboard	10	high	Karthikayini A Aboorva S Abisha M Dhanalakshmi K
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	5	High	Karthikayini A Aboorva S Abisha M
		USN-3	As a user, I can register for the application through Gmail	I can register & access the dashboard with Gmail Login	5	High	Aboorva S Abisha M Dhanalakshmi K
Sprint-2	Login	USN-4	After Registration Login page will appear, the user will login using the login credentials	I can register & access the dashboard with Gmail Login	20	High	Karthikayini A Aboorva S Abisha M Dhanalakshmi K
Sprint3	Dashboard	USN-5	The user is allowed to view or update is profile	I can see the profile	10	Medium	Karthikayini A Aboorva S Abisha M Dhanalakshmi K
		USN-7	The user can change password	I can able to change the password.	10	Medium	Karthikayini A Aboorva S Abisha M Dhanalakshmi K
Sprint-4	Classified result	USN-8	Home - Analzse your Heart	I can detec the heart condition from where ever I want.	5	High	Karthikayini A Abisha M Dhanalakshmi K
		USN-9	The user will have to fill in the 13 required	This will prevent the user to	10	High	Karthikayini A Aboorva S Abisha M

			fields for the system to predict a heart disease	predict whether I has heart disease or not based on the values I entered			Dhanalakshmi K
		USN10	The report is generated based on the condition	The user can able to view/download the report if needed	5	Medium	Aboorva S Karthikayini A Dhanalakshmi K

6.2Sprint Delivery Schedule

Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date(Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint Releas Date(Actual)**
Sprint-1	20	6Days	24Oct2022	29Oct2022	20	29Oct2022
Sprint-2	20	6Days	30Oct2022	04Nov2022	17	04Nov2022
Sprint-3	20	6Days	05Nov2022	11Nov2022	18	11Nov2022
Sprint-4	20	6Days	12Nov2022	17Nov2022	19	17Nov2022

6.3JiraReport



7. Coding And Solution

7.1 Machine Learning

Learning which model is best for the given Dataset

Out[ ]:

	Estimators	Accuracy
0	Linear Regression	0.565830
3	K-Nearest Neighbor	0.729167
4	Random Forest	0.854167
5	Bagging Decision Tree	0.854167
6	Hard coting classifier	0.854167
2	Gaussian Naive Bayes	0.875000
1	Logistic Regression	0.895833

From the above result we can conclude that Logistic Regression has the hisgest accuracy for this particular dataset.

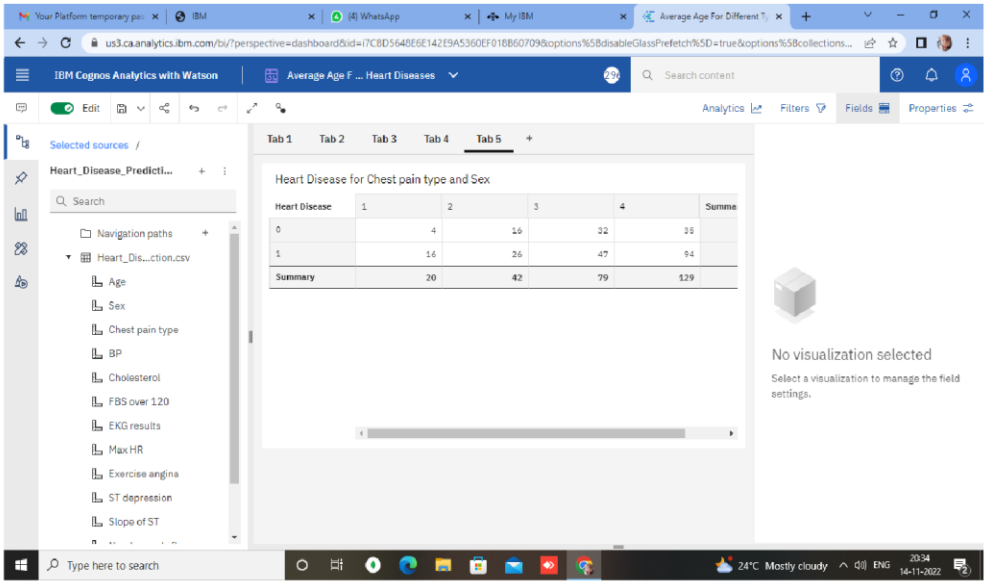
Comparing it with the accuracy gotten from Decision Tree:

```
TP=cm[0][0]#cm=Confusion Matrix TN=cm[1][1]
        FN=cm[1][0]FP=cm[0][1]
        print('Testing Accuracy for Decision
        Tree:',(TP+TN)/(TP+TN+FN+FP))print('TestingSensitivityforDecisionTr
        ee:',(TP/(TP+FN)))print('TestingSpecificity for Decision
        Tree:',(TN/(TN+FP))) print('Testing Precision for Decision
        Tree:',(TP/(TP+FP)))
```

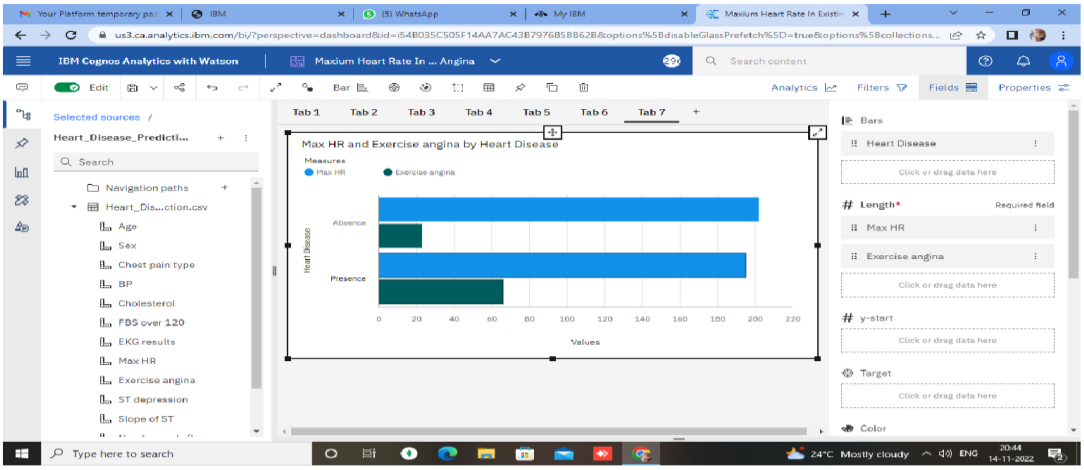
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

7.2. Dashboard

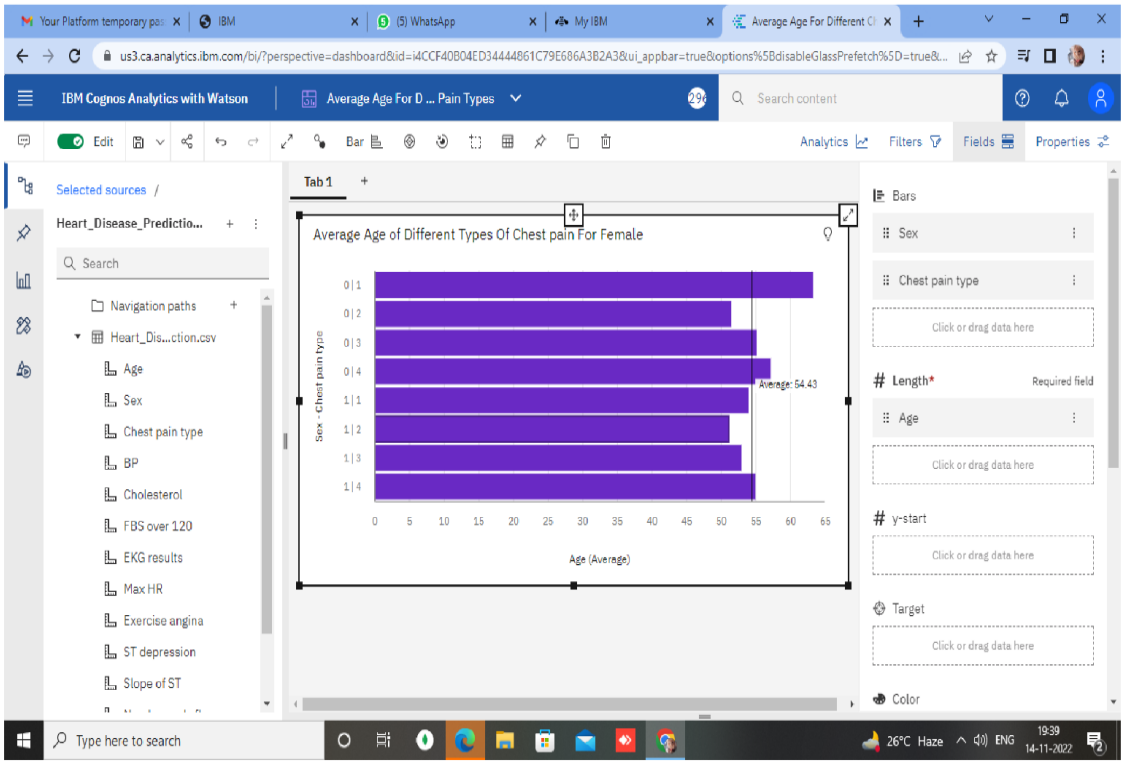
Average Age For Different Types Of Chest Pain In Existing Heart Diseases



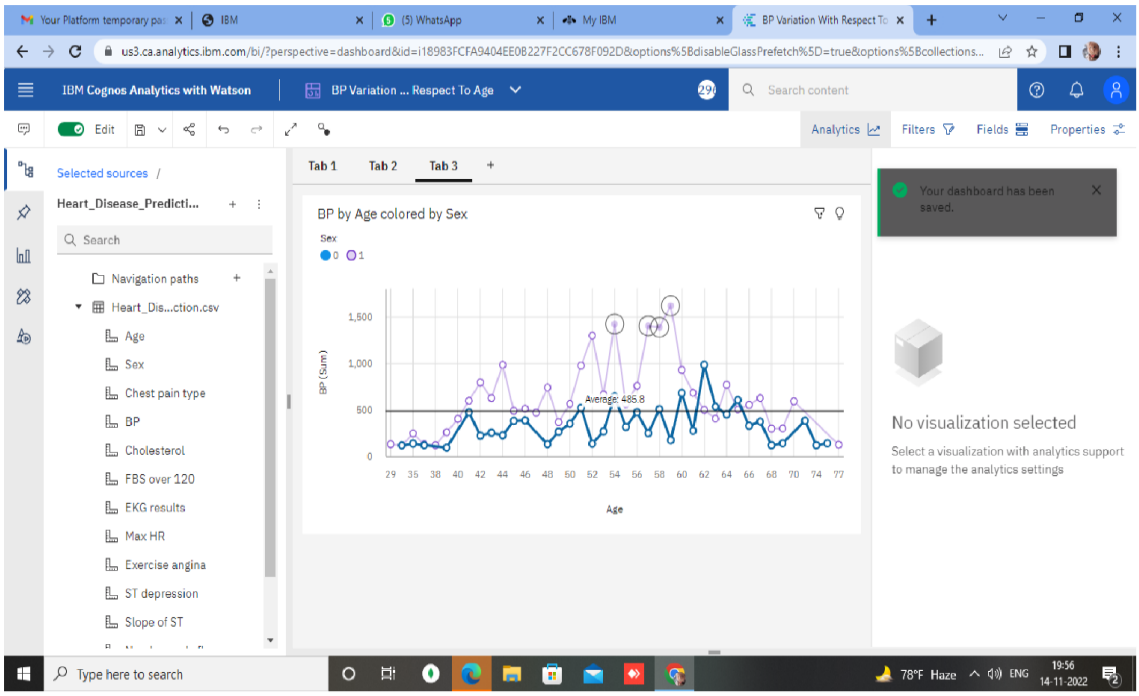
Maximum Heart Rate In Existing Heart Disease By Exercise Angina



Average Age For Different Chest Pain Types



BP Variation With Respect To Age

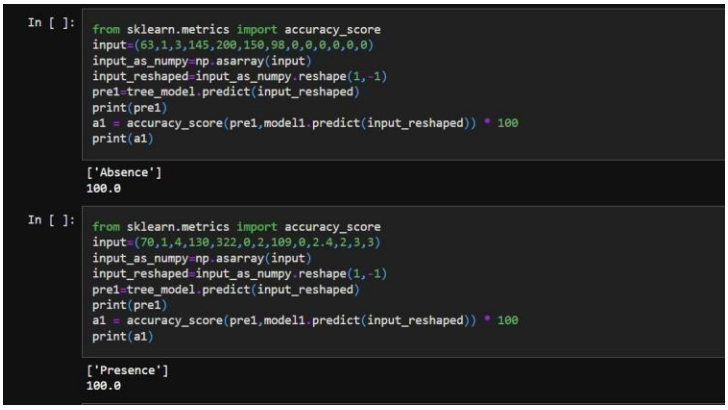


Dashboard Showing Different Types Of Visuals:

8. Testing

8.1 Test Cases

Testing the data model for various input values.



```
In [ ]: from sklearn.metrics import accuracy_score
        input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped)) * 100
        print(a1)

        ['Absence']
        100.0

In [ ]: from sklearn.metrics import accuracy_score
        input=(70,1,4,130,322,0,2,109,0,2.4,2,3,3)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped)) * 100
        print(a1)

        ['Presence']
        100.0
```

8.2 UseracceptanceTesting

Testing a case where user has heart disease

1. Purpose of Document

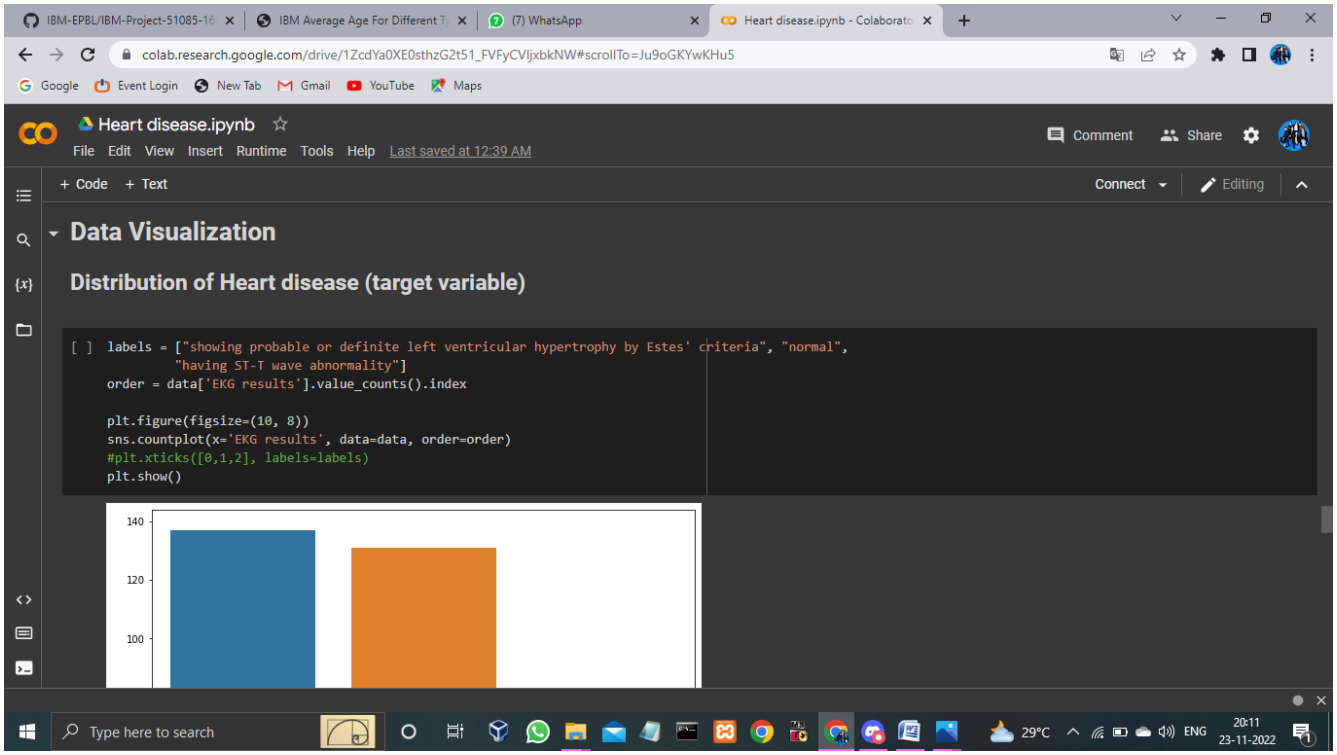
The purpose of this document is to briefly explain the test coverage and open issues of the Visualizing and Predicting Heart Diseases with an Interactive Dashboard project at the time of the release to User Acceptance Testing (UAT).

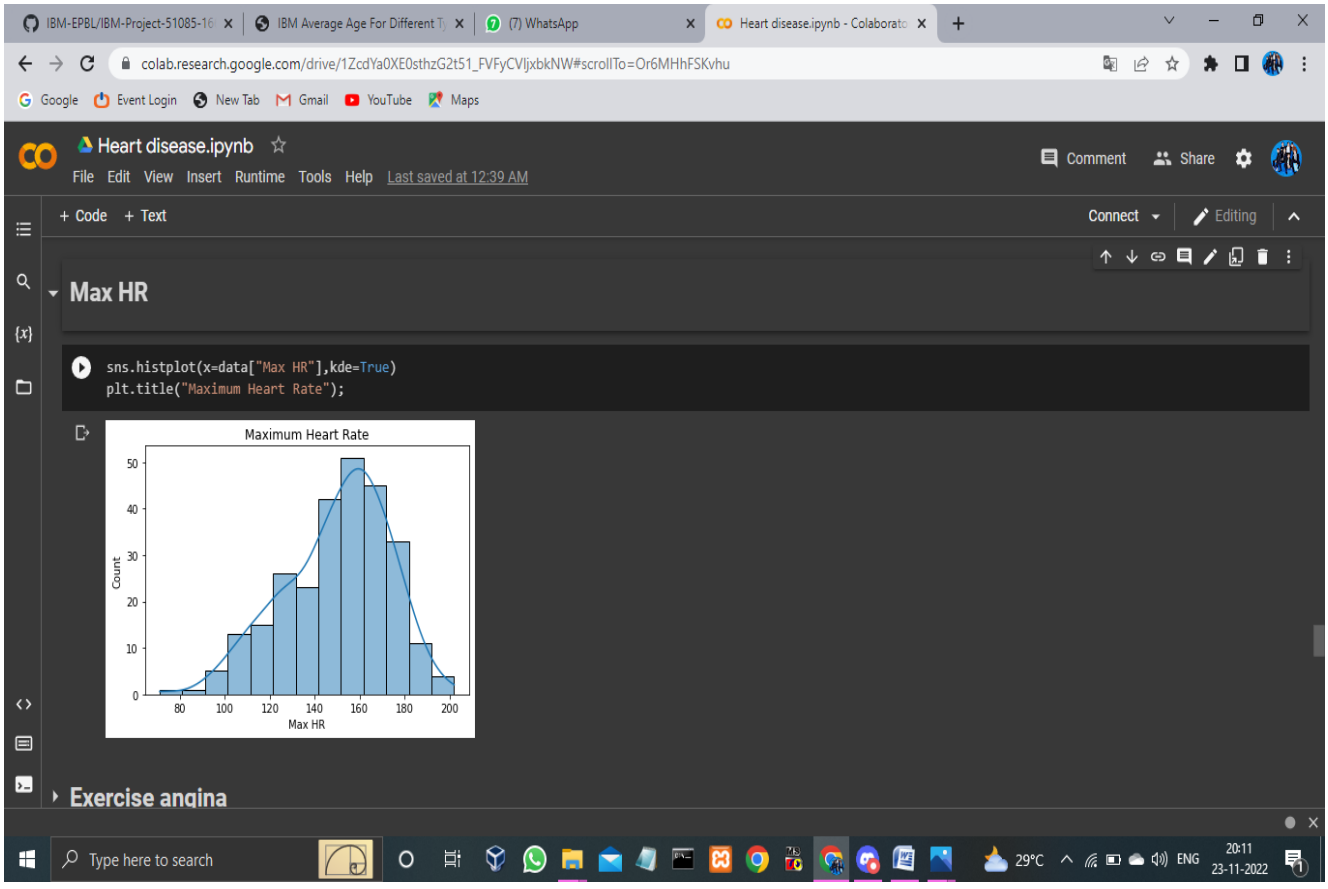
2. Defect Analysis

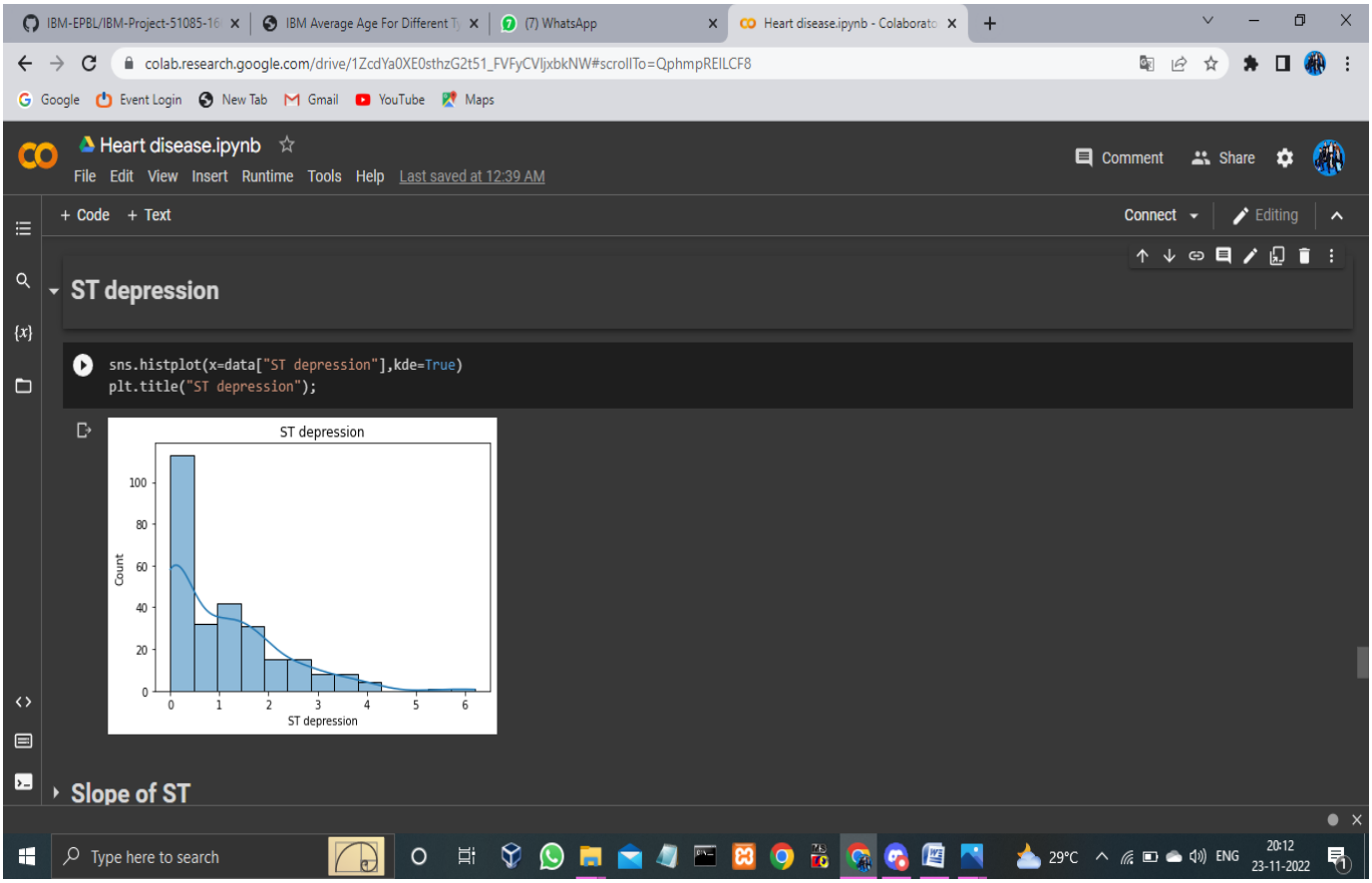
This report shows the number of resolved or closed bugs at each severity level, and how they wereresolved.

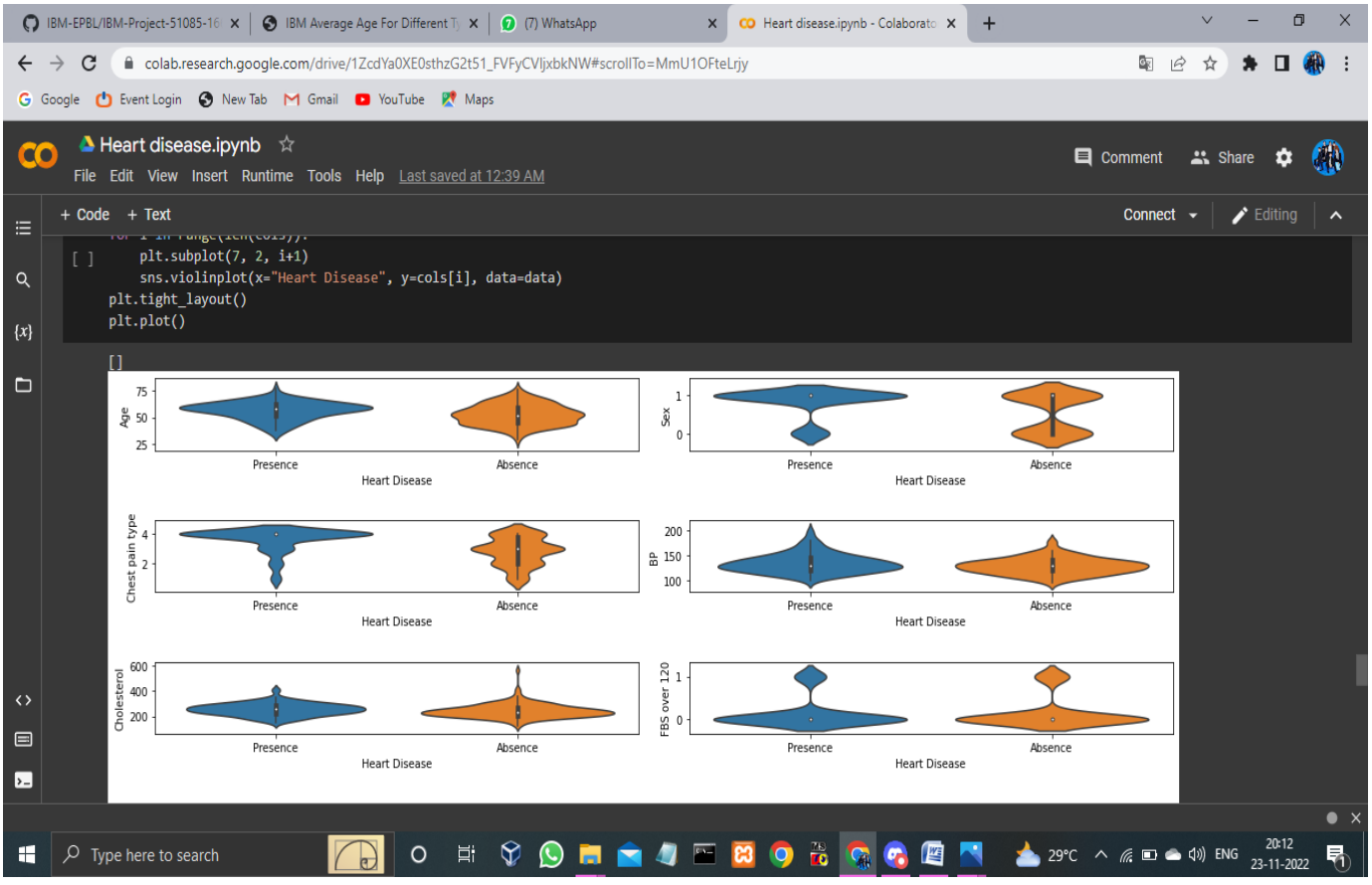
Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	7	5	1	1	14
Duplicate	1	0	2	0	3
External	2	1	1	1	5
Fixed	7	5	1	1	12
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	17	11	4	3	34

Coding:


Heart disease.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:39 AM
Comment
Share
+ Code
+ Text
Connect
Editing
Data Visualization
Distribution of Heart disease (target variable)
[ ] labels = ["showing probable or definite left ventricular hypertrophy by Estes' criteria", "normal",
"having ST-T wave abnormality"]
order = data['EKG results'].value_counts().index
plt.figure(figsize=(10, 8))
sns.countplot(x='EKG results', data=data, order=order)
#plt.xticks([0,1,2], labels=labels)
plt.show()
140
120
100
Type here to search
29°C
ENG
20:11
23-11-2022


Heart disease.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:39 AM
Comment
Share
+ Code
+ Text
Connect
Editing
Max HR
sns.histplot(x=data["Max HR"],kde=True)
plt.title("Maximum Heart Rate");
Maximum Heart Rate
Count
Max HR
0
10
20
30
40
50
80
100
120
140
160
180
200
Exercise angina
Type here to search
29°C
ENG
20:11
23-11-2022


IBM-EPBL/IBM-Project-51085-16
IBM Average Age For Different T
(7) WhatsApp
Heart disease.ipynb - Colaborato
colab.research.google.com/drive/1ZcdYa0XE0sthzG2t51_FVFyCVIjxbkNW#scrollTo=QphmpREILCF8
Google
Event Login
New Tab
Gmail
YouTube
Maps
Heart disease.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:39 AM
Comment
Share
+ Code + Text
Connect
Editing
ST depression
sns.histplot(x=data["ST depression"],kde=True)
plt.title("ST depression");
ST depression
Count
ST depression
Slope of ST
Type here to search
29°C
ENG
20:12
23-11-2022


IBM-EPBL/IBM-Project-51085-16
IBM Average Age For Different T
(7) WhatsApp
Heart disease.ipynb - Colaborato
colab.research.google.com/drive/1ZcdYa0XE0sthzG2t51_FVFyCVIjxbkNW#scrollTo=MmU1OFteLrjy
Google
Event Login
New Tab
Gmail
YouTube
Maps
Heart disease.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:39 AM
Comment
Share
+ Code + Text
Connect
Editing
plt.subplot(7, 2, i+1)
sns.violinplot(x="Heart Disease", y=cols[i], data=data)
plt.tight_layout()
plt.plot()
Age
Sex
Chest pain type
BP
Cholesterol
FBS over 120
Presence
Absence
Heart Disease
Type here to search
29°C
ENG
20:12
23-11-2022

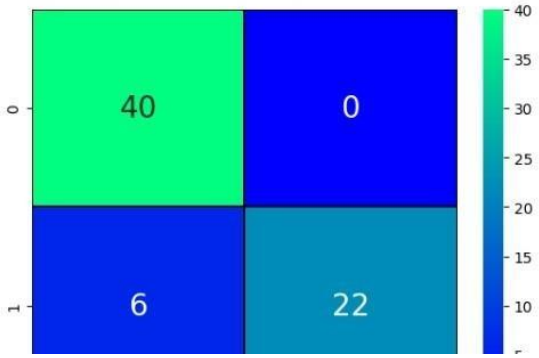
9. Result

9.1 Performance Metrics

The confusion matrix below shows the performance metrics of the machine learning model.

```
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

tree_model = DecisionTreeClassifier(max_depth=5,criterion='entropy')
cv_scores = cross_val_score(tree_model, x, y, cv=10, scoring='accuracy')
m=tree_model.fit(x, y)
prediction=m.predict(X_test)
cm= confusion_matrix(y_test,prediction)
sns.heatmap(cm, annot=True,cmap='winter',linewidths=0.3, linecolor='black',annot_kws={"size": 20})
print(classification_report(y_test, prediction))

TP=cm[0][0]
TN=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN)))
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))
print('Testing Precision for Decision Tree:',(TP/(TP+FP)))
```

```
              precision    recall  f1-score   support

      Absence       0.87      1.00      0.93        40
     Presence       1.00      0.79      0.88        28

     accuracy                           0.91        68
    macro avg       0.93      0.89      0.91        68
 weighted avg       0.92      0.91      0.91        68

Testing Accuracy for Decision Tree: 0.9117647058823529
Testing Sensitivity for Decision Tree: 0.8695652173913043
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```



10. Advantages Disadvantages

Advantages:

- This is one of the fastest ways to determine if a personal likely to suffer from a heart disease or not.
- Use full for medical practitioners to easily classify the in patients.
- User Friendly
- Easy to understand
- Secure
- Dash board provides in sight full in formations

Disadvantages:

- Needs work
- Users need to know all the fields
- Does Not take null value as input

- Does not provide suggestions to user

11. Conclusion

Complications of heart disease include heart attack and stroke .You can reduce there is complications with early diagnosis and treatment. So the suggestion that we get from the website might help save patients .It is always to get treated in the early stages of heart disease.

12. Future Scope

Like the saying goes “Prevention is better than cure”. We have to look in to methods to prevent heart diseases all together other than just predicting it in early stages.
To use this website we need to take a lot of tests beforehand. So it would be better if we require less attributes and still give an effective result.

Reference Link:

https://drive.google.com/file/d/1y8ptpgiRoscERpYMpfpM_rah0jYhWmZv/view?usp=drivesdk