Project Report Format

PROJECT REPORT

1.INTRODUCTION

1.1 PROJECT OVERVIEW

An Interactive Dashboard which uses Machine Learning algorithm to predict the heart condition of a person by providing some inputs about the person health like age, gender, blood pressure, cholesterol level etc built using IBM Cognos

As being a Data and ML enthusiast I have tried many different projects related to the subject but what I have realized is that Deploying your machine learning model is a key aspect of every ML and Data science project. Everything thing I had studied or been taught so far in my Data science and ML journey had mostly focused on defining problem statement followed by Data collection and preparation, model building and evaluation process which is of course important for every ML/DS project but what if I want different people to interact with my models, how can I make my model available for end-users? I can't send them jupyter notebooks right!. That's why I wanted to try my hands on complete end-to-end machine learning project.

1.2 PURPOSE

Training Of neutral networks is performed using back propagation to evaluate the prediction system. In the testing place approximately 95% accuracy is achieved on testing set. Practical use of data collected from previous record is time consuming. Low accuracy rate.

So to overcome this we are implementing Random Forests Algorithm in order to achieve accurate result in time. Machine learning is given a major priority in modern life in many application and in health care sector. Prediction is one of the area where machine learning plays a vital role, our topic is to predict heart diseases by processing patient’s dataset and a data of patients i.e., user of whom we need to predict the chance of occurrence of a heart diseases.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

S.NO	AUTHOR&YEAR	TITLE	DESCRPTION	ACCURACY
01.	Dhai Eddine Salhi, abdelkamel tari, Tahar Kechadi 2021	using machine learning for heart disease prediction	found neural network are easier to configure and obtain much good results	93%
02.	Xiao-yangao, Eman M.Anwar 2020	improving the accuracy for analyzing heart diseases prediction based on the ensemble method	the bagging ensemble learning algorithm with DT and PCA feature extraction method had achieved the best performance	83%
03.	Shriniket Dixit, Pilla Vaishno Mohan, Shrishail Ravi Terni 2013	Prediction of heart disease using machine learning algorithms	comparative analysis of the results of various machine learning algorithms	89%
04.	Nabaouia Louridi, Meryem Amar, Bouabid EI Ouahidi 2019	Identification of cardio vascular diseases using machine Learning	improved the quality of cardiovascular disease prediction using a better processing phase	90%
05.	Ankita Dewan , Meghna Sharma 2015	Prediction of heart disease using a hybrid technique in data mining Classification	important query of how to make useful information out of the data	85%

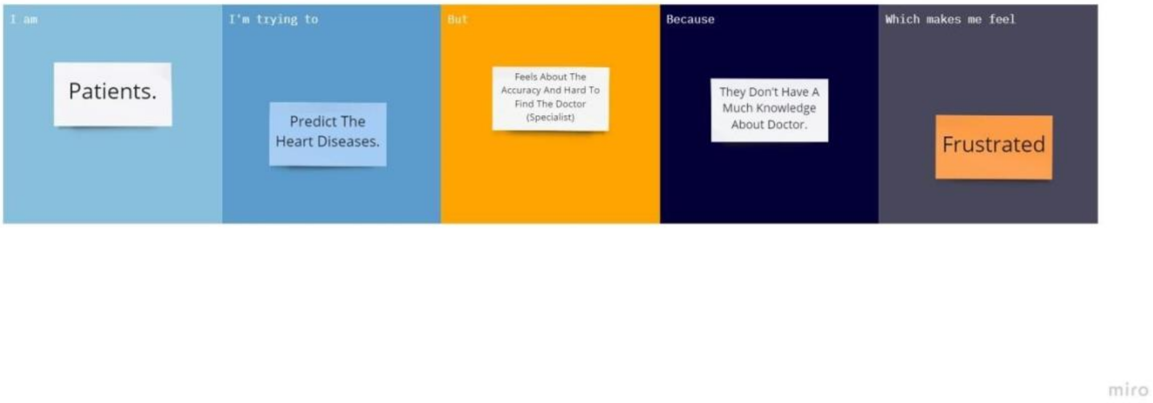
2.2 REFERENCE

- K. V. S. H. Gayatri Sarman
- Tenneti Madhu
- A. Mallikharjuna Prasad
- Vivek Sharma
- Sandeep Kumar
- Aarti Chugh
- Charu Jain

2.3 PROBLEM STATEMENT DEFINITION.

Heart Diseases remain the biggest cause of deaths for the last two decades. Recently computer technology and machine learning techniques are used to develop software to assist doctors in making appropriate decision of heart disease in an early stage. The diagnosis of heart disease depends on clinical and pathological data. Heart disease prediction system can assist medical professionals in predicting status of heart disease, based on the clinical data of patients.

Doctors may sometime fail to take an accurate decision in predicting heart disease risk level, therefore heart disease prediction systems are useful in such cases to get accurate results. There are many tools available for performing this task but all of them have some flaws. Most of the tools cannot handle big data and hence predicting heart disease would be a tedious task. In this project we are making an effort to predict the risk level of the huge datasets of patients.



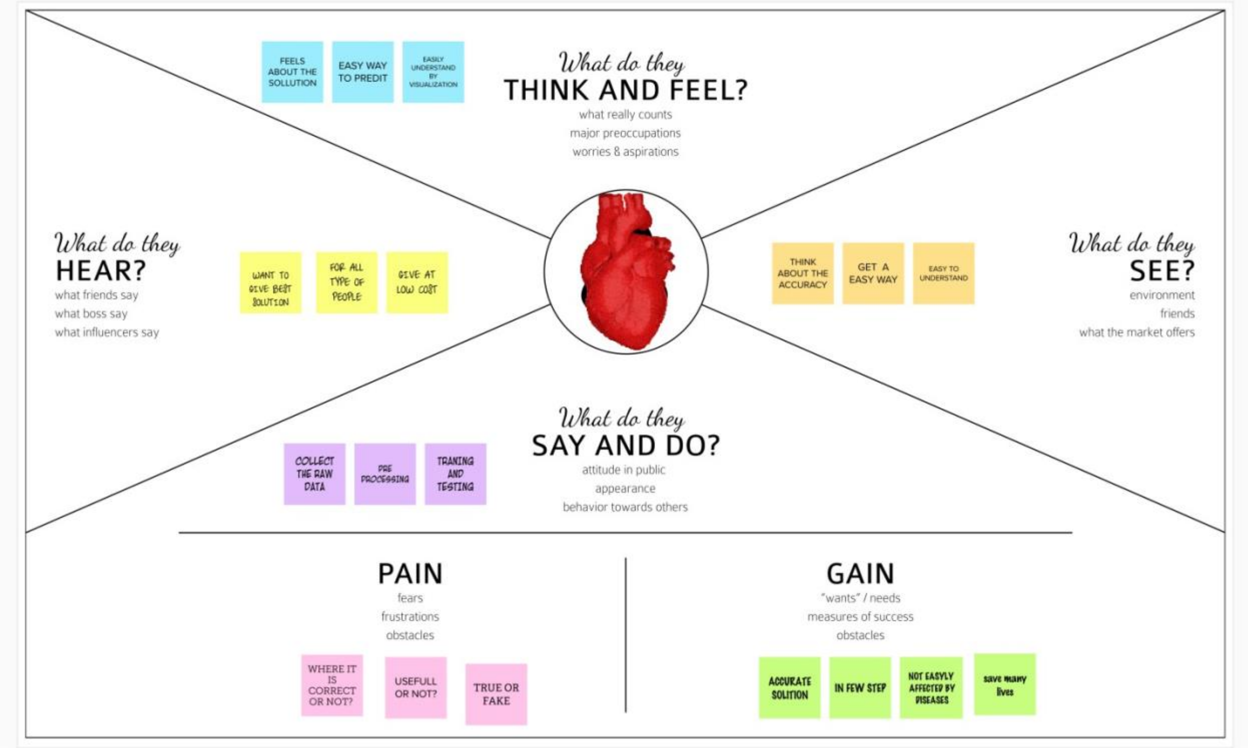
I am	I'm trying to	But	Because	Which makes me feel
Patients.	Predict The Heart Diseases.	Feels About The Accuracy And Hard To Find The Doctor (Specialist)	They Don't Have A Much Knowledge About Doctor.	Frustrated

3.IDEATION & PROPOSED SOLUTION

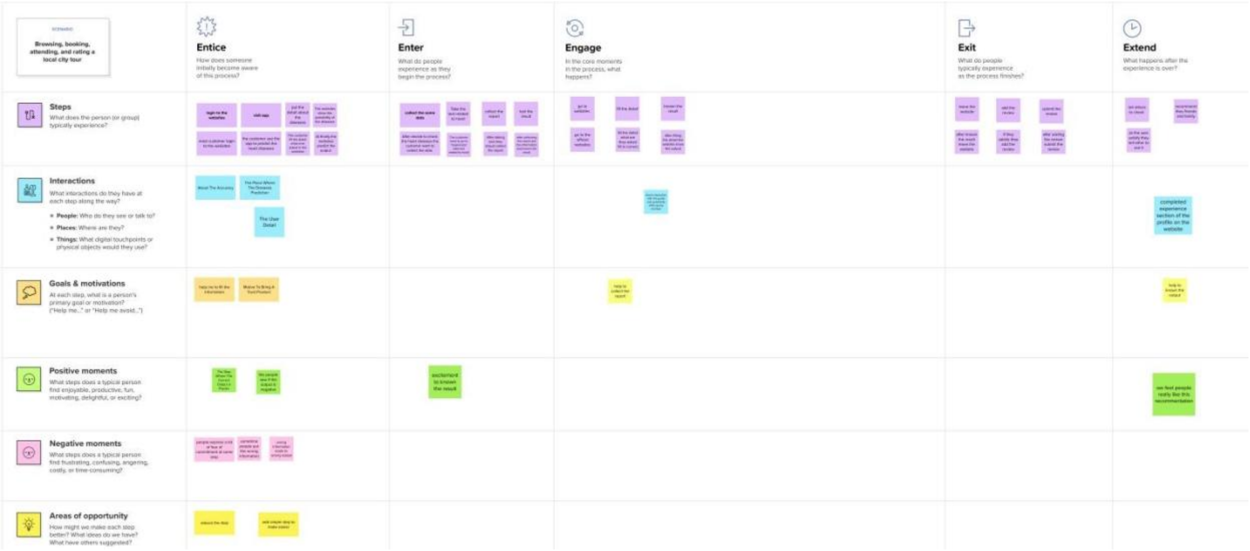
3.1 EMPATHY MAP CANVAS

It’s easy to jump straight into value proposition design. That is the core of your business and where the revenue or exchange of value will come from. However, trying to provide value to a misunderstood customer is very risky business. Do you have your blinkers on? Try using this canvas before you design your value proposition to make sure your offer nails exactly what your customer wants, needs, or may pleasantly surprise them! Keep asking yourself “why would they care?”. What problem are you solving? What opportunity are you creating?

In this empathy map what customer think and feels. this map shows the pain and gain of the customer and what do their hear about the problem. this is the easy way to understand the problem statement



3.2 IDEATION & BRAINSTORMING


Browsing, booking, attending, and rating a local city tour
Entice
How does someone initially become aware of this process?
Enter
What do people experience as they begin the process?
Engage
In the core moments in the process, what happens?
Exit
What do people typically experience as the process finishes?
Extend
What happens after the experience is over?
Steps
What does the person (or group) typically experience?
Interactions
What interactions do they have at each step along the way?
People: Who do they see or talk to?
Places: Where are they?
Things: What digital touchpoints or physical objects would they use?
Goals & motivations
At each step, what is a person's primary goal or motivation? ("Help me..." or "Help me avoid...")
Positive moments
What steps does a typical person find enjoyable, productive, fun, motivating, delightful, or exciting?
Negative moments
What steps does a typical person find frustrating, confusing, angering, costly, or time-consuming?
Areas of opportunity
How might we make each step better? What ideas do we have? What have others suggested?
The User Detail
excitement to known the result
completed experience section of the profile on the website
we feel people really like this recommendation

Brainstorming is **a method of generating ideas and sharing knowledge to solve a particular commercial or technical problem, in which participants are**

encouraged to think without interruption. Brainstorming is a group activity where each participant shares their ideas as soon as they come to mind.

3.3 PROPOSED SOLUTION

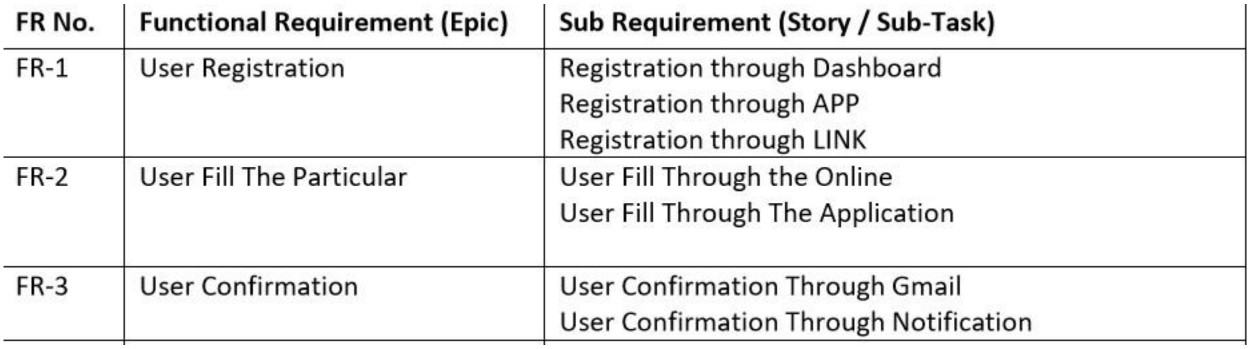
To overcome this we are implementing random forests in order to achieve accurate results in less time . Machine learning is given a major priority in modern life in many application and in healthcare sector . Prediction is one of areas where machine learning plays a vital role , Our topic is to predict heart diseases by processing patient’s dataset and a data of patients i.e., user of whom we need to predict the chances of occurrence of the heart diseases.

Our aim is to build an application of heart diseases prediction system using Flask and deployed on Heroku. A csv file is given as input . After the successful completion of operation the result is predicted and displayed.

4. REQUIREMENT ANALYSIS

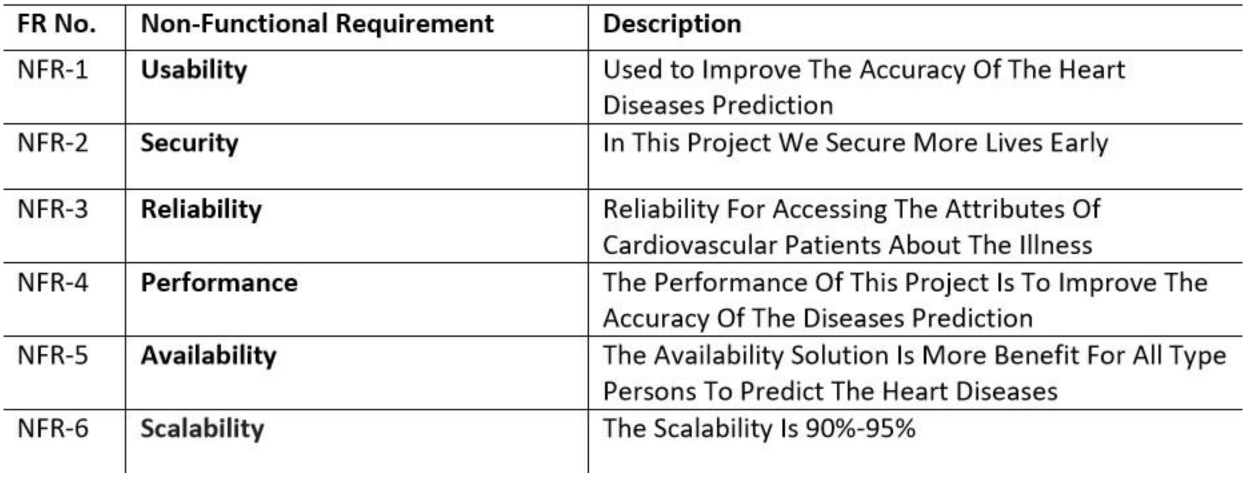
A functional requirement defines a system or its component. A non-functional requirement defines the quality attribute of a software system. It specifies “What should the software system do?” It places constraints on “How should the software system fulfill the functional requirements?”

4.1 FUNCTIONAL REQUIREMENT



FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Dashboard Registration through APP Registration through LINK
FR-2	User Fill The Particular	User Fill Through the Online User Fill Through The Application
FR-3	User Confirmation	User Confirmation Through Gmail User Confirmation Through Notification

4.2 NON-FUNCTIONAL REQUIREMENT

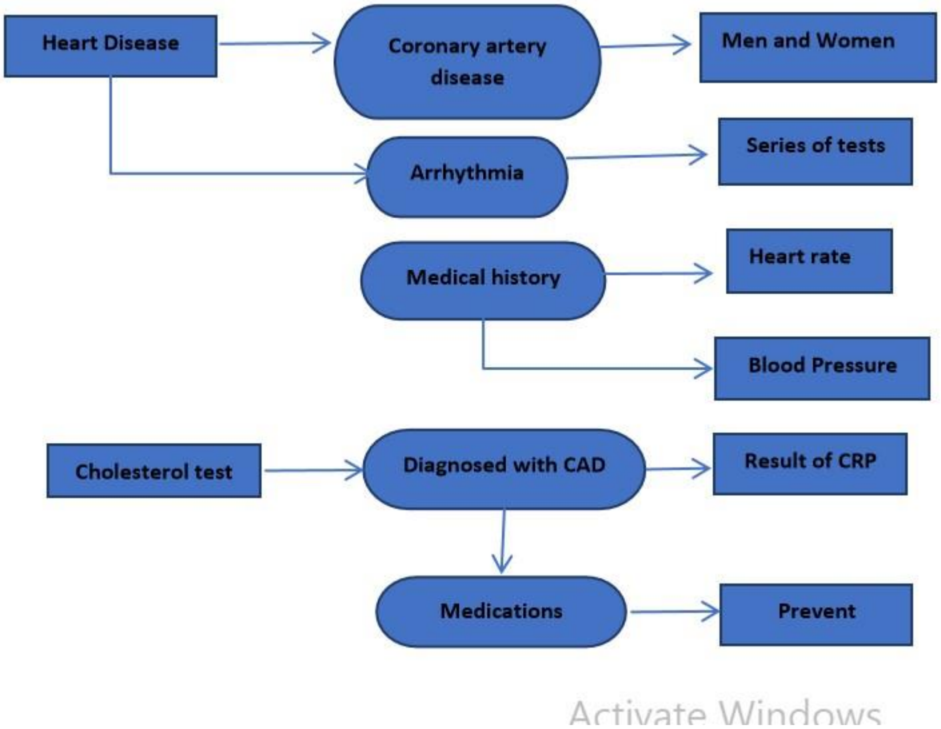


FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Used to Improve The Accuracy Of The Heart Diseases Prediction
NFR-2	**Security**	In This Project We Secure More Lives Early
NFR-3	**Reliability**	Reliability For Accessing The Attributes Of Cardiovascular Patients About The Illness
NFR-4	**Performance**	The Performance Of This Project Is To Improve The Accuracy Of The Diseases Prediction
NFR-5	**Availability**	The Availability Solution Is More Benefit For All Type Persons To Predict The Heart Diseases
NFR-6	**Scalability**	The Scalability Is 90%-95%

5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

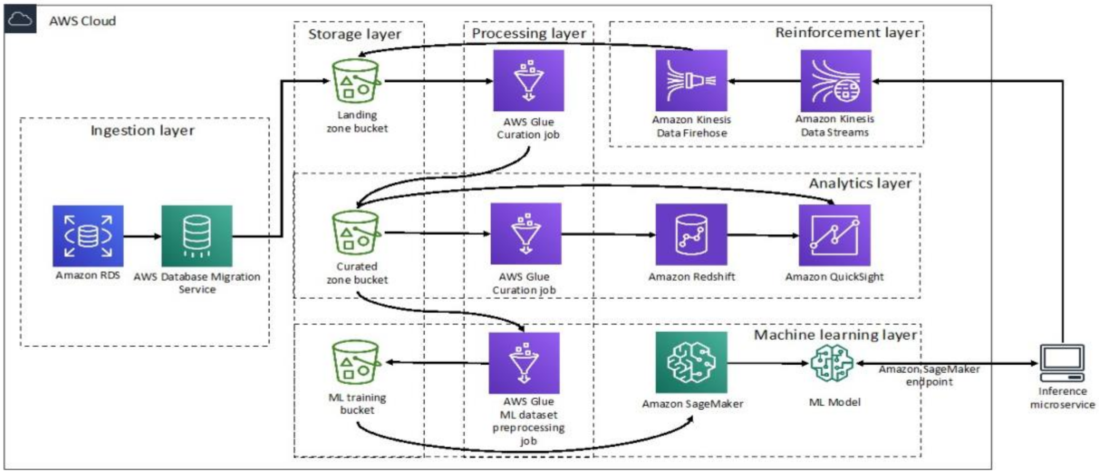
A data flow diagram (DFD) is **a graphical or visual representation using a standardized set of symbols and notations to describe a business's operations through data movement**.



In this flow diagram we are showing that the heart diseases prediction.

5.2 SOLUTION AND TECHNICAL ARCHITECTURE

A solution architecture (SA) is **an architectural description of a specific solution**. SAs combine guidance from different enterprise architecture viewpoints (business, information and technical), as well as from the enterprise solution architecture (ESA).



5.3 USER STORIES

Story 1:

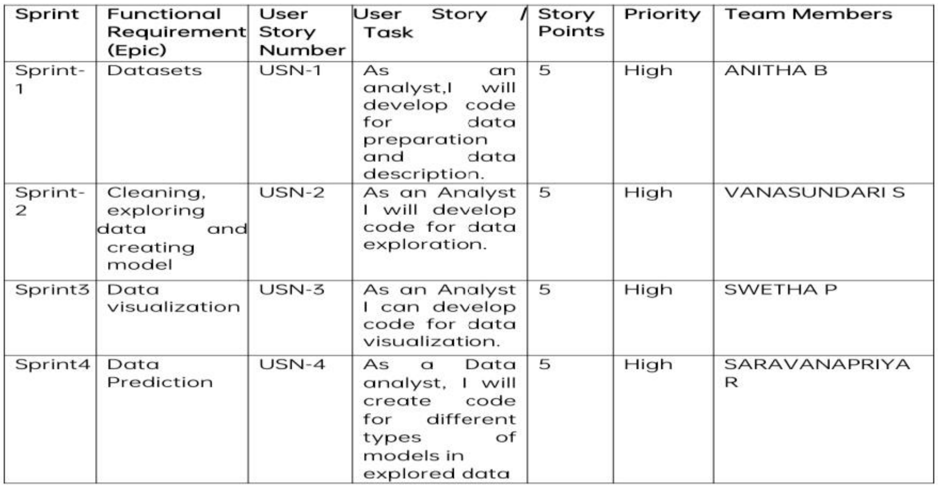
As an aging individual, I want an application so that I could predict my cardiac health. Let’s break this down one step further, As the user is an aging individual, we are building a heart disease-predicting application that enables the user to predict their cardiac health immediately within a few seconds. The app has the user login and signup for the authentication of information, and it uses the Logistic Regression algorithm to predict the result. We have visualized the user’s query for their requirement only concerning the Age and Cholesterol of the user. The prediction gives a result if the disease could be present or not. As a working person, I want an application so that I can predict my own cardiac health even while I am at work.

Story 2:

When we address this user issue, As the user is a working person, our heart disease predicting application would enable the user to predict his health just by simply typing the values into the boxes given. The process is very simple, it takes only a few seconds to view the results of the process. The app has the user login and signup for the authentication of information, and it uses the Logistic Regression algorithm to predict the result. The users can customize their range of values and can visualize the entire result in a graphical representation. It allows the user to have a clear knowledge of what must be done and what to be neglected.

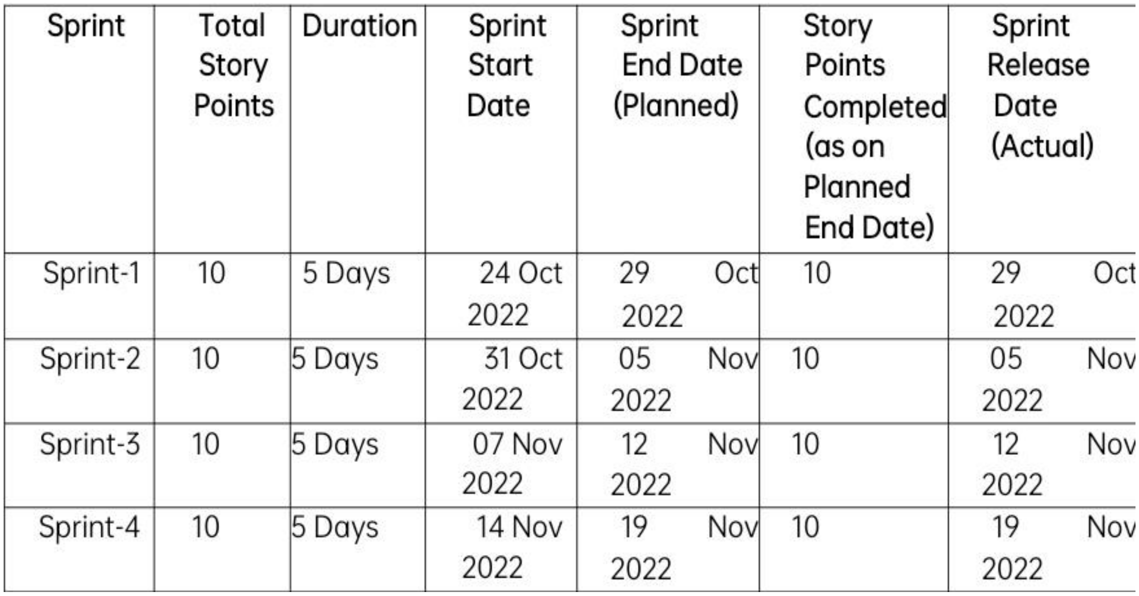
6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION



Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Datasets	USN-1	As an analyst,I will develop code for data preparation and data description.	5	High	ANITHA B
Sprint-2	Cleaning, exploring data and creating model	USN-2	As an Analyst I will develop code for data exploration.	5	High	VANASUNDARI S
Sprint3	Data visualization	USN-3	As an Analyst I can develop code for data visualization.	5	High	SWETHA P
Sprint4	Data Prediction	USN-4	As a Data analyst, I will create code for different types of models in explored data	5	High	SARAVANAPRIYA R

6.2 SPRINT DELIVERY SCHEDULE



Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	5 Days	24 Oct 2022	29 Oct 2022	10	29 Oct 2022
Sprint-2	10	5 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	10	5 Days	07 Nov 2022	12 Nov 2022	10	12 Nov 2022
Sprint-4	10	5 Days	14 Nov 2022	19 Nov 2022	10	19 Nov 2022

9. RESULTS

9.1 PERFORMANCE METRICS

Metrics are measurements and parameters obtained throughout the quality assurance procedure. They may make reference to several test types. As you might have guessed, performance testing data gives you the ability to evaluate the efficiency of performance testing. Alternatively said, these measurements demonstrate how well software reacts to user scenarios and manages user flow in real time.

The following two categories of data are appropriate:

1. Measurements are data that are kept track of when testing, such as how long it takes to react to a request.
2. Metrics, which include various types of percentages, average indicators, and other metrics, are computations performed with the aid of certain formulas.

Accuracy Percentages:

The number of wrong predictions on the test set as a whole divided by all of the test set predictions yields the error rate. Since accuracy and error rate are complementary quantities, we can always compute one from the other

$$\text{Accuracy} = 1 – \text{Error Rate}$$

$$\text{Error rate}=1-\text{Accuracy}$$

Logistic Regression: 0.82	KNN: 0.61	Naïve Bayes: 0.77	Decision Tree: 0.79

10.ADVANTAGE & DISADVANTAGE

ADVANTAGE

- The advantage of this model are high performance and accuracy rate.
- It is very flexible and high rates of success are achieved
- The application when implemented using random forests has more accuracy rate when compare to other algorithm. In this system, we achieve around 98%.

DISADVANTAGES

- Data analytics techniques do not help to. Provide effective decision making
- Cannot handle enormous datasets
- Prediction of cardiovascular disease results is not accurate

11. CONCLUTION

The primary objective of the proposed algorithm is to minimize Makespan and improve fitness function. Improving the load balance process through task Scheduling can result in efficient utilization of cloud resources. The objective of this proposed work was to provide an enhanced load balancing algorithm. Result proved that our algorithm reduce makespan and provide efficient resources utilization of compared to existing dynamic LBA (load balancing algorithm). It also shows that the proposed algorithm can function in a dynamic cloud environment where user requests arrive in random order and where there are many changes in the length of the user requests. The algorithm is also to handle large size requests compared to the existing approach.

12. FUTURE SCOPE

In the future, various other metrics like throughput, average time, resources utilizing, waiting time, etc. can be considered. In the future, author will work to optimize the cloud resources further and enhance cloud-based application performance, such as considering more SLA (service level agreement) parameters. For example, the algorithm will be tested based on the number of violation and the migration count for better performance. Also, the algorithm will be comprehensively compared to other existing algorithm in the literature.

13. APPENDIX

PYTHON

Python is a computer programming language often used to **build websites and software, automate tasks, and conduct data analysis**. Python is a generalpurpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

SOURCE CODE

```
Dataset = pd.read_csv(‘Heart_Disease_Prediction.csv’,sep=’,’,encoding=”utf-8”)
Dataset.shape
Type(dataset)
Dataset.info()
Dataset.columns
Import pandas as pd
Import numpy as np
Import Matplotlib.pyplot as plt
From matplotlib import rcParams
from matplotlib.cm import rainbow import seaborn as sns %matplotlib
inline
target =df[‘Heart Disease’].map({‘Presen ce’:1, ‘Absence’:0})
inputs = df.drop([‘Heart Disease’], axis=1)
plt.suptitle(“CorrelationMap/Pearsoncorrelation coefficient”)
sns.heatmap(df.iloc[:,1 :-1].corr())
plt.show()
data=dfsns.barplot(x=data.Age.value_counts()[:10].index,y=data.Age.value_counts()[:1
0].values) plt.show()
plt.suptitle(“Age”) sns.scatterplot(data=df, x=’Age’, y=np.zeros(len(df[‘Age ‘])),
hue=target) plt.show()
minAge=min(data.Age)
maxAge=max(data.Age)
meanAge=data.Age.m ean()
print(‘MinAge :’,minAge)
Print(‘MaxAge :’,maxAge)
Print(‘MeanAge :’,meanAge)
Plt.suptitle(‘Age histogram’,
Fontweight=’heavy’)
Plt.title(‘The age average is around 54’)
sns.histplot(data=df, x=’Age’
plt.show()
labels = [‘Male’, ‘Female’]
Order = df[‘Sex’].value_counts().index
Plt.figure(figsize=(10,5))
plt.suptitle(“Sex (Gender)”)
plt.subplot(1,2,1) plt.title(‘Pie chart’) plt.pie(df[‘Sex’].value_ counts(),labels=labels,
textprops={‘fontsize’:12})
plt.subplot(1,2,2) plt.title(‘Count plot’) sns.countplot(x=’Sex’, data=df, order=order)
plt.xticks([0, 1], labels)
plt.show()
print(df[‘Sex’].value_c ounts()) print(“It can be noticed that predictor (Gender)
is imbalance”)
labels = [“typical angina”, “atypical angina”, “non-anginal pain”, “asymptomatic”]
order = df[‘Chest pain type’].value_counts().index
```

```
plt.figure(figsize=(10,5))
Plt.suptitle(“Chest pain Type”)
Plt.subplot(1,2,1) plt.title(‘Pie chart’)
plt.pie(df[‘Chestpaintype’].value_counts(), textprops={‘fontsize’:1 2})
plt.subplots_adjust(left =0.125)
plt.subplot(1,2,2) plt.title(‘Count plot’)
sns.countplot(x=’Chest pain type’, data=df, order=order) plt.xticks([0,1,2,3],labels,
rotation=45)
plt.show()
df[‘Chest pain type’].value_counts()
knn_classifier= KNeighborsClassifier( n_neighbors=31,leaf_s ize=30)
knn_classifier.fit(X_tra in,y_train)
Y_pred_knn = knn_classifier.predict( X_test)
Score_knn = round(accuracy_score( Y_pred_knn,y_test)*1 00,2) score_knn
Y_pred_knne    = knn_classifier.predict( X_test)
Plt.figure(figsize=(10, 8))
CM=confusion_matrix (y_test,y_pred_knne) sns.heatmap(CM, annot=True)
TN = CM[0][0] FN = CM[1][0]
TP = CM[1][1] FP = CM[0][1]
Specificity = TN/(TN+FP)
Loss_log =log_loss(y_test, y_pred_knne) acc= accuracy_score(y_test, y_pred_knne)
roc=roc_auc_score(y_tEst, y_pred_knne)
Prec = precision_score(y_test, y_pred_knne)
Rec = recall_score(y_test, y_pred_knne)
f1 = f1_score(y_test, Y_pred_knne)
mathew   = matthews_corrcoef(y_t est, y_pred_knne)

model_results =pd.DataFrame([['KNearest Neighbors ',acc, prec,rec,specificity,
f1,roc, loss_log,mathew]],columns = ['Model','Accuracy','Precision',
'Sensitivity','Specificit y', 'F1 Score','ROC','Log_Los s','mathew_corrcoef'])

model_results

from sklearn import metrics

Y_pred_knn = np.around(Y_pred_knn)
Print(metrics.classificat
Ion_report(y_test,Y_pr ed_knn))
From Sklearn.metrics._plot.r
Oc_curve
```

```
import Plot_roc_curve
Plot_roc_curve(knn_cl
Assifier,X_test,y_test)
Plot_roc_curve plt.xlabel(‘False Positive Rate’) plt.ylabel(‘True Positive Rate’)
plt.title(‘Receiver Operating Characteristic Curve’);
Plt.savefig(“KNN.png”)
```

GitHub & Project Demo Link:

https://github.com/IBM-EPBL/IBM-Project-32980-1660213350