Visualizing and Predicting Heart Diseases with an InteractiveDash Board

Team ID: PNT2022TMID02974

Faculty Mentor:

Ratheeshkumar AM

Team Leader: L Preethi

Team Member: S R Kanimozhi

Team Member: V Kaviya

Team Member: V OviyaSvapna

PROJECT REPORT

1. INTRODUCTION

1.1 Project Overview:

A simple web application which uses Machine Learning algorithms to Predict the heart condition of a person by providing some inputs about the Person health like age, blood pressure, cholesterol, sex etc.

Heart related diseases or Cardiovascular Diseases are the main reason for a huge number of deaths in the world over the last few years and has emerged as the most life threatening disease, not only in India but in whole world. Many researchers have been trying to find a solution to predict the heart disease and to help care industry using several machine learning techniques. This indicates a need of reliable, accurate and feasible system to continuously monitor and diagnose for CVD for timely action and treatment. The major cause of death in developed world is heart diseases. To analyze and predict which patients are most likely to suffer from heart disease in the near future we have to find out some solution.

1.2 Purpose:

So for the above mentioned problem statement, we can create or develop an interactive dashboard for visualizing the people who might have the possibilities are high chances of getting Cardiovascular Disease through a collection of dataset. Most of all the heart diseases can be identified and treated using ECG in medical field, and the theory of curing can be in handwritten and they get research to it and finally implement it in practical.

Prediction is one of area where machine learning plays a vital role, our topic is to predict heart diseases by processing patient's dataset and data of patients i.e., user of whom we need to predict the chance of occurrence of a heart disease.

2. LITERATURE SURVEY

2.1 Existing problem:

The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to heart diseases. About 25% deaths in the age group of 25-69 year occur because of heart diseases. In urban areas, 32.8% deaths occur because of heart ailments, while this percentage in rural areas is 22.9%WHO estimated by 2030, almost 23.6 million people will be affected by Cardiovascular disease. The treatment for this disease is quite high and not affordable by most of the patients particularly in India. The heart disease prediction can be carried out using various algorithms as Support Vector Machine (SVM) classifier, Decision Tree and

Random Forest algorithm. But these are supervised ML algorithms. Each one of them has their advantage and disadvantages.

2.2 References:

PAPER 1: Prediction of Heart Disease Using Machine Learning algorithms by Mr.Santhana Krishnan.J, Dr.Geetha.S.

In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of occurring
heart disease of the patients in terms of percentage. This is performed through data mining classification techniques.
ALGORITHMS USED: Decision Tree and Naive Bayes.

PAPER 2: A Hybrid Machine Learning Approach for Prediction of Heart Diseases by Sanchayita Dhar, Pritha Datta, Ankur Biswas
Tanusree Dey, Krishna Roy.

In this paper, to develop a prediction system that be capable to envisage heart diseases based on measurements, are
extracted from THE ERIC laboratory consisting of 209 test cases.
ALGORITHMS USED: Naive Bayes, Decision Tree and Random Forest.

2.3 Problem Statement Definition:

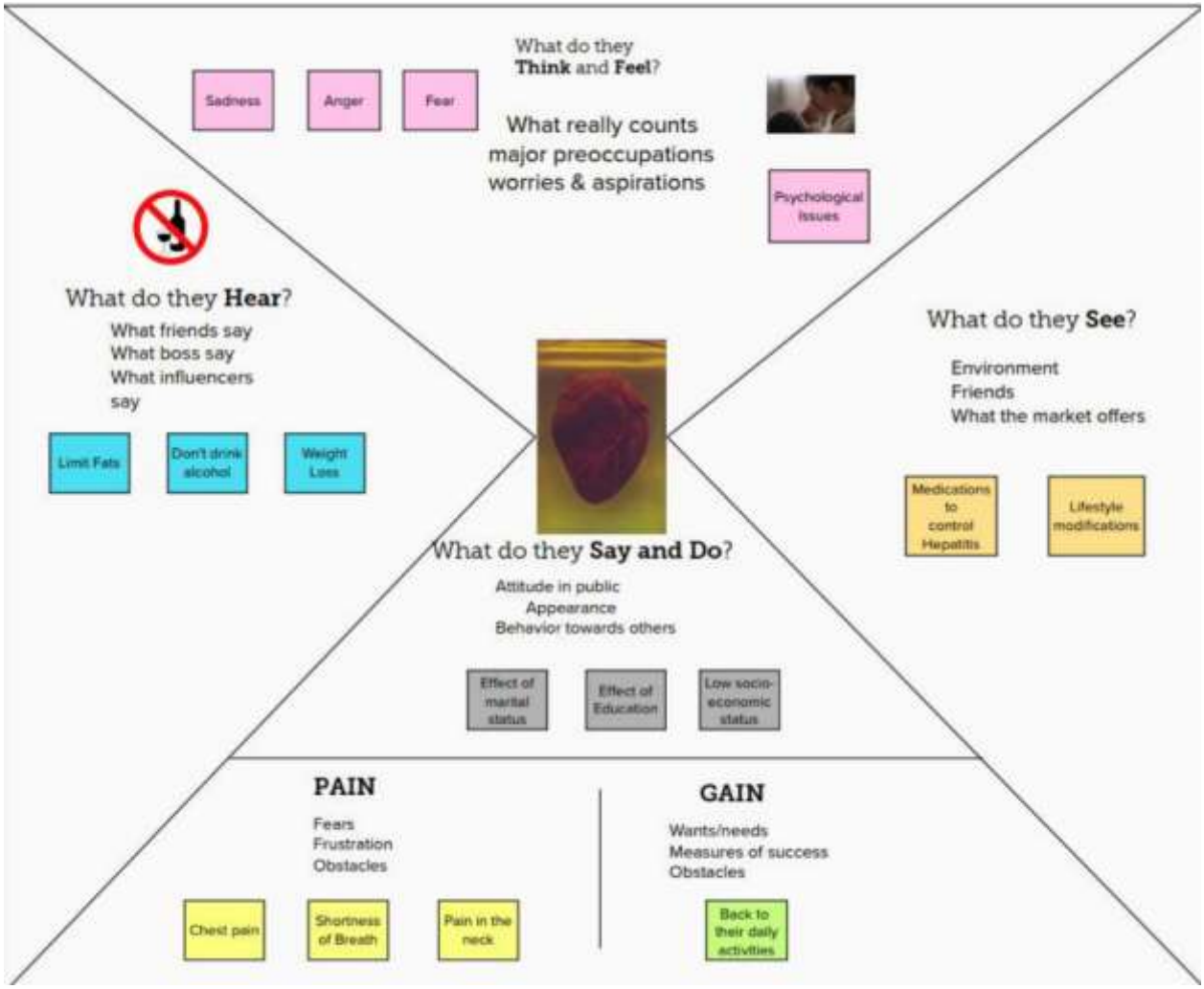
Heart Diseases remain the biggest cause of deaths for the last two decades. Recently computer technology and machine learning techniques are used to develop software to assist doctors in making appropriate decision of heart disease in an early stage. The diagnosis of heart disease depends on clinical and pathological data. Heart disease prediction system can assist medical professional 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 may 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 tedious task. In this project we are making an effort to predict the risk level of the huge datasets of patients.

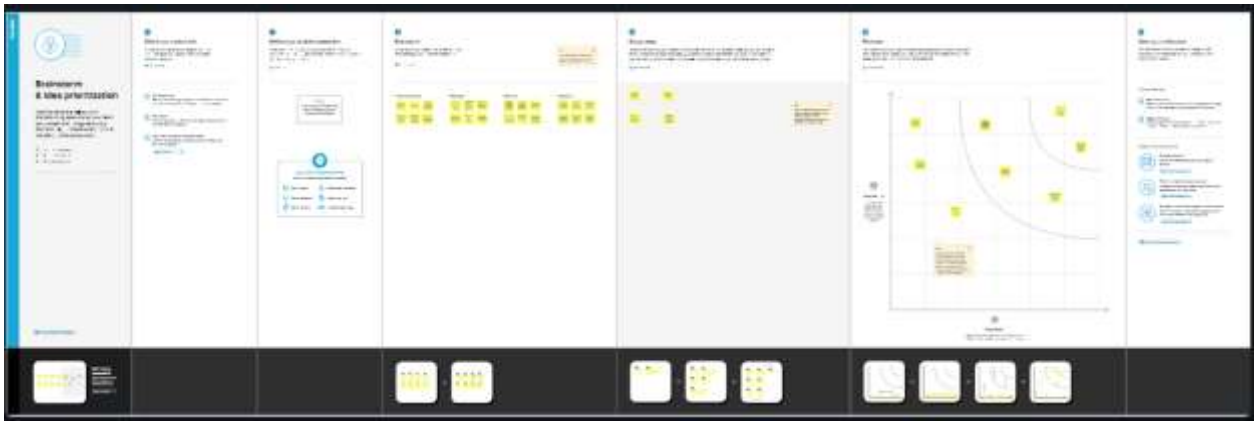
3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS:

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



3.2 IDEATION &BRAINSTORMING



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 predict the heart attack disease quickly and efficiently

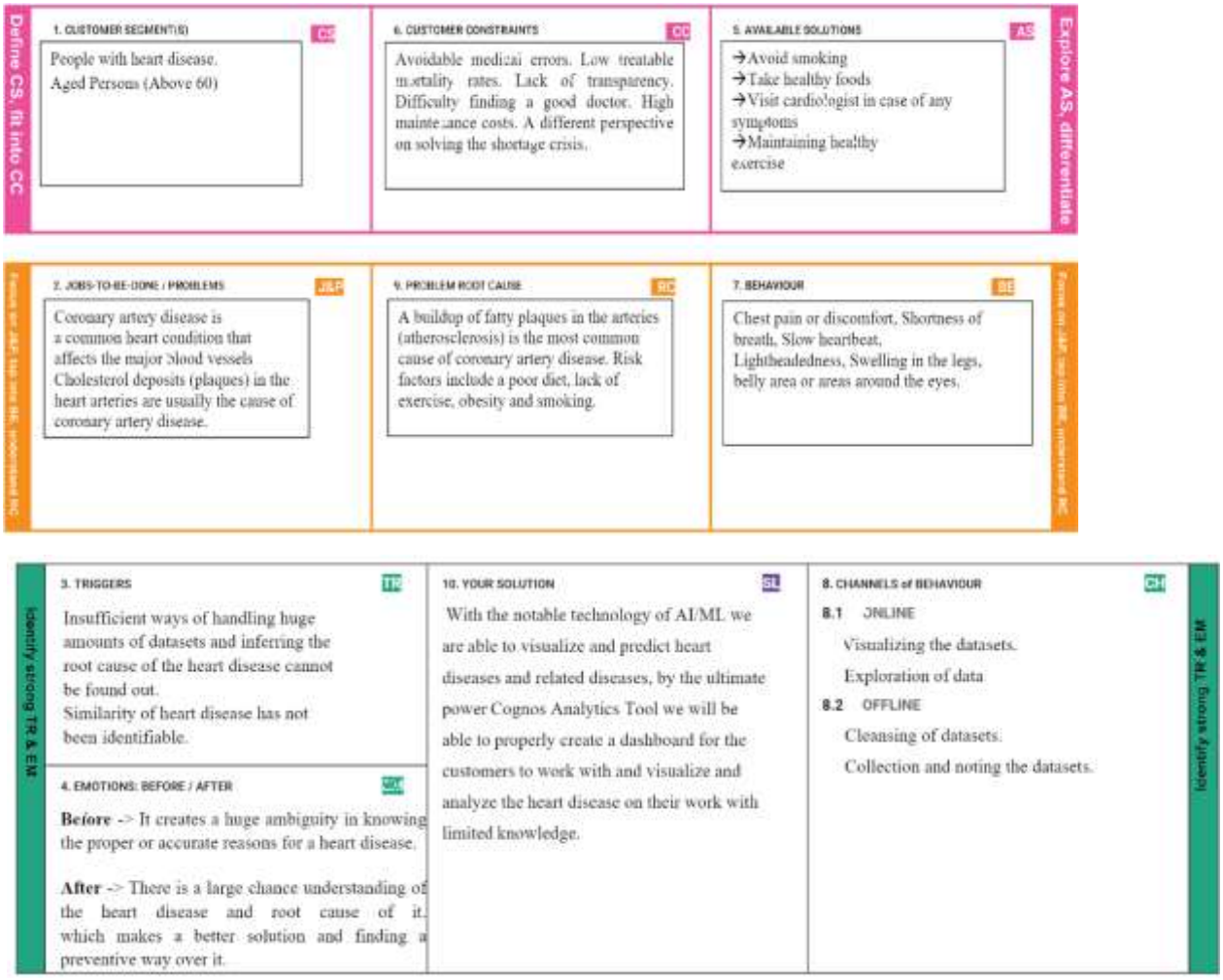
It helps in reducing treatment costs by providing effective treatments.

To develop a system that can be quite low and will be affordable by most of the patients around the world. Machine learning is given a major priority in modern life in many applications and in healthcare sector. Prediction is one of areas where machine learning plays a vital role.

Our aim is to develop an interactive dashboard to predict the heart disease. A csv file is given as input. After the successful completion of operations the result is predicted and displayed.

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 for it actually solves the customer's problem.



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:

Following are the functional requirements of the proposed solution.

FR No.	Function 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 application ○ User fill through online
FR-3	User Confirmation	○ User confirmation through Gmail ○ User confirmation through Notification ○ User confirmation through Call

4.2 Non-Functional Requirements:

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

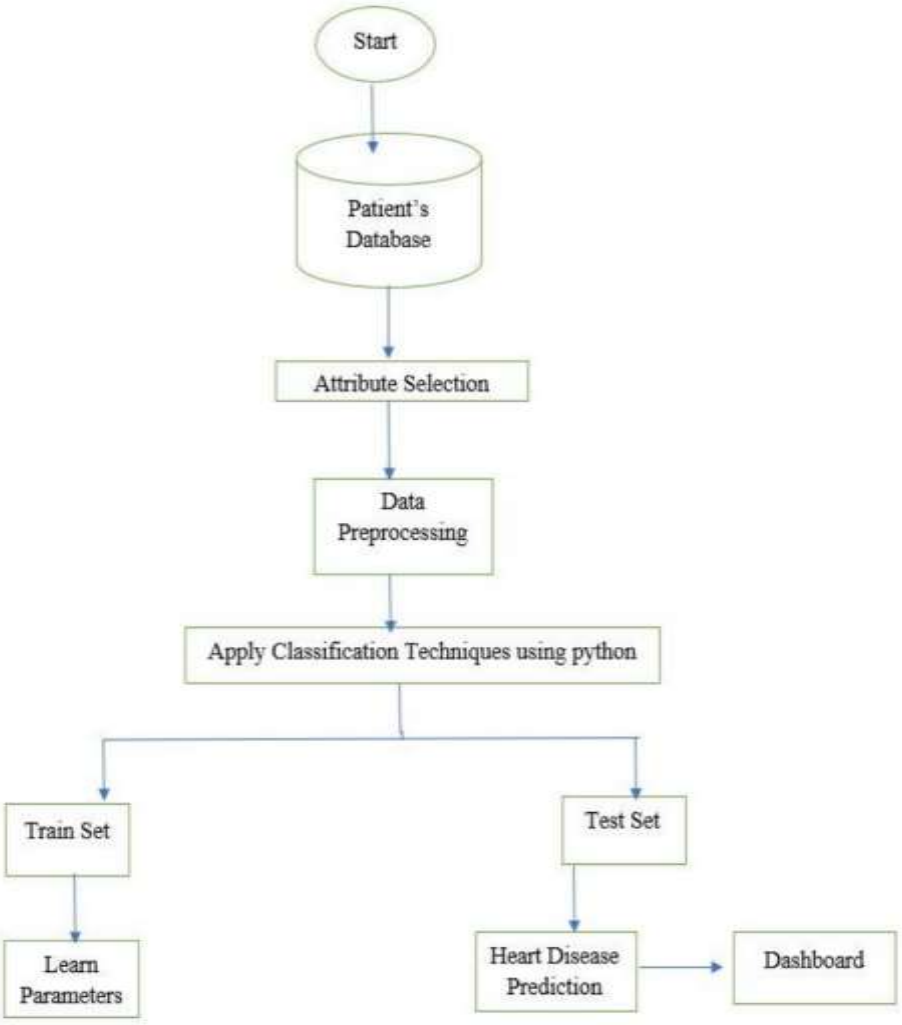
NFR NO	Non-functional requirements	Description
NFR-1	Usability	It is used to improve the prediction of heart disease of a patient
NFR-2	Security	Using this project, we can secure more lives early
NFR-3	Reliability	It is more reliable to access the patient's cardiovascular attributes to measure the illness.

NFR-4	Performance	The performance of this project is to improve the accuracy of disease prediction
NFR-5	Availability	The project is designed so that it will be available for all people to identify their disease.
NFR-6	Scalability	The Scalability is 90%

5. PROJECT DESIGN

5.1 Data Flow Diagram:

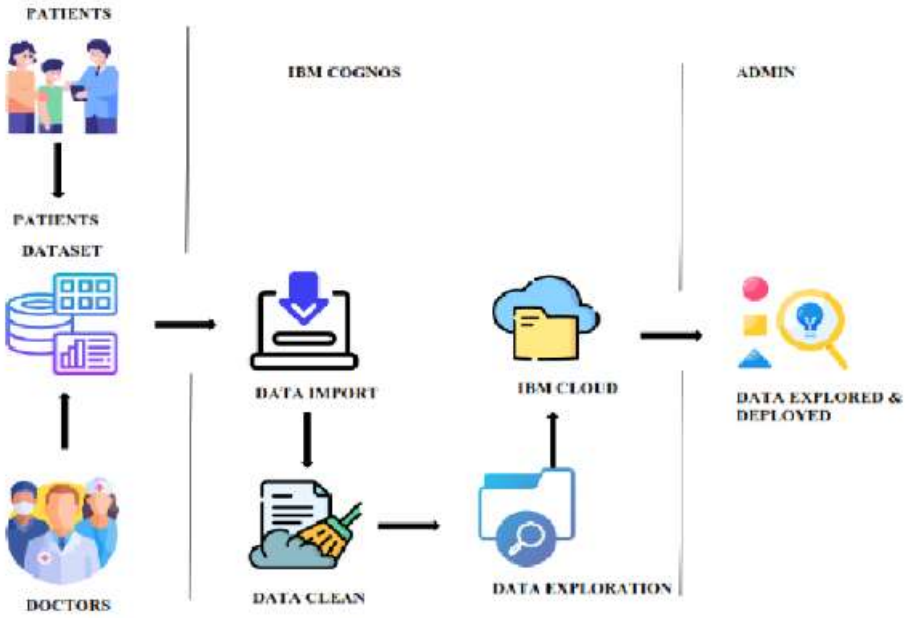
A data flow diagram is a graphical or visual representation using a standardized set of symbol and notations to describe a business's operations through data movement.



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

5.2. Solution and Technical Architecture:

A solution architecture is an architectural description of a specific solution. SAs combine guidance from different enterprise architecture view points.



5.2 User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story/ Task	Acceptance Criteria	Priority	Release
Customer	Registration	USN-1	As a user, I can register for the application by entering my email, password and confirm my password	I can access my account / dashboard	High	Sprint -1
		USN-2	As a user, I can receive a confirmation email while I registered for the application	I can receive confirmation Mail	High	Sprint - 1
		USN-3	As a user, I can register for application using	I can access the dashboard using facebook login	Low	Sprint-1

			facebook			
		USN-4	As a user, I can register forthe application through gmail	I can register andaccess the dashboard using gmail	Mediu m	Sprint-1
	Login	USN-5	As a user, I can log into application by entering email and password	I can register & access dashboard with gmail login and password	High	Sprint-1
Customer	Dashboard	USN-6	As a user, I can view a patient's medical analysis and accuracy of disease prediction.	I can view medical analysis in the dashboard	High	Sprint-2
		USN-7	As auer, I can analysis the risk factors and preventiontips	I can read the prevention tips	High	Sprint-2
		USN-8	As a user, I can check thetreatment options	The user can readthe treatment options	High	Sprint-2
Customer Care Executi ve	HelpDesk	USN-9	As a customer care, he/she can view the customerqueries	I can post my queries in the dashboard	High	Sprint-3
		USN-10	As a customer care executive, he can answer thequeries	I can get supportfrom helpdesk	High	Sprint-3

Administrator	UserProfile	USN-11	As an admin, he can update the health details of users	I can view my health details	High	Sprint-4
		USN-12	As an admin, he can add ordelete users	I can access my account when logged in	High	Sprint-4
		USN-13	As an admin, he can update the risk and prevention tips	I can update the risk and prevention tips	High	Sprint-4

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation:

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint1	Data Preprocessing and Exploratory Data Analysis (EDA)	USN -1	Data Cleaning is implemented to check whether there are any null values, or any outliers are found	10	Medium	Preethi L Kanimo_hi S R Kaviya V Oviya Svapna V
		USN-2	Testing and training the data model is implemented using Jupyter Notebook	10	High	Preethi L Kanimozhi S R Kaviya V Oviya Svapna V
Sprint2	Working with dataset	USN-3	1. Working with dataset. 2. Understand dataset 3. Load the Dataset 4. Explore the Data 5. Visualize the Data	20	Medium	Preethi L Kanimozhi S R Kaviya V Oviya Svapna V
Sprint3	Data Visualization	USN-4	We plan to create various graphs and charts to highlight the	20	High	Preethi L Kanimozhi S R Kaviya V Oviya Svapna V

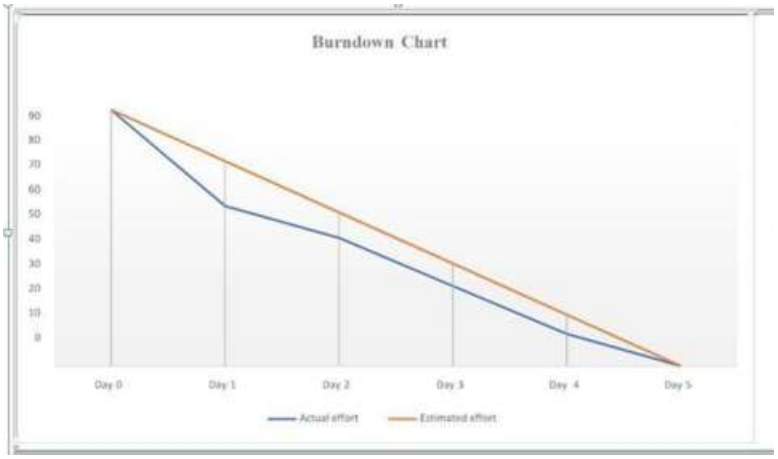
			insights and visualizations with given attributes			
Sprint4	Dashboard	USN-5	Dashboard showing different types of visuals	15	High	Preethi L Kanimozhi S R Kaviya V Oviya Svapna V
		USN-6	User can be able to generate Report and Story	5	Medium	Preethi L Kanimozhi S R Kaviya V Oviya Svapna V

6.2 Sprint Delivery Schedule:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date	Story points completed	Sprint Release Date
Sprint-1	20	6 Days	04 November 2022	06 November 2022	20	10 November 2022
Sprint-2	20	6 Days	06 November 2022	08 November 2022	20	10 November 2022
Sprint-3	20	6 Days	08 November 2022	10 November 2022	20	10 November 2022
Sprint-4	20	6 Days	10 November 2022	12 November 2022	20	10 November 2022

6.3 Report From Jira:

A burndown chart shows the amount of work that has been completed in an epic or sprint, and the total work remaining. Burndown charts are used to predict your team's likelihood of completing their work in the time available.



7. CODING & SOLUTIONING

7.1 Feature 1

```
categorial = [('Sex', ['female', 'male']),('Chest pain type', ['typical angina', 'atypical angina', 'non-anginal pain', 'asymptomatic']),
        ('FBS over 120', ['fbs > 120mg', 'fbs < 120mg']),
        ('EKG results', ['normal', 'ST-T wave', 'left ventricular']),
        ('Exercise angina', ['yes', 'no']),
        ('Slope of ST', ['upsloping', 'flat', 'downsloping']),
        ('Thallium', ['normal', 'fixed defect', 'reversible defect'])]
def plotGrid(isCategorial):
    if isCategorial:
        [plotCategorial(x[0], x[1], i) for i, x in enumerate(categorial)]
    else:
        [plotContinuous(x[0], x[1], i) for i, x in enumerate(continuous)]
def plotCategorial(attribute, labels, ax_index):
    sns.countplot(x=attribute, data=dataset, ax=axes[ax_index][0])
    sns.countplot(x='Heart Disease', hue=attribute, data=dataset, ax=axes[ax_index][1])
    avg = dataset[[attribute, 'Heart Disease']].groupby([attribute], as_index=False).mean()
    sns.barplot(x=attribute, y='Heart Disease', hue=attribute, data=dataset, ax=axes[ax_index][2])

    for t, l in zip(axes[ax_index][1].get_legend().texts, labels):
        t.set_text(l)
    for t, l in zip(axes[ax_index][2].get_legend().texts, labels):
        t.set_text(l)
fig_categorial, axes = plt.subplots(nrows=len(categorial), ncols=3, figsize=(15, 30))
plotGrid(True)
```

7.2 Feature 2

```
continuous = [('BP', 'blood pressure in mm Hg'),
              ('Cholesterol', 'serum cholestoral in mg/d'),
              ('Max HR', 'maximum heart rate achieved'),
              ('ST depression', 'ST depression by exercise relative to rest'),
              ('Slope of ST', '# major vessels: (0-3) colored by flourosopy')]
def plotContinuous(attribute, xlabel, ax_index):
   sns.distplot(dataset[[attribute]], ax=axes[ax_index][0])
   axes[ax_index][0].set(xlabel=xlabel, ylabel='density')
   sns.violinplot(x='Heart Disease', y=attribute, data=dataset, ax=axes[ax_index][1])
fig_continuous, axes = plt.subplots(nrows=len(continuous), ncols=2, figsize=(15, 22))
plotGrid(isCategorial=False)
```

8. TESTING

8.1 Test Cases:

A test case is nothing but a series of step executed on a product, using a predefined set of input data, expected to produce a pre-defined set of outputs, in a given environment. I t describes “how” to implement those test cases. Test case specifications are useful as it enlists the specification details of the items.

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provide a way to check the functionality of component, sub assemblies, assemblies and

a finished product. It is the process of exercising software with the intent of ensuing that the software system meets its requirement and does not fail in an unacceptable manner. There are various types of testing. Each test and does not fail in an unacceptable manner. There are various types of testing. Each test type addressing a specific testing requirement.

8.2 User Acceptance Testing:

User acceptance testing is a critical phase of any project and requires significant participant by the end user. It also ensures that the system meets the functional requirement.

9. RESULT

9.1 Performance Metrics

Classification Report

	precision	recall	f1-score	support
0	0.98	1.00	0.99	132
1	1.00	0.98	0.99	125
accuracy			0.99	257
macro avg	0.99	0.99	0.99	257
weighted avg	0.99	0.99	0.99	257

Accuracy: 98.83%

10. ADVANTAGE & DISADVANTAGE

ADVANTAGE:

- The advantage of this model are high performance and accuracy rate.
- It is very flexible ad high rates of success are achieved.
- The application when implemented using Naïve bayes has more accuracy when compare to other algorithms.

11. CONCLUSION

The primary objective of the proposed algorithm is to minimize Make span 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 make span 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 manychanges 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 themigration 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, automated tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

SOURCE CODE:

```
y = dataset["Heart Disease"]

rcParams['figure.figsize'] = 8,6

plt.bar(dataset['Heart Disease'].unique(), dataset['Heart Disease'].value_counts(), color = ['blue', 'green'])

plt.xticks(["Presence","Absence"])

plt.xlabel('Heart Disease(Presence,Absence)')

plt.ylabel('Samples')

plt.title('Count of each Heart Disease Class')

target_temp = dataset['Heart Disease'].value_counts()

print(target_temp)
```

PROJECT DEMO LINK:

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-24469-1659943256

DEMO LINK:

https://drive.google.com/file/d/1sJksvoYdnMvBbaouW3uKYn4vG_7NP9bO/view?usp=sharing