VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

PROJECT REPORT

1.INTRODUCTION

1.1 PROJECT OVERVIEW

A simple web application 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 Flask and deployed on Heroku.

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-toend 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

2. Literature Survey

Bo Jin, Chao Che et al. (2018) proposed a "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling" model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the sequential nature of clinical records [1].

Aakash Chauhan et al. (2018) presented "Heart Disease Prediction using Evolutionary Rule Learning". This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. This will facilitate (help) in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of coronary sickness [2].

Ashir Javeed, Shijie Zhou et al. (2017) designed "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection". This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. Two forms of experiments are used for cardiovascular disease prediction. In the first form, only random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed. This methodology is efficient and less complex than conventional random forest model. Comparing to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection[3].

"Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" proposed by Senthilkumar Mohan, Chandrasegar Thirumalai et al. (2019) was efficient technique using hybrid machine learning methodology. The hybrid approach is combination of random forest and linear method. The dataset and subsets of attributes were collected for prediction. The subset of some attributes were chosen from the pre-processed knowledge(data) set of cardiovascular disease .After prep-processing , the hybrid techniques were applied and disgnosis the cardiovascular disease [4].

K.Prasanna Lakshmi, Dr. C.R.K.Reddy (2015) designed "Fast RuleBased Heart Disease Prediction using Associative Classification Mining". In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-square testing and arranging the rules in an order to form a classifier. Using these phase to predict the heart disease easily [5].

M.Satish, et al. (2015) used different Data Mining techniques like Rule based, Decision Tree, Navie Bayes, and Artifical Neural Network. An efficient approach called pruningclassification association rule (PCAR) was used to generate association rules from cardiovascular disease warehouse for prediction of Heart Disease. Heart attack data warehouse was used for pre-processing for mining. All the above discussed data mining technique were described [6].

Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik (2015) "An Intelligent Decision Support System for Cardiac Disease Detection", designed a cost efficient model by using genetic algorithm optimizer technique. The weights were optimized and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks [7].

"Prediction and Diagnosis of Heart Disease by Data Mining Techniques" designed by Boshra Bahrami, Mirsaeid Hosseini Shirvani. This paper uses various classification methodology for diagnosing cardiovascular disease. Classifiers like KNN, SVO classifier and Decision Tree are used to divide the datasets. Once the classification and performance evaluation the Decision tree is examined as the best one for cardiovascular disease prediction from the dataset[8].

Mamatha Alex P and Shaicy P Shaji (2019) designed “Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique”. This paper uses techniques of Artificial Neural Network, KNN, Random Forest and Support Vector Machine. Comparing with the above mentioned classification techniques in data mining to predict the higher accuracy for diagnosing the heart disease is Artificial Neural Network[9]

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

What do they THINK AND FEEL?

what really counts
major preoccupations
worries & aspirations

They may get Exhausted easily

The quality service implies diagnosing patients properly and administering effective treatments

Nausea,Stomach pain,Indigestion

What do they HEAR?

what friends say
what boss say
what influencers say

Simple blood test help to predict cardiovascular disease

This disease leads to bluish skin ,shortness of breath,and feeling tired

Shortness of breath,chest pain, are the symptoms of cardiovascular disease

What do they SEE?

environment
friends
what the market offers

Positive result=1,for patient with heart disease

Negative result=1,for patient with no heart disease

Unhealthy diet,physical inactivity is the major risk factor

What do they SAY AND DO?

attitude in public
appearance
behavior towards others

Seek emergency medical care if an person suffer from chest paint.

Don't smoke

Maintain a healthy weight and reduce stress

PAIN

fears
frustrations
obstacles

Chest pain,chest tightness,chest pressure and chest discomfort(angina)

Most heart attacks involve discomfort in the center or left side of the chest that lasts for more than a few minutes or goes away and come back

squeezing ,fullness or pain in the center of your chest

GAIN

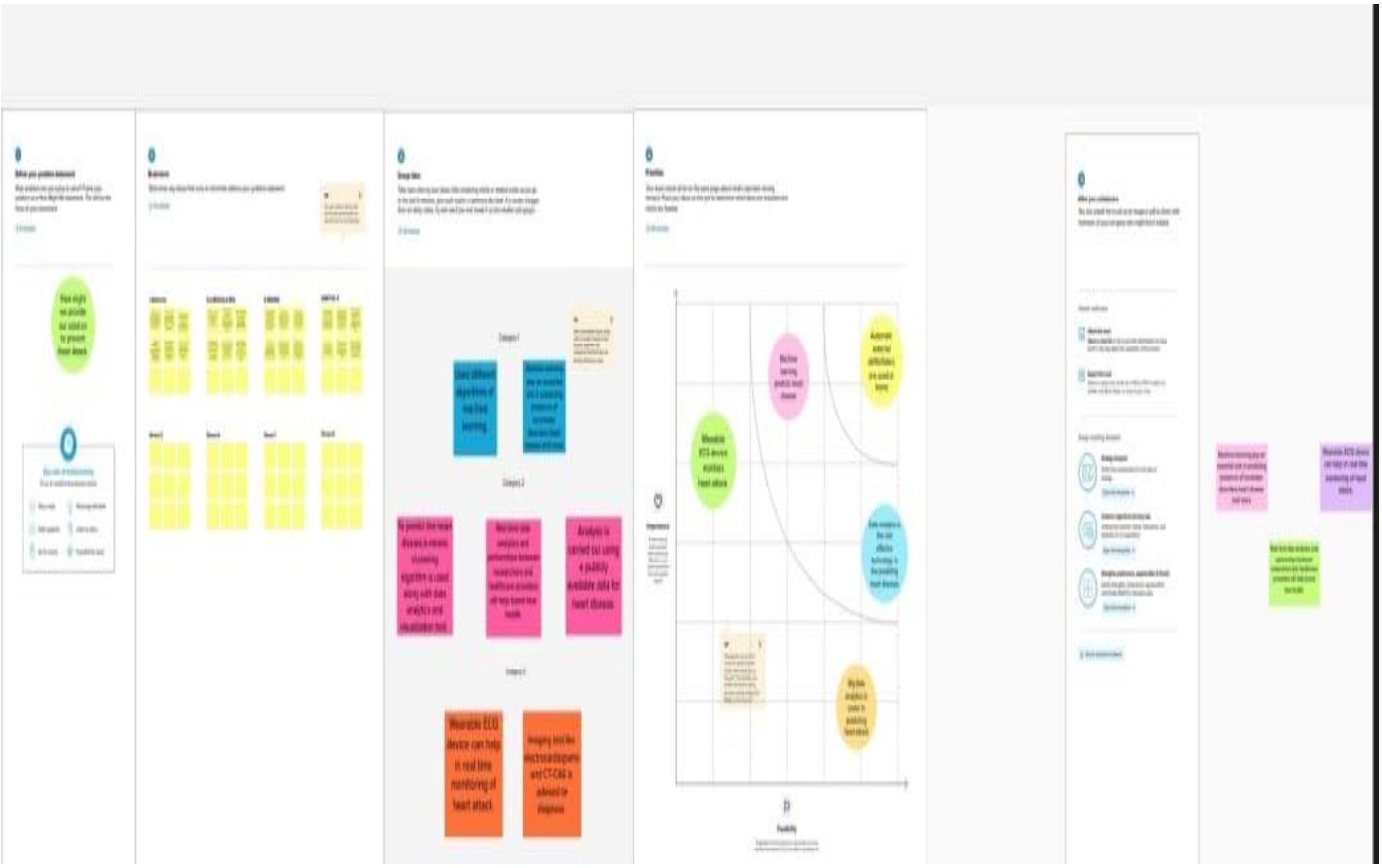
"wants" / needs
measures of success
obstacles

It enables significant knowledge

Increased accuracy for effective heart disease diagnosis

The performance and correctness have been surveyed due to do measuring some performance indices to reduce the death rate

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

Proposed Solution Template

S.No.	Parameter	Description
1.	**Problem Statement (Problem to be solved)**	➢ Coronary Artery Disease(CAD) is the most common heart problem. ➢ With CAD, a person may get blockages in their coronary arteries, the vessels that supply blood to their heart. ➢ That can lead to a decrease in the flow of blood to their heart muscle and reduce the oxygen level too.
2.	**Idea / Solution description**	➢ Stop eating unhealthy foods. ➢ Measure the amount of calcium in the walls of coronary arteries. ➢ Try to workout physical activity in your daily life.
3.	**Novelty / Uniqueness**	➢ Cardiac catheterization, inserting tubes into patients coronary arteries to evaluate or confirm CAD. ➢ Electrocardiogram records heart's electrical activity which can detect heart issues.
4.	**Social Impact / Customer Satisfaction**	➢ Treatment for CAD often includes lifestyle changes, risk factor management and medications. ➢ Some people may also benefit from a procedure or surgery.
5.	**Business Model (Revenue Model)**	➢ Aim for an eating plan that's low in saturated fats. ➢ Keep blood pressure under control.
6.	**Scalability of the Solution**	➢ A lifestyle changes should be made to reduce the risk of developing further health problems.

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.

After having identified the target customer segment, it's time to **investigate their needs**.

One of the cheapest, fastest and most informative things to do at this stage is to meet with customers through **customer discovery interviews** (more about customer discovery interviews here and here) until we keep hearing the same things from customers. Meeting with a customer is an invaluable source of insights, much more valuable than a survey.Besides, as entrepreneurs, our job is to meet and pitch to customers all the time, we'd be better off to start earlier rather than later.

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	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

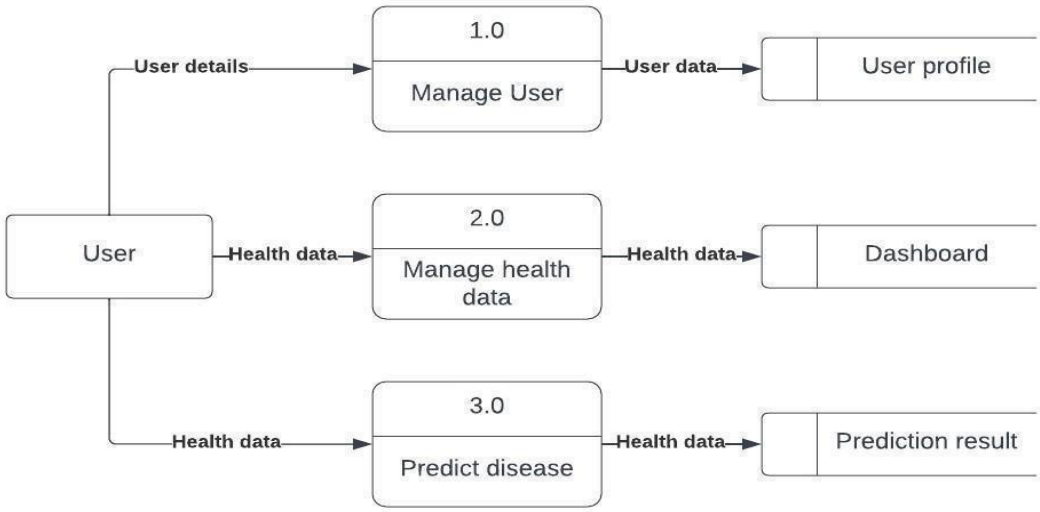
4.2 NON-FUNCTIONAL REQUIREMENT

NFR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user-friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	**Security**	For security of the application the technique known as database replication should be used so that all the important data should be kept safe. Incase of crash, the system should be able to backup and recover the data
NFR-3	**Reliability**	The application has to be consistent at every scenario and has to work without failure in any environment
NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
NFR-5	**Availability**	The application has to be available 24 x 7 for users without any interruption
NFR-6	**Scalability**	The application can withstand the increase in the no. of users and has to be able to develop Higher versions

5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

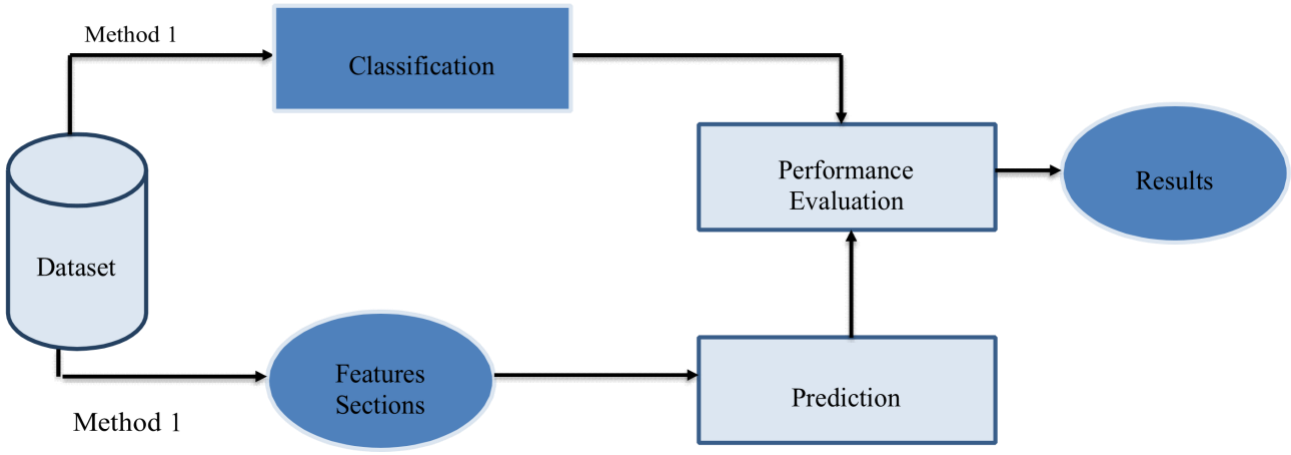
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



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).

ARCHITECTURE:



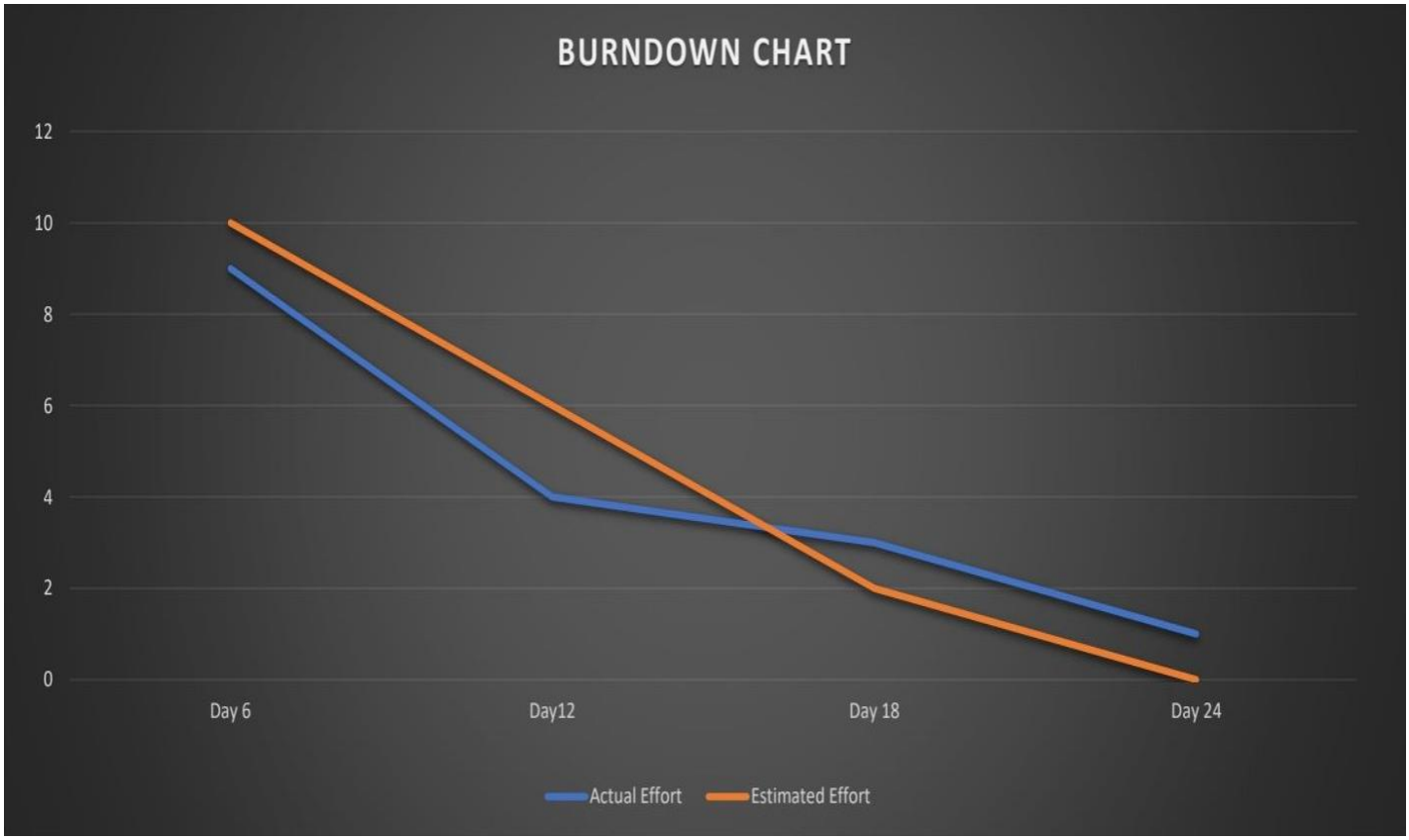
6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-I	Registration	USN-I	As a user, I can register for the application by entering my email, password, and confirming my password.	8	High	Sabitha R, Retheesha
		USN-2	As a user, I will receive confirmation email once I have registered for the application	8	High	Sabitha R, Retheesha
	Login	USN-3	As a user, I can log into the application by entering email & password	4	Medium	Reghula mol, Reshma
Sprint-2	Working with the dataset	USN-4	To work on the dataset, understand and load the dataset	10	High	Sabitha R, Reghula mol R A,
		USN-5	Exploration of BP vs chest pain type and gender, maximum heart rate during the chest an	5	High	Sabitha R, Retheesha J, Reghula mol R A, Reshma R
		USN-6	BP by age, Cholesterol by agent gender	5	High	Retheesha J, Reshma R
Sprint-3	Data Visualization	USN-7	Visualization of average age for chest pain types, average exercise angina curing chest pain	2	Medium	Sabitha R, Reghula mol R A
		USN-8	BP variation with respect to age, Effect of existing heart disease on average of Exercise Angina	6	High	Sabitha R, Retheesha J, Reghula mol R A, Reshma R
		USN-9	Average age for different types of chest pain in existing heart disease, serum cholesterol levels vs age	6	High	Reghula mol R A, Reshma R
		USN-10	Maximum heart rate in Existing heart disease by Exercise Angina	6	High	Sabitha R, Retheesha J, Reghula mol R A, Reshma R
Sprint-4	Dashboard Creation	USN-11	Dashboard showing different types of visualization	20	High	Sabitha R, Retheesha J, Reghula mol R A, Reshma R

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.



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.TESTING

7.1 TEST CASES

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
LoginPage_TC_001	Functional	Home Page	Verify user is able to see the Login/Signup popup when user clicked on My account button		1.Enter URL and click go 2.Click on My Account dropdown button 3.Verify login/Signup popup displayed or not		Login/Signup popup should display	Working as expected	Pass			BUG-1235	
LoginPage_TC_002	UI	Home Page	Verify the UI elements in Login/Signup popup		1.Enter URL and click go 2.Click on sign 3.Verify login/Signup popup with below UI elements: a.email text box b.password text box c.Login button d.New customer? Create account link e.Last password? Recovery password link		Application should show below UI elements: a.email text box b.password text box c.Login button with orange colour d.New customer? Create account link e.Last password? Recovery password link	Working as expected	Fail	Steps are not clear to follow		BUG-1234	
LoginPage_TC_003	Functional	Home page	Verify user is able to log into application with Valid credentials		1.Enter URL 2.Click on sign 3.Enter Valid username/email in Email text box 4.Enter valid password in password text box 5.Click on login button	Username: adhu23@gmail.com password: 124	User should navigate to user account homepage	working as expected	fail	Incorrect mail id/ password		BUG 123456789 US 1237	
LoginPage_TC_004	Functional	Login page	Verify user is able to log into application with InValid credentials		1.Enter URL(https://shopenzer.com/) and click go 2.Click on sign in 3.Enter InValid username/email in Email text box 4.Enter valid password in password text box 5.Click on login button	Username: chalam@gmail password: Testing@123	Application should show 'Incorrect email or password ' validation message.	working as expected	pass				
LoginPage_TC_004	Functional	Login page	Verify user is able to log in		1.Enter URL(https://shopenzer.com/) and click go 2.Click on sign 3.Enter Valid username/email in Email text box 4.Enter Invalid password in password text box 5.Click on login button	Username: chalam@gmail.com password: Testing@12367666786676 676	Application should show login successfully	working as expected	pass				
LoginPage_TC_005	Functional	Login page	Verify user is able to log into application with InValid credentials		user can directly login once sign in		application should be login in	Working as expected	pass				

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

The testing report are submitted in github account.

7.2 USER ACCEPTANCE TESTING

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

8. RESULTS

8.1 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%
```

9.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%.

10.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.

11. 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.

12. 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.

PROJECT DEMO LINK