Analytics for Hospitals' Health-Care Data

TEAM ID – PNT2022TMID11633

A PROJECT REPORT

SUBMITTED BY

Dharani Kumar C

Keerthivasan B

Mohammed Upayathulla

Akil V

Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   - 7.1 Feature 1
   - 7.2 Feature 2
   - 7.3 Database Schema (if Applicable)
8. **TESTING**
   - 8.1 Test Cases
   - 8.2 User Acceptance Testing
9. **RESULTS**
   - 9.1 Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**

    Source Code

PROJECT REPORT

1.INTRODUCTION

1.1 PROJECT OVERVIEW

Healthcare organizations are under increasing pressure to improve patient care outcomes and achieve better care. While this situation represents a challenge, it also offers organizations an opportunity to dramatically improve the quality of care by leveraging more value and insights from their data. Health care analytics refers to the analysis of data using quantitative and qualitative techniques to explore trends and patterns in the acquired data. While healthcare management uses various metrics for performance, a patient‟s length of stay is an important one.

Being able to predict the length of stay (LOS) allows hospitals to optimize their treatment plans to reduce LOS, to reduce infection rates among patients, staff, and visitors.

1.2 PURPOSE

Hospital admission data was analyzed to accurately predict the patient‟s Length of Stay at the time of admit so that the hospitals can optimize resources and function better. Built 3 models in Python to predict the length of stay,

- A supervised algorithm Naïve Bayes which was classifying with an accuracy of 34.55%.
- An ensemble method XGBoost which was predicting with an accuracy of 43.05%.
- A dense neural network with 6 layers which yields an accuracy of 42.5%.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

1.Big Data Analytics in Healthcare Systems

Big Data analytics can improve patient outcomes, advance and personalize care, improve provider relationships with patients, and reduce medical spending. This paper introduces healthcare data, big data in healthcare systems, and

applications and advantages of Big Data analytics in healthcare. We also present the technological progress of big data in healthcare, such as cloud computing and stream processing. Challenges of Big Data analytics in healthcare systems are also discussed.

Keywords- Big data, Big Data analytics, Healthcare, Personalized medicine, Precision medicine, Cloud computing,
Stream processing.

AUTHOR : **Mohammad Alkhatib, Amir Talaei-Khoei**, **Amir Ghapanchi**

Big data in healthcare: management, analysis and future prospects

The term "*big data*" has become extremely popular across the globe in recent years. Almost every sector of research, whether it relates to industry or academics, is generating and analyzing *big data* for various purposes. The most challenging task regarding this huge heap of data that can be organized and unorganized, is its management. Given the fact that big data is unmanageable using the traditional software, we need technically advanced applications and software that can utilize fast and cost-efficient high-end computational power for such tasks. Implementation of artificial intelligence (AI) algorithms and novel fusion algorithms would be necessary to make sense from this large amount of data. Indeed, it would be a great feat to achieve automated decision-making by the implementation of machine learning (ML) methods like neural networks and other AI techniques. However, in absence of appropriate software and hardware support, big data can be quite hazy. We need to develop better techniques to handle this „endless sea" of data and smart web applications for efficient analysis to gain workable insights. With proper storage and analytical tools in hand, the information and insights derived from big data can make the critical social infrastructure components and services (like healthcare, safety or transportation) more aware, interactive and efficient [3]. In addition, visualization of big data in a user-friendly manner will be a critical factor for societal development.

AUTHOR : Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma & Sandeep Kaushik

3. Exploring big data analytics in health care

Health care Industries are facing lot of challenges in maintaining patient information across various databases due to storage issues. In order to extract patient information, preprocessing techniques can be applied in the process of data mining across databases. But as the data is growing enormously with rapid speed, data mining techniques are becoming obsolete due to issues such as Storage, Speed. So, cost optimization has become one of the major requirements in health industry as there is huge burden in maintaining large volumes of patient‟s information using traditional databases. Here Big Data plays a vital role in storing huge volumes of patient information using storage mechanisms such as HDFS, HBase. Many issues in health care are discussed in this paper such as prediction of diseases, getting patients information across databases as a single view.

AUTHOR : T.Ramesh[a]V.Santhi[b]

4. Big data analytics in healthcare: promise and potential

Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Its potential is great; however there remain challenges to overcome.

Keywords: Big data, Analytics, Hadoop, Healthcare, Framework, Methodology

AUTHOR : Wullianallur Raghupathi and Viju Raghupathi

5. An overview of healthcare data analytics with applications to the COVID-19 pandemic

In the era of big data, standard analysis tools may be inadequate for making inference and there is a growing need for more efficient and innovative ways to collect, process, analyze and interpret the massive and complex data. We provide an overview of challenges in big data problems and describe how innovative analytical methods, machine learning tools and metaheuristics can tackle general healthcare problems with a focus on the current pandemic. In particular, we give applications of modern digital technology, statistical methods,data platforms and data integration systems to improve diagnosis and treatment of diseases in clinical research and novel epidemiologic tools to tackle infection source problems, such as finding Patient Zero in the spread of epidemics. We make the case that analyzing and interpreting big data is a very challenging task that requires a multi-

disciplinary effort to continuously create more effective methodologies and powerful tools to transfer data information into knowledge that enables informed decision making.

AUTHOR : Zhe Fei, Yevgen Ryeznik, Alex Sverdlov, Chee Wei Tan

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.

Predictive analytics can create patient journey dashboards and disease trajectories that can lead to effective, and result-driven healthcare. It improves treatment delivery, cuts costs, improves efficiencies, and so on. To achieve this, make sure you have access to clean, scaled, formatted, and quality data from external as well as internal resources. Providers can improve their data capture schedules by organizing important data types for their specific projects to ensure that data is meaningful for downstream analytics.

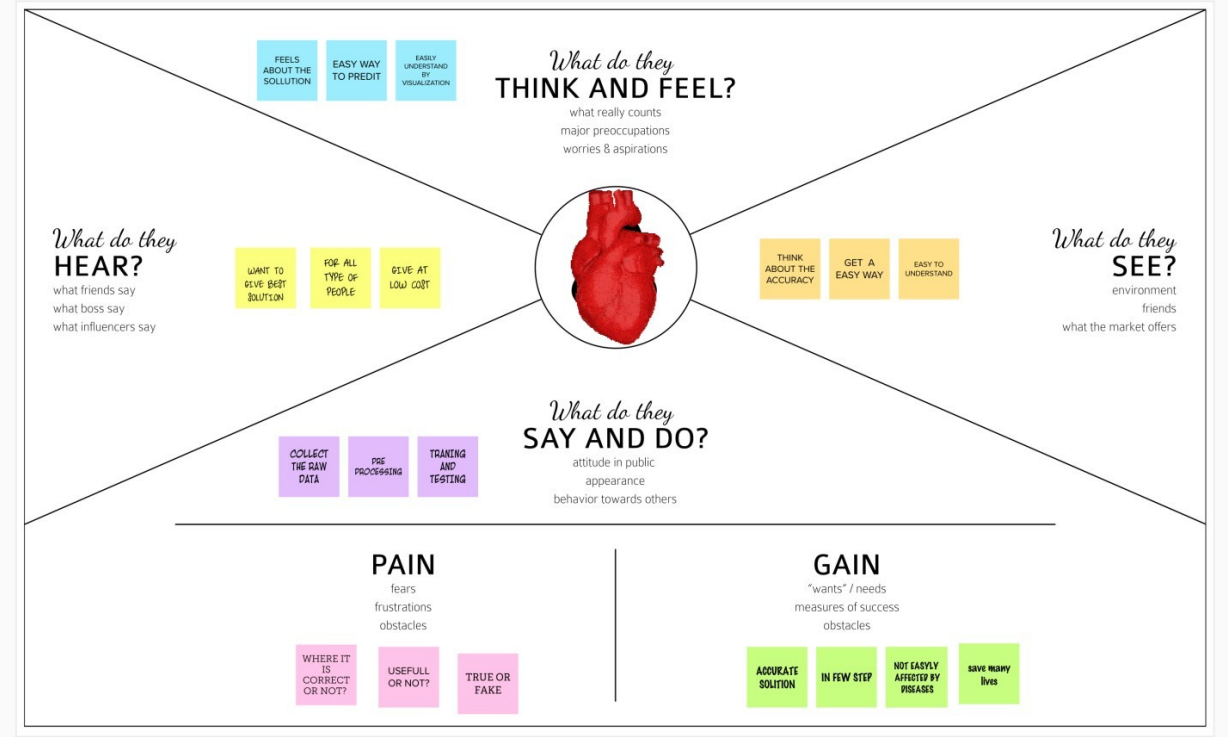
3.IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

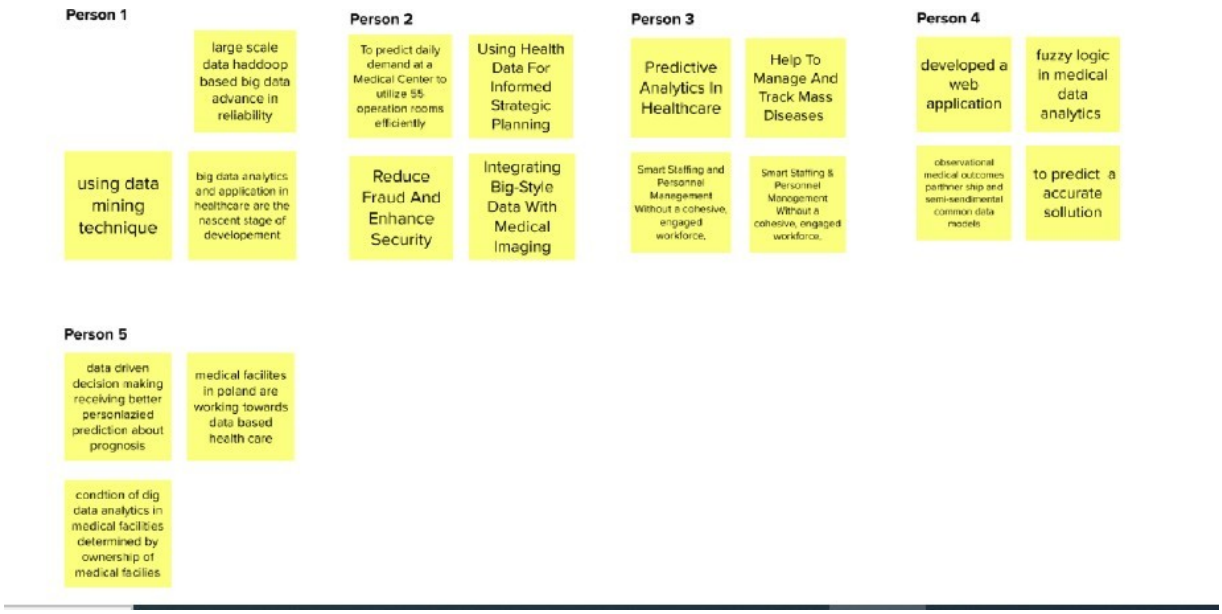
It‟s easy to jump straight into value proposition design. That is the core of your business and where the revenue or exchange of value will come from. However, trying to provide value to a misunderstood customer is very risky business. Do you have your blinkers on? Try using this canvas before you design 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?

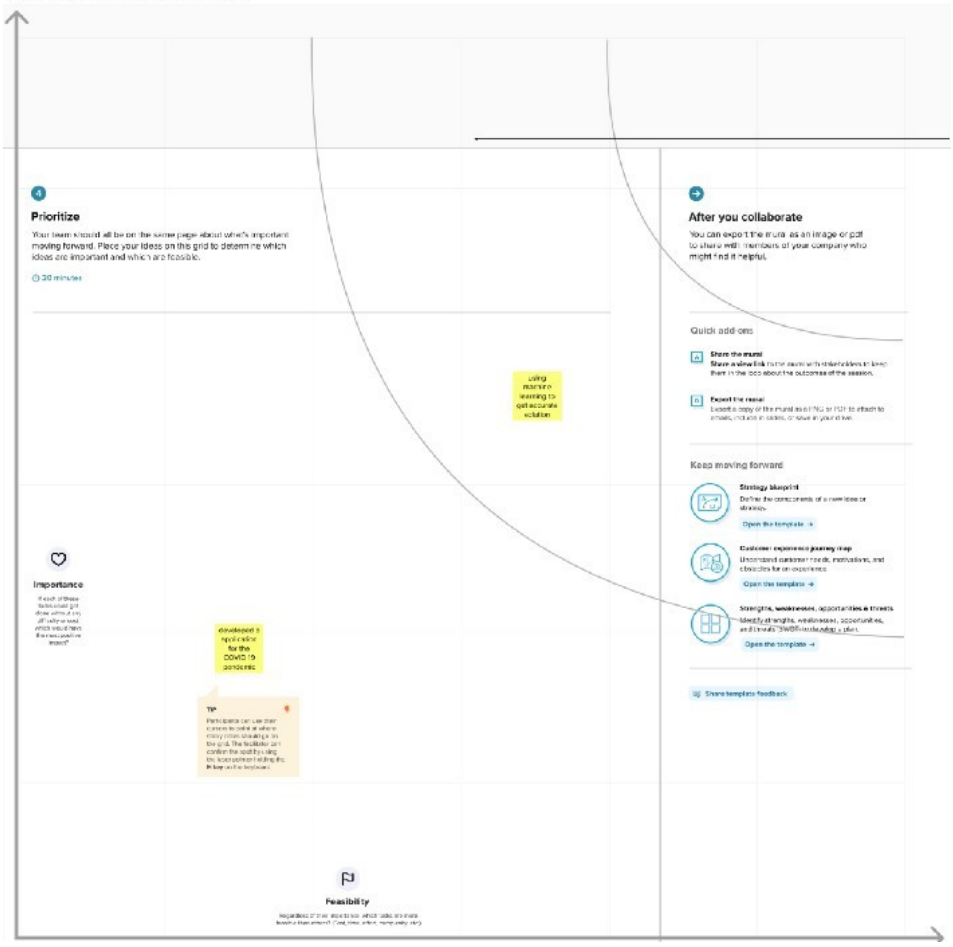
In this empathy map what customer think and feels. this map shows the pain and gain of the customer and what do their hear about the problem. this is the easy way to understand the problem statement



3.2 IDEATION & BRAINSTORMING



Step-3: Idea Prioritization



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 drive reliable experiences, AI and ML algorithms need credible information without duplications and mistakes. This helps specialists to get real-time predicted data that seems relevant to the patient‟s health history. And, based on that the right treatment is prescribed to them.

Hence, HCOs should exclude data governance and master data management solutions to enhance data quality. They should develop multidisciplinary groups

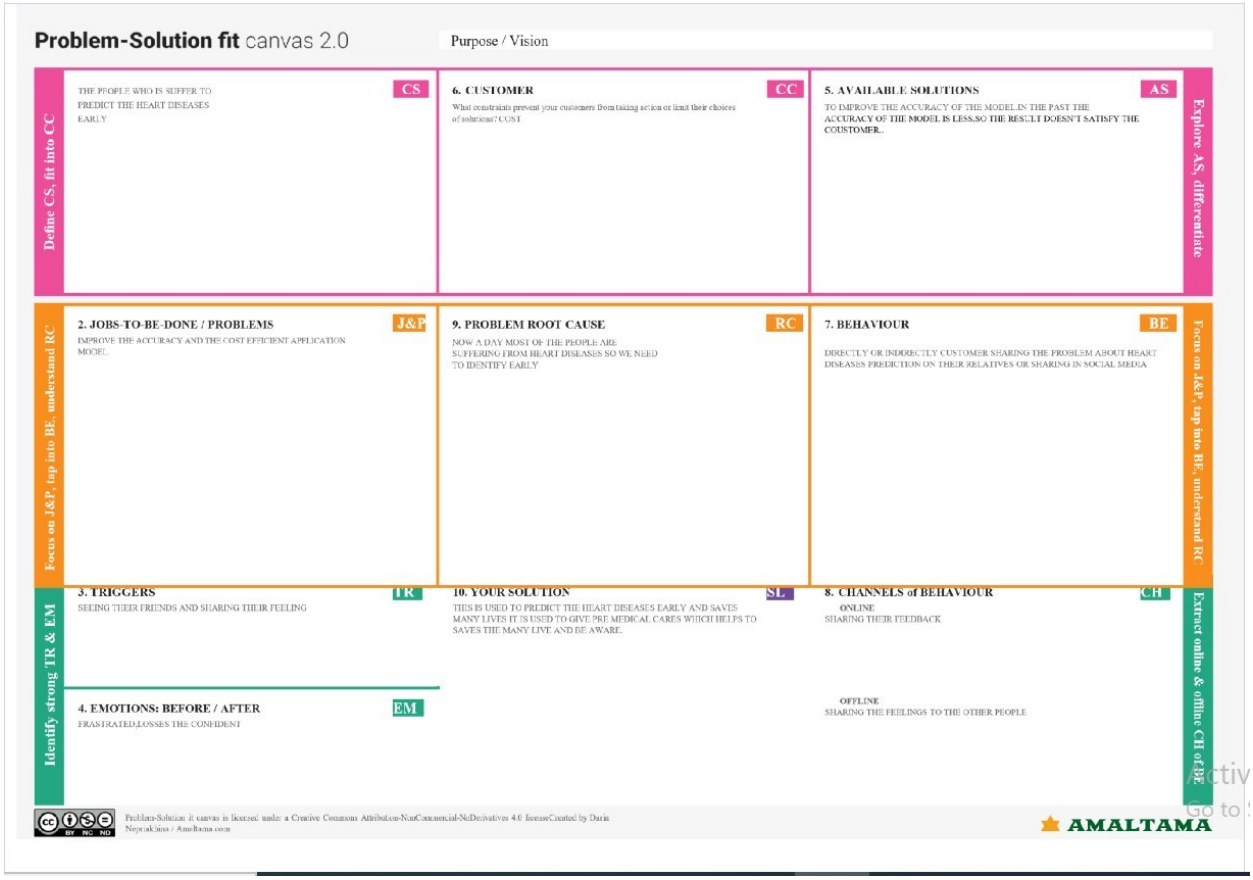
that separate the shackles debilitating the healthcare services, processes, and clinicians.

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	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

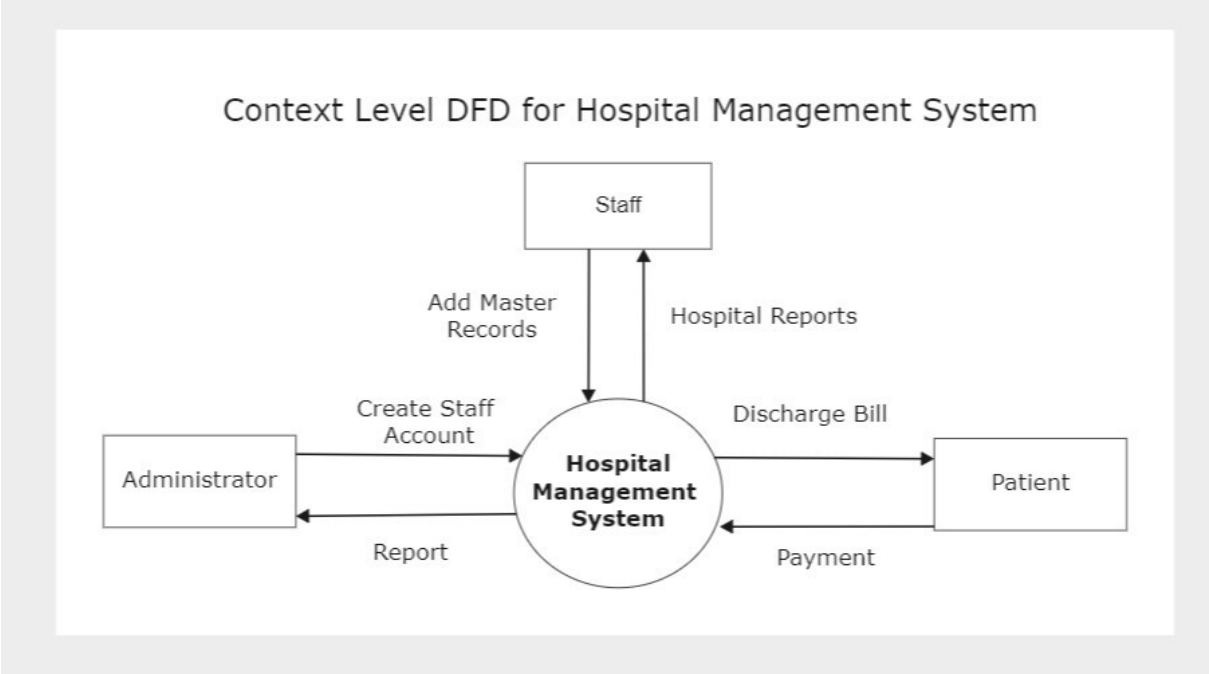
4.2 NON-FUNCTIONAL REQUIREMENT

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

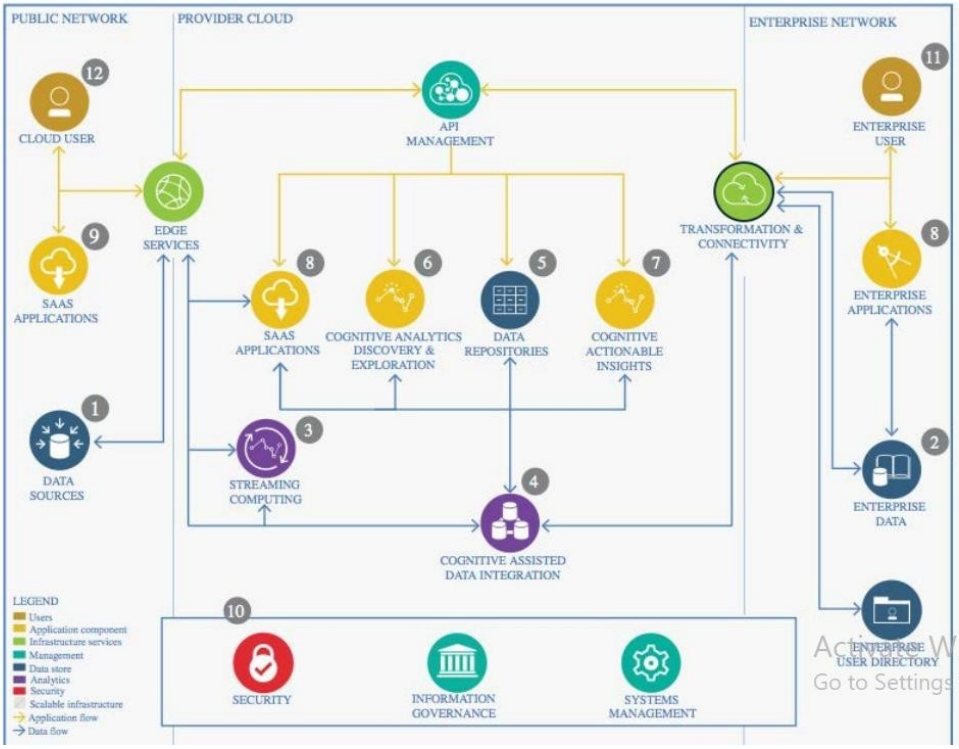
A data flow diagram (DFD) is **a graphical or visual representation using a standardized set of symbols and notations to describe a business's operations through data movement**.



In this flow diagram we are showing the analytics for hospital health care

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

User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	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	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard through facebook	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register through gmail	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard					
Customer (Web user)		USN-6	As a user I can fill the detail asked here	I can register the asked detail	high	Sprint 1
Customer Care Executive		USN-6	As a user I executive the given detail	I can accept the terms	medium	Sprint-1
Administrator		USN-7	As a administer it predict the output	Show the result	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING ESTIMATION

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

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	
Sprint-1		USN-3	As a user, I can register for the application through Gmail	2	Medium	
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email	1	High	
	Dashboard	USN-5	As a user/patient I can request for Organ/Blood Donation .	2	Medium	
		USN-6	As a user/donor I can view my appointments	1	High	

6.2 SPRINT DELIVERY SCHEDULE

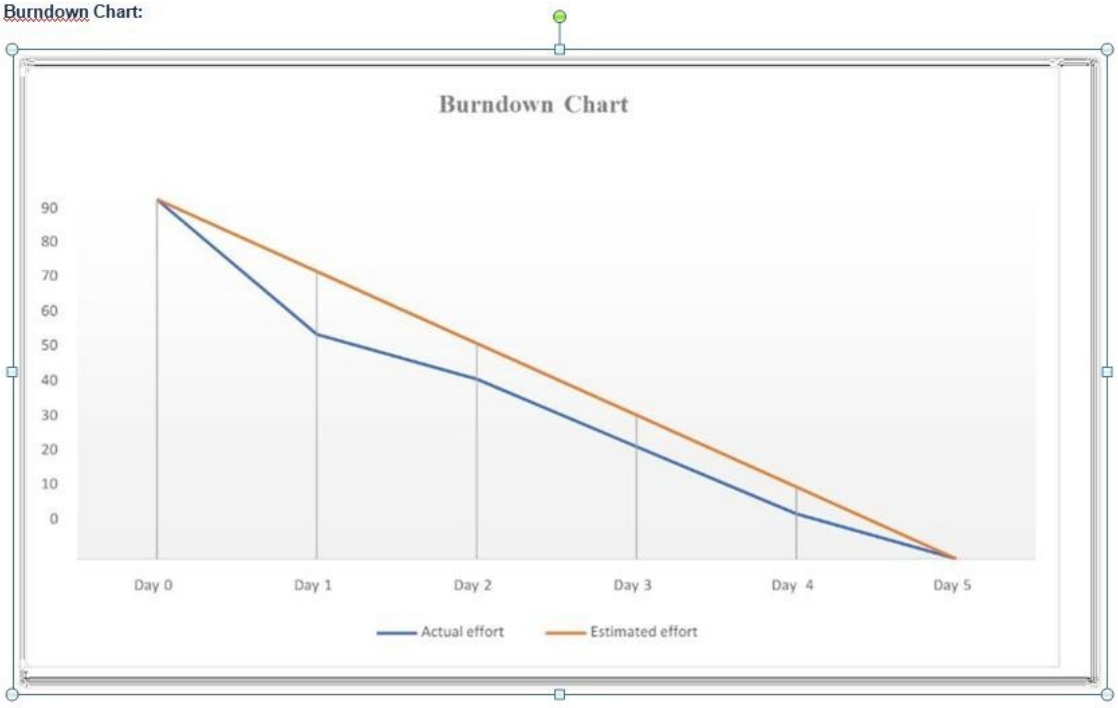
Project Tracker, Velocity & Burndown Chart: (4 Marks)

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	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	07 Nov 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.

Burndown Chart:



7.CODING & SOLUTIONING

7.1 FEATURE 1

* Main app.js file for web app of Example Health Analytics

*/

```
var createError = require('http-errors');

var express = require('express');
```

```
var bodyParser = require('body-parser');
var logger = require('morgan');

var app = express();

var indexRouter = require('./routes/index');

app.use(logger('dev'));

app.use(express.static('./public')); // load UI from public folder
app.use(bodyParser.json())

app.use('/data', indexRouter);

// catch 404 and forward to error handler
app.use(function(req, res, next) {
next(createError(404));
});

// error handler
app.use(function(err, req, res, next) {
  // set locals, only providing error in development
  res.locals.message = err.message;
```

```
  res.locals.error = req.app.get('env') === 'development' ? err : {};


  res.status(err.status || 500);
});


module.exports = app;
```

FEATURE 2

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: dataservice
spec:
  selector:
      matchLabels:
        name: dataservice-deployment
        version: v1
  replicas: 1
  template:
    metadata:
      labels:
        name: dataservice-deployment
        version: v1
    spec:
```

```
    containers:
      - name: dataservice
        image: maxshapiro32/examplehealthanalyticsdata
        imagePullPolicy: Always
        resources:
            limits:
              cpu: 200m
            memory: 500Mi
            requests:
              cpu: 100m
              memory: 200Mi
        ports:
          - containerPort: 3000
        env:
          - name: HOST_IP
            value: <IP_OR_SUBDOMAIN>
          - name: SCHEME
            value: <SCHEME>
          - name: MAPBOX_ACCESS_TOKEN
            value: <MAPBOX_ACCESS_TOKEN>
          - name: DATA_SOURCE_API
            value: <DATA_SOURCE_API>
---
```

```
apiVersion: v1
kind: Service
metadata:
  name: dataservice-service
spec:
  ports:
    - port: 3000
      targetPort: 3000
      nodePort: 32000
      name: http
  type: NodePort
  selector:
    name: dataservice-deployment
```

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**. 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.

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 ensure that the system meets the functional requirement.

9. RESULTS

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 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%.

11. CONCLUTION

- Small Healthcare provider stretched to meet ARRA EMR Deadline
- Many ethical and privacy issues are unresolved
- Industry is still developing and maturing
- Significant Medical Benefits Exists Outside The Individual Healthcare Provider

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 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

```
* Main app.js file for web app of Example Health Analytics
 */


var createError = require('http-errors');

var express = require('express');

var bodyParser = require('body-parser');

var logger = require('morgan');


var app = express();


var indexRouter = require('./routes/index');
```

```
app.use(logger('dev'));

app.use(express.static('./public')); // load UI from public folder
app.use(bodyParser.json())

app.use('/data', indexRouter);

// catch 404 and forward to error handler
app.use(function(req, res, next) {
next(createError(404));
});

// error handler
app.use(function(err, req, res, next) {
  // set locals, only providing error in development
  res.locals.message = err.message;
  res.locals.error = req.app.get('env') === 'development' ? err : {};

  res.status(err.status || 500);
});

module.exports = app;
```