PROJECT REPORT ON

ANALYTICS FOR HOSPITAL'S HEALTH CARE DATA

DATA ANALYTICS

Submitted By,

JESWIN W - 917719IT040
JEYA GANESH A V - 917719IT041
RAHUL HARIESH B - 917719IT074
SANJAYKUMAR S - 917719IT082

TEAM ID : PNT2022TMID21545

Department Of Information Technology

Thiagarajar College Of Engineering, Madurai - 625014.

INSTITUTION MENTOR : Dr.S.Sridevi

INDUSTRIAL MENTOR : Shivam Shivare, Hari Prabu

1. ABSTARCT

The purpose of Analytics for Hospital Healthcare data is to provide the information to patients about the average length of hospital stays. Patient duration of stay is one crucial statistic to monitor and forecast if one wishes to increase the effectiveness of health care management in a hospital. Data science has several applications in the field of health care management. At the time of admission, this metric aids hospitals in identifying patients who are at high LOS-risk (patients who will stay longer). Once identified, patients at high risk for LOS can have their treatment plans improved to reduce LOS and reduce the risk of nfection in staff or visitors. Additionally, prior awareness of LOS might help with planning logistics like room and bed allotment.

Keywords :

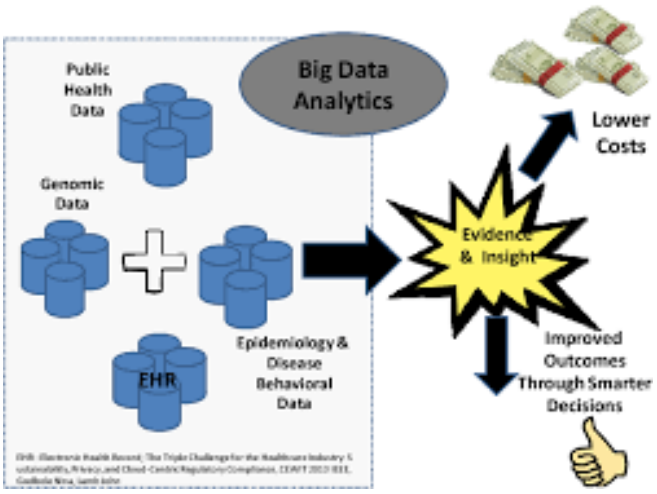
LOS - Length of Stay
Data Analytics
Severity of illness
Bed Allotment

2. INTRODUCTION

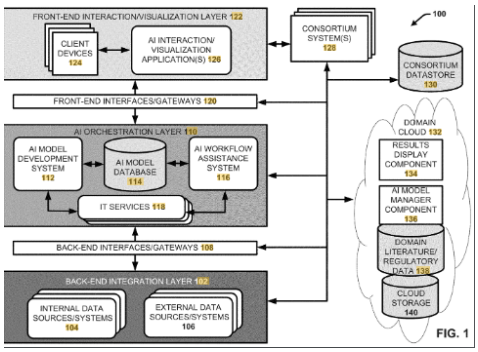
The Analytics for Hospital Health care data is to make the patients to know about the length of stay in the hospital. The healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital. This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning. To accurately predict the Length of Stay for each patient on a case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days. By predicting LOS we can plan the required beds, checking the future availability of beds in hospitals, monitoring length of stay of patients.If we can predict the length of stay the hospital management can prepare the requirements for the patient. The patients can easily get the things and clothes required for the stay. The hospital management can increase the beds available with the data.

3. LITERATURE SURVEY

3.1 Research Paper:

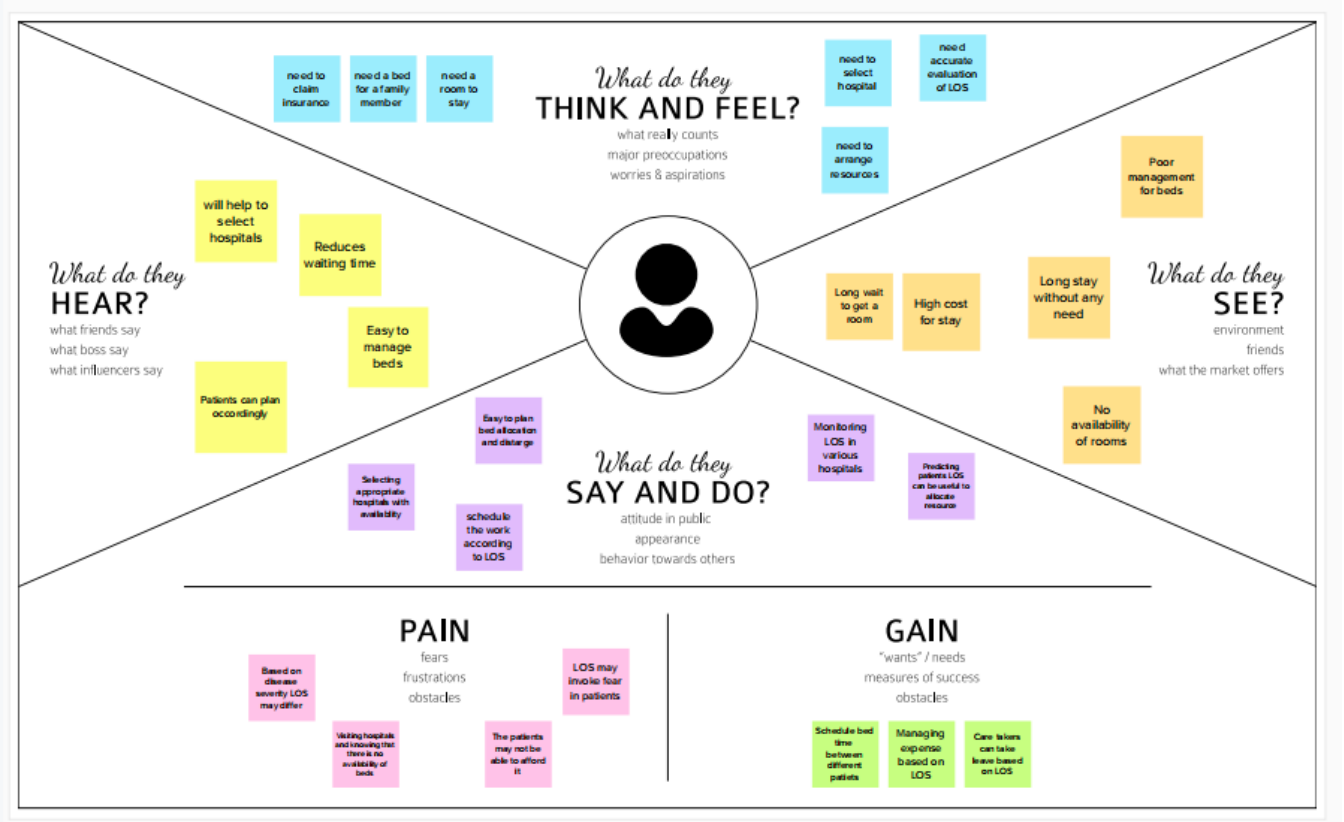
Journal	IEEE
Title	Big data analytics in healthcare
Authors	Sohail Imran, Tariq Mahmood, Ashan Morshed, Timos Sellis
Volume/ Issue Year	Volume: 8, Issue: 1, January 2021
Description	Big data analytics (BDA) in healthcare can, for instance, help determine causes of diseases, generate effective diagnoses, enhance QoS guarantees by increasing efficiency of the healthcare delivery and effectiveness and viability of treatments, generate accurate predictions of read missions, enhance clinical care, and pinpoint opportunities for cost savings.We initially determine big data characteristics for healthcare and then review BDA applications to healthcare in academic research focusing particularly on NoSQL databases. 

3.2 Patent:

Title	Facilitating artificial intelligence integration into systems using a distributed learning platform
Patent no	US10957442B2
Inventor	John Kalafut, Keith Dreyer, Mark Michalski, Stuart Pomerantz, Sean Doyle, Neil Tenenholtz
Description	Techniques are described that facilitate integrating artificial intelligence informatics in healthcare systems using a distributed learning platform. In one embodiment, a computer-implemented is provided that comprises interfacing, by a system operative coupled to a processor, with a medical imaging application that provides for viewing medical image data. The method further comprises, facilitating, by the system, generation of structured diagnostic data according to a defined ontology in association with usage of the imaging application to perform a evaluation of the medical image data.  FIG. 1

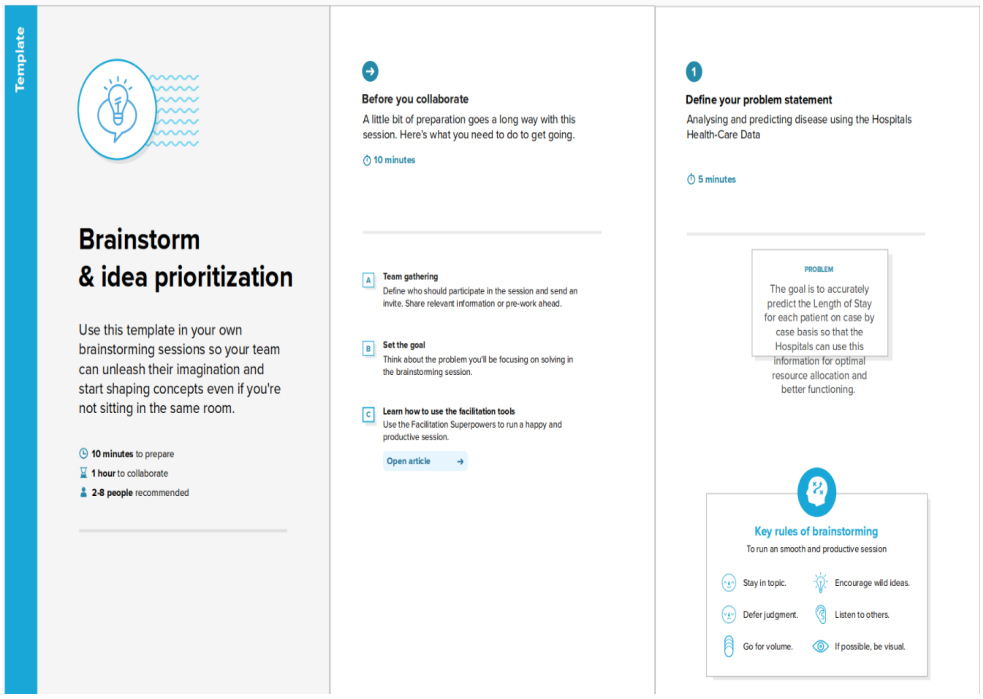
4. IDEATION AND PROPOSED SOLUTION

4.1 Empathy Map

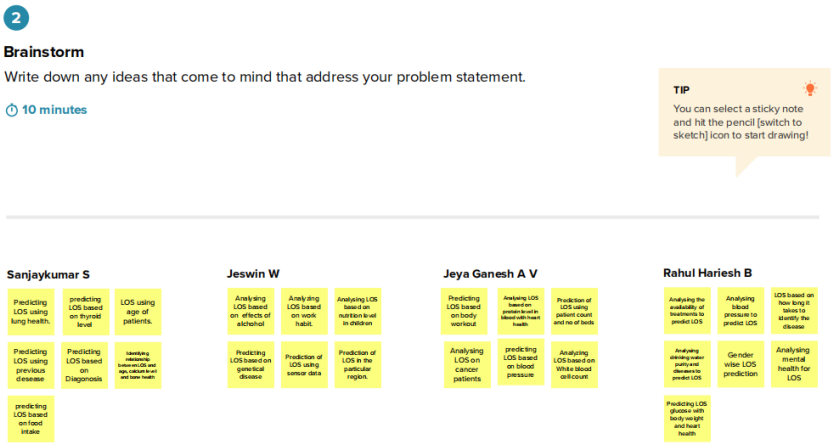


4.2 Brainstorming & Idea Prioritization

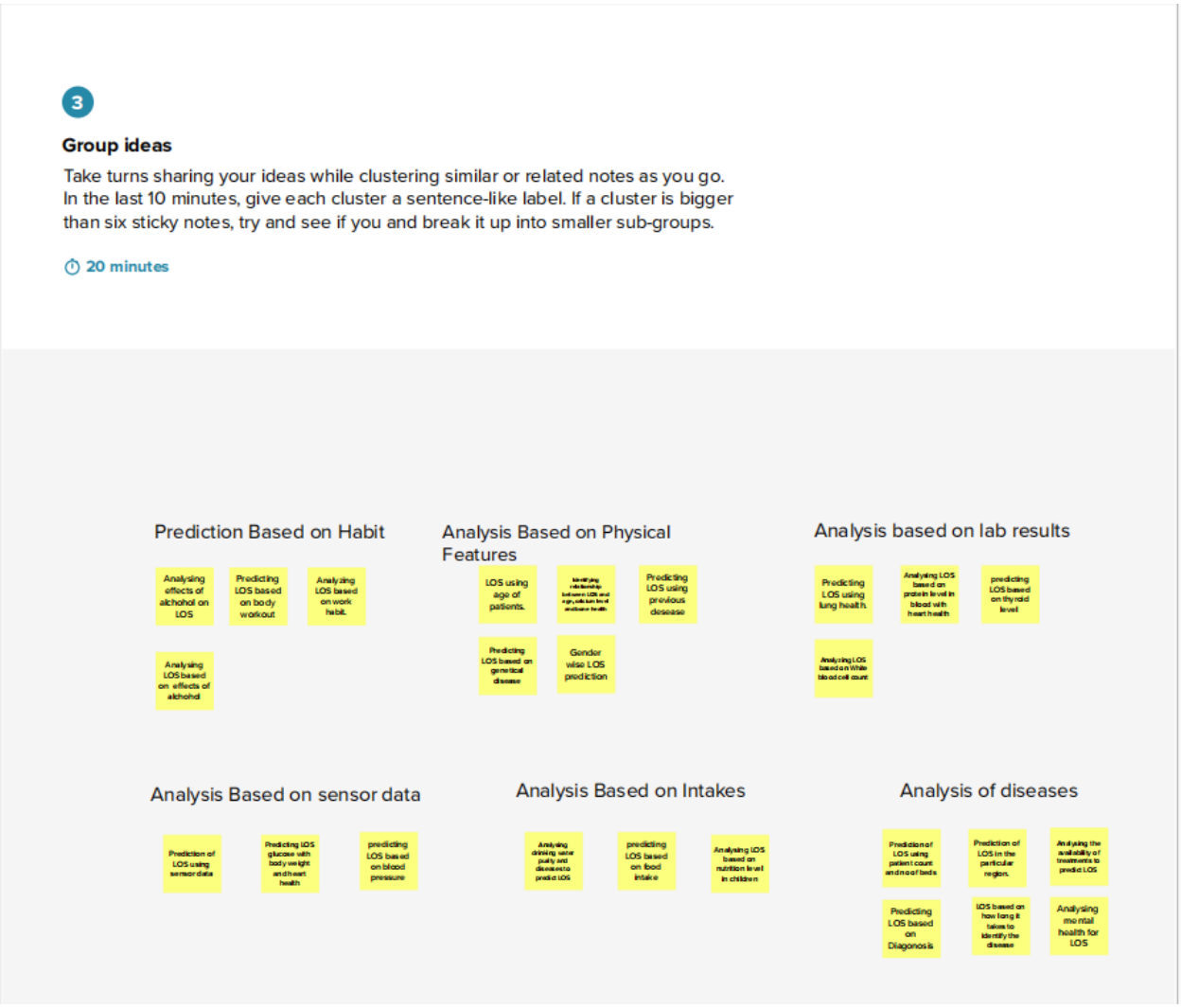
Step-1: Team Gathering, Collaboration and Problem statement Selection



Step-2: Brainstorming



Step-3: Grouping Ideas



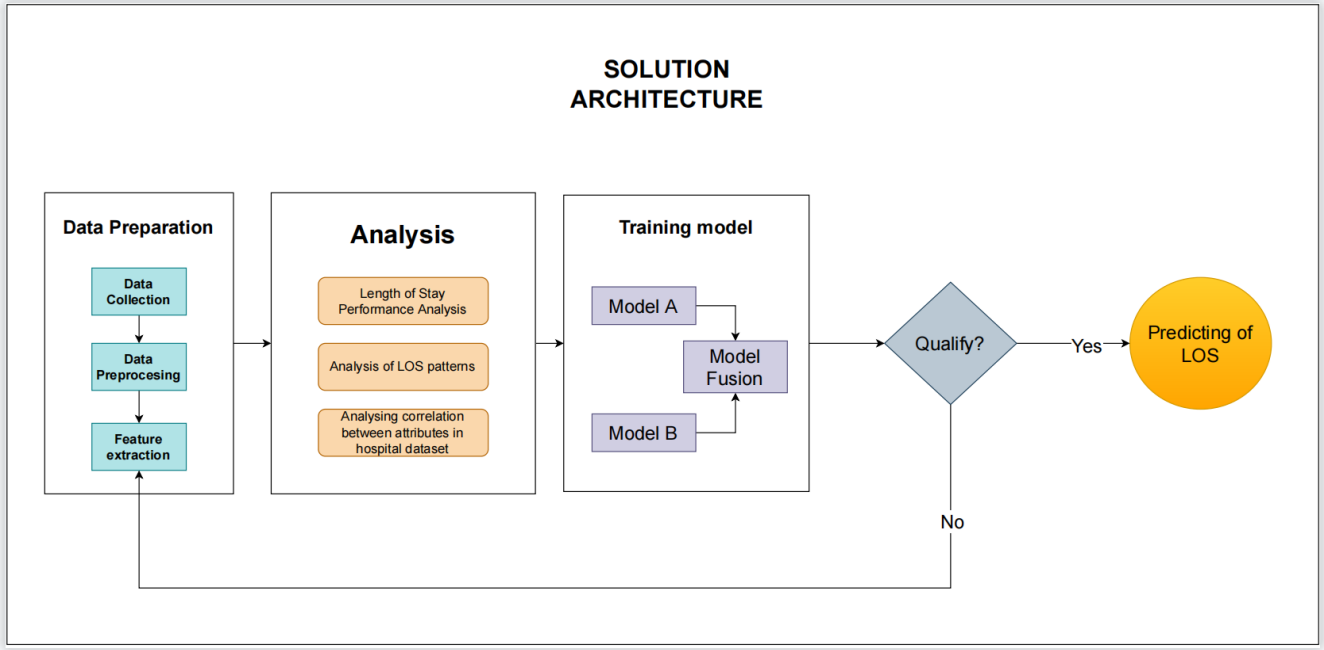
Step-4: Idea Prioritization

4.3 Proposed Solution

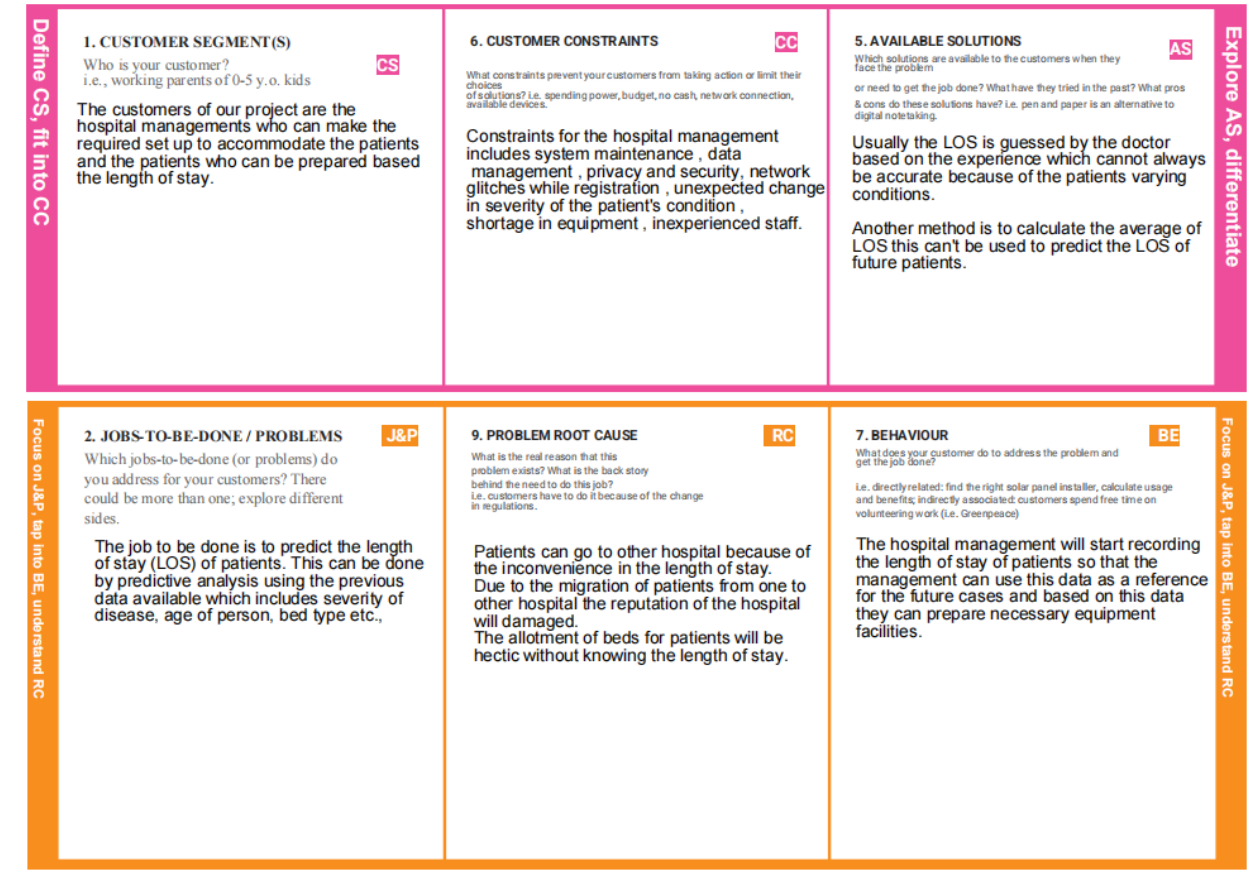
S.No	Parameter	Description
1.	Problem Statement	● While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital. ● This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning
2.	Idea / Solution description	● To accurately predict the Length of Stay for each patient on a case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. ● The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.
3.	Novelty / Uniqueness	● Predicting the Length Of Stay; based on disease diagnosed lets the hospital reduce the LOS by optimizing treatment ● Classifying patients LOS used to plan the bed availability accordingly. ● Predicting LOS using disease severity, disease type, hospital department etc. ● LOS can be used to book beds in hospitals.

4.	Social Impact / Customer satisfaction	● If we can predict the length of stay the hospital management can prepare the requirements for the patient. ● The patients can easily get the things and clothes required for the stay ● The hospital management can increase the beds available with the data
5.	Business Model	Customer selection: ● Hospital management. ● Patients/Public ● Government ● Insurance Value proposition: By predicting LOS we can plan the required beds, checking the future availability of beds in hospitals, monitoring length of stay of patients.
6.	Scalability of the Solution	The solution can be used in every hospital and by patients and can take required measures for the length of stay.

4.4 Solution Architecture



4.5 Solution Fit



5. REQUIREMENT ANALYSIS

5.1 Functional Requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Analyzing and Visualizing Hospital health care data	● Visualizing Analysis result on application dashboard. ● Analyze the relationship between various attributes in the dataset and Length of stay. ● Interactive dashboard that users can easily understand the insights.
FR-2	Prediction of LOS	● Predict the Length of Stay using the user's hospital data like Severity of disease, hospital type, hospital location, hospital name, emergency or not , etc. ● System should predict the LOS with any number of given attributes.
FR-3	Obtaining User Response for prediction	● Get the user's response after the prediction. ● This helps us to find how accurate our prediction is from the user's point of view. ● Bad user experience can be noted by doing this.So that we can improve the prediction accuracy.
FR-4	Monitoring user response and satisfaction	● Real Time monitoring of user response. ● Monitoring user satisfaction through various visualizations like barchart, pie chart etc
FR-5	Monitoring System accuracy	● The accuracy of the prediction should be monitored every time there is a change in dataset. ● If the accuracy becomes low the model should be redesigned for higher accuracy. ● This way the predictions will be up to date.

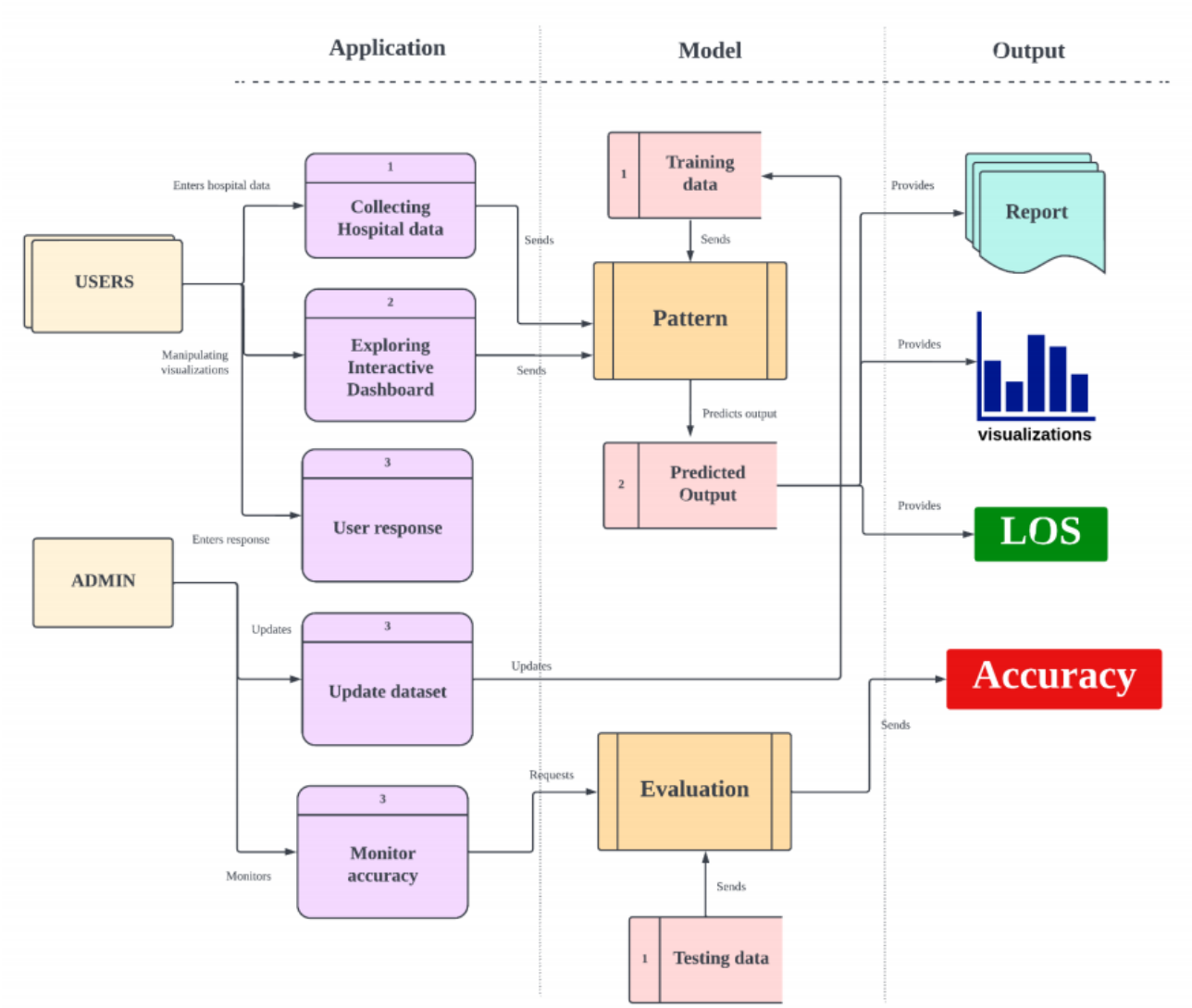
5.2 Non-Functional Requirement

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

FR No	Non-Functional Requirement	Description
NFR-1	Usability	The goals of the users are easily accomplished quickly by interactive design and less error.
NFR-2	Security	The dataset is accessed only by the administrators and the user's input is encrypted and it is protected.
NFR-3	Reliability	It works without a failure at the prediction time because of less bugs in the code it is because of using good trained data.
NFR-4	Performance	It supports at most 1000 patients queries at a time and after prediction is done it will be fastly communicated to the users.
NFR-5	Availability	The application is 99% available 24/7
NFR-6	Scalability	The application should support all browser types and it can handle maximum users.

6. PROJECT DESIGN

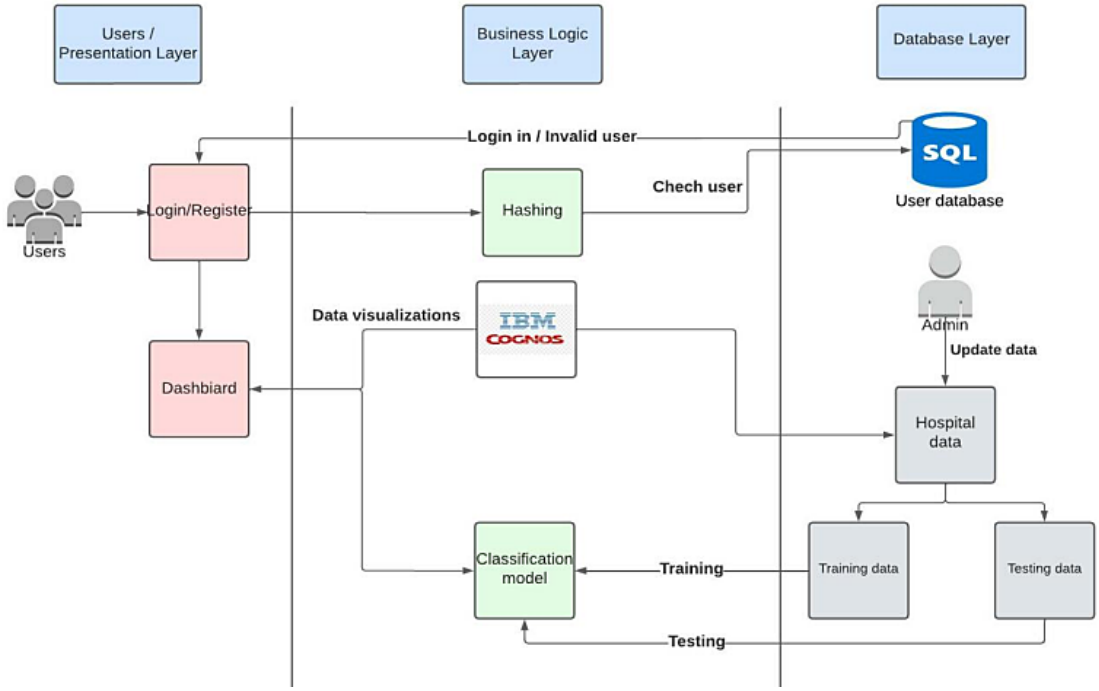
6.1 Data Flow Diagrams



6.2 Solution and Technology Architecture

Technical Architecture:

Architectural Diagram:



1. Component and Technologies			
S.No	**Component**	**Description**	**Technology**
1.	User Interface	user interacts with application in Web UI.	HTML, CSS, JavaScript / Angular Js etc.
2.	Data Visualization	Data is visualized so that the users can understand the important patterns in data.	IBM Watson
3.	Data Classification	Data is classified using classification algorithms to classify the data into 10 classifications	IBM Watson , colab
4.	Data Prediction	Logic for a process in the application colab,	IBM Watson
5.	Database	All the datasets of the patients and the hospital	MySQL, etc.
6.	File Storage	File storage requirements	Local Filesystem
7.	External API-1	Build models and helps in predict the data	IBM Watson api
8.	Machine Learning Model	Helps in developing the model	Classification algorithms
9.	Infrastructure	The application is deployed in cloud	IBM cloud

2.Application Characteristics			
S.No	**Characteristics**	**Description**	**Technology**
1	Open-Source Frameworks	The data prediction is done in open-source framework	Colab ,python
2	Security Implementations	The login and sign in purpose are implemented with security concerns	Salt hashing
3	Scalable Architecture	The application is done 3 tier architecture	Presentation layer-HTML/CSS javascript Business Logic Layer-colab, IBM cognos Database layer-IBM db2
4	Availability	The application is available for all the users at anytime	IBM Cognos
5	Performance	The application provides various visualization types in the dashboard	IBM Cognos

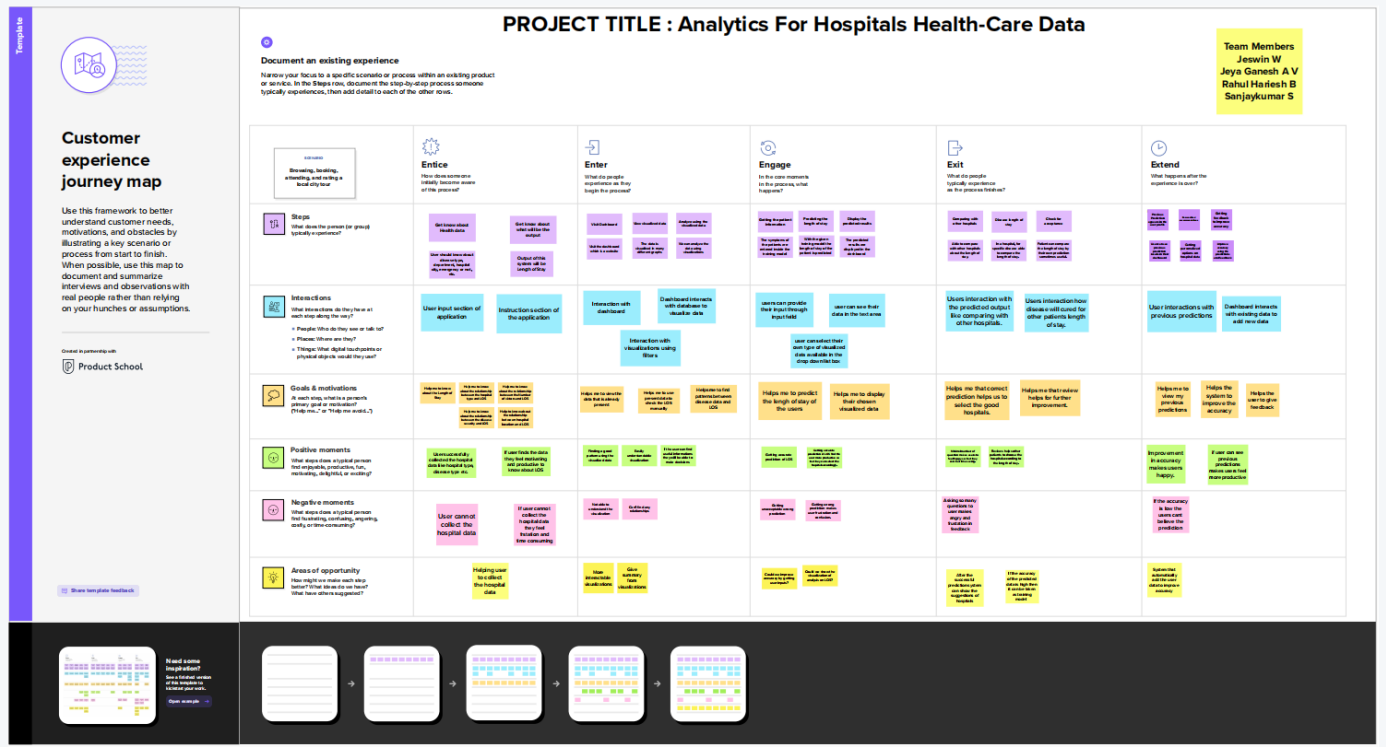
6.3 User Story

User Type	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Acceptance criteria**	**Priority**	**Release**
Patient	Analyzing and Visualizing Hospital health care data	USN-1	As a patient, I want to visualize the hospital health care data	I can visualize health care data.	Medium	Sprint 2
Patient		USN-2	As a patient, I want the relationship between various attributes in dataset	I can understand the relationships between various	Medium	Sprint 2

				attributes		
Patient	Prediction of LOS	USN-3	As a patient, I want to predict length of stay so that I can plan accordingly	I can get the predicted LOS	High	Sprint 1
Hospital Room allotment Manager		USN-4	As a manager, I want to predict the length of stay so that I can allot the hospital room accordingly	I can get the predicted LOS.	High	Sprint 1
Admin	Obtaining User Response for prediction	USN-5	As a admin, I want to obtain user response for prediction, so that I can improve the accuracy	I can obtain response data from the user.	Low	Sprint 3
Patient		USN-6	As a patient, I want to send my suggestions so that admin can improve the application accuracy.	I can send response to the admin	Low	Sprint 3
Admin	Monitoring user response and satisfaction	USN-7	As a admin, I want to monitor user response and satisfaction so that I can improve application	I can monitor the user responses and satisfaction.	Low	Sprint 3

			experience.			
Patient	Monitoring Model accuracy	USN-8	As a Patient, I want to monitor system accuracy so that I can believe prediction is correct.	I can check the prediction model accuracy.	Medium	Sprint 3
Admin		USN-9	As a Admin, I want to monitor system accuracy so that I can improve the prediction model.	I can check the prediction model accuracy	Medium	Sprint 3

6.3 Customer Journey



7. PROJECT PLANNING AND SCHEDULING

7.1 Sprint Delivery Schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team members
Sprint 1	Analysing and Visualizing Hospital health care data	USN-1	As a patient, I want to visualize the hospital health care data	10	Medium	Sanjaykumar, Jeya Ganesh
Sprint-1		USN-2	As a patient, I want the relationship between various attributes in the dataset	5	Medium	Jeswin, Rahul Hariesh
Sprint-2	Prediction of LOS	USN-3	As a patient, I want to predict length of stay in different hospitals so that I can plan accordingly	7	High	Jeswin, Sanjaykumar
Sprint-2		USN-4	As a Hospital manager, I want to predict the length of stay so that I can allot the hospital room accordingly	3	Medium	Jeya Ganesh

Sprint-2		USN-5	As a user, I want a easily understandable UI to get my prediction	2	Low	Rahul Hariesh
Sprint-3	Dashboard	USN-6	As a user, I want an interactive dashboard to understand the data easily	5	High	Sanjaykumar
Sprint-3		USN-7	As a patient, I want to find the available rooms in each hospital	3	Medium	Jeswin
Sprint-3		USN-8	As a user, I want to see be able to change the visualizations to my convenience	2	Low	Jeya Ganesh
Sprint-4	Monitoring user response and Model Accuracy	USN-9	As a Patient, I want to know the system accuracy so that I can believe prediction is correct	5	Medium	Rahul Hariesh
Sprint-4		USN-10	As a Patient, I want to give user response	2	Medium	Sanjaykumar
Sprint-4	Admin Login	USN-11	As the admin,I want to login to the admin dashboard.	3		Jeswin
Sprint-4	Admin Dashboard	USN-12	As the admin ,I need to be able to monitor the user	2	Medium	Sanjaykumar

			responses.			
Sprint-4		USN-13	As an admin , I want to be able to update the dataset for the model training and monitor the accuracy.	3	High	Jeya Ganesh

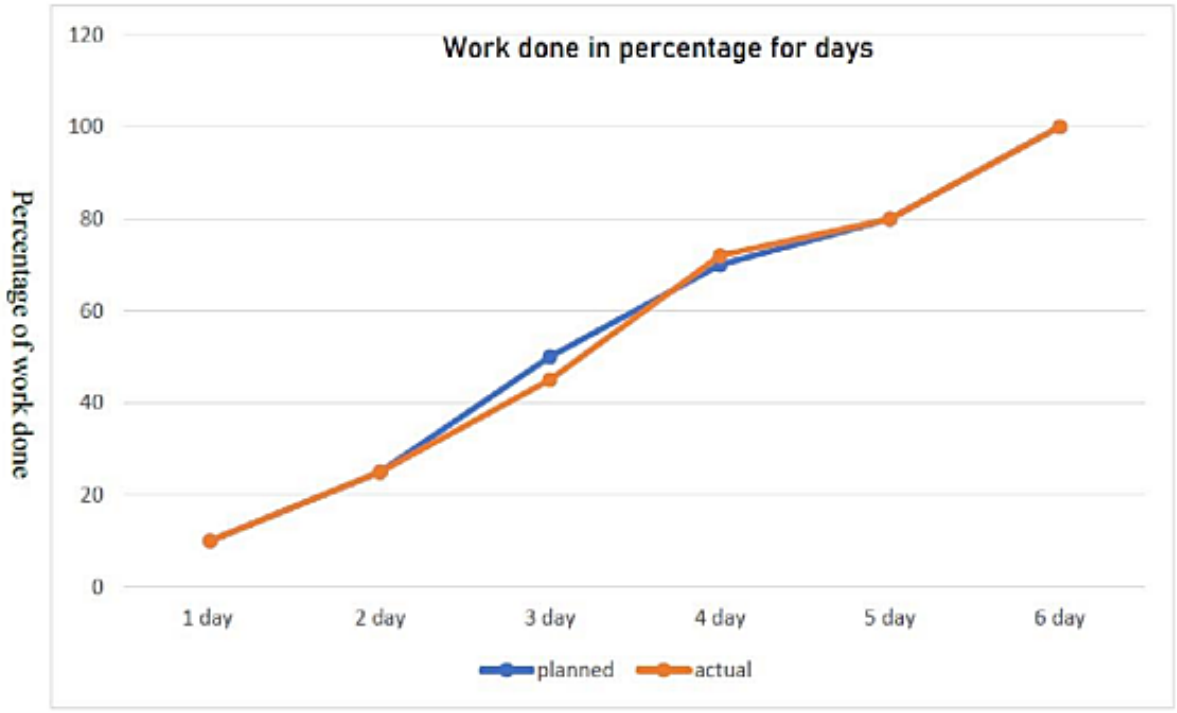
7.2 Project Tracker

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	15	6 Days	24 Oct 2022	29 Oct 2022	15	29 Oct 2022
Sprint-2	10	6 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	13	6 Days	07 Nov 2022	12 Nov 2022	13	12 Nov 2022
Sprint-4	15	6 Days	14 Nov 2022	19 Nov 2022	15	19 Nov 2022

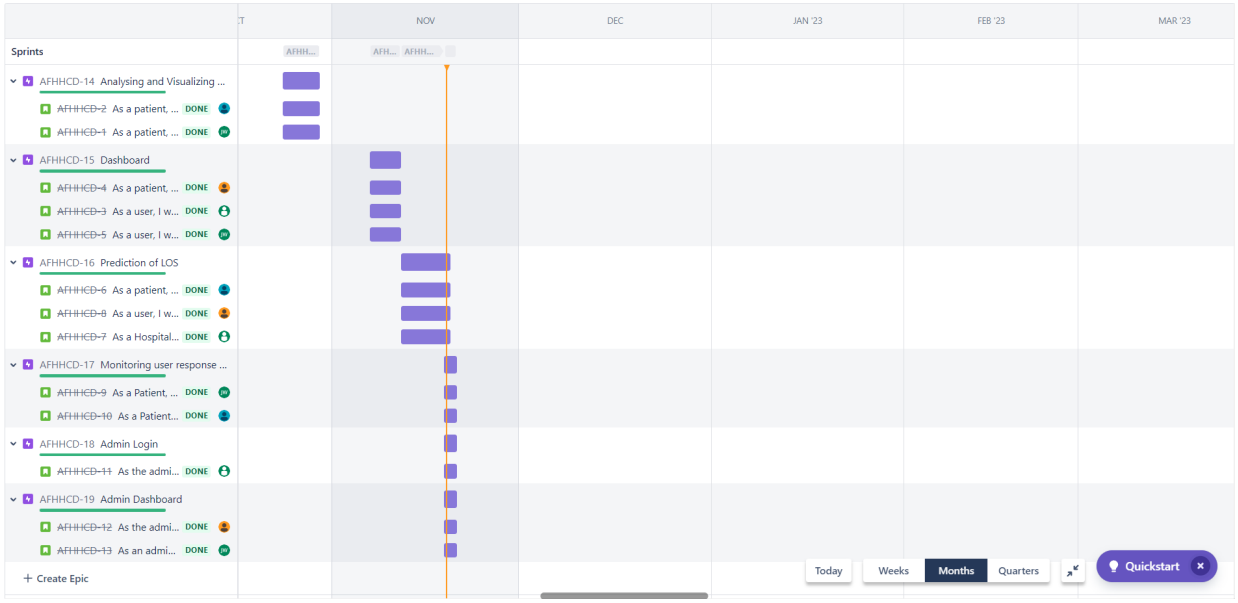
7.3 Velocity:

Sprint	Total Story points	Sprint duration	Average velocity
Sprint -1	15	6 days	15/6=2.5
Sprint -2	10	6 days	10/6=1.67
Sprint -3	13	6 days	13/6=2.16
Sprint -4	15	6 days	15/6=2.5

7.4 Burndown Charts



7.5 Reports from JIRA



8. IMPLEMENTATION AND OUTPUT SCREENSHOTS

Data Cleaning and Preprocessing

[9]
```
df["Bed Grade"].value_counts()
```
```
2.0    123671
3.0    110583
4.0     57566
1.0     26505
Name: Bed Grade, dtype: int64
```

[10]
```
df["Bed Grade"].fillna(2.0, inplace=True)
```

[11]
```
df["Bed Grade"].isna().sum()
```
```
0
```

[12]
```
df["City_Code_Hospital"].value_counts()
```
```
1     55351
2     51809
6     46991
7     35463
3     31569
5     31105
9     26277
11    17137
4     13857
10     5249
13     3630
Name: City_Code_Hospital, dtype: int64
```

[13]
```
df["City_Code_Patient"].fillna(1, inplace=True)
```

```
df.isna().sum()
```

```
Hospital_code                      0
Hospital_type_code                 0
City_Code_Hospital                 0
Hospital_region_code               0
Available Extra Rooms in Hospital  0
Department                         0
Ward_Type                          0
Ward_Facility_Code                 0
Bed Grade                          0
City_Code_Patient                  0
Type of Admission                  0
Severity of Illness                0
Visitors with Patient              0
Age                                0
Admission_Deposit                  0
Stay                               0
dtype: int64
```

Dropping unnecessary attributes

```
[8]
df.drop(['case_id', 'patientid'], axis=1, inplace=True)
```

Correlations in Dataset

```
[7]
df.corr()
```

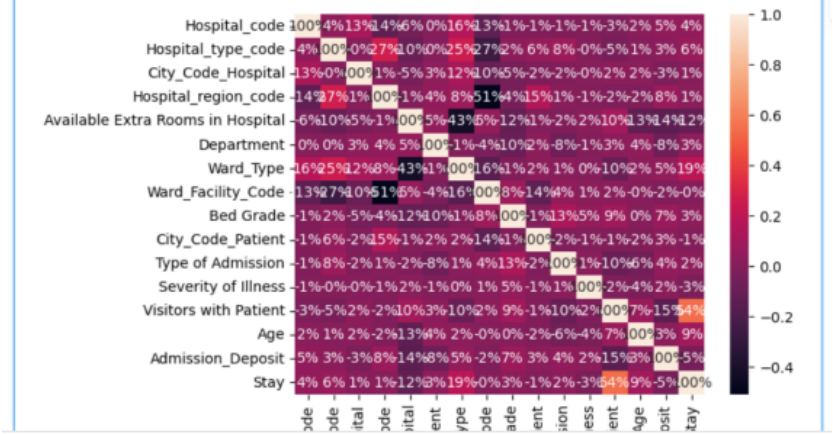
Visualize

	case_id float64 -0.045972258824...	Hospital_code fl... -0.059638371212...	City_Code_Hosp... -0.049309083896...	Available Extra ... -0.143739099539...	Bed Grade float64 -0.115867687...
case_id	1.0	-0.043022506716140094	-0.011351672166076155	0.04258049297268307	0.0137019181
Hospital_code	-0.043022506716140094	1.0	0.12829362483898515	-0.059638371212683096	-0.0137389593
City_Code_Ho...	-0.011351672166076155	0.12829362483898515	1.0	-0.04577097039417465	-0.0493090830
Available Extr...	0.04258049297268307	-0.059638371212683096	-0.04577097039417465	1.0	-0.115867682
Bed Grade	0.013701912168819724	-0.013738959637931984	-0.049309083896303674	-0.11586768762021678	
patientid	-0.004149891023962355	0.0022906150968089894	0.0007503734599816436	0.000920916992943411	0.00164487639
City_Code_Pa...	0.06519606662981017	-0.01552984445298396	-0.023988370312 9584	-0.009680986398408104	-0.0081054475
Visitors with...	0.001308943118 4283259	-0.028500291008472723	0.018184441177590032	0.09671435368821796	0.0889453 64
Admissi	-0.045972258824	0.045445524001	-0.034455292791	-0.143739099539	0.07383255

```
[42]
import seaborn

seaborn.heatmap(df.corr(), annot=True, fmt='.0%')
```

```
<AxesSubplot: >
```



Data Transformation

```
[16]
from sklearn import preprocessing
```

```
[17]
le = preprocessing.LabelEncoder()
```

```
[18]
le.fit(df["Stay"])
```

```
LabelEncoder()
```

```
[19]
transformed = le.transform(df["Stay"])
```

```
[20]
df["Stay"] = transformed
```

```
[21]
le.fit(df["Age"])
```

```
LabelEncoder()
```

```
[22]
age_transformed = le.transform(df["Age"])
```

```
[23]
df["Age"] = age_transformed
```

```
[24]
le.fit(df["Hospital_region_code"])
hrc_transformed = le.transform(df["Hospital_region_code"])
df["Hospital_region_code"] = hrc_transformed
```

```
[25]
le.fit(df["Department"])
dept_transformed = le.transform(df["Department"])
df["Department"] = dept_transformed
```

```
[26]
le.fit(df["Ward_Type"])
wt_transformed = le.transform(df["Ward_Type"])
df["Ward_Type"] = wt_transformed
```

```
[27]
le.fit(df["Ward_Facility_Code"])
wfc_transformed = le.transform(df["Ward_Facility_Code"])
df["Ward_Facility_Code"] = wfc_transformed
```

Transformed Data

```
[31]
df.head()
```

	Hospital_code	Hospital_type_c...	City_Code_Hosp...	Hospital_region...
0	8	2	3	2
1	2	2	5	2
2	10	4	1	0
3	26	1	2	1
4	26	1	2	1

5 rows, showing 10 per page Page 1 of 1

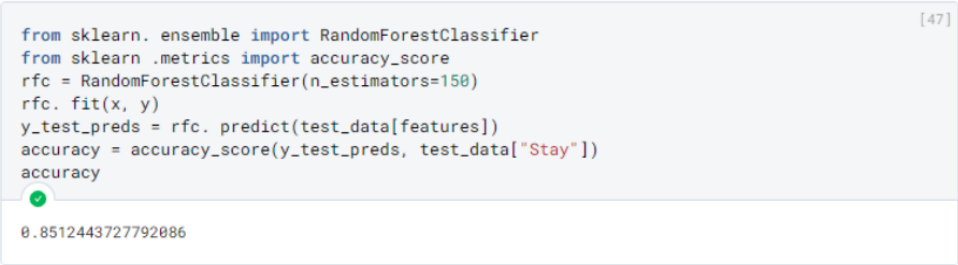
Prediction Using Models

Algorithm Used

1.Random Forest

● Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.
● It can be used for both Classification and Regression problems in ML.
● It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

Code and Accuracy



```
[47]
from sklearn. ensemble import RandomForestClassifier
from sklearn .metrics import accuracy_score
rfc = RandomForestClassifier(n_estimators=150)
rfc. fit(x, y)
y_test_preds = rfc. predict(test_data[features])
accuracy = accuracy_score(y_test_preds, test_data["Stay"])
accuracy

0.8512443727792086
```

The accuracy for prediction of length of stay using the algorithm Random forest is 85.12%

2.Random Forest

● Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique.
● It can be used for both Classification and Regression problems in ML.
● It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

Code and Accuracy

```
[47]
from sklearn. ensemble import RandomForestClassifier
from sklearn .metrics import accuracy_score
rfc = RandomForestClassifier(n_estimators=150)
rfc. fit(x, y)
y_test_preds = rfc. predict(test_data[features])
accuracy = accuracy_score(y_test_preds, test_data["Stay"])
accuracy

0.8512443727792086
```

The accuracy for prediction of length of stay using the algorithm Random forest is 85.12%

3.MLP Classifier

- MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network.
- Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.
- MLP classifier is a very powerful neural network model that enables the learning of non-linear functions for complex data.

Code and Accuracy

```
[49]
from sklearn.neural_network import MLPClassifier
clf = MLPClassifier(random_state=1, max_iter=1000).fit(x, y)
y_pred = clf.predict(test_data[features])
accuracy=accuracy_score(y_pred,test_data["Stay"])
accuracy

/shared-libs/python3.9/py/lib/python3.9/site-packages/sklearn/neural_network/_multilayer_perceptron.py:709: U
  warnings.warn("Training interrupted by user.")

0.5744106466652561
```

The accuracy for prediction of length of stay using the algorithm MLP classifier is 57.44%

4.Gaussian NB

● Naïve Bayes is a probabilistic machine learning algorithm used for many classification functions and is based on the Bayes theorem.
● Gaussian Naïve Bayes is the extension of naïve Bayes.
● While other functions are used to estimate data distribution, Gaussian or normal distribution is the simplest to implement as you will need to calculate the mean and standard deviation for the training data.

Code and Accuracy

The accuracy for prediction of length of stay using the algorithm Gaussian NB is 48.64%

DASHBOARD

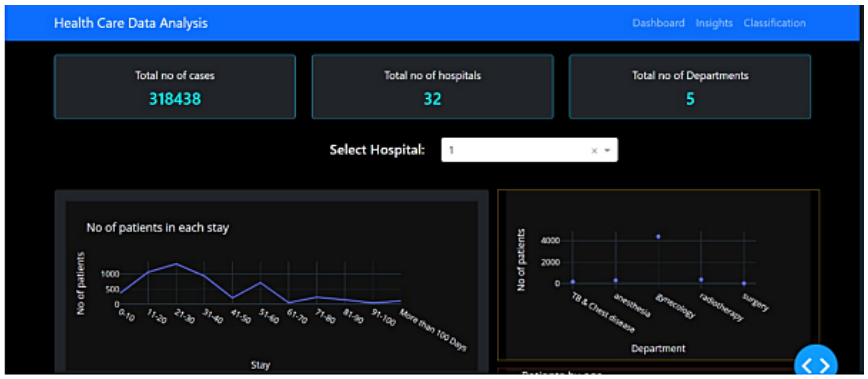
Tools Used - DASH - Python Framework

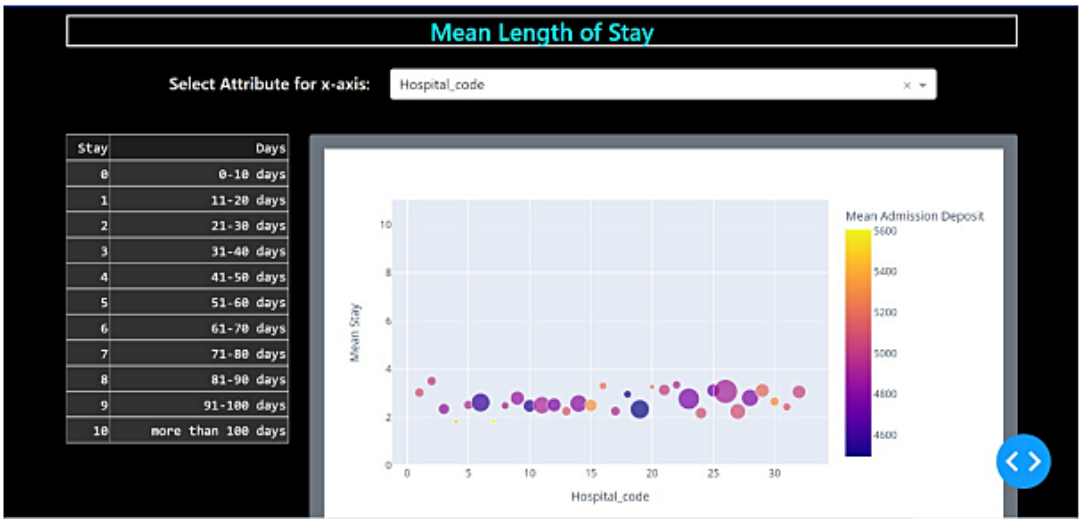
- Dash is an open-source Python framework used for building analytical web applications.
- It is a powerful library that simplifies the development of data-driven applications.
- It's especially useful for Python data scientists who aren't very familiar with web development.
- Users can create amazing dashboards in their browser using dash.

Description

This is the home page of our dashboard. It has navigation s like Dashboard, Insights and the Classification. It depicts the number of cases, number of hospitals, and the number of departments there are in that region. Here we can select the hospital id to look into more insights about the hospital.There are several graphs which show the patient's stay, and their particular department, severity of the particular patient, age wise category and finally the mean length of stay.

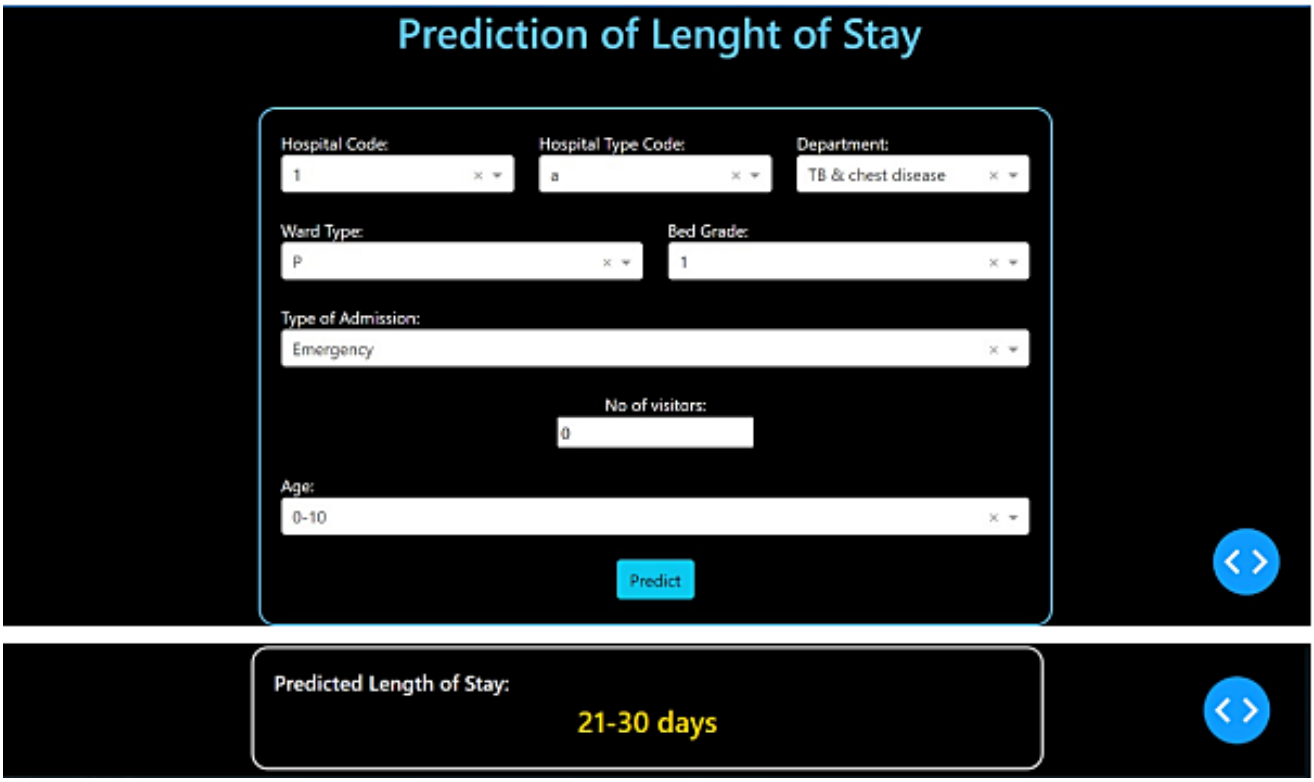
DASHBOARD PAGE:





Stay	Days
0	0-10 days
1	11-20 days
2	21-30 days
3	31-40 days
4	41-50 days
5	51-60 days
6	61-70 days
7	71-80 days
8	81-90 days
9	91-100 days
10	more than 100 days

PREDICTION OF LENGTH OF STAY PAGE:



ADMIN DASHBOARD

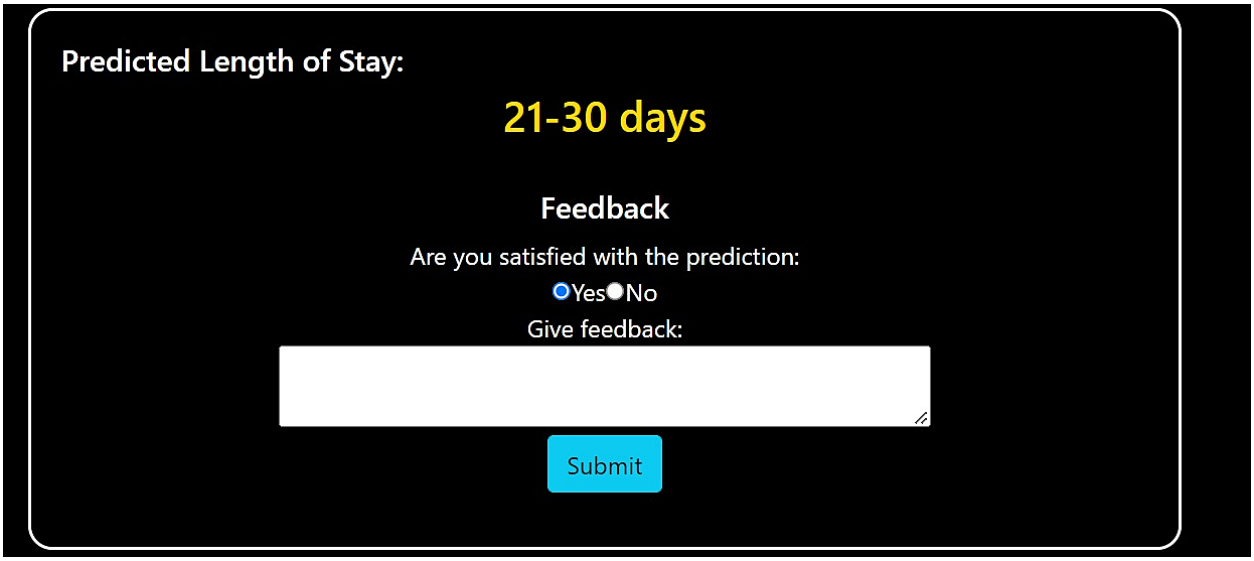
Tools Used

Firebase

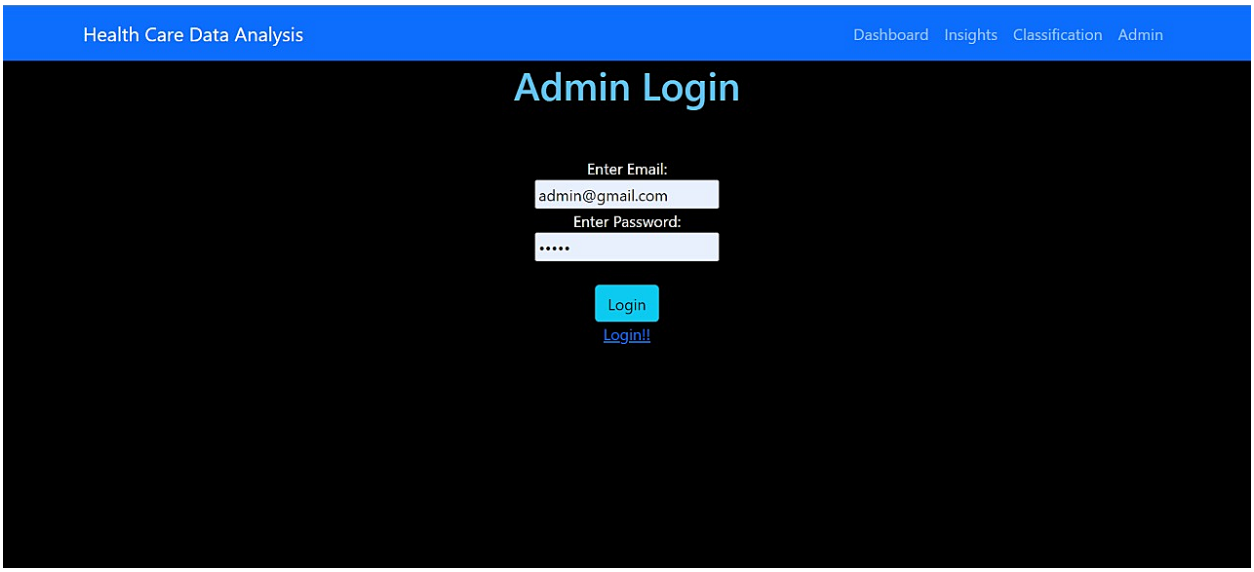
- Firebase is a set of hosting services for any type of application.
- It offers NoSQL and real-time hosting of databases, content, social authentication, and notifications, or services, such as a real-time communication server.
- Firebase helps you develop high-quality apps, grow your user base, and earn more money. Each feature works independently, and they work even better

together.

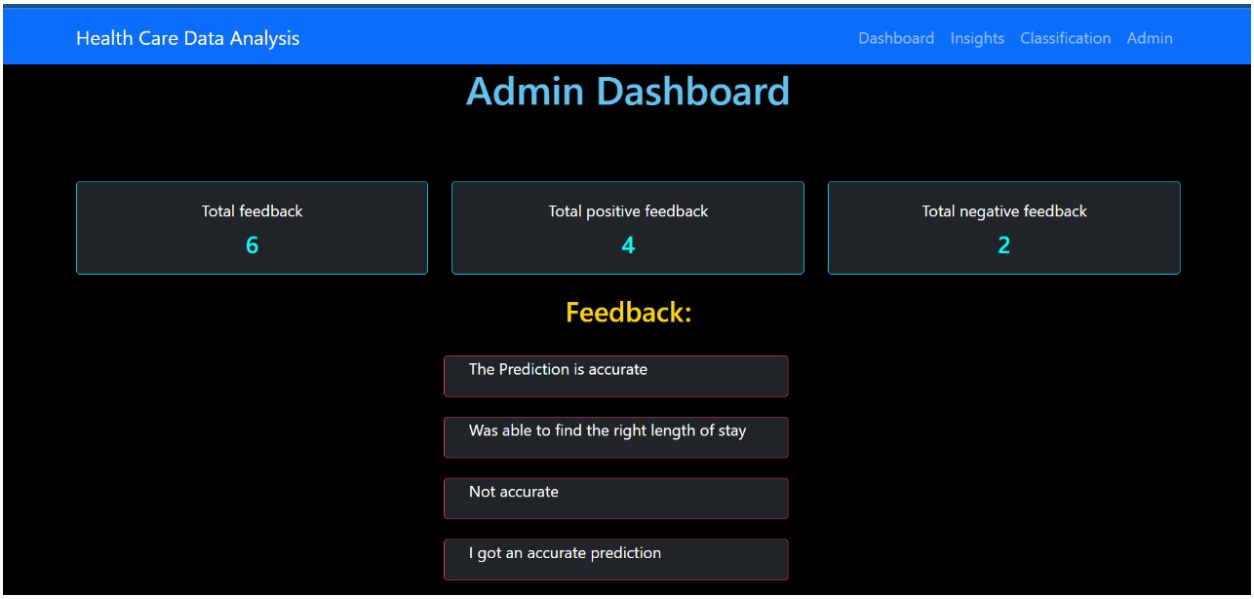
SATISFIED WITH THE PREDICTION PAGE



ADMIN LOGIN



ADMIN DASHBOARD - FEEDBACK OF THE PATIENTS ABOUT PREDICTION



9. TESTING

User Acceptance Testing

Purpose

- The purpose of this document is to briefly explain the test coverage and open issues of the Analytics for Hospitals Health-Care data project at the time of the release to User Acceptance Testing (UAT).
- This document mainly covers the severity of each resolution in the system and contains the severity score of the resolution mentioned below in the table.

Defect Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Model	50	0	3	47
Client Application	37	0	3	34

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	5	1	2	3	11
Duplicate	0	0	3	0	3
External	0	3	0	1	4
Fixed	7	2	4	2	15
Not Reproduced	0	0	1	0	1
Skipped	0	0	0	3	3
Won't Fix	0	1	0	1	2
Totals	12	7	10	10	39

Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Homepage	4	0	0	4
Insights	2	0	0	2
Classification	3	0	0	3
Login	2	0	0	2
Admin page	4	0	0	4

Performance Testing

Project team shall fill the following information in the model performance testing template.

S.No	Parameter	Values
1.	Metrics	Regression Model: Random Forest Accuracy: 85.12%
2.	Tune the Model	Hyperparameter Tuning 1) Learning Rate: [0.01, 0.03, 0.05, 0.07] 2) Max features: ['auto','sqrt'] 3) Number of Estimators: [10,20,30,50] 4) min_samples_leaf : [2,4,6] Validation Method: Grid Search Cross Validation Best Parameters: Learning Rate – 0.07 Number of Estimators - 300

Load Testing

No.Of Users:1

```
Get Method status is  200
Post Method status is  200
Get Method status is  200
Post Method status is  200
Get Method status is  200
Post Method status is  200
Get Method status is  200
Post Method status is  200
Get Method status is  200
Post Method status is  200
Get Method status is  200
Post Method status is  200
Get Method status is  200
```

Statistics | Charts | Failures | Exceptions | Download Data

Type	Name	# Requests	# Fails	Median (ms)	90%ile (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS	Current Failures/s
GET	/	17	0	36	330	85	28	336	1256	0.5	0
POST	/?status=success	17	0	120	290	143	73	410	1256	0.4	0
	Aggregated	**34**	**0**	**110**	**290**	**114**	**28**	**410**	**1256**	**0.9**	**0**

No.Of Users:20

Type	Name	# Requests	# Fails	Median (ms)	90%ile (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS	Current Failures/s
GET	Homepage	17	0	140	530	218	134	535	1155	0.1	0
GET	Homepage - Search	15	0	150	300	192	138	316	5119	0.1	0
GET	Classification	15	0	150	270	175	139	280	4108	0.1	0
POST	Classification data	15	0	290	600	399	271	1210	4730	0.1	0
GET	feedback	7	0	140	270	159	137	267	3391	0.1	0
GET	Admin login	2	0	140	140	141	140	141	3387	0	0
GET	Admin homepage	2	0	140	140	140	140	140	3937	0	0
GET	Hompage - visualization	2	0	141	1200	667	141	1192	3653	0.1	0
GET	Insights	2	0	150	320	234	150	318	3393	0	0
GET	Insights - select box	15	0	280	490	345	274	661	4503	0.1	0
	Aggregated	**92**	**0**	**160**	**420**	**259**	**134**	**1210**	**3794**	**0.7**	**0**

10. RESULTS

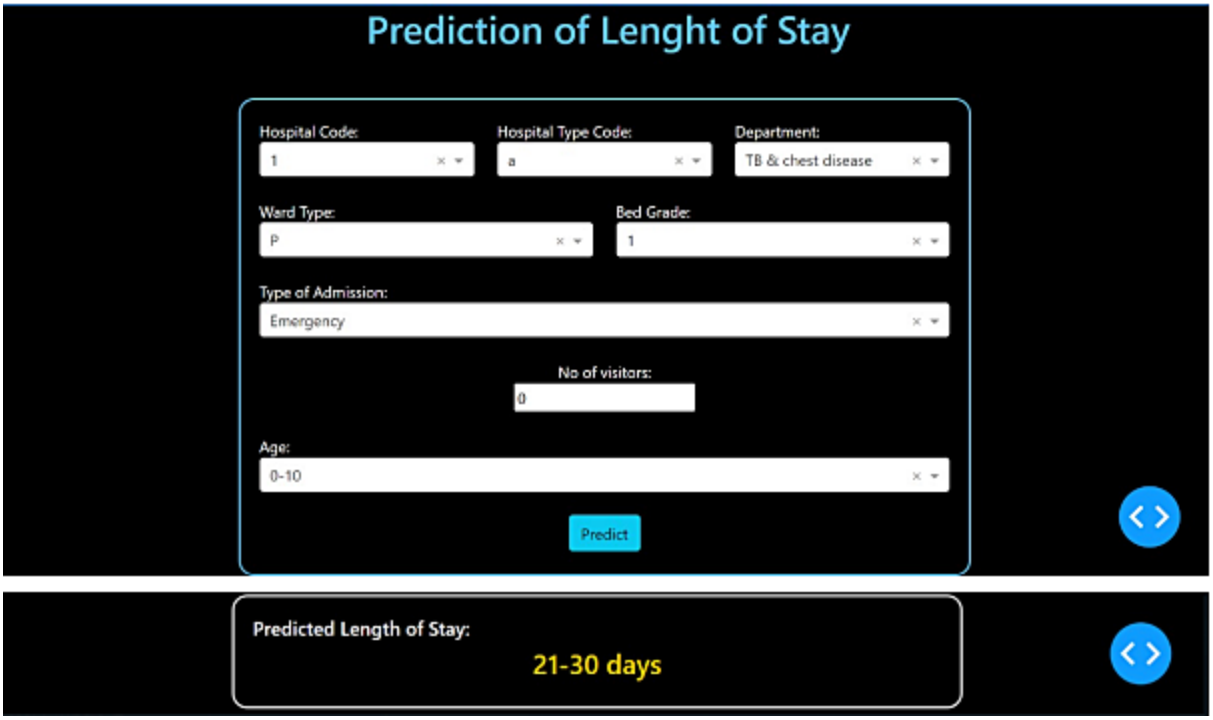
Using the different prediction models, Random Forest Algorithm has highest accuracy, so we chose this for our prediction of length of stay.

```
[47]
from sklearn. ensemble import RandomForestClassifier
from sklearn .metrics import accuracy_score
rfc = RandomForestClassifier(n_estimators=150)
rfc. fit(x, y)
y_test_preds = rfc. predict(test_data[features])
accuracy = accuracy_score(y_test_preds, test_data["Stay"])
accuracy

0.8512443727792086
```

The accuracy for prediction of length of stay using the algorithm Random forest is 85.12%

Github Repo Link :https://github.com/IBM-EPBL/IBM-Project-31564-1660202873.git



11. BENEFITS

- Accurate prediction of patient LOS may aid the healthcare specialists to take medical decisions and allocate medical team and resources.
- The patient and insurance companies may use this prediction to manage their budget.
- Patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of getting a hospital-acquired condition such as staph infection.
- Prior knowledge of LOS can aid in logistics such as room and bed allocation planning.

12. CONCLUSION

Thus this project of Analytics for hospital health care data helps the patients in making plans of staying in the hospital and also it helps the other patients to know the capacity of bed available in hospitals during the pandemic times.This study uses data visualization and analytics to show analytics for hospital and healthcare data. This data is acquired from various health information systems and other technical tools used by government agencies, insurance providers, and healthcare professionals. Real-time analysis of the data being gathered allows for a better understanding of the virus's effects and the forecasting of future trends, which will help us contain the spread and stop further outbreaks. If used appropriately, health care data management could result in better treatment.The collection and analysis of data from the healthcare industry with the aim of gaining insights and influencing decision-making can be referred to as healthcare analytics.

13. FUTURE SCOPE

The data analytics market in the healthcare space has only increased over the last few years.Decision-making is improved since guessing and manual duties are eliminated by data analytics. whether it be selecting the appropriate content, organizing marketing initiatives, or creating products.

Organizations can use the data analytics insights they uncover to make wise decisions. resulting in improved results and customer satisfaction. After the Affordable Act was passed, the necessity for data analytics to meet business goals of pharmaceutical firms, payers, insurance companies, physicians, hospitals, medical equipment companies, sales reps, and other players in the healthcare industry only grew.

14. REFERENCE

[1] R. Vargheese, "Dynamic Protection for Critical Health Care Systems Using Cisco CWS: Unleashing the Power of Big Data Analytics," 2014 Fifth International Conference on Computing for Geospatial Research and Application, 2014, pp. 77-81, doi: 10.1109/COM.Geo.2014.28.

[2] M. Panda, S. M. Ali and S. K. Panda, "Big data in health care: A mobile based solution," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), 2017, pp. 149-152, doi: 10.1109/ICBDACI.2017.8070826.

[3] S. Balaji and V. Prasathkumar, "Dynamic Changes by Big Data in Health Care," 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-4, doi: 10.1109/ICCCI48352.2020.9104168.

[4] A. Alahmar and R. Benlamri, "Optimizing Hospital Resources using Big Data Analytics with Standardized e-Clinical Pathways," 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2020, pp. 650-657, doi: 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00112.

[5] Z. Yu et al., "Health Service Decision Toolbox (HSDT): Delivering the Right Treatment to the Right Patient with Health Information Technology and Data Analytics," 2018 15th International Conference on Service Systems and Service Management (ICSSSM), 2018, pp. 1-6, doi: 10.1109/ICSSSM.2018.8465110.