ANALYTICS FOR HOSPITALS HEALTH -CARE DATA

NALAIYATHIRAN PROJECT REPORT

TEAMID-PNT2022TMID30370

D.LEONY ELAVARASI	611419104035
E.PAVITHRA BHARATHI	611419104050
S.SATHYA	611419104073
N.SNEKA	611419104080
M.VEERAJOTHI	611419104098

In partial fulfillment of the award of the degree

of

BACHELOR OF ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

MAHENDRA ENGINEERING FOR WOMEN

ANNA UNIVERSITY - CHENNAI 600 025

INTRODUCTION

1.1 Projectoverview

Recent Covid-19 Pandemic has raised alarms over one of the mostoverlooked areas to focus: Healthcare Management. While healthcaremanagement has various use cases for using data science, patient length ofstayisonecriticalparametertoobserveandpredictifonewantstoi mprovethe efficiency of the healthcare management in a hospital. This parameterhelps hospitals to identify patients of high LOS-risk (patients who will staylonger) at the time of admission. Once identified, patients with high LOSriskcanhavetheirtreatmentplanoptimizedtominimizeLOSan dlowerthe chance of staff/visitor infection. Also, prior knowledge of LOS can aidin logistics such as room and bed allocation planning. Suppose you havebeen hired as Data Scientist of Health Man – a not for profit organizationdedicated to manage the functioning of Hospitals in a professional andoptimal manner. While healthcare management has various use cases forusingdatascience,patientlengthofstayisonecriticalparameterto observe and predict if one wants to improve the efficiency of the healthcaremanagement in a hospital. This parameter helps hospitals to identifypatients of high LOS-risk (patients who will stay longer) at the time ofadmission. Once identified, patients with high LOS risk can have theirtreatment plan optimized to minimize LOS and lower the chance ofstaff/visitor infection. Also, prior knowledge of LOS can aid in logisticssuch as room and bed allocation planning. Suppose you have been hired asData Scientist of Health Man – a not for profit organization dedicated tomanagethefunctioningofHospitalsinaprofessionalandoptimal manner.

1.2 Purpose

Data analytics in health care is vital. It helps health care organizationsto evaluate and develop practitioners, detect anomalies in scans and predictoutbreaks in illness, per the Harvard Business School. Data analytics canalso lower costs for health care organizations and boost businessintelligence. Hospitaldataanalyticscanlookoverpatientdataandanyprescribed medication to alert doctors and patients of incorrect dosages orwrong prescriptions, which lessens human error and the cost to yourhospital.

2. LITERATURESURVEY

2.1 Existing Problem

i. Thealreadyexistingmodelistrainedwithmini malparametersbyleavingthenecessary parameter
ii. Lowaccuracyinprediction
iii. Nofeatureextractiondone
iv. Highcomplexity.

2.2 References

1. Yang J.-J., Li J., Mulder J., Wang Y., Chen S., Wu H., Wang Q., Pan H.Emerginginformation technologies for enhanced healthcare. *Comput. Ind.*2015;69:3–11.doi:10.1016/j.compind.2015.01.012.[CrossRef] [GoogleScholar]
2. Cortada J.W., Gordon D., Lenihan B. *The Value of Analytics in Healthcare.* IBMInstitute for Business Value; Armonk, NY, USA: 2012. Report No.: GBE03476-USEN-00.[Google Scholar]
3. Center for Medicare and Medicaid Services. [(accessed on 1 August 2017)];Availableonline: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-andReports/NationalHealthExpendData/NationalHealthAccountsHistorical.html
4. Berwick D.M., Hackbarth A.D. Eliminating waste in US health care. *J. Am. Med.Assoc.* 2012;307:1513–1516. doi: 10.1001/jama.2012.362. [PubMed] [CrossRef][GoogleScholar]
5. Makary M.A., Daniel M. Medical error-the third leading cause of death in the US.*Br. Med. J.* 2016;353:i2139. doi: 10.1136/bmj.i2139. [PubMed] [CrossRef][GoogleScholar]
6. Prokosch H.-U., Ganslandt T. Perspectives for medical informatics. *Methods Inf.Med.* 2009;48:38–44.doi:10.3414/ME9132.[PubMed][CrossRef][GoogleScholar]
7. Simpao A.F., Ahumada L.M., Gálvez J.A., Rehman M.A. A review of analyticsandclinicalinformatics in health care.*J.Med.Syst.*2014;38:45.doi:

10.1007/s10916-014-0045-x.[PubMed][CrossRef][GoogleScholar]

8. Ghassemi M., Celi L.A., Stone D.J. State of the art review: The data revolutionincritical care. *Crit. Care.* 2015;19:118. doi: 10.1186/s13054-015-0801-4. [PMCfreearticle] [PubMed] [CrossRef][GoogleScholar]

9. Tomar D., Agarwal S. A survey on Data Mining approaches for Healthcare. *Int. J.Bio-Sci. Bio-Technol.* 2013;5:241–266. doi: 10.14257/ijbsbt.2013.5.5.25. [CrossRef][GoogleScholar]

10. Panagiota Galetsia , Korina Katsaliakia , Sameer Kumarb,* a School ofEconomics,BusinessAdministration&LegalStudies,InternationalHellenicUniversity,14thkm
Thessaloniki-N. Moudania, Thessaloniki, 57001, Greece b Opus College of Business,University of St. Thomas Minneapolis Campus, 1000 LaSalle Avenue, Schulze Hall435, Minneapolis,MN55403,USA

11. K. Jee and G. H. Kim, "Potentiality of big data in the medical sector: Focus onhow to reshape the healthcare system," *Healthc. Inform. Res.*, vol. 19, no. 2, pp.79–85,Jun.2013.doi:10.4258/hir.2013.19.2.79

12. J. King, V. Patel, and M. F. Furukawa, "Physician adoption of electronic healthrecord technology to meet meaningful use objectives: 2009–2012," The Office ofthe National Coordinator for Health Information Technology, Tech. Rep., Dec.2012.

13. V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That WillTransformHowWe Live,Work,andThink*.EamonDolan,2014.

14. S. Axryd. Why 85% of big data projects fail. [Online].

Available:https://www.digitalnewsasia.com/insights/why-85-big-data-projects-fail.Accessedon:Apr.16,2019.

15. W.H. Organization et al., ”How can hospital performance be measured andmonitored?,” in How can hospital performance be measured and monitored? 2003.p.17–17.

16. Y.Varabyova,J.Schreyögg

Internationalcomparisonsofthetechnicalefficiencyofthehospitalsecto r:paneldata analysis of oecd countries using parametric and non-parametric approachesHealth Policy,112(1) (2013),pp.70-79

17. D.Heyland,D.Cook,S.M.Bagshaw,A.Garland,H.T.Stelfox,S.M ehta,P.Dodek,J.

Kutsogiannis,K.Burns,J.Muscedere,*etal.*

The very elderly admitted to icu: a quality finish?CritCare Med,43(7)(2015),pp.1352-1360

18. Teno JM, Fisher E, Hamel MB, Wu AW, Murphy DJ, Wenger NS, et al. Decision-making and outcomes of prolonged icu stays in seriously ill patients. J Am GeriatrSoc2000;vol.48,no.S1.

19. A.Hunter,L.Johnson,A.Coustasse

Reduction of intensive care unit length of stay: the case of early mobilizationHealth Care Manager,33 (2) (2014),pp.128-135

20. ve Maliyetleri S. ”Characteristics, outcomes and costs of prolonged stay icupatients;”2011.

21. A.A.Kramer,J.E.Zimmerman

Apredictivemodelfortheearlyidentificationofpatientsatriskforaprolongedi ntensivecareunitlengthofstay

BMCMed InformDecis Making,10 (1)(2010),p.27

22. A.Pérez,W.Chan,R.J.Dennis

Predictingthelengthofstayofpatientsadmittedforintensivecareusingafirststepanalysis
HealthServOutcomesResMethod,6(3–4)(2006),pp.127-138

23. J.Rapoport,D.Teres,Y.Zhao,S.Lemeshow

Length of stay data as a guide to hospital economic performance for icu patientsMed Care,41(3) (2003),pp.386-397

24. F. Barili, N. Barzaghi, F.H. Cheema, A. Capo, J. Jiang, E. Ardemagni, M.Argenziano,C.Grossi

An original model to predict intensive care unit length-of stay after cardiac surgery inacompetingriskframework

IntJ Cardiol,168 (1)(2013),pp.219-225

25. M.N.Diringer,N.L.Reaven,S.E.Funk,G.C.Uman

ElevatedbodytemperatureindependentlycontributestoincreasedlengthofstayinneurologicintensivecareunitpatientsCrit CareMed,32 (7)(2004),pp.
1489-1495

26. R. Paterson, D. MacLeod, D. Thetford, A. Beattie, C. Graham, S. Lam, D.BellPredictionofin-hospitalmortalityandlengthofstayusinganearlywarningscoring system:clinicalaudit

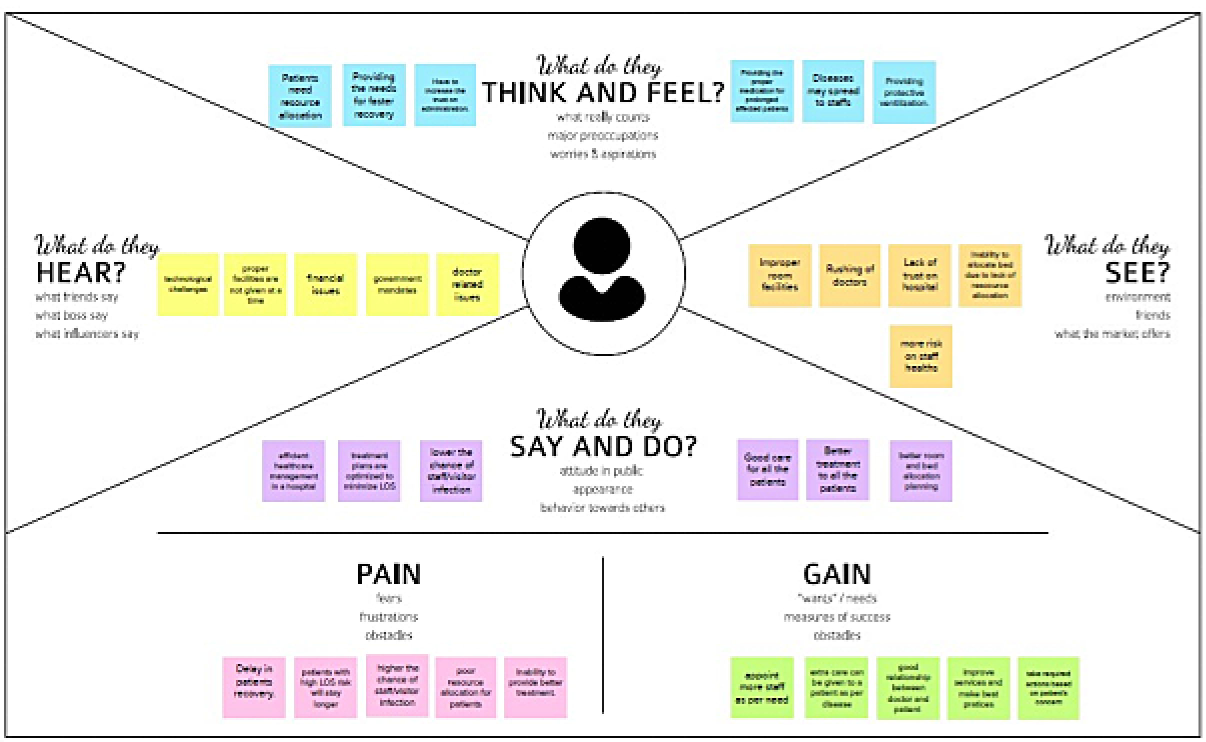
ClinMed,6(3)(2006),pp.281-284

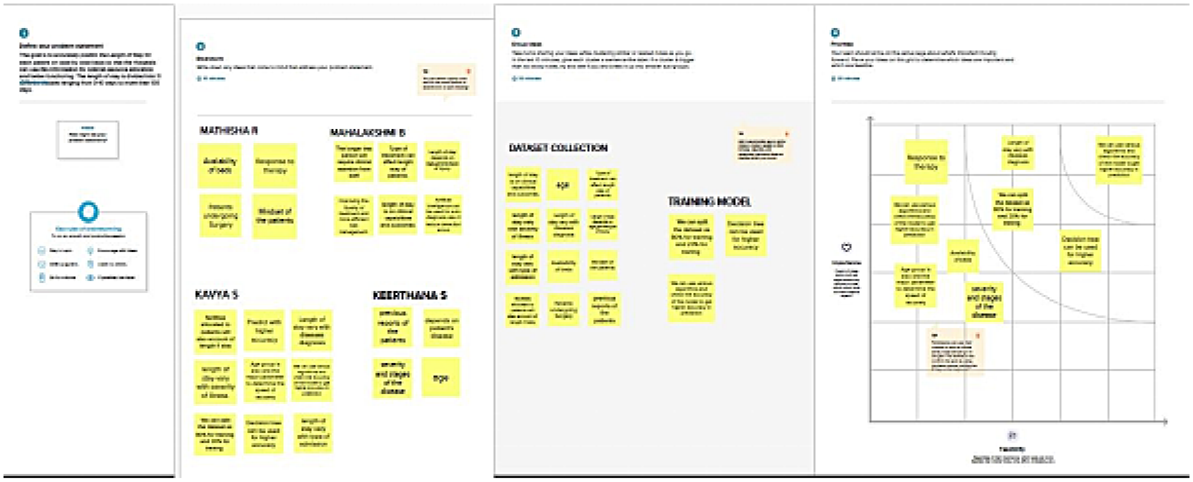
2.3 ProblemStatementDefinition

v. ThegoalistoaccuratelypredicttheLengthofStayfor eachpatientoncaseby casebasissothattheHospitalscanusethisinformationforoptimalresourceallocationandbetterfunctioning.

vi. The length of stay is divided into 11 different classes ranging from 0-10daysto morethan 100 days.

3. IDEATIONANDPROPOSEDSOLUTION





3.3ProposedSolution

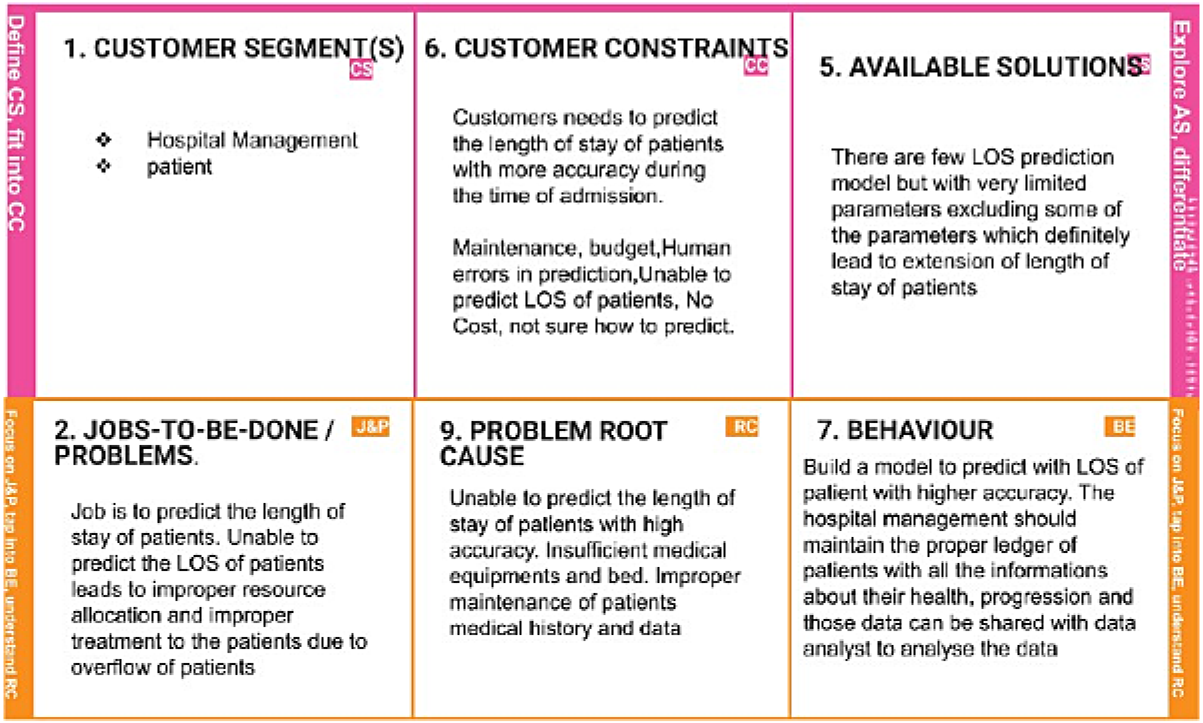
Predictthelengthofstayofpatients.

The length of the stay can be predicted using either Random forest orDecision Tree for more accuracy. Certain parameters like age, stage of thediseases, disease diagnosis, severity of illness, type of admission, facilitiesallocated, etc., are used for prediction. IBM Cognos will be used for dataanalytic s. The model will be trained using colab.It predicts the length ofstay(LOS)ofthepatientswithmoreaccuracy.Asaresultproperresourcesand therapy can be provided.Patients can get proper treatment and bettermedical care than before which helps them for their faster recovery. So theprediction minimizes the overflow of patients and helps in resourcemanagement and optimize their resource utilization. Hence this leads tofaster recovery and lower the expenses for treatment. It improves the trustin hospital management. It avoids the major risk of spreading

infectionamong the hospital staff. This leads tosafety of hospital staff andpatients.Resource consumption is optimized.This model can be used by allgovernment hospitals, private hospitals, and even in The model is trainedwiththerealworldhospitalsurveyforbetterpredictiosmallc linics.Lengthofthestaywillbepredictedwithmoreaccuracy.This model predicts the length of the stay for all kinds of patients and predictswithmoreaccuracy.

3.4 Problem Solution fit

a.



4. REQUIREMENT ANALYSIS

4.1 Functionalrequirement

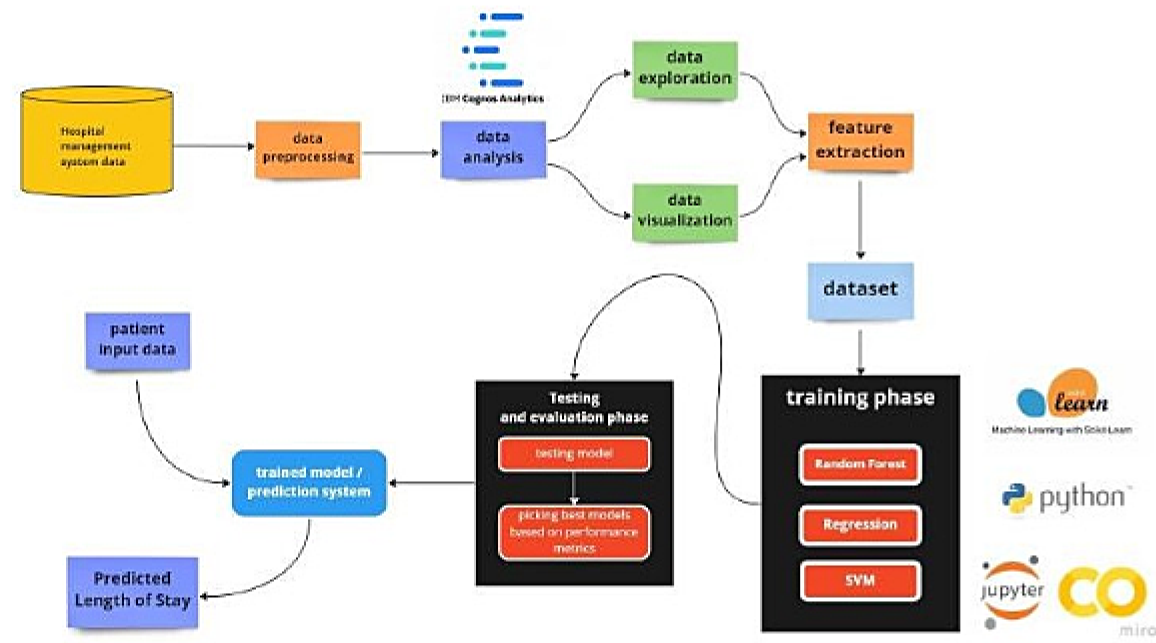
FunctionalRequirement(Epic)	SubRequirement(Story/Sub-Task)
collectDataset	Datafromdifferentsourcesarecollectedinordertogetoptimized result
Datacleaning	When combining data from multiple sources theareduplicateddata andhencewecleanthedata1st
Datamodelling	Identifytherelationshipbetweenvariousparameters.
Predictionandanalysis	ThelengthofstayispredictedwiththeMachinelearningalgorithm

4.2 Non-Functionalrequirements

Non-FunctionalRequirement	Description
Usability	Usercanviewandvisualisethedatathroughthe interactive dashboard and predict the length ofstayofpatientswithmachinelearningalgorithm
Security	IBM Cognos provides better security. Thedataset uploaded to the dashboard cannot bedownloadedoraccessedbyexternalsources
Reliability	Thedashboardandthepredictionisveryreliableandprovide predictionwith moreaccuracy
Performance	Thelengthofstayofpatientsispredictedwithmoreaccuracy
Availability	The predictedlengthofstayandthevisualizationwillbeavailableincognosanalys is
Scalability	The software is scalable and extendable.Becauseitallowmultipleusertohandlethedataatthesame time

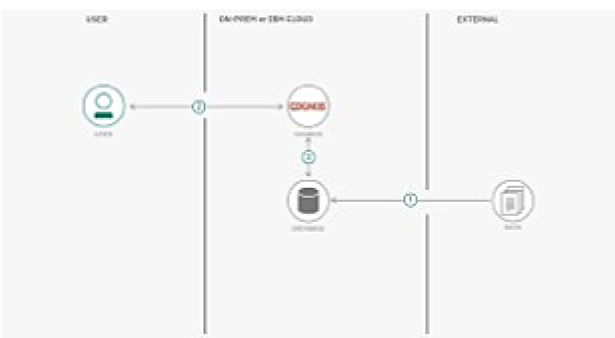
5. PROJECTDESIGN

5.1 DataFl



Example: (Simplified)

5.2 Solution&TechnicalArchitecture



5.3 UserStories

UserType	FunctionalRequirement (Epic)	UserStory Number	UserStory/Task	Acceptancecriteria
Customer	Dashboard	USN-1	As a user, I can uploadthedataset to the dashboard	Icanaccessdashboard
	View	USN-2	Asauser,Icanviewthepatientdetails	I can visualizethedata
Admin	Analyse	USN-3	As a user, I will analysethegiven dataset	I can analysethedataset

	Predict	USN-4	Asauser,Iwillpredictthelength ofstay	I can predictthelengthof stay	
	Collectdata	USN-5	AsaanalystIneedtocollectthedataset		
	Preparedata	USN-6	AsananalystIneedtodofeatureextraction	Icanextractthe parametersthat haveimpactthe lengthofstay	
Visualization	Dashboard	USN-7	As a user I can preparedata by usingvisualizationtechnique	I can preparethe data withvisualization technique	

6. PROJECTPLANNING

6.1 SprintPlanning & Estimation

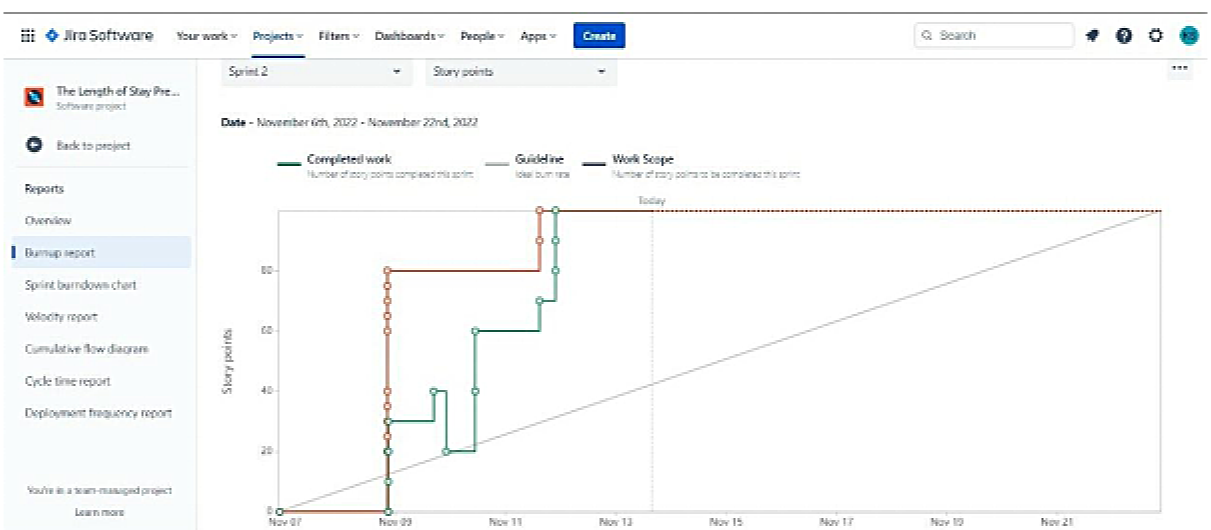
Sprint	FunctionalRequirement(Epic)	User StoryNumber	UserStory/Task	StoryPoints	Priorty
Sprint-1	Registration	USN-1	AsahealthcareproviderI cancreateanaccountin IBMcloudandthedata arecollected..	10	High
Sprint-1	Analyze	USN-2	As a health care provider allthedatathatarecollected is learnedanduploadedinthe databaseorIBMcloud.	10	Medium
Sprint-1	Feature Extraction	USN-3	AsahealthcareproviderI can visualizehowvariousparameters affectthelengthofstayofpatients anddofeatureextractionforbetter prediction	10	Medium

6.2 SprintDelivery Schedule

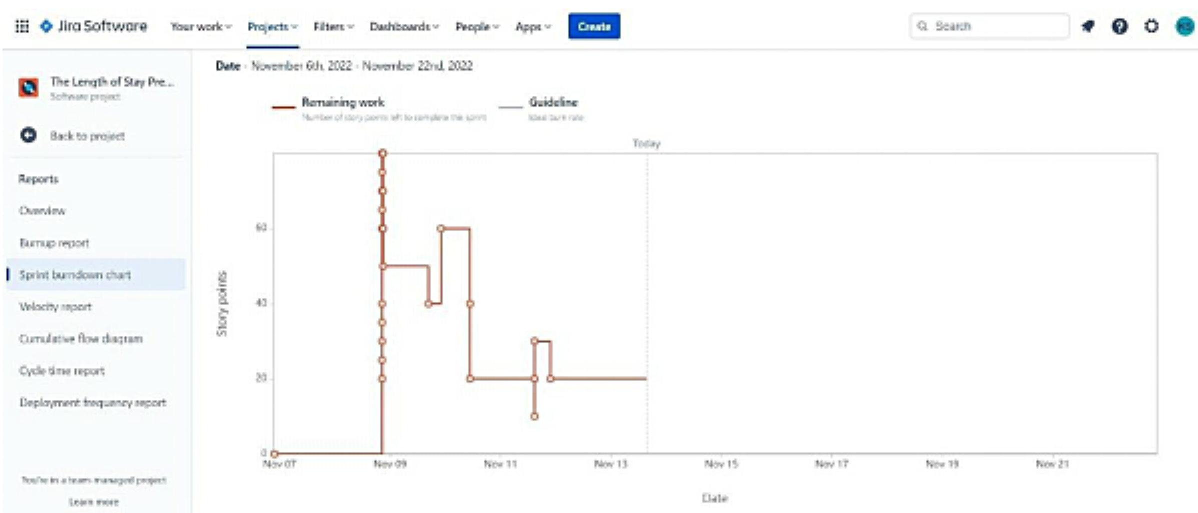
FunctionalRequirement(Epic)	User StoryNumber	UserStory/Task	StoryPoints	Priority	TeamMembers
Visualization	USN-4	Asahealthprovider Ican preparedataformy visualization.	2 0	Medium	Maha LakshmiB, Keerthana S,KavyaS, MathishaR
Dashboard	USN-5	AsahealthcareproviderIcanuse my accountinmy dashboardforuploading dataset.	2 0	High	MahaLakshmiB, Keerthana S,KavyaS,MathishaR
Prediction	USN-6	As a health care provider Icanpredictthelengthof stay	2 0	HIgh	MahaLakshmiB, Keerthana S,KavyaS, MathishaR

6.3 ReportsfromJIRA

Burnt UpChart

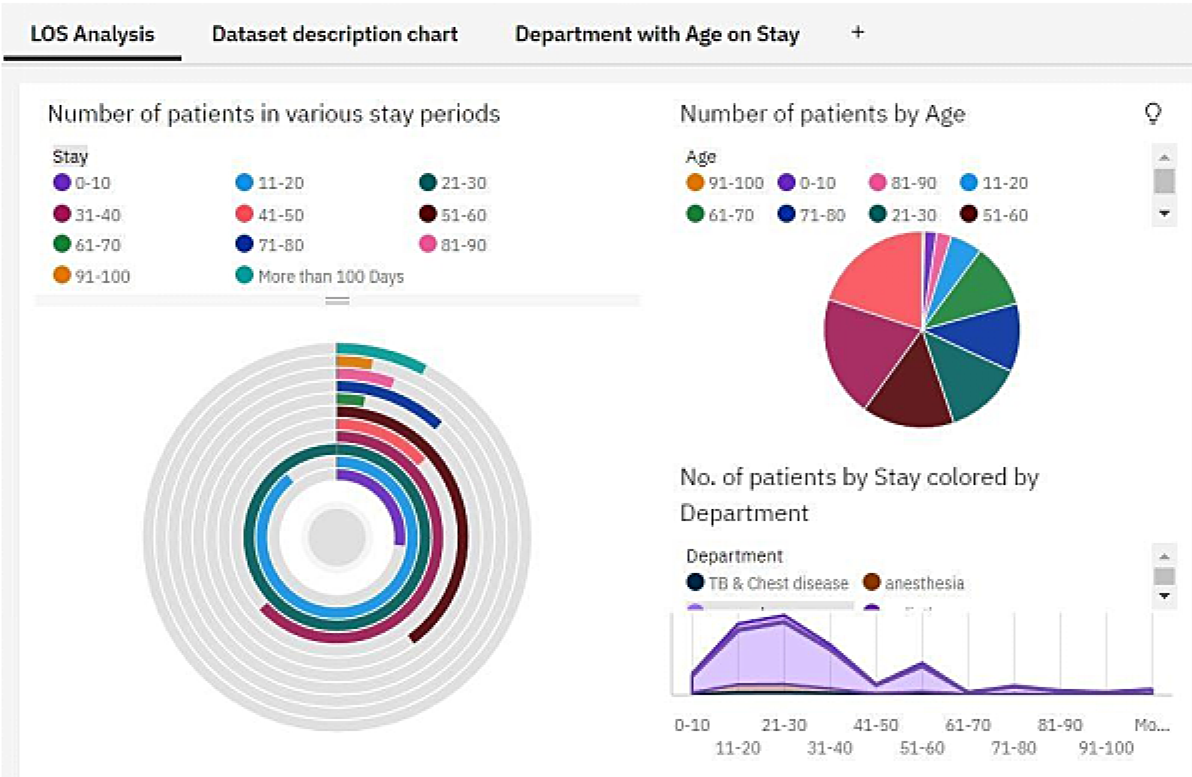


BurntDownChart



7. CODING & SOLUTIONING (Explain the features added inthe projectalong with code)

7.1 Feature1




```
X_train.fillna(0,inplace=True)
Y_train.fillna(0,inplace=True)
X_test.fillna(0,inplace=True)
```

K-Nearest Neighbor Algorithm

```
knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_train, Y_train)
Y_pred = knn.predict(X_test)
acc_knn = round(knn.score(X_train, Y_train) * 100, 2)
acc_knn
```

53.99

Descision Tree Algorithm

```
decision_tree = DecisionTreeClassifier()
decision_tree.fit(X_train, Y_train)
Y_pred = decision_tree.predict(X_test)
acc_decision_tree = round(decision_tree.score(X_train, Y_train) * 100, 2)
acc_decision_tree
```

99.76

Random Forest Algorithm

```
random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(X_train, Y_train)
Y_pred = random_forest.predict(X_test)
random_forest.score(X_train, Y_train)
acc_random_forest = round(random_forest.score(X_train, Y_train) * 100, 2)
acc_random_forest
```

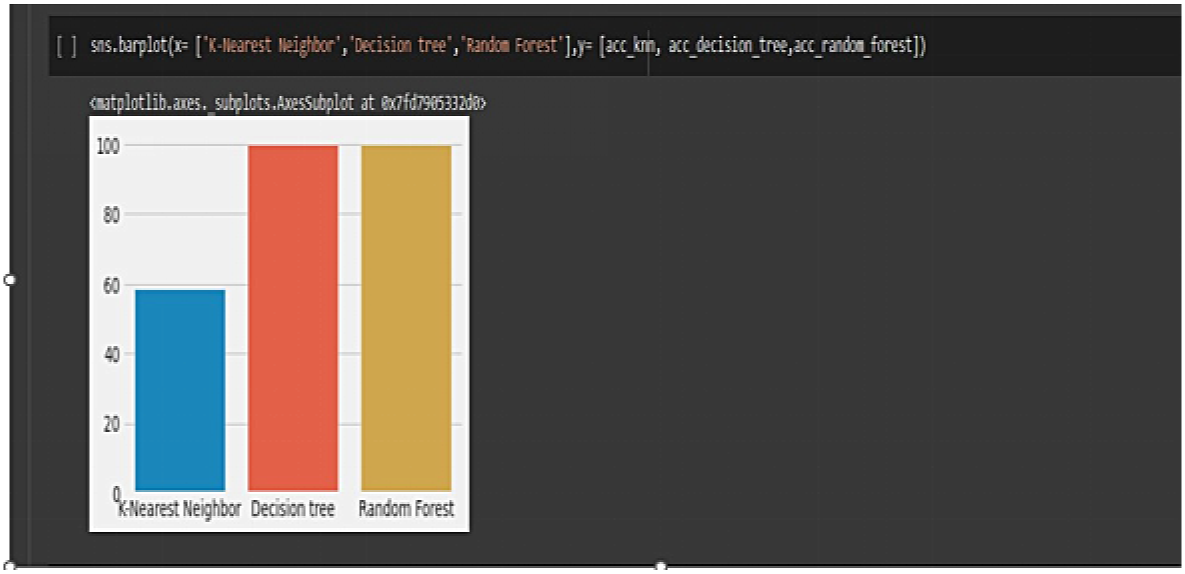
99.76

Prediction accuracy comparison

7.2 Feature2

8. RESULTS

8.1 PerformanceMetrics



9.ADVANTAGES

1. Analysingclinicaldatatoimprovemedicalresearch
2. Usingpatientdatatoimprove health outcomes
3. Gainingoperationalinsightsfromhealthcareproviderdata
4. Improvedstaffingthroughhealthbusinessmanagementanalytics
5. Researchandpredictionofdisease.
6. Automationofhospitaladministrativeprocesses.
7. Earlydetectionofdisease.
8. Preventionofunnecessarydoctor'svisits.
9. Discoveryofnew drugs.

10. Moreaccurate calculationofhealthinsurancerates.

Disadvantages

Replacing Medical Personnel

Application of technology in every sphere of human life is improvingthe way things are done. These technologies are are also posing some threattoworld ofworks.Roboticsarereplacinghuman labour.

DataSafety

Data security is another challenge in applying big data in healthcare.Big data storage is usually targets of hackers. This endangers the safety ofmedicaldata.Healthcareorganisationsareverymuchconcerneda boutthesafety of patients' sensitive personal data. For this, all healthcareapplications must meet the requirement for data security and be HIPAAcompliantbeforetheycanbedeployedforhealthcareservice

Privacy

One of the major drawbacks in the application of big data inhealthcare industry is the issue of lack of privacy. Application of big datatechnologies involves monitoring of patient's data, tracking of medicalinventoryandassets,organizingcollecteddata,andvisualiz ationofdataonthe dashboard and the reports. So visualization of sensitive medical dataespecially that of the patients creates negative impression of big data as itvioletsprivacy

ManPower

`Applyingbigdatasolutionsinhealthcarerequiresspecialskills,andsuch kills are scarce. Handling of big data requires the combination ofmedical, technologicalandstatisticalknowledge.

10. CONCLUSION

Data analytics is the science of analysing raw datasets in order toderive a conclusion regarding the information they hold. It enables us todiscover patterns in the raw data and draw valuable information from them.Tosome,thedomainofhealthcaredataanalytics maylooknew,butithasalot of potential, especiallyifyouwishto engage inchallengingjobrolesand build a strong data analytics profile in the upcoming years. In this blog,we have covered some of the major topics such as what is healthcare dataanalytics, its applications, scope, and benefits, etc. We hope it helps you inyourdecision-makingasahealthcaredataanalyticsprofessional

11. FUTURESCOPE

The Future of Healthcare, Intel provides a foundation for big dataplatforms and AI to advance health analytics.

Predictive data analytics ishelping health organizations enhance patient care, improve outcomes, andreducecostsbyanticipatingwhen,where,andhowcareshouldbe provided.Thefutureofbigdatainhealthcarewillbedeterminedbytechnological breakthroughs from 2022 to 2030. Complete patient care andcost-effective prescription procedures are required for population healthmanagement. To assess clinical and claims data, they must be combined onthesameplatform.

Countries around the world have started to invest more capital inmedical infrastructure, pharmaceuticals, and healthcare smart analyticssolutions. The market is growing and will continue to expand, given thebenefitsofhealthcaredataanalytics.Ithasalsorisenasagoodcareeroption for fresh data science and data analytics graduates or professionalswhowishtobuildtheircareerinthehealthcaresector.Duetothesensitivityoftheprofession, thesalaryoffersforhealthcaredataanalystsare lucrative around the world. Apart from the remuneration, theopportunities to work with some of the biggest names in the healthcaresector is also worth mentioning. Hence, healthcare data analytics is growingto be one of the most rewarding branches of data analytics in the comingfuture.

12. APPENDIX

Source Code

Importing required Packages

In [72]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set_style("darkgrid")
plt.style.use("dark_background")
```

Importing the dataset

In [73]:

```
train = pd.read_csv('/content/input/training_data.csv')
test = pd.read_csv('/content/input/testing_data.csv')
Paramters_Description = pd.read_csv('/content/input/parameter_description.csv')
sample = pd.read_csv('/content/input/testing_target.csv')
```

Viewing dataset

In [74]:

```
train.head(5)
```

Out[74]:

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available_Extra_Rooms_in_Hospital	Department	Ward_Type	Ward_Facility_Code	Bed_Grade
0	1	8	c	3	Z	3	radiotherapy	R	F	2.0
1	2	2	c	5	Z	2	radiotherapy	S	F	2.0
2	3	10	e	1	X	2	anesthesia	S	E	2.0
3	4	26	b	2	Y	2	radiotherapy	R	D	2.0
4	5	26	b	2	Y	2	radiotherapy	S	D	2.0

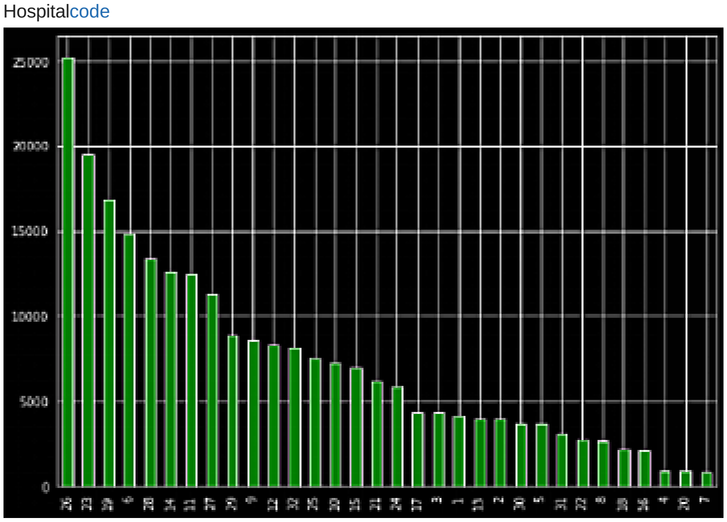
Dataset Column Description

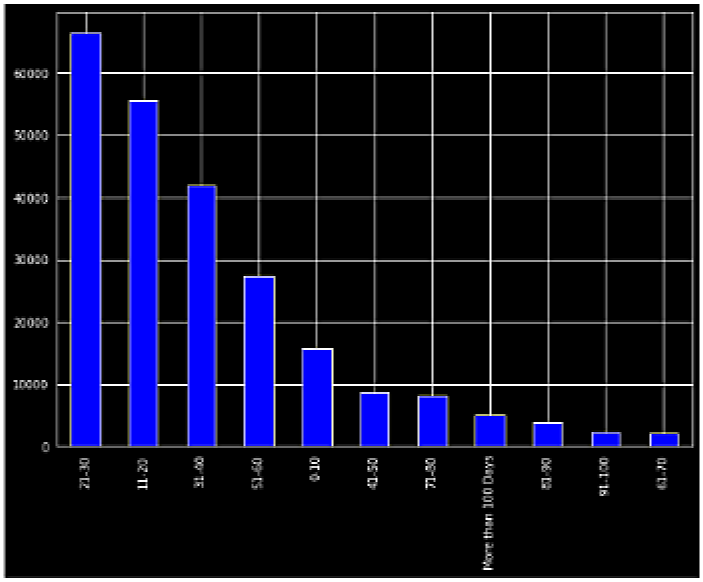
```
Paramters_Description
```

	Column	Description
0	case_id	It is identity number given by hospital admini...
1	Hospital_code	It is the code (identity number) given to the ...
2	Hospital_type_code	It is the unique code given to the type of hos...
3	City_Code_Hospital	It is the code given to the city where the hos...
4	Hospital_region_code	It is the code given to the region where the h...
5	Available_Extra_Rooms_in_Hospital	It will display the number of rooms that are s...
6	Department	The department that is overlooking the patient...
7	Ward_Type	The unique code given to the type of ward to w...
8	Ward_Facility_Code	The unique code given to the facility in the w...
9	Bed_Grade	It is the quality or condition of the bed in t...
10	patientid	It is the unique identity value given to the p...
11	City_Code_Patient	It is the unique identity code given to the ci...
12	Type_of_Admission	It is the admission type registered in the hos...
13	Severity_of_Illness	It is the severity level of the patients' illn...
14	Visitors_with_Patient	Number of the visitors with the patients to ta...
15	Age	It is the age of patients. It is given in peri...
16	Admission_Deposit	It is the deposit amount that the patient paid...
17	Stay	It is the Length Of Stay (LOS) of patients. I...

Analysisofdataset

Hospitalcode





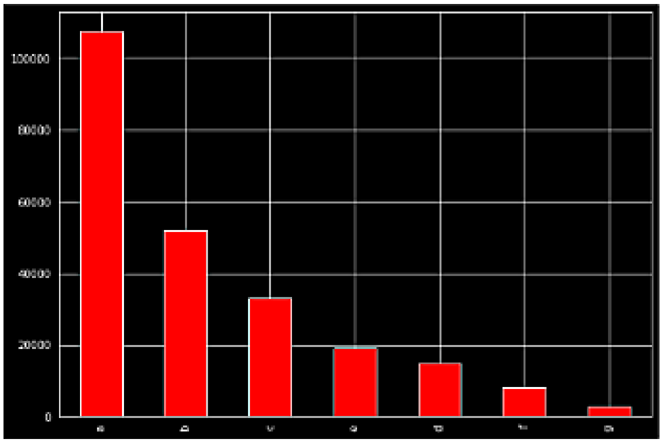
Age

```
train.Age.value_counts()
```

```
41-50      48272
31-40      48106
51-60      36969
21-30      28555
71-80      28552
61-70      26139
11-20      10141
```

```
c     32985
e     19185
d     14833
f      8166
g      2740
Name: Hospital_type_code, dtype: int64
```

```python
#Hospital_type_code distribution
plt.figure(figsize=(10,7))
train.Hospital_type_code.value_counts().plot(kind="bar", color = ['Red'])
```

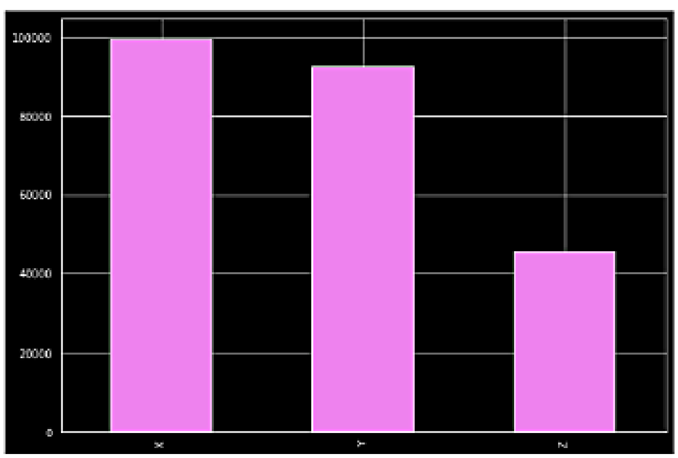


Hospital_region_code

```python
train.Hospital_region_code.value_counts()
```

```
X     99568
Y     92214
Z     45527
Name: Hospital_region_code, dtype: int64
```

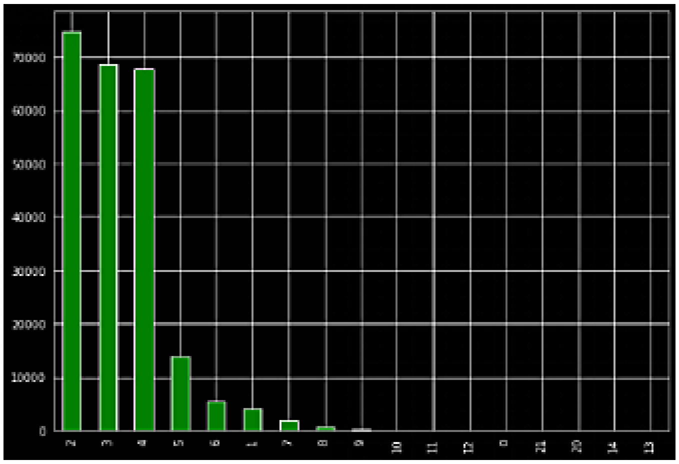
```python
#Hospital_region_code distribution
plt.figure(figsize=(10,7))
train.Hospital_region_code.value_counts().plot(kind="bar", color = ['Violet'])
```



Available_Extra_Rooms_in_Hospital

```python
train.Available_Extra_Rooms_in_Hospital.value_counts()
```

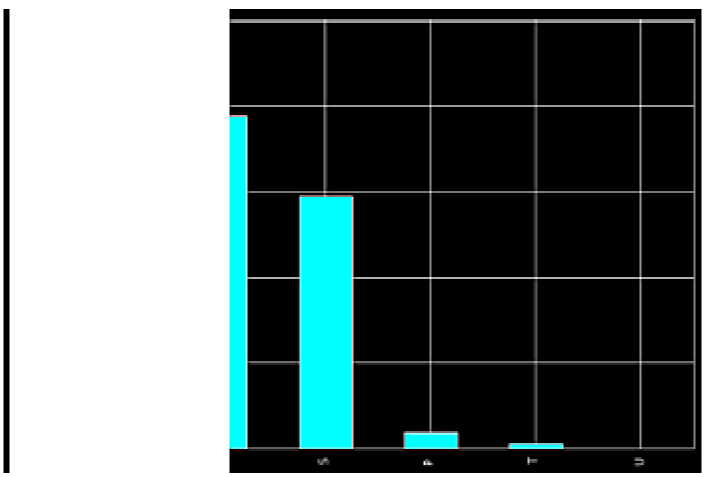
```
2     74877
3     68517
4     67756
5     13879
6      5344
1      4288
7      1876
8       622
9       144
10       46
```




```
                185862
```

```
R      9
D      77 7
P      3
T
U            9
```

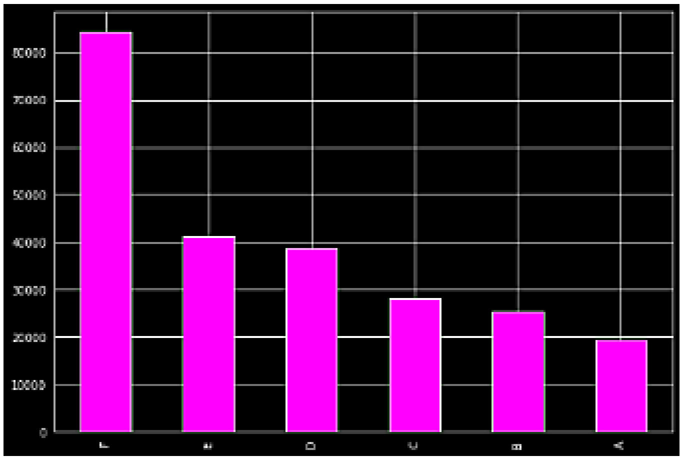
```
#Ward_Type distribution
plt.figure(figsize=(10,7))
train.Ward_Type.value_count
```



Ward_Facility_Code

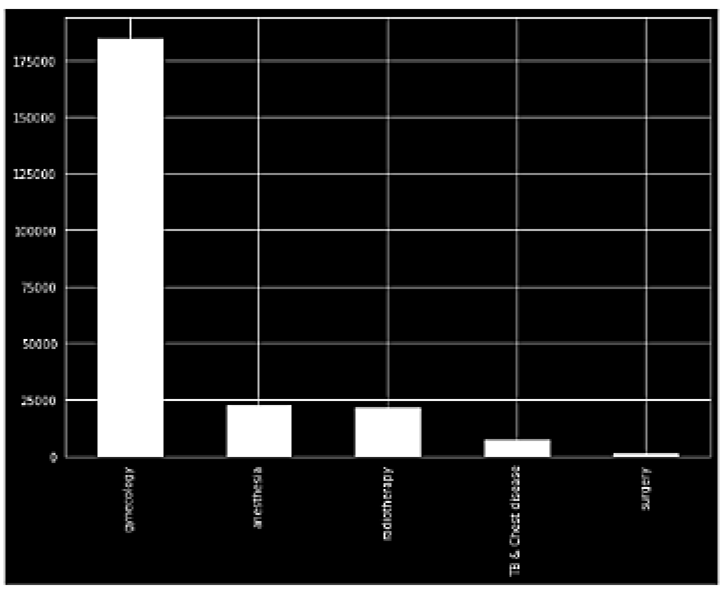
```
      19411
ame: Ward_Facility_Code, dtype: int64
```

```
#Ward_Facility_Code distribution
plt.figure(figsize=(10,7))
train.Ward_Facility_Code.value_counts().plot(kind="bar", color = ['magenta'])
```



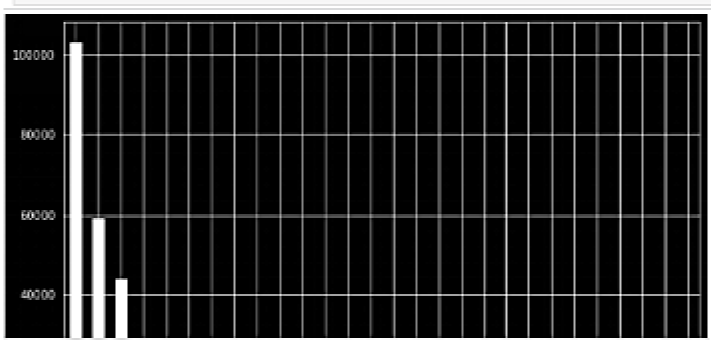

```
anesthesia            22557
radiotherapy          21725
TB & Chest disease     7017
surgery                 948
Name: Department, dtype: int64
```

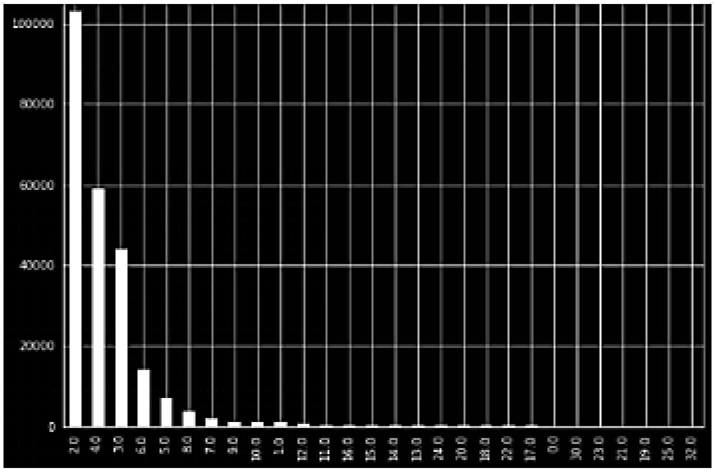
```
#Department distribution
plt.figure(figsize=(10,7))
train.Department.value_counts().plot(kind="bar", color = ['white'])
```



Ward_Type

```
train.Ward_Type.value_counts()
```



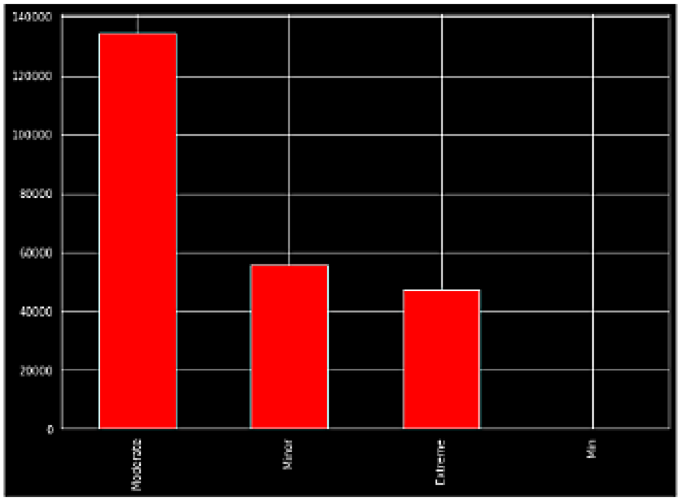



```
train.Severity_of_Illness.value_counts()
```

```
oderate
inor        55665
xtreme      47319

   : Severity_of_Illness, dtype: int64
```

```
#Severity_of_Illness distribution
plt.figure(figsize=(10,7))
train.Severity_of_Illness.value_counts().plot(kind="bar", color = ['red
```



Unique values of columns

```
for features in train.columns:
    print('*--------------------------------------------------------------------------------------------*')
    print(' Unique Values for {}'.format(features))
    print(train[features].unique())
    print('*--------------------------------------------------------------------------------------------*')
    print()
```

```
*------------------------------------------------------------------------------------------*
 Unique Values for case_id
[     1      2      3 ... 237307 237308 237309]
*------------------------------------------------------------------------------------------*
```

```
*------------------------------------------------------------------------------------------*
 Unique Values for Hospital_code
[ 8  2 10 26 23 32  1 22 16  9  6 29 12  3 21 28 27 19  5 14 13 31 24 17
 25 15 11 30 18  4  7 20]
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Hospital_type_code
['c' 'e' 'b' 'a' 'f' 'd' 'g']
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for City_Code_Hospital
[ 3  5  1  2  6  9 10  4 11  7 13]
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Hospital_region_code
['Z' 'X' 'Y']
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Available_Extra_Rooms_in_Hospital
[ 3  2  1  4  6  5  7  8  9 10 12  0 11 20 14 21 13]
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Department
['radiotherapy' 'anesthesia' 'gynecology' 'TB & Chest disease' 'surgery']
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Ward_Type
['R' 'S' 'Q' 'P' 'T' 'U']
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Ward_Facility_Code
['F' 'E' 'D' 'B' 'A' 'C']
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for Bed_Grade
[ 2.  3.  4.  1. nan]
*------------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------------*
 Unique Values for patientid
[31397 63418  8088 ... 37502 73756 21763]
*                                                                                          *
```

```
*-----------------------------------------------------------------------------------------*
 Unique Values for City_Code_Patient
[ 7.  8.  2.  5.  6.  3.  4.  1.  9. 14. nan 25. 15. 12. 10. 28. 24. 23.
 20. 11. 13. 21. 18. 16. 26. 27. 22. 19. 31. 34. 32. 30. 29. 37. 33. 35.
 36.]
*-----------------------------------------------------------------------------------------*

*
*---------------------------------------------------------------------------------------*  *

*---------------------------------------------------------------------------------------*
 Unique Values for Admission_Deposit
[4911. 5954. 4745. ... 2710. 2236.   nan]
*---------------------------------------------------------------------------------------*

*---------------------------------------------------------------------------------------*
 Unique Values for Stay
['0-10' '41-50' '31-40' '11-20' '51-60' '21-30' '71-80'
 'More than 100 Days' '81-90' '61-70' '91-100' nan]
*                                                                                       *
```

DataPreprocessing&FeatureEngineering

The following features may have relevance with the Length of Stay of a patient

Department:ItRelatestothetypeofd-sease.Hence”t\¥fIlhavefrrpaclartheengthofstayofthepatients

Type of Admission: It Re ates to patients’reBsorofadm-ssionto the Wasp-talard definitely ft 'wi I

have impact ouength of stay opftbe patientsSmm1tyoflllr«zss:ItRelatetothecurab-l”tyofdisease

AgeRelatetothecurabfl”tyofdiseaseThefola*'fiiing*featuresmayPa\erelevance'with

theLengthofStayofapat”entDepartment:ItRelatestothetypeofd-

sease.Hence”t\¥fIlhavefrrpaclartheengthofstayofthepatients

Type of Admission: It Re ates to patients'reasorofadm”ssionto the bosp”talard definitelyft'xi I

haveimpact ouengthof stay opftbe patientsSeverityafhIrless:ItRelatetothecurab”l”tyofdisease

AgeRelatetothecurabfl-tyofdisease
U¥ård_Type:Reatestothedurabilityofdisease

befollowing features doesn1 have relevancewfth It+e

Length Of 5tayQo5}of Patients’Hospitalreqirxieerie:It-

s*cade*giventothehospitalreg-onvA-chis-

rreleventtotheLerpthofStay.

Bed Grade:Itisthegradegiventothequalityoftheked ir\Yardftisalsolrrelevanttothelengthofstay.
patierrbd:It -s the identity numberor code gi\xrfor the -dentificatian of the

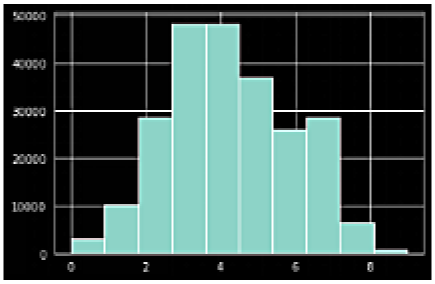
patient\Yhich Is irrel evartto theeqgthof

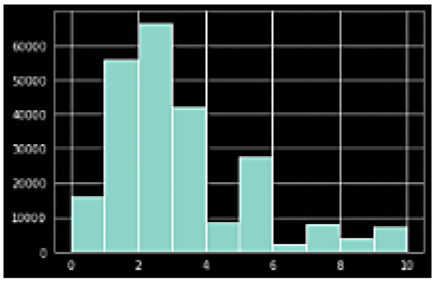
stay.City_Ccde_Fatient:Itisthecitycodeand”rrelevanttothelenpthofstayofp

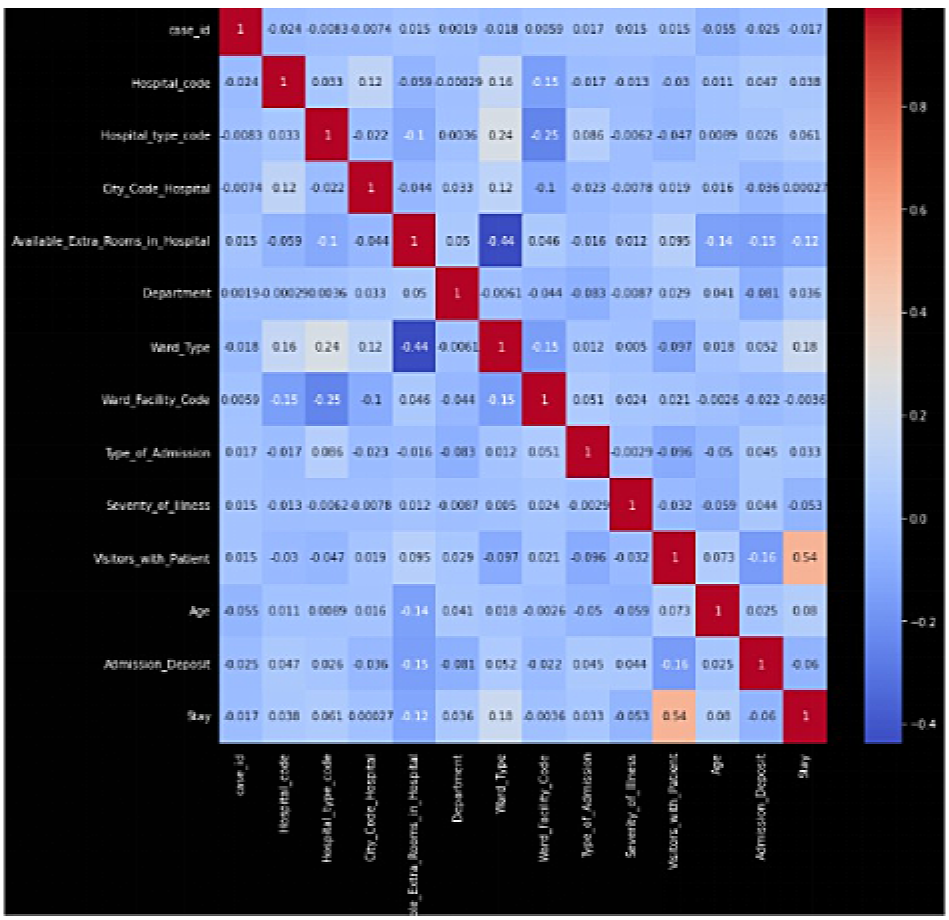
atierts.

as'Hosp1taZ_re:Icn_czde',"ded_firade','patientid','c:Ity_czde_patient'ar'elnreJevanttctheheaLthor Ja'ngtho+stayo+patbentsso2etsdropthezepararetersl-rantrainingan3ta'stEngdatasetto:tmprovetheper+'smaca'o+mdeL(highaczurract')

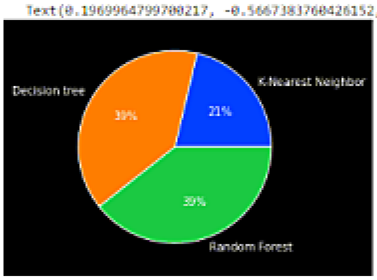
```
tratn=tratn.drap, ['HospitaJ_ra\g1on_zade”,’bed_urade'’,’patigntid,'
'E:1ty_C:o3e_I‘atbent”,]
axes=2)
```


50000
40000
30000
20000
10000
0
0
2
4
6
8

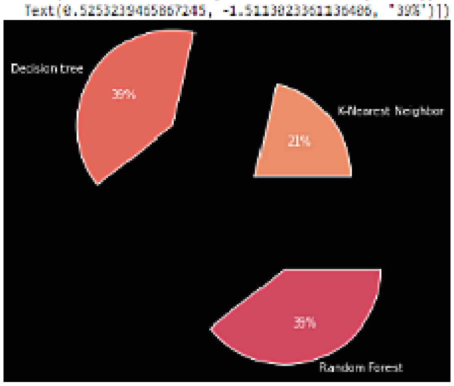

60000
50000
40000
30000
20000
10000
0
0
2
4
6
8
10


case_id
Hospital_code
Hospital_type_code
City_Code_Hospital
Available_Extra_Rooms_in_Hospital
Department
Ward_Type
Ward_Facility_Code
Type_of_Admission
Severity_of_Illness
Visitors_with_Patient
Age
Admission_Deposit
Stay
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

```
[Text(0.8628423642631272, 0.682277842548633, 'K-Nearest Neighbor'),
 Text(-0.9277499083745313, 0.590999244932723, 'Decision tree'),
 Text(0.3611602132783717, -1.0390203560781281, 'Random Forest')],
[Text(0.4706412895980693, 0.3721515504810725, '21%'),
 Text(-0.5060454045679261, 0.322363224508758, '39%'),
 Text(0.1969964799700217, -0.5667383760426152,
```



```
 ],
[Text(0.8706863857564283, 0.6884803683899842, 'K-Nearest Neighbor'),
 Text(-1.7711589159877414, 1.1282712857806532, 'Decision tree'),
 Text(0.689487679805876, -1.0835843161491535, 'Random Forest')],
[Text(0.4784853110913704, 0.37835407632242374, '21%'),
 Text(-1.3494544121811365, 0.8596352653566088, '39%'),
 Text(0.5253239465867245, -1.5113023361136406, '39%')])
```




```
output = pd.DataFrame({
        "case_id": test["case_id"],
        "Stay": Y_pred
})
```

```
output['Stay'] = output['Stay'].replace(stay_labels.values(), stay_labels.keys())
```

```
output.to_csv('LOS_Prediction.csv', index = False)
```

```
output
```

	case_id	Stay
0	318439	0-10
2	318441	21-30
3	318442	11-20
4	318443	31-40
...	...	...
137052	455491	0-10
137053	455492	0-10
137054	455493	21-30
137055	455494	21-30
137056	455495	51-60

137057 rows × 2 columns

```
data=np.array([[29,0,4,2,2,3,5,1,2,4,7,4018]])
p=random_forest.predict(data)
p
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not have valid feature names, but RandomForestClassifier was fitted wi
th feature names
  "X does not have valid feature names, but"
```

```
array([5.])
```

```
def prediction(p):
  if(p[0]==0):
    print("The predicted LOS of patient is : 0-10")
  elif(p[0]==1):
    print("The predicted LOS of patient is : 11-20")
  elif(p[0]==2):
    print("The predicted LOS of patient is : 21-30")
  elif(p[0]==3):
    print("The predicted LOS of patient is : 31-40")
  elif(p[0]==4):
    print("The predicted LOS of patient is : 41-50")
  elif(p[0]==5):
    print("The predicted LOS of patient is : 51-60")
  elif(p[0]==6):
    print("The predicted LOS of patient is : 61-70")
  elif(p[0]==7):
    print("The predicted LOS of patient is : 71-80")
  elif(p[0]==8):
```

```
  elif(p[0]==8):
    print("The predicted LOS of patient is : 81-90")
  elif(p[0]==9):
    print("The predicted LOS of patient is : 91-100")
  elif(p[0]==10):
    print("The predicted LOS of patient is : More than 100 Days")
```

```
data=np.array([[29,0,4,2,2,3,5,1,2,4,7,4018]])
p=random_forest.predict(data)
print(p)
```

```
prediction(p)
```

```
The predicted LOS of patient is : 51-60
```

GitHub&Project DemoLinks

GitHub link: IBM-EPBL invited you to IBM-EPBL/IBM-Project-41108-1660639510

Projectdemolink:

https://colab.research.google.com/drive/1DpGcjD6aJZENhHU-iDWnwIjFAbk0I3ux?usp=sharing