PREDICTIVE AND PRESCRIPTIVE STATISTICAL MODELS FOR INVENTORY MANAGEMENT IN HEALTH CARE

“ANEKANT EDUCATION SOCIETY”

TULJARAM CHATURCHAND COLLEGE OF ARTS, SCIENCE & COMMERCE, BARAMATI-413102

(Autonomous)

A PROJECT REPORT ON

“Predictive and Prescriptive Statistical Models for Inventory Management in Health Care”

SUBMITTED TO

DEPARTMENT OF STATISTICS

SAVITRIBAI PHULE PUNE UNIVERSITY

By

Mr. Kshirsagar Tushar Deepak (17029)

Mr. Gujar Hrishikesh Ankush (17027)

Mr. Bhoite Nikhil Suryakant (17028)

UNDER THE GUIDENCE OF

Miss. N.A. Jagtap

2022-2023

“ANEKANT EDUCATION SOCIETY”

TULJARAM CHATURCHAND COLLEGE OF ARTS, SCIENCE &COMMERCE, BARAMATI-413102

(Autonomous)

CERTIFICATE

This is to certify that **Mr. Kshirsagar Tushar Deepak, Mr. Gujar Hrishikesh Ankush, Mr. Bhoite Nikhil Suryakant,** regular students of class M.Sc.-II (Statistics) has completed project on **“Predictive and Prescriptive Statistical Models for Inventory Management in Health Care”** in the partial fulfillment of the program in M.Sc. II as per the syllabus laid by Savitribai Phule Pune University, during the academic year 2022-2023.

Miss. N.A. Jagtap

Dr. Jagtap A.S.

PROJECT GUIDE	**EXTERNAL EXAMINER**	**HEAD OF DEPARTMENT**

PLACE: BARAMATI

DATE:

ACKNOWLEDGEMENT

We have satisfaction upon completion of the project work entitled **“Predictive and Prescriptive Statistical Models for Inventory Management in Health Care”** at Department of Statistics, Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati during the academic year of 2022-23.

In the accomplishment of this project successfully, many people have best owned upon us their guidance & assistance. All the work that we have done in this project is only due to such supervision & assistance, this time we are utilizing to thank all the people who have been concerned with this project.

We thank to Dr. A. S. Jagtap Sir, Head of the Department of Statistics for providing us an opportunity to do this project and giving us all the support and their valuable guidance for successful completion of this project work duly.

The timely and successful completion of this project could hardly be possible without the guidance and assistance of our guide **Miss. N. A. Jagtap** Madam, whose valuable guidance has been the ones that helped us patch this project work and make it full proof success.

We would like to thank the Department of Statistics for giving priceless guidance their providing all necessary facilities, good co-operation and their valuable discussion during completion of our project.

We are especially thankful to the Shripal Hospital, Baramati who gave us the permission of data collection in hospital. Also, we would like to express whole hearted thanks to my friends for their help and support for completion of project work. Last but not the list we would like to thank non-teaching staff of Department of Statistics for providing laboratory at any time during the project.

INDEX

ABSTRACT

Recently, the healthcare industry has started using inventory management applications to achieve efficiency and effectiveness in its supply chains. There is a rapid growth in the demand of drugs and diagnostic systems within the healthcare industry. The biggest challenge for health care supply chains is to manage inventory efficiently and keep up the satisfactory service level at the same time. Our objective of this study is to build a system to manage the inventories in hospital wards using Predictive and Prescriptive Analysis with the help of various machine learning techniques and to conclude which techniques are effective and accurate. A secondary data is collected from Shripal Hospital, Baramati. Modern Machine learning algorithms such as Decision Tree, Random Forest, Gradient boosting, Ada boosting, KNN classifier etc. are used to classify this dataset. The accuracy of the model is calculated and then one of them with good accuracy is taken as the model for predicting recovery time. The model we propose makes use of machine learning by taking into consideration both historical usage patterns of the ward's current situation to minimize inventory levels as well as the necessity for emergency.

KEYWORDS: Epidemic Disease, Abdominal Disease, Antibiotic, Beta Blockers, Random Forest, Decision tree.

INTRODUCTION

The Inventory Management System in hospital wards is a process in which, it has a real-time status of all the available beds and the bed occupied, to plan for the efficient use of beds. It helps the hospital staff and management by reducing the time of counting and recording the availability of beds. Healthcare inventory management is incredibly important in Hospital wards. Inventory management systems can help hospitals and other medical organizations streamline their processes to save time while providing quality care to patients. Inventory management of drugs and bed in hospitals became a pivotal practice after the National Health Service's statement rule on patient emergency trolley waiting hours. After an accident, inconvenient waiting hours in hospital waiting halls due to improper operational capacity planning and control over managing, allotting, and maintaining beds may turn frustrating when emergency care is required. The lack of real-time data analysis provides a disadvantage resulting in patient overcrowding and malefactions in inpatient bed management services during emergencies. Monitoring and managing inpatient-outpatient capacity & patient inflows can be challenging tasks during emergencies. Unplanned inventory, the inability to beds, and other scarce resources may lead to poor operations, and that affects the hospital's goodwill.

Predictive Analysis refers to the use of different statistics and machine learning techniques to make predictions about future events taking present and future data patterns into consideration. Prescriptive Analysis is a process that what actions we have to take to achieve the predicted results, and suggests actions and possible results so we can easily achieve the predicted results. The availability of huge amounts of data is creating new opportunities for decision making under uncertainty, and the renewed attention towards machine learning provides new tools that can be used to generate predictions of quantities that may be of interest for inventory management. In this project we deal with the situation of beds and drugs inventories inside a hospital ward.

The main purpose of this study is to explore the different machine learning techniques that will help with the managing the inventories in hospitals. The current study focuses on patients records with their medical history collected from different hospitals. Record contains which type of disease patient diagnosed with, and which kind of treatment was given to them. The record also contains the date of admission and discharge of patient and from that we concluded the recovery time of patient.

Our main objective in this study is to propose an efficient model for predicting how much time a patient might requires to recover which help in managing the inventories in hospital wards. And to correctly predict the output model needs to take some important actions using prescriptive analysis.

MOTIVATION

A hospital inventory management system is an innovative method that helps hospitals to manage their beds, inventories, patient flow, and resource allocation. A hospital inventory management system can track the availability of beds, patient admission and discharge, and bed cleaning schedules. The main purpose of this study is to help hospitals avoiding inconvenient waiting hours in hospital waiting halls due to improper operational capacity planning, and using different machine learning technique to check how this inventory management system actually works with the help of predictive and prescriptive Analysis. To help in early detection, technology can be used very reliably and efficiently. Managing the inflow and outflow of the patients is very big task for hospitals. The inventory management system comes handy in these situations. The real motivation of this study for hospital inventory management system is to make easy process of all the management process like patients' registration, patients discharge etc.

OBJECTIVES

- To propose an innovative machine learning model to predict Recovery Time of patient using Predictive and Prescriptive Analysis.
- To determine the factors like Age, Gender, Medical History, Diagnosis etc. which affects the Recovery Time of patient.
- To find the best fit model on the basis of Accuracy, Precision, Recall and F1-score among various other models.

METHODOLOGY

Predictive Analysis applies mathematical models to the current data to inform future behavior. It is the 'What could happen". Prescriptive Analysis utilizes similar modeling structures to predict outcomes and then utilizes combination of machine learning, and algorithms to simulate various approaches to these numerous outcomes. It then suggests the best possible actions to optimize inventories in hospital. It is the "What should happen".

We are interested in predicting recovery time of patients, by taking into consideration their Age, Gender, Medical History, and Diagnosis etc. So, for data collection purpose we have collected data from a reputed hospital. We collected the data of 2065 patients from Shripal Hospital which is located in Baramati. In this study we have used predictive and prescriptive Analysis for effective management of inventories in hospitals. We have used various machine learning algorithms and selected one of best among them with greater accuracy and best precision and applied Input Method on that selected model.

The first stage in data mining is Data preprocessing which includes cleaning of missing and noisy values and irrelevant observation from the collected data. Then we performed Exploratory Data Analysis where we checked the normality of the dataset by using the probability plot. For the visualization purpose we have drawn pie chart, bar graphs, tree map, and word cloud for various features such as Age Group, Gender, Sugar Level, Type of disease, Type of Medicine, Locality and Recovery Time etc. For checking the dependency of various features with the recovery time of patient, we have used Chi-Square test. After checking the dependency, we then further moved for finding the accuracy of the various Machine Learning models.

MATERIALS AND METHODS

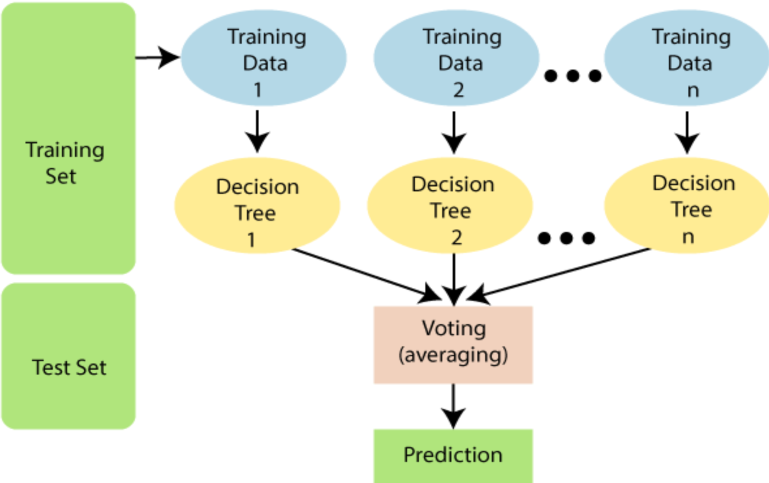
Software, Libraries and Packages Used:

Jupyter Notebook	1) Predictive and Prescriptive analysis and Exploratory Data Analysis (EDA) is performed using Python language. 2) Basically, Jupyter notebook is an application which provides us with an environment where we can write and execute python code. 3) Importing data in python and performing data pre-processing methods on it, so that data becomes ready for analysis and we get more accurate analysis. 4) After performing pre-processing on the data, we can also implement different machine learning algorithms in order to classify the sentiments. 5) And finally, Jupyter notebook will also make it easier to visualise and analyse the data. 6) Code scripts, research-related text, visualisation plots and graphs, machine learning models, and other materials may all be kept in one Notebook or document and shared effortlessly across several platforms. 7) These are all important uses of the Jupyter notebook and these are the reasons behind choosing Jupyter Notebook.
Pandas	The data which was collected from the Hospitals was stored in CSV files or formats. So, in order to work with CSV files and import them in the python we will require a Special library of python which is PANDAS LIBRARY. So, basically when we want to work with the datasets, we will be using Pandas library. It is not just used to import the CSV files but there are many other uses such as: I. This library can clean up unreadable and irrelevant data sets. Relevant data is most important thing for a Data Analyst. II. This library can be used to answer some basic questions about data such as, • If there are two columns, so is there any correlation between these two columns? • What is an average value of a particular column? • What is the minimum or maximum value that is occurring in a dataset?
Sklearn	Sklearn basically stands for Scikit Learn Library. As we are aware of the fact that there are many Machine Learning Libraries used in Python. Scikit learn is one of them, in fact this library is one the best known. This library is responsible to support both of the machine learning approaches,

	Supervised and Unsupervised. This library also provides many different algorithms for the classification, dimensionality reduction, clustering and regression purposes. The combination of two libraries i.e., NumPy and SciPy is Sklearn library. Additionally, it functions nicely with other libraries like Pandas and Seaborn. We can use this library for pre-processing such as feature encoding, feature extraction and we can also use this library to split the data into train and test, then we can use this library to implement different machine learning models and finally in order to check the accuracy of these models we can use this library.
WordCloud	When we have textual data, and you have no idea how to visualise it then "Word Cloud" library is the most important library which will be used. The magnitude of each word in a word cloud, shows its frequency or relevance. That is, the word which is mostly repeated will be displayed most bold and large in size. In the image below is a Word Cloud. By looking at that image we can understand that words like **"Antibiotic", "inhibitor"** are bolder and larger in size, this is because frequency of these words in the whole text is highest. Python modules such as matplotlib, pandas, and Word Cloud are required for creating word clouds.
Pickle	The pickle module implements binary protocols for serializing and de-serializing a Python object structure. "**Pickling**" is the process whereby a Python object hierarchy is converted into a byte stream and "**unpickling**" is the inverse operation, whereby a byte stream (from a binary file or bytes-like object) is converted back into an object hierarchy. Pickling is alternatively known as "serialization", "**marshalling**," 1 or "flattening"; however, to avoid confusion, the terms used here are "**pickling**" and "**unpickling**".
Excel	Most of the visualization is performed using excel. Excel is the excellent tool for effective visualization. Excel has charts of many kinds, including column charts, bar charts, pie charts, line charts, area charts, scatter charts, surface charts, and many more. In visualization, pie chart, bar graphs, and tree map are drawn for various features.
Word	Report writing is done using MS Word.
PowerPoint	PowerPoint Presentation is made using MS PowerPoint.

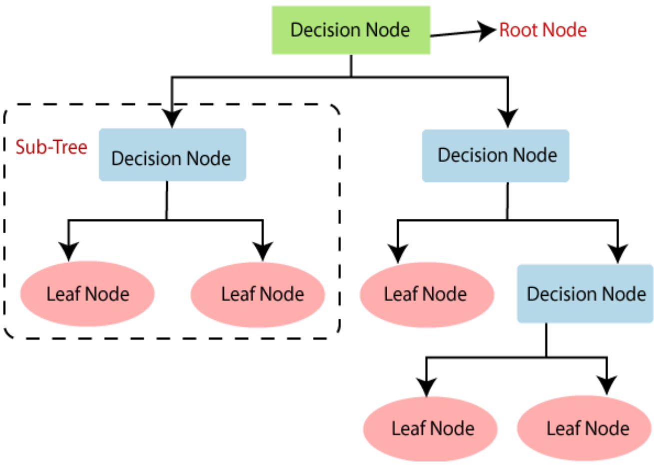
STATISTICAL TERMS

RANDOM FOREST - A random forest is a Machine Learning technique that is used to solve regression and classification problems. It utilizes ensemble learning technique, that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms. The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.



DECISION TREE- Decision tree is the most powerful and popular tool for classification and prediction. Decision tree is a type of Supervised Machine Learning where data is continuously split according to certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or final outcomes and the decision nodes are where the data is split.

In its simplest form, a decision tree is a type of flowchart that shows a clear pathway to a decision. In terms of data Analysis, it is a type of algorithm that includes conditional 'control' statements to classify data. A decision tree starts at a single point (or 'node') which then branches (or 'splits') in two or more directions. Each branch offers different possible outcomes, incorporating a variety of decisions and chance events until a final outcome is achieved. When shown visually, their appearance is tree-like…hence the name!



- **Root nodes:**

  In the diagram above, the blue decision node is what we call a 'root node.' This is always the first node in the path. It is the node from which all other decision, chance, and end nodes eventually branch.

- **Leaf nodes:**

  In the diagram above, the end nodes are what we call 'leaf nodes.' These show the end of a decision path (or outcome). You can always identify a leaf node because it doesn't split, or branch any further. Just like a real leaf!

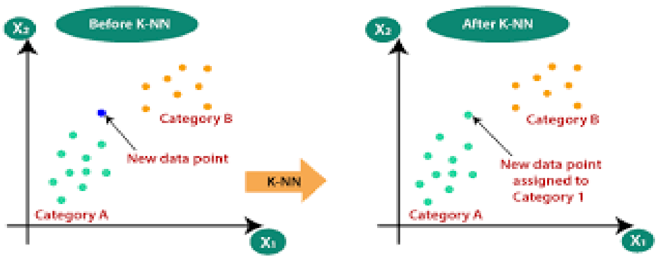
- **Internal nodes**:

  Between the root node and the leaf nodes, we can have any number of internal nodes. These can include decisions and chance nodes. It's easy to identify an internal node—each one has branches of its own while also connecting to a previous node.

- **Splitting:**

  Branching or 'splitting' is what we call it when any node divides into two or more sub-nodes. These sub-nodes can be another internal node, or they can lead to an outcome (a leaf/ end node.)

KNN- The k-nearest neighbours algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. It classifies the data point on how its neighbor is classified. KNN classifies the new data points based on the similarity measure of the earlier stored data points.



For example, if we have a dataset of Apple and Bananas. KNN will store similar measures like shape and color. When a new object comes it will check its similarity with the color (red or yellow) and shape. K in KNN represents the number of the nearest neighbors we used to classify new data points.

GRADIENT BOOSTING – Gradient boosting is popular boosting algorithm. The main idea behind this is to build models sequentially and these subsequent models try to reduce the errors of the previous model. Gradient boosting is a method standing out for its prediction speed and accuracy, particularly with large and complex datasets. When the target column is continuous, we use Gradient Boosting Regressor whereas when it is a classification problem, we use Gradient Boosting Classifier.

ADA BOOSTING- AdaBoost, also called Adaptive Boosting, is a technique in Machine Learning used as an Ensemble Method. It is called Adaptive Boosting as the weights are reassigned to each instance, with higher weights assigned to incorrectly classified instances. Boosting is used to reduce bias as well as variance for supervised learning. It works on the principal of learners growing sequentially. In simple words, weak learners are converted into strong ones. The AdaBoost algorithm works on the same principle as boosting with a slight difference.

REPRESENTATION OF DATASET

ID	Age	AgeGroup	Gender	BloodPressure	Hemoglobin	BloodGlucose	SugarLevel	Arthritis	BreathingProblem	Diagnosis	TypeofDisease	Medicine
1	62	Adults	Female	70	12.60	122	Prediabetic	No	No	Dengue	Epidemic disease	Inj C-Tri 1 gm ,NS 100ml
2	43	Adults	Male	60	17.80	174	High	No	No	Chikungunya	Epidemic disease	Inj C-Tri 1 gm, NS 100ml ,Tab HCQS 200
3	62	Adults	Female	70	12.60	122	Prediabetic	No	No	Dengue	Epidemic disease	Inj C-Tri 1 gm ,NS 100ml
4	43	Adults	Male	60	17.80	165	High	No	No	Chikungunya	Epidemic disease	Inj C-Tri 1 gm, NS 100ml ,Tab HCQS 200
5	56	Adults	Female	60	13.70	97	Normal	No	No	Viral Fever	Epidemic disease	Inj C-Tri 1 gm, Tab-Montegress LC
...	...	...	...	...	...	...	...	...	...	...	...	...
2061	2	Children	Male	80	12.65	121	Prediabetic	No	No	Bacterial Infection	Communicable disease	Inj C-Tri 500 mg
2062	57	Adults	Female	60	15.40	132	High	No	Yes	dm type 2 COPD pulmonary emobolisum	Respiratory Disease	beta-2 agonist, thephylline
2063	2	Children	Male	80	12.65	121	Prediabetic	No	No	Bacterial Infection	Communicable disease	Inj C-Tri 500 mg
2064	4	Children	Male	80	12.00	99	Normal	No	No	Bacterial Infection in Stomach	Abdominal disease	Inj C-Tri 700 gm, Inj Ondem 0.7 ml
2065	4	Children	Male	80	12.00	99	Normal	No	No	Bacterial Infection in Stomach	Abdominal disease	Inj C-Tri 700 gm, Inj Ondem 0.7 ml

BASIC STATISTICAL DESCRIPTION AND UNDERSTANDING OF DATA

The Statistics of the dataset gives us the basic idea about the spread of the data. Therefore, we need to go for description data set like count, mean, standard deviation, min, max, quartiles, etc. This image shows the data description.

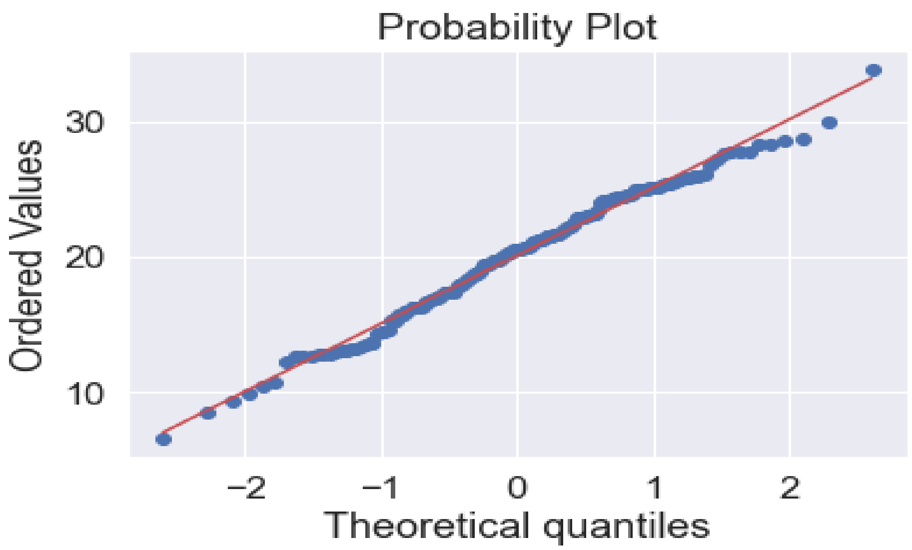
	count	mean	std	min	25%	50%	75%	max
Age	2065.0	20.537530	23.973190	1.0	2.0	7.0	44.00	90.0
BloodPressure	2065.0	63.235351	18.625330	50.0	50.0	50.0	80.00	180.0
Hemoglobin	2065.0	11.777094	1.822262	4.8	11.3	11.3	12.65	17.8
BloodGlucose	2065.0	121.106053	19.180292	52.0	121.0	121.0	121.00	234.0

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2065 entries, 0 to 2064
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ID                2065 non-null   int64
 1   Age               2065 non-null   int64
 2   AgeGroup          2065 non-null   object
 3   Gender            2065 non-null   object
 4   BloodPressure     2065 non-null   int64
 5   Hemoglobin        2065 non-null   float64
 6   BloodGlucose      2065 non-null   int64
 7   SugarLevel        2065 non-null   object
 8   Arthritis         2065 non-null   object
 9   BreathingProblem  2065 non-null   object
 10  Diagnosis         2065 non-null   object
 11  TypeofDisease     2065 non-null   object
 12  Medicine          2065 non-null   object
 13  TypeofMedicine    2065 non-null   object
 14  Locality          2065 non-null   object
 15  RecoveryTime      2065 non-null   object
dtypes: float64(1), int64(4), object(11)
memory usage: 258.2+ KB
```

EXPLORATORY DATA ANALYSIS

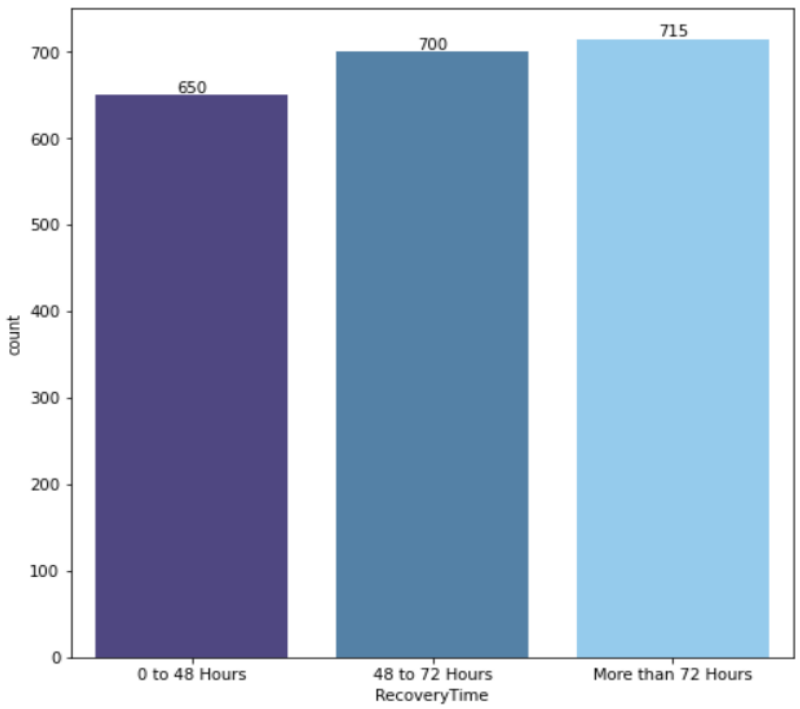
1) DESCRIPTIVE STATISTICAL MODEL:

Probability Plot:



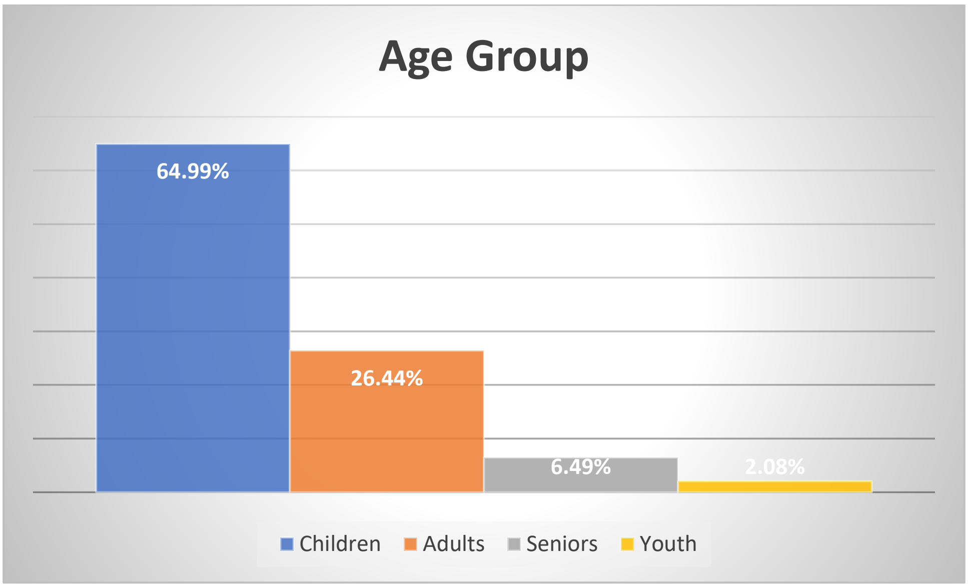
Conclusion: The Probability plot shows that all the points lies near or on the line. Hence, the data is normally distributed. The points located along the line represents the normal and points which are distant from this line represents suspected values or outliers.

Count Plot of Target Variable:



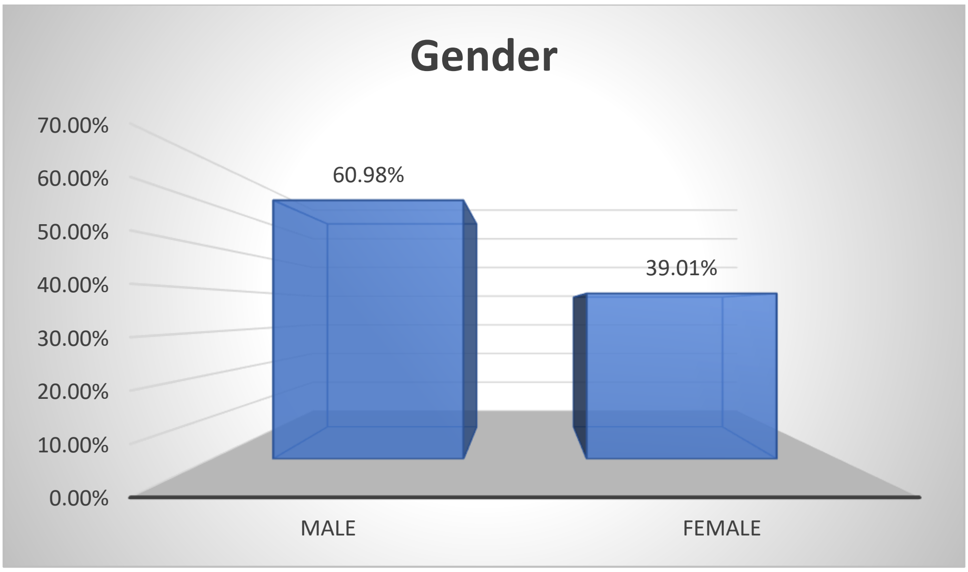
Conclusion: The above Count plot shows that target labels are nearly balanced.

Bar Plot for Age Group:



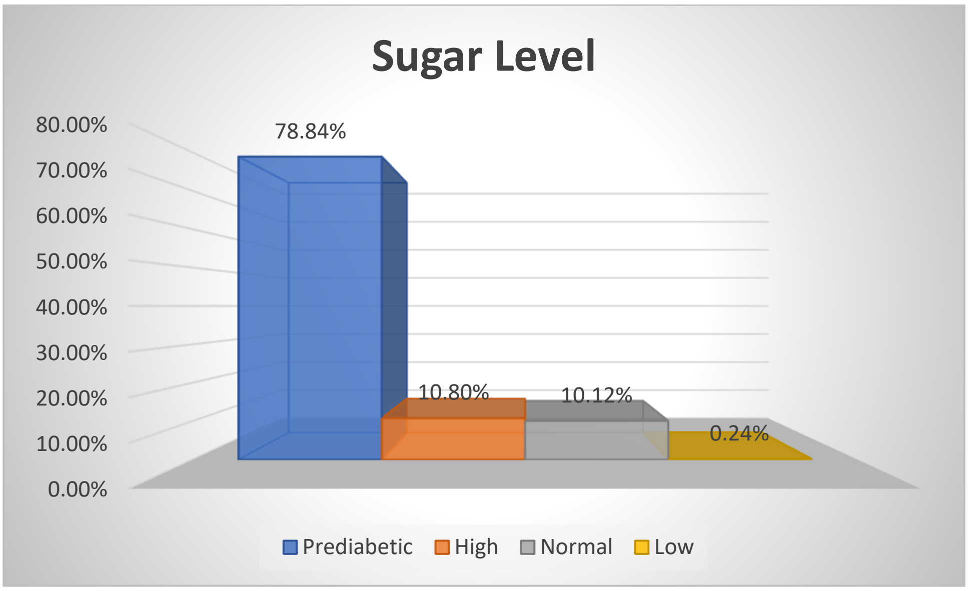
Conclusion: From the above Bar plot, we can see that, proportion of patients belongs to age group "Children" is more as compared to other age groups.

Bar Plot for Gender:



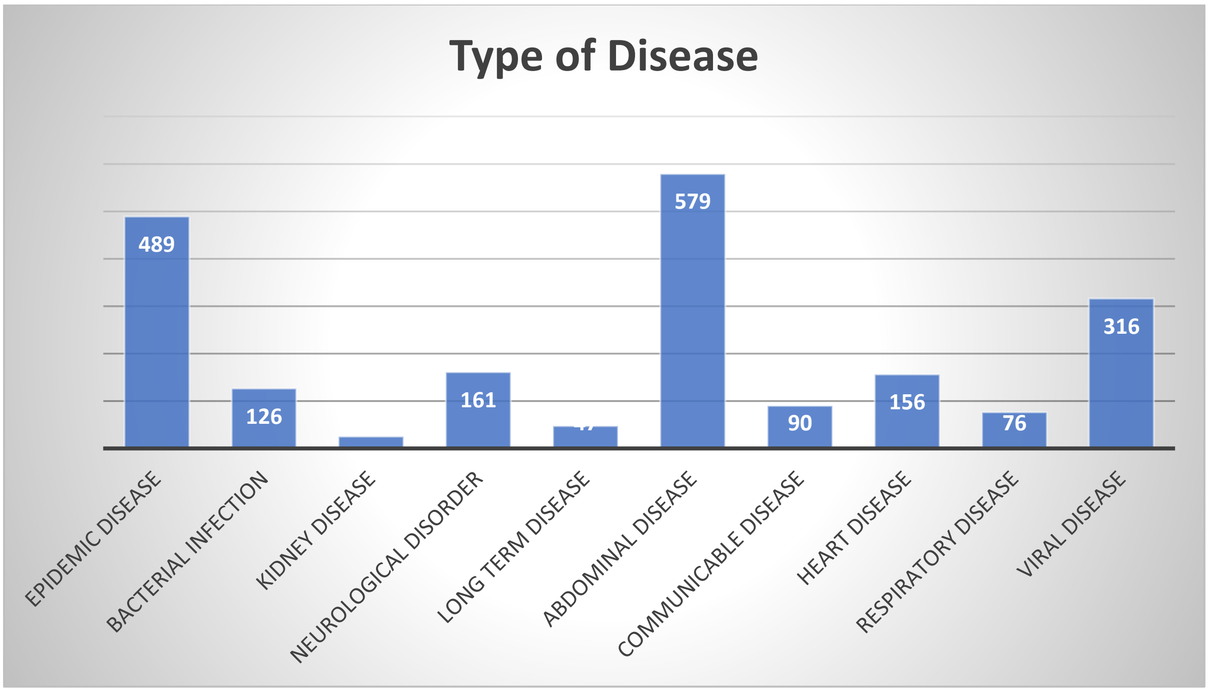
Conclusion: From the above Bar plot, we can see that, proportion of "Male" patients is more as compared to "Female".

Bar Plot for Sugar Level:



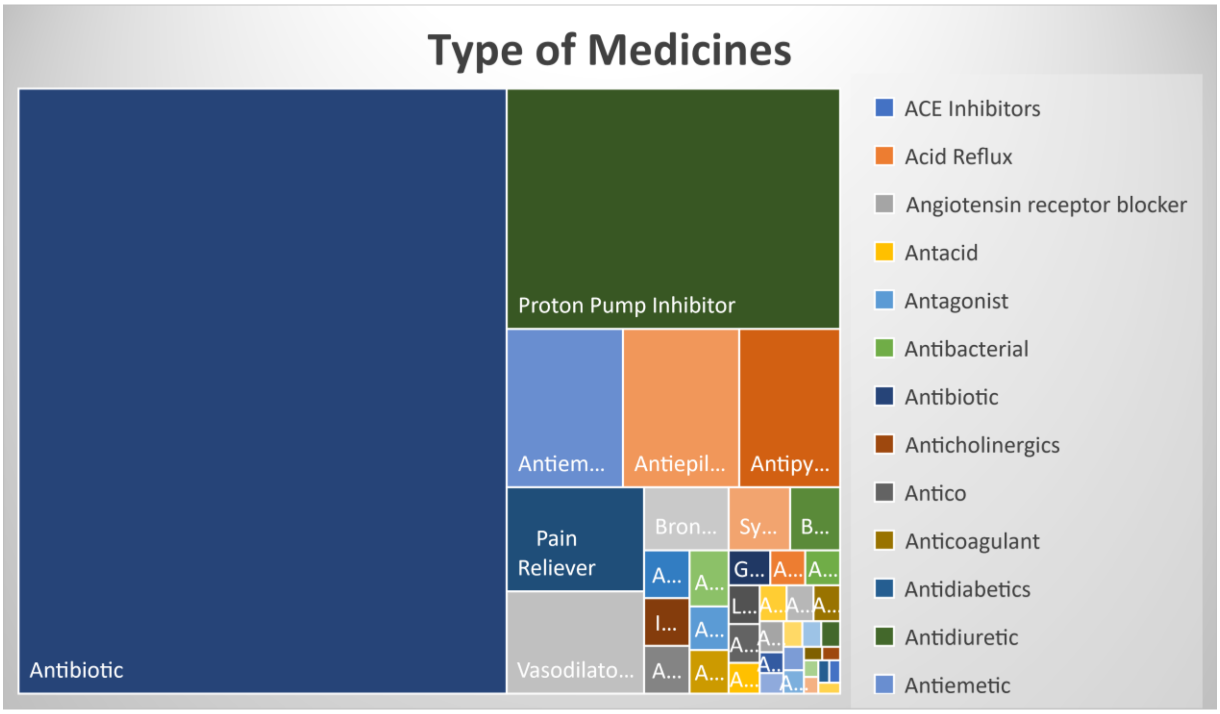
Conclusion: From the above Bar plot, we can see that, proportion of "Prediabetic" patients is more as compared to others.

Bar Plot for Type of Disease:



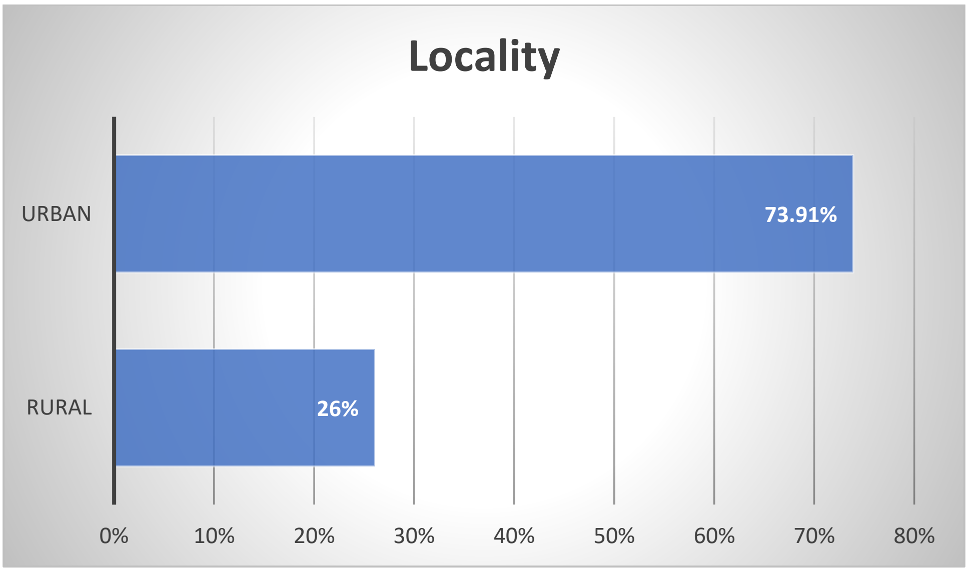
Conclusion: From the above Bar plot, we can see that, the disease type "Abdominal", "Epidemic" and "Viral" has high number of cases as compared to other types of disease.

Tree Map for Type of Medicine:



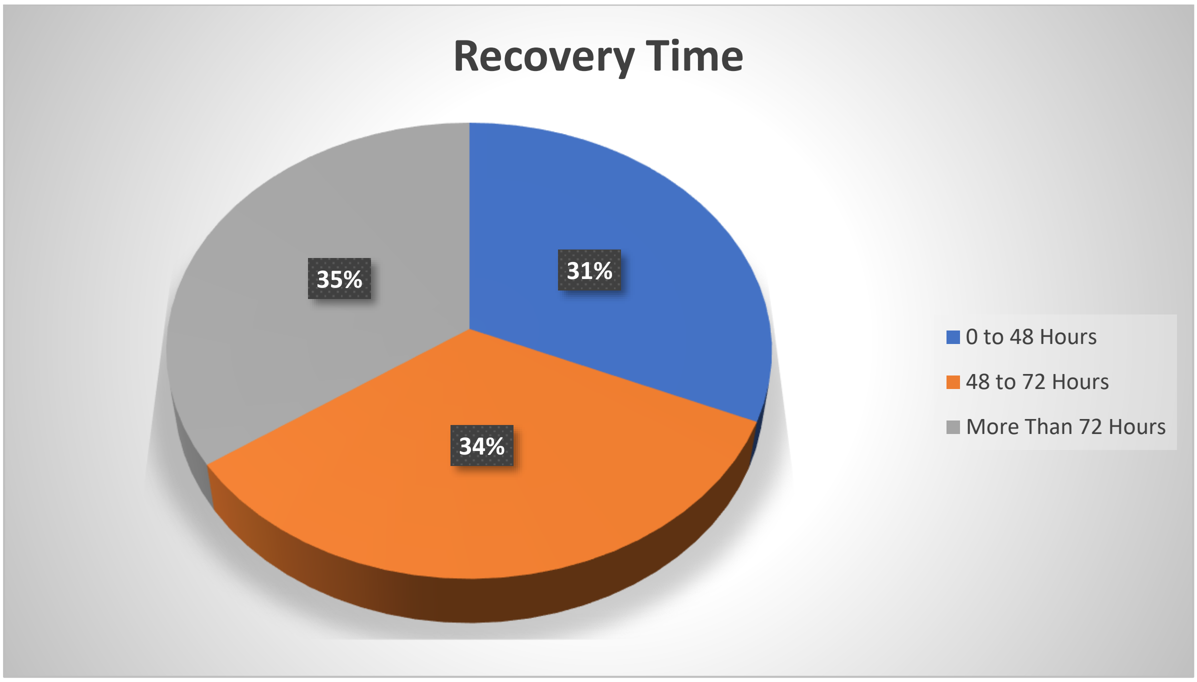
Conclusion: From the above Tree Map, we can see that, mostly patients are treated with "Antibiotic" and "Proton Pump Inhibitor" type of medicines.

Bar Plot for Locality:



Conclusion: From the above Bar plot, we can see that, proportion of patients from "Rural" area is more as compared to "Urban" area.

Pie Chart of Recovery Time of Patient:

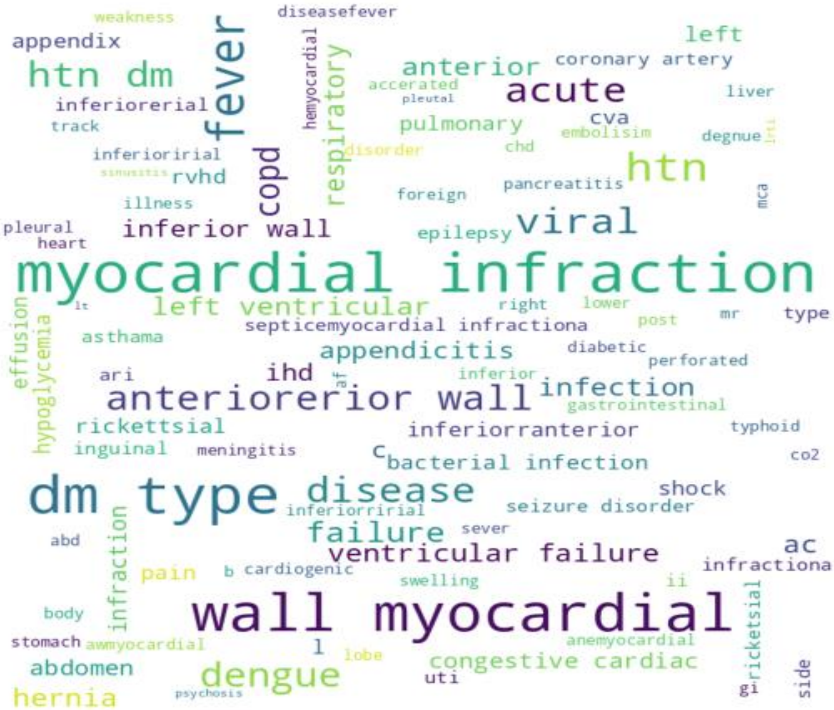


Conclusion: From the above Pie chart, we can observe that, there are 31% patients who recovered within 48 hours, 34% patients who recovered in between 48 to 72 hours and 35% patients who took more than 72 hours to recover. So, we can say that proportion of patients who took more than 72 hours to recover is more as compared to others.

Word Cloud for Medicines:

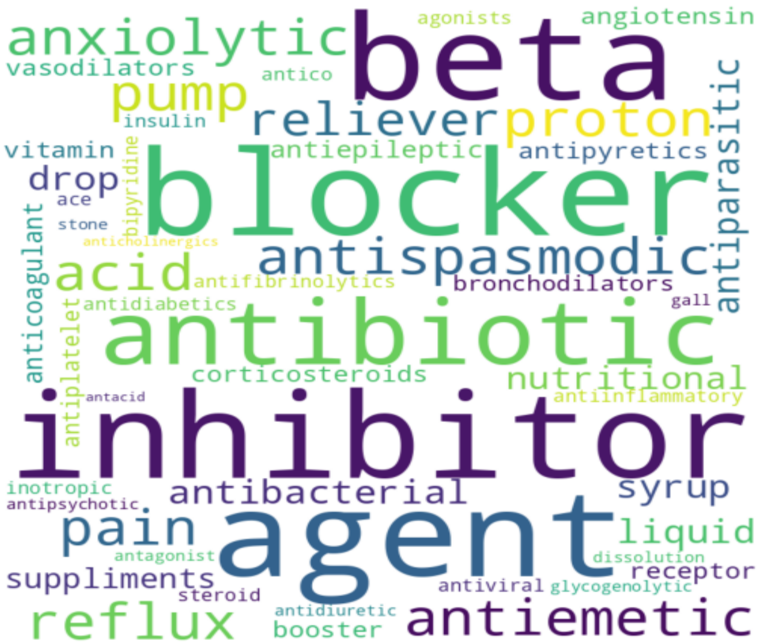
Conclusion: From the above Word cloud, we can observe that, most commonly used medicines are “Injection C-Tri” and “Injection Ondem”.

Word Cloud for Disease:



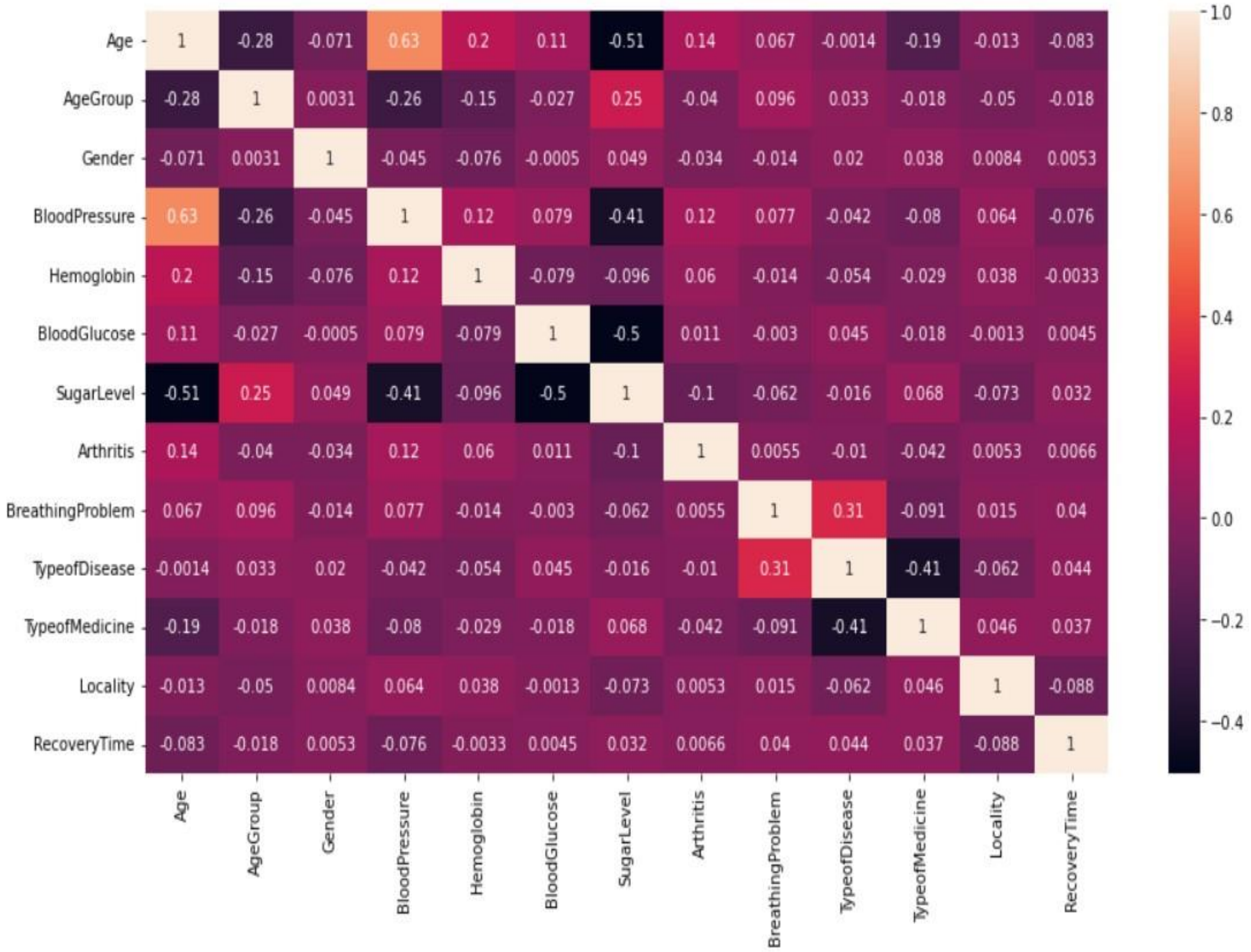
Conclusion: From the above Word cloud, we can observe that, most of the patients are affected by “Myocardial Wall Infraction” and “Diabetes”.

Word Cloud for Type of Medicines:



Conclusion: From the above Word cloud, we can observe that, most of the patients are treated with “Beta Blockers”, “Inhibitors”, and “Antibiotic”.

Correlation Plot:



Interpretation: Based on the correlation plot, it is clear that "Age", "Blood Pressure", "Sugar Level", "Breathing Problem", "Type of Disease", "Type of Medicine", and "Locality" are the features having higher correlation with the "Recovery Time".

2) DIAGNOSTIC STATISTICAL MODEL:

Age	Gender	BloodPressure	Hemoglobin	BloodGlucose	SugarLevel	Arthritis	BreathingProblem	TypeofDisease	TypeofMedicine	Locality	RecoveryTime
62	0	70	12.6	122	3	0	0	6	6	0	0
43	1	60	17.8	174	0	0	0	6	6	0	0
62	0	70	12.6	122	3	0	0	6	6	0	0
43	1	60	17.8	165	0	0	0	6	6	0	0
56	0	60	13.7	97	2	0	0	6	6	0	1

The above data consist of some important features and their description is explained below:

Age: The complete age of patient in years.

Gender: In this feature, there are two classes on the basis of whether the patient is Male or Female.

Female: - “0”

Male: - “1”

Blood Pressure: Blood pressure is the measurement of the pressure or force of blood inside your arteries. Each time your heart beats, it pumps blood into arteries that carry blood throughout your body. This happens 60 to 100 times a minute, 24 hours a day. Arteries deliver oxygen and nutrients to your whole body so it can function.

Hemoglobin: Hemoglobin (Hb) is the protein contained in red blood cells that is responsible for delivery of oxygen to the tissues. To ensure adequate tissue oxygenation, a sufficient hemoglobin level must be maintained. The amount of hemoglobin in whole blood is expressed in grams per deciliter (g/dl). The normal Hb level for males is 14 to 18 g/dl; that for females is 12 to 16 g/dl. Old Peak: ST depression induced by exercise relative to rest.

Blood Glucose: Blood glucose, or blood sugar, is the main sugar found in your blood. It is your body's primary source of energy. It comes from the food you eat. Your body breaks down most of that food into glucose and releases it into your bloodstream. When your blood glucose goes up, it signals your pancreas to release insulin. Insulin is a hormone that helps the glucose gets into your cells to be used for energy.

Sugar Level: In this feature, there are four classes on the basis of sugar level of patient.

High: - “0”

Low: - “1”

Normal: - “2”

Prediabetic: - “3”

Arthritis: Arthritis means joint inflammation. Arthritis is a medical condition in which the patient feels swelling and pain in their joints. Among adults aged 18 to 44. 7.1% ever reported having diagnosed with arthritis, and adults aged 45 to 64 29% reported having diagnosed with arthritis.

No: - “0”

Yes: - “1”

Breathing Problem: In this feature, we have taken two classes on the basis of patient having breathing problem or not.

No: - “0”

Yes: - “1”

Type of Disease: In this data, there are total 10 different types of disease. Epidemic Disease, Bacterial Infection, Kidney Disease, Neurological Disorder, Long Term Disease, Communica ble Disease, Respiratory Disease, Abdominal Disease, Heart Disease, Viral Disease.

Type of Medicines: The data consist of total 41 various types of Medicines. Antibiotic, Anti -inflammatory, Proton Pump Inhibitor, Antispasmodic, Acid reflux, Antacid, Antiemetic, An xiolytic, Pain Reliever, Antibacterial, Drop, Syrup, Antiphrastic, Liquid, Nutritional Supplem ents, Antiepileptic, Anticoagulant, Corticosteroids, Vasodilators, Antipyretics, Vitamin Boost er, Angiotensin receptor blocker, Bronchodilators, Antiplatelet, Antifibrinolytics, Beta Block ers, ACE Inhibitors, Antidiabetics, Beta Agonists, Antico, Steroid, Bipyridine Inotropic, Anti viral, Insulin, 'Glycogenolytic Agent, Antidiuretic, Antagonist, Antipsychotic, Gall Stone Di ssolution Agents, Anticholinergics.

Locality: In this feature, there are two classes on the basis of patient’s residence.

Rural: - “0”

Urban: - “1”

Recovery Time of Patient: The target feature consists of, in how much time a patient will recover, in this feature there are three classes of recovery time.
0 to 48 Hours: - “0”
48 to 72 Hours: - “1”
More than 72 Hours: - “2”

STATISTICAL ANALYSIS

CHI-SQUARED TEST:

Here we are testing the dependency of Recovery Time on the other features.

1) **Between Age Group and Recovery Time**

H_0: Recovery Time does not depend on Age Group.

v/s

H_1: Recovery Time depends on Age Group.

Contingency table:

Recovery Time / Age Group	0 to 48 Hours	48 to 72 Hours	More Than 72 Hours
Children	412	421	509
Youth	14	20	9
Adults	171	208	167
Seniors	53	51	30

Chi-2 Statistic: 11.1217
Degrees of Freedom:6
p-value: 0.08468

Decision Criteria: Here p-value: 0.08468 > 0.05

Decision:

We accept H_0 at 5% level of significance.

Conclusion:

Recovery Time may not depend on Age Group.

2) **Between Gender and Recovery Time**

H_0: Recovery Time does not depend on Gender.

v/s

H_1: Recovery Time depends on Gender.

Contingency table:

Recovery Time / Gender	0 to 48 Hours	48 to 72 Hours	More Than 72 Hours
Female	256	272	277
Male	394	428	438

Chi-2 Statistic: 0.04873
Degrees of Freedom: 2
p-value: 0.9759

Decision Criteria: Here p-value: 0.9759 > 0.05

Decision:

We accept H_0 at 5% level of significance.

Conclusion:

Recovery Time may not depend on Gender.

3) Between Sugar Level and Recovery Time

H_0: Recovery Time does not depend on Sugar Level.

v/s

H_1: Recovery Time depends on Sugar Level.

Contingency table:

Recovery Time / Sugar Level	0 to 48 Hours	48 to 72 Hours	More Than 72 Hours
High	72	84	67
Low	0	0	5
Normal	68	91	50
Prediabetic	510	525	593

Chi-2 Statistic: 11.3764
Degrees of Freedom: 4
p-value: 0.02264

Decision Criteria: Here p-value: 0.02264 < 0.05

Decision:

We reject H_0 at 5% level of significance.

Conclusion:

Recovery Time may depend on Sugar Level.

4) **Between Breathing Problem and Recovery Time**

H_0: Recovery Time does not depend on Breathing Problem.

v/s

H_1: Recovery Time depends on Breathing Problem.

Contingency table:

Recovery Time / Breathing Problem	0 to 48 Hours	48 to 72 Hours	More Than 72 Hours
No	613	658	657
Yes	37	42	58

Chi-2 Statistic: 1.3861
Degrees of Freedom: 2
p-value: 0.5000

Decision Criteria: Here p-value: 0.5000 > 0.05

Decision:

We accept H_0 at 5% level of significance.

Conclusion:

Recovery Time may not depend on Breathing Problem.

5) **Between Type of Disease and Recovery Time**

H_0: Recovery Time does not depend on Type of Disease.

v/s

H_1: Recovery Time depends on Type of Disease.

Contingency table:

Recovery Time / Type of Disease	0 to 48 Hours	48 to 72 Hours	More Than 72 Hours
Abdominal Disease	196	188	195
Bacterial Infection	54	43	29
Communicable Disease	42	41	7
Epidemic Disease	152	198	139
Heart Disease	14	42	100
Kidney Disease	12	7	6
Long Term Disease	4	16	27
Neurological Disorder	100	43	18
Respiratory Disease	16	26	34
Viral Disease	60	96	160

Chi-2 Statistic: 745.2438
Degrees of Freedom: 2
p-value: 0.0

Decision Criteria: Here p-value: 0.0 < 0.05

Decision:

We reject H_0 at 5% level of significance.

Conclusion:

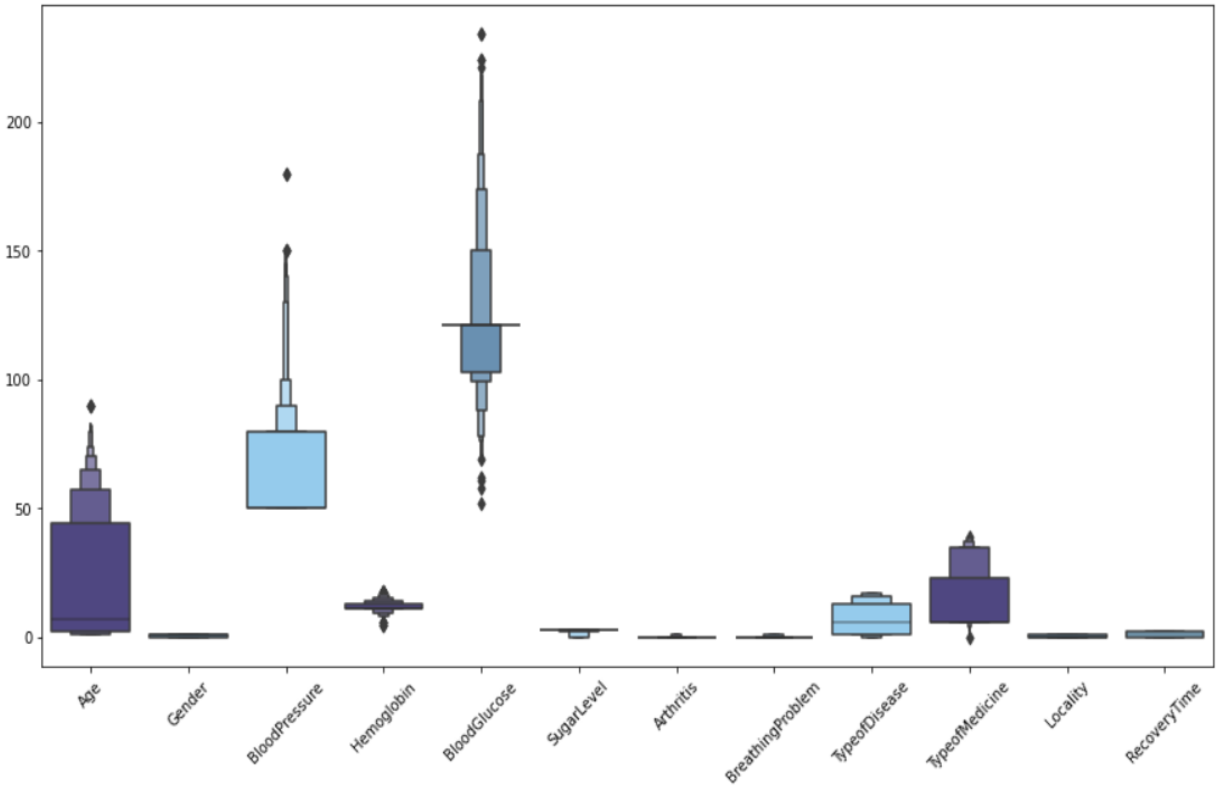
Recovery Time may depend on Type of Disease.

CONCLUSION:

From the above Chi-Squared Tests, we can conclude that, Recovery Time of patient depends on Sugar Level of patient and Type of Disease but it does not depend on Age Group, Gender, and Breathing Problem.

FEATURE SCALING

Plotting the Features using Boxen Plot:



Interpretation:

The above plot shows the range of features in the data. All the features are in different ranges.

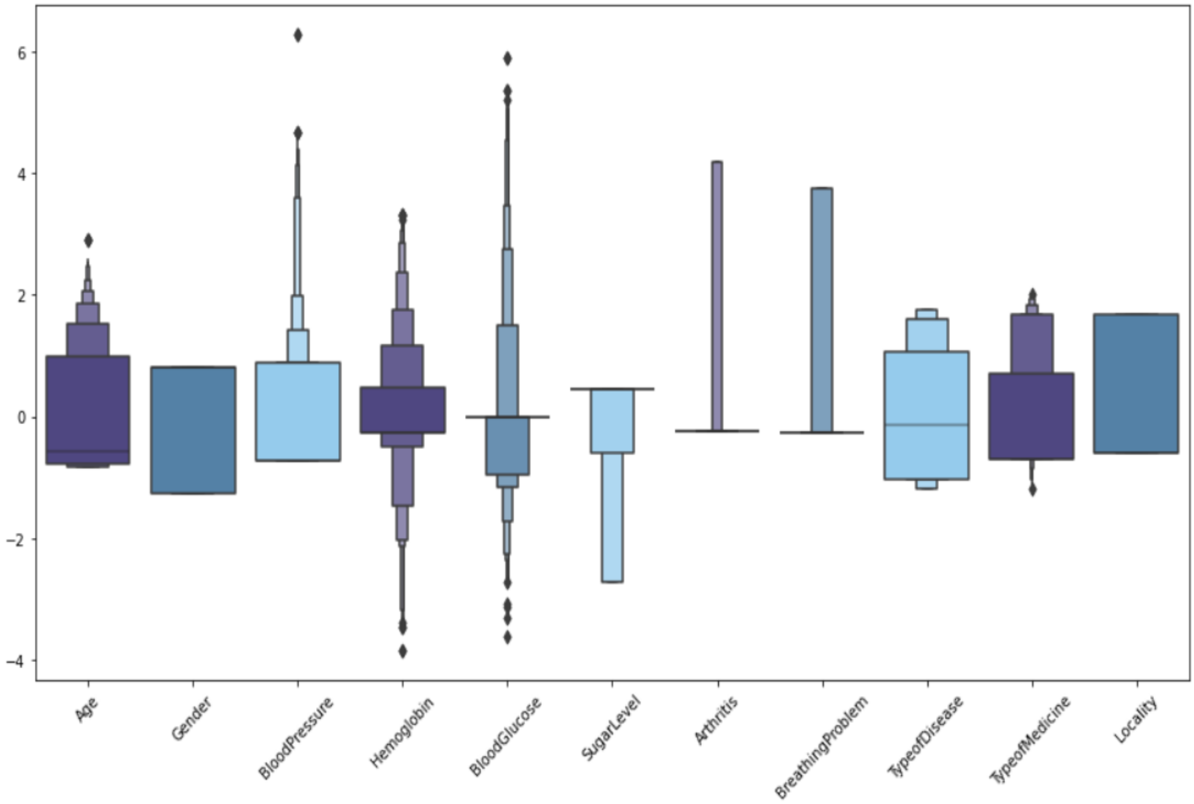
To fit this in a model we must scale it to the same range.

In the model building, we will preprocess the features to do the same.

Scaling the Data by setting up a Standard Scaler for the Features:

	count	mean	std	min	25%	50%	75%	max
Age	2065.0	1.517215e-16	1.000242	-0.815172	-0.773448	-0.564831	0.978933	2.898208
Gender	2065.0	2.095714e-16	1.000242	-1.251086	-1.251086	0.799305	0.799305	0.799305
BloodPressure	2065.0	5.592513e-16	1.000242	-0.710782	-0.710782	-0.710782	0.900317	6.270650
Hemoglobin	2065.0	-1.105572e-15	1.000242	-3.829736	-0.261878	-0.261878	0.479139	3.305981
BloodGlucose	2065.0	5.926924e-16	1.000242	-3.603845	-0.005531	-0.005531	-0.005531	5.887360
SugarLevel	2065.0	1.208826e-15	1.000242	-2.712177	0.453818	0.453818	0.453818	0.453818
Arthritis	2065.0	2.105822e-15	1.000242	-0.238341	-0.238341	-0.238341	-0.238341	4.195665
BreathingProblem	2065.0	4.747346e-17	1.000242	-0.266567	-0.266567	-0.266567	-0.266567	3.751399
TypeofDisease	2065.0	7.529086e-16	1.000242	-1.190737	-1.016645	-0.146187	1.072456	1.768823
TypeofMedicine	2065.0	2.000337e-15	1.000242	-1.184896	-0.693914	-0.693914	0.697203	2.006489
Locality	2065.0	9.997921e-16	1.000242	-0.593570	-0.593570	-0.593570	1.684723	1.684723

Plotting the Scaled Features using Boxen Plots:



Interpretation: The above plot clearly indicates that, now all the features are in same range since we have scaled the data.

3) PREDICTIVE STATISTICAL MODEL:

Model Building-

Splitting Train & Test Set:

Taking randomly 80% dataset as train dataset and 20% dataset as test dataset. Based on this we fit some following models.

Model Selection:

- **Decision Tree Classifier-**

  Output: Accuracy of Decision Tree Classifier is **83.29%.**

- **Random Forest Classifier-**

  Output: Accuracy of Random Forest classifier is **84.98%.**

- **Gradient Boosting -**

  Output: Accuracy of Gradient Boosting classifier is **74.81%.**

- **Ada boost -**

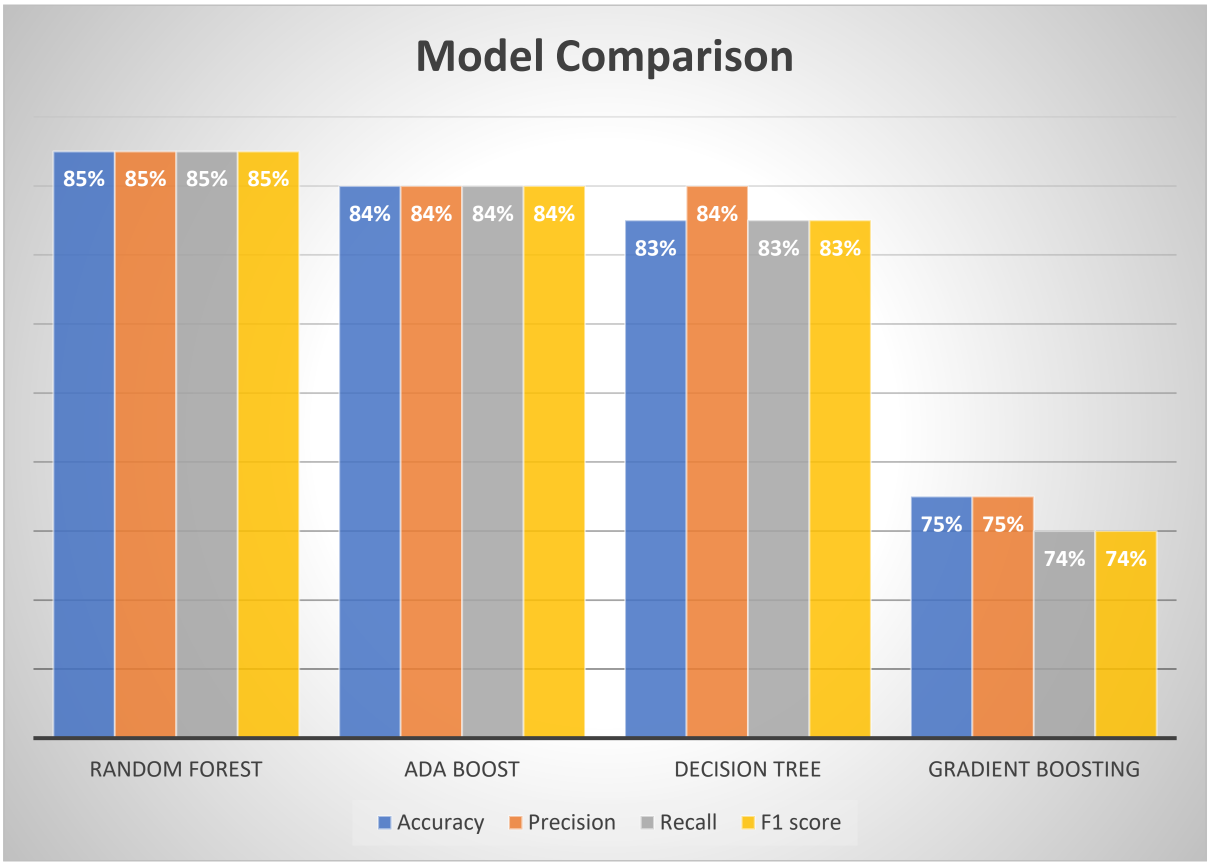
  Output: Accuracy of Ada Boost classifier is **84.26%.**

Conclusion:

Random Forest Classifier has greater accuracy than the other models. So Random Forest does best amongst the models to be the most accurate.

MODEL COMPARISON

Model	Accuracy	Precision	Recall	F1 score
Random Forest	85%	85%	85%	85%
Ada boost	84%	84%	84%	84%
Decision Tree	83%	84%	83%	83%
Gradient Boosting	75%	75%	74%	74%



Conclusion:

Above graph also shows that, Random Forest performs best among all the other models, with greater Accuracy, Precision, Recall and F1 score.
Hence, we will fit Random Forest Model for predictions.

RANDOM FOREST MODEL

Classification Report of Fitted Model:

```
              precision    recall  f1-score   support

           0       0.76      0.86      0.80       127
           1       0.81      0.78      0.79       138
           2       0.99      0.91      0.95       148

    accuracy                           0.85       413
   macro avg       0.85      0.85      0.85       413
weighted avg       0.86      0.85      0.85       413
```

Conclusion: From the above classification report, we can see that, the accuracy of fitted model on test set is 85%.

Confusion Matrix for Fitted Model:

	Predicted class			
Actual Class		**0**	**1**	**2**
	0	109	18	0
	1	29	107	2
	2	6	7	135

Conclusion:
From the above Confusion Matrix, we can conclude that,
86% of class 0, truly predicted 0,
78% of class 1, truly predicted 1,
91% of class 2, truly predicted 2.

4) PRESCRIPTIVE STATISTICAL MODEL:

How does this model work?

To build this model, we have used input method. Model works on Random Forest Classifier. The model requires patient's age, gender, medical history like blood pressure, hemoglobin, blood glucose, sugar level, arthritis, and what type of disease they are suffering from and their locality. We provide these inputs to the model, and the model informs us about how much time patient might need to recover. The output is in the terms of hours, in how many hours the patient will recover? So, the hospital management will get an idea about how to manage the beds, inventories, and all the other necessary equipment's needed for the patient to recover.

Demonstration:

Suppose a case of a male patient from rural area having **age 56**, with medical history i.e**., Blood Pressure: 80, Hemoglobin: 15.9, Blood Glucose: 121, Sugar Level: Prediabetic "3", Arthritis: No "0", Breathing Problem: No "0"** suffering from **Thrombocytopenia** which belongs to class of **Bacterial Infection "2",** arises at a hospital for treatment. The user will provide these inputs to the model, and then the model will predict how long will it take for a patient to recover.

Model will predict Recovery Time of Patient as one of the following:

- Patient might recover within 48 hours.
- Patient might recover in between 48 to 72 hours.
- Patient might take more than 72 hours to recover.

Using the output given by the model, the hospital management will get an idea about how to manage the beds, inventories and other necessary equipment's required for the recovery of the patient. This Prescriptive Model will help hospitals to avoid improper management of inventories, will also help to control over managing, allotting, and maintaining beds.

Model works as given below:

```
Age :56
Gender :1
Blood Pressure :80
Hemoglobin :15.9
Blood Glucose :121
Sugar Level :3
Arthritis :0
Breathing Problem :0
Type of Disease :2
Type of Medicine :6
Locality :0
Patient might take more than 72 hours to recover.
```

From above output, we can see that, patient of **Age: 56**, **Gender: Male “1”, Blood Pressure: 80**, **Hemoglobin 15.9**, **Blood Glucose: 121, Sugar Level: Prediabetic “3”, Arthritis: No “0”, Breathing Problem: No “0”** suffering from **Thrombocytopenia** which belongs to class of **Bacterial Infection “2”**, if the patient is treated with **Antibiotic “6”** type of medicine. Then the model predicts that patient might take more than 72 hours to recover, and using this result the hospital management can manage their beds, inventories, and other necessary equipment’s.

CONCLUSIONS

- The proportion of patients who took more than 72 hours to recover is slightly more as compared to others.
- Most of the patients are treated with "Antibiotic" and "Proton Pump Inhibitor" type of medicines.
- The features "Age", "Blood Pressure", "Sugar Level", "Breathing Problem", "Type of Disease", "Type of Medicine", and "Locality" having higher correlation with the "Recovery Time".
- The Recovery Time of patient depends on Sugar Level of patient and Type of Disease but it does not depend on Age Group, Gender and Breathing Problem.
- Random Forest performs best among all the other models, with greater Accuracy, Precision, Recall and F1 score.
- The accuracy of Random Forest Model on test set is 85%.

DISCUSSION

In Healthcare sector inventory management is incredibly important in Hospital wards. Inventory management systems can help hospitals and other medical organizations streamline their processes to save time while providing quality care to patients. This study proposes an innovative model for hospitals to manage their beds, equipment's inventories with the help of predictive and prescriptive Analysis integrated with python and other machine learning software. The lack of real-time data analysis provides a disadvantage resulting in patient overcrowding and malfunctions in in-patient bed management services during emergencies. Monitoring and managing inpatient-outpatient capacity & patient inflows can be challenging tasks during emergencies. Thus, the proposed models give the best predicted values with greater accuracy and precision. In future all Hospitals may require Inventory management system to provide quality care for their patients.

SUGGESTIONS

- On the basis of this study, we suggest hospitals to deploy this model or models similar to this in their management system to avoid overcrowding and inconvenience to the patient and their relatives.

- From this study, we suggest government hospitals to frame facilities to improve their inventory management so that the needy people will get an advantage of all healthcare facilities without any delay.

SCOPE & LIMITATIONS

SCOPE

- In this study, we have proposed an innovative machine learning model which can be very helpful for all local and multispecialty hospitals, so they can avoid the crowding and effectively manage all necessary inventories in their ward.

- This model can also be very helpful for Government Hospitals. From this study, Government Hospitals can change their policies towards inventory management system, so common people will get all the necessary medical services in government hospitals.

LIMITATIONS

- In this study, we have collected the data from a single hospital but which is not enough for being this model as global solution for inventory management system.
- This study is based only on handful of diseases and the data is collected from a hospital belonging to specific city.

REFERENCES

- Prescriptive analytics for inventory management in health care (Leonardo Galli, Tommaso Levato, Fabio Schoen & Luca Tigli)

- Inventory Management Applications for Healthcare Supply Chains (Laquanda Leaven, Kamal Ahmmad and Demesha Peebles)

- Data Mining by Mathew A. Russell

- Baron, O. (2010). Managing perishable inventory. Wiley Encyclopedia of Operations Research and Management Science.

- Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive Analysis. Management Science, 66(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253