

Hospital Recommendation System based on multiple diseases using ML

Submitted in partial fulfilment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

CH. SAIBHAGAVAN MOURYA (37110127)
B. YESHWANTH (37110083)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH - 2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with 'A' grade by NAAC
Jeppiar Nagar, Rajiv Gandhi Salai, Chennai - 600 119
www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **CH. SAI BHAGAVAN MOURYA** (37110127) and **B. YESHWANTH** (37110083) who carried out the project entitled "**Hospital Recommendation System based on multiple diseases using ML**" under my supervision from October 2019 to April 2020.

Internal Guide

Mr. S. DHAMODHARAN., M. Tech., (Ph.D.,)

Head of the Department

Dr.L.Lakshmanan M.E., Ph.D.,

Dr. S.Vigneshwari M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We **CH.SAIBHAGAVAN MOURYA** (Reg. No. 37110127) and **B. YESHWANTH** (Reg. No. 37110083) hereby declare that the Project Report entitled “**Hospital Recommendation System based on multiple diseases using ML**” done by us under the guidance of **Mr. S. DHAMODARAN., M. Tech., (Ph.D.,)** is submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

- 1.**
- 2.**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala, M.E., Ph.D., Dean**, School of Computing, **Dr. S. Vigneswari, M.E., Ph.D.**, and **Dr. L. Lakshmanan, M.E., Ph.D.**, Heads of the **Department** of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mr. S. DHAMODARAN., M. Tech., (Ph.D.,)** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE NO
	ABSTRACT	
	LIST OF FIGURES	
1	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 SCOPE OF THE PROJECT	2
	1.3 MOTIVATION	2
2	LITERATURE REVIEW	3
	2.1 LITERATURE REVIEW	3
	2.2 EXISTING SYSTEM	4
	2.2.1 DISADVANTAGES	4
	2.3 PROPOSED SYSTEM	4
	2.3.1 ADVANTAGES	5
3	AIM AND SCOPE OF PROJECT	6
	3.1 AIM OF PROJECT	6
	3.2 SCOPE OF THE PRESENT INVESTIGATION	6
	3.3 SYSTEM ARCHITECTURE	7
	3.4 SYSTEM REQUIREMENTS	8
	3.4.1 HARDWARE REQUIREMENTS	8
	3.4.2 SOFTWARE REQUIREMENTS	8
4	ALGORITHMS AND MATERIALS	9
	4.1 REQUIREMENT ANALYSIS	9
	4.2 MODULES	9
	4.2.1 INPUT DATASET	9
	4.2.2 BACKEND	10
	4.3 ALGORITHMS	10
	4.3.1 DECISION TREE	10
	4.3.2 NAÏVE BAYES	12
	4.3.3 FRONTEND	13

	UML DIAGRAMS	13
	USE CASE DIAGRAM	14
	ACTIVITY DIAGRAM	16
5	RESULTS AND DISCUSSION	17
	RESULTS	17
	SCREENSHOTS	17
6	SUMMARY AND CONCLUSION	22
	CONCLUSION	22
	FUTURE WORK	22
	APPENDIX	22
	REFERENCES	24

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
3.1	System Architecture	7
4.1	Input Data set	9
4.2	Example of Decision Tree	11
4.3	Example of Naïve Bayes	12
4.4	Example of Use case Diagram	14
4.5	Example of Activity Diagram	15
4.6	Example of Collaboration Diagram	16
5.1	Set Location image	17
5.2	Select Hospital image	18
5.3	Select Disease image	18
5.4	Select Another Hospital	19
5.5	Success Rate	19
5.6	Select All Hospital	20
5.7	Top Two Hospital Success Rates	20
5.8	Success rate for another disease selected	21
6.1	Python Sample Code	22
6.2	Java Sample Code	23

CHAPTER 1

INTRODUCTION

INTRODUCTION:

In this venture, our motivation is to give an instrument to assist patients with deciding and pick the most appropriate emergency clinic for their sickness. In request to accomplish this objective, we made a framework that permits clients to inquiry clinics that meet certain conditions, for instance, regarding the quantity of patients in the emergency clinic that treats the sickness, a specific number of patients have been effectively restored, and patient data can be seen.

For instance, an older diabetic female patient is youthful in Chennai and does not know which emergency clinic she should go to for treatment. For this situation, utilize our arrangement. He can discover a medical clinic close to him with a high pace of accomplishment in treating diabetes. To address this inquiry, we gather information from different clinics in the city. At that point, we determined the achievement pace of every sickness treated by the medical clinic.

For improving patient satisfaction, issues of health care provided at the hospital level and the factors that originate those issues from patient's point of view should be discovered. Therefore, survey data should be manually analyzed by examining each possible pattern in the data set using conventional methods is needed to do the analysis with least amount of human interaction. Such methodology should get the satisfaction survey data, find patterns that are repeated among patients' demographics and their satisfaction level in different fields, validate the patterns and compile them into a set of recommendations to help hospitals improve satisfaction within various patient communities.

To this end, a new hybrid methodology is proposed that differs from these conventional approaches in that it is not bound to a single outcome. The focus of this approach is to find patterns in patients' responses to all satisfaction questions and relate them to patients' demographics. The proposed methodology is focused on discovering

Issues of the health care provided at the hospital level and the factors that originate those issues from the patients' point of view. The associations are validated by using standard statistical models and turned into useful recommendations for hospitals to improve patients' experience, save cost, and build long-term patient loyalty. The methodology can be generalized to any complex multi-level survey analysis.

SCOPE OF THE PROJECT:

Present we have several recommendation systems regarding healthcare, but it will be recommend based on patient surveys. Unlike those in our project recommend the hospital based on success rate of a particular disease for any hospital.

MOTIVATION:

Whenever we travel to new place and if we have any health problems then we need to go to hospital, there may many hospitals nearby, but we do not know which hospital is the best for our problem.

In such cases we usually ask people nearby to us for a suggestion, and we may or may not get a correct answer for that. So, we developed a web interactive page here one can get a suggestion about the hospital. By showing the success rate of that hospital which is best one to recommend.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW:

Title 1: "A Hospital Recommendation System Based on Patient Satisfaction Survey/2017". Mohammad Reza Khoie, Tannaz Sattari Tabrizi, and Nina Marhamati are the authors of this paper. Here they are using K- means Clustering technique in order to analyze the given unstructured data of patient surveys from various sources to recommend the hospital.

Title 2: "A temporal matching between poi popularity and user regularity," in 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 549- 558, IEEE, 2016. The following where the authors of this paper Z. Yao, Y. Fu, B. Liu, Y. Liu, and H. Xiong. They used Decision making as the process of making choices by identifying a decision, gathering information, and assessing alternative resolutions. Using a step-by-step decision- making process can help you make more deliberate, by organizing relevant information and defining Alternatives.

Title 3: "Social topic modeling for point-of-interest recommendation in location-based social networks" in 2014 IEEE International Conference on Data Mining (ICDM), pp. 845-850, IEEE, 2014. The following are the authors of this B. Hu and M. Ester. They used Naive Bayes approach to recommend the point of interest in location based social networks .A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. These classifiers are widely used for machine learning because they are simple to implement

EXISTING SYSTEM:

Hospital uses only manual system, and they would not release the information of the patients that should be a critical information for hospital and patient records. Using numerous papers to store the patient information to use of that to get a survey of that to recommend the best one based on the patient survey who get cured to the disease based on that other related persons to recommend that. It will use the patient feedback survey to recommend a hospital.

DISADVANTAGES:

- More Paperwork to storing the information.
- Workload for employee's and increasing the hospital admissions.

PROPOSED SYSTEM:

The hospital recommendation system is designed for any hospital to replace their manual based storing mechanism and will be using the online method to recommend the hospital based on the patient information to nearby hospitals or with the success rate to recommend it for users to reduce the time and using these resources to save time and money. Using this proposed algorithm apply the proposed model to certain hospitals and constructed an online hospital recommendation system which can provide users with a hospital rank list basing the real-time population density information and the hospitals' basic information such as their distances

ADVANTAGES:

- Reduce the paperwork.
- Easy to access and user friendly.
- Cost Reduction.
- Increasing patient satisfaction in and out.

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

AIM OF PROJECT:

The main aim of this project is to develop a Hospital Recommendation system based on multiple diseases, In which we recommend hospitals based on the success rate of a particular disease in the selected locality.

SCOPE OF THE PRESENT INVESTIGATION:

Here we gather Patient datasets from hospitals in a particular locality and now using these datasets we calculate the success rate of every disease treated by that hospital using Machine learning algorithms. Now when a patient enters locality and disease, he is suffering from our system predicts the hospital with higher success rate for that particular disease. Unlike many existing Hospital recommendation Systems available online who predict the hospitals based on patient feedback surveys our system will predict the hospital based on the success rate of that hospital.

System Architecture

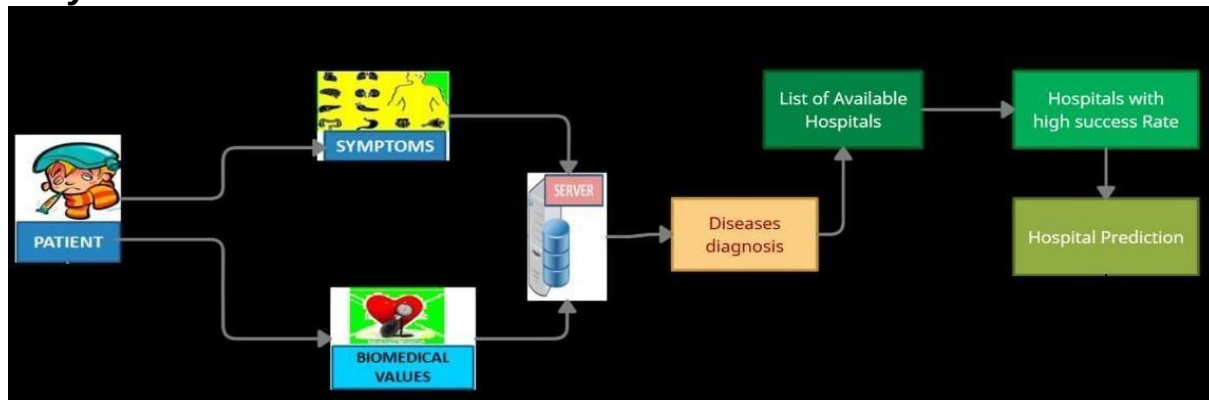


Fig.3.1. System Architecture

Our main aim of the project to describe through this system architecture it will shows the patient or person to use our web interface to access it. While access the interface of the web application if he or she have any medical obligations or any doubts of him or her health to enter the symptoms and if possible, to add biomedical values. Our server will analyze and validate it. To recommend best hospital on him or her disease to recommend the best hospital to basis on the success rate and it would be the feedback of the patients in a survey.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS

System	: Pentium Dual Core.
Hard Disk	: 120 GB.
Monitor	: 15 LED
Input Devices	: Keyboard, Mouse
Ram	: 1GB.

SOFTWARE REQUIREMENTS

Operating system	: Windows 7/10.
Coding Language	: Java, Python
IDE	: Eclipse, PyCharm

CHAPTER 4

ALGORITHMS AND MATERIALS USED

REQUIREMENT ANALYSIS:

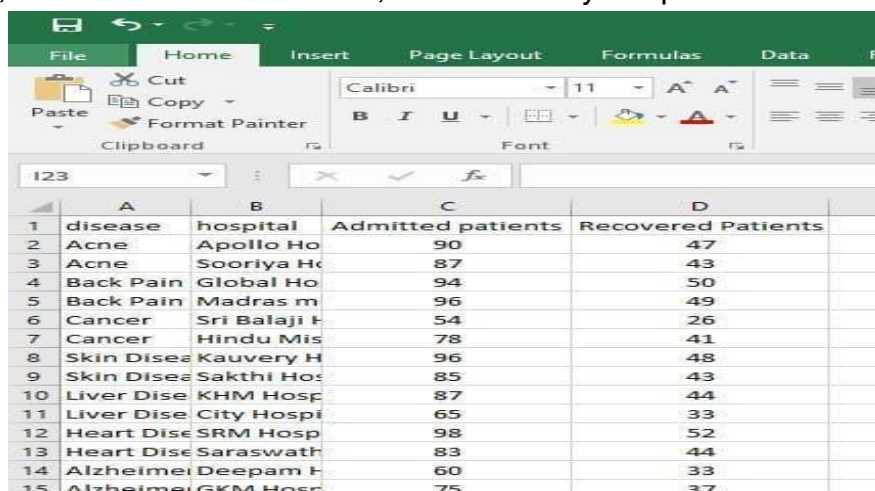
Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

MODULES:

- Input Dataset
- Backend
- User Interface or Frontend

INPUT DATASET:

- Here Input Data sets will be collected from various Hospitals in a particular city.
- The Input Data sets consists of no of patients admitted for each disease, Disease name, no of Patients Recovered, Time taken by the patients to recover.



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
	disease	hospital	Admitted patients	Recovered Patients
2	Acne	Apollo Ho	90	47
3	Acne	Sooriya Ho	87	43
4	Back Pain	Global Ho	94	50
5	Back Pain	Madras m	96	49
6	Cancer	Sri Balaji H	54	26
7	Cancer	Hindu Mis	78	41
8	Skin Disea	Kauvery H	96	48
9	Skin Disea	Sakthi Hos	85	43
10	Liver Dise	KHM Hosp	87	44
11	Liver Dise	City Hospi	65	33
12	Heart Dise	SRM Hosp	98	52
13	Heart Dise	Saraswath	83	44
14	Alzheimer	Deepam H	60	33
15	Alzheimer	GKM Hosp	75	37

Fig 4.2. Input Dataset

BACKEND

- In this program the backend process of our web App is made using python. Analysis of the hospital data sets is done by using machine learning algorithms Decision Tree algorithms and Naïve Bayes Algorithm. The success rate of a disease in all hospitals in our locality is calculated here.
- Based on the Input Data Set Success rate for each disease is calculated by the system.
- Now when a user enters the Disease, he is suffering from the system will suggest the best hospital for the treatment of that disease near him.

ALGORITHMS USED:

Decision Tree algorithm

Naïve Bayes Algorithm

DECISION TREE ALGORITHM

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split. For example



Fig no: 4.3 Example of Decision Tree

An example of a decision tree can be explained using above binary tree. Let's say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc. The decision nodes here are questions like 'What's the age?', 'Does he exercise?', and 'Does he eat a lot of pizzas? And the leaves, which are outcomes like either 'fit', or 'unfit'. In this case this was a binary classification problem (a yes, no type of problem). There are two main types of Decision Trees:

CLASSIFICATION TREES (Yes/No types)

What we have seen above is an example of classification tree, where the outcome was a variable like 'fit' or 'unfit'. Here the decision variable is Categorical.

REGRESSION TREES (Continuous data types)

Here the decision or the outcome variable is Continuous.

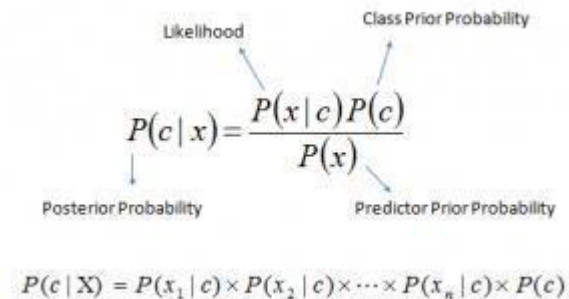
Now that we know what a Decision Tree is, we will see how it works internally. There are many algorithms out there which construct Decision Trees, but one of the best is called as ID3 Algorithm. ID3 Stands for Iterative Dichotomize 3.

NAIVE BAYES ALGORITHM

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

For example, some fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. Look at the equation below:



The diagram shows the Naive Bayes equation with labels for its components:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Labels and arrows:

- Likelihood** points to $P(x|c)$.
- Class Prior Probability** points to $P(c)$.
- Posterior Probability** points to $P(c|x)$.
- Predictor Prior Probability** points to $P(x)$.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Fig no: 4.3. Naïve Bayes Equation

Above,

- $P(c|x)$ is the posterior probability of class (c, target) given predictor (x, attributes).
- $P(c)$ is the prior probability of a class.
- $P(x/c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

APPLICATIONS OF NAIVE BAYES ALGORITHM:

- **Real time Prediction:** Naive Bayes is an eager learning classifier, and it is sure fast. Thus, it could be used for making predictions in real time.
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System:** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data

Mining techniques to filter unseen information and predict whether a user would like a given resource or not.

USER INTERFACE OR FRONTEND

The front end or the User Interface of the project is done using java in eclipse IDE. Here there are three steps.

1. Enter your locality.
2. Select any hospital to see the success rate of the disease you are suffering from. Or select all hospitals to find the best of all.
3. Now select the Disease. After selecting the disease, the system will show the success rate of hospital you selected or if you selected the option All hospitals the system will show the top two hospitals with highest success rate.

UML DIAGRAMS

UML is simply another graphical representation of a common semantic model. UML provides a comprehensive notation for the full lifecycle of object-oriented development.

ADVANTAGES:

- To represent complete systems (instead of only the software portion) using object oriented concepts
- To establish an explicit coupling between concepts and executable code
- To take into account the scaling factors that are inherent to complex and critical systems.
- To creating a modelling language usable by both humans and machines.

UML defines several models for representing systems.

- The class model captures the static structure.
- The state model expresses the dynamic behavior of objects.
- The use case model describes the requirements of the user.
- The interaction model represents the scenarios and messages flows.
- The implementation model shows the work units.

- The deployment model provides details that pertain to process allocation.

USECASE DIAGRAM

Use case diagrams overview the usage requirement for system. They are useful for presentations to management and/or project stakeholders, but for actual development you will find that use cases provide significantly more value because they describe the meaning of the actual requirements. A use case describes a sequence of action that provides something of measurable value to an action and is drawn as a horizontal ellipse.

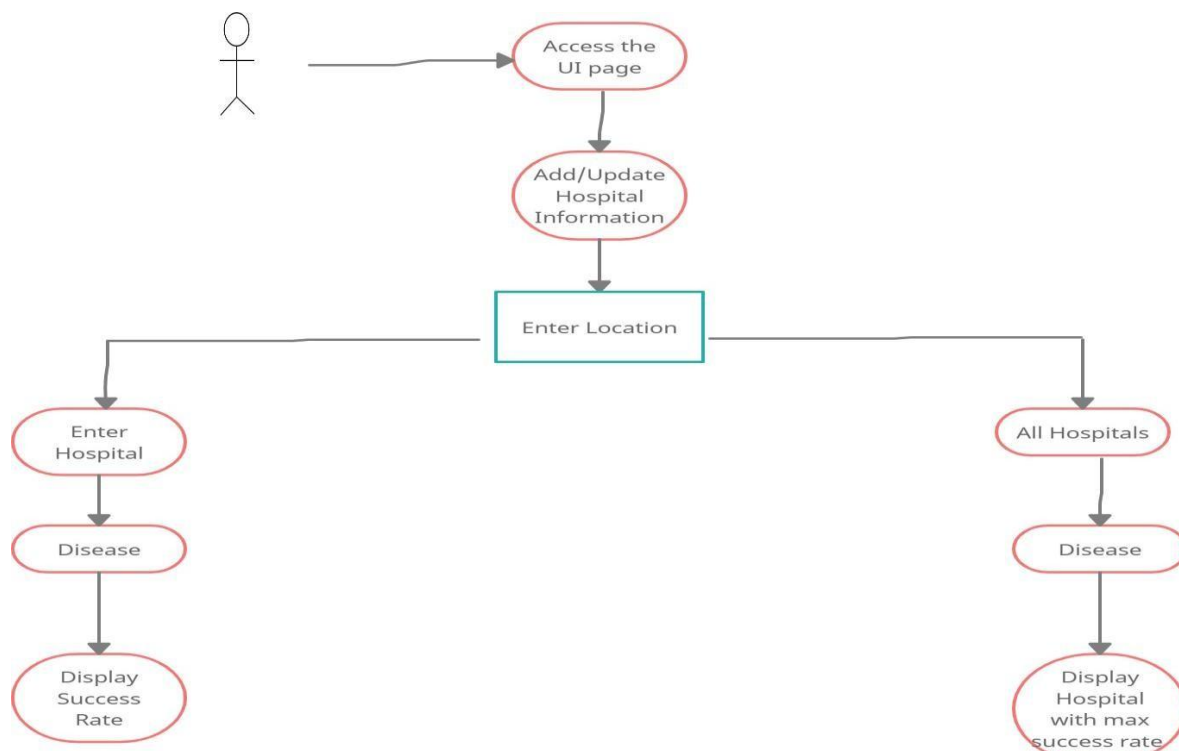


Fig No: 4.4 Example of Use Case Diagram

ACTIVITY DIAGRAM

Activity diagram are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. The activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. Activity diagrams consist of Initial node, activity final node and activities in between.

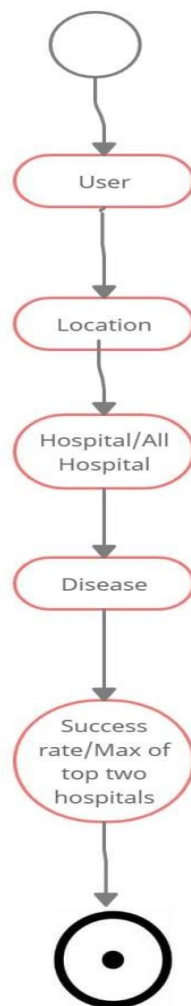
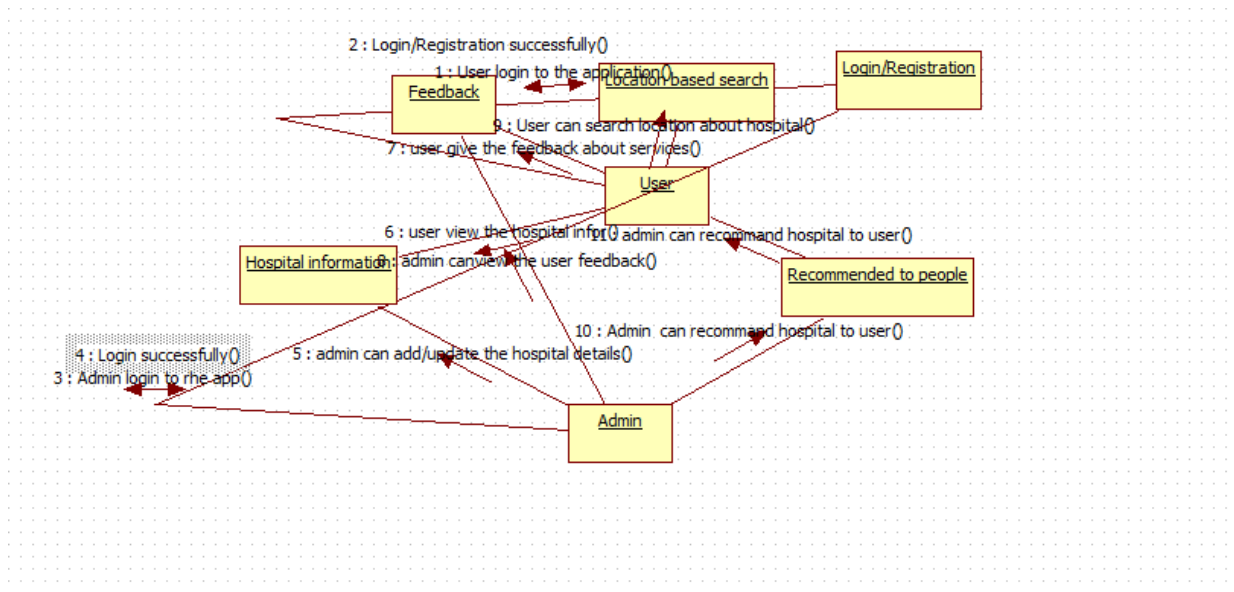


Fig No: 4.5 Example of Activity Diagram

Collaboration Diagram:

It is a type of interaction diagram and how to collaborate it within this. Set of objects to relate in a particular context and interaction which is a set of messages exchange among the objects within the collaboration to achieve a desired outcome.



CHAPTER 5

RESULTS AND DISCUSSION

Results:

Using our approach, the patient or person access the web application based on the symptoms to recommend the best hospital. Using Decision tree and Naïve Bayes approach analyze and validate the data. Initially we used to train the system and test it and get to calculate the success rate of each hospital of disease.

After that based on the symptoms and biomedical values to recommend the hospital along with the success rate of the correspondent hospital.

SCREENSHOTS:

Step 1: Selecting the location

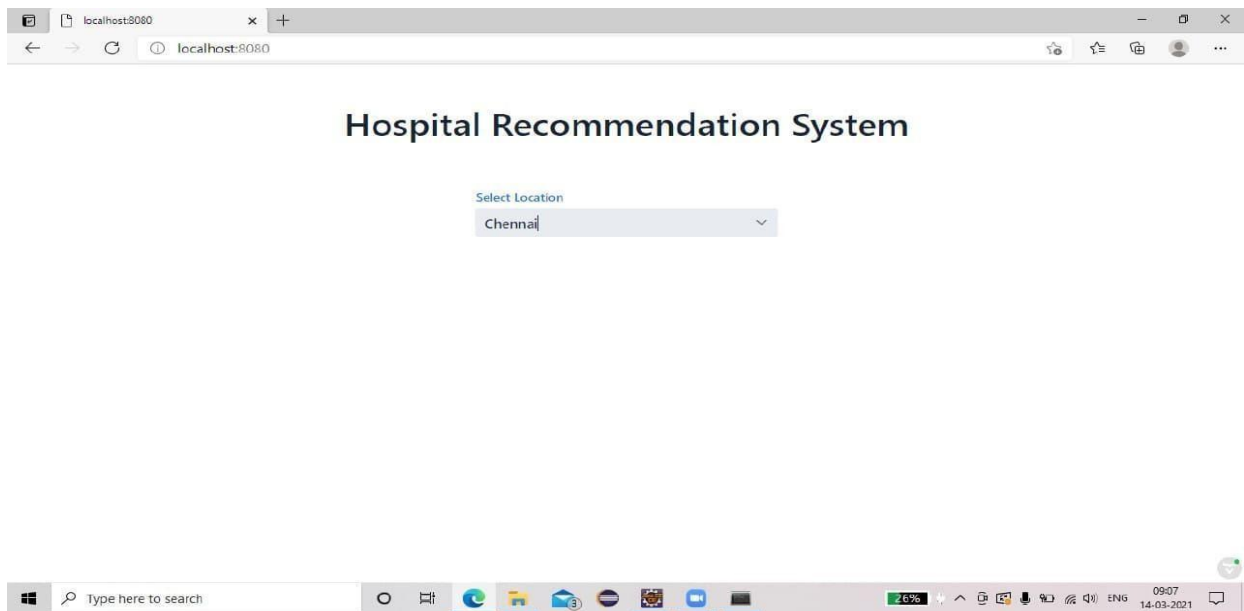


Fig. 5.1. Set Location

Step 2: Select the hospital

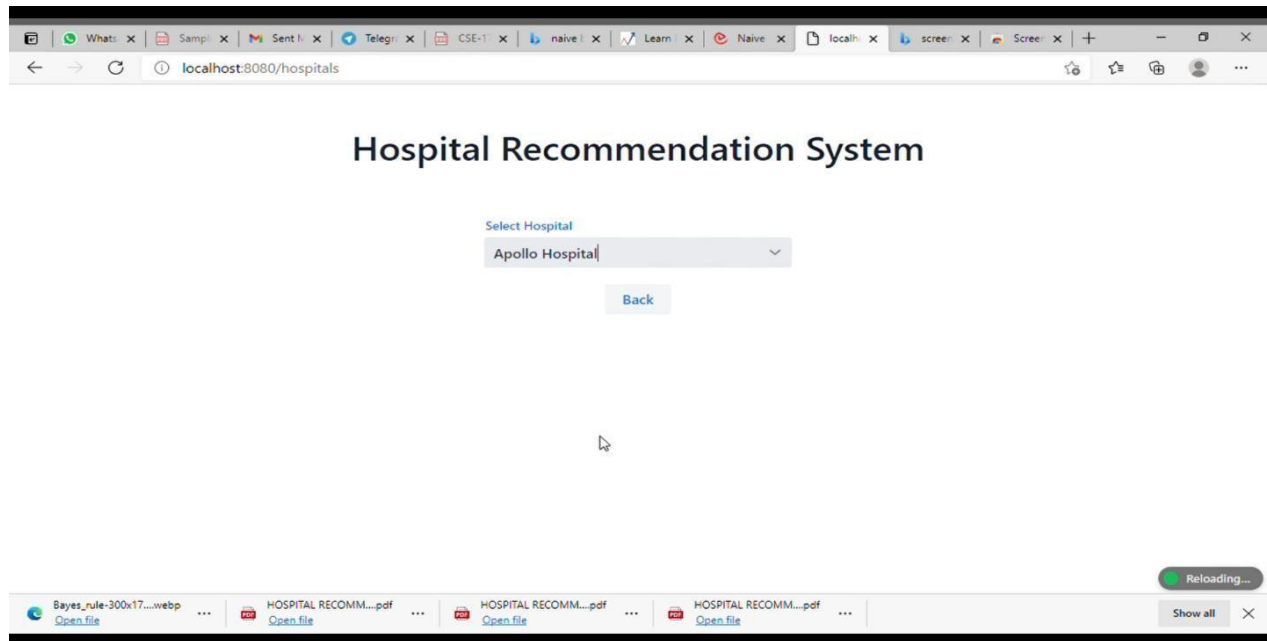


Fig 5.2. Select Hospital

Step 3: Select the Disease in particular hospital

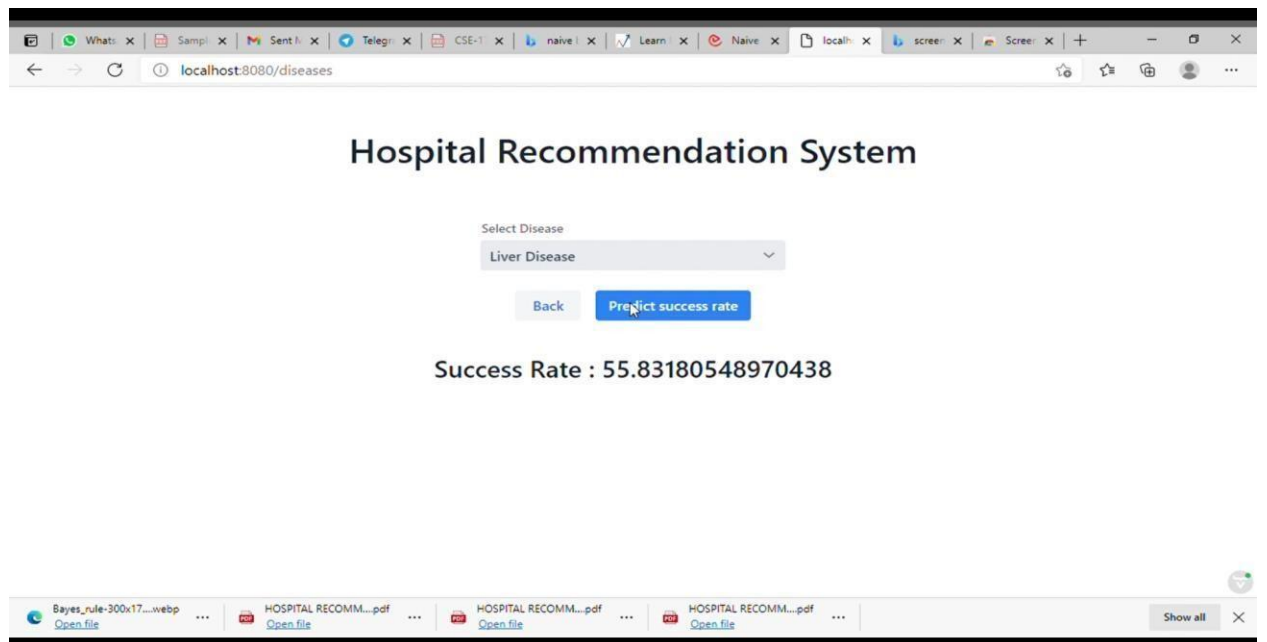


Fig 5.3. Select Disease

Step 4: Select the hospital

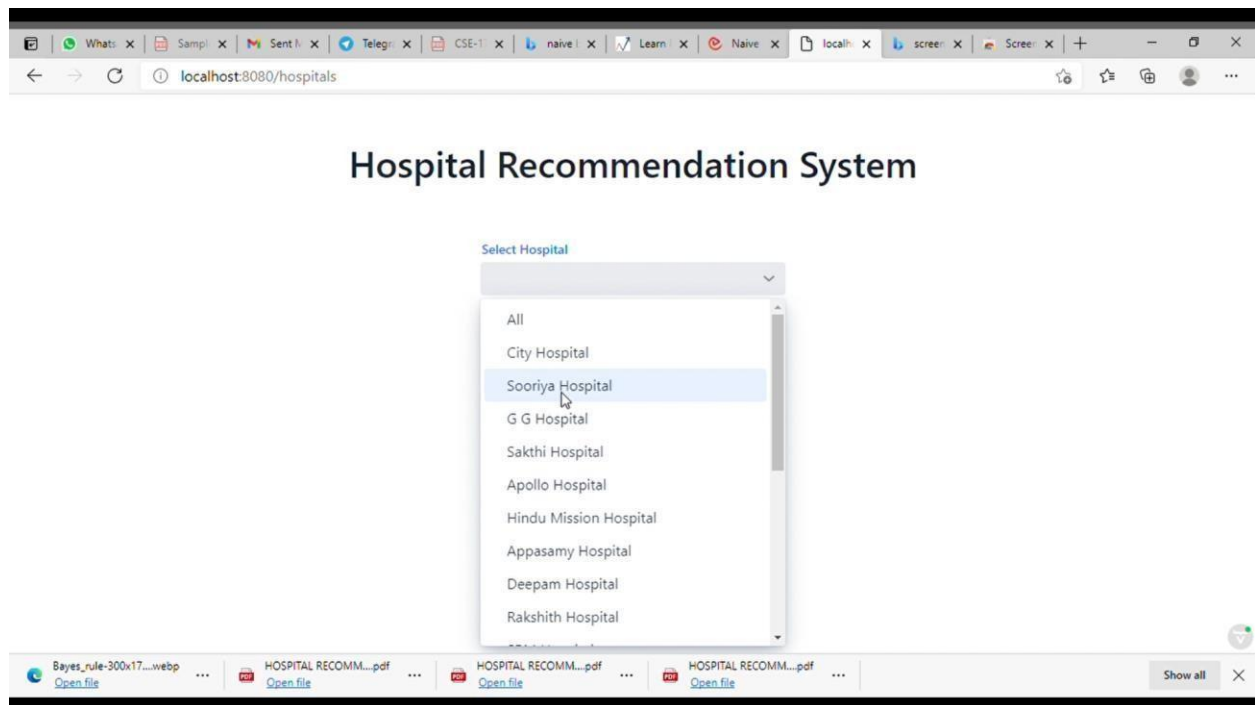


Fig 5.4. Select Hospital

Step 5:

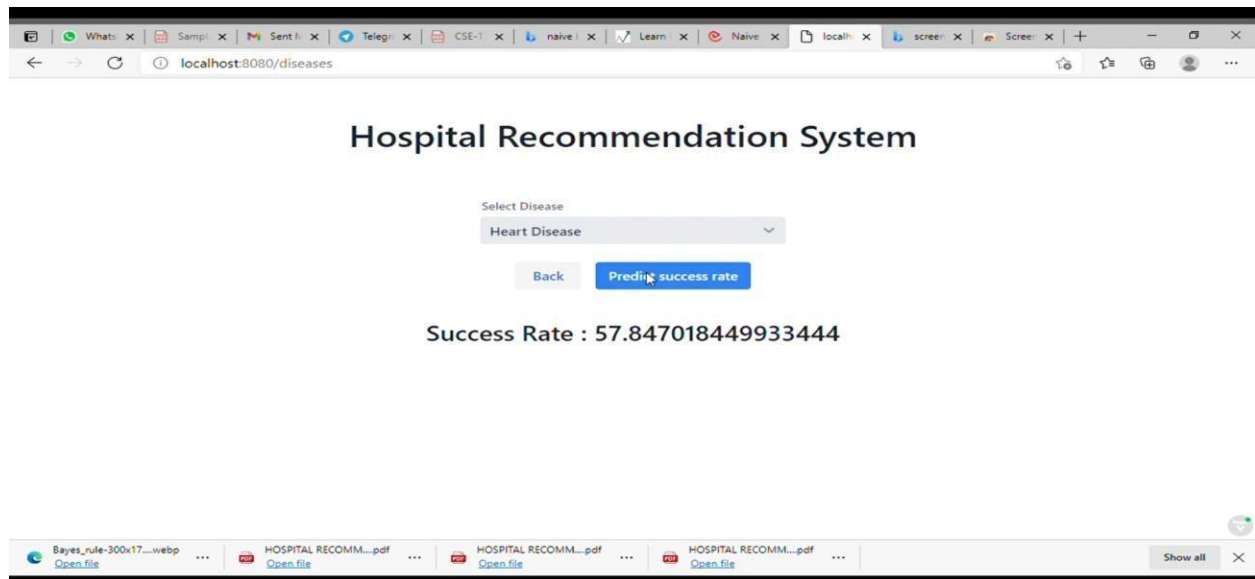


Fig 5.5. Success rate

Step 6: Select All hospitals

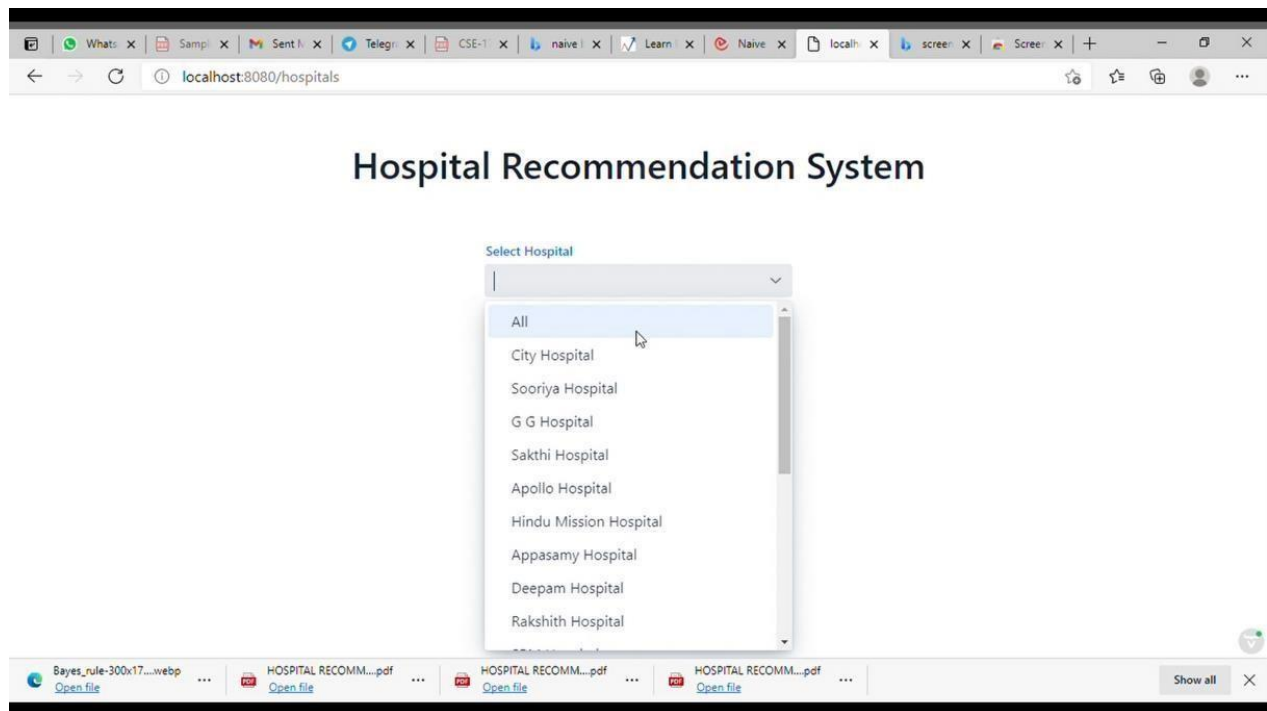


Fig 5.5. Select All Hospitals

Step 7: Select the disease and it will display the top two hospitals of maximum success rate.

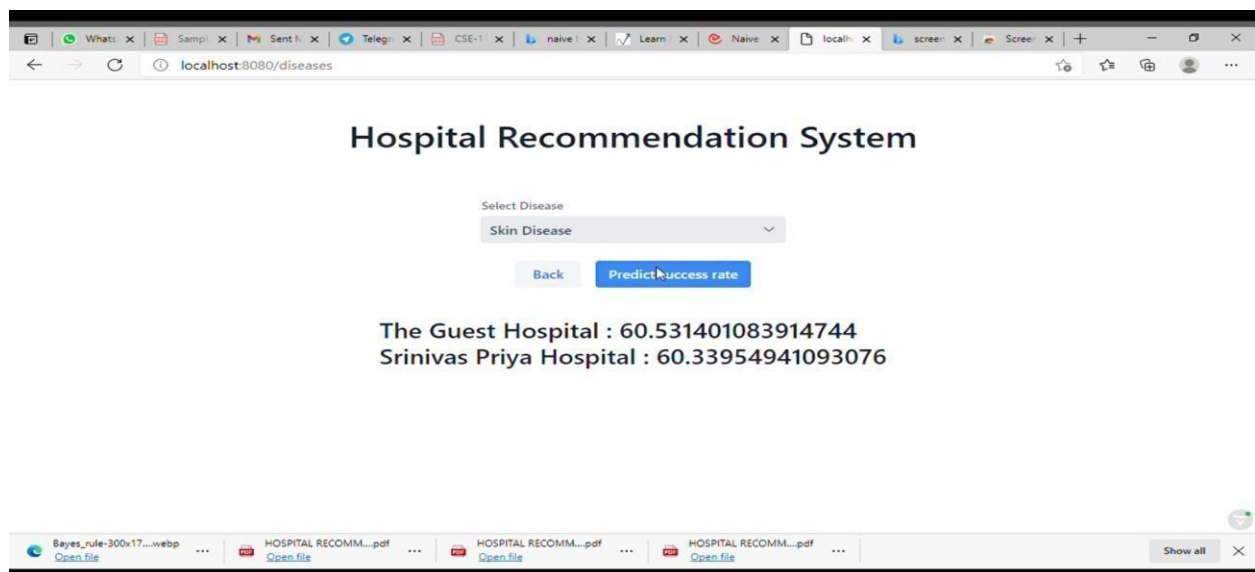


Fig 5.6. Top two success rates

Step 8:

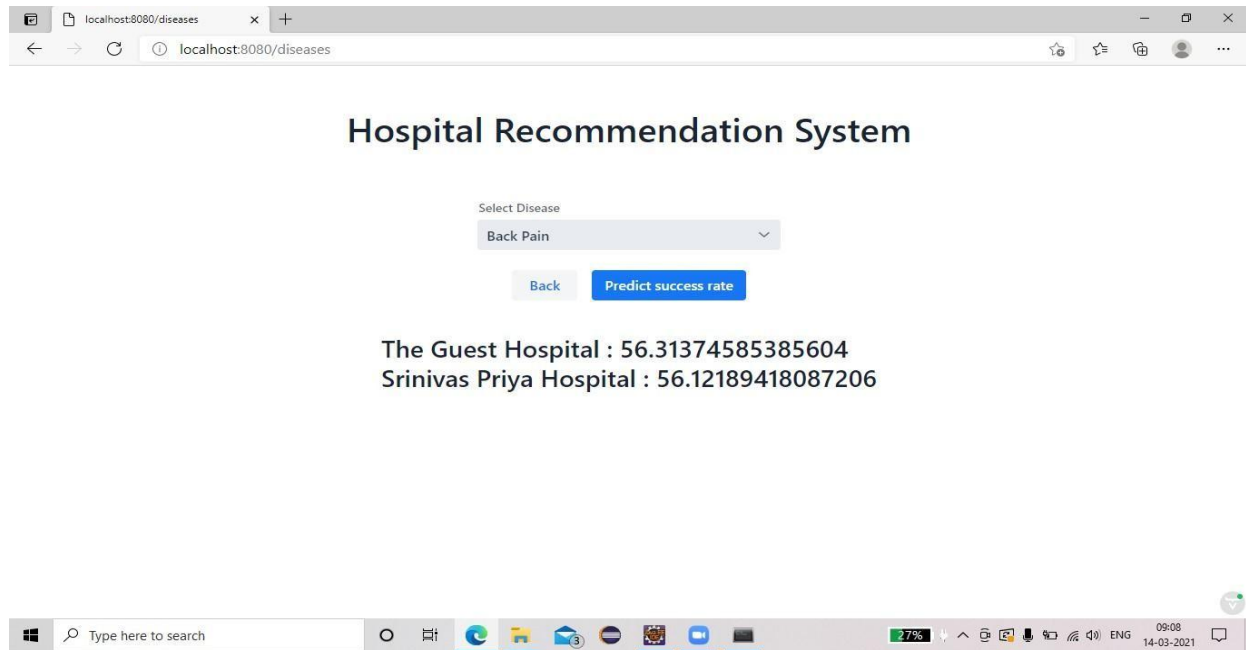


Fig 5.7 Another disease success rate

CHAPTER 6

SUMMARY AND CONCLUSION

Conclusion:

We presented a novel approach for hospital population density estimation based on Symptoms and Bio Medical values. To get the best hospital to recommend on basis of disease along with success rate and about to know the success rate and to choose the best one basis on disease. Using some approaches to analyze and validate it to recommend it. In this paper, we propose an approach for hospital recommendation based on disease.

Future Work:

For the future work to create a login and admin module to elaborate it and it should be obtained and rank answers based on incomplete information and provide personalization. To cope with incomplete and noisy data, we allow both exact and close matches when answering queries. We also present an intuitive approach to display answers to users, which aims to help users to understand the ranked results and possibly refine their queries. To resolving the queries and ranking system of the hospital on any disease with customer satisfaction in a survey to get that.

APPENDIX:

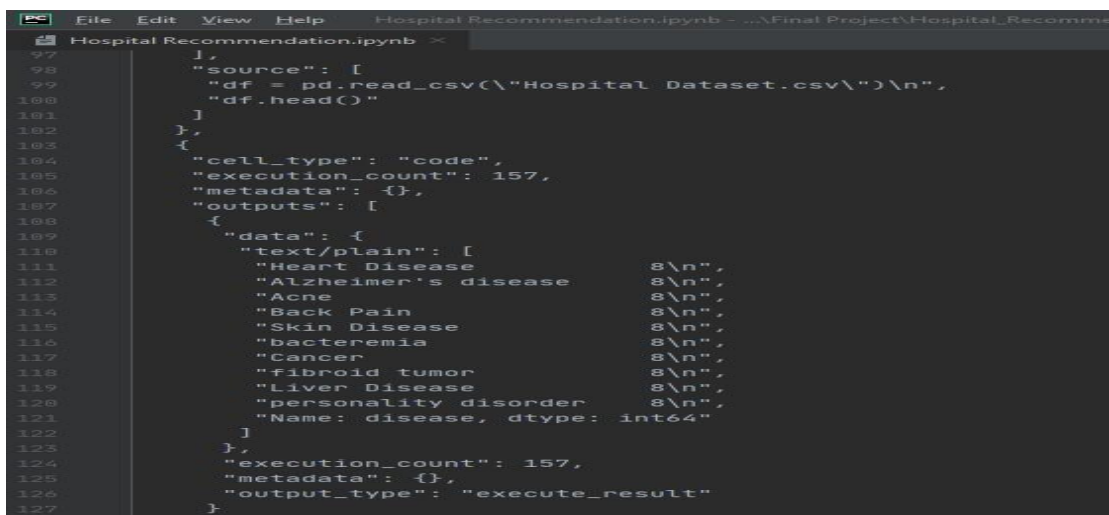
The image shows a Jupyter Notebook interface with a dark theme. The title bar at the top reads 'Hospital Recommendation.ipynb'. The code is written in a JSON-like format, typical of Jupyter notebooks. It includes a 'source' section with Python code that reads a CSV file and displays its head. The 'outputs' section contains a 'data' object with a 'text/plain' field listing various diseases: Heart Disease, Alzheimer's disease, Acne, Back Pain, Skin Disease, bacteremia, Cancer, fibroid tumor, Liver Disease, personality disorder, and Name: disease, dtype: int64. The code also includes metadata such as 'execution_count': 157 and 'output_type': 'execute_result'.

Fig No: 6.1 Python sample code

```

1 package in.hospital.recommendation.view;
2
3 import static com.vaadin.flow.component.orderedlayout.FlexComponent.Alignment.CENTER;
4
5 @SuppressWarnings("serial")
6 @Route(value = "")
7 public class LocationSelectionView extends VerticalLayout {
8
9     public LocationSelectionView() {
10
11         super();
12
13         setAlignItems(CENTER);
14
15         final ComboBox<String> locationCombo = new ComboBox<>("Select Location");
16         locationCombo.setWidth("25%");
17         locationCombo.setItems(Arrays.asList("Chennai"));
18
19         locationCombo.addValueChangeListener(l -> {
20             UI.getCurrent().navigate(HospitalSelectionView.class);
21         });
22         add(new H1("Hospital Recommendation System"),
23             locationCombo);
24     }
25 }

```

Fig No: 6.2 Java Sample Code

REFERENCES:

- Carey RG, Seibert JH. "A patient survey system to measure quality improvement: Questionnaire reliability and validity". *Med Care*, vol.31, pp. 834–845, 1993.
- Jatulis DE, Bundeck NI, Legorreta AP. "Identifying predictors of satisfaction with access to medical care and quality of care". *Am J Med Qual*, vol.12, pp. 11-18, 1997.
- Y. Zhang, R. Padman, L. Wasserman, N. Patel, P. Teredesai, and Q. Xie, "On clinical pathway discovery from electronic health record data", *IEEE Intelligent Systems*, vol. 30, no. 1, pp. 70-75, 2015.