A PRELIMENERY REPORT ON

HEALTHCARE CHATBOT

SUBMITTED TO THE VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF TECHNOLOGY (COMPUTER ENGINEERING)

SUBMITTED BY

RAJ SHAH	**Exam Seat No. :17U193**
SHRIYA CHIVATE	**Exam Seat No. :17U544**
SUHANI CHANDOLIKAR	**Exam Seat No. :17U539**
TASNEEM CHOPDAWALA	**Exam Seat No. :17U327**

DEPARTMENT OF COMPUTER ENGINEERING

BRACT'S
VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY

SURVEY NO. 3/4, KONDHWA (BUDRUK), PUNE – 411048, MAHARASHTRA (INDIA).

BRACT'S
VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, PUNE
2020 -2021

CERTIFICATE

This is to certify that the project report entitles

" HEALTHCARE CHATBOT"

Submitted by

RAJ SHAH	**Exam Seat No. :17U193**
SHRIYA CHIVATE	**Exam Seat No. :17U544**
SUHANI CHANDOLIKAR	**Exam Seat No. :17U539**
TASNEEM CHOPDAWALA	**Exam Seat No. :17U327**

Is a bonafide student of this institute and the work has been carried out by him/her under the supervision of P**rof. Priyanka.D.More** and it is approved for the partial fulfillment of the requirement of VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, for the award of the degree of **Bachelor of Technology** (Computer Engineering).

(Prof. Priyanka.D.More)
Guide,Department of
Computer Engineering

(Dr. S. R. Sakhare)
Head,Department of
Computer Engineering

(Dr. Vivek Deshpande)
Director,
BRACT's Vishwakarma Institute of Information Technology, Pune-48

Place : Pune
Date :

CERTIFICATE FROM INDUSTRY

Proost Solutions Pvt. Ltd.
12 Ramkunj, ICS Colony, Pune - 411007
www.drprax.com

PROOST SOLUTIONS

Project Completion Certificate Date: 7th December 2020

This is to certify that **Mr. Raj Shah, Ms. Shriya Chivate, Ms. Suhani Chandolikar & Ms. Tasneem Chopdawala** of Vishwakarma Institute of Information Technology, Pune had undertaken the project "**Healthcare Chatbot**" from 15th July 2020 to 1st December 2020 in our organization. The project has been completed to our satisfaction.

We wish them all the best in their future endeavors.

With best regards,

Mr. Atul Shah
Founder
Proost Solutions Private Limited, Pune.

Proost Solutions Private Limited
CIN-U72100PN2016PTC164182
12, Ramkunj Society, ICS Colony,
Shivajinagar Pune MH 411007 IN
Phone No: +91-9823092100
www.drprax.com

DrPrax
Cheers to Good Health

ACKNOWLEDGEMENT

This project would not have been possible without the constant guidance and support of the faculty members of the Computer Engineering Department at VIIT, Pune. Especially, the support of our project guide, P**rof. Priyanka. D. More**, without whose esteemed guidance and support, this project would not have been possible. We cannot thank her enough.

We would like to thank **Proost Solutions Pvt Ltd** for providing us with this opportunity, and we would like to express our gratitude towards our industry mentors **Mr. Atul Shah, Mrs. Vinita Navadgi** and **Mr. Gururaj Padnis** for constantly guiding and motivating us.

Also, we would like to thank P**rof. Sachin Sakhare**, our Head of Department, for his encouragement.

RAJ SHAH
SHRIYA CHIVATE
SUHANI CHANDOLIKAR
TASNEEM CHOPDAWALA

ABSTRACT

A bot or chatbot is meant to make a conversation between humans and machines. The machine has been embedded with the knowledge to identify the sentences and to make a decision itself as a response to answer a question. The objective of the project is to build a Health Care System (Disease Prediction) using the chatbot concept for patients or anyone who suspects they have a particular disease. This chatbot will be built by using Google's Dialogflow for NLP, React JS for the user interface, Python for building the Machine Learning Model and Mongo DB as the database to store and retrieve information.

The connections of these layers will be established by writing a webhook in fulfilments of Dialogflow and interaction through the Web App. Many companies like Dominos, KLM Royal Dutch Airlines, Ticket master have used Dialogflow for building their interactive bots. Our end product will be a Web App that will be able to interact with the patients. Thus, the proposed system will be able to process the user input using natural language processing, match the input with the Machine Learning Model and then generate a suitable response.

TABLE OF CONTENTS

LIST OF FIGURES i
LIST OF TABLES ii

CHAPTER TITLE PAGE NO.

Sr. No.	Title of Chapter	Page No.
01	**Introduction**	1
1.1	Overview	1
1.2	Motivation	1
1.3	Problem Definition and Objectives	1
1.4	Project Scope & Limitations	2
1.5	Methodologies of Problem solving	2
02	**Literature Survey**	3
03	**Software Requirements Specification**	5
3.1	Assumptions and Dependencies	5
3.2	Functional Requirements	5
3.3	External Interface Requirements	6
3.3.1	User Interfaces	6
3.3.2	Hardware Interfaces	6
3.3.3	Software Interfaces	6
3.4	Nonfunctional Requirements	6
3.4.1	Performance Requirements	6
3.4.2	Safety Requirements	7
3.4.3	Security Requirements	7
3.4.4	Software Quality Attributes	7
3.5	System Requirements	7
3.5.1	Database Requirements	7
3.5.2	Software Requirements	8
3.5.3	Hardware Requirements	9
3.6	Analysis Models: SDLC Model to be applied	9
04	**System Design**	14
4.1	System Architecture	14
4.2	Mathematical Model	15
4.3	Data Flow Diagrams	16
4.4	Entity Relationship Diagrams	18
4.5	UML Diagrams	19
05	**Project Plan**	22
5.1	Project Estimate	22
5.1.1	Reconciled Estimates	22
5.1.2	Project Resources	23

	5.2	Risk Management		24
		5.2.1	Risk Identification	24
		5.2.2	Risk Analysis	24
		5.2.3	Overview of Risk Mitigation, Monitoring, Management	25
	5.3	Project Schedule		26
		5.3.1	Project task set	26
		5.3.2	Task Network	26
		5.3.3	Timeline Chart	27
	5.4	Team Organization		28
		5.4.1	Team structure	28
		5.4.2	Management reporting and communication	29
06		**Project Implementation**		30
	6.1	Overview of Project Modules		30
	6.2	Tools and Technologies Used		31
	6.3	Algorithm Details		32
		6.3.1	Decision Tree Algorithm	32
07		**Software Testing**		34
	7.1	Type of Testing		34
	7.2	Test cases & Test Results		35
08		**Results**		37
	8.1	Outcomes		37
	8.2	Screen Shots		37
09		**Conclusions**		44
	9.1	Conclusions		44
	9.2	Future Work		44
	9.3	Applications		44
		Appendix		45
		Appendix A		45
		Appendix B		47
		References		55

LIST OF FIGURES

FIGURE	ILLUSTRATION	PAGE NO.
1	System Architecture	14
2	Mathematical Model	15
3	DFD Level – 0	16
4	DFD Level – 1	17
5	ER Diagram	18
6	Activity Diagram	19
7	Use Case Diagram	20
8	State Machine Diagram	21
9	Task Network	27
10	Timeline Chart	28
11	Team Structure	29

LIST OF TABLES

TABLE	ILLUSTRATION	PAGE NO.
1	Functional Requirements	5
2	Risk Analysis	24
3	Test cases & Test Results	35-36

CHAPTER 1. INTRODUCTION

1.1 OVERVIEW

There are many tools related to disease prediction. But generally, there are no such tools that are used for the prediction of general diseases. So, our Disease Predictor helps for the prediction of these general diseases. It takes in input from the user such as their symptoms if they have any heart problems etc. and gives a probable disease that they may be suffering from, along with some immediate suggestions as well as which doctor, they should refer to. It also gives a list of nearby hospitals in the user's vicinity.

1.2 MOTIVATION

At present, when someone suffers from a particular disease, they have to visit a doctor, which is time-consuming and costly. If the user is out of reach of doctors and hospitals it may be difficult for the user as the disease cannot be identified. So, if the above process can be completed using an automated program which can save time as well as money, it can be very helpful to the user and can make the process easier. The aim of Disease prediction is to predict the disease of patient provides the symptoms. As there is advancements in digital technology the use of internet is at pace, people are curious towards new things. People always attempt to ask the web if any problem arises. People have access to the internet faster than hospitals and doctors. So, this system can be helpful to people to know how serious their symptoms are and which doctor they need to go to.

1.3 PROBLEM DEFINITION AND OBJECTIVES:-

The objective of the project is to build a chatbot that can predict diseases and provide information on the disease as well as nearby hospitals.

1. Provide accurate prediction of diseases.
2. Save the efforts of getting unnecessary checkups from different doctors.
3. Interactive and easy chatbot.

1.4 PROJECT SCOPE & LIMITATIONS:-

1.4.1. Scope

This project aims to provide a web platform to predict the occurrences of disease based on various symptoms. The user can type in their symptoms, answer a few questions and find the probable disease with some immediate suggestions as well as which doctor, they should refer to. They can also check nearby hospitals.

1.4.2. Limitations

1. History of diseases and genetic history is not considered.
2. Model not taught to be continuously learning.

1.5 METHODOLOGIES OF PROBLEM SOLVING

We can implement the chatbot using various approaches. Development of Machine Learning Based chatbots is becoming popular and easier for software developers using Dialogflow .Developers can use it to make interfaces (e.g.chatbots) that are capable of natural and rich interactions between users and businesses. It is powered by machine learning to perceive the intent and context of what a user says, allowing the conversational interface to produce highly efficient and accurate responses. Thus, we are implementing our chatbot using Dialogflow.

Disease Prediction has been already being implemented using different techniques like Neural Network, the Naive Bayes algorithm, etc. From the analysis, it was found that the Decision Tree is more accurate than other techniques. Kommunicate provides customized User interface for the chatbot. Kommunicate was being integrated with React js.

CHAPTER 2. LITERATURE SURVEY

K.M. Al-Aidaroos, A.A. Bakar and Z. Othman have researched the best medical diagnosis mining technique. For this author compared Naïve Bayes with five other classifiers i.e. Logistic Regression (LR), KStar (K*), Decision Tree (DT), Neural Network (NN) and a simple rule-based algorithm (ZeroR). For this, 15 real-world medical problems from the UCI machine learning repository (Asuncion and Newman, 2007) were selected for evaluating the performance of all algorithms. In the experiment, it was found that NB outperforms the other algorithms in 8 out of 15 data sets so it was concluded that the predictive accuracy results in Naïve Bayes are better than other techniques.

(Al-Aidaroos, Bakar, & Othman, 2012)

Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni have done this research paper into providing a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. The number of experiment has been conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering is not performing well.

(JyotiSoni, Ansari, Sharma, & Soni, 2011)

Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, Nicholas Christakis, Albert-Laszlo Barabasi have found that global treatment of chronic disease is neither time nor cost-efficient. So the authors conducted this research to predict future disease risk. For this CARE was used (which relies only on a patient's medical history using ICD- 9-CM codes to predict future disease risks). CARE combines collaborative filtering methods with clustering to predict each patient's greatest disease risks based on their medical history and that of similar patients. Authors have also described an Iterative version, ICARE, which incorporates ensemble concepts for improved

performance. These novel systems require no specialized information and provide predictions for medical conditions of all kinds in a single run. The impressive future disease coverage of ICARE represents more accurate early warnings for thousands of diseases, some even years in advance. Applied to full potential, the CARE framework can be used to explore a broader disease Disease Predictor 7 histories, suggest previously unconsidered concerns, and facilitating discussion about early testing and prevention.

(A.Davis, V.Chawla, Blumm, Christakis, & Barbasi, 2008)

CHAPTER 3. SOFTWARE REQUIREMENTS SPECIFICATION

3.1 ASSUMPTIONS AND DEPENDENCIES:-

The user needs to have an active internet connection to access the contents of the website. The user must have a laptop or a desktop for a mobile phone for using this web application. In case the client or user is using a laptop or desktop then he should be on an operating system that supports internet browsing. This Web App won't be violating any of the internet rules.

3.2 FUNCTIONAL REQUIREMENTS:-

FR NO	Name
FR1	New user should Register.
FR2	Allow users to login into the system if already registered.
FR3	Record the email and password of the new user.
FR4	While registering with a new email, user should get email verification which contains a verification token.
FR5	Allows the user to view Doctors and their details in nearby location once logged in.
FR6	Allows the user to have a proper conversation with the chatbot.
FR7	Record the symptoms of the user when he talks with the chatbot.
FR8	Analyze the symptoms and produce results.
FR9	Display the resulting disease and related information to the user.
FR10	Allows users to log out of the system.

Table 1 :Functional Requirements

3.3 EXTERNAL INTERFACE REQUIREMENTS:-

3.3.1 User Interface

The interface provided to the user will consist of a menu for the user to select the required category for which the function should be performed. After selecting a category, the user either acts as virtual consultation or a nearby hospital location or the user can talk with a chatbot.

3.3.2 Hardware Interface

To run the website on any device, a browser with CGI support working on any operating system (Windows, Linux or Mac). Since it is not a resource-hungry program, it will run on most of the systems without hassle. The user needs to have a laptop/desktop/Mobile Phone with an internet connection.

3.3.3 Software Interface

The system will be built using MERN Framework along with Flask which is used to handle the webhook request. Various machine learning libraries like NumPy, pandas and scikit-learn. Also, a database to store user data as per requirements. Google API interface is used to provide nearby doctors to the user. Dialogflow API is used to create the chatbot and Kommunicate is used for designing the chatbot UI.

3.4 NON-FUNCTIONAL REQUIREMENTS:-

3.4.1 Performance Requirements:

We have considered important the fact that the system must perform as what every user expects. So there are no immediate delays. In case of opening different webpages, or giving error messages, etc. there is a delay much below 2 seconds, so the throughput is considerably high. The response time is quite less. The web application gives instant results and the web pages also load fast with a considerable internet bandwidth speed. In case of high workloads to it is intended to prove great in terms of response time. The system may go into failures only when the bandwidth is very less. The response from the chatbot is quick and the doctors in a nearby location to are viewed instantly.

3.4.2 Safety Requirements:

The system will be distributed among various nodes to ensure maximum safety from threats and attacks. Secure Logins will protect user data.

3.4.3 Security Requirements:

The user in the application is provided with a username and password for authenticating before the login process. Email authentication improves the degree of safety. invalid users are blocked from logging in. Encryption and decryption facility is being provided for the passwords to increase authentication and security.

3.4.3 Software Quality Attributes:

1.**Availability**: We have been Checking that the system always has something to function and always show error messages in case of problems. So here, error messages appear when something goes wrong.

2.**Maintainability**: The diseases and their symptom data is checked and maintained.

3.**Usability**: Checking from time to time that the system is easy to handle and works most expectedly with no delays. In that case, the system program reacts accordingly.

4.**Accessibility**: Better fonts and colors are used for the webpages.

3.5 SYSTEM REQUIREMENTS:

3.5.1 Database Requirements

A classical database built-in Mongo server is needed to be integrated within the project which is secure as well as portable over the internet. Its features should be such that there is ease of access to data from remote locations as well.

3.5.2 Software Requirements:

React Js for the front end of web pages

React is a JavaScript library used to build single-page applications ie. User interfaces. We can create reusable UI components using the React Js framework.

Kommunicate for chatbot Ui:

We have used Kommunicate to design the chatbot UI. It is BOT+ HUMAN hybrid support. It helps manage customer conversations, website chat, etc.

Dialog flow for building chatbot:

Dialog flow is a service of Google Cloud Platform and it is used to build conversational chatbots. These chatbots can be customized and added to websites or mobile apps, etc. Through the use of intents and entities, a chatbot is made and integrated.

Nodejs for backend:

Node.js is a free open source server environment that runs on various platforms.It can generate the dynamic page content. also, node js can create, open, read, write, delete, and close files on the server. NodeJS is a cross-platform runtime environment highly used for developing server-side and networking applications.

Python(flask) for connecting webhook:

Flask is a web framework that will provide us with tools, libraries and technologies through which we can build a web application. It is written in Python. It is a lighter framework and easier to use as compared to frameworks such as Django.

Webhook:

Webhook is a web call back for an app to provide other applications with real-time information. Whenever some event happens on the trigger application, data about that event is serialized and send to a webhook URL from the active application where some other task will be performed.

MongoDB as Database:

MongoDB is an object-oriented database that is based on the NoSQL document store model It is simple and Dynamic to use. It provides a better and reliable platform for data storage. Data objects in mongo dB are stored as separate documents and

documents and inside collections. The Traditional Data Storage system of rows and columns is eliminated here.

3.5.2 Hardware Requirements:

A Desktop/ Laptop with an internet connection and an operating system that supports web browsing.

3.6 ANALYSIS MODEL:

Classical methods of software development have many disadvantages:

1.Huge effort during the planning phase

2.Poor requirements conversion in a rapidly changing environment

3.Treatment of staff as a factor of production

New methods: Agile Software Development

SCRUM is an agile, lightweight process for managing and controlling software and product development in rapidly changing environments.

a. Iterative, incremental process

b. Team-based approach

c. developing systems/ products with rapidly changing requirements

d. Controls the confusion of conflicting interests and needs

e. Improve communication and maximize cooperation

f. Protecting the team from disruptions

g. A way to maximize productivity

Components of Scrum

a. Scrum Roles

b. Process

c. Scrum Artifacts

Scrum Master

1. Represents management to the project

2. Typically filled by a Project Manager or Team Leader

3. Responsible for enacting Scrum values and practices
4. The main job is to remove impediments

The Scrum Team

1. Typically 5-10 people
2. Cross-functional (QA, Programmers, UI Designers, etc.)
3. Members should be full-time
4. Team is self-organizing
5. Membership can change only between sprints

Product Owner

1. Acts like one voice (in any case)
2. Knows what needs to be built and in what sequence this should be done
3. Typically a product manager

The Process

1. Sprint Planning Meeting
2. Sprint
3. Daily Scrum
4. Sprint Review Meeting

Sprint Planning Meeting

1. A collaborative meeting at the beginning of each Sprint between the Product Owner,the Scrum Master and the Team
2. Takes 8 hours and consists of 2 parts ("before lunch and after lunch")

Parts of Sprint Planning Meeting

1. 1st Part:
    a. Creating Product Backlog
    b. Determining the Sprint Goal.
    c. Participants: Product Owner, Scrum Master, Scrum Team

2. 2nd Part:
    a. Participants: Scrum Master, Scrum Team
    b. Creating Sprint Backlog

Pre-Project/Kickoff Meeting

1. A special form of Sprint Planning Meeting
2. Meeting before the beginning of the Project

Sprint

1. A month-long iteration, during which is incremented a product functionality
2. No outside influence can interference with the Scrum team during the Sprint
3. Each Sprint begins with the Daily Scrum Meeting

Daily Scrum

1. Is a short (15 minutes long) meeting, which is held every day before the Team starts working.
2. Participants: Scrum Master (which is the chairman), Scrum Team
3. Every Team member should answer 3 questions

Questions

1. What progress is thereafter the last Scrum?
2. What are you doing until subsequent Scrum?
3. What is stopping you from getting on with the work?

Sprint Review Meeting

1. Is held at the end of each Sprint
2. Business functionality which was created during the Sprint is demonstrated to the Product Owner
3. Informal, should not distract Team members from doing their work

Scrum Artifacts

1. Product Backlog
2. Sprint Backlog
3. Burndown Charts

Product Backlog

Requirements for a system are usually expressed as a list of Backlog Items according to priority.

1. Is managed and owned by a Product Owner
2. Spreadsheet (typically)
3. Usually is created during the Sprint Planning Meeting
4. Can be changed and re-prioritized before each Planning Meeting

Estimation of Product Backlog Items

1. Establishes team's velocity (how many efforts a Team can handle in one Sprint)
2. Determining units of complexity.
    a. Size-category ("T-Shirt size")
    b. Story points
    c. Workdays/work hours
3. Methods of estimation:
    a. Expert Review
    b. Creating a Work Breakdown Structure (WBS)

Sprint Backlog

1. A subclass of Product Backlog Items, which is the work of Sprint.
2. Is created ONLY by Team members
3. Each Item has its status
4. Should be updated every day
5. No more than 300 tasks in the list
6. If a task requires more than 16 hours, it should be broken down
7. The team can add or subtract items from the list. The Product Owner isn't allowed to try and do it.

Burndown Chartsa

1. Are used to represent “work done”.
2. Are wonderful Information Radiators
3. 3 Types:
    a. Sprint Burndown Chart (progress of the Sprint)
    b. Release Burndown Chart (progress of release)
    c. Product Burndown chart (progress of the Product)

CHAPTER 4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

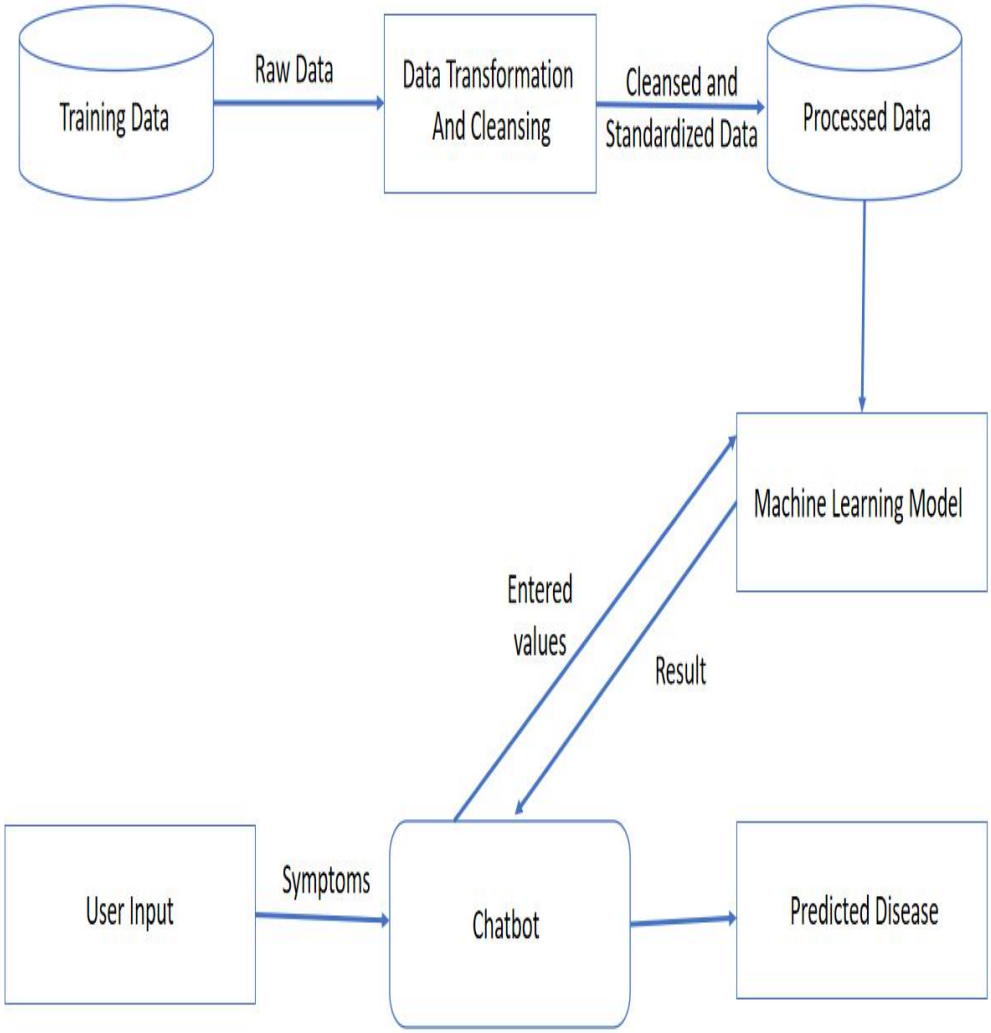


Figure 1 :System Architecture

4.2 MATHEMATICAL MODEL

Mathematical model is a description of a system using mathematical concepts and language. The process of developing a mathematical model is termed mathematical modeling. A model may help to explain a system and to study the effects of different components, and to make predictions about behaviour.

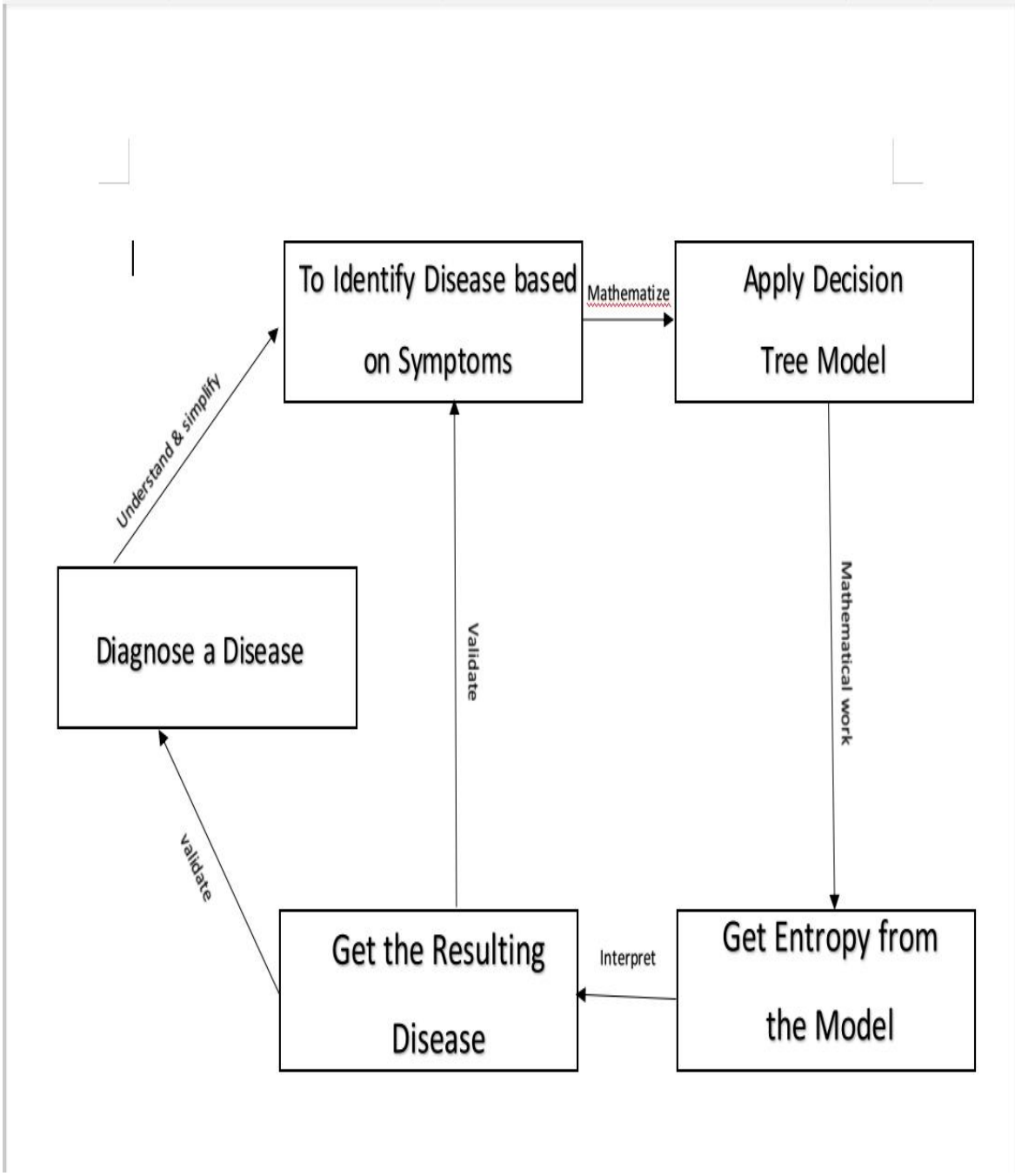


Figure 2: Mathematical Model

4.3 DATA FLOW DIAGRAMS

The objective of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communication tool between a system analyst and any person who plays a part in the order that acts as a starting point for redesigning a system.

DFD Level-0:

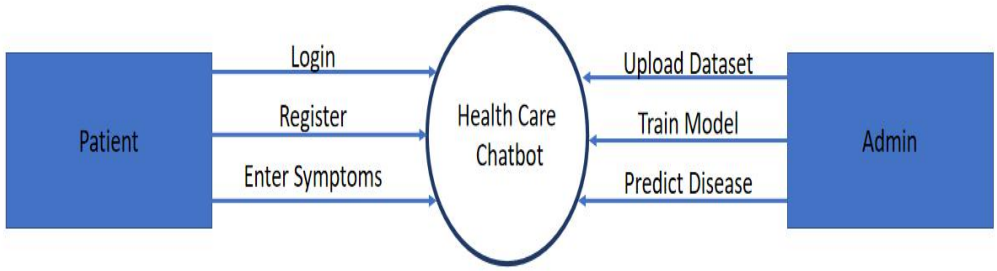


Figure 3: DFD Level - 0

DFD Level-1:

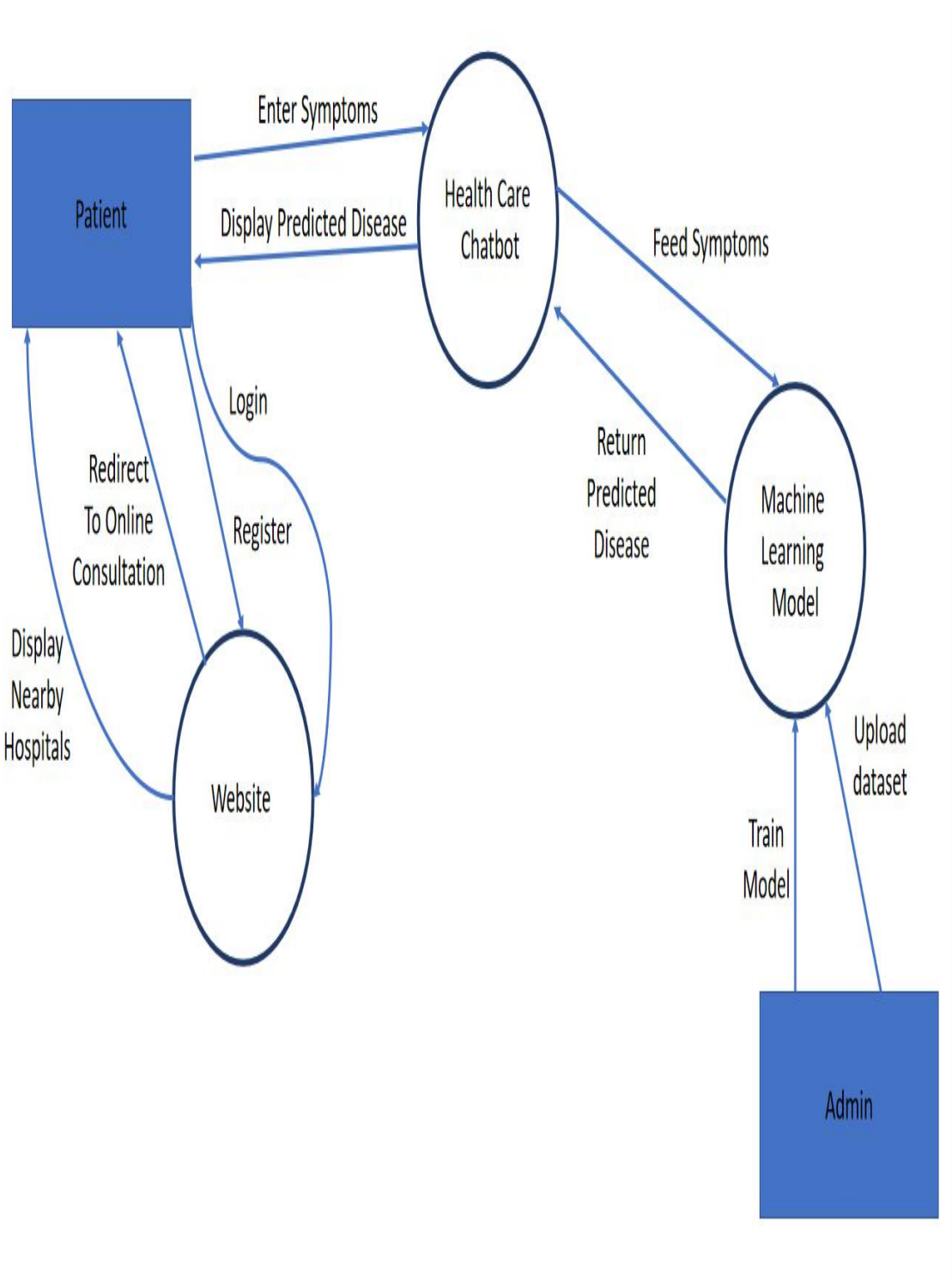


Figure 4: DFD Level - 1

4.4 ENTITY RELATIONSHIP DIAGRAMS

An Entity–relationship model (ER model) describes the structure of a database with the help of a diagram, which is known as Entity Relationship Diagram (ER Diagram). An ER model is a design or blueprint of a database that can later be implemented as a database. The main components of E-R model are: entity set and relationship set.

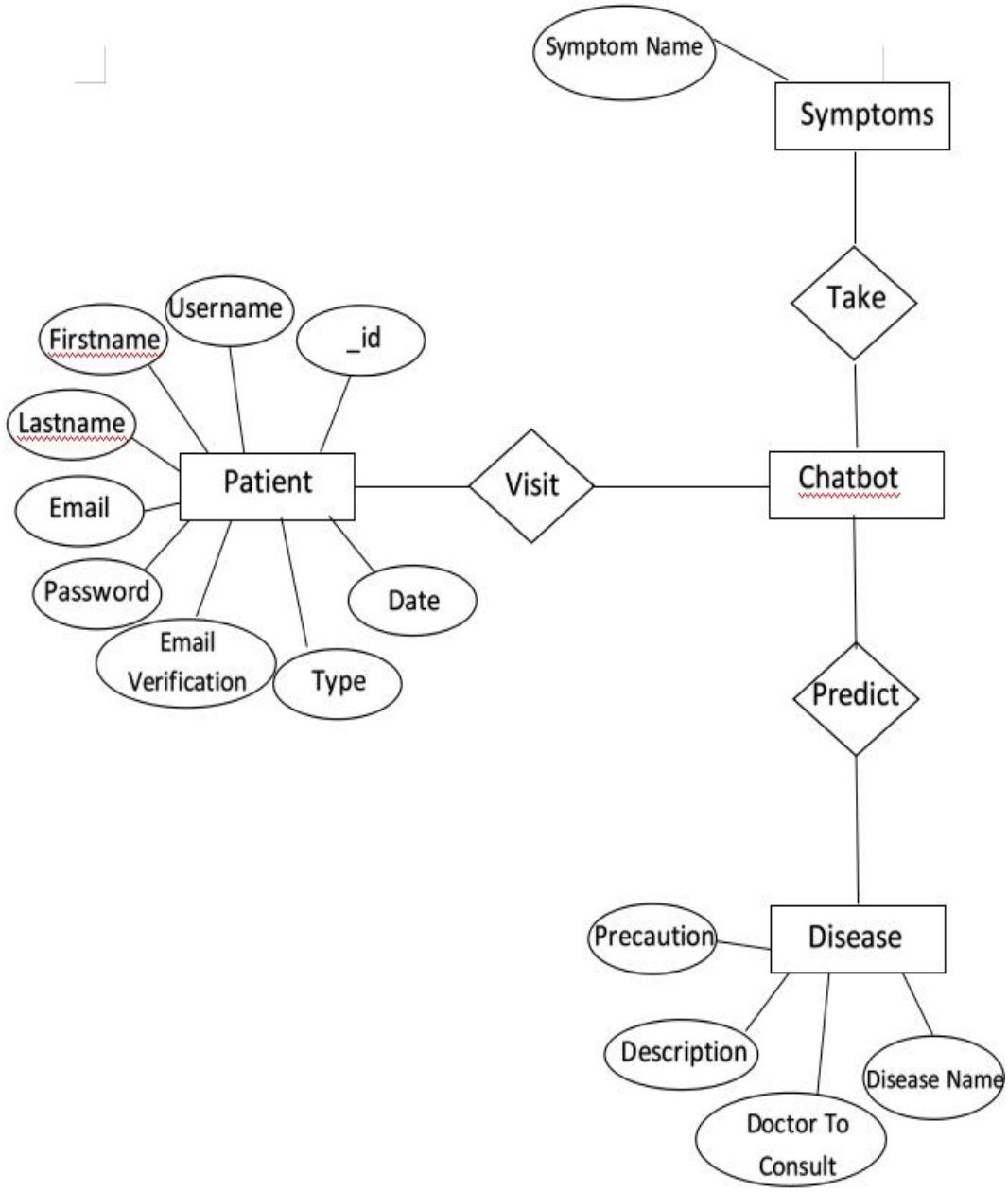


Figure 5 : ER Diagram

4.5 UML DIAGRAMS

4.5.1 Activity Diagram

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system.It activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed. We can depict both sequential processing and concurrent processing of activities using an activity diagram. They are used in business and process modelling where their primary use is to depict the dynamic aspects of a system.

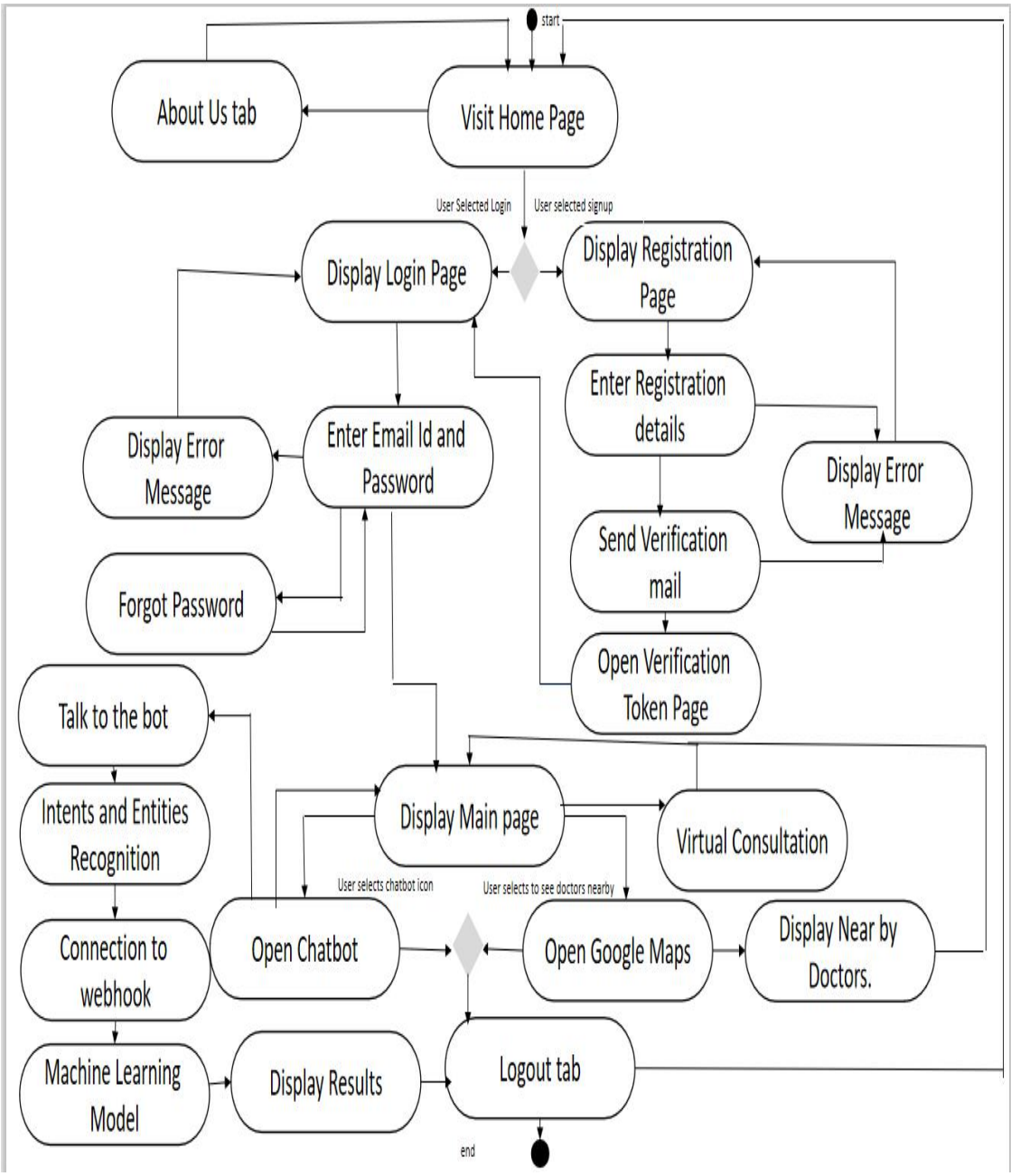


Figure 6: Activity Diagram

4.5.2 Use Case Diagram

Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified.

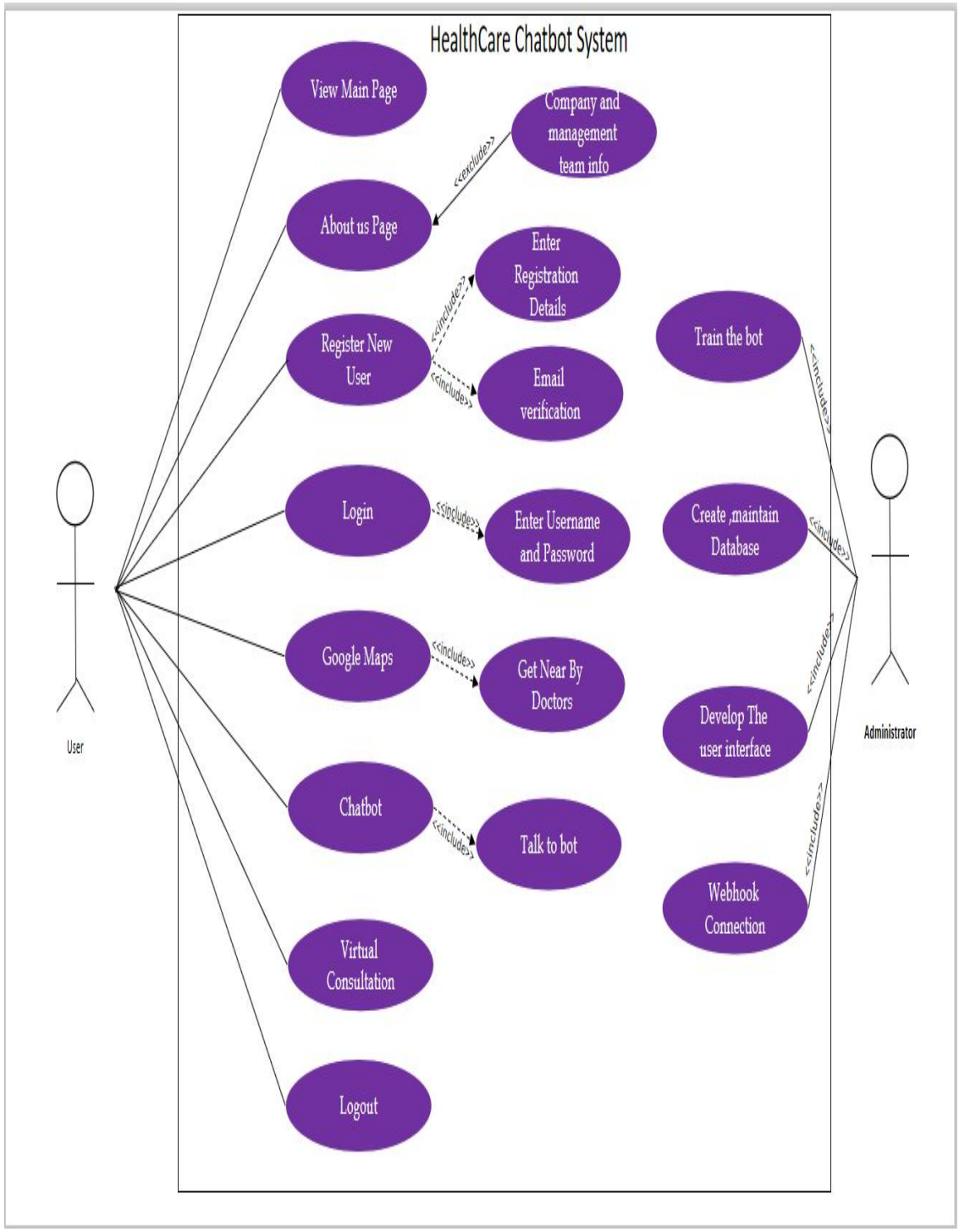


Figure 7:Use Case Diagram

4.5.3 State Machine Diagram

State machine diagram typically are used to describe state-dependent behavior for an object. An object responds differently to the same event depending on what state it is in. State machine diagrams are usually applied to objects but can be applied to any element that has behavior to other entities such as: actors, use cases, methods, subsystems systems and etc.

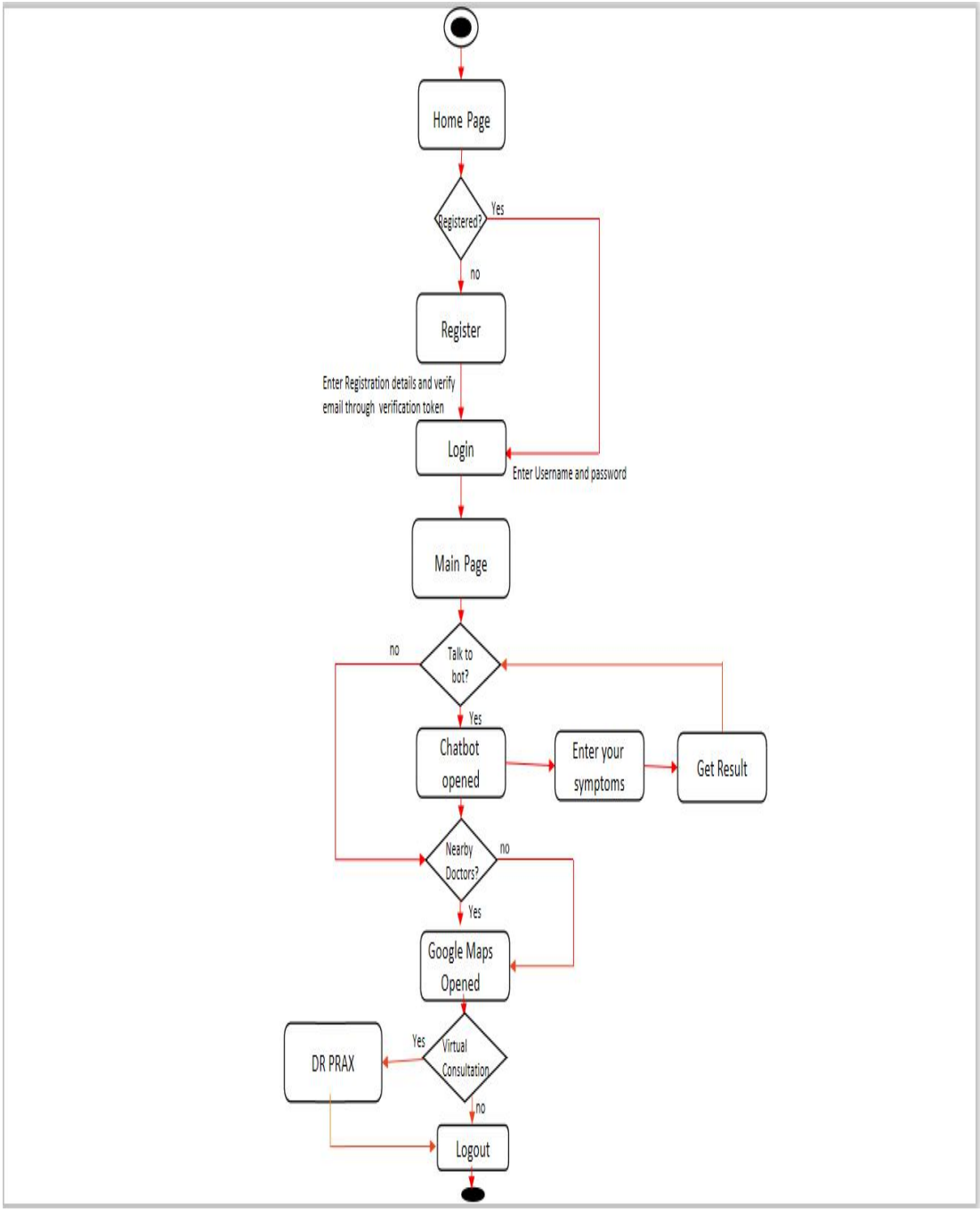


Figure 8:State Machine Diagram

CHAPTER 5. PROJECT PLAN

5.1 PROJECT ESTIMATE

5.1.1 Reconciled Estimates

Cost Estimate

In the real world, there are 3 types of models included in estimation using the COCOMO Model. They can be described as follows:

1.Organic: Small team size, 2-50 KLOC, stable and little innovation required

2.Semi Detached: 50-300 KLOC, medium-sized, average abilities, medium time-constraints

3.Embedded: > 300 KLOC, large project team, complex, innovative, severe constraints

The proposed system will utilize the Organic Model for Cost Estimation.

Step 1:

$$E=a*KLOC^{b}$$

Where, E = Effort in staff-months

a, b = Coefficients to be determined

KLOC = Kilo lines of Code

Step 2:

For organic model a = 2.4 and b = 1.05

Assuming KLOC = 6

$$\mathbf{E=2.4*(4)^{1.05}=10.3\ Staff\text{-}Months}$$

Step 3:

Project Duration

Where TDEV = Time for Development

$$TDEV=c*(E)^{d}$$

For Organic c = 2.5 and d = 0.38

$$\mathbf{TDEV=2.5*(10.3)^{0.38}=6.0\ Months}$$

Step 4:

Average Staff Size=E/TDEV=10.3/6 = 2 to 3 Staff

Step 5:

Productivity=P=KLOC/E=4000/10.3=388.35 Loc/staff-month

Step 6:

Assuming cost per month = ₹ 5000 /-

Total cost=E*cost per Month=10.3*5000=Rs 51500 /-

Time Estimates

Project Duration

$$\mathbf{TDEV=c*(E)^d}$$

Where TDEV = Time for Development

For Organic c = 2.5 and d = 0.38

$$\mathbf{TDEV=2.5*(10.3)^{0.38}=6.0\ Months}$$

5.1.2 Project Resources

There are certain external requirements for the project to successfully run lifetime like Google Api ,Kommunicate , MailGun.But ,currently we have used free version of these requirements.

-

5.2 RISK MANAGEMENT

5.2.1 Risk Identification

Projects undergo risks and these need to be identified early to avoid difficulties in the project cycle. Risk identification is very important when it comes to Risk Management Planning. It involves determining which risks might affect our project and it's quality and functionality and documents their characteristics. Our Project Data should be thoroughly analyzed along with its implementation to understand the potential risks in our project. This will in turn help in proper utilization of our resources and obviously improve the reliability.

After doing an extensive search concerning all the modules from our project, the following risks are identified:

1.Verification email can be received a few seconds late.

2.Not all spelling errors can be identified by the chatbot

3.User may input improper or incorrect symptoms.

5.2.2 Risk Analysis

After successful identification of these potential risks, we analyzed them carefully and calculated their probability of occurrence and severity so that we could categorize them into low, medium, high. This helped us in identifying how these risks could be mitigated.

Risk No.	**Potential Risks**	**Probability of Occurrence**	**Severity**	**Impact**
1.	Late Verification mail.	high	medium	Increased Response Time
2.	Not all spelling errors identified.	medium	low	Decreased accuracy
3.	Incorrect stated symptoms	low	high	Incorrect analysis

Table 2:Risk Analysis

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Risk Mitigation:

Risk mitigation actually refers to ways in which we can avoid the occurrence of any potential risk. The project team can together look into it and find ways to mitigate these risks.

Risk 1: Check with the concerned software authorities for the reasons to same.

Risk 2: We can train the chatbot for more accuracy. Can take feedback from people regarding the same.

Risk 3: Make the user understand the importance to have a clear vision of his symptoms

Risk Monitoring:

Risk monitoring refers to the process of actually checking if these risks occur or not and that the steps which are defined to avoid these risks are properly executed or not. We can have more users involved which could tell us if the risks still prevail or have chances to occur shortly after they are managed. And if so, what is their severity to avoid any reverse impact on the project.

Risk 1: It may happen that after talking to the concerned authorities this problem still continues. We can check the frequency of occurrence and compare it to the previous frequency to check if there is some improvement.

Risk 2: We can have a regular check to ensure that whenever a spell check is identified, it is immediately included into the training phase so that next time it could be properly identified. We can talk with more people to get an idea about the errors which people can normally make while using some particular word or group of words.

Risk 3 : We can have a manual talk to some people about their symptoms and then ask them to use the system so that we can compare if user actually enters the correct symptoms or not. We can make them understand the importance of adding correct symptoms.

Risk Management:

Project Risk management mainly refers to practices we follow when predicted risks have occurred.

Risk 1: We can use any alternative service to have the verification mails in time.

Risk 2: Understand the common spelling errors and keep including them in the database.

Risk 3: The user should talk with a doctor related to his symptoms etc, to have a better understanding of the disease suffering after taking with the chatbot.

5.3 PROJECT SCHEDULE

5.3.1 Project task set

1. Requirement Gathering and Analysis
2. Planning
3. Design and Modeling
4. Coding
5. Testing
6. Delivery
7. Documentation
8. Operations
9. Meetings
10. Scheduling and Formal Technical Reviews

5.3.2 Task Network

Task network (activity network) is a graphic representation of the task flow for a project. The task network depicts major software engineering tasks.

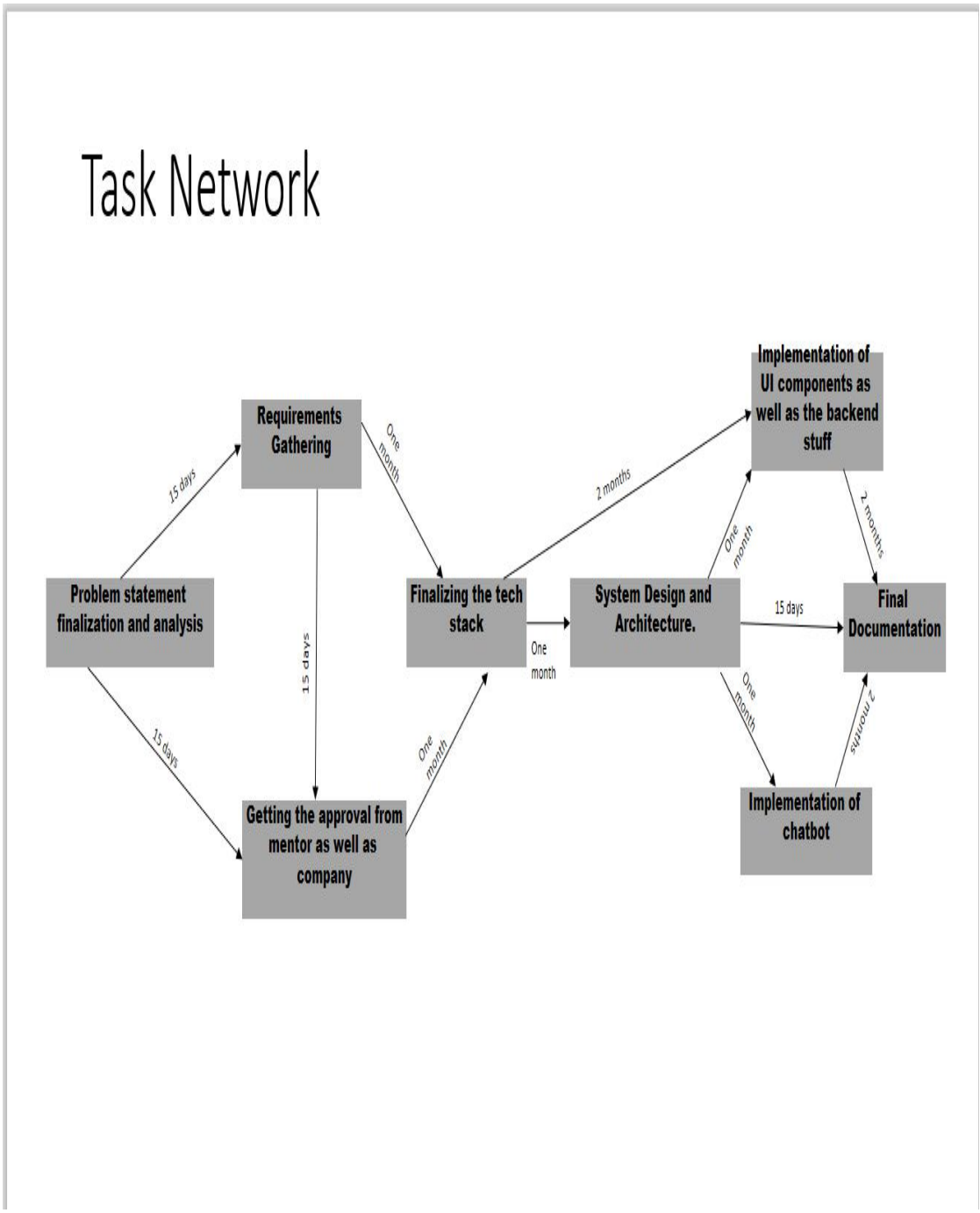


Figure 9: Task Network

5.3.2 Timeline Chart

A timeline chart is an effective way to visualize a process using chronological order. Since details are displayed graphically, important points in time can be easy seen and understood. Often used for managing a project's schedule, timeline charts function as a sort of calendar of events within a specific period of time

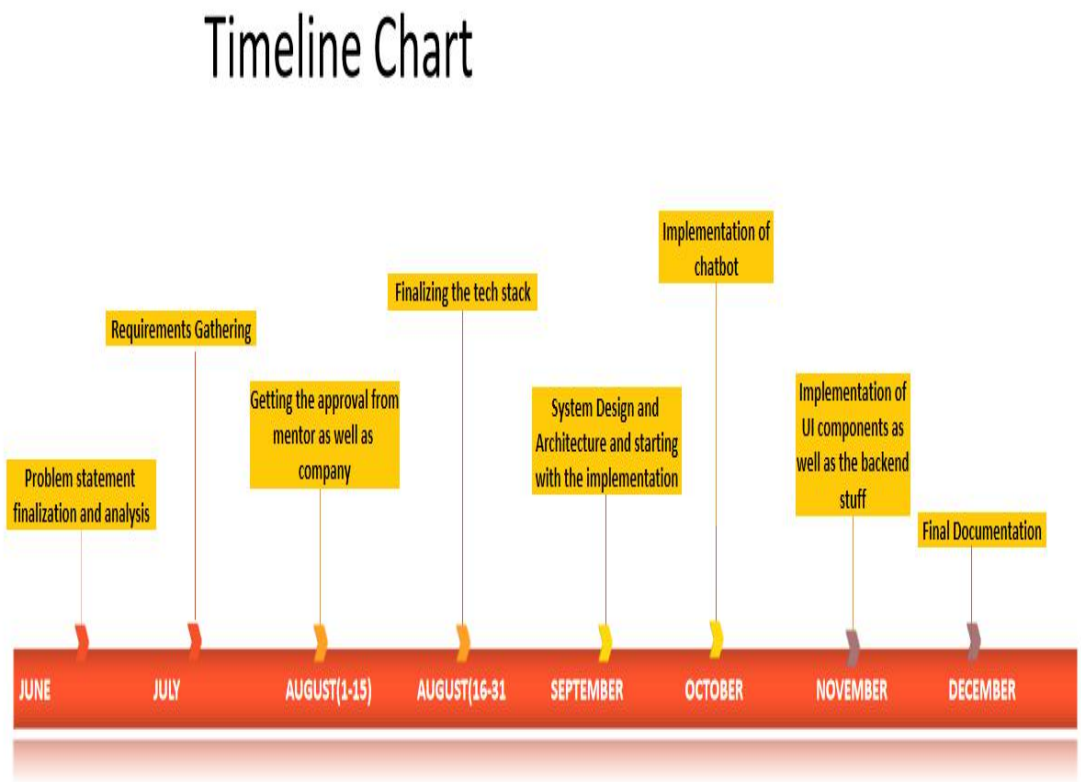


Figure 10: Timeline Chart

5.4 TEAM ORGANIZATION

5.4.1 Team Structure

Team structure is an integral part of the teamwork process.

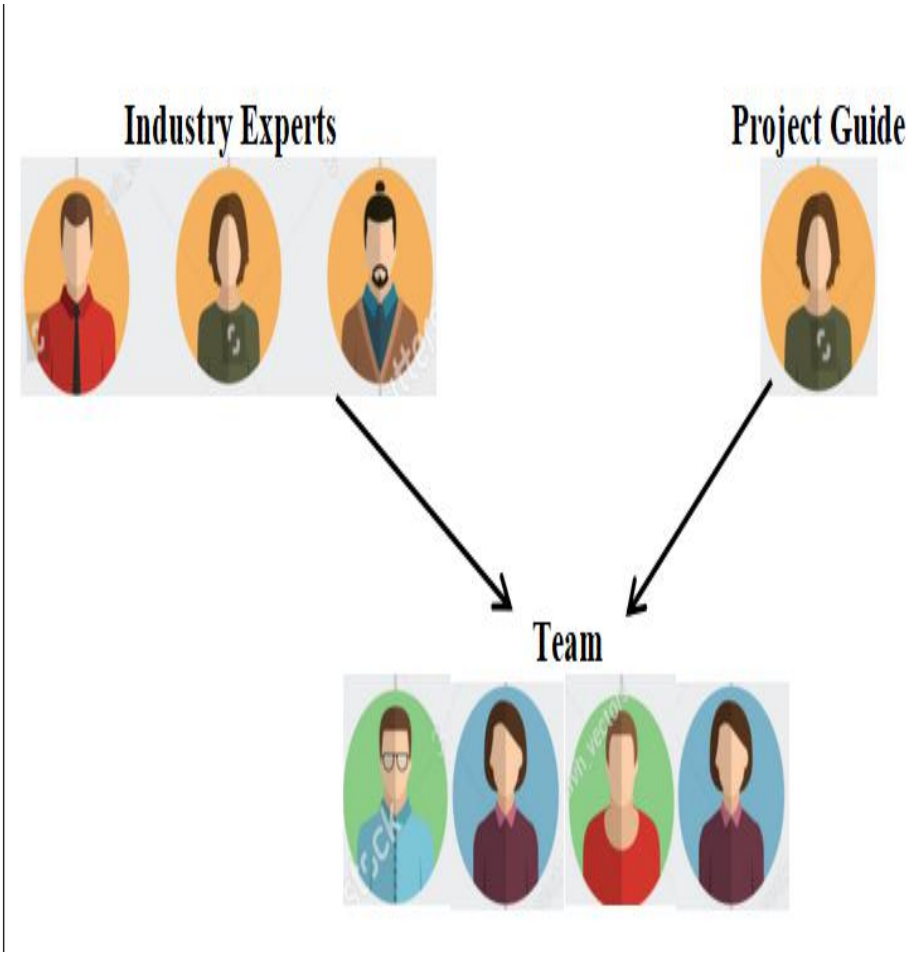


Figure 11: Team Structure

Industry Experts:-

1. Vinita Navadgi, Founder
2. Atul Shah, Founder
3. Gururaj Phadnis, VP Engineering

Project Guide:-

1. Prof.Priyanka.D.More

Team Members:-

1. Raj Shah
2. Shriya Chivate
3. Tasneem Chopdawala
4. Suhani Chandolikar

5.4.2 Management reporting and communication

To sustain improvements and creativity in the idea of the project, communication with stakeholders was maintained throughout the lifecycle of the project development. Timely reporting and formal reviews regarding the project guide were submitted to the project guide, Prof. P.D.More. Interactions with domain experts were conducted to further refine the idea and scope of the project.

CHAPTER 6. PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

We have designed a linear based retrieval based chatbot which acts as a tool of communication to predict disease based on the input, the symptoms given by the user. it has 2 phases of conversation i.e extraction of symptoms and diagnosis of the symptoms. The input given to the chatbot is in the form of unstructured text and thus we have used natural language understanding of Dialogflow to extract symptoms from it. These symptom vector arrays are now passed to the machine learning model, the model shortlists several diseases based on the input. Here one of the models used is a decision tree whose root node returns the most likely disease. Another 3 model used is Gaussian Naive Bayes, Support Vector Machine, and K Nearest Neighbour Classifier to match which model finds the best response based on the highest accuracy. to get a better understanding of the disease the chatbot responds with several questions and thus finally the most likely disease is predicted based on the number of symptoms matched. The front end of the system is designed using React js. A chatbot is being customized using Kommunicate.

1 Symptom Extraction

Symptom extraction is performed to identify symptoms from the user's text input. A results of misspellings may lead to no or wrong symptoms mapping and may not always map to the database's symptoms. To handle this, a variety of symptoms are extracted and compared with the symptoms in the database using its synonyms. which are defined in entities in Dialogflow.

2 Diagnosis of Symptoms

Various machine learning models were trained on the disease-symptom database. Once symptoms are extracted from the user, the decision Tree model was applied to the set of symptoms, and the predicted disease is specified to the end-user by the chatbot.

Database

The database is required for the security as well as the integrity of the project. The machine learning module requires input data which is fed from the database. All the results obtained from the machine learning module extract final data from the database. It provides a reliable groundwork for machine learning to train and store important aspects of the data. The database consists of three different master tables of all categories which in turn are connected to their respective attributes.

User Interface

The user interface is the only module through which the users get connected to the system. it's a React-based user interface. it is integrated into the system through the Node js and Python (Flask)server.

6.2 TOOLS AND TECHNOLOGIES USED

Tools and technologies used for Web Development:

1. Visual Studio Code
2. React js
3. Node js
4. Kommunicate

Tools and technologies used for Machine Learning:

1.Dialogflow
2. Python 3.8
3. Scikit-learn
4. Pandas
5. Decision Tress Library
6. Ngrok

Database : MongoDB

API : Google API, MailGun

6.3 ALGORITHM DETAILS

6.3.1 Decision Tree Algorithm

Decision Tree comes under Supervised learning technique which is able to be used for both classification and Regression problems, but mostly it's preferred for solving Classification problems. it's a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent main decision rules and every leaf node represents the result. within a Decision tree, there are two nodes, which are the principal decision Node and Leaf Node. Decision nodes are wont to make any decision and have multiple branches, whereas Leaf nodes are the output of these decisions and don't contain any more branches.

The decisions or the test are performed based on features of the given dataset. it's a graphical representation for getting all the possible solutions to a problem/decision subjected to given conditions. it's called a decision tree because almost sort of a tree, it starts with the decision node, which expands on further branches and constructs a tree-like structure. to make a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. a decision tree simply asks a problem and suggests the answer (Yes/No), it further split the tree into subtrees.

Initially, we start by asking the patients a set of questions like what his/her is undergoing. After that, we ask 3 fixed questions to have a better understanding of the patients. Symptoms extraction takes place from Dialogflow, Webhook and finally the trained Decision Tree Model helps to predict the disease, it’s description, precautions and doctor the patients may consult.

CHAPTER 7. SOFTWARE TESTING

7.1 TYPES OF TESTING

This document is a high-level overview defining our testing strategy for web application. The tests were done were unit tests and validation tests, where features created were unit-tested and also tested for valid inputs. Apart from this, integration tests were also done to ensure all the systems were completely integrated.

Unit Testing

The entire project consists of many small modules where each one needs to be checked and tested if working as per design and requirement. Thus, unit testing was required. We tested each piece of code individually so that in the future, we could find fewer bugs.

Integration Testing

Every module was tested as and when integration was performed. After the integration of each module into the main project, the altogether performance should meet expectations, hence this testing was carried out. We identified issues such as data type inconsistency while passing it through different modules that were resolved after the testing. Major bugfixes took place during this phase of testing.

Compatibility Testing

As our system is a Web Application, browser compatibility testing was necessary. We as developers ensured that this web application runs in every browser regardless of the operating system including various versions of the browser. Rigorous testing of the application was performed in all the browsers on Windows, Linux.

Back–end Testing

To check if the given input from GUI was appropriately stored in the database, this testing was carried out. Table structure, schema, stored procedure, data structure were the parameters considered while carrying out this testing. Few queries were given to authenticate the entire environment.

System Testing

The entire system was tested as per the requirements thoroughly. Overall requirement specifications were taken into consideration while performing this testing.

Alpha Testing

The main aim of this testing was to identify all the defects before releasing it to the end-users. The development team of this project was responsible to do this testing. All bugs including font issues in the User Interface along with relevancy of the output displayed were checked.

Performance Testing

The overall performance including stress and load on the server using different platforms was done. Many requests were sent to the server together to check the amount of load it is capable of handling.

7.2 TEST CASES & TEST RESULTS

TestID	Test Case Condition	Result
1	Login with correct credentials	Should opens the Main interface.
2	Login with incorrect credentials	Shows message “User Not found”.
3	New user and correct credentials	Verification Mail followed by the Main interface.
4	New user and Incorrect credentials.	Verification Mail followed by “Invalid Password” or ” User Not Found” message.
5	User Clicks Nearby Hospital Button	Provides Hospitals nearby the user.
6	User Clicks Online Consultation Button	Opens the Drprax Website.
7	User clicks on log out	Logout from the system.

8	User Clicks on Chatbot Icon	Open the Chatbot Interface.
9	The test installs required software and checks the module integration	A complete Website.
10	Provide Symptoms	Ask the next question in order to understand user.
11	Username,symptoms not understood	Ask the same question again from chat interface.

Table 3:Test Cases & Test Results

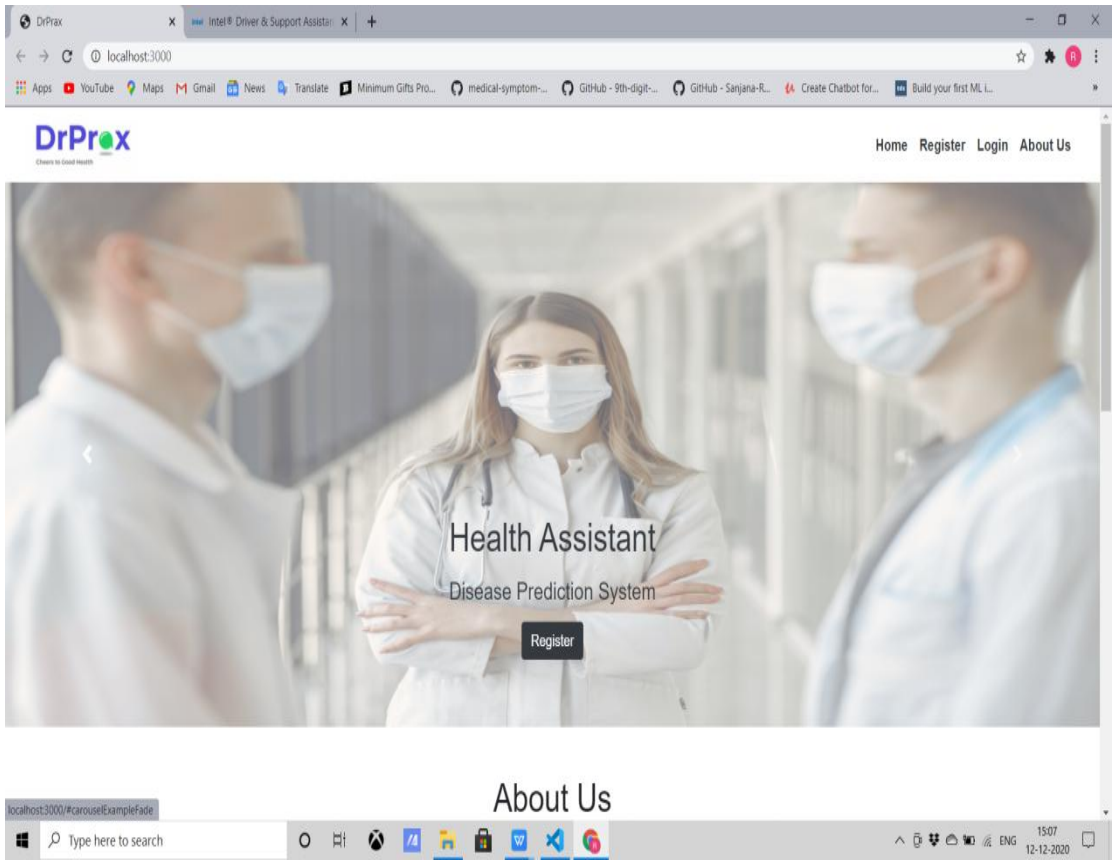
CHAPTER 8. RESULTS

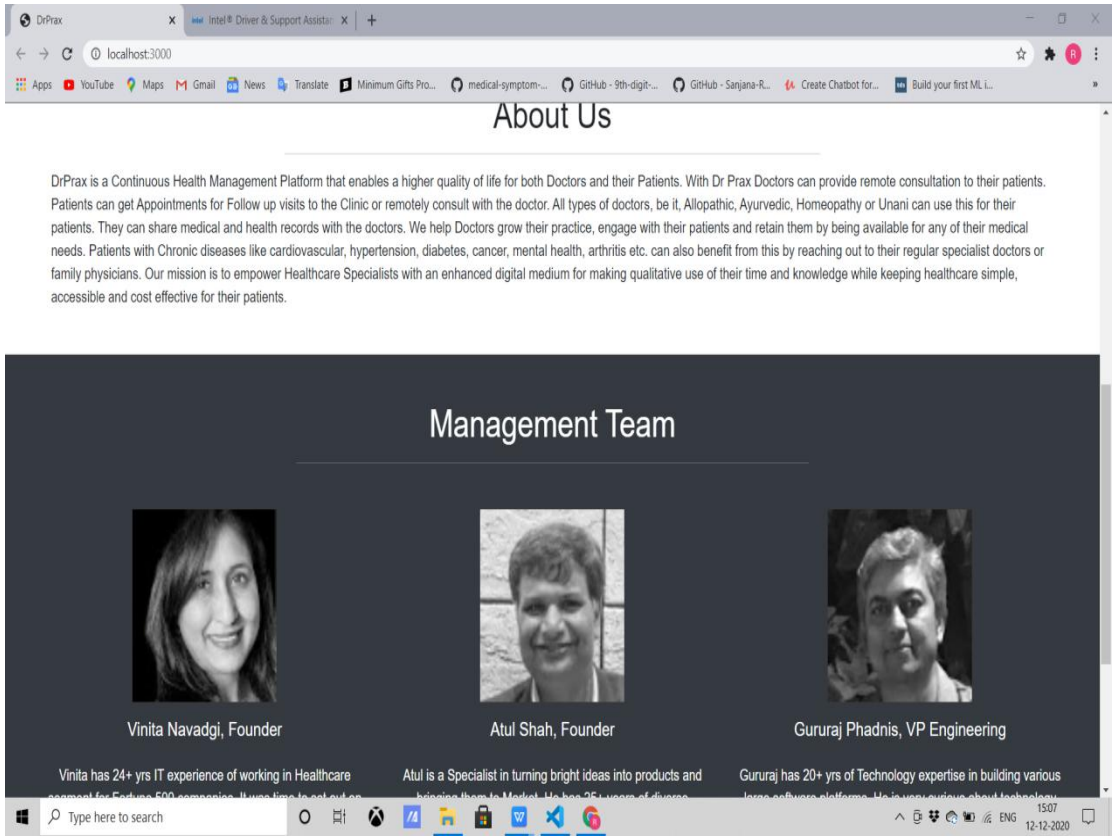
8.1 OUTCOMES

The user will have a text to a text conversation with the chatbot and the chatbot will identify the symptoms the user is undergoing through. The user needed to enter his/her symptom and answer a few fixed questions to help the chatbot identify the disease. Once the symptoms are entered the Machine Learning algorithm determines the most probable disease, its description, precautions and the doctor you may consult. The application also provides another feature of a nearby hospital and online consultation.

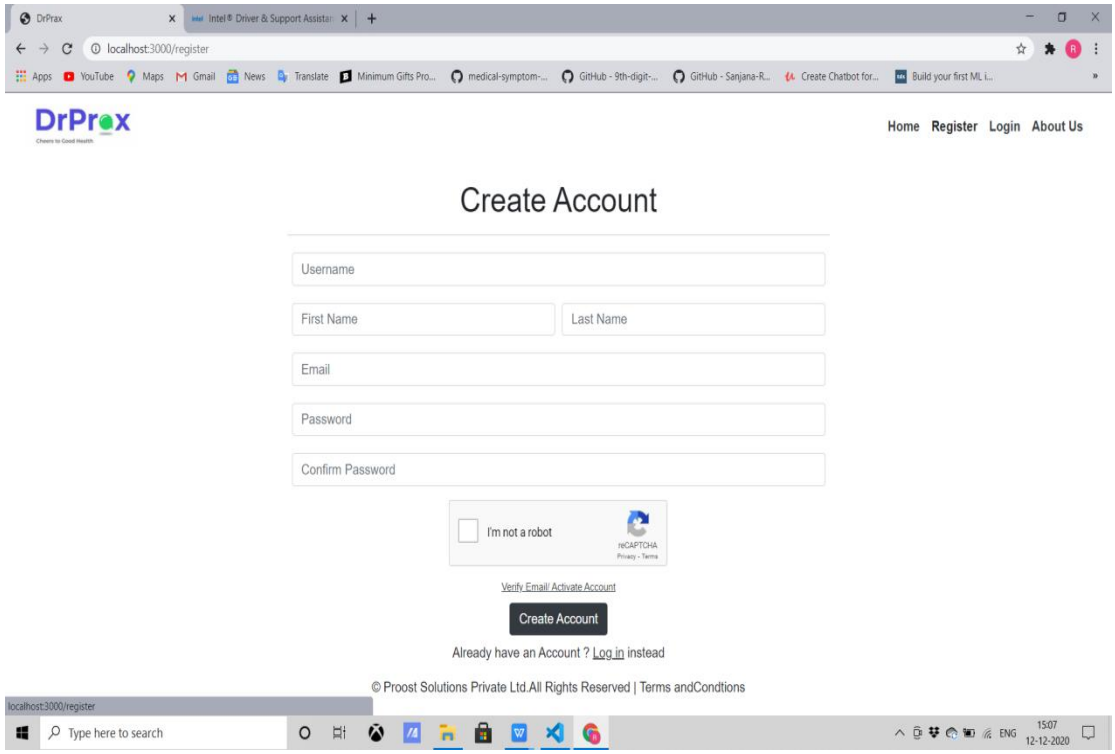
8.2 SCREEN SHOTS

Default Page:

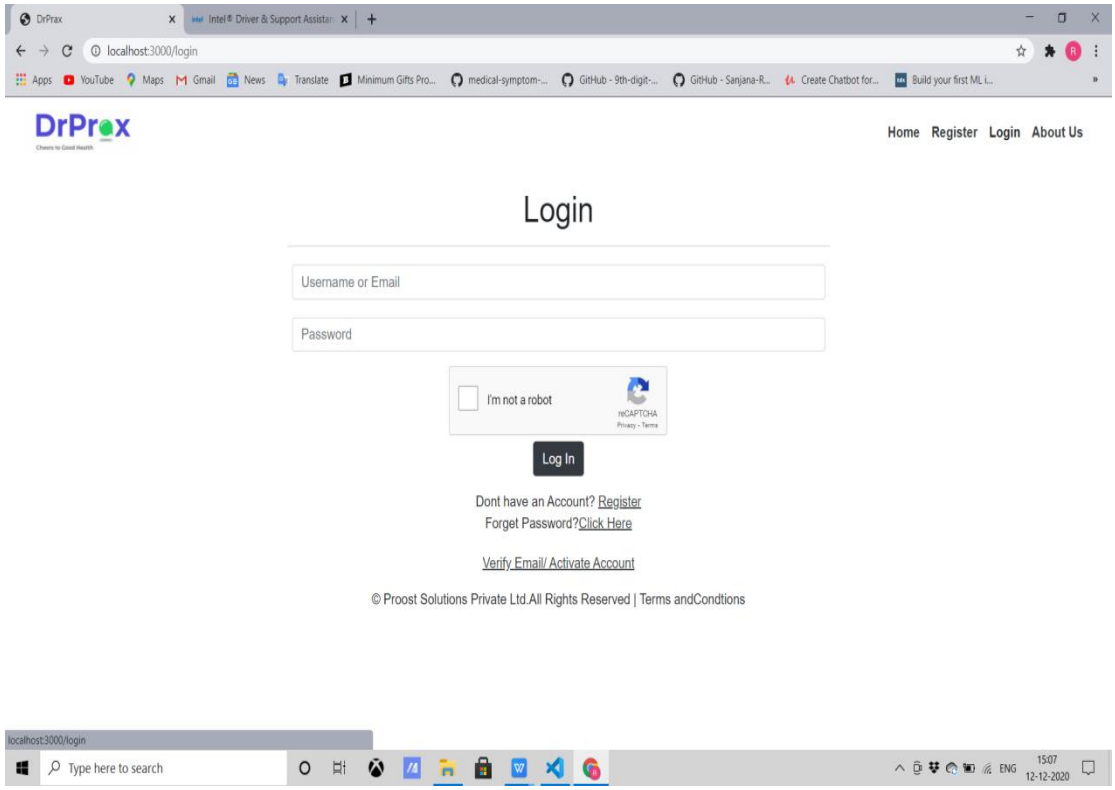




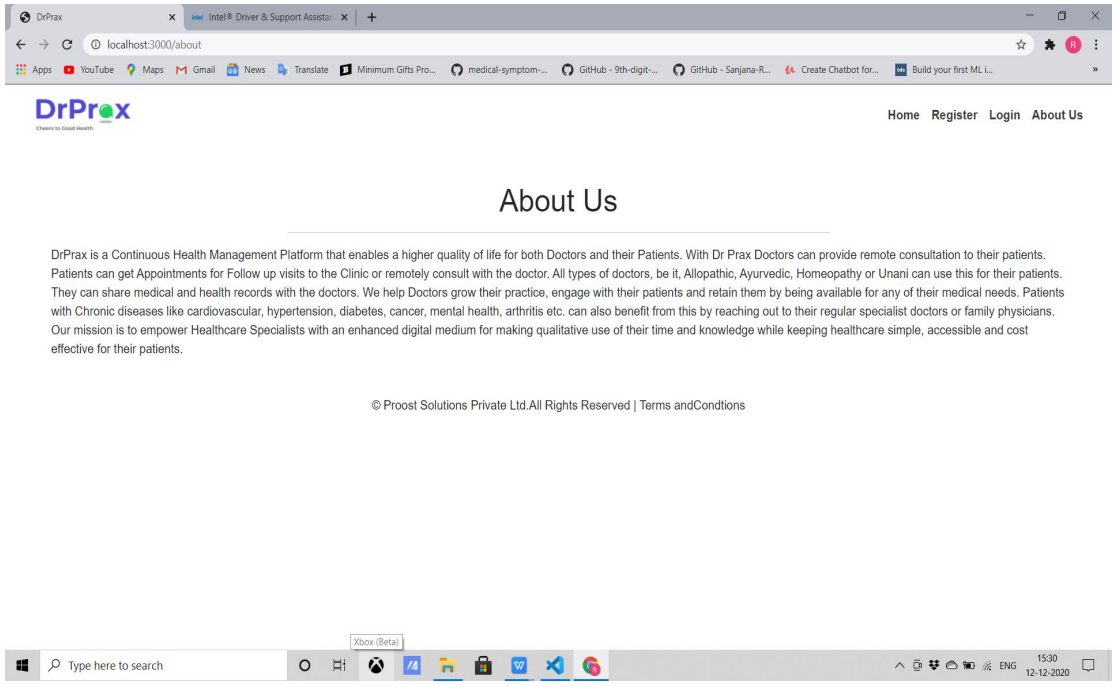
Register Page:



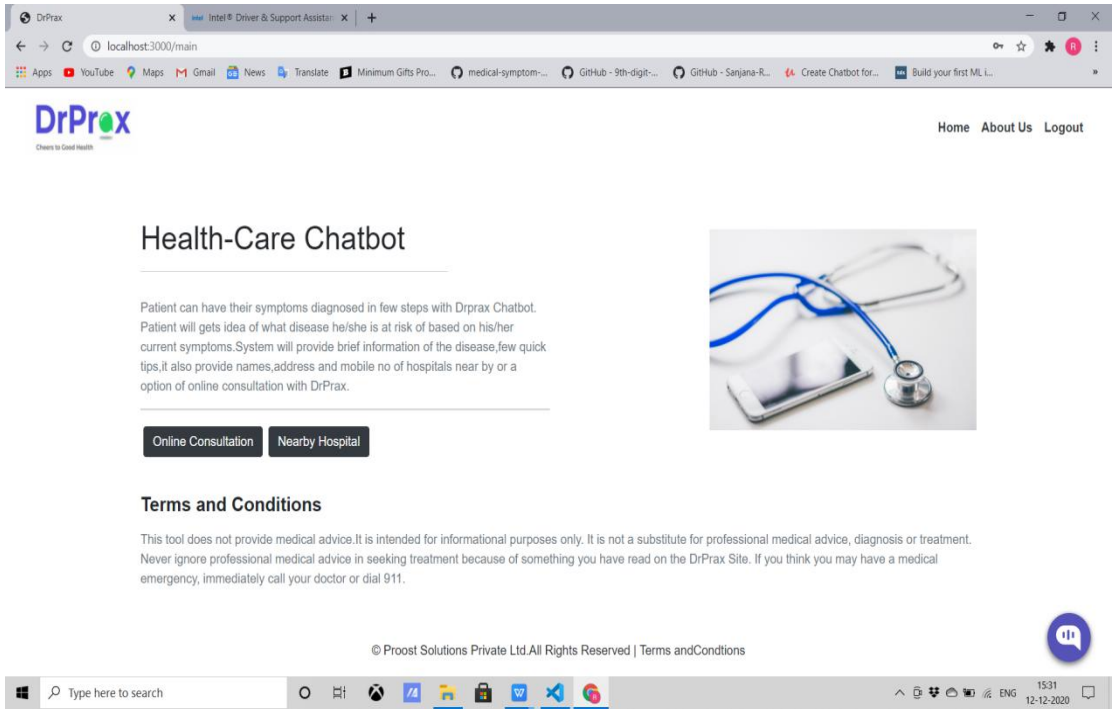
Login Page:



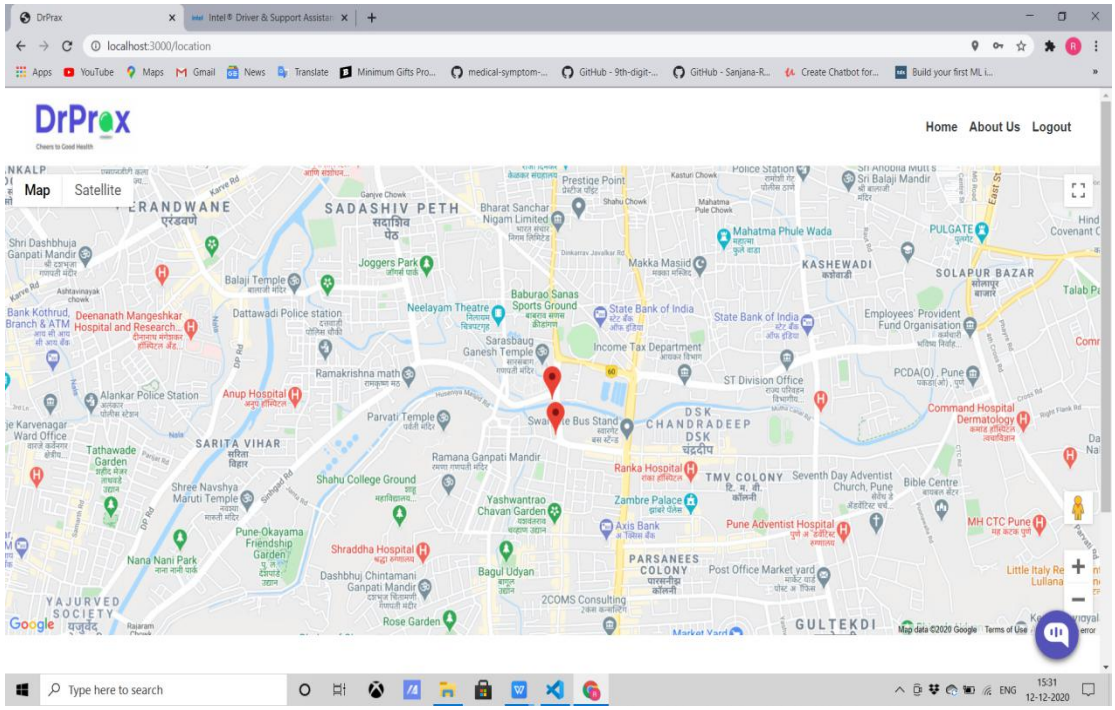
AboutUs Page:



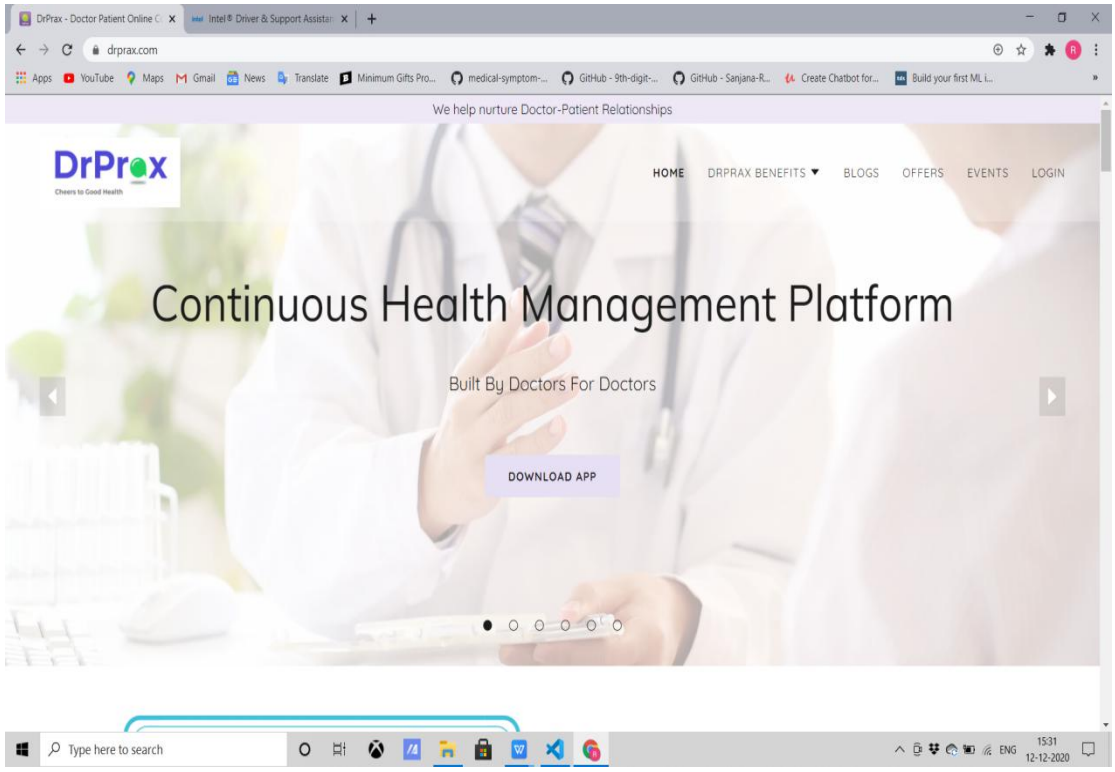
Main Page:



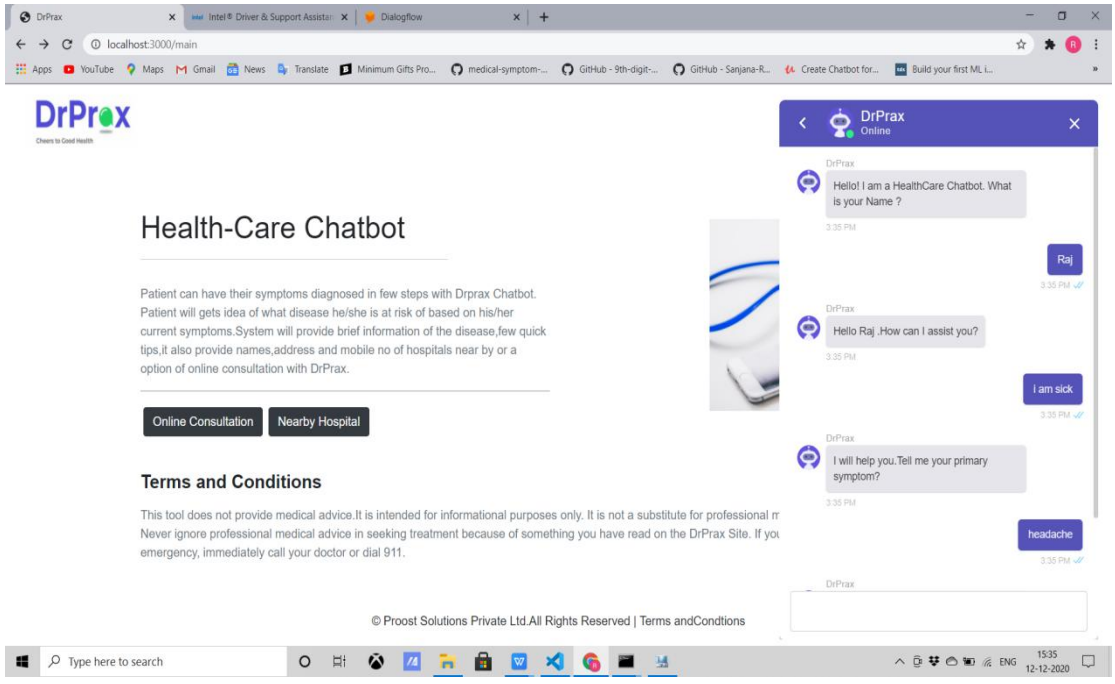
Location page:

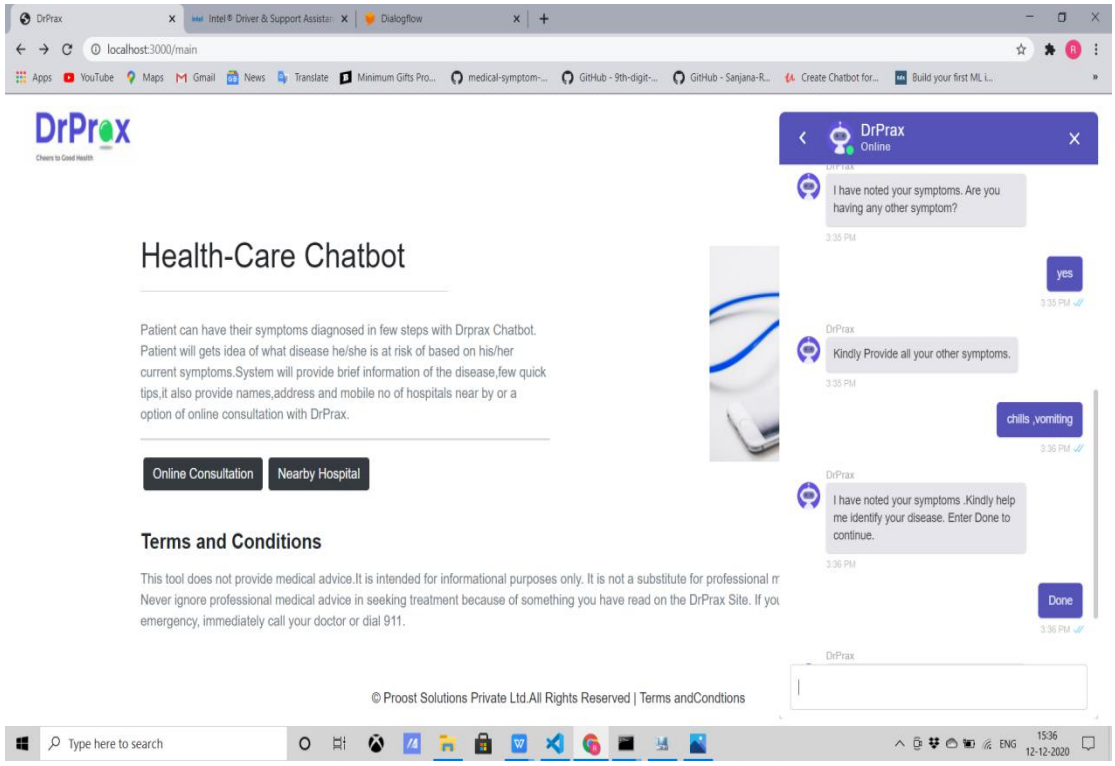


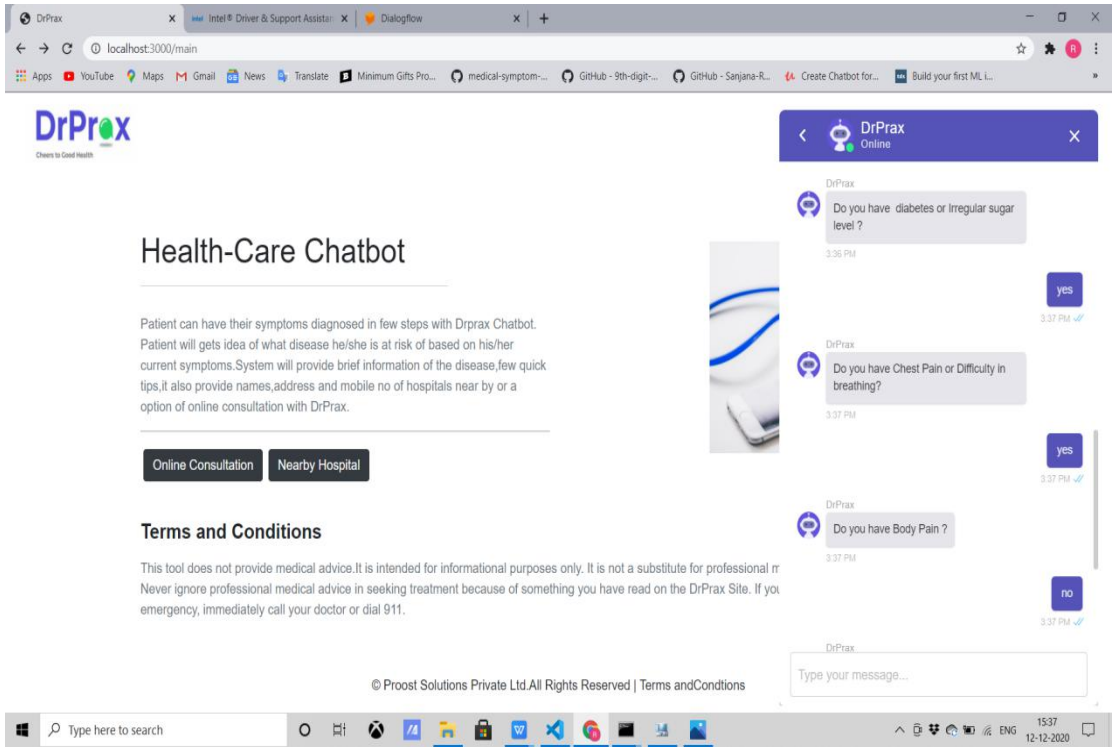
Online Consultation Page:

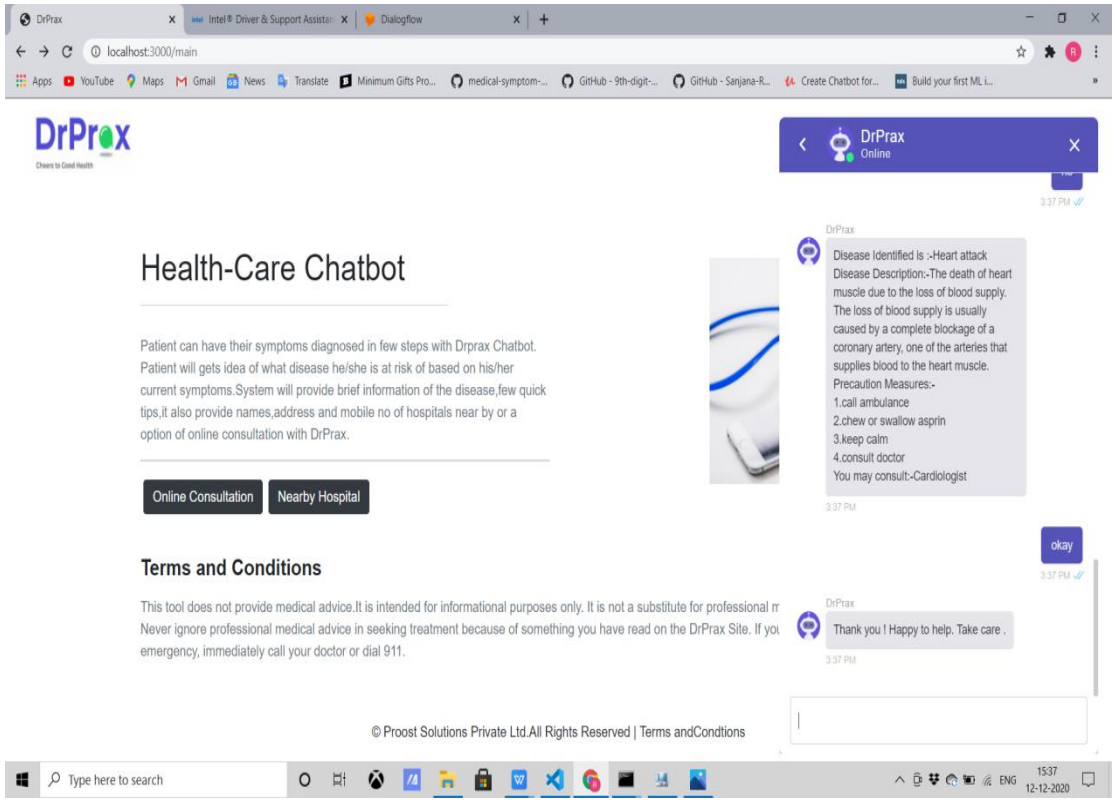


Chatbot Interface :




DrPrax
Intel® Driver & Support Assistan
Dialogflow
localhost:3000/main
DrPrax
Cheers to Good Health
Health-Care Chatbot
Patient can have their symptoms diagnosed in few steps with Drprax Chatbot. Patient will gets idea of what disease he/she is at risk of based on his/her current symptoms.System will provide brief information of the disease,few quick tips,it also provide names,address and mobile no of hospitals near by or a option of online consultation with DrPrax.
Online Consultation
Nearby Hospital
Terms and Conditions
This tool does not provide medical advice.It is intended for informational purposes only. It is not a substitute for professional m
Never ignore professional medical advice in seeking treatment because of something you have read on the DrPrax Site. If you
emergency, immediately call your doctor or dial 911.
© Proost Solutions Private Ltd.All Rights Reserved | Terms andCondtions
DrPrax
Online
I have noted your symptoms. Are you having any other symptom?
3:35 PM
yes
3:35 PM
DrPrax
Kindly Provide all your other symptoms.
3:35 PM
chills ,vomiting
3:36 PM
DrPrax
I have noted your symptoms .Kindly help me identify your disease. Enter Done to continue.
3:36 PM
Done
3:36 PM
DrPrax
Type here to search
ENG
15:36
12-12-2020


DrPrax
Intel® Driver & Support Assistan
Dialogflow
localhost:3000/main
DrPrax
Cheers to Good Health
Health-Care Chatbot
Patient can have their symptoms diagnosed in few steps with Drprax Chatbot. Patient will gets idea of what disease he/she is at risk of based on his/her current symptoms.System will provide brief information of the disease,few quick tips,it also provide names,address and mobile no of hospitals near by or a option of online consultation with DrPrax.
Online Consultation
Nearby Hospital
Terms and Conditions
This tool does not provide medical advice.It is intended for informational purposes only. It is not a substitute for professional m
Never ignore professional medical advice in seeking treatment because of something you have read on the DrPrax Site. If you
emergency, immediately call your doctor or dial 911.
© Proost Solutions Private Ltd.All Rights Reserved | Terms andCondtions
DrPrax
Online
DrPrax
Do you have diabetes or Irregular sugar level ?
3:36 PM
yes
3:37 PM
DrPrax
Do you have Chest Pain or Difficulty in breathing?
3:37 PM
yes
3:37 PM
DrPrax
Do you have Body Pain ?
3:37 PM
no
3:37 PM
DrPrax
Type your message...
Type here to search
ENG
15:37
12-12-2020


DrPrax
Intel® Driver & Support Assistan
Dialogflow
localhost:3000/main
Apps YouTube Maps Gmail News Translate Minimum Gifts Pro... medical-symptom-... GitHub - 9th-digit-... GitHub - Sanjana-R... Create Chatbot for... Build your first ML i...
DrProx
Cheers to Good Health
Health-Care Chatbot
Patient can have their symptoms diagnosed in few steps with Drprax Chatbot. Patient will gets idea of what disease he/she is at risk of based on his/her current symptoms.System will provide brief information of the disease,few quick tips,it also provide names,address and mobile no of hospitals near by or a option of online consultation with DrPrax.
Online Consultation
Nearby Hospital
Terms and Conditions
This tool does not provide medical advice.It is intended for informational purposes only. It is not a substitute for professional m
Never ignore professional medical advice in seeking treatment because of something you have read on the DrPrax Site. If you
emergency, immediately call your doctor or dial 911.
© Proost Solutions Private Ltd.All Rights Reserved | Terms andCondtions
DrPrax
Online
3:37 PM
DrPrax
Disease Identified is :-Heart attack
Disease Description:-The death of heart muscle due to the loss of blood supply. The loss of blood supply is usually caused by a complete blockage of a coronary artery, one of the arteries that supplies blood to the heart muscle.
Precaution Measures:-
1.call ambulance
2.chew or swallow asprin
3.keep calm
4.consult doctor
You may consult:-Cardiologist
3:37 PM
okay
3:37 PM
DrPrax
Thank you ! Happy to help. Take care .
3:37 PM
Type here to search
ENG
15:37
12-12-2020

CHAPTER 9. CONCLUSIONS

9.1 CONCLUSION

As the field of Healthcare is moving towards AI technology, Machine learning is now being adopted into this field and a wide variety of methods are being used for various tasks in the diagnosis of diseases in patients.We have built a working medical bot that is user friendly who knows how to type in the English language and provides a personalized diagnosis of the user's health. It heavily relies on machine learning algorithms as well as the data to train the chatbot and also predict diseases. Patients get a better understanding before-hand about the disease they might be facing so they could consult doctors and ensure their safety. The portal also provides the feature of Nearby Hospitals and Virtual consultation.

9.2 FUTURE SCOPE

- In the future, the conversational flow can be improved with better APIs.
- Given, the disease is predicted accurately, the chat-bot can prescribe medicine to the user.
- The database can be improved by adding all sorts of diseases and symptoms to make it easier to use.
- Additional details like gender, blood group, hereditary information and the patient's previous medical condition could also be incorporated for the analysis.
- Once a user takes the test his/her a proper summary can be generated from the results of the test.

9.3 APPLICATIONS

This chatbot will benefit the patients by ensuring their safety with a simple quick test and also preserve time, effort, and other overheads.

APPENDIX

APPENDIX A

Problem statement feasibility assessment using, satisfiability analysis and NP Hard, NP-Complete or P type using modern algebra and relevant mathematical models.

Theory :-

What is P?

P is set of all decision problems which can be solved in polynomial time by a deterministic.Since it can be solved in polynomial time, it can be verified in polynomial time.Therefore P is a subset of NP.

What is N?

'N' in 'NP' refers to the fact that you are not bound by the normal way a computer works, which is step-by-step. The 'N' actually stands for 'Non-deterministic'. This means that you are dealing with an amazing kind of computer that can run things simultaneously or could somehow guess the right way to do things, or something like that.So this 'N' computer can solve lots more problems in 'P' time – for example it can just clone copies of itself when needed.So, programs that take dramatically longer as the problem gets harder (i.e not in 'P') could be solved quickly on this amazing 'N' computer and so are in 'NP'. Thus 'NP' means 'we can solve it in polynomial time if we can break the normal rules of step-by-step computing'.

What is NP?

'NP' means 'we can solve it in polynomial time if we can break the normal rules of step-by-step computing'.

What is NP-Complete?

Since this amazing 'N' computer can also do anything a normal computer can,we know that 'P' problems are also in 'NP'.So, the easy problems are in 'P' (and 'NP'), but the really hard ones are only in 'NP', and they are called 'NP-complete'.It is like saying there are things that People can do ('P'), there are things that Super People can do ('SP'), and there are things only Super People can do('SP-complete').

What is NP Hard?

A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP-problem (non-deterministic polynomial time) problem. NP-hard therefore means 'at least as hard as any NP-problem,' although it might, in fact,be harder.

Economic Feasibility

Economic analysis is the most frequently used technique for evaluating the effectiveness of a proposed system. More commonly known as Cost / Benefit analysis, the procedure is to determine the benefits and savings that are expected from a proposed system and compare them with costs. If benefits outweigh costs, a decision is taken to design and implement the system. Otherwise,further justification or alternative in the proposed system will have to be made if itis to have a chance of being approved. This is an outgoing effort that improves inaccuracy at each phase of the system life cycle.

Operational Feasibility

This is mainly related to human organizational aspects.The points to be consider As most network packet forwarding systems are hidden from end user so direct interaction of end use are not concern with this system but they will get better service over the network. This feasibility study is carried out by a small group of people who are familiar with information system technique and are skilled in system analysis and design process.

Conclusion:

Hence we have successfully stated feasibility assessment using NPhard/NP-complete and mathematical model.

APPENDIX B

Plagiarism Report of project report.

Abstract

Page 1

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	191	Date	December 15,2020
Characters	1171	Exclude URL	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	8 Unique Sentences

Content Checked For Plagiarism

ABSTRACT

A bot or chatbot is meant to make a conversation between humans and machines. The machine has been embedded with the knowledge to identify the sentences and to make a decision itself as a response to answer a question. The objective of the project is to build a Health Care System (Disease Prediction) using the chatbot concept for patients or anyone who suspects they have a particular disease. This chatbot will be built by using Google's Dialogflow for NLP, React JS for the user interface, Python for building the Machine Learning Model and Mongo DB as the database to store and retrieve information.

The connections of these layers will be established by writing a webhook in fulfilments of Dialogflow and interaction through the Web App. Many companies like Dominos, KLM Royal Dutch Airlines, Ticket master have used Dialogflow for building their interactive bots. Our end product will be a Web App that will be able to interact with the patients. Thus, the proposed system will be able to process the user input using natural language processing, match the input with the Machine Learning Model and then generate a suitable response.

Introduction

Page 1

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	521	Date	December 16,2020
Characters	3318	Exclude URL	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	22 Unique Sentences

Content Checked For Plagiarism

1.INTRODUCTION
1.1 OVERVIEW:-
There are many tools related to disease prediction. But generally, there are no such tools that are used for the prediction of general diseases. So, our Disease Predictor helps for the prediction of these general diseases. It takes in input from the user such as their symptoms if they have any heart problems etc. and gives a probable disease that they may be suffering from, along with some immediate suggestions as well as which doctor, they should refer to. It also gives a list of nearby hospitals in the user's vicinity.
1.2 MOTIVATION:-
At present, when someone suffers from a particular disease, they have to visit a doctor, which is time-consuming and costly. If the user is out of reach of doctors and hospitals it may be difficult for the user as the disease cannot be identified. So, if the above process can be completed using an automated program which can save time as well as money, it can be very helpful to the user and can make the process easier. The aim of Disease prediction is to predict the disease of patient provides the symptoms. As there is advancements in digital technology the use of internet is at pace, people are curious towards new things. People always attempt to ask the web if any problem arises. People have access to the internet faster than hospitals and doctors. So, this system can be helpful to people to know how serious their symptoms are and which doctor they need to go to.
1.3 PROBLEM DEFINITION AND OBJECTIVES:-
The objective of the project is to build a chatbot that can predict diseases and provide information on the disease as well as nearby hospitals.
1. Provide accurate prediction of diseases.
2. Save the efforts of getting unnecessary checkups from different doctors.
3. Interactive and easy chatbot.
1.4 PROJECT SCOPE & LIMITATIONS:-
1.4.1. Scope
This project aims to provide a web platform to predict the occurrences of disease based on various symptoms. The user can type in their symptoms, answer a few questions and find the probable disease with some immediate suggestions as well as which doctor, they should refer to. They can also check nearby hospitals.
1.4.2. Limitations
1. History of diseases and genetic history is not considered.
2. Model not taught to be continuously learning.
1.5 METHODOLOGIES OF PROBLEM SOLVING
We can implement the chatbot using various approaches. Development of Machine Learning Based chatbots is becoming popular and easier for software developers using Dialogflow .Developers can use it to make interfaces (e.g., chatbots) that are capable of natural and rich interactions between users and businesses. It is powered by machine learning to perceive the intent and context of what a user says, allowing the conversational interface to produce highly efficient and accurate responses. Thus, we are implementing our chatbot using Dialogflow.
Disease Prediction has been already being implemented using different techniques like Neural Network, the Naive Bayes algorithm, etc. From the analysis, it was found that the Decision Tree is more accurate than other techniques. Kommunicate provides customized User interface for the chatbot. Kommunicate was being integrated with React js.

Literature Survey

Page 1

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	369	Date	December 16,2020
Characters	2554	Exclude URL	

0%	100%	0	17
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

2.LITERATURE SURVEY

K.M. Al-Aidaroos, A.A. Bakar and Z. Othman have researched the best medical diagnosis mining technique. For this author compared Naïve Bayes with five other classifiers like Decision tree, K star, Logistic regression, ANN and ZeroR. For evaluating the performance of all algorithms. ,15 medical problems from the UCI repository (Asuncion and Newman, 2007) were selected. In the experiment, it had been found that NB outperforms the opposite algorithms in 8 out of 15 data sets so it had been concluded that the predictive accuracy leads to Naïve Bayes are better than other techniques.

(Al-Aidaroos, Bakar, & Othman, 2012)

Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni have done this research paper into providing a survey of current techniques of data discovery in databases using data processing techniques that are in use in today's medical research particularly in heart condition Prediction. The number of experiment has been conducted to match the performance of predictive data processing technique on an equivalent dataset and therefore the outcome reveals that call Tree outperforms and some time Bayesian classification has similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification supported clustering isn't performing well.

(JyotiSoni, Ansari, Sharma, & Soni, 2011)

Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, Nicholas Christakis, Albert-Laszlo Barabasi have found that global treatment of chronic disease is neither time nor cost-efficient. So the research was conducted by authors to predict future disease risk. For this CARE was used (which relies only on a patient's medical record using ICD- 9-CM codes to predict future disease risks). CARE combines collaborative filtering methods with clustering to predict each patient's greatest disease risks supported their medical record which of comparable patients. For improved performance, authors have also described an Iterative version, ICARE. These novel systems require no specialized information and supply predictions for medical conditions of all types during a single run. The future disease scope of ICARE is providing more accurate warnings for diseases beforehand . Applied to full potential, the CARE framework are often wont to explore a broader disease Disease Predictor 7 histories, suggest previously unconsidered concerns, and facilitating discussion about early testing and prevention.

(A.Davis, V.Chawla, Blumm, Christakis, & Barbasi, 2008)

Sources	Similarity

Software Requirements Specification

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	924	Date	December 15,2020
Characters	5971	Exclude URL	

0%	100%	0	50
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

3.1ASSUMPTIONS AND DEPENDENCIES:-
The user needs to have an active internet connection to access the contents of the website. The user must have a laptop or a desktop for a mobile phone for using this web application. In case the client or user is using a laptop or desktop then he should be on an operating system that supports internet browsing. This Web App won't be violating any of the internet rules.
3.3 EXTERNAL INTERFACE REQUIREMENTS:-
3.3.1 User Interface
The interface provided to the user will consist of a menu for the user to select the required category for which the function should be performed. After selecting a category, the user either acts as virtual consultation or a nearby hospital location or the user can talk with a chatbot.
3.3.2 Hardware Interface
To run the website on any device, a browser with CGI support working on any operating system (Windows, Linux or Mac). Since it is not a resource-hungry program, it will run on most of the systems without hassle. The user needs to have a laptop/desktop/Mobile Phone with an internet connection.
3.3.3 Software Interface
The system will be built using MERN Framework along with Flask which is used to handle the webhook request. Various machine learning libraries like NumPy, pandas and scikit-learn. Also, a database to store user data as per requirements. Google API interface is used to provide nearby doctors to the user. Dialogflow API is used to create the chatbot and Kommunicate is used for designing the chatbot UI.
3.4 NON-FUNCTIONAL REQUIREMENTS:-
3.4.1 Performance Requirements:
We have considered important the fact that the system must perform as what every user expects. So there are no immediate delays. In case of opening different webpages, or giving error messages, etc. there is a delay much below 2 seconds, so the throughput is considerably high. The response time is quite less. The web application gives instant results and the web pages also load fast with a considerable internet bandwidth speed. In case of high workloads to it is intended to prove great in terms of response time. The system may go into failures only when the bandwidth is very less. The response from the chatbot is quick and the doctors in a nearby location to are viewed instantly.
3.4.2 Safety Requirements:
The system will be distributed among various nodes to ensure maximum safety from threats and attacks. Secure Logins will protect user data.
3.4.3 Security Requirements:
The user in the application is provided with a username and password for authenticating before the login process. Email authentication improves the degree of safety. invalid users are blocked from logging in. Encryption and decryption facility is being provided for the passwords to increase authentication and security.
3.4.3 Software Quality Attributes:
Availability: We have been Checking that the system always has something to function and always show error messages in case of problems. So here, error messages appear when something goes wrong.
Maintainability: The diseases and their symptom data is checked and maintained.
Usability: Checking from time to time that the system is easy to handle and works most expectedly with no delays. In

that case, the system program reacts accordingly.
Accessibility: Better fonts and colors are used for the webpages.
3.5 SYSTEM REQUIREMENTS:
3.5.1 Database Requirements
A classical database built-in Mongo server is needed to be integrated within the project which is secure as well as portable over the internet. Its features should be such that there is ease of access to data from remote locations as well.
3.5.2 Software Requirements:
React Js for the front end of web pages:
React is a JavaScript library used to build single-page applications ie. User interfaces. We can create reusable UI components using the React Js framework.
Kommunicate for chatbot Ui:
We have used Kommunicate to design the chatbot UI. It is BOT+ HUMAN hybrid support. It helps manage customer conversations, website chat, etc.
Dialog flow for building chatbot:
Dialog flow is a service of Google Cloud Platform and it is used to build conversational chatbots. These chatbots can be customized and added to websites or mobile apps, etc. Through the use of intents and entities, a chatbot is made and integrated.
Nodejs for backend:
Node.js exists as a free open source server environment that runs on various platforms.
It can generate the dynamic page content. also, node js can create, open, read, write, delete, and close files on the server. NodeJS is a cross-platform runtime environment highly used for developing server-side and networking applications.
Python(flask) for connecting webhook:
Flask is a web framework that will provide us with tools, libraries and technologies through which we can build a web application. It is written in Python. It is a lighter framework and easier to use as compared to frameworks such as Django.
Webhook:
Webhook exists as a web call back for an app to issue other applications with real-time information. Whenever some event happens on the trigger application, data about that event is serialized and send to a webhook URL from the active application where some other task will be performed.
MongoDB as Database:
MongoDB is an object-oriented database that is based on the NoSQL document store model It is simple and Dynamic to use. It provides a better and reliable platform for data storage. Data objects in mongo dB are stored as separate documents and documents and inside collections. The Traditional Data Storage system of rows and columns is eliminated here.
3.5.2 Hardware Requirements:
A Desktop/ Laptop with an internet connection and an operating system that supports web browsing.

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	685	Date	December 15,2020
Characters	4865	Exclude URL	

0%	100%	0	25
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

3.6 Analysis Models: SDLC Model to be applied
Classical methods of software development have many disadvantages:
Huge effort during the planning phase
Poor requirements conversion in a rapidly changing environment
Treatment of staff as a factor of production
New methods: Agile Software Development
Scrum - an agile process
SCRUM is an agile, lightweight process for managing and controlling software and products development in rapidly changing environments.
a. Iterative, incremental process
b. Team-based approach
c. developing systems/ products with rapidly changing requirements
d. Controls the confusion of conflicting interests and needs
e. Improve communication and maximize cooperation
f. Protecting the team from disruptions
g. A way to maximize productivity
Components of Scrum
a. Scrum Roles
b. Process
c. Scrum Artifacts
Scrum Master
1. Represents management to the project
2. Typically filled by a Project Manager or Team Leader
3. Responsible for enacting Scrum values and practices
4. The main job is to remove impediments
The Scrum Team
1. Typically 5-10 people
2. Cross-functional (QA, Programmers, UI Designers, etc.)
3. Members should be full-time
4. Team is self-organizing
5. Membership can change only between sprints
Product Owner
1. Acts like one voice (in any case)
2. Knows what needs to be built and in what sequence this should be done
3. Typically a product manager
The Process
1. Sprint Planning Meeting
2. Sprint
3. Daily Scrum

4. Sprint Review Meeting
Sprint Planning Meeting
1. A collaborative meeting at the beginning of each Sprint between the Product Owner, the Scrum Master and the Team
2. Takes 8 hours and consists of 2 parts ("before lunch and after lunch")
Parts of Sprint Planning Meeting
1. 1st Part:
a. Creating Product Backlog
b. Determining the Sprint Goal.
c. Participants: Product Owner, Scrum Master, Scrum Team
2. 2nd Part:
a. Participants: Scrum Master, Scrum Team
b. Creating Sprint Backlog
Pre-Project/Kickoff Meeting
1. A special form of Sprint Planning Meeting
2. Meeting before the beginning of the Project
Sprint
1. A month-long iteration, during which is incremented a product functionality
2. NO outside influence can interference with the Scrum team during the Sprint
3. Each Sprint begins with the Daily Scrum Meeting
Daily Scrum
1. Is a short (15 minutes long) meeting, which is held every day before the Team starts working
2. Participants: Scrum Master (which is the chairman), Scrum Team
3. Every Team member should answer 3 questions
Questions
1. What progress is thereafter the last Scrum?
2. What are you doing until subsequent Scrum?
3. What is stopping you from getting on with the work?
Sprint Review Meeting
1. Is held at the end of each Sprint
2. Business functionality which was created during the Sprint is demonstrated to the Product Owner
3. Informal, should not distract Team members from doing their work
Scrum Artifacts
1. Product Backlog
2. Sprint Backlog
3. Burndown Charts
Product Backlog
Requirements for a system are usually expressed as a list of Backlog Items according to priority.
1. Is managed and owned by a Product Owner
2. Spreadsheet (typically)
3. Usually is created during the Sprint Planning Meeting
4. Can be changed and re-prioritized before each Planning Meeting
Estimation of Product Backlog Items
1. Establishes team's velocity (how many efforts a Team can handle in one Sprint)
2. Determining units of complexity.
a. Size-category ("T-Shirt size")
b. Story points
c. Workdays/work hours
3. Methods of estimation:
a. Expert Review
b. Creating a Work Breakdown Structure (WBS)
Sprint Backlog
1. A subclass of Product Backlog Items, which is the work of Sprint.
2. Is created ONLY by Team members
3. Each Item has its status
4. Should be updated every day
5. No more than 300 tasks in the list
6. If a task requires more than 16 hours, it should be broken down
7. The team can add or subtract items from the list. The Product Owner isn't allowed to try and do it.
Burndown Charts
1. Are used to represent "work done".
2. Are wonderful Information Radiators
3. 3 Types:

Project Plan

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	223	Date	December 16,2020
Characters	1698	Exclude URL	

0% Plagiarism	100% Unique	0 Plagiarized Sentences	9 Unique Sentences

Content Checked For Plagiarism

5.PROJECT PLAN
5.1 PROJECT ESTIMATE
5.1.1 Reconciled Estimates
5.1.1.1 Cost Estimate
In the real world, there are 3 types of models included in estimation using the COCOMO Model. They can be described as follows:
1. Organic: Small team size, 2-50 KLOC, stable and little innovation required
2. Semi Detached: 50-300 KLOC, medium-sized, average abilities, medium time-constraints
3. Embedded: > 300 KLOC, large project team, complex, innovative, severe constraints
The proposed system will utilize the Organic Model for Cost Estimation.
Step 1:
E=a*KLOCb
Where, E = Effort in staff-months
a, b = Coefficients to be determined
KLOC = Kilo lines of Code
Step 2:
For organic model a = 2.4 and b = 1.05
Assuming KLOC = 6
E=2.4*(4)1.05=10.3
Step 3:
Project Duration
Where TDEV = Time for Development
TDEV=c*(E)d
For Organic c = 2.5 and d = 0.38
TDEV=2.5*(10.3)0.38=6.0 Months
Step 4:
Average Staff Size=E/TDEV=10.3/6 = 2 to 3 Staff
Step 5:
Productivity=P=KLOC/E=4000/10.3=388.35 Loc/staff-month
Step 6:
Assuming cost per month = ₹ 5000 /-
Total cost=E*cost per Month=10.3*5000=Rs 51500 /-
5.1.1.2 Time Estimates
Project Duration
TDEV=c*(E)d
Where TDEV = Time for Development
For Organic c = 2.5 and d = 0.38
TDEV=2.5*(10.3)0.38=6.0 Months
5.1.2 Project Resources

There are certain external requirements for the project to sucessfully run lifetime like Google Api ,Kommunicate , MailGun.But ,currently we have used free version of these requirements. Page 2

Sources	Similarity

Project Implementation

Page 1

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	913	Date	December 15,2020
Characters	6002	Exclude URL	

0%	100%	0	40
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

6.PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

We have designed a linear based retrieval based chatbot which acts as a tool of communication to predict disease based on the input, the symptoms given by the user. it has 2 phases of conversation i.e extraction of symptoms and diagnosis of the symptoms. The input given to the chatbot is in the form of unstructured text and thus we have used natural language understanding of Dialogflow to extract symptoms from it. These symptom vector arrays are now passed to the machine learning model, the model shortlists several diseases based on the input. Here one of the models used is a decision tree whose root node returns the most likely disease. Another 3 model used is Gaussian Naive Bayes, Support Vector Machine, and K Nearest Neighbour Classifier to match which model finds the best response based on the highest accuracy. to get a better understanding of the disease the chatbot responds with several questions and thus finally the most likely disease is predicted based on the number of symptoms matched. The front end of the system is designed using React js. A chatbot is being customized using Kommunicate.

1 Symptom Extraction

Symptom extraction is performed to identify symptoms from the user's text input. A results of misspellings may lead to no or wrong symptoms mapping and may not always map to the database's symptoms. To handle this, a variety of symptoms are extracted and compared with the symptoms in the database using its synonyms. which are defined in entities in Dialogflow.

2 Diagnosis of Symptoms

Various machine learning models were trained on the disease-symptom database. Once symptoms are extracted from the user, the decision Tree model was applied to the set of symptoms, and the predicted disease is specified to the end-user by the chatbot.

Database

The database is required for the security as well as the integrity of the project. The machine learning module requires input data which is fed from the database. All the results obtained from the machine learning module extract final data from the database. It provides a reliable groundwork for machine learning to train and store important aspects of the data. The database consists of three different master tables of all categories which in turn are connected to their respective attributes.

User Interface

The user interface is the only module through which the users get connected to the system. it's a React-based user interface. it is integrated into the system through the Node js and Python (Flask)server.

6.2 TOOLS AND TECHNOLOGIES USED

Tools and technologies used for Web Development:

1. Visual Studio Code
2. React js
3. Node js
4. Kommunicate

Tools and technologies used for Machine Learning:

1.Dialogflow
2. Python 3.8

Page 2

3. Scikit-learn
4. Pandas
5. Decision Tress Library
6. Ngrok

Database:MongoDB

API: Google API, MailGun

6.3 ALGORITHM DETAILS

6.3.1 Naive Bayes Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which depends on the Bayes theorem and used for solving classification problems. it is mainly utilized in text classification that has a high-dimensional training dataset.

Naïve Bayes Classifier is one of the straightforward and best Classification algorithms which helps in building the fast machine learning models which can make quick predictions. it is a probabilistic classifier, which suggests it depends on the idea of the probability of an object.

6.3.2 Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the foremost popular Supervised Learning algorithms, which comes under Classification and as Regression problems. the primary purpose of SVM is, it is used for Classification problems in Machine Learning. the main objective of the SVM algorithm is to form the decision boundary which will segregate n-dimensional space into classes so we will easily put the new information in the right category within the future. This best decision boundary is named a hyperplane.

SVM chooses the ultimate points/vectors that help in creating the hyperplane. These extreme cases are called support vectors.

6.3.3 K-Means Clustering Algorithm

K-Means Clustering is one among the unsupervised learning algorithms whose aim is to urge solutions to the clustering problems in machine learning or data science. Determines the most effective value for K center points or centroids by an iterative process. Assigns each datum to its closest k-center. Those data points which are around the actual k-center, create a cluster.

6.3.4 Decision Tree Algorithm

Decision Tree comes under Supervised learning technique which is able to be used for both classification and Regression problems, but mostly it's preferred for solving Classification problems. it's a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent main decision rules and every leaf node represents the result. within a Decision tree, there are two nodes, which are the principal decision Node and Leaf Node. Decision nodes are wont to make any decision and have multiple branches, whereas Leaf nodes are the output of these decisions and don't contain any more branches.

The decisions or the test are performed based on features of the given dataset. it's a graphical representation for getting all the possible solutions to a problem/decision subjected to given conditions. it's called a decision tree because almost sort of a tree, it starts with the foundation node, which expands on further branches and constructs a tree-like structure. to make a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. a decision tree simply asks a problem and suggests the answer (Yes/No), it further split the tree into subtrees.

Sources	Similarity

Software Testing,Result & Conclusion

Page 1

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	864	Date	December 15,2020
Characters	5853	Exclude URL	

0%	100%	0	48
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

7.SOFTWARE TESTING

7.1 TYPES OF TESTING

This document is a high-level overview defining our testing strategy for web applications. The tests were done were unit tests and validation tests, where features created were unit-tested and also tested for valid inputs. Apart from this, integration tests were also done to ensure all the systems were completely integrated.

7.1.1 Unit Testing

The entire project consists of many small modules where each one needs to be checked and tested if working as per design and requirement. Thus, unit testing was required. We tested each piece of code individually so that in the future, we could find fewer bugs.

7.1.2 Integration Testing

Every module was tested as and when integration was performed. After the integration of each module into the main project, the altogether performance should meet expectations, hence this testing was carried out. We identified issues such as data type inconsistency while passing it through different modules that were resolved after the testing. Major bugfixes took place during this phase of testing.

7.1.3 Compatibility Testing

As our system is a Web Application, browser compatibility testing was necessary. We as developers ensured that this web application runs in every browser regardless of the operating system including various versions of the browser. Rigorous testing of the application was performed in all the browsers on Windows, Linux.

7.1.4 Back–end Testing

To check if the given input from GUI was appropriately stored in the database, this testing was carried out. Table structure, schema, stored procedure, data structure were the parameters considered while carrying out this testing. Few queries were given to authenticate the entire environment.

7.1.5 System Testing

The entire system was tested as per the requirements thoroughly. Overall requirement specifications were taken into consideration while performing this testing.

7.1.6 Alpha Testing

The main aim of this testing was to identify all the defects before releasing it to the end-users. The development team of this project was responsible to do this testing. All bugs including font issues in the User Interface along with relevancy of the output displayed were checked.

7.1.7 Performance Testing

The overall performance including stress and load on the server using different platforms was done. Many requests were sent to the server together to check the amount of load it is capable of handling.

7.2 TEST CASES & TEST RESULTS

TestID Test Case Condition Result

1 login with correct credentials Should open the Main interface.

2 Login with incorrect credentials Shows the message "User Not found".

3 New user and correct credentials Verification Mail followed by the Main interface.

4 New users and Incorrect credentials. Verification Mail followed by "Invalid Password" or " User Not Found" message.

5 User Clicks Nearby Hospital Button Provides Hospitals nearby the user.

Page 2

6 User Clicks Online Consultation Button Opens the Drprax Website.

7 User clicks on log out Logout from the system.

8 User Clicks on Chatbot Icon Open the Chatbot Interface.

9 The test installs required software and checks the module integration A complete Website.

10 Provide Symptoms Ask the next question in order to understand the user.

11 Username, symptoms not understood Ask the same question again from chat interface.

8.RESULTS

8.1OUTCOMES

The user will have a text to a text conversation with the chatbot and the chatbot will identify the symptoms the user is undergoing through. The user needed to enter his/her symptom and answer a few fixed questions to help the chatbot identify the disease. Once the symptoms are entered the Machine Learning algorithm determines the most probable disease, its description, precautions and the doctor you may consult. The application also provides another feature of a nearby hospital and online consultation.

8.2SCREEN SHOTS

Default Page:

Register Page:

Login Page:

AboutUs Page:

Main Page:

Location page:

Online Consultation Page:

Chatbot Interface :

9.CONCLUSIONS

9.1 CONCLUSION

As the field of Healthcare is moving towards AI technology, Machine learning is now being adopted into this field and a wide variety of methods are being used for various tasks in the diagnosis of diseases in patients. We have built a working medical bot that is user friendly who knows how to type in the English language and provides a personalized diagnosis of the user's health. It heavily relies on machine learning algorithms as well as the data to train the chatbot and also predict diseases. Patients get a better understanding before-hand about the disease they might be facing so they could consult doctors and ensure their safety. The portal also provides the feature of Nearby Hospitals and Virtual consultation.

9.2 FUTURE SCOPE

- In the future, the conversational flow can be improved with better APIs.
- Given, the disease is predicted accurately, the chat-bot can prescribe medicine to the user.
- The database can be improved by adding all sorts of diseases and symptoms to make it easier to use.
- Additional details like gender, blood group, hereditary information and the patient's previous medical condition could also be incorporated for the analysis.
- Once a user takes the test his/her a proper summary can be generated from the results of the test.

9.3 APPLICATIONS

This chatbot will benefit the patients by ensuring their safety with a simple quick test and also preserve time, effort, and other overheads.

Sources	Similarity

REFERENCES

1. K.M. Al-Aidaroos, A.A. Bakar and Z. Othman “Medical Data Classification with Naive Bayes Approach ”.
2. Jyoti Soni Ujma Ansari Dipesh Sharma Sunita Soni “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.
3. Darcy A. Davis Nitesh V. Chawla ·Nicholas A. Christakis Albert-László Barabási”Time to CARE: a collaborative enginefor practical disease prediction”.
4. Anil Parajuli“Disease Predictor-Department of Computer Science and Information Technology DWIT College| August 26
5. Gajendra Prasad K. C. Satvik Ranjan Tathagat Ankit Vivek Kumar“A Personalized Medical Assistant Chatbot: MediBot”IJSTE - International Journal of Science Technology & Engineering | Volume 5 | Issue 7 | January 2019.
6. Arash Shaban-Nejad Martin Michalowski David L Buckeridge“Health intelligence: How Artifcial Intelligence Transforms Population and Personalized Health”.
7. Kobrinskii,O.G.Grigoriev,A.I.Molodchenkov,I.V.Smirnov,N.A.Blagosklonov“Artif icial Intelligence Technologies Application for Personal Health Management”IFAC PapersOnLine 52-25 (2019) 70–74.
8. Dimitrios Mantzaris George Anastassopoulos Dimitrios Lymberopoulos“Medical disease prediction using Arti?cial Neural Networks”.
9. Swapnil Chaudhari“A Parallel System Implementation of Classification and Disease Prediction on Machine Learning from Healthcare Communities”.
10. Hiba Hussain, Komal Aswani, Mahima Gupta, Dr. G.T.Thampi”Implementation of Disease Prediction Chatbot and Report Analyzer using the Concepts of NLP, Machine Learning and OCR” International Research Journal of Engineering and Technology (IRJET)Volume:07 Issue: 04 | Apr 2020 .