knowledge about the prescriptions and they can involve in a good debate with the physicians.

1.1 Scope

As part of achievable scope, we propose to create a system that takes as input the symptoms that the patient notices. Based on those symptoms proper medicines will be prescribed to the patients. Big Data Analytics is used to understand the symptoms. patient can also get the information about the alternate cheaper medicines.

Chapter 2

Review of Literature

2.1 Domain Explanation

2.1.1 Big Data Analytics

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zettabytes. Big data is a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process the data with low-latency. And it has one or more of the following characteristics – high volume, high velocity, or high variety. Big data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media - much of it generated in real time and in a very large scale.

Analysing big data allows analysts, researchers, and business users to make better and faster decisions using data that was previously inaccessible or unusable. Using advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyse previously untapped data sources independent or together with their existing enterprise data to gain new insights resulting in significantly better and faster decisions.

With the increasing production of information from various initiatives, there is also the need to transform a large volume of unstructured data into useful information for society. All this information should be easily accessible and made available in a meaningful and effective way to achieve semantic interoperability in electronic services. The concept of variety often underlies use of the term big data. Big Data refers to technologies and initiatives that involve data that is too diverse, fast-changing or massive for conventional technologies, skills and infra- structure to address efficiently. It’s about the ability to make better decisions and take meaningful actions at the right time.

Innovations in technology and greater affordability of digital devices have presided over today's Age of Big Data, an umbrella term for the explosion in the quantity and diversity of high frequency digital data. The world is filled with data. Several other authors often refer to the "Three V's" of big data: volume, variety, and velocity, originally discussed by Laney in 2001, to distinguish big data. Volume refers to the actual size of the dataset(s) analysed, variety to the various types of datasets possibly combined to produce new insights, and velocity to the frequency with which data is recorded and/or analysed for action. When trying to understand the concept of Big Data, the words "MapReduce" and "Hadoop" cannot be avoided.

2.1.1.1 MapReduce

MapReduce is a huge hit. Google faced with making sense of the largest collection of data in the world and took on this challenge. The result was MapReduce: a software framework that breaks big problems into small, manageable tasks and then distributes them to multiple servers. Google makes very heavy use of MapReduce internally, and the Apache Software Foundation turned to MapReduce to form the foundation of its Hadoop implementation. MapReduce can work with raw data that's stored in disk files, in relational databases, or both. The data may be structured or unstructured, and is commonly made up of text, binary, or multi-line records. The most common MapReduce usage pattern employs a distributed file system known as Hadoop Distributed File System (HDFS).

Data is stored on local disk and processing is done locally on the computer with the data. The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

At its core, MapReduce is composed of two major processing steps: Map and Reduce.

- The Map task takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key-value pairs).
- The Reduce task takes the output from the Map as an input and combines those data tuples (key-value pairs) into a smaller set of tuples. The reduce task is always performed after the map job.

2.1.2 Recommendation System

2.1.2.1 Definition

A recommendation system is a software program which attempts to narrow down selections for users based on their expressed preferences, past behaviour, or other data which can be mined about the user or other users with similar interests.

2.1.2.2 Overview

Recommendation systems have impacted or even redefined our lives in many ways. One example of this impact is how our online shopping experience is being redefined. As we browse through products, the Recommendation system offer recommendations of products we might be interested in. Regardless of the perspective — business or consumer, Recommendation systems have been immensely beneficial. And big data is the driving force behind Recommendation systems. A typical Recommendation system cannot do its job without sufficient data and big data supplies plenty of user data such as past purchases, browsing history, and feedback for the Recommendation systems to provide relevant and effective recommendations. In a nutshell, even the most advanced Recommenders cannot be effective without big data.

2.1.2.4 How does a Recommendation system work?

A Recommendation system works in well-defined, logical phases which are data collection, ratings, and filtering. These phases are described below.

2.1.2.4.1 Data collection

Let us assume that a user of Medicine Recommendation system is looking for some medicine details. Each time the user enters a name of medicine or enters the symptoms of a disease, an event such as an Ajax event could be fired. The event type could vary depending on the technology used. The entry is technical in content but in layman's language could read something like "User A preferred (in case of symptoms) or looked-for medicine Z details once". That is how user details get captured and stored for future recommendations.

How does the Recommendation system capture the details? If the user has logged in, then the details are extracted either from an http session or from the system cookies. In case the Recommendation system depends on system cookies, then the data is available only till the time the user is using the same terminal. Events are fired almost in every case — a user preferring a medicine or rating a medicine and providing reviews for it. So that is how user details are stored. But that is just one part of what Recommenders do.

2.1.2.4.2 Ratings

Ratings are important in the sense that they tell you what a user feels about a product. User's feelings about a product can be reflected to an extent in the actions he or she takes such as likes, adding to shopping cart, purchasing or just clicking. Recommendation systems can assign implicit ratings based on user actions. The maximum rating is 5. For example, effective can be assigned a

rating of 4, should work can get 3, preferring/clicking can get 2 and so on. Recommendation systems can also consider ratings and feedback users provide.

2.1.2.4.3 Filtering

Filtering means filtering products based on ratings and other user data. Recommendation systems use three types of filtering: collaborative, user-based and a hybrid approach. In collaborative filtering, a comparison of users' choices is done and recommendations given. For example, if user X likes products A, B, C, and D and user Y Likes products A, B, C, D and E, the it is likely that user X will be recommended product E because there are a lot of similarities between user's X and Y as far as choice of products is concerned. In user-based filtering, the user's browsing history, likes, purchases and ratings are considered before providing recommendations.

2.1.2.5 Role of big data

As stated earlier, big data drives what Recommenders do primarily. Recommenders cannot do a thing without the constant supply of data. However, the role of big data goes beyond just data. The above operations require a high-capacity CPU which can work for hours. To realize this, Hadoop can be used.

So, the role of big data can be summed in providing meaningful, actionable data fast and providing necessary setup to quickly process the data. It is obvious that traditional technologies are not meant to process such large volumes of data so quickly. So, it will not suffice to just have big data to provide strong recommendations.

2.1.2.6 Limitations of Recommendation systems

- For all their efficiencies, Recommendation Systems are not a full proof system. Recommenders have been known to suffer from the following limitations:

- Recommenders depend totally on data and their hirers must constantly supply them with large volumes of data. That is why; smaller firms are more disadvantaged then the bigger firms such as Google and Amazon.
- Recommenders may find it difficult to exactly identify user choice patterns if the user preferences tend to vary quickly, as in fashion. Recommenders depend a lot on historic data but that may not be suitable for certain product niches.
- Recommenders face problems with unpredictable items. For example, there are certain movie types that evoke extreme reactions such as love or hate.

2.2 Existing Solution

- MedicineNet is a medical website that provides detailed information about diseases, conditions, medications and general health. MedicineNet.com launched on October 17, 1996 William Shiel co-founded MedicineNet and continues today as the Chief Medical Editor. Melissa Stoppler also serves on the MedicineNet Editorial Board and she and Shiel were co-editors-in-chief of Webster's New World Medical Dictionary, Year 2008, Third Edition.

- MedicineNet is an owned and operated site in the WebMD Consumer Network and was acquired by WebMD in December 2004. MedicineNet is a network of U.S. Board-Certified Physicians and Allied Health Professionals working together to provide the public with current, comprehensive medical information, written in easy to understand language.

- MedicineNet, as part of the WebMD Consumer Network, adheres to the same privacy policy as WebMD.com and is certified by the TRUSTe online privacy certification program. In addition, MedicineNet is HONcode certified to be in compliance with the Health on the Net health website principles.

- MedicineNet is ranked #5 in the January 2014 eBizMBA Top 15 Most Popular Health Websites.

2.3 H/W and S/W requirement

The application requires certain hardware and software requirements to fulfil the needs of the project
They are as follows:

2.3.1 Hardware Requirements

During Development:

- Windows XP enabled Computer
- RAM-2GB
- Hard Disk Space- a minimum of 1GB of space which is needed

During Usage:

- Windows XP7/VISTA enabled Computer
- RAM-2GB
- Hard Disk Space- minimum of 512 MB of free space which is required on system's drive.

2.3.2 Software Requirements

During Development:

- NetBeans/Eclipse, Web Browser, AWS (as a front end for GUI design)
- HTML, CSS, Java (as a programming language)

During Usage:

- HTML file of our system, Web Browser.
- Text pad (for taking as input the database from FLAT files)

Chapter 3

Analysis

3.1 Functional Requirements

- Search medicine according to symptoms.
- Search medicine according to medicine name.
- Can view alternatives for noxious medicines.
- Can view the dosage requirements of each medicine.
- The recommender system will provide knowledge about medication with excellent efficiency and scalability.

3.2 Non-Functional Requirements

- **Learnability**: A new user should be able to use the recommender engine without putting too much efforts, and, in case of doubt, there must be some help to solve their doubts.
- **Reusability**: The recommender engine should be reusable when someone wants to reuse and search for a different medicine.
- **Performance**: The recommender engine should generate recommendations within a time frame of 500 milliseconds.
- **Reliability**: Integrity and consistency of the recommender engine and all its transactions should be ensured.
- **Correctness**: The system should be able to predict results with a very small margin of error.
- **Privacy**: The privacy of the information provided by user should be guaranteed in the system.

3.3 Proposed System

We design and implement a universal medicine recommender system framework that applies Big Data technologies to the recommendation system. The medicine recommender system consists of database system module, data preparation module, recommendation model module and model evaluation. We investigate the medicine recommendation algorithms of the MapReduce algorithm, Collaborative Filtering algorithm and Content Based algorithm based on the diagnosis data.

Experiments are done to tune the parameters for each algorithm to get better performance. Finally, in the given open dataset, recommendation model is selected for the medicine recommendation module to obtain a good trade-off among model accuracy, model efficiency, and model scalability. We also propose a mechanism to recommend the medicine to a user based on another user with similar health conditions. Experimental results show our system can give medication recommendation with an excellent efficiency, accuracy and scalability.

Chapter 4

Design

4.1 Design Details

4.1.1 Block Diagram

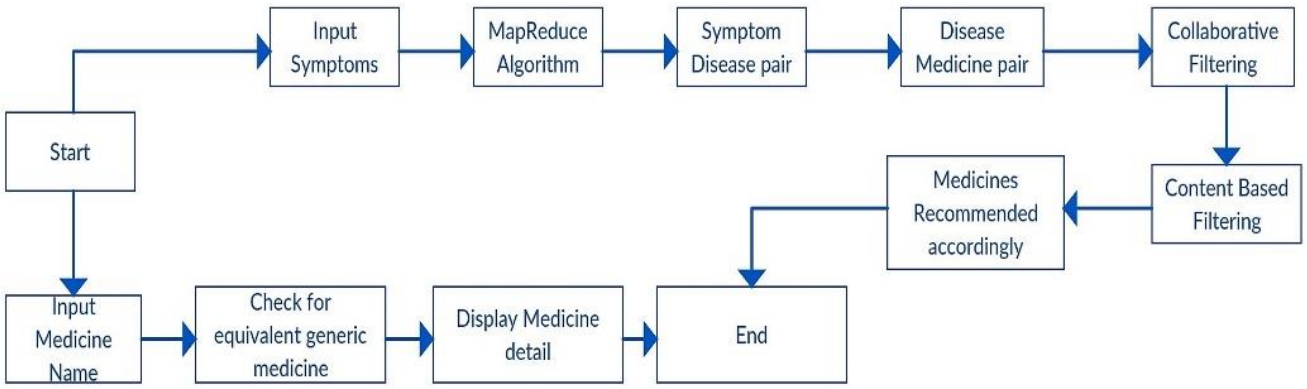


Fig 4.1.1

4.1.2 Use Case Diagram

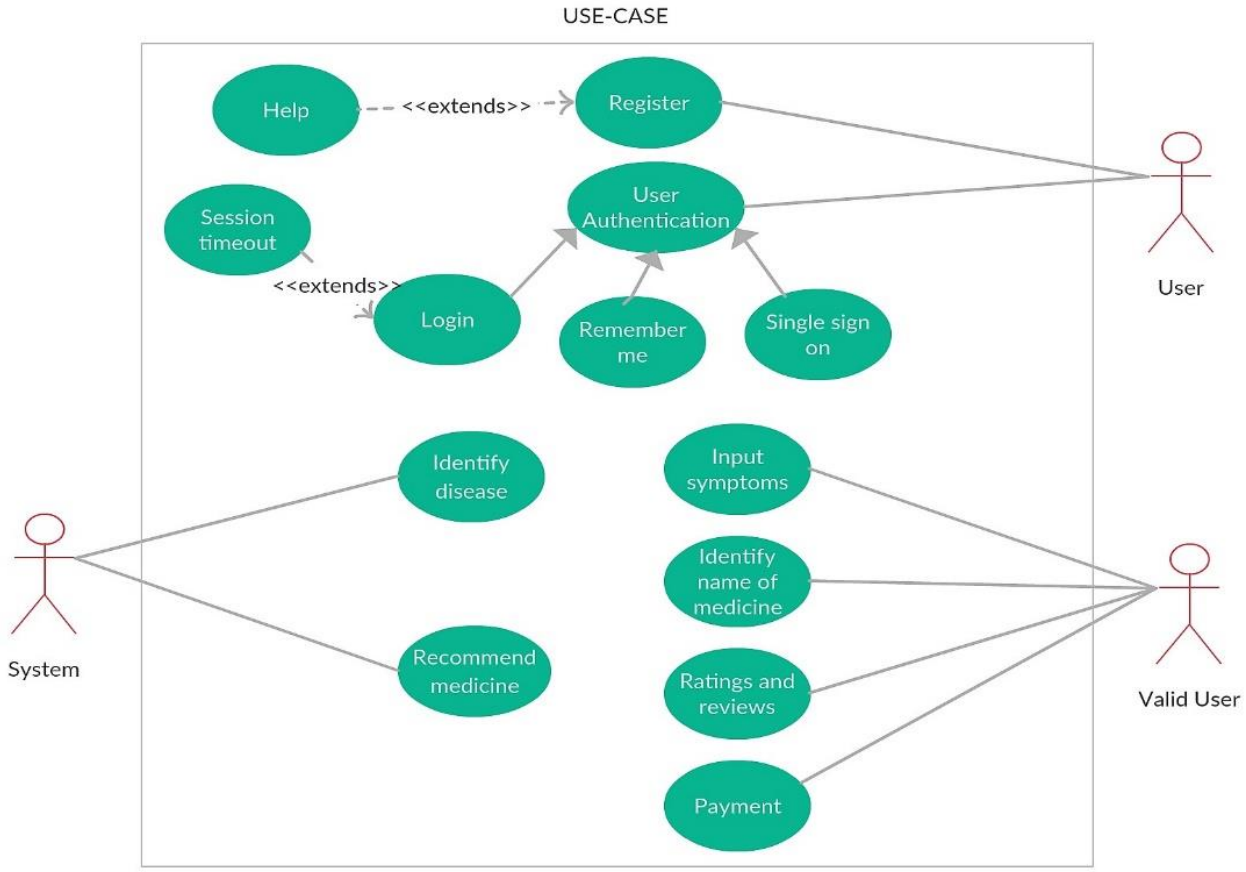


Fig 4.1.2-Use Case Diagram for recommending medicine

4.1.3 Activity Diagram

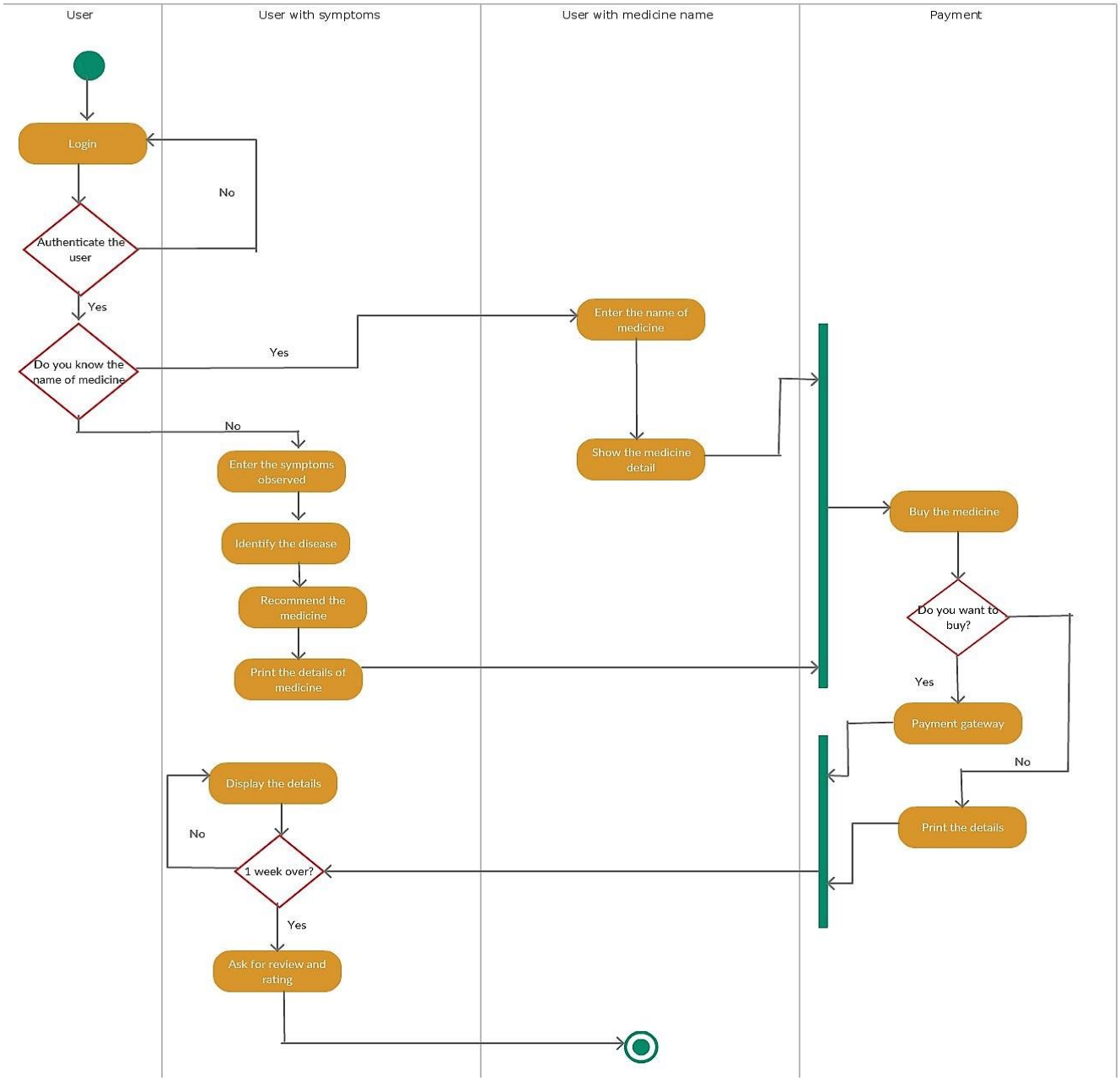


Fig 4.1.3 Activity Diagram for recommending medicine

4.1.4 Class Diagram

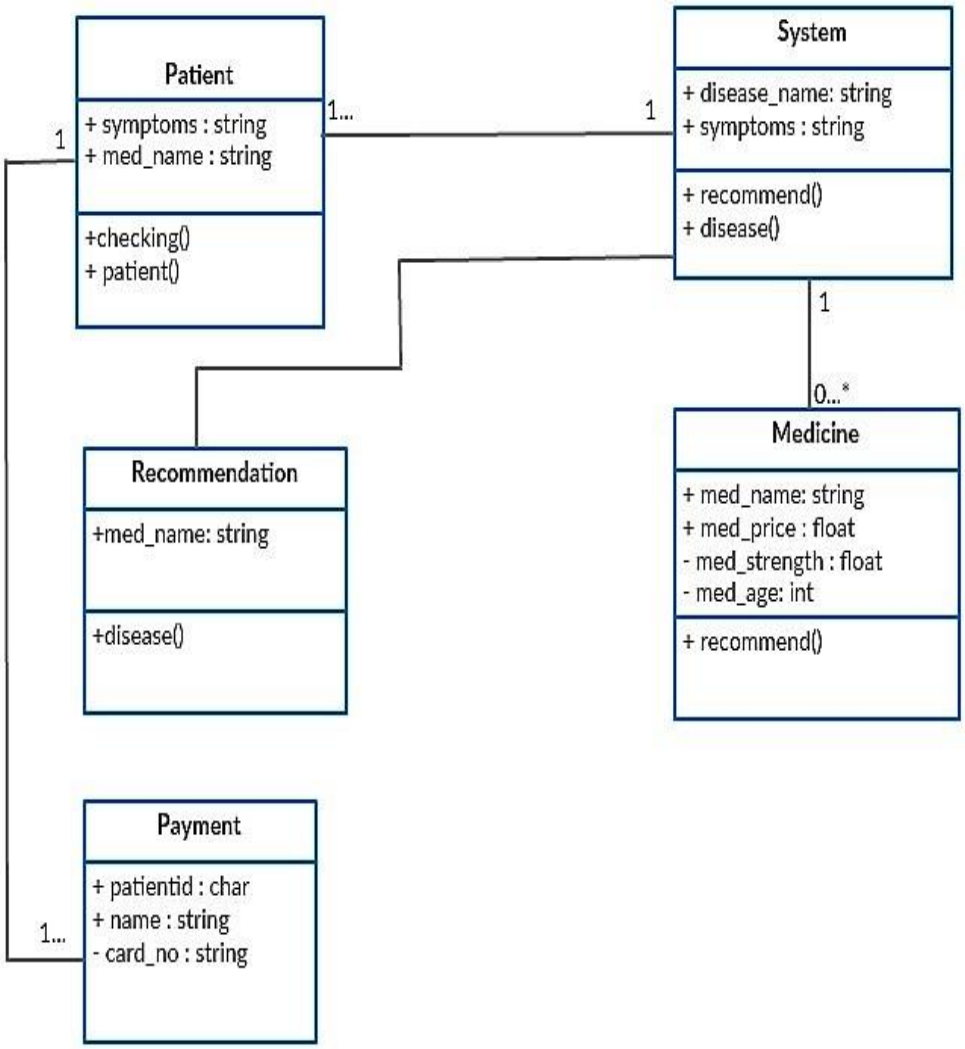


Fig 4.1.4 Class Diagram for recommending medicine and data analysis.

4.1.5 Sequence Diagram

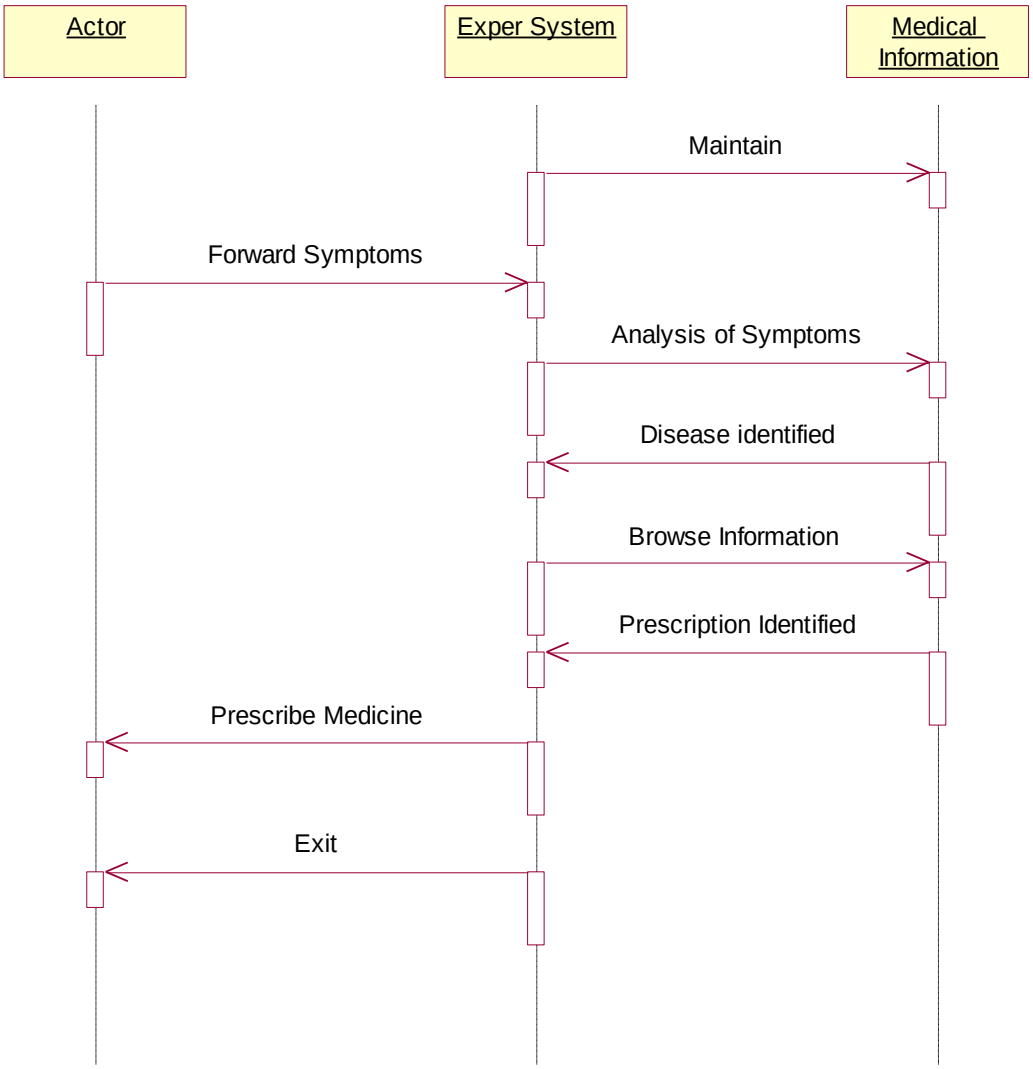


Fig 4.1.5 Sequence Diagram for identifying disease

4.1.6 State Chart Diagram

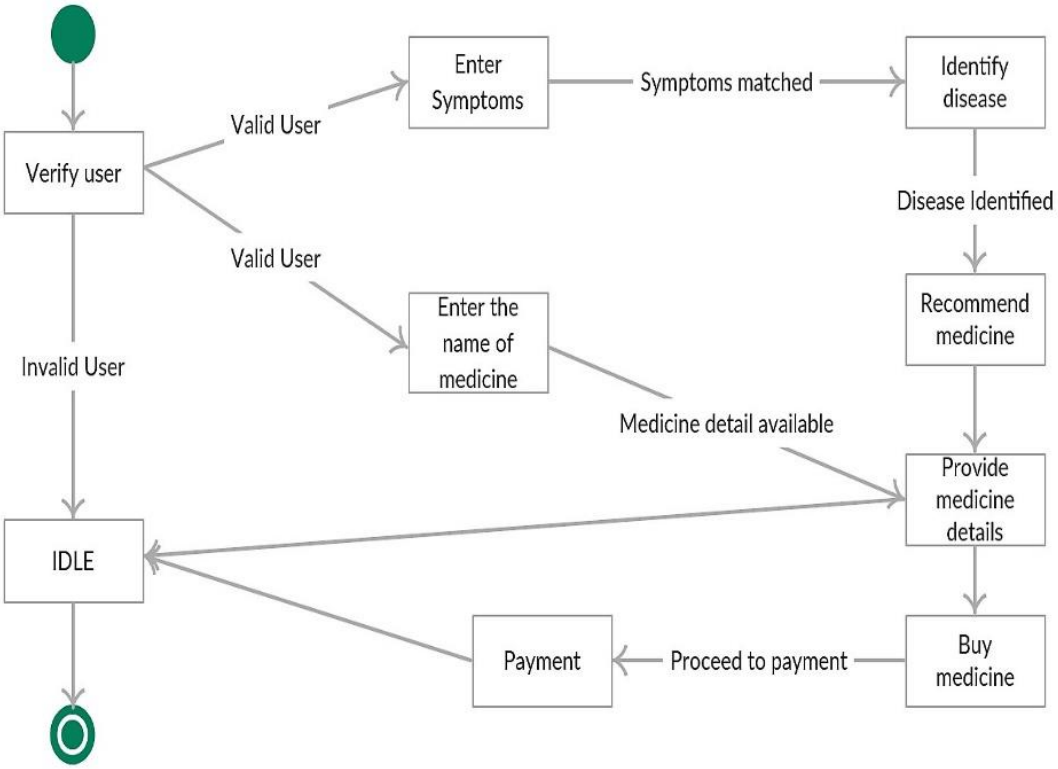


Fig 4.1.6 State Chart Diagram for recommending medicine

4.1.7 Data Flow Diagram

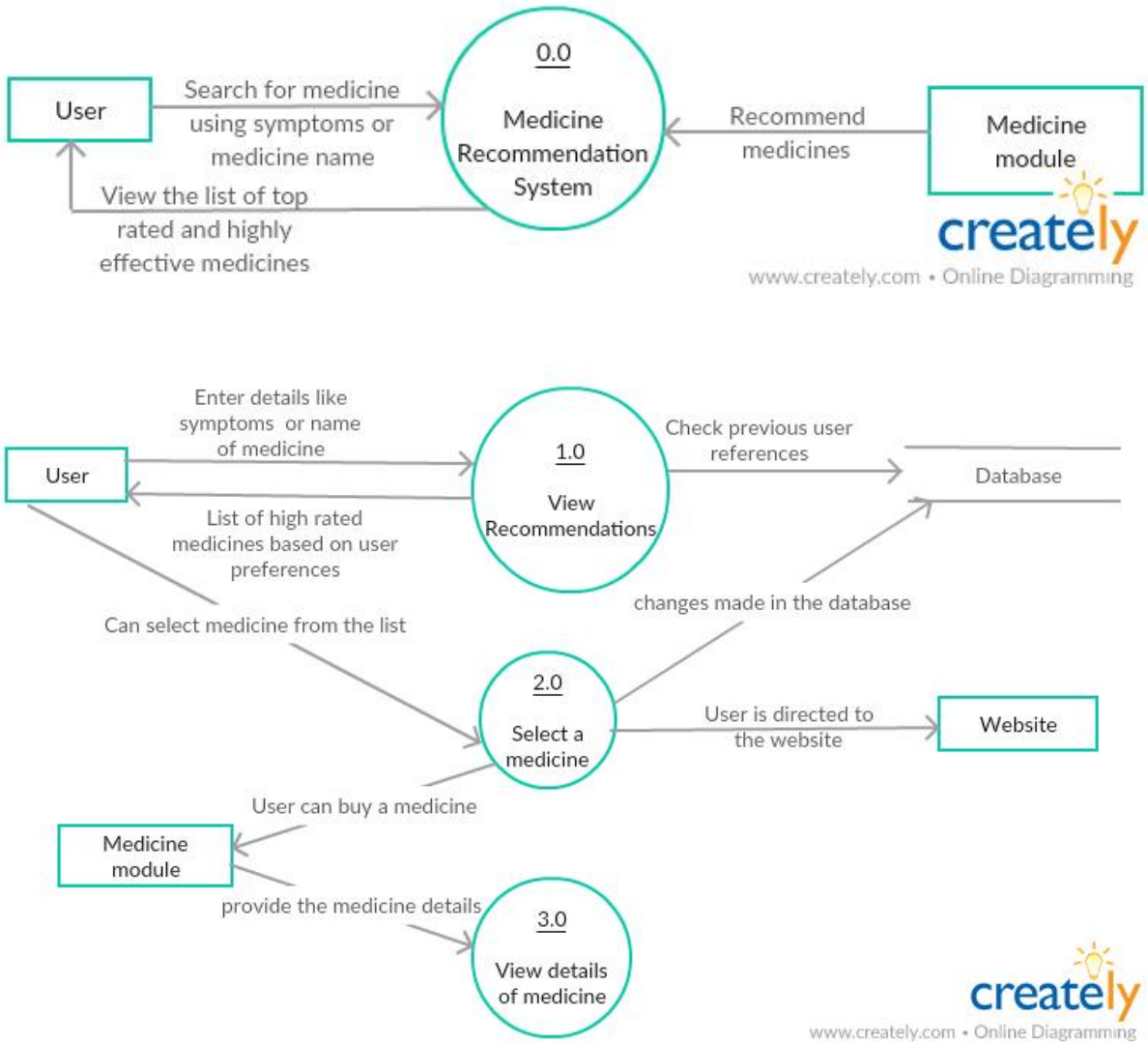
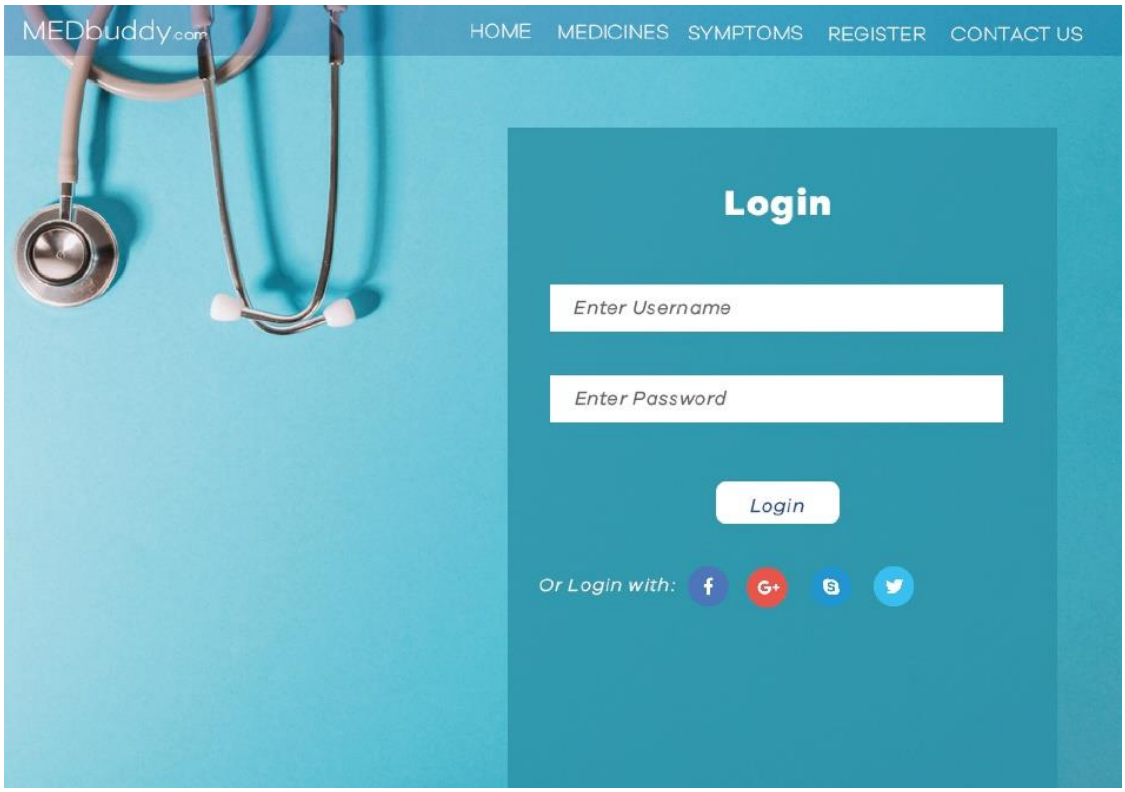


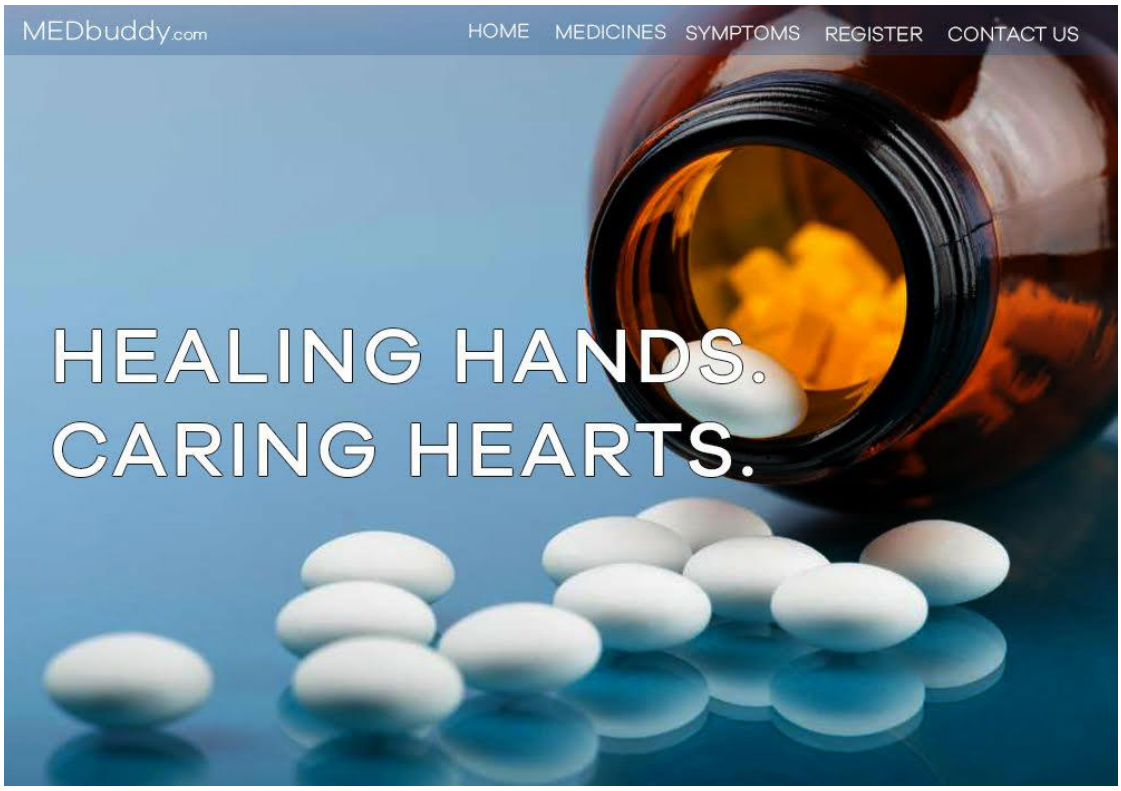
Fig 4.1.7 Data Flow Diagram for view recommendations

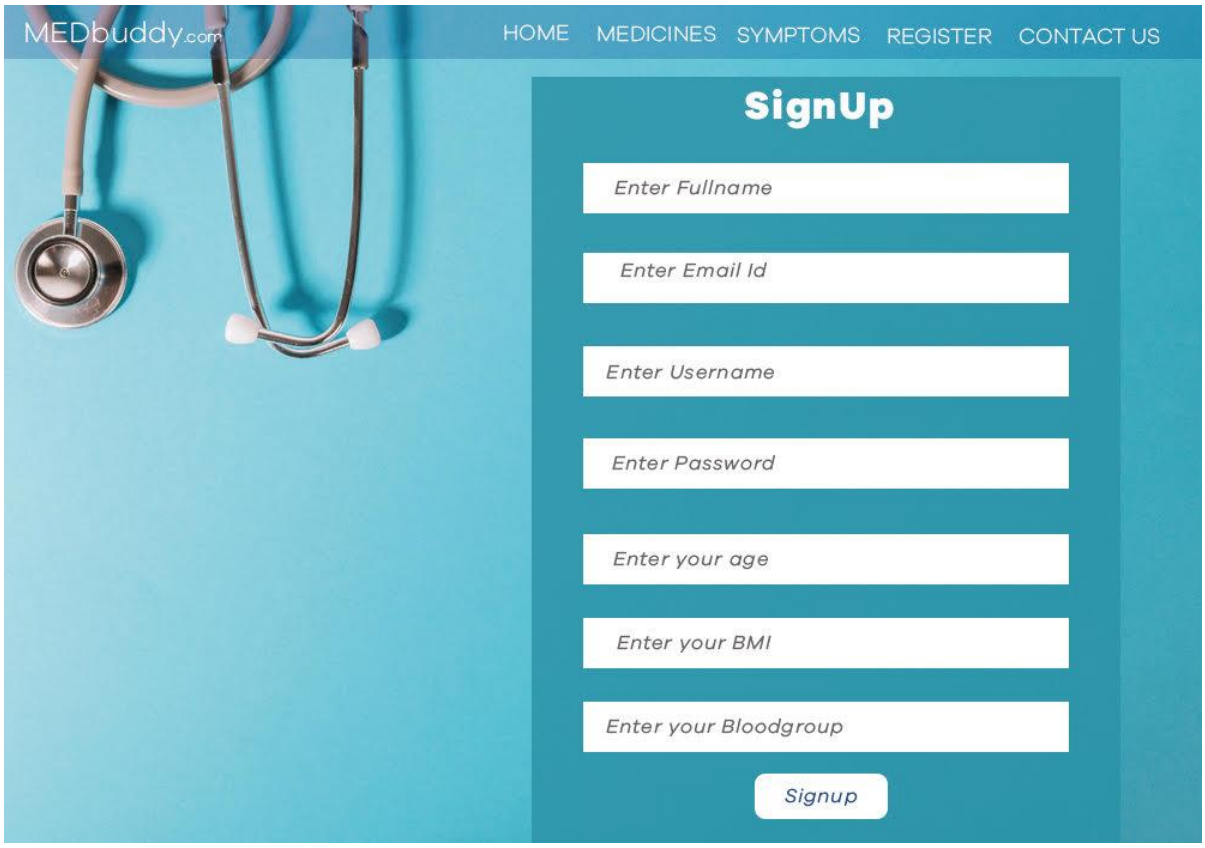
4.2 DESIGN CONSIDERATIONS:

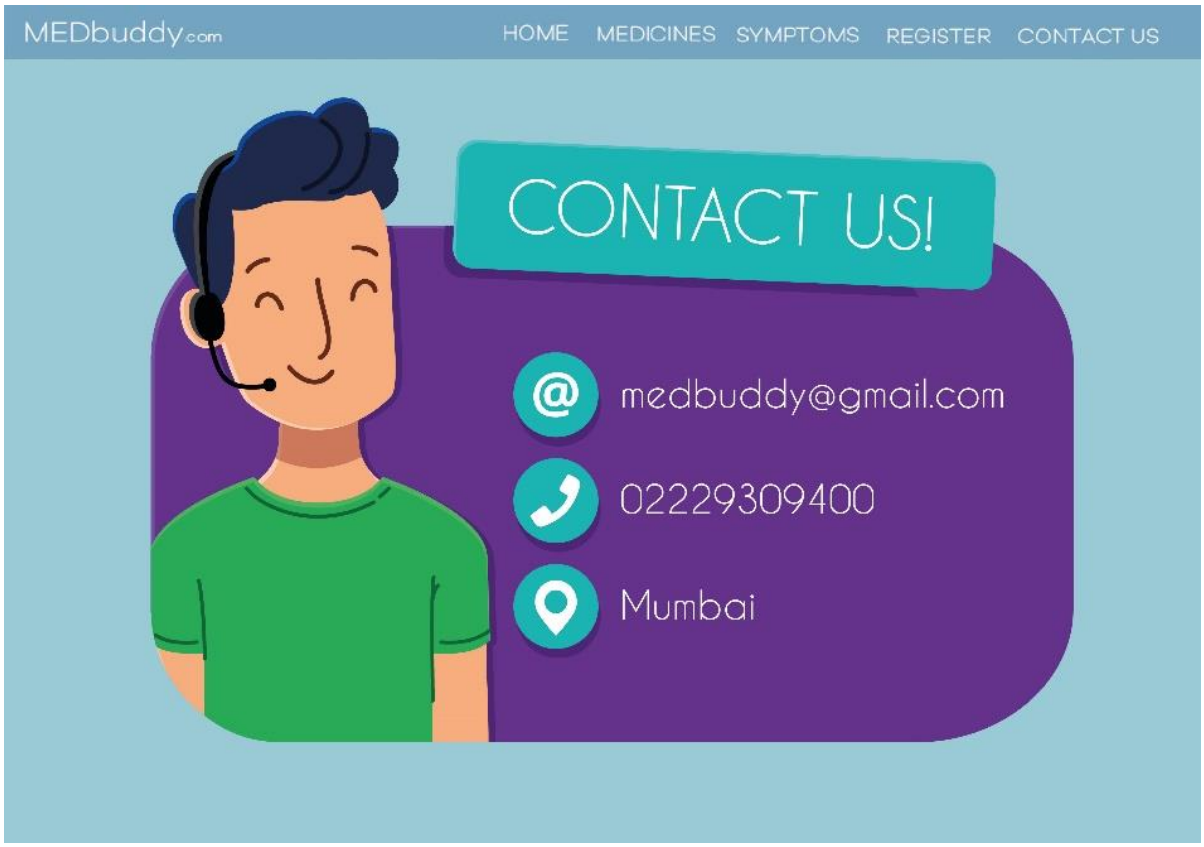
- **Technology**: The technology is fast and scalable so that there are no bottlenecks in the future. Server speed and resources are optimised so that users can get an immediate response to their requests and there are no latencies.
- **Target audience**: The web app is designed and developed keeping in mind the target audience for our platform. The look and feel of the entire platform is simple, minimalistic and clean. All the design elements used have been beta tested for user feedback.
- **Navigability**: The entire platform has been designed to help the user navigate easily and quickly. Visual aids have been used wherever possible and content has been arranged in a linear fashion so that the user can understand the flow.
- **Usability**: The platform uses the principles of SPA (Single Page Application) so that all page's load quickly. The website response time has been optimised for less than 3s so that users can get access to information fast and so that they don't have to wait for too long.

4.3 GUI Design






MEDbuddy.com
HOME
MEDICINES
SYMPTOMS
REGISTER
CONTACT US
SignUp
Enter Fullname
Enter Email Id
Enter Username
Enter Password
Enter your age
Enter your BMI
Enter your Bloodgroup
Signup


MEDbuddy.com
HOME
MEDICINES
SYMPTOMS
REGISTER
CONTACT US
CONTACT US!
medbuddy@gmail.com
02229309400
Mumbai

MEDbuddy.com HOME MEDICINES SYMPTOMS REGISTER CONTACT US

Enter your symptoms

Vomiting, Diarrhea, Muscle Pain, Headache, Fever

Disease = MALARIA

Medicines
1) Chloroquine
2) Doxycycline
3) Quinine
4) Mefloquine

Suggestions

Pramod had almost same health card and his preferred medicine was Quinine.

MEDbuddy.com HOME MEDICINES SYMPTOMS REGISTER CONTACT US

Enter the medicine name

Glucophage

RESULTS

-GLUCOPHAGE
Disease: Diabetes

Side Effects: Muscle Pain or Weakness, Stomach Pain, Vomiting

Dosage: 500 mg twice a day

Generic Medicine : Metformin

Chapter 5

Implementation

5.1 Plan for Implementation

Work Done this Semester

1. Studying the available solution.
2. Improvement of the available solution by introducing real time environment situation.
3. System & UML diagrams.
4. Requirement gathering.
5. Developing a methodology and GUI of the application.

TIMELINE:

Table 5.1 Timeline

FROM	TO	WORK TO BE DONE
16th Oct,2017	30th Oct,2017	Developing sign up page for users.
18th Dec,2017	24th Dec,2017	Developing a search method for medicines
25th Dec,2017	30th December,2017	Developing a search engine which uses filtering techniques
8th Jan,2018	14th Jan,2018	Incorporating MapReduce Techniques and testing the system for maximum accuracy
15th Jan,2018	28th Jan,2018	Deploying the application on website.
29th Jan,2018	5th Feb.2018	Detailing for finalized project.

CHAPTER 6

Conclusion

We have successfully implemented Medicine Recommendation System based on Big Data and Machine Learning approaches. This system deals with personalized suggestions by analysing user preferences and provides the user with the information of medicines that are highly recommended so as the user is completely satisfied with the variety of choices which is been received. The system provides not just one recommendation, but to review the predictions and return the appropriate medicine for the given symptoms with utmost perfection. We use multiple models and techniques to arrive at our best solution.

References

IEEE Standard paper:

1. Kumbhar, A. A., Wanjara, K. S., Trivedi, D. H., Khairatkar, A. U., & Sharma, D. (2014). “Automated Attendance Monitoring System using Android Platform.”, International Journal of Current Engineering and Technology, Vol. 4, No. 2, pp1096-1099.

2. Bandra.U,Bandra.P,”Tagciti:A practical approach for location-aware and socially relevant information creation and discovery for mobile users,”IEEE International symposium on Wireless communication systems,Reykjavik,pp.118-122,2008.

3. Gupta.A.kumar,S.Qadeer,M.A.,”Location based services using android(LBSOID),” IEEE International conference on Multimedia services architecture and applications,pp 1-5,2009.

4. Nitin Agarwal, Ehtesham Haque, Huan Liu, and Lance Parsons, "Research Paper Recommender Systems: A Subspace Clustering Approach, " WAIM 2005, LNCS 3739, pp. 475-491, 2005.

5. Joseph Kramer, Sunil Noronha, John Vergo, "A User- centred design approach to personalization " Communications of the ACM, vo.43, Issue 8, pp.44-48, 2000.

Acknowledgement

We have taken efforts in this project. However, it would not have been possible without the kind support and help of our guide. We would like to extend our sincere thanks to all of them. We are highly indebted to Ms. Anagha Durugkar for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project. We would also like to express our gratitude towards our H.O.D Dr.Tanuja Sarode for her kind co-operation and encouragement which help us in completion of this project.