DAYANANDA SAGAR UNIVERSITY

KUDLU GATE, BANGALORE – 560068

Bachelor of Technology
in
COMPUTER SCIENCE AND ENGINEERING

Major Project Phase-II Report

SMART HEALTHCARE DISEASE PREDICTION SYSTEM

By

Akshit Kumar – ENG18CS0029
Anirban Saha – ENG18CS0037
Arjun Upadhayay – ENG18CS0045
Kanishk Raj – ENG18CS0124
Kinshuk Kishore – ENG18CS0135

Under the supervision of

Prof. Shwetha G.S.
Assistant Professor, CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY,

(2021-2022)

DAYANANDA SAGAR UNIVERSITY

School of Engineering
Department of Computer Science & Engineering

Kudlu Gate, Bangalore – 560068
Karnataka, India

CERTIFICATE

This is to certify that the Phase-II project work titled **"SMART HEALTHCARE DISEASE PREDICTION SYSTEM"** is carried out by **Akshit Kumar (ENG18CS0029), Anirban Saha (ENG18CS0037), Arjun Upadhayay (ENG18CS0045), Kanishk Raj (ENG18CS0124), Kinshuk Kishore (ENG18CS0135),** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2021-2022**.

Prof Shwetha G.S.	**Dr Girisha G S**	**Dr. A Srinivas**
Assistant Professor Dept. of CS&E, School of Engineering Dayananda Sagar University	Chairman CSE School of Engineering Dayananda Sagar University	Dean School of Engineering Dayananda Sagar University
Date:	Date:	Date:

Name of the Examiner	**Signature of Examiner**
1.	
2.	

DECLARATION

We, **Akshit Kumar (ENG18CS0029), Anirban Saha (ENG18CS0037), Arjun Upadhayay (ENG18CS0045), Kanishk Raj (ENG18CS0124), Kinshuk Kishore (ENG18CS0135),** are student's of the seventh semester B.Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the phase-II project titled **"SMART HEALTHCARE DISEASE PREDICTION SYSTEM"** has been carried out by us and submitted in partial fulfillment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2021-2022**.

Student **Signature**

Name1: Akshit Kumar

USN: ENG18CS0029

Name2: Anirban Saha

USN: ENG18CS0037

Name3: Arjun Upadhayay

USN: ENG18CS0045

Name4: Kanishk Raj

USN: ENG18CS0124

Name5: Kinshuk Kishore

USN: ENG18CS0135

Place : Bangalore

Date :

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank **Dr. A Srinivas. Dean**, **School of Engineering & Technology**, **Dayananda Sagar University** for his constant encouragement and expert advice.It is a matter of immense pleasure to express our sincere thanks to **Dr. Girisha G S, Department Chairman**, **Computer Science, and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our guide **Prof. Shwetha G S**, **Assistant Professor**, **Dept. of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing her valuable time to extend help in every step of our project work, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Project Coordinator Dr. Meenakshi Malhotra and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Project work.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The wide adaptation of computer-based technology in the health care industry resulted in the accumulation of electronic data. Due to the substantial amounts of data, patients are facing challenges to analyze symptoms accurately and identify diseases before visiting a doctor. However, Intelligent data mining (DM) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the detection of diseases. In this literature, the aim is to predict health disease according symptoms stored in database. The most prominently discussed supervised Data Mining algorithm is Naïve Bayes (NB). As per findings, Naïve Bayes (NB) is most effective in predicting the disease in real time as it is faster to through large dataset.

CHAPTER 1

INTRODUCTION

It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason. The Health Prediction system is an end user support and online consultation project. Here we propose a system that allows users to get instant guidance on their health issues through an intelligent health care system online. The system is fed with various symptoms and the disease/illness associated with those systems. The system allows user to share their symptoms and issues. It then processes user's symptoms to check for various illnesses that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's symptoms.

In doctor module when doctor login to the system doctor can view his patient details and the report of that patient. Doctor can view details about the patient search what patient searched for according to their prediction. Doctor can view his personal details. Admin can add new disease details by specifying the type and symptoms of the disease into the database. Based on the name of the disease and symptom the data-mining algorithm works. Admin can view various disease and symptoms stored in database. This system will provide proper guidance when the user specifies the symptoms of his illness.

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed decisions. Data mining is a key part of data analytics overall and one of the core disciplines in data science, which uses advanced analytics techniques to find useful information in data sets. At a more granular level, data mining is a step in the knowledge discovery in databases (KDD) process, a data science methodology for gathering, processing and analyzing data. Data mining and KDD are sometimes referred to interchangeably, but they're more commonly seen as distinct things.

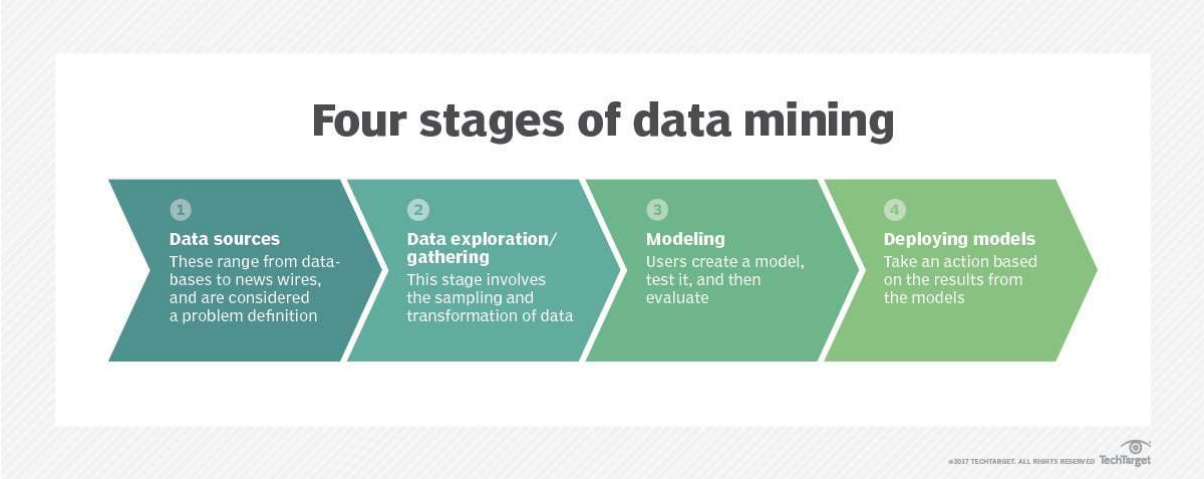


Fig 1.1 Stages of Data Mining

Naive Bayes is a set of supervised learning algorithms based on the Bayes' theorem with the "naïve" assumption of independence between every pair of features. Despite its simplicity, it often outperforms more sophisticated classification methods.

If there are input variables x and output variable y, Bayes' theorem states the following relationship.

$p(y|x) = p(y).p(x|y)/ p(x)$

In this project, Gaussian Naïve Bayes algorithm has been implemented. In case of Gaussian Naïve Bayes, the likelihood of the features us assumed to be Gaussian i.e. all continuous values x associated with class y are distributed according to Gaussian distribution.

Given a continuous attribute x in training data, the data is first segmented by the class y. Then, the mean and variance of x in each class is computed. If μ be the mean of the values in x associated with class y, then let d2 be the variance of the values in x associated with class y. Suppose there is some observation value v then, the probability distribution of v given by class y, $p(x=v \mid y)$, can be computed by plugging into the equation for a normal distribution.

CHAPTER 2

PROBLEM DEFINITION

Prediction of health disease may seem tricky, but this is part of user service system (application support direct contact with user). The core idea behind the project is to propose a system that allows users to get instant guidance on their health issues. This system is fed with various symptoms and the disease/illness associated with those systems. This system allows user to share their symptoms and issues It then processes user's symptoms to check for various illnesses that could be associated with it If the system is not able to provide suitable results, it informs the user about the type of disease or disorder it feels user's symptoms are associated with and also suggest the doctor to whom he or she can contact.

CHAPTER 3

LITERATURE SURVEY

Smart Health Prediction System with Data Mining (2020)

This study proposes the application of data mining algorithm for health prediction that can eventually shape a suitable health prediction system for patients. Even though, some hospitals are well equipped to provide the best healthcare services to its citizens, some of the hospitals are still lack in certain qualities. Consequently, patients are doubtful and uncertain when it comes to picking which hospital suits them. Numerous issues are faced by patients pertinent to hospitals such as, being unable to provide medical services, insufficient number of qualified medical staffs, poor communication between doctors and patients, and unorganized health records and data. Eventually, these issues impede the opportunity for hospitals to handle both their management and their duties steadily in order to maintain the health of every citizen and community. Therefore, application of data mining in health prediction is considered in this paper as the best practice to facilitate better healthcare system.

Smart Health Care System using Data Mining (2018)

The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Data mining has been a current trend for attaining diagnostic results. The health management system is an end user support and online consultation project. Here we propose a system that allows users to get guidance on their health issues through an intelligent health care online system. The objective of our paper is to predict Chronic Kidney Disease (CKD), Heart Disease and Liver Disease using clustering technique, K-means algorithm.

Data mining and visualization for prediction of multiple diseases in healthcare (2017)

This paper highlights the application of classifying and predicting a specific disease by implementing the operations on medical data generated in the field of medical and healthcare. In this project an efficient multiclass Naïve Bayes algorithm is used for

prediction of a particular disease by training it on a set of data before implementation. The proposed system can solve difficult queries for detecting a particular disease and also can assist medical practitioners to make smart clinical decisions which traditional decision support systems were not able to. The decisions taken by medical practitioners with the help of technology can result in effective and low cost treatments.

Survey on Big Data Analytics in Health Care (2019)

Massive amount of data in different forms need to be handled in any healthcare applications. Because of its significance, there is need of developing efficient and better performing algorithms, techniques and tools to analyze medical big data. Whereas, the traditional algorithms are not capable for analyzing such complex data. We discussed about big data analytics in major healthcare areas like EHR maintenance, disease diagnose, prediction of emergency condition of patients, etc... Also stated different machine algorithms usage in disease diagnose and patient's data analysis and discussed about importance of various machine learning algorithms.

Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach (2019)

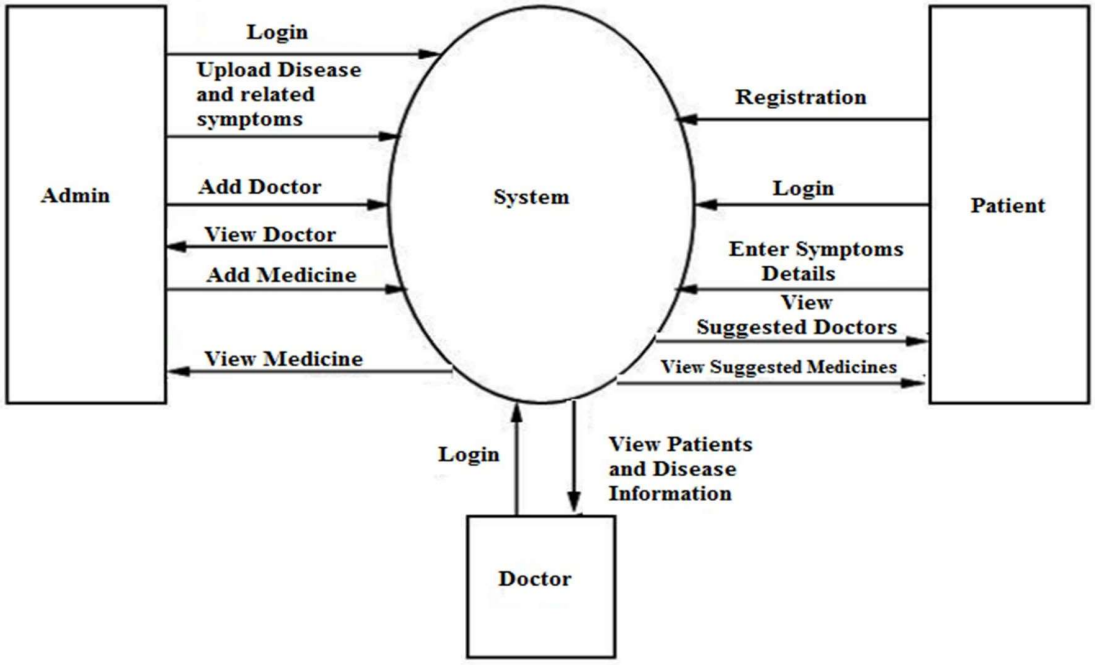
Urban living in modern large cities has significant adverse effects on health, increasing the risk of several chronic diseases. Here, they focused on the two leading clusters of chronic diseases, heart disease and diabetes, and develop data-driven methods to predict hospitalizations due to these conditions. They base these predictions on the patients' medical history, recent and more distant, as described in their Electronic Health Records (EHRs) and formulated the prediction problem as a binary classification problem and consider a variety of machine learning methods, including kernelized and sparse Support Vector Machines (SVMs), sparse logistic regression, and random forests. They validated the algorithms on large data sets from the Boston Medical Center, the largest safety-net hospital system in New England.

CHAPTER 4

PROJECT DESCRIPTION

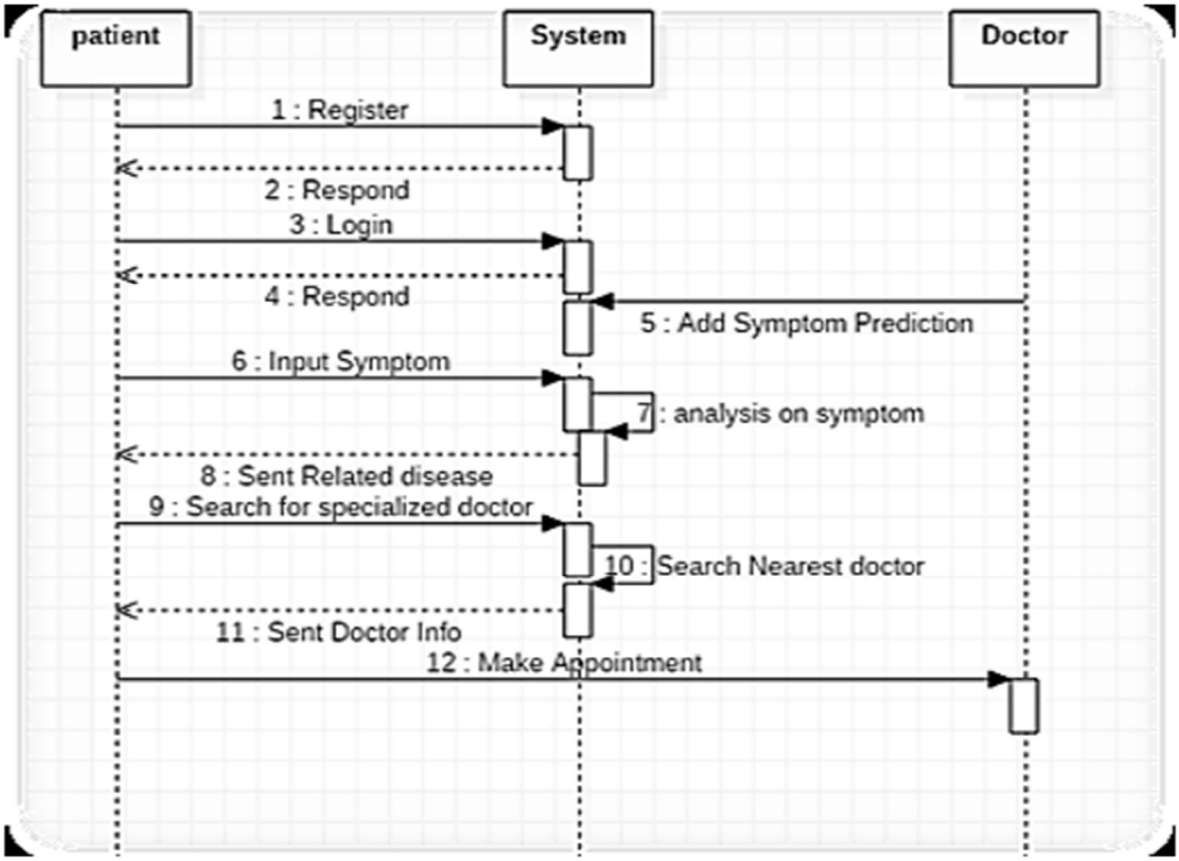
To beat the downside of existing framework we have created smart health prediction System. We have built up a specialist framework called Smart Health Prediction framework, which is utilized for improving the task of specialists. A framework checks a patient at initial level and proposes the possible diseases. It begins with getting some information about manifestations to the patient, in the event that the framework can distinguish the fitting sickness, at that point it proposes a specialist accessible to the patient in the closest conceivable territory. On the off chance that the framework isn't sufficiently sure, it asks few questions to the patients, still on the off chance that the framework isn't sure; at that point it will show a few tests to the patient. In light of accessible total data, the framework will demonstrate the result. Here we utilize some intelligent methods to figure the most precise disorder that could be associated with patient's appearances and dependent on the database of a couple of patient's restorative record, calculation (Naïve Bayes) is connected for mapping the side effects with conceivable diseases. This framework improves undertaking of the specialists as well as helps the patients by giving vital help at a soonest organize conceivable.

4.1 PROPOSED DESIGN



4.1 DATA FLOW DIAGRAM

- Patient Login: - Patient Login to the system using his ID and Password.
- Patient Registration: -If Patient is a new user he will enter his personal details and he will user Id and password through which he can login to the system.
- My Details: - Patient can view his personal details.
- Disease Prediction: - Patient will specify the symptoms caused due to his illness. System will ask certain question regarding his illness and system predict the disease based on the symptoms specified by the patient and system will also suggest doctors based on the disease.
- Search Doctor: - Patient can search for doctor by specifying name, address or type.
- Feedback:-Patient will give feedback this will be reported to the admin.
- Doctor Login: - Doctor will access the system using his User ID and Password.
- Patient Details: Doctor can view patient's personal details.
- Notification: Doctor will get notification how many people had accessed the system and what all are the diseases predicted by the system.
- Admin Login: Admin can login to the system using his ID and Password.
- Add Doctor: Admin can add new doctor details into the database.
- Add Disease: Admin can add disease details along with symptoms and type.
- View Doctor: Admin can view various Doctors along with their personal details.
- View Disease: Admin can view various diseases details stored in database.
- View Patient: Admin can view various patient details who had accessed the system.
- View Feedback: Admin can view feedback provided by various users.



4.2 Activity Diagram

4.2. ASSUMPTIONS AND DEPENDENCIES

1. We are assuming that whatever the symptoms the patient has, it is present in the database.
2. The patient has a high-speed internet.
3. We are assuming that the patient is able to identify his/her symptoms clearly.

CHAPTER 5

REQUIREMENTS

5.1. FUNCTIONAL REQUIREMENTS

Registration Process

Adding Patients: The system enables the patients to register themselves as patient.

Adding Doctors: The system enables the doctors to register and sign up with their specialty as doctors. Prediction of Diseases

Symptoms of Diseases: Every patient has to enter the symptoms they are feeling and system will predict the results and help the patient to contact the doctor according to disease.

5.2. NON FUNCTIONAL REQUIREMENTS

Security:

Logon ID: Any users who make use of the system need to hold a Logon ID and password.

Modifications: Any modifications like insert, delete, update, etc. for the database can be synchronized quickly and executed only by the site administrator.

Response Time: The system provides acknowledgment in very fast once the patient's symptoms are checked.

Capacity: The system needs to support many users. Reliability: The system needs to available all the time.

5.3. SOFTWARE REQUIREMENTS:

Technology: Python Django

IDE: VS Code/Atom

Client Side Technologies: HTML, CSS, JavaScript, Bootstrap

Server Side Technologies: Python

Data Base Server: SQLite Operating System: Microsoft Windows/Linux

CHAPTER 6

METHODOLOGY

Data collection

Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. The primary data collected from the online sources remains in the raw form of statements, digits and qualitative terms. The raw data contains error, omissions and inconsistencies. It requires corrections after careful scrutinizing the completed questionnaires. The following steps are involved in the processing of primary data. A huge volume of raw data collected through field survey needs to be grouped for similar details of individual responses.

Data Pre-Processing

Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Therefore, certain steps are executed to convert the data into a small clean data set. This technique is performed before the execution of Iterative Analysis. The set of steps is known as Data Pre-processing. It includes -

Data Cleaning

Data Integration

Data Transformation

Data Reduction

Data Pre-processing is necessary because of the presence of unformatted real-world data. Mostly real-world data is composed of -

• Inaccurate data (missing data) - There are many reasons for missing data such as data is not continuously collected, a mistake in data entry, technical problems with biometrics and much more.

• The presence of noisy data (erroneous data and outliers) - The reasons for the existence of noisy data could be a technological problem of gadget that gathers data, a human mistake during data entry and much more.

• Inconsistent data - The presence of inconsistencies are due to the reasons such that existence of duplication within data, human data entry, containing mistakes in codes or names, i.e.,

violation of data constraints and much more.

Model Testing

First, we started with collecting input as a symptom given by the user. Then for verification of the input, we matched it with different list of symptom present in the database and returned "invalid input" if the user entered wrong. Then we carried out the work by giving list of disease matched against the symptom. Then again, we check if there is any more symptoms or not. Then we predict the disease using set of unions of symptoms.

We searched for various renowned doctors across the cities in various different websites spread across the internet. Moreover, we collected the information about the doctors (like – name, specialization) and stored it in our database. We have also gathered the symptoms and diseases dataset after researching many different sites like 'kaggle.com' and more and stored it in the database after data preprocessing.

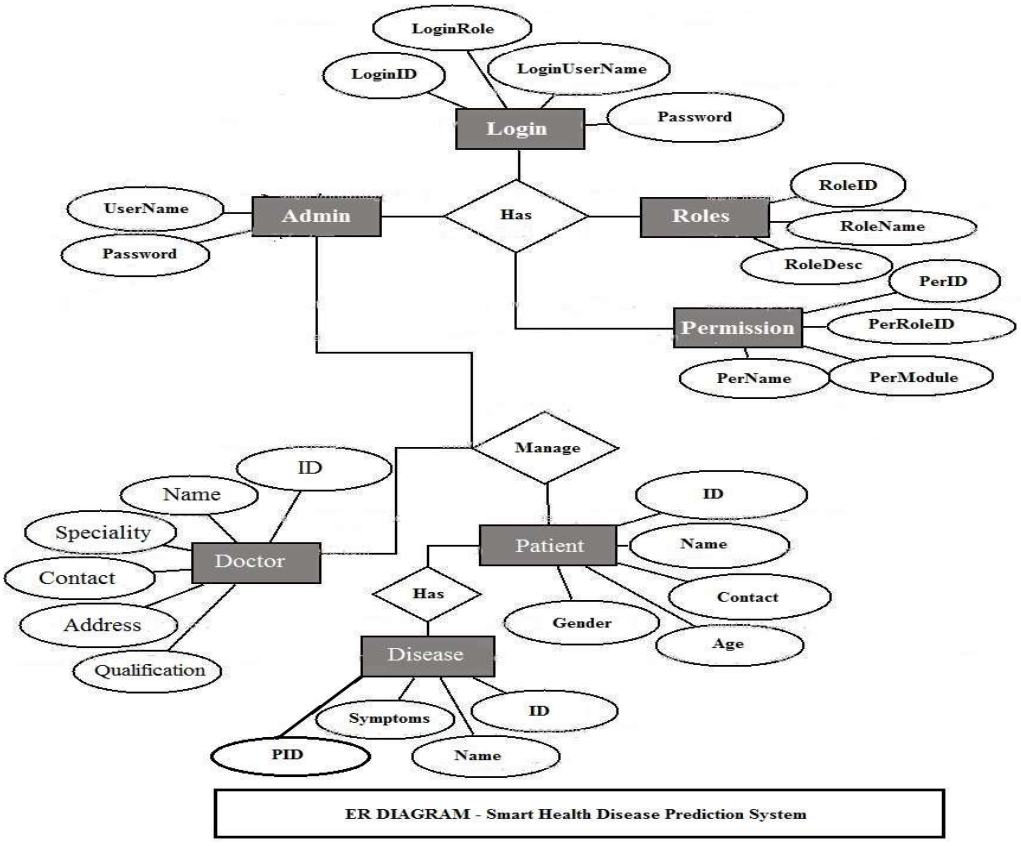


Fig:- 6.1

CHAPTER 7

EXPERIMENTATION

We started with a plan to implement our model in node js framework during the implementation we came across following issues: -

- Connectivity with the database
- Usage of Frontend framework was needed

So, to overcome the above issues we have used python Django framework which provides build in database support of SQLite. Since Django uses MVT Structure it can render frontend pages easily and there will be no need of any external frontend framework.

ALGORITHM:

Start:

Step 1: Request the user for a symptom as input.

Step 2: Check the input symptom against list of symptoms in database for validity.

Step 3: If input symptom is not a valid symptom, prompt the user with “Invalid symptom” and go to step 1.

Step 4: Use the finalized input symptom from step 3 to find possible diseases.

Step 5: If possible, disease is only one then present the user with that disease and corresponding doctor.

Step 6: Else Take union of sets of symptoms of each disease found in step 4.

Step 7: If set generated in step 6 is empty then present the user with multiple possible disease list generated in step 4 and corresponding doctor(s).

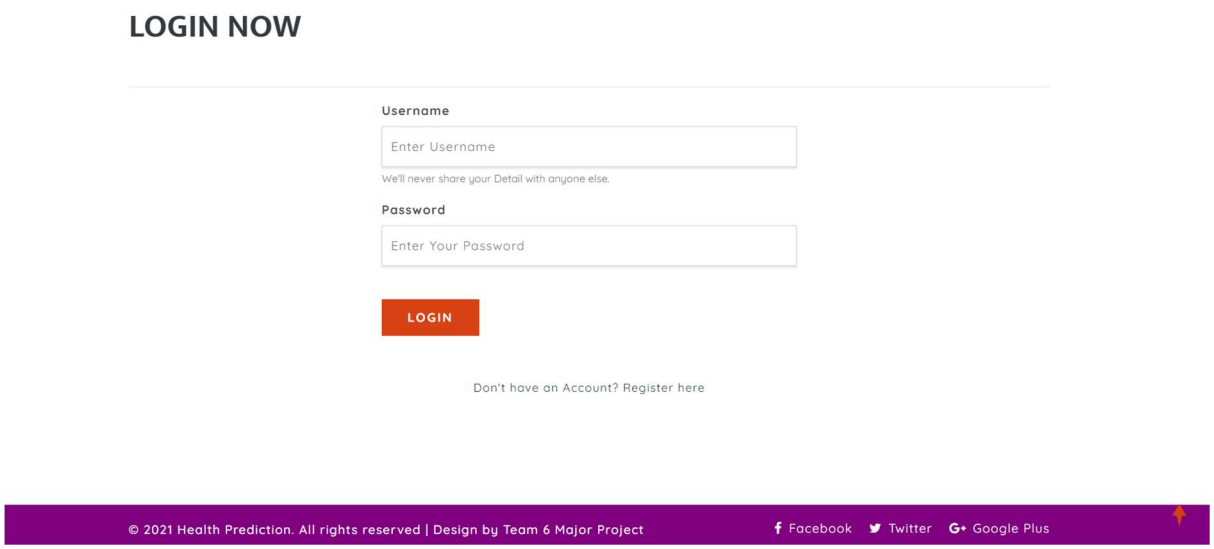
Step 8: Else Ask the user to select one more symptom from the set of symptoms generated in step 6.

Step 9: Go to step 4 with symptom selected by user in step 8 as input symptom.

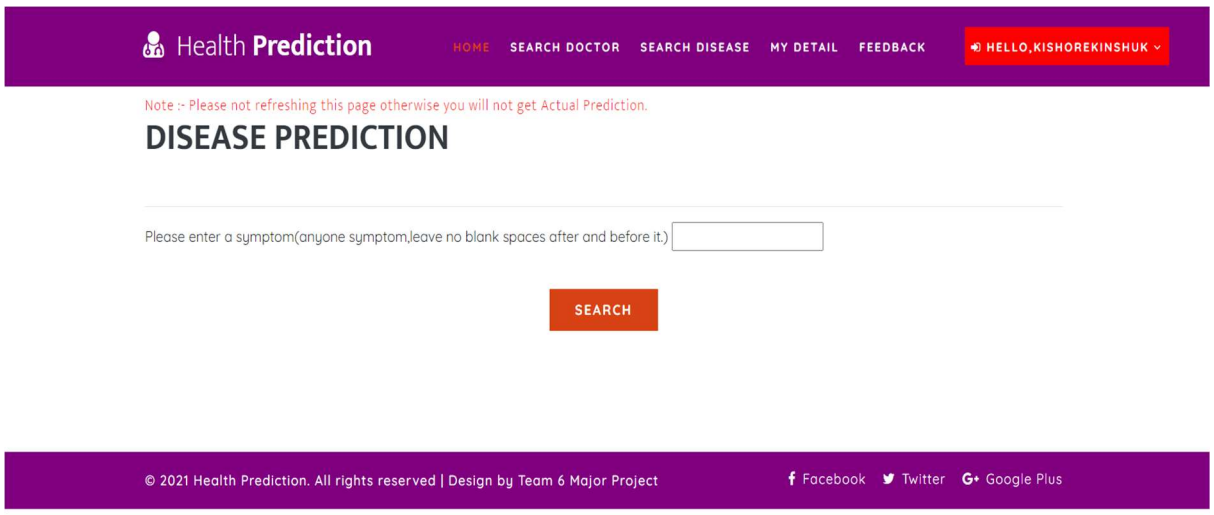
NOTE: Maximum times Loop from step 4 to 9 must not execute more than 9 times.

CHAPTER 8

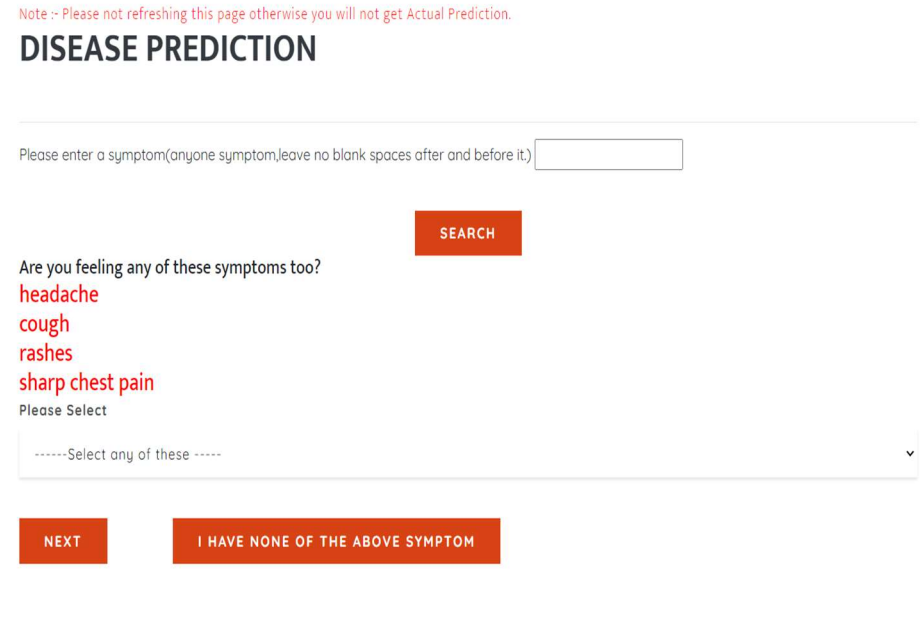
TESTING AND RESULTS



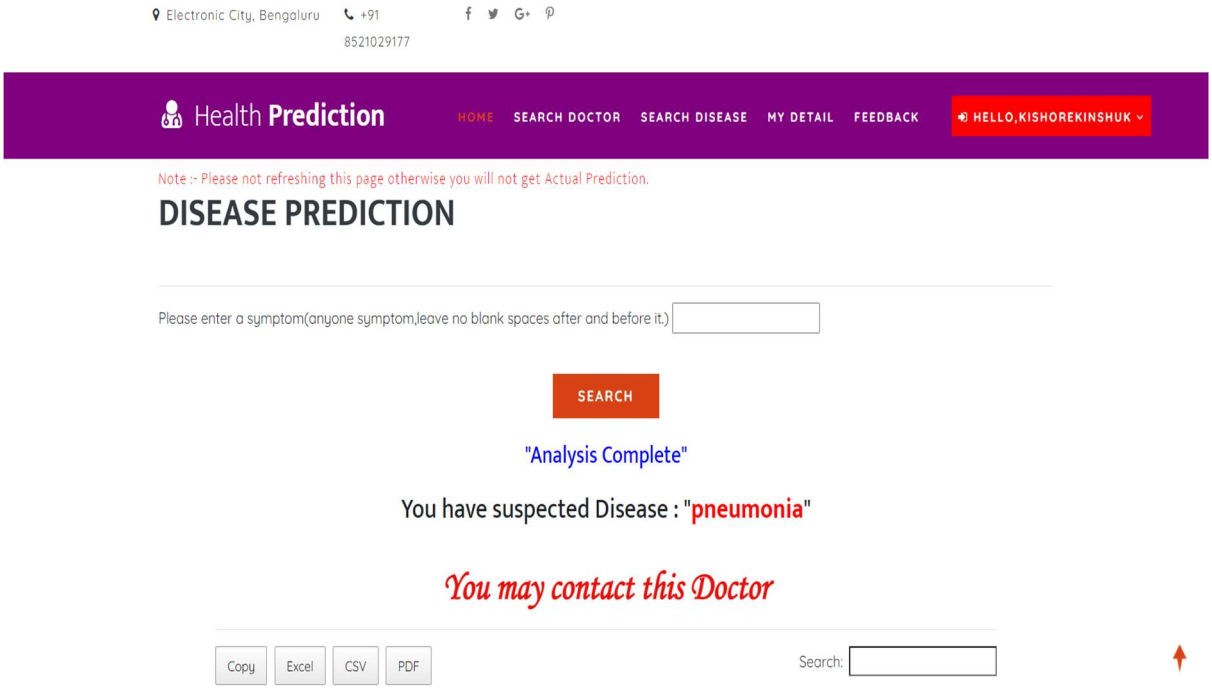
8.a Login Page



8.b Symptom Page



8.c Selection of Symptoms



8.d Predicted disease

Test Case	Input	Output
Test 1	username: Kinshuk Pass: 5692	Username & password are not matching
Test 2	Patient login:- user name : kinshukkishore Password: 12345678	Login successful
Test 3	Doctor login Username: drakshukla Password: password	Shows patients and disease information.
Test 4	Admin login Username: kinshukk Password: 12345678	Add disease, doctor, symptoms.
Test 5	Symptoms: Fever, Headache, Joint Pain.	The disease predicted is dengue.
Test 6	Symptoms: Nausea, shortness of breath.	The disease Predicted is Coronary Artery Disease.

REFERENCES

[1] V. Kirubha and S. Priya, "Survey on data mining algorithms in disease prediction," International Journal of Computer Trends and Technology (IJCTT), vol. 38, no. 3 , pp. 24-128, 2016. [Online].
Available:
https://pdfs.semanticscholar.org/0d05/05b5b057333888de85043ae9b41259f07d85.pdf.
[Accessed 8 February 2020]

[2] S. Chowdhury, "A Review Paper on Health Data Integration with Secured Record Linkage in National Health Data Warehouse and Privacy & Compatible Solution for its Privacy and Security Problems in Bangladesh and other developing countries," International Islamic University Chittagong (IIUC), 2017. [Online]. Available at: https://works.bepress.com/sayembinsarwarchowdhury/ 2/download/. [Accessed 18th November 2019]

[3] J. Tulsky, M. Beach, P. Butow, S. Hickman, J. Mack, R. Morrison, R. Street, D. White and K. Pollak, "A research agenda for communication between health care professionals and patients living with serious illness," JAMA internal medicine, vol. 177, no. 9, pp.1361-1366, 2017. [Online]. Available at:
https://sci-hub.tw/https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2635330#. [Accessed 19th November 2019],

[4] C. Kuziemsky, "Decision-making in healthcare as a complex adaptive system. Healthcare management forum," The Canadian College of Health Leaders, vol. 29, no. 1, pp. 4-7, 2016. [Online].
Available at: https://sci-hub.tw/https://doi.org/10.1177/0840470415614842#. [Accessed 20th November 2019]

[5] A. H. M. S. U. Marjia Sultana, "Analysis of Data Mining Techniques for Heart DiseasePrediction," 2018.

[6] M. I. K. ,. A. I. ,. S. Musfiq Ali, "Heart Disease Prediction Using Machine LearningAlgorithms".

[7] K. Bhanot, "towarddatascience.com," 13 Feb 2019. [Online]. Available:https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c. [Accessed 2 March 2020].[4] [Online]. Available: https://www.kaggle.com/ronitf/heart-disease-uci#heart.csv.. [Accessed05 December 2019].

[8] M. A. K. S. H. K. M. a. V. P. M Marimuthu, "A Review on Heart Disease Prediction usingMachine Learning and Data Analytics Approach".

[9] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, “Diabetes Disease Prediction Using Data Mining”, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

[10] Zhilbert Tafa, Nerxhivane Pervetica, Bertran Karahoda, “An Intelligent System for Diabetes Prediction”, 4thMediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro.

GITHUB LINK

Link Address – https://github.com/kinshukkishore/Smart-Healthcare-Disease-Prediction-System