

DWIT COLLEGE
DEERWALK INSTITUTE OF TECHNOLOGY



ECHECKUP - A DISEASE PREDICTION
SYSTEM FROM SYMPTOMS

A MINI PROJECT REPORT

Submitted to
Department of Computer Science
DWIT College

Submitted by
Nischal Badal
Class of 2020
October, 2019

DWIT College
DEERWALK INSTITUTE OF TECHNOLOGY

Supervisor's Recommendation

I hereby recommend that this project prepared under my supervision by NISCHAL BADAL entitled “**ECHECKUP - A DISEASE PREDICTION SYSTEM FROM SYMPTOMS**” in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Information Technology be processed for the evaluation.

.....

Ritu Raj Lamsal
Head, Electronics Department
Deerwalk Institute of Technology
DWIT College

DWIT College
DEERWALK INSTITUTE OF TECHNOLOGY

Student's Declaration

I hereby declare that I am the only author of this work and that no sources other than that listed here have been used in this work.

.....

Nischal Badal

October, 2019

DWIT College
DEERWALK INSTITUTE OF TECHNOLOGY

LETTER OF APPROVAL

This is to certify that this project prepared by NISCHAL BADAL entitled “**ECHECKUP - A DISEASE PREDICTION SYSTEM FROM SYMPTOMS**” in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Information Technology has been well studied. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

<p style="text-align: center;">.....</p> <p style="text-align: center;">[Supervisor]</p>	<p style="text-align: center;">.....</p> <p style="text-align: center;">[Examiner]</p>
--	--

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who provided me the possibility to complete this project. A special gratitude I give to our project supervisor, Mr. Ritu Raj Lamsal, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project and also in writing this report.

Furthermore, I would also like to acknowledge with much appreciation the crucial role of Mr Sunil Chaudhary, Mr Bhas Raj Pathak and Mr. Hitesh Karki who gave the permission to use all required equipment and the necessary materials to complete the task “Earthquake Analyzer”. I have to appreciate the guidance given by other supervisors as well as the panels especially in our project presentation that has improved our presentation skills thanks to their comment and advises.

Nischal Badal

631, Batch of 2020

October, 2019

ABSTRACT

This project is developed as an informative system which maps the possible symptoms that a person is facing into the most probable disease. It is a Machine Learning application which makes use of multiple algorithms to draw a conclusions based on the user symptoms. The user can select multiple symptoms and they are collectively directed towards one of each disease based on algorithms. Multiple predictions are made based on the data set by which the model is trained. Classification method is used and three different algorithms with different accuracy and complexity is used and a final prediction is obtained with acceptable error rate. Data is one of the major player in this project and data are accessed through a study conducted at Columbia University. This project aims to provide an open web based application to general people who can pre analyze their symptoms before visiting to a medical professional.

Keywords: *disease, classification, prediction, symptoms, application*

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....	IV
ABSTRACT.....	V
TABLE OF CONTENTS.....	VI
LIST OF FIGURES.....	VII
LIST OF ABBREVIATIONS.....	VIII
CHAPTER 1: INTRODUCTION.....	1
1.1 Overview And Motivation.....	1
1.2 Problem Statement.....	1
1.3 Objective Of The Project.....	1
1.3.1 General Objective.....	1
1.3.2 Specific Objective.....	1
1.4 Scope Of The Project.....	2
1.5 Outline.....	2
CHAPTER 2: BACKGROUND RESEARCH.....	3
2.1. Literature Review.....	3
2.2. Current System.....	3
2.3. The Problem With Current System.....	3
CHAPTER 3: SPECIFICATION AND DESIGN.....	4
3.1. Requirement Elicitation And Analysis.....	4
3.1.1. Functional Requirement.....	4
3.2.2. Non-Functional Requirement.....	4
3.3. System Design.....	4
3.3.1. Use Case Diagram.....	6
CHAPTER 4: IMPLEMENTATION AND EVALUATION.....	7
4.1. Tool And Technology.....	7
4.2. Implementation.....	7
4.3. Evaluation And Result.....	8
CHAPTER 5: CONCLUSION.....	9
CHAPTER 6: LIMITATION.....	10
REFERENCES.....	11
APPENDIX I.....	12

LIST OF FIGURES

Figure 1 Project Block Diagram..... 2

Figure 2 Work Flow diagram..... 5

Figure 3 Use Case Diagram..... 6

LIST OF ABBREVIATIONS

ML	Machine Learning
GUI	Graphical User Interface
IDE	Integrated development environment
ER	Entity Relationship

CHAPTER 1: INTRODUCTION

1.1 Overview And Motivation

With wide coverage of Internet technology today, people around the globe are familiar with search engines and internet surfing techniques. This upgrade in technology has many good aspects but some downsides also. People often tend to search every problem they have on the internet. Internet Technology have updated many fields and made the working easier.

Talking precisely about Medical field, technology is improving medical care with the use of Health wearable, Artificial organs, Telemedicines, etc. Talking about the other side of medical technology, we receive unnecessary search results in internet which makes us think once about that matter.

1.2 Problem Statement

When we type in the basic symptoms that we might be observing for few days in Google or Yahoo or any other search engines, we often get through unnecessary conclusions or predictions that mostly ends in death. Search engines are vague and provides both authorized and unauthorized results.

eCheckUp is intended to solve that particular problem by pre-analyzing the symptoms to conclude the possibility of one or more particular disease so that we can proceed into further treatment consulting medical professionals.

1.3 Objective Of The Project

1.3.1 General Objective

To implement different machine learning algorithms into a specific data set by training a model and compare the outputs of the algorithms with different accuracy rate and determine the best algorithm into the scenario.

1.3.2 Specific Objective

To develop a web application which maps your occurring symptoms into one or more possible disease which helps to minimize vague disease prediction.

To know where your symptoms are leading to and proceed with the necessary checkups and meet the specific medical professionals.

1.4 Scope Of The Project

The scope of the project are the general people who have certain recognized symptoms and are searching for any possible disease prediction. Displaying them some particular disease based on their symptoms is the main working of the project.

1.5 Outline

This outline of the project is organized as shown in block diagram in Figure 1:

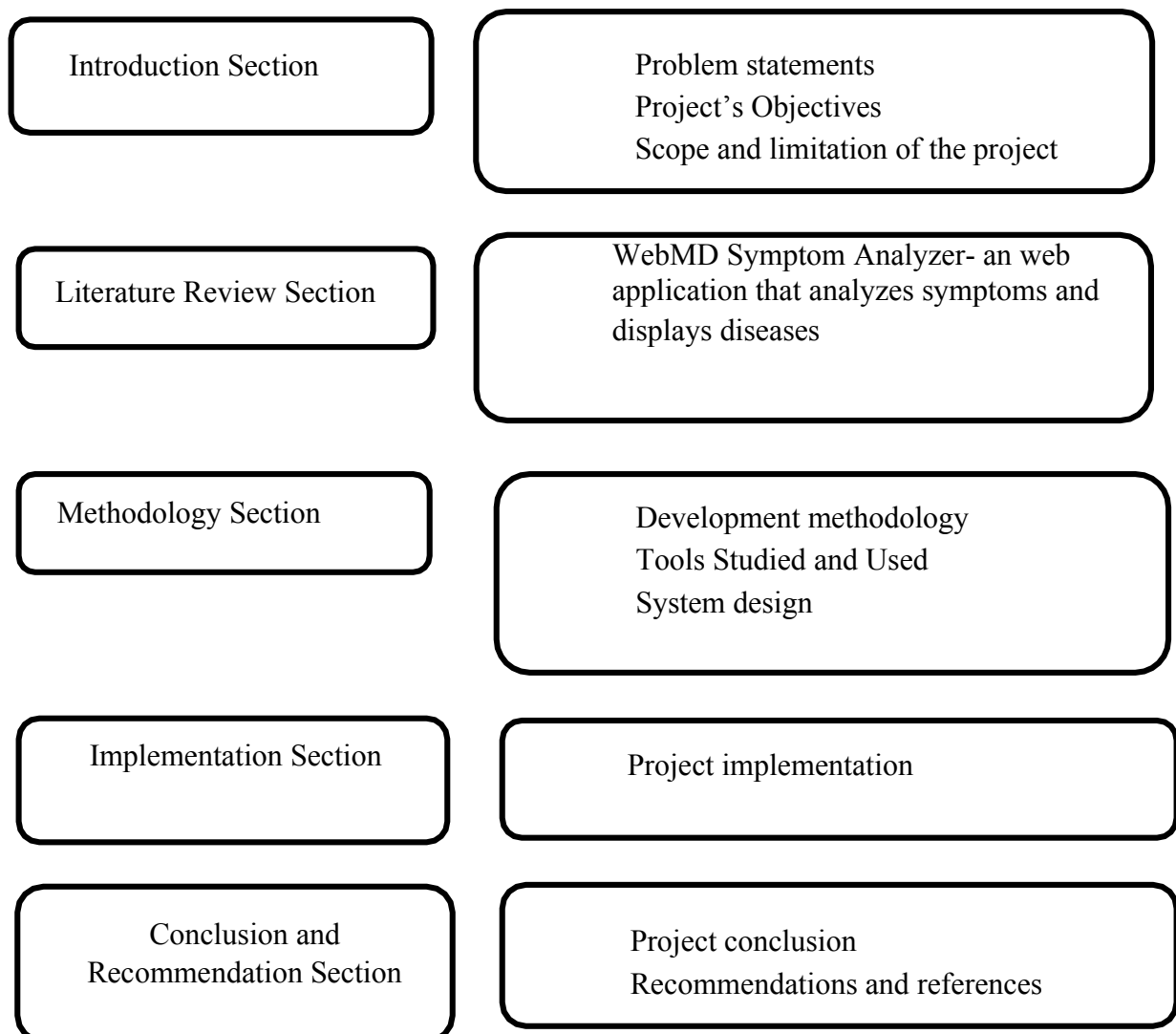


Figure 1 Project Block Diagram

CHAPTER 2: BACKGROUND RESEARCH

2.1. Literature Review

Web MD Symptom Checker (<https://symptoms.webmd.com/default.htm#/info>)

There is an online Symptom Checker application called WebMD Symptom Checker. It consists a full data set of symptoms and probable disease with conditions and treatments. eCheckUp is similar to WebMD Symptom Checker.

2.2. Current System

Talking about the current system i.e. Web MD Symptom Checker, it analyzes the symptoms of a user based on various factors as age, gender, symptoms, questions, conditions and details. It is a fully functioning application with big data set than eCheckUp. This project is a mini project inspired by the Web MD Symptom Checker.

2.3. The Problem With Current System

Since Web MD Symptom Checker is a complete working web application, there are no reported problem in it. It clearly states that it does not provide medical advice but shows some treatment methods. Web MD is in beta version and it does not show which data set it is making use of and accuracy. This is the problem with the current system but as a whole, it is a practical application with easy to use GUI and Treatment feature included.

CHAPTER 3: SPECIFICATION AND DESIGN

3.1. Requirement Elicitation And Analysis

The basic requirement of this project is a computer system which allows python applications to run. This project is based up on Python3, so the computer must support Python program to be executed successfully. Data set and training of the model is also major requirement of this project. The data set is divided into two parts: Training data set and Testing data set. The training data set is used to train the model while the testing data set is used to test the model. The accuracy obtained is approximately 95%.

3.1.1. Functional Requirement

Functional requirements of eCheckUp can be:

- The User should open the web application
- The system should ask for the Name of the user and 5 symptoms as input
- The user should click on predictions to create 3 predictions based on 3 different algorithms
- The user should be able to understand the predictions and understand them for further processing and also save the data to save the user info in the system.
- Python3 and libraries as numpy, tkinter and pandas.

3.2.2. Non-Functional Requirement

Similarly, some of the Non-functional requirements are:

- The data must be in CSV format to train the model
- The format of data must be in correct order as it is places
- The GUI must be designed correctly to provide correct output

3.3. System Design

The design of this system was easy compared to other system. This project did not contain any databases so diagrams like ER, Relational can be omitted. The system was designed using web markup languages and the processing part was done using python while GUI was made on a framework of python called as Tkinter.

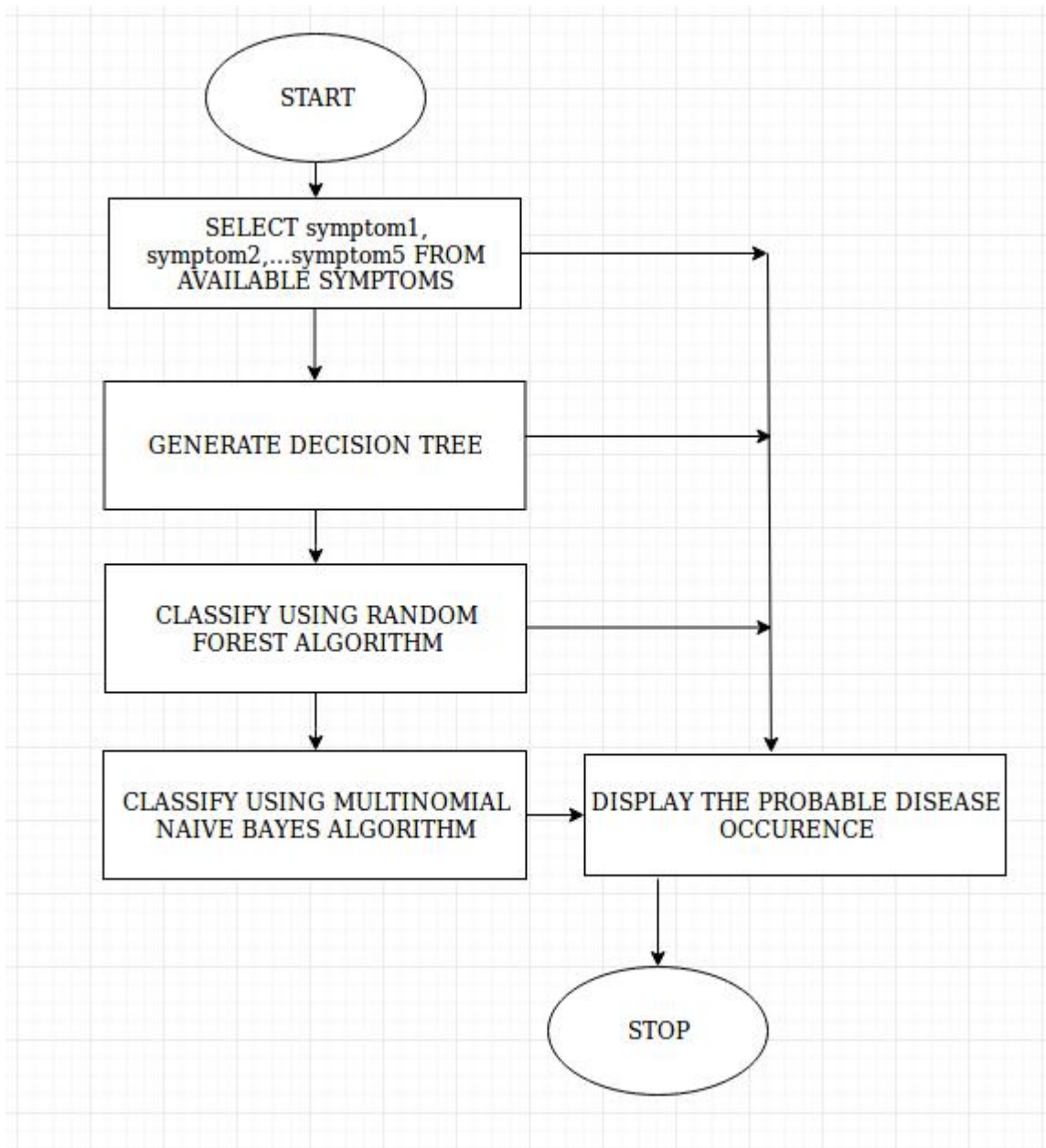


Figure 2: Work flow diagram

This is the work flow chart of the system. The user can select upto 5 symptoms from a given list of symptoms and can proceed to obtain the predictions. The program calls 3 functions: Decision Tree, Random Forest and Naive Bayes algorithms generate the outputs and the probable disease based on all of them individually is displayed. There is a save option which enters the user info into a file.

3.3.1. Use Case Diagram

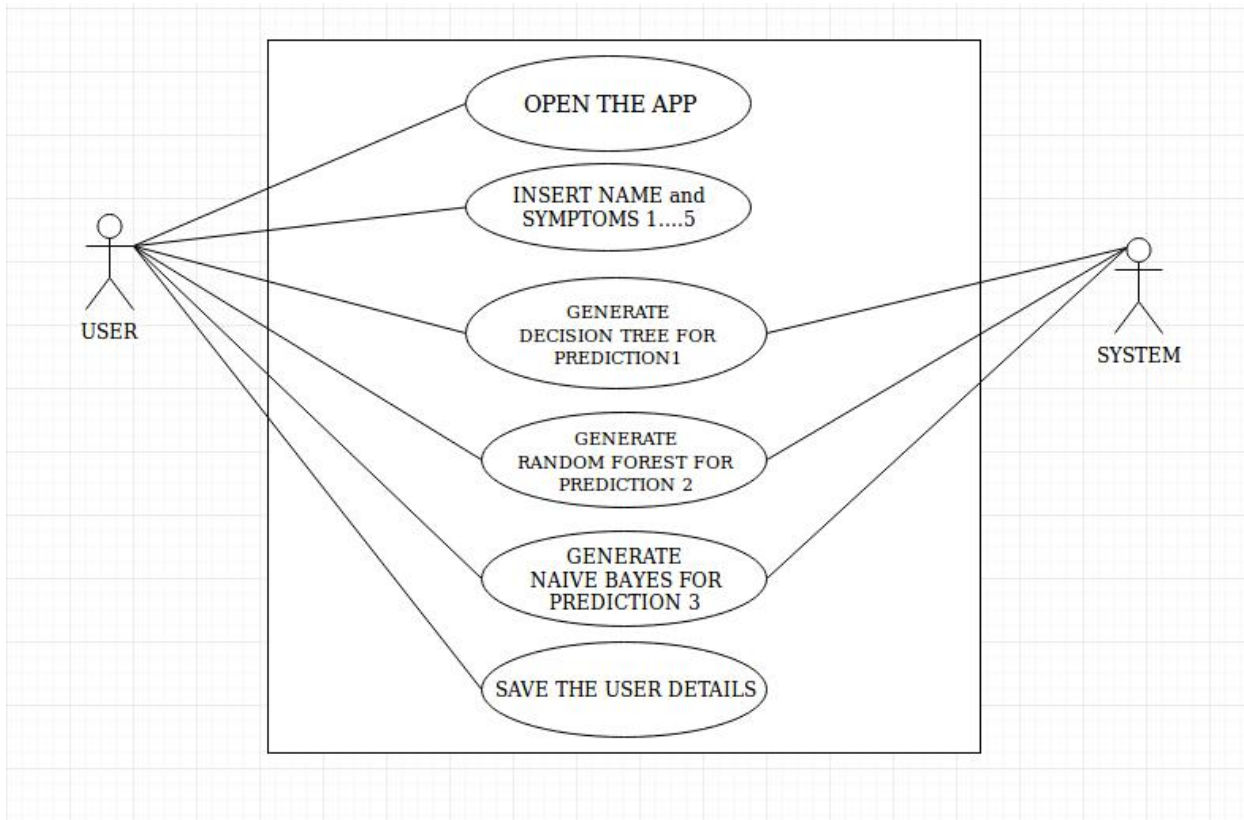


Figure 3: Use-Case Diagram

There are two actors in this project: User and System. User has access to the GUI part of the project while System can generate the predictions that can be viewed by the User. The above is the use case diagram of the project.

CHAPTER 4: IMPLEMENTATION AND EVALUATION

4.1. Tool And Technology

The tools and technology used in this project along with their function are listed below:

- Python3 is used as the prime coding language
- Python Libraries as numpy, pandas and tkinter is used to ease the working of the project
- CSV format is used to place the training data, testing data and save records of patients
- Any IDE that supports python is used to develop the project

4.2. Implementation

The Implementation of this project had all to deal with mapping user inputs i.e. symptoms into probable disease with an acceptable error rate. The Classification method ML was used including three different algorithms: Decision Tree, Random Forest and Naive Bayes. These all algorithms have their different complexity measure and accuracy level. Working with three algorithms helped to compare the outputs on a given set of input. The main challenge was to obtain accurate result based on the trained model using the data set.

A **Decision Tree** is a map of the possible outcomes of a series of related choices. It allows an individual or organization to weigh possible actions against one another based on their costs, probabilities, and benefits. They can be used either to drive informal discussion or to map out an algorithm that predicts the best choice mathematically.

The **Random Forest** algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest".The following are the basic steps involved in performing the random forest algorithm:

1. Pick N random records from the dataset.
2. Build a decision tree based on these N records.
3. Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
4. In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

A **Naive Bayes** classifier is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume strong, or naive, independence between attributes of data points. Popular uses of naive Bayes classifiers include spam filters, text analysis and medical diagnosis.

A naive Bayes classifier is not a single algorithm, but a family of machine learning algorithms that make use of statistical independence. These algorithms are relatively easy to write and run more efficiently than more complex Bayes algorithms. The formula used for naive Bayes is:

$$P(c|x) = (P(x|c) * P(c)) / P(x)$$

These three algorithms are implemented in the project with their functions to predict the disease.

4.3. Evaluation And Result

This application was intended to show the disease from symptoms and it was made so. It contained a simple and easy-to-use design which had no difficulty in using. Users after inserting up to 5 symptoms by selecting from the list were able to predict the probable disease they might be suffering from. This application provides no medical evidences but however it could be as a pre requisite before visiting to any medical organization or professionals. Talking about one of the predictions that could be made was if you are suffering from mild fever, belly pain, loss of smell, congestion and diarrhoea could be possibly typhoid and user can proceed to test for it by visiting a doctor.

There was cases where all of the algorithms showed same prediction but there was condition where all of them were different as well. In case of different, it could be the reason of different working of algorithms. This can be evaluated as there are different possibility of same symptoms. The most efficient algorithm was naive Bayes algorithm as the accuracy rate was approx. 95% with minimum error. So naive Bayes prediction was considered as final prediction.

CHAPTER 5: CONCLUSION

eCheckUp is a Disease prediction symptom based on ML classification technique. This project is developed for personal use and creates no medical evidences. It is not a substitute for professional medical advice, diagnosis or treatment. It is like a pre-requisite for medical professional. This project measures an accuracy of around 95% with very less error rate. Going through this project, one can know about where their symptoms are taking them towards. The three algorithms are compared and there different predictions which include two normal predictions and one final prediction.

Also, there is a data saving option which allows user to save data and store in a file including information as Name of the user, symptoms and predictions that are made. In this way, many other predictions can be extracted from the project and we can be aware of our probable diseases.

CHAPTER 6: LIMITATION

This project is based on Machine Learning technique and main player of ML project is data on which it is working on. This project only recognizes diseases listed on the data. Also, in it user cannot insert their manual symptoms but they have to select from the pre-listed ones. These are the limitations of the project. But as stated earlier, ML project depends upon data set, the data set can be further appended and model can be trained again to obtain good results and higher accuracy.

The current data set is scrapped from a study held at University of Columbia.
(<http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html>)

Another limitation of the project is that in case of multidimensional symptoms that may lead to multiple diseases, this project might not give correct prediction as it depends directly on the dataset.

REFERENCES

[1]"Symptom Checker from WebMD. Check Your Medical Symptoms.", *WebMD*, 2019. [Online]. Available: <https://symptoms.webmd.com/default.htm#/info>. [Accessed: 6- Oct- 2019]

[2]"Random Forest," *R*. [Online]. Available: https://www.tutorialspoint.com/r/r_random_forest.htm. [Accessed: 1-Oct-2019].

[3]"Decision Tree Tutorials & Notes: Machine Learning," *HackerEarth*. [Online]. Available: <https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/>. [Accessed: 19-Oct-2019].

[4]"How Naive Bayes Algorithm Works ? (with example and full code): ML ," *Machine Learning Plus*, 09-Nov-2018. [Online]. Available: <https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/>. [Accessed: 31-Oct-2019].

[5]*Disease*. [Online]. Available: <http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html>. [Accessed: 31-Oct-2019].

APPENDIX I

tk

eCheckUp - Disease Prediction System using Machine Learning

Please go through the available symptoms and various predictions based on your symptoms.

Name of the Patient

Symptom 1

Symptom 2

Symptom 3

Symptom 4

Symptom 5

Prediction 1

Prediction 2

Final Prediction

tk

eCheckUp - Disease Prediction System using Machine Learning

Please go through the available symptoms and various predictions based on your symptoms.

Name of the Patient

Symptom 1

Symptom 2

Symptom 3

Symptom 4

Symptom 5

Prediction 1

Prediction 2

Final Prediction

*users.txt

~/Desktop/Disease-prediction-using-Machine-Learning

Save

```
Name: TEST Symptoms: blister chest pain constipation depression dizziness PROGNOSIS: ERD Hypertension Hypertension
Name: asdf Symptoms: weakness in limbs chest pain coma depression belly pain PROGNOSIS: piles Hypertension Fungal infection
Name: TEST Symptoms: blister chest pain constipation depression dizziness PROGNOSIS: GERD Hypertension Hypertension
Name: TEST Symptoms: blister chest pain constipation depression dizziness PROGNOSIS: GERD Hypertension Hypertension
Name: TEST Symptoms: blister chest pain constipation depression dizziness PROGNOSIS: GERD Hypertension Hypertension
```