AN ENGINEERING PROJECT REPORT
ON
An E-Health Care Chatbot with Mental Health Consultant using Decision Tree Algorithm

Project Supervisor:

Ankit Bhattarai
Department for IT and Computer
Cosmos College of Management and Technology

Submitted By:

Hrishabh Mahaju (160311)
Mann Shrestha (160315)
Muskan Kafle (160316)
Nabin Hamal (160317)

Submitted in partial fulfillment of requirement
of
Bachelors in Computer Engineering
(BCE 2016-2020)

Cosmos College of Management and Technology
Affiliated to Pokhara University
Tutepani, Lalitpur
Date of Submission: 1 August, 2021

PREFACE

Pokhara University, offers four-years course on Bachelor's degree in Computer Engineering. In the final semester, the students have to submit a project report to implement their knowledge acquired during the course of four years. Project on different topics is performed which may be allocated by the institute or the students may bring the project by themselves.

Health, as seen, has been one of the dominant if not the most dominant constraints of life. The latest COVID-19 (A disease caused by a new strain of corona virus killing millions of people since it originated from Wuhan, China in 2019) pandemic and the magnitude of impact it has caused are all prominent till this day as we are still grappling to overcome the situation. It has made us wide open our eyes to the fact that health is the most important factor of human lives. In the context of Nepal, there has been thousands of deaths and over half million people infected by the virus itself. The use of AI chatbot using ML algorithms will be most practical to confront this situation. We created a chatbot which is a scientific application that can help users to gain health counseling promptly. The project for "E-Health care chatbot and Mental Health Consultant" is chosen so that we will be able to contribute during this challenging period and also be able to help the community in a long run.

This project work has been undertaken for the partial fulfillment of requirements for the Bachelor's Degree in Computer Engineering. This project work contains use of decision tree algorithm, a very classical machine learning algorithm that uses tree-like model of decisions and their possible consequences, including chance event outcomes, resource cost and utility. All the theoretical knowledge on analysis and design acquired during the four-years engineering course are utilized with practical application. The main objective of the project work is to acquaint us with the practical aspects of Computer Engineering. This report has been prepared by presenting all the works regarding analysis, design and detailing of Decision tree Algorithm including all related subjects in simple, clear and concise manner.

The advanced software: Jupiter Notebook and PyCharm have been used for the development and deployment of model. Manually prepared datasheets i.e., csv files on MS-Excel have been used to integrate with the model.

ACKNOWLEDGEMENTS

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. The successful completion of our project has been possible only due to such supervision and assistance and we would not forget to thank them.

We respect and thank Cosmos College of Management and Technology, for providing us the platform where we can showcase our skills and knowledge through this project. The support and guidance throughout the process has made us complete the project duly. We are extremely thankful to the Cosmos administration.

We owe our deep gratitude to our project supervisor Er. Ankit Bhattarai, Department of Computer and IT Engineering, Cosmos College of Management and Technology, Lalitpur, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information and support. His proficient guidance, incisive comments and rewarding technical suggestions has all been encouragement to us.

We would also like to express our gratitude towards our Vice Principal Er. Yubaraj Khatiwada, Cosmos College of Management and Technology, Lalitpur, for his direction and leadership through the initial stages of our project. Not to forget our Head of Department Er. Bibek Ropakheti, who has been all but supportive for the past four years of Engineering. His supervision has always been encouragement for the completion of our project. We heartily thank entire department of IT and Computer Engineering for the guidance and suggestions during the project.

Finally, we want to present an honorary mention to our friends and families. We are thankful and fortunate enough to get constant encouragement, support and guidance which helped us in successfully completing our project.

Hrishabh Mahaju (160311)

Mann Shrestha (160315)

Muskan Kafle (160316)

Nabin Hamal (160317)

ABSTRACT

This project "An E-Health Care Chatbot with Mental Health Consultant using Decision Tree Algorithm" is proposed to change the conventional method of receiving health treatments. Today, hospitals are the most widely used means by which a sick person gets medical checkups, disease diagnosis and treatment recommendation. This proposed system is to create an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get diagnosis. This project intends to apply the concepts of machine learning algorithm (Decision Tree) to create a chatbot application. Users can interact with the chatbot just like they do with another human and through a series of yes/no queries, chatbot will classify the symptoms of the user and thereby, predicts the disease and recommends treatment. This system can be of great use to people in conducting daily check-ups, makes people aware of their health status and encourages people to make proper measures to remain healthy. This project also has added feature of some mental health diseases. Users can enter the valid input about their mental condition and the system will predict the probable disease they might be suffering. Executing this proposed framework can help users avoid the time-consuming method of visiting hospitals by using this free of cost application from anywhere.

Keywords—Chatbot; Health; Mental Health; Machine Learning; Classification; Decision Tree Algorithm;

Table of Contents **Page no**

Tables of Figures Page No.

List of tables Page No.

1. INTRODUCTION

1.1. Background

A Chatbot is a system or a software that resembles itself as a human being. In that it can respond to the user's text-based questions or sentences like human being. A chatbot is a type of software that works seamlessly with a user by mimicking the natural language and behavior of a human being. It can respond to text-based questions and sentences. The history of chatbots goes back to the very beginning. They were first introduced to the public in 2009 by QChat and are now available for various platforms. At this time, there are a large number of artificial chat bots available, the main purposes of these chat bots are, however, limited to entertainment, client support and advertising. Some of the chatbots are also used for teaching purpose and better learning of students.

Our project specifically aims to overcome the health issues prevailing on the society along with mental health issues with the use of chat bot. The deliverables of the project majorly focus on the combination of AI and Medicine. The sight is to create a chatbot that can analyze both physical and mental health condition of any patient and suggest medication.

According to WHO "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." The important implication of this definition is that health is not just absence of disorders or disabilities. Health is a state of well- being in which an individual realizes his or her own abilities, can cope with the normal stress of life, can work productively and is able to make a contribution to his or her community. Likewise, mental health can be referred to our emotional, psychological, and social well-being. It affects how we think, feel, and act. It also determines how we handle stress, relate to others, and make choices. Mental health is an important factor at every stage of life, from childhood and adolescence through adulthood.

Our chatbot basically focus on asking questions to the users then predict the disease and also suggest necessary precautions. A chatbot may come handy in many circumstances where the user is hesitant of discussing their disease with any human and provides a new dimension and scientific way on tackling diseases.

1.1.1. Origin of Idea

The aftermath of the COVID-19 (an infectious disease caused by severe acute respiratory syndrome coronavirus 2 SARS-CoV-2) has saddened all of us. The impact during and after the pandemic seems disastrous. The disease affected not only the COVID patients but also the patients that suffered from other disease keeping them deprived of necessary treatment required. Basically, our idea is to aim that the users do not need to visit the health service centers exposing themselves to the virus and rather use the chatbot to gain basic treatments and remedy.

For months, the whole world went under lockdown because of the Corona virus disease or COVID-19. The pandemic has led the economic activities on hold all over the world and impacted the daily lives of the people. Specifically, marginalized groups are edged on the thin line from dying of resource scarcity rather than the virus itself. As a result, a total of 1,647 cases of suicide were filed as of June 27-2020 in police stations all over Nepal, with an average of 18 people killing themselves every day since the enforcement of country-wide lockdown came into effect on 24 March 2020. Thus, we came to realize that "Mental Health" has been one of the misperceived diseases that can really impact on human life. The impact of COVID has ignited a compulsion that we take the matter as a cornerstone of health for humans. That's when we realized not only the physical disease, we must also look upon the mental health related issues prevailing in the society and thus included the subject in our project.

We took this opportunity to work as a team and discussed how we can contribute to this cause and help the society. We came up with an idea of developing a bot that can suggest users to overcome their health issues. A fully integrated AI system operating with classic algorithm (Decision Tree) to chat with the users and provide a meaningful response.

1.1.2. Scope

The main purpose of this project is to deliver the support system for different diseases prevailing on the society with the help of chatbot. Any individual can benefit themselves if they are suffering from any health conditions. The required treatment for the illness can be suggested to the patient by this system after analyzing the symptoms. Recently, COVID crisis has led the entire global population to months of lockdown. This has led the human population towards

degradation of health condition. Lack of exercise and physical activity during months of lockdown caused even more exposure to disease. So, this project can prove to become fruitful for those seeking help and encourage the users to have tendency towards healthy life.

This system can eradicate the old fashion of taking treatment as the patient is not required to visit hospital physically. The system manipulates the entered data of an individual virtually and can get the treatments as per the requirement. This system can be more beneficial to the differently abled peoples as well as old age groups. The system can effortlessly operate 24 hours and each and every individual can enjoy its features to make their life healthier. The virtual way of treatment and consultant can be a boon in the medical field.

1.2. Literature Review

Science and medicinal operation have been moving hand on hand since the beginning. Nonetheless, this mutual growth of science and medical field has benefitted human beings. Here, we have discussed some of the researches that has been carried out over the years on the subject.

i. **Rafael A. Calvo, David N. Milne, M. Sazzad Hussain and Helen Christensen [2]** of Cambridge University has published a report on 30 January 2017 with the topic NLP in mental health applications using non- clinical texts. This review has aimed to provide a taxonomy of how NPL has been used in mental health applications and potential future opportunities for its integration into online mental health tools. They have used data from Twitter, Facebook, blogs and other social media services to learn about people's behavior, emotions, social communications and more. They have majorly covered these three areas:
   - Data Collection
   - Processing or diagnosis
   - Generation of automated mental health interventions

ii. **Benjamin L. Cook, Ana M. Progovac, Pei Chen, Brian Mullin, Sherry Hou, And Enrique Baca-Garcia [3]** has published a report on 26 September 2016 with the topic - Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. They used natural language processing (NLP) and machine learning to predict suicidal ideation and heightened psychiatric symptoms among adults recently discharged from psychiatric inpatient or emergency room settings in Madrid, Spain. Participants responded to structured mental and physical health instruments at multiple follow-up points. Outcome variables of interest were suicidal ideation and psychiatric symptoms. Predictor variables included structured items (e.g., relating to sleep and well-being) and responses to one unstructured question, "how do you feel today?" We compared NLP-based models using the unstructured question with logistic regression prediction models using structured data. They concluded that it is possible to use natural language processing (NLP) based machine learning prediction methods to predict suicide risk as well as heightened psychiatric symptoms in free-text responses sent via mobile phone. This will contribute to the ultimate goal of reducing suicides and suicide attempts, even in low-resource settings.

iii. **Robert Stewart and Sumithra Velupillai, King's College London**, **(Institute of Psychiatry, Psychology and Neuroscience), London, UK [4]** published an article on 7

September 2020, pertinacious about how big data and NLP has created a huge impact over 10 years building a substantial progress in applying NLP within the Clinical Record Interactive Search (CRIS) platform to enhance research at the South London and Maudsley Trust (SLaM): a large mental healthcare provider serving an urban catchment of around 1.3 million residents.

iv. **Azy Barak, Department of Psychology, University of Haifa, Aba Hushi Road, Mount Carmel, Haifa 31905, Israel [6]** published a report on 19 September 2005 with topic "Emotional support and suicide prevention through the Internet: A field project report". In this report the author suggests that the internet can be efficiently exploited to help people in severe emotional distress, including those contemplating suicide. The idea was to initiate an anonymous, confidential online environment that would attract people in a crisis situation and offer them a listening ear, mental support, and warmth, provided by anonymous, skilled helpers in Israel. It came to success and the website is accessed more than 10,000 times a month, or 350 times a day, a considerable number relative to Israel's small population. Feedback by users also indicates the success of the research as a unique psychological application on the Internet.

v. **Adeline Abbe, Cyril Grouin, Pierre Zweigenbaum and Bruno Falissard (17 July 2015) [7]** reviewed he text mining in psychiatry and explored its advantage and limitations. Applying text mining and ontent analysis, they identified four major areas of application. The areas are Psychopathology, the Patient perspective, medical records and medical literature. The systematic literature search to identify applications of TM was performed in the following electronic databases in: the Cochrane Library, MEDLINE (using PubMed platform), Embase, PsycINFO and CINAHL. Their work demonstrates that the text mining can contribute to complex research task in psychiatry.

vi. **Alshawwa, Izzeddin A.; Elkahlout, Mohammed; El-Mashharawi, Hosni Q.; Abu-Naser, Samy S. [8]** has prepared a report on the topic "An Expert System for Depression Diagnosis". The proposed expert system presents an overview about depression disease, the cause of diseases is outlined and the treatment of disease whenever possible is given out. The main goal of this expert system is to get the appropriate diagnosis of disease and the correct treatment and give the appropriate method of treatment through several tips that concern the disease and how to treat it and is seen through the application on the expert system. In this paper, the design of the proposed Expert System which was produced to

help Psychologist in diagnosing depression disease through its symptoms such as: a loss of energy, a change in appetite, sleeping more or less, anxiety, reduced concentration, indecisiveness, restlessness, feelings of worthlessness, guilt or hopelessness and thoughts of self-harm or suicide is discussed. SL5 Object Expert System language was used for designing and implementing the proposed expert system. The Simpler Level 5 (SL5) Object Expert System Language is a rule-based language for specifying expert systems. The proposed depression disease diagnosis expert system was evaluated by psychologist students and they were satisfied with its performance.

vii. **Rohit Binu Mathew, Sandra Varghese, Sera Elsa Joy and Swanthana Susan Alex; Department of Computer Science and Engineering, Saintgits College of Engineering, Kottayam, Kerala, India** [**9]** published an research paper with topic "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning" in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) conducted by Institute of Electrical and Electronics Engineer (IEEE). The conference was held on 23 – 25 April 2019. Here, the authors have proposed a system, a medical chatbot, built to be a conversational agent that motivates users to discuss about their health issues and based on the symptoms provided by them; chatbot returns the diagnosis. This chatbot system will be able to identify symptoms from user interaction. Using these extracted symptoms, chatbot predicts the disease and recommends treatment. The algorithm used here is K- Nearest Neighbor (KNN).

viii. **Vidit Laijawala, Aadesh Aachaliya, Hardik Jatta and Vijaya Pinjarkar; students of University of Mumbai, K.J Somaiya Institute of Engineering and Information Technology, Department of Information Technology** [**10]** proposed a report on April 8, 2020 with the topic "Mental Health Prediction using Data Mining: A Systematic Review". They collected data from online available datasets. The data has been label encoded for better prediction. The data is being subject to various machine learning techniques to obtain labels. These classified labels will then be used to build a model to predict the mental health of an individual. They have planned to apply Random Forest or Decision Tree classification algorithm for classification of the data. Their target population is in the working class i.e., people above the age of 18.

ix. **Sana Mujeeb, Muhammad Hafeez Javed and Tayyaba Arshad from Islamabad, Pakistan [11]** has developed Aquabot: A Diagnostic Chatbot for Achluophobia and Autism. This paper emphasizes on the use of a chatbot in the diagnosis of Achluophobia – the fear of darkness and autism disorder. A chatbot has been developed in this work which can diagnose the severity of disease based on user's text-based questions. It performs Natural Language Processing (NLP) for meaning extraction and uses Decision Trees to characterize a patient in terms of possible disease. NLP unit extracts meaning of keywords defining intensity of disease's symptoms, from user's chat. After that similarity matching of sentence containing keywords is performed. Depth First Search (DFS) technique is used for traversing Decision Tree and making decision about severity of disease. The proposed system namely Aquabot, proves to be an efficient technique in diagnosing Achluophobia and Autism. Aquabot is useful for practitioner psychologists to assist a human psychologist. Aquabot not only saved time and resources but also achieved an accuracy of 88 percent when compared against human psychologist's diagnosed results.

2. CASE STUDY OF AQUABOT: A DIAGNOSTIC CHATBOT FOR ACHLUOPHOBIA AND AUTISM

2.1. Introduction

Aquabot is a robotic assistant that works seamlessly with a human psychologist. It can detect emotion and provide accurate results. This paper describes the use of chatbot in the diagnosis of Achluophobia and autism disorder. The chatbot can analyze the severity of disease based on the text-based questions that the user has sent to it. It uses Natural Language Processing (NLP) to extract meaning from the query. Aquabot is trained to determine if a person has the fear of darkness or if they are suffering from achluophobia.

The paper under study focuses on Achluophobia (fear of the dark), and Autism (conditions that make an individual's social life challenging by affecting one's speech, repetitive behavior and non-verbal communication). These disorders are mostly seen in children. Autism is a pervasive disorder and to treat it, robots are used. To identify these disorders through state-of-the-art techniques takes a lot of time and hence the paper approaches the diagnostics by implementing a Chabot that aims to reduce the time taken for diagnosis. In this Chabot method, a question-answer format is used where a user's answers are analyzed using Natural Language Processing method (NLP) to extract meanings from the text, and uses Decision Trees to categorize the subjects of study (i.e., individuals) suffering under different potential diseases. Depth-First Technique (DFS) is used to traverse through the decision tree, and hence assisting to make decisions about the severity of diseases. The Chabot, Aquabot, claims to achieve an accuracy of 88% compared to diagnostic results from a human counterpart. Aquabot finds its main use in assisting a human psychologist.

2.2. Related Works for Diagnosis and Treatment

To diagnose and treat diseases related to phobias, chatting is mostly used. Psychiatrists help identify and treat these diseases whereas in automated treatment of diseases, virtual reality (VR) technique is used which mainly involves the use of a special device that can be used to interact with a three-dimensional virtual world with the help of a joystick or a click of a mouse. The main motive is to let the subject experience a virtual world. The study also may consist of measuring the speed of keystrokes while playing a game. Non-Chabot methods already exist in the industry that uses the client and service-provider model where a client subscribes to an online portal by

paying a regular fee or using an on-demand pricing model for psychological counseling. Expert systems also exist to diagnose the disease that has some standard knowledgebase that can be extended or interfaced with some artificial intelligent systems. To address autism, works have been done with automated intelligent solutions. In such solutions, the modalities of the machine are limited and hence autistic people may develop some abilities to interact with the machine using those modalities, as they are the only way to interact with the machine and since autistic people focus on single details, interaction with robots might help them gain an extra ability. Another method used to diagnose autism is by creating social situations designed to create a particular social response with the use of interactive social robots and see how the subject behaves, and measures the response-time. The chatbot system, although gives a slight picture of what might be going inside a subject's psyche, hinders a broad view of what is actually going on inside a subject. A Chabot's questions are narrow and fixed, and hence they are useful for a preliminary assessment of a psychological condition before a real psychiatrist is involved.

2.3. Proposed System

Aquabot mainly asks questions to its users to collect replies from the user and generate replies to user input, and processes the reply to narrow down the issues with the subject, based on a decision tree. Chabot's are similar to a two-way communication; however, a Chabot can't address broad or off-topic replies from the user, and is limited strictly by the context of the conversation. The replies from the users are processed, and Aquabot traverses the decision tree until a leaf is reached.

The Chabot is interactive, i.e., the users respond to the Chabot's questions, and the Chabot also processes the response and is supposed to know what to ask next so as to gather more data from the user. The Chabot decides at every input and proceeds further until it reaches the end.

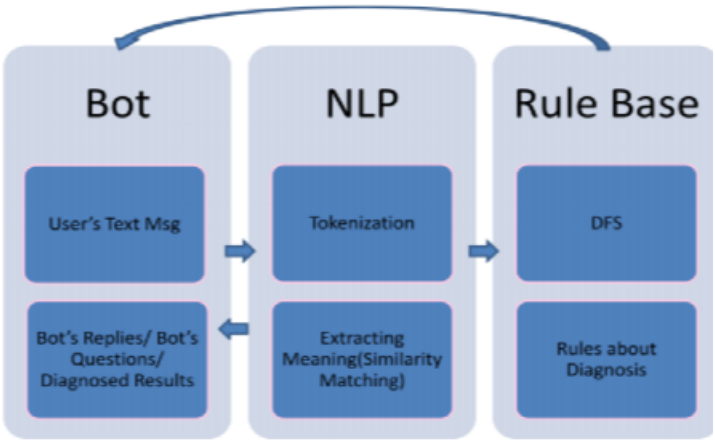


Figure 2.3. 1 Framework of Aquabot

NLP forms the major part of the Aquabot which processes the sentences by breaking them into words and phrases and then figuring out important keywords and the context of the conversation. Similarity is matched from the degree of intensity of the extracted keywords and seen if it matches the context, i.e., the severity of the symptoms is decided. The following picture describes the framework of Aquabot.

2.4. Working Mechanism of Aquabot

The working mechanism consists of: -

1. **Natural Language Processing (NLP)**
   Whenever any question is asked from Aquabot to the users, during diagnosis process, the answer is processed using the NLP methodologies.

Figure 2.4. 1 Aquabot showing features of NLP

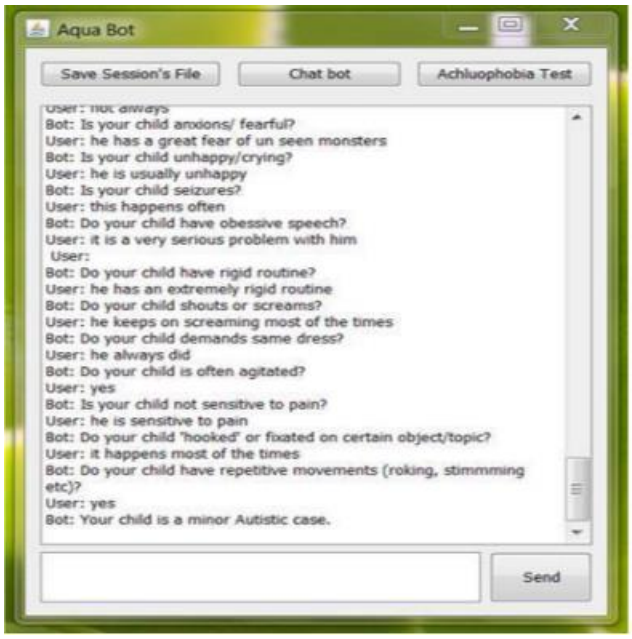


Figure 2.4. 2 Working of Aquabot

2. **Tokenization**

   The answer given by the user is splatted into tokens. Then, the splatted tokens is saved for the comparison.

3. **Keyword Extraction**

   Every token are scrutinize and compared with the words that can express the intensity of symptoms.

4. **Similarity Matching of Sentence**

   The entire sentence is compared with other sentence to check either the meaning of the keywords is same or not.

5. **Understanding Meaning of Keywords**

   The keywords are matched to the severity of the system (i.e., normal, minor, average and severe).

   When the keyword explaining the symptom's severity is found in user's answer, the keyword is mapped to severity level. The question regarding diagnosis is asked. When all symptoms are gained through the user it is saved and the output is given with the use of Decision Tree algorithm.

2.5. Results

The paper shows that the diagnostic accuracy of Aquabot to diagnose Achluophobia was on average greater than 86%, and that of autism was on average slightly greater than 87%, assuming that the human counterpart diagnosed the problem 100%.

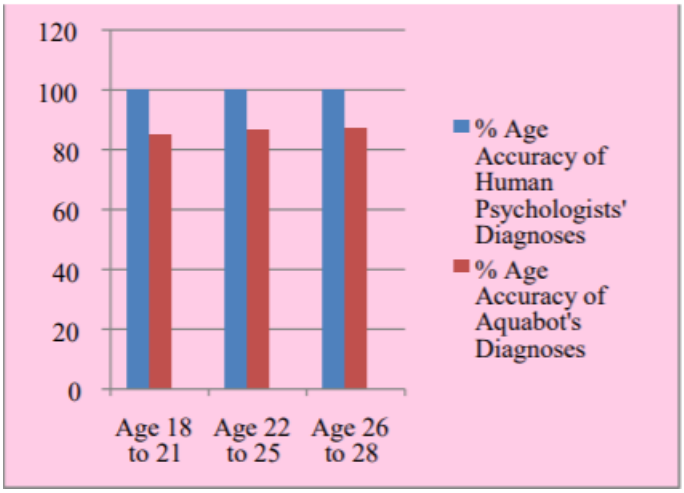


Figure 2.5 1 Percentage accuracy for achluophobia diagnosis

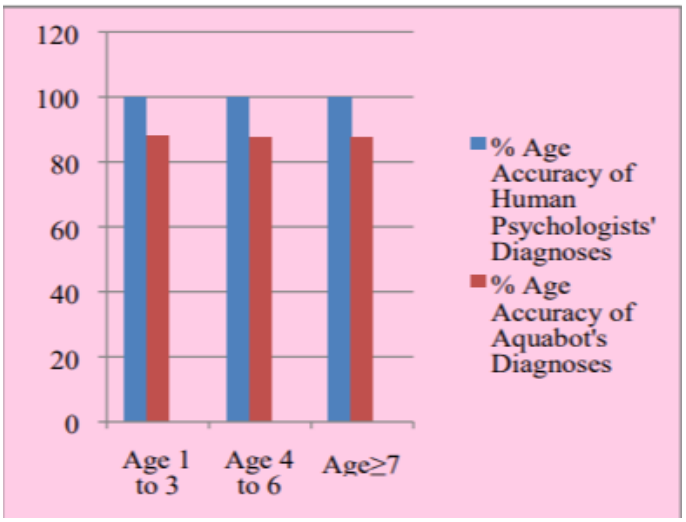


Figure 2.5 2 Percentage accuracy for autism diagnosis

This shows that Aquabot is able to save time and energy of a human psychologist by automating the initial diagnostic steps.

The paper claims that Aquabot can be extended to cover other psychological diagnosis, but some psychological issues might require more immediate attention thereby, bypassing the logistics of using an application like Aquabot, nonetheless, bots like Aquabot have proven that it can save a psychologist's time and energy in doing the repetitive preliminary diagnostic task that can be outsourced to a machine.

2.6. Comparison Between Aquabot and an E-Health Care Chatbot with Mental Illness Counselor using Decision Tree Algorithm

We compared this project with the Aquabot. Here, we have presented the table comparing these two systems under various parameters.

Table 1. Table comparing various features of Aquabot and E- Health Care Chatbot with Mental Illness using Decision Tree Algorithm

S. N	Diagnostic Chatbot	Technique Used	Diagnostic Accuracy	Response Time (Tn)
1.	Aquabot	NLP, Decision Tree, DFS	88%	[n(n+1)]/3
2.	An E-Health Care Chatbot with Mental Health Consultant using Decision Tree Algorithm	Decision Tree Classifier, Scikit learn	97. 78%	-

3. STATEMENT OF PROBLEM AND SOLUTION

First and foremost, a country like Nepal where the majority of people are marginalized and have a scarcity on everyday basic needs, it is evident that the obtaining necessary health facility is a critical subject of discussion. Furthermore, the lack of education and awareness has guided society towards having poor health system and facility.

Since, the COVID-19 pandemic started it came to our knowledge how impoverished our health system is and how backward we are in this development sector. The major thing the COVID-19 pandemic taught us that there is nothing more important than our health. Over these two years of pandemic, we heard lots of cases where patients who were suffering from different sort of diseases other than COVID-19 were also deprived of getting proper health care service even in the urban areas due to chaos caused by the pandemic. Besides, we also came to recognition about the cases where people were so terrorized and panicked by the situation that they misinterpret and related everything to COVID. It caused more disruption of the disease as people got more and more exposed to the virus in longing for treatment in the hospital.

On the other hand, over the years, mental illness has been misinterpreted as different superstitions and still not considered as a serious subject. It is very necessary to aware people about the impact of mental health. The increase in number of mental health patients is inadequate. The fact that Corona virus pandemic has made it even more difficult to deal with is undeniable. Number of research and campaigns are carried out all over the world on command of mental health but the fact is it still has been as intriguing as always. There are currently multiple websites that provide necessary information about the disease but no further experiment has been seen promising. With the development of science and technology there is a need of permanent solution to this subject but in context of Nepal there seems no progress at all. Different organizations come up with practical ideas of campaigns and awareness but there is no effective web-based platform that can help mental health patient to overcome their concern.

As quoted by the great ancient Greek philosopher Plato “Necessity is the mother of invention.”, it is necessary to come with the solution for this long prevailing problem. This project helps to

overcome all the issues as mentioned above. With the help of AI, we can deal with the subject with a scientific approach. Some of the solutions that AI can address are:

a) AI can help support health professionals in doing their jobs. Algorithms can analyze data much faster than humans, can suggest possible treatments, monitor a patient's progress and alert the human professional to any concerns. In many cases, AI and a human clinician would work together.

b) Due to the lack of human mental health professionals, it can take months to get an appointment. If patients live in an area without enough mental health professionals, their wait will be even longer. AI provides a tool that an individual can access all the time, 24/7 without waiting for an appointment.

c) The cost of care prohibits some individuals from seeking help. Artificial intelligent tools could offer a more accessible solution.

d) While it might take some people time to feel comfortable talking to a bot, the anonymity of an AI algorithm can be positive. In case of mental disease, what might be difficult to share with a therapist in person is easier for some to disclose to a bot.

4. OBJECTIVES

4.1. General Objectives

a) To have the importance of good health reach out to people.

b) To establish a new scientific approach, reach out to people and make them believe the advancement of technology in the field of health and science.

4.2. Major Objectives

a) To model a chatbot applying decision tree algorithm with optimal accuracy (more than 95% for training set of more than 5000 data) that can help people suffering from different diseases to have a virtual approach to treatment. An important aspect, especially, when it comes to mental health is that talking with a non-human entity provides a sense of security as it remains as a confidential meeting with the diagnosis being available only to the user.

5. System Development

This chatbot is helpful for anyone who needs to get the hang of something about wellbeing. The user can interact with chatbot and can rely on them to get timely diagnosis. This framework helps users to present their symptoms that affect their wellbeing. The genuine welfare of the chatbot is to encourage the general population by giving appropriate direction in regards to the great and sound living.

We have collected data from online. The dataset mainly consists of data of different disease and their symptoms. The data consists of string attributes which we later encoded to numeric attributes for better prediction. It consists of more than 140 attributes for prediction and 1 predicting label. We designed a platform where a user will be communicating to the bot through text. The bot will ask the questions based on the dataset gathered. The user will answer the yes/no questions and a result about his/her health condition will be provided as per the inputs provided. The bot makes use of the model that we build using the decision tree learning algorithms to provide the output.

Table 2 Comparison of Machine Learning Algorithms

Algorithm	Accuracy (%)	Mean Absolute Error	Precision	Time Taken
Decision Tree	82.2	0.256	0.827	0.3 sec
Random Forest	79.3	0.316	0.793	0.6 sec
Logistical Regressic	81.4	0.231	0.826	0.5 sec
One R	82.1	0.178	0.834	0.3 sec
Naive Bayes	78.7	0.24	0.787	0.5 sec
Nnge	75.8	0.242	0.758	0.3 sec

From the above comparison of algorithms, we have found that the most optimal algorithm is Decision Tree Algorithm. It has the highest accuracy and the lowest execution time. Decision Tree algorithm is a supervised learning algorithm. In this project it is used to solve classification

problems. It uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the trees.

5.1. Use case diagram

The proposed Chatbot works similar as a question answer machine. Users are required to query about their problem through the user-friendly chatbot. As soon as query for the mental issue is passed, the algorithm comes into play (i.e., Decision Tree). The Decision Tree algorithm retrieves and classification of the symptoms is done. Whenever the required nodes of the tree are found, the algorithm displays it to the user interface and finally the user gets the output. Users may be asked for further multiple questions for the better understanding of the whole situation. Hence, the entire system suggests the users their health situation and advice for the required medication.

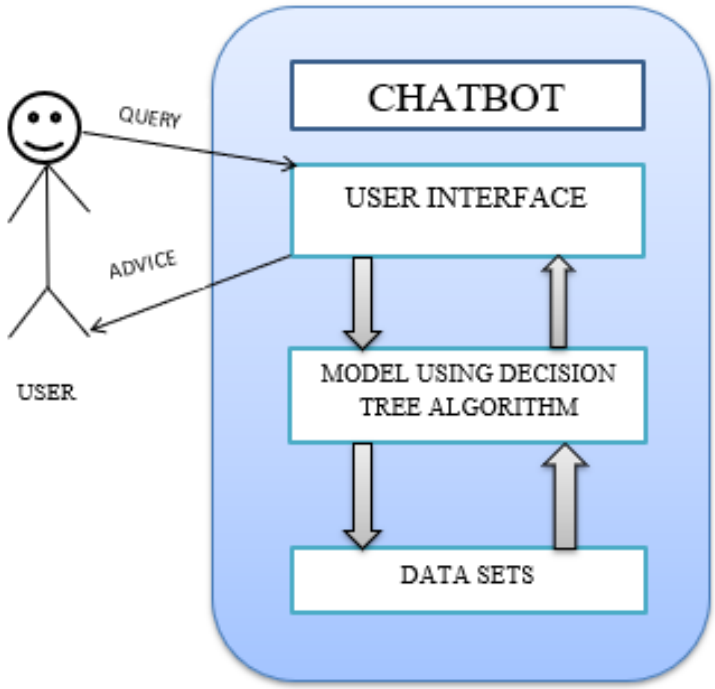


Figure 5. 1 Figure representing the use case for an E-Health Care Chatbot with Mental Illness Counselor using Decision Tree Algorithm

6. METHODOLOGY

6.1. Decision Tree Algorithm:

Decision tree is a tree like structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node will hold a class label. In other words, a decision tree is a decision support tool using a tree like structure model of decisions to see what consequences occurs from each decision.

Decision tree classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. During execution of algorithm, first of all, the best attributes of the data set is placed at the root of the tree. After this, the training sets are split into subsets (which is made in such a way that each subset contains data with same value of attribute). Then, these processes are repeated on each subset until the leaf nodes in all the branches of the tree are found. Furthermore, for our project we can list the following steps involves in the decision tree algorithm:

a) Select the best attribute using Attribute Selection Measures (ASM) to split the records.
b) Make the attribute a decision node and breaks the dataset into smaller subsets.
c) Starts tree building by replacing this process recursively for each child until one of the following conditions will match:
    - All tuples belong to the same attribute value.
    - There are no more remaining attributes.
    - There are no more instances.

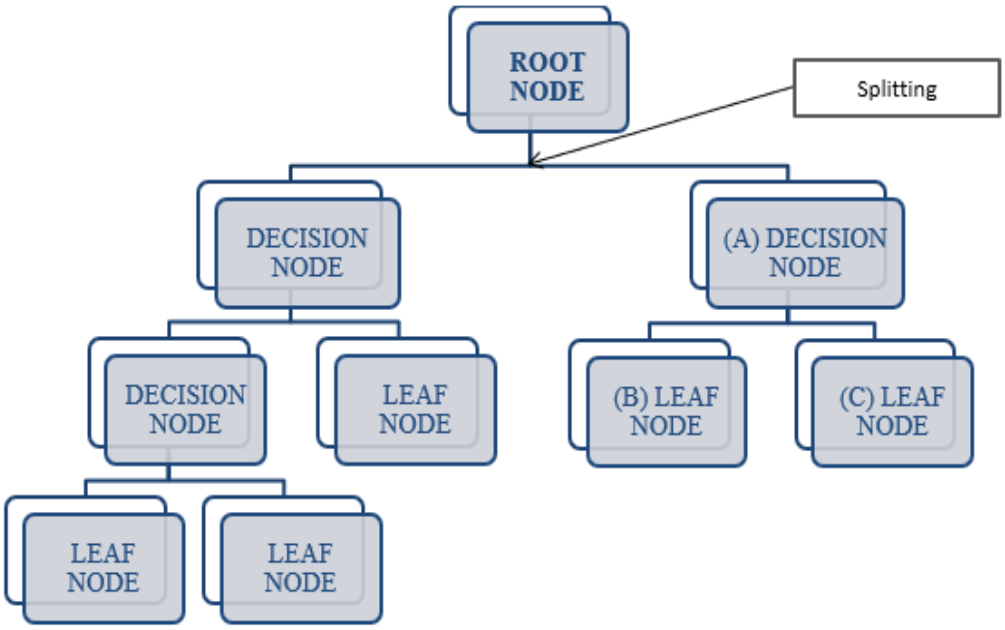


Figure 6. 1 Figure Representing Decision Tree Algorithm Structure

Note: A is parent node of B and C.

Here are some important terminologies related to Decision trees:

- **Root Node:** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
- **Splitting:** It is a process of dividing a node into two or more sub-nodes.
- **Decision Node:** When a sub-node splits into further sub-nodes, then it is called the decision node.
- **Leaf / Terminal Node:** Nodes do not split is called Leaf or Terminal node.
- **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. It can be said as the opposite process of splitting.
- **Branch / Sub-Tree:** A subsection of the entire tree is called branch or sub-tree.
- **Parent and Child Node:** A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.

6.2. Tools and Technology Used

6.2.1. Python: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. It is easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

The reason why we chose python programming language over other is because python offers concise and readable code. Python is commonly used for developing machine learning algorithms, its simplicity makes it possible to develop systems that are reliable and easy to implement. Python's extensive selection of machine learning-specific libraries and frameworks simplify the development process and cut development time.

6.2.2. Integrated Development Environment (IDE): We used Jupyter notebook as our main IDE (Integrated Development Environment) for the entirety of this project. The IDE was basically supported by Anaconda software. Anaconda is an open source, cross-platform, language-agnostic package manager and environment management system for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that installs, runs, and updates packages and their dependencies. It was created specifically for Python programs.

The Jupyter Notebook is an open-source web application that allowed us to create and share documents that contain live code, equations, visualizations and narrative text. Its uses include: data cleaning and transformation, numerical simulation,

statistical modeling, data visualization, machine learning, and much more which were very essential prospects of our project.

6.2.3. Kaggle: Kaggle is an online community of data scientists and machine learning practitioners which allows users to publish data sets, explore and build models in a web-based data-science environment. Our project required different sets of data related to health symptoms, precautions and description. We found Kaggle very helpful tool to gather enough information and data for our project.

6.2.4. Libraries

a) Scikit-Learn

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. Our project majorly focuses on Decision tree classifier and scikit- learn library is primitively very necessary for this project.

b) Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. Data in pandas is often used to feed statistical analysis in SciPy, plotting functions from Matplotlib, and machine learning algorithms in Scikit-learn. In our project, we use Pandas library for data analysis. We used it to import data from csv files and run manipulation operation such as merging, reshaping, selecting and data cleaning.

c) NumPy

NumPy brings the computational power of languages like C and Fortran to Python. NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Our project acquires more than 4000 data. NumPy's memory efficiency has proven very efficient for our project.

6.3. Project Block Diagram

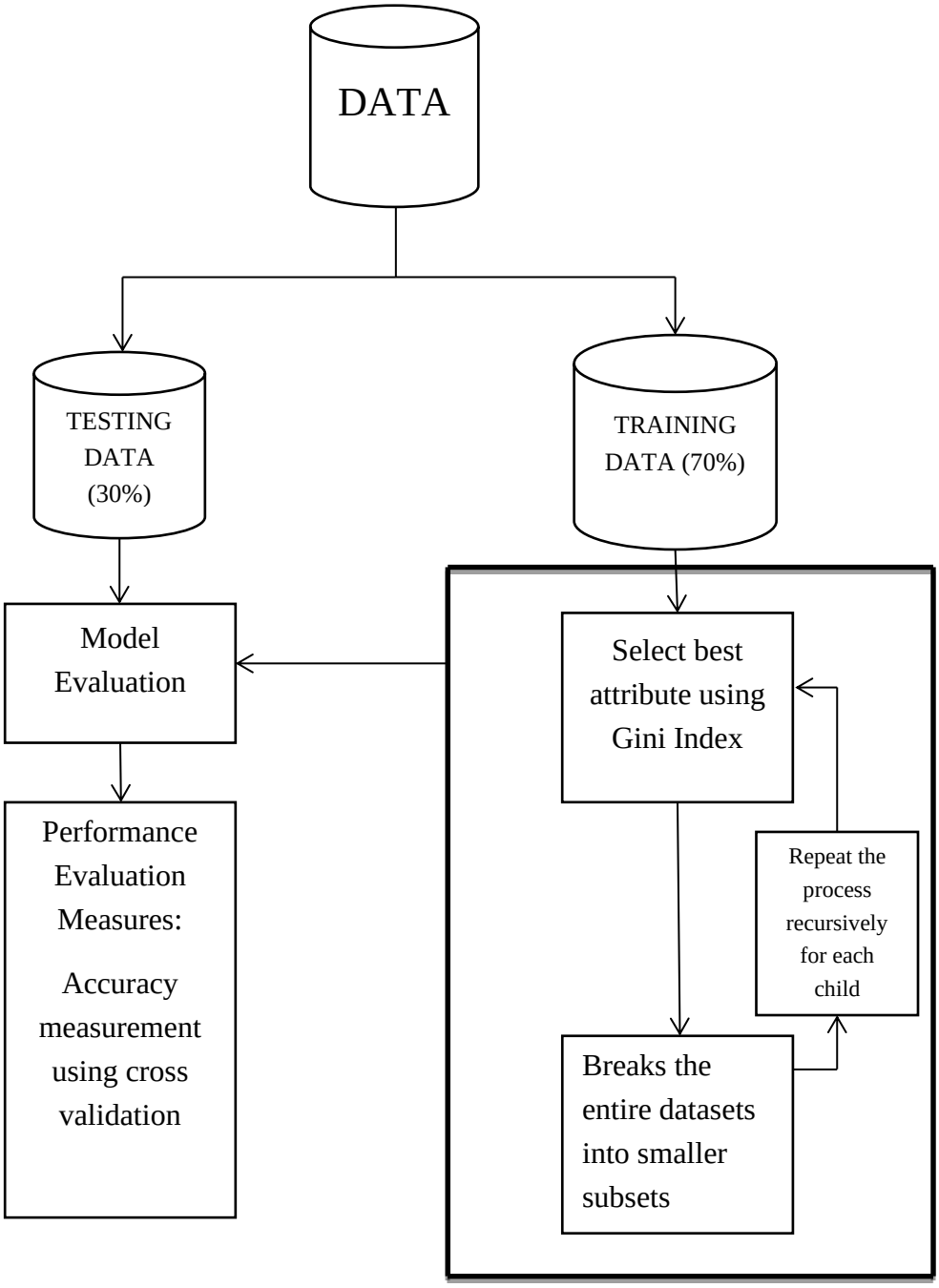


Figure 6.3. 1 An E-Health Care Chatbot with Mental Health Consultant using Decision Tree Algorithm Block Diagram

6.4. Working Principle

6.4.1. Data Extraction

Kaggle is a platform where online community of data scientists and machine learning practitioners publish data sets, explore and build models in a web-based data-science environment. It claims to consist over 50,000 public datasets and 400,000 public notebooks required for several machine learning and data science projects. Our project required a set of data related to health symptoms, precautions and description. We found Kaggle very helpful tool to gather enough of these data for our project. We extracted the dataset containing more than 40 diseases (rows) with their more than 140 symptoms (columns) which was the most suitable for our project.

The user sends messages and as a response the chatbot gives appropriate message. For this to happen smoothly, the chatbot will be trained with some possible questions and answers predefine, that the user can ask. For that, we will use the data extracted from the above-mentioned platform.

6.4.2. Pre-processing

After the data extraction, the next step in the project was to preprocess the data. Before that we needed to make sure that the data, we are using does not contain any impurities and anomalies. We made sure the data is clean and ready for the use.

Thereafter, we carried out these steps in the project:

- **Dimensionality Reduction:** It is not feasible to analyze each and every variable at microscopic level, we might lose tons of time doing so in our project. Considering the amount of computational power, we used this concept to reduce the number of features in our dataset without having to lose too much of information and improve the model's performance.

  We used groupby() function to group large amounts of data and compute operations on these groups.

- **Feature Selection**: Feature selection is the process of reducing the number of input variables when developing a predictive model. Predictive modeling problems have a large number of variables that can slow the development and training of models and require a large amount of system memory. So, we divided given columns into two types of variables dependent (or target variable) and independent variable (or feature variables).

  In this project, feature variable represents the different symptoms in the dataset provided in the columns and target variable refers to the disease.

- **Label Encoding:** There are three types of data i.e., structured (data stored in database having rows and column), semi-structured (data in XML, JSON format) and unstructured data (emails, images, log data). If we are working on only structured data and collected data is combination of categorical variables and continuous variables, then many machines learning algorithm do not recognize it as machine prefer number system for training and testing. Label encoder changes the categorical value to a numeric value between 0 and the number of classes minus 1. Suppose categorical values of 5 distinct classes are given, then we use 0,1,2,3 and 4 only. As a result, label encoder helps to enhance the accuracy and machine learning algorithm will perform better when the data are presented as a number.

  We achieved Label Encoder in our project using Sklearn library. This approach is very simple and it involves converting each value in a column to a number. We did so that we could map each disease on our datasets on the basis of number provided by label encoder.

6.4.3. Splitting Data

Before moving any further, let us discuss about the data sets of this project and their properties. We strategically, split the extracted dataset into a training set and test set to better understand the model performance. There are basically 6 sets of data that we use. They are:

- Dataset.csv: This is the main data set that we extracted from the online platform; Kaggle. It contains different combinations of symptoms to their specific diseases.

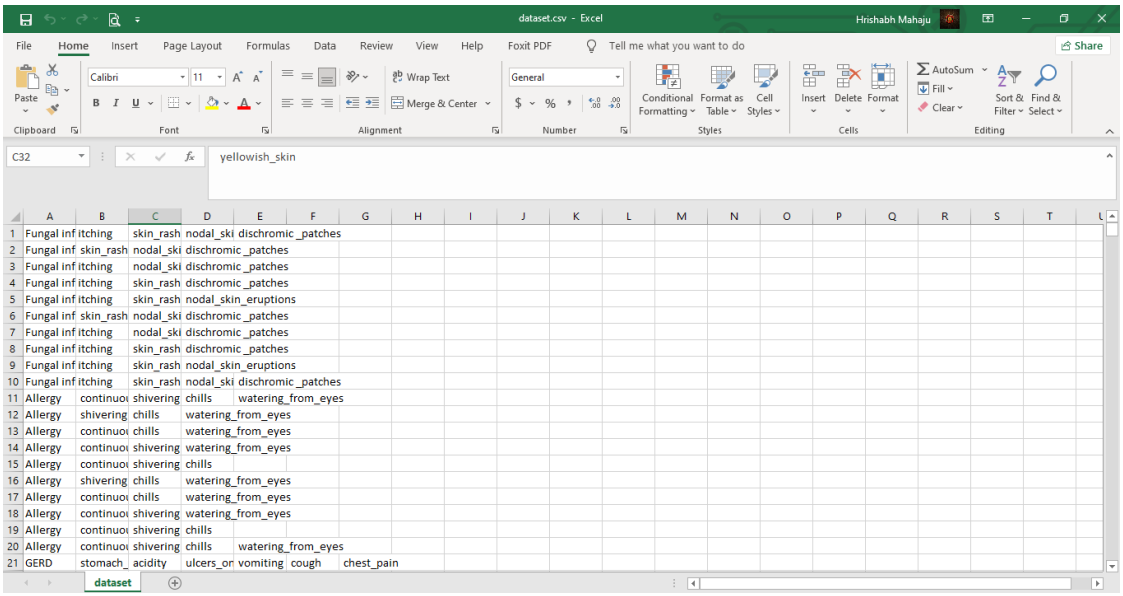


Figure 6.4.3. 1 Snapshot of dataset.csv

- Training.csv: This data set is used to train the model.

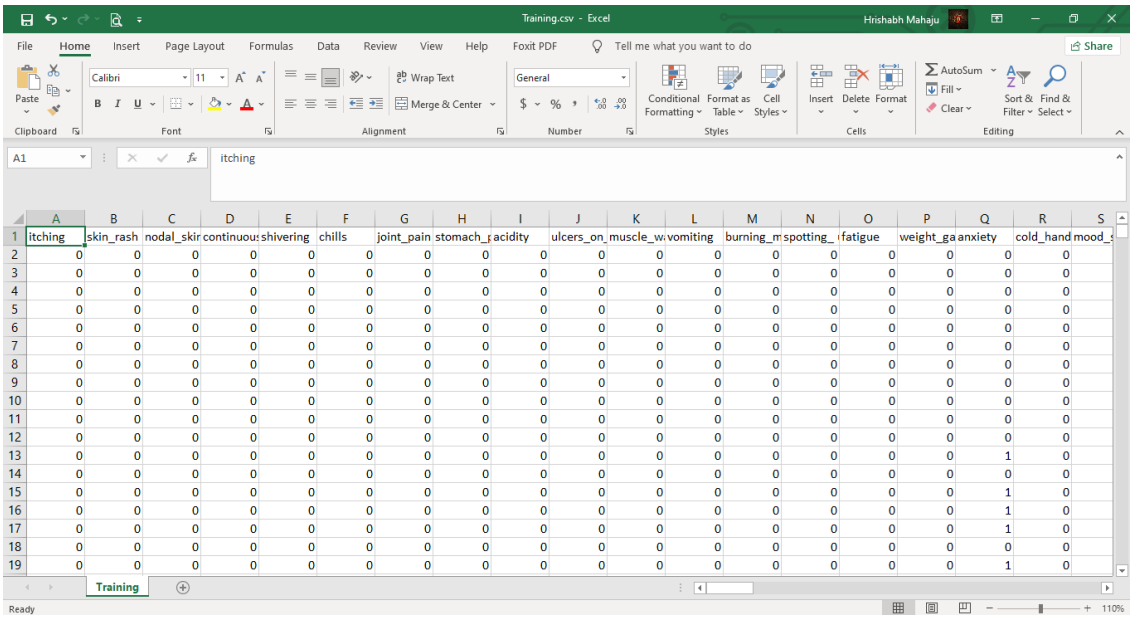


Figure 6.4.3. 2 Snapshot of training.csv

- Testing.csv: This data set is used to test the model.

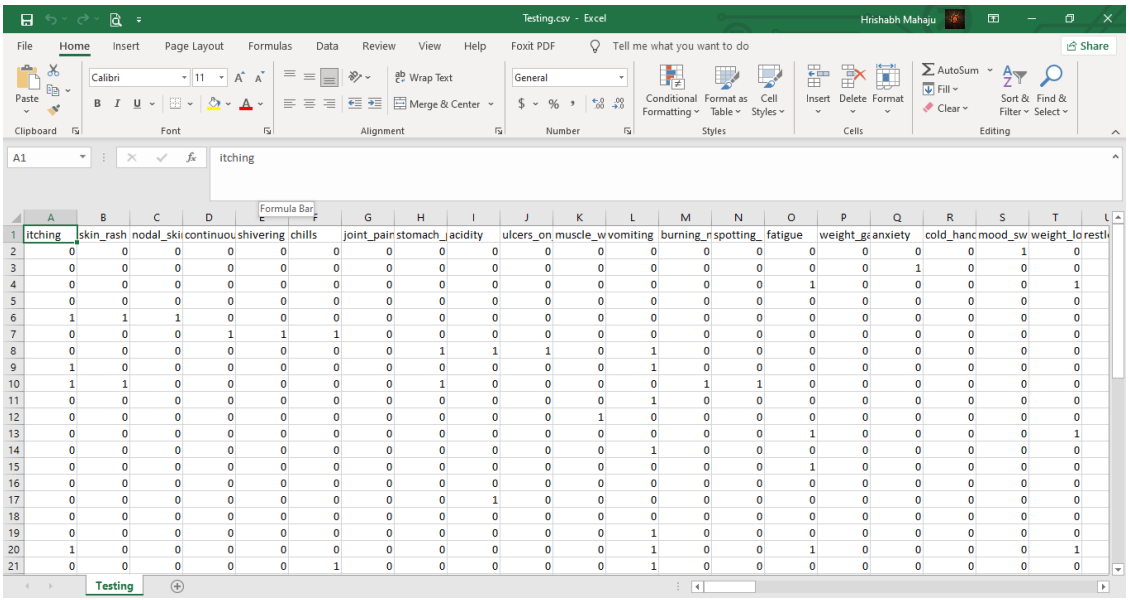


Figure 6.4.3. 3 Snapshot of testing.csv

- Symptom_description.csv: This set contains the description of all the diseases.

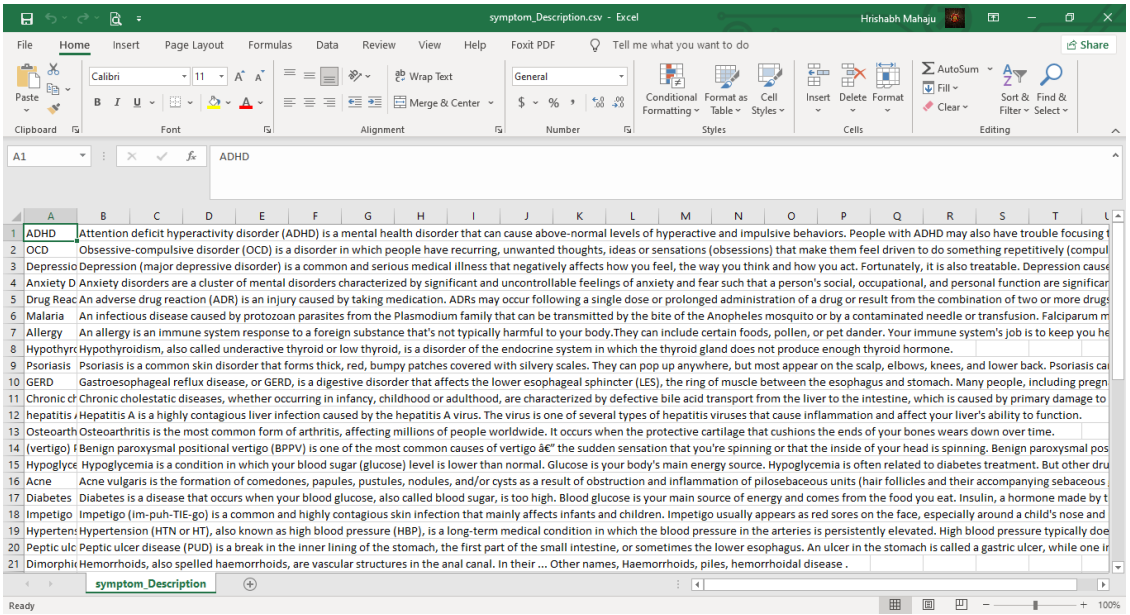


Figure 6.4.3. 4 Snapshot of symptom_description.csv

- Symptom_precaution.csv: This set contains list of precautions required to overcome the diseases.

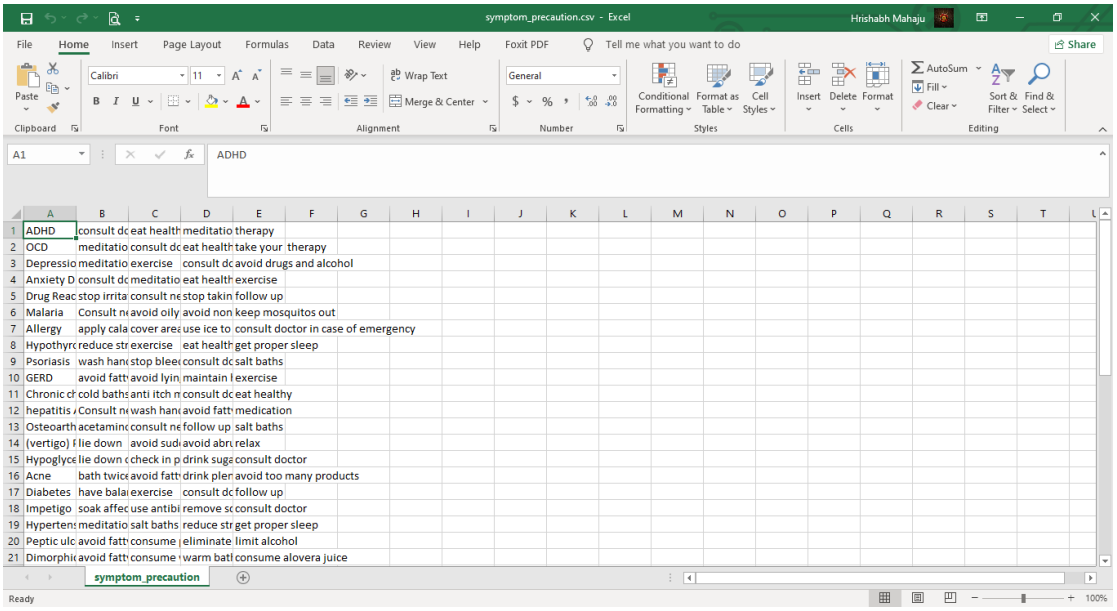


Figure 6.4.3. 5 Snapshot of symptom_precaution.csv

- Symptom_severity.csv: This set contains the disease and the severity of the disease.

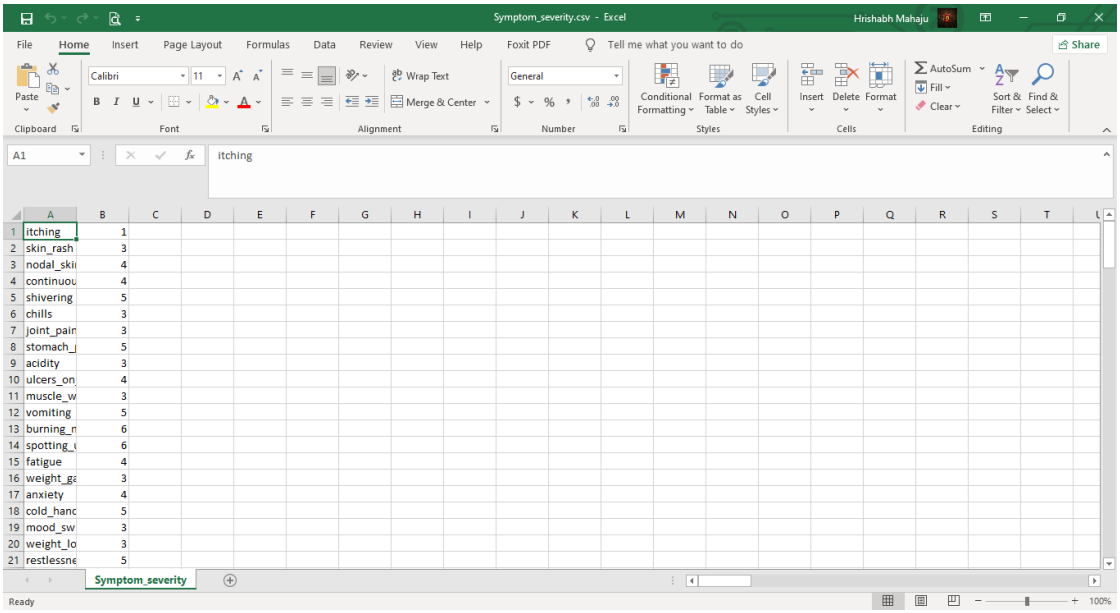


Figure 6.4.3. 6 Snapshot of symptom_severity.csv

6.4.4. Model Development

This system requires the need for an algorithm that can help in predicting the disease from the symptoms entered by the user. For this to happen, it uses a machine learning algorithm. For the chatbot to provide the user with a response to the messages received, we used Decision tree classifier. Classification works in two steps, learning step and prediction step. Here, in learning step, the model is developed based on training data. Then, the model is used to predict the response for given data.

Decision Tree Algorithm belongs to the family of supervised learning algorithm. The goal of using Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data.

There are basically two types of decision trees; Categorical Variable Decision Tree and Continuous Variable Decision Tree. Let's understand these two with the following example:

Example - Let's assume we have a problem to predict whether a customer will pay his renewal premium with an insurance company (yes/ no). Here we know that the income of customers is a significant variable but the insurance company does not have income details for all customers. Now, as we know this is an important variable, then we can build a decision tree to predict customer income based on occupation, product, and various other variables. In this case, we are predicting values for the continuous variables. If the decision Tree has a continuous target variable, then it is called Continuous Variable Decision Tree.

In this project, the decision tree algorithm includes categorical target variables that are divided into categories. The categories can be yes and no. The user either has (Yes) the symptom or doesn't (No). This means that every stage of the decision process falls into one category and there are no in-betweens.

How does it work?

Decision Tree is a machine learning algorithm which is very simple and efficient. It is a widely used classification algorithm because, it similar to a tree, and starts with the root node, which expands on further branches and constructs a tree-like structure. It classifies the examples by sorting them down the trees from the root to some leaf/terminal node, with leaf/terminal node providing the classification of the example.

Each node in the tree acts as a test case for some attribute, and each edge descending from the node corresponds to the possible answers to the test case. This process is recursive in nature and is repeated for every subtree rooted at the new node.

Here are some Assumptions we make while using decision tree:

i. The whole training set is considered as the root in the beginning.
ii. Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
iii. Records are distributed recursively on the basis of attribute values.
iv. Order to placing attributes as root or internal node of the tree is done using some statistical approach.

In Decision Trees, it follows Sum of Product (SOP) representation. It is also known as Disjunctive Normal Form (DNF). For a class, every branch from the root of the tree to a leaf node having the same class is conjunction (product) of values, different branches ending in that class form a disjunction (sum).

The primary challenge in the decision tree implementation is to identify which attributes do we need to consider as the root node and each level. Handling this is to know as the attribute's selection. We have different attributes selection measures (ASM) to identify the attribute which can be considered as the root note at each level.

Moreover, for predicting a class label for a record, we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

In context of this project, following example can be considered:

Example - Headache symptom can be distinguished by having normal fever or a tumor. In reality, there can be any number of predictors and the example can be extended to incorporate any number of characteristics. For example, if we take four kinds of diseases relating headache, they are classified to a proper disease in to sub-root and first check the symptoms of one sub-root if it is true, gives the required precautions and if not, it goes for another sub-root for searching the symptoms.

Decision Tree can classify properly the symptoms and map them to their corresponding diseases.

Attribute Selection Measures

A dataset may contain N attributes, then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step. Random selection of any node to be root may cause bad results and low accuracy. There are different criteria suggested for attribute selection problem. Some of them are:

i. Entropy
ii. Information Gain
iii. Gini Index
iv. Gain Ratio
v. Reduce in Variance, etc.

For this specific project of classification, we use Gini Index method to create split points. The Gini index can be understood as a cost function used to evaluate splits in the dataset. It is calculated by subtracting the sum of the squared probabilities of each class from one. It favors larger partitions and easy to implement whereas information gain favors smaller partitions with distinct values.

Mathematically, Gini Index can be represented as:

$$Gini = 1 - \sum_{i=1}^{c} (Pi)^2$$

Here, Gini index can also be written as,

$$Gini(D) = 1 - \sum_{i=1}^{c} (Pi)^2$$

Where, Pi = Probability that a tuple in D belongs to class Ci.

Gini Index works with the categorical target variable "Success" or "Failure". It performs only Binary splits for each attribute.

We can compute a weighted sum of the impurity of each partition. If a binary split on attribute A partitions data D into D1 and D2, the Gini index of D is:

$$Gini_A(D) = \frac{|D1|}{|D|} Gini(D1) + \frac{|D2|}{|D|} Gini(D2)$$

In case of a discrete-valued attribute, the subset that gives the minimum Gini index for that chosen is selected as a splitting attribute. In the case of continuous-valued attributes, the strategy is to select each pair of adjacent values as a possible split-point and point with smaller Gini index chosen as the splitting point.

$$\Delta Gini(A) = Gini(D) - Gini_A(D)$$

The attribute with minimum Gini index is chosen as splitting attribute.

Mapping and Output

In case of this project, the keywords gained from the user as their symptoms are mapped with the keywords explaining the symptom's severity. In this fashion, chatbot asks multiple symptoms and the users respond with yes/ no. When entire questioning session is completed the decision tree classify the disease and publishes the result. It maps the disease with the disease_precaution.csv and disease_description.csv data sets and provides the output accordingly.

6.4.5. Model Evaluation

For model evaluation we chose Cross-Validation. Cross validation is a resampling procedure used to evaluate machine learning models on a limited data sample. This procedure is a statistical method used to estimate the skill of machine learning models.

In this project we used Cross-validation to calculate the accuracy of the model. We used cross_val_score() function. It uses the scoring provided in the given estimator. It is one of the simplest appropriate scoring methods provided by the sklearn. The function cross_val_score() suggests if the model is over fitting or underfitting.

From our calculations, we observed the mean score of this model is 0.9778. It suggests that the model has 97.78% accuracy. This model is based on only classification of the data so it is normal to have this level of accuracy.

7. PROBLEMS ENCOUNTERED:

During the course of this project, we faced numbers of problems and issues from the very beginning of the project. Firstly, we targeted our system to be largely focused on mental health diseases and its precautions but the lack of availability of useful data we were compelled to work with the limited resources. However, we managed to cover some of the most important mental health issues prevailing in the community.

The effects of COVID pandemic have overdone everyone for over a year now and we are no different to the situation. Time and again we were forced to halt the project due to vicious situation created by the pandemic. Beyond these challenges we overcame and successfully met our objectives with very less comptonization.

Some of the technical problems we encountered during the course of this project are:

a) Difficulties with data collection.
b) Problem during integrating project to multiple IDE's.
c) Mapping issues regarding to multiple datasets.

8. OUTCOME:

The outcome of this project is a responsive chatbot based on sklearn where we can give a major symptom and in return the bot asks us multiple yes/no questions related to disease and finally gives the output of the disease description that the user may possess and few preventive measures that the user can follow to feel better. This system is an outcome of productive use of Decision tree classifier providing accuracy of 97.78 %. As mentioned in the objectives for this set of data of more than 5000 entities we targeted to achieve more than 95% accuracy and we have managed to do so. Here are some of the tests we run for different symptoms in our model.

Test 1: Stress

```
Your Name
                                                ->Laxman
hello  laxman
Enter the symptom you are experiencing
                                                ->stress
searches related to input:
0 ) stress
Okay. From how many days ? : 2
Are you experiencing any
anxiety ? : yes
depression ? : no
aggressive ? : yes
insomnia ? : yes
stress ? : yes
fear_of_contamination ? : no
It might not be that bad but you should take precautions.
You may have  OCD or  Migraine
Obsessive-compulsive disorder (OCD) is a disorder in which people have recurring, unwanted
thoughts, ideas or sensations (obsessions) that make them feel driven to do something repetitively
(compulsions). The repetitive behaviors, such as hand washing, checking on things or cleaning, can
significantly interfere with a person’s daily activities and social interactions.
A migraine can cause severe throbbing pain or a pulsing sensation, usually on one side of the head.
It's often accompanied by nausea, vomiting, and extreme sensitivity to light and sound. Migraine
attacks can last for hours to days, and the pain can be so severe that it interferes with your daily
activities.
Take following measures :
1 ) meditation
2 ) consult doctor
3 ) eat healthy
```

Test 2: Headache

```
Your Name
                                                    ->George
hello  george
Enter the symptom you are experiencing
                                                    ->headache
searches related to input:
0 ) headache
Okay. From how many days ? : 5
Are you experiencing any
itching ? : yes
vomiting ? : yes
yellowish_skin ? : no
nausea ? : yes
loss_of_appetite ? : yes
abdominal_pain ? : yes
yellowing_of_eyes ? : no
You should take the consultation from doctor.
You may have  Chronic cholestasis
Chronic cholestatic diseases, whether occurring in infancy, childhood or adulthood, are
characterized by defective bile acid transport from the liver to the intestine, which is
caused by primary damage to the biliary epithelium in most cases
Take following measures :
1 ) cold baths
2 ) anti itch medicine
3 ) consult doctor
4 ) eat healthy
```

Test 3: Headache

```
Your Name
                                                    ->
hello
Enter the symptom you are experiencing
                                                    ->George
Enter valid symptom.
Enter the symptom you are experiencing
                                                    ->Headache
searches related to input:
0 ) headache
Okay. From how many days ? : 7
Are you experiencing any
acidity ? : no
indigestion ? : no
headache ? : yes
blurred_and_distorted_vision ? : yes
excessive_hunger ? : yes
stiff_neck ? : no
depression ? : yes
irritability ? : yes
visual_disturbances ? : yes
You should take the consultation from doctor.
You may have  Migraine
A migraine can cause severe throbbing pain or a pulsing sensation,
usually on one side of the head. It's often accompanied by nausea,
vomiting, and extreme sensitivity to light and sound. Migraine attacks
can last for hours to days, and the pain can be so severe that it
interferes with your daily activities.
Take following measures :
1 ) meditation
2 ) reduce stress
3 ) use poloroid glasses in sun
4 ) consult doctor
```

Note: Test 2 and Test 3 shows the system providing different prediction of disease for the same symptom. This happens due to model working on the basis of severity of the symptom.

Test 4: Red spots over body

```
Your Name
                                                    ->Ravi
hello  ravi
Enter the symptom you are experiencing
                                                    ->red_spots_over_body
searches related to input:
0 ) red_spots_over_body
Okay. From how many days ? : 3
Are you experiencing any
vomiting ? : yes
yellowish_skin ? : yes
abdominal_pain ? : no
swelling_of_stomach ? : no
distention_of_abdomen ? : no
history_of_alcohol_consumption ? : yes
fluid_overload ? : no
It might not be that bad but you should take precautions.
You may have  Alcoholic hepatitis
Alcoholic hepatitis is a diseased, inflammatory condition of the liver caused by heavy
alcohol consumption over an extended period of time. It's also aggravated by binge
drinking and ongoing alcohol use. If you develop this condition, you must stop
drinking alcohol
Take following measures :
1 ) stop alcohol consumption
2 ) consult doctor
3 ) medication
4 ) follow up
```

9. WORKING SCHEDULE:

This major project is done as a part of four years Bachelors in Computer Engineering program at the final semester. We worked on the project according to the working schedule presented below:

		Months										
		December- 2020				June- 2021	July-2021				August-2021	
	Date	11	16	20	25	1 to 30	1	9	25	30	1	2
S.N	Task/ Deliverables											
1	Literature Review											
2	Deliverable: Proposal Report											
3	Proposal Defence											
4	Determining the methods to be used											
5	Implementation											
6	Code Documnetation											
7	Modelling											
8	Deliverable: Mid- Term report											
10	Extracting Results (Simulation)											
11	Results Analysis											
12	Concluding											
13	Final Report Writing (Dissertation)											
14	Deliverable: Final Report											
15	Final Report Defence											

Figure 9. 1 Gantt chart showing working schedule of an E-Health Care Chatbot with Mental Health Consultant using Decision Tree Algorithm

.

10. FUTURE ENHANCEMENTS

In future, this project can be extended. Here in this project work we use limited number of diseases to be diagnosed. Like this this bot can be used for more problems diagnoses. It can even be extended for taking and saving case history of physical diseases. In this way, human expert doesn't have to take case history manually, which will of course save time. One of the lacking we feel in this system is not being able to recommend any doctors. The feature to recommend the specialist doctor and their contact by creating a suitable dataset would be very helpful.

Another future direction could be using chatbots for surveying the mental health of users regularly for some period(months) and counselling them. And passing the result at the end of survey so the user can get proper consultation. In case of mental health, this method is more productive. The human psychologist uses the method of counseling for treatment of psychological problems. Same type of counseling can be done using a chatbot. Using chatting through a chatbot for relaxing a person having some psychological problems like depression etc. is possible. In future Chatbot can be also combined with some techniques which capture facial expression of user; it will be helpful in drawing true picture of user's psychological condition

In terms of technical advancement, we would like to put forward some of the recommendations as follows:

i. This system is based on a very classical machine learning algorithm; Decision tree Classifier. We have seen the Machine Learning field progress to the use of complex algorithms such as ANN (Artificial Neural Network) and CNN (Convolutional Neural Network). With the application of modern deep learning algorithms this system could be improvised even better.

ii. This system is based on Yes/No conversation with the user and is built with a very primitive concept. In the future, this system can be built into having smart conversations using NLP and taking sentences as input.

iii. At this moment, there is no graphical touch to this project but it is to surely possible to make use of different frameworks such as FLASK to create an attractive GUI.

11. CONCLUSION

This project explains a medical chatbot which can be used to replace the conventional method of disease diagnosis and treatment recommendation. Chatbot can act as a doctor. The chatbot acts as a user application. The user of this application can specify their symptoms to the chatbot and in turn, chatbot will specify the health measures to be taken. The information about symptom and diseases are available in the dataset and thus the chatbot instance can provide information about disease and treatment to the user. After classifying the symptoms of the different users, it finally maps the disease to the user and provides details about the disease and precaution.

A smart medical chatbot can be useful to patients by identifying the symptoms as described by them, giving proper diagnosis and providing with suitable treatment for the disease. In the busy life, it is rare for people to frequently visit hospitals for check-ups. Chatbot is of great importance in such situations as they provide diagnostic assistance with a single click of button. Chatbot doesn't require the help of any physician to give proper health measures to the users and this is one of the major advantages of chatbot. Moreover, the cost-effectiveness in using chatbot is a major attractiveness to users. The chat with users is completely personal and this helps users to be more open with their health matters and paves way for chatbot to efficiently identify the disease.

12.REFERENCES

[1] Valderrama, E., Rullán, M., Sánchez, F., Pons, J., Mans, C., Giné, F., Jiménez, L. and Peig, E., 2009, October. Guidelines for the final year project assessment in engineering. In 2009 39th IEEE Frontiers in Education Conference (pp. 1-5). IEEE.

[2]Calvo, R.A., Milne, D.N., Hussain, M.S. and Christensen, H., 2017. Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 23(5), pp.649-68

[3] Cook, B.L., Progovac, A.M., Chen, P., Mullin, B., Hou, S. and Baca-Garcia, E., 2016. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Computational and mathematical methods in medicine, 2016.

[4] Stewart, R. and Velupillai, S., 2020. Applied natural language processing in mental health big data. Neuropsychopharmacology, 46(1), p.252.

[5] Pan, K.Y., Kok, A.A., Eikelenboom, M., Horsfall, M., Jörg, F., Luteijn, R.A., Rhebergen, D., van Oppen, P., Giltay, E.J. and Penninx, B.W., 2020. The mental health impact of the COVID-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: a longitudinal study of three Dutch case-control cohorts. The Lancet Psychiatry.

[6] Barak, A., 2007. Emotional support and suicide prevention through the Internet: A field project report. Computers in Human Behavior, 23(2), pp.971-984.

[7] Abbe, A., Grouin, C., Zweigenbaum, P. and Falissard, B., 2016. Text mining applications in psychiatry: a systematic literature review. International journal of methods in psychiatric research, 25(2), pp.86-100.

[8] Alshawwa, I.A., Elkahlout, M., El-Mashharawi, H.Q. and Abu-Naser, S.S., 2019. An Expert System for Depression Diagnosis.

[9] R. B. Mathew, S. Varghese, S. E. Joy and S. S. Alex, "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 851-856, doi: 10.1109/ICOEI.2019.8862707.

[10] Laijawala, V., Aachaliya, A., Jatta, H. and Pinjarkar, V., 2020. Mental Health Prediction using Data Mining: A Systematic Review. Available at SSRN 3561661.

[11] Mujeeb, S., Hafeez, M. and Arshad, T., 2017. Aquabot: a diagnostic chatbot for achluophobia and autism. Int J Adv Computer Sci Appl, 8(9), pp.39-46.

13.APPENDICES

Loading all the basic Libraries:

```
import pandas as pd
from sklearn import preprocessing
from sklearn.tree import DecisionTreeClassifier,_tree
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.svm import SVC
import csv
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

```

Loading the dataset:

```
df = pd.read_csv('dataset.csv', header=None)
df
```

Preprocessing:

```
training = pd.read_csv('Training.csv')
testing= pd.read_csv('Testing.csv')

#Saving the information of columns
cols= training.columns
cols= cols[:-1]

#Slicing and Dicing the dataset to separate features from predictions
x = training[cols]
y = training['prognosis']
y1= y

#Dimentionality Reduction for removing redundancies
reduced_data = training.groupby(training['prognosis']).max()

#Encoding String values to integer constants
le = preprocessing.LabelEncoder()
le.fit(y)
y = le.transform(y)
print(y)
```

```
[ 1  1  1 ... 42 39 30]
```

Splitting the dataset into training and testing set:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42)
testx    = testing[cols]
testy    = testing['prognosis']
testy    = le.transform(testy)
```

Implementing Decision Tree Classifier:

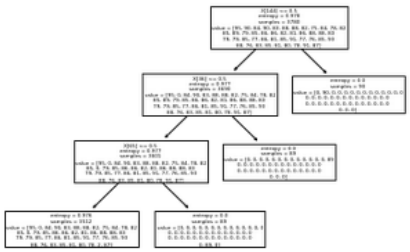
```
model  = DecisionTreeClassifier()
model.fit(x_train,y_train)
```

Model Evaluation using Cross Validation Method:

```
#Model Evaluation Using Cross Validation Method
scores = cross_val_score(clf, x_test, y_test, cv=3)
print (scores.mean())
```

Visualization of decision tree:

```
#Visualization of a decision tree
from sklearn import tree
import matplotlib.pyplot as plt
tree.plot_tree(clf)
plt.show()
plt.savefig('HC-Chatbot.png')
```



```
<Figure size 432x288 with 0 Axes>
```

Decision Tree Model:

```
def print_disease(node):
    node = node[0]
    val  = node.nonzero()
    disease = le.inverse_transform(val[0])
    return disease
def tree_to_code(tree, feature_names):
    tree_ = tree.tree_
    feature_name = [
        feature_names[i] if i != _tree.TREE_UNDEFINED else "undefined!"
        for i in tree_.feature
    ]

    chk_dis=",".join(feature_names).split(",")
    symptoms_present = []
```

```
while True:

    print("Enter the symptom you are experiencing  \n\t\t\t\t\t\t\t",end="->")
    disease_input = input("").lower()
    conf,cnf_dis=check_pattern(chk_dis,disease_input)
    if conf==1:
        print("searches related to input: ")
        for num,it in enumerate(cnf_dis):
            print(num,")",it)
        if num!=0:
            print(f"Select the one you meant (0 - {num}):  ", end="")
            conf_inp = int(input(""))
        else:
            conf_inp=0

        disease_input=cnf_dis[conf_inp]
        break

    else:
        print("Enter valid symptom.")

while True:
    try:
        num_days=int(input("Okay. From how many days ? : "))
        break
    except:
        print("Enter number of days.")
```

```
while True:
    try:
        num_days=int(input("Okay. From how many days ? : "))
        break
    except:
        print("Enter number of days.")
def recurse(node, depth):
    indent = "  " * depth
    if tree_.feature[node] != _tree.TREE_UNDEFINED:
        name = feature_name[node]
        threshold = tree_.threshold[node]

        if name == disease_input:
            val = 1
        else:
            val = 0
        if  val <= threshold:
            recurse(tree_.children_left[node], depth + 1)
        else:
            symptoms_present.append(name)
            recurse(tree_.children_right[node], depth + 1)
    else:
        present_disease = print_disease(tree_.value[node])
        red_cols = reduced_data.columns
        symptoms_given = red_cols[reduced_data.loc[present_disease].values[0].nonzero()]

        print("Are you experiencing any ")
        symptoms_exp=[]
        for syms in list(symptoms_given):
            inp=""
            print(syms,"? : ",end='')
            while True:
                inp=input("")
```

```
            else:
                print("You may have ", present_disease[0], "or ", second_prediction[0])
                print(description_list[present_disease[0]])
                print(description_list[second_prediction[0]])

            #print(description_list[present_disease[0]])
            precution_list=precautionDictionary[present_disease[0]]
            print("Take following measures : ")
            for  i,j in enumerate(precution_list):
                print(i+1,")",j)


    recurse(0, 1)
getSeverityDict()
getDescription()
getprecautionDict()
getInfo()
tree_to_code(clf,cols)
```

User Interface:

```
Your Name
                                                    ->
```

```
Your Name
                                                    ->Kevin
hello  kevin
Enter the symptom you are experiencing
                                                    ->Headache
searches related to input:
0 ) headache

Okay. From how many days ? :
```

```
Your Name
                                                    ->Kevin
hello  kevin
Enter the symptom you are experiencing
                                                    ->Headache
searches related to input:
0 ) headache
Okay. From how many days ? : 7
Are you experiencing any
vomiting ? :
```

Output:

```
Your Name
                                                ->Test
hello  test
Enter the symptom you are experiencing
                                                ->fatigue
searches related to input:
0 ) fatigue
Okay. From how many days ? : 3
Are you experiencing any
vomiting ? : no
headache ? : no
nausea ? : yes
spinning_movements ? : yes
loss_of_balance ? : no
unsteadiness ? : no
It might not be that bad but you should take precautions.
You may have  (vertigo) Paroymsal  Positional Vertigo or  Chronic cholestasis
Benign paroxysmal positional vertigo (BPPV) is one of the most common causes of vertigo â€" the sudden sensati
on that you're spinning or that the inside of your head is spinning. Benign paroxysmal positional vertigo caus
es brief episodes of mild to intense dizziness.
Chronic cholestatic diseases, whether occurring in infancy, childhood or adulthood, are characterized by defec
tive bile acid transport from the liver to the intestine, which is caused by primary damage to the biliary epi
thelium in most cases
Take following measures :
1 ) lie down
2 ) avoid sudden change in body
3 ) avoid abrupt head movment
4 ) relax
```