Jyothi Hills, Panjal Road,
Vettikattiri PO, Cheruthuruthy , Thrissur,
Kerala 679531

Jyothi Engineering College
NAAC Accredited College with NBA Accredited Programmes*
Approved by AICTE & affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKATTIRI P.O, CHERUTHURUTHY, THRISSUR. PIN-679531 PH : +91- 4884-259000, 274423 FAX : 04884-274777
NBA accredited B.Tech Programmes in Computer Science & Engineering, Electronics & Communication Engineering, Electrical & Electronics Engineering and Mechanical Engineering valid for the academic years 2016-2022. NBA accredited B.Tech Programme in Civil Engineering valid for the academic years 2019-2022.

QRTBOT - Chatbot application for people in quarantine

MAIN PROJECT REPORT

JOSHUA JOSEPH (JEC17CS059)

JOSIN GEORGE (JEC17CS060)

SANGEETHA C P (JEC17CS086)

SHILPA SIVADAS (JEC17CS093)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY (B.Tech)

in

COMPUTER SCIENCE & ENGINEERING

of

A P J ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Under the guidance of

Mr. UNNIKRISHNAN P

CREATING TECHNOLOGY LEADERS OF TOMORROW

JANUARY 2021

Department of Computer Science & Engineering

Jyothi Hills, Panjal Road,
Vettikattiri PO, Cheruthuruthy , Thrissur,
Kerala 679531

Jyothi Engineering College
NAAC Accredited College with NBA Accredited Programmes*
Approved by AICTE & affiliated to APJ Abdul Kalam Technological University
A CENTRE OF EXCELLENCE IN SCIENCE & TECHNOLOGY BY THE CATHOLIC ARCHDIOCESE OF TRICHUR
JYOTHI HILLS, VETTIKATTIRI P.O, CHERUTHURUTHY, THRISSUR. PIN-679531 PH : +91- 4884-259000, 274423 FAX : 04884-274777
NBA accredited B.Tech Programmes in Computer Science & Engineering, Electronics & Communication Engineering, Electrical & Electronics Engineering and Mechanical Engineering valid for the academic years 2016-2022. NBA accredited B.Tech Programme in Civil Engineering valid for the academic years 2019-2022.

QRTBOT - Chatbot application for people in quarantine

MAIN PROJECT REPORT

JOSHUA JOSEPH (JEC17CS059)

JOSIN GEORGE (JEC17CS060)

SANGEETHA C P (JEC17CS086)

SHILPA SIVADAS (JEC17CS093)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY (B.Tech)

in

COMPUTER SCIENCE & ENGINEERING

of

A P J ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Under the guidance of

Mr. UNNIKRISHNAN P

CREATING TECHNOLOGY LEADERS OF TOMORROW

JANUARY 2021

Department of Computer Science & Engineering

Department of Computer Science and Engineering

JYOTHI ENGINEERING COLLEGE, CHERUTHURUTHY

THRISSUR 679 531

JANUARY 2021

BONAFIDE CERTIFICATE

This is to certify that the main project report entitled **QRTBOT - Chatbot application for people in quarantine** submitted by **Joshua Joseph (JEC17CS059), Josin George (JEC17CS060), Sangeetha C P (JEC17CS086) and Shilpa Sivadas (JEC17CS093)** in partial fulfillment of the requirements for the award of **Bachelor of Technology** degree in **Computer Science and Engineering** of **A P J Abdul Kalam Technological University** is the bonafide work carried out by them under our supervision and guidance.

Mr. Unnikrishnan P	**Mr. Shaiju Paul**	**Dr. Vinith R**
Project Guide	Project Coordinator	Head of The Dept &
Assistant Professor	Assistant Professor	Associate Professor
Dept. of CSE	Dept. of CSE	Dept. of CSE

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

COLLEGE VISION

Creating eminent and ethical leaders through quality professional education with emphasis on holistic excellence.

COLLEGE MISSION

- To emerge as an institution par excellence of global standards by imparting quality engineering and other professional programmes with state-of- the-art facilities.
- To equip the students with appropriate skills for a meaningful career in the global scenario.
- To inculcate ethical values among students and ignite their passion for holistic excellence through social initiatives.
- To participate in the development of society through technology incubation, entrepreneurship and industry interaction.

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

DEPARTMENT VISION

Creating eminent and ethical leaders in the domain of computational sciences through quality professional education with a focus on holistic learning and excellence.

DEPARTMENT MISSION

- To create technically competent and ethically conscious graduates in the field of Computer Science & Engineering by encouraging holistic learning and excellence.
- To prepare students for careers in Industry, Academia and the Government.
- To instill Entrepreneurial Orientation and research motivation among the students of the department.
- To emerge as a leader in education in the region by encouraging teaching, learning, industry and societal connect

PROGRAMME OUTCOMES (POs)

1. **Engineering Knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem Analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/Development of Solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations of Complex Problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern Tool Usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The Engineer and Society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and Team Work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project Management and Finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-Long Learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

1. The graduates shall have sound knowledge of Mathematics, Science, Engineering and Management to be able to offer practical software and hardware solutions for the problems of industry and society at large.
2. The graduates shall be able to establish themselves as practising professionals, researchers or Entrepreneurs in computer science or allied areas and shall also be able to pursue higher education in reputed institutes.
3. The graduates shall be able to communicate effectively and work in multidisciplinary teams with team spirit demonstrating value driven and ethical leadership.

PROGRAMME SPECIFIC OUTCOMES (PSOs)

1. An ability to apply knowledge of data structures and algorithms appropriate to computational problems.
2. An ability to apply knowledge of operating systems, programming languages, data management, or networking principles to computational assignments.
3. An ability to apply design, development, maintenance or evaluation of software engineering principles in the construction of computer and software systems of varying complexity and quality.
4. An ability to understand concepts involved in modeling and design of computer science applications in a way that demonstrates comprehension of the fundamentals and trade-offs involved in design choices.

COURSE OUTCOMES (COs)

C410.1 The students will be able to analyse a current topic of professional interest and present it before an audience.

C410.2 Students will be able to identify an engineering problem, analyse it and propose a work plan to solve it.

C410.3 Students will have gained thorough knowledge in design, implementations and execution of Computer science related projects.

C410.4 Students will have attained the practical knowledge of what they learned in theory subjects.

C410.5 Students will become familiar with usage of modern tools.

C410.6 Students will have ability to plan and work in a team.

ACKNOWLEDGEMENT

We take this opportunity to express our heartfelt gratitude to all respected personalities who had guided, inspired and helped us in the successful completion of this interim project. First and foremost, we express our thanks to **The Lord Almighty** for guiding us in this endeavour and making it a success.

We take immense pleasure in thanking the **Management** of Jyothi Engineering College and **Dr. Sunny Joseph Kalayathankal**, Principal, Jyothi Engineering College for having permitted us to carry out this seminar. Our sincere thanks to **Dr. Vinith R**, Head of the Department of Computer Science and Engineering for permitting us to make use of the facilities available in the department to carry out the interim project successfully.

We express our sincere gratitude to **Mr. Shaiju Paul & Dr. Swapna B Sasi**, project coordinators for their invaluable supervision and timely suggestions. We are very happy to express our deepest gratitude to our mentor **Mr. Unnikrishnan P**, Assistant Professor,Department of Computer Science and Engineering, Jyothi Engineering College for his able guidance and continuous encouragement.

Last but not least, we extend our gratefulness to all teaching and non-teaching staff who directly or indirectly involved in the successful completion of this interim project work and to all friends who have patiently extended all sorts of help for accomplishing this undertaking.

ABSTRACT

COVID-19 was first discovered in December 2019 and has continued to rapidly spread across countries worldwide infecting thousands and millions of people. The virus is deadly, and people who are suffering from prior illnesses or are older than the age of 60 are at a higher risk of mortality. Medicine and Healthcare industries have surged towards finding a cure, and different policies have been amended to mitigate the spread of the virus. The project aims at introducing a medical chatbot called QRTBOT for COVID-19 assistance during quarantine using Machine Learning and Natural Language Processing. The chatbot is built to be a conversational agent that motivates users to discuss about their health issues and returns the diagnosis of disease through a series of queries. It also identifies the emotional state of the user through the text. The main goal of this project is to keep the users updated regarding this infectious disease COVID 19 and report to the health department and also help the people to stay from loneliness which further doesn't lead to depression and keep them mentally fit.

Keywords: Natural Language Processing, chatbot, COVID-19, Symptom prediction, Sentiment Analysis, Machine Learning.

CONTENTS

List of Figures

List of Abbreviations

COVID	: *Novel Corona Virus Disease*
ML	: *Machine Learning*
NLP	: *Natural Language Processing*
SVM	: *Support Vector Machine*
CDC	: *Centers for Disease Control and Prediction*
WHO	: *World Health Organization*
API	: *Application Programming Interface*
GCP	: *Google Cloud Platform*
VUI	: *Voice User Interface*
SSML	: *Speech Synthesis Markup Language*
NLU	: *Natural Language Understanding*
NLG	: *Natural Language Generation*
LSTM	: *Long Short Term Memory*
RNN	: *Recurrent Neural Network*
KNN	: *K − Nearest Neighbour*
POS	: *Part of Speech*
SDLC	: *System Development Life Cycle*
NLTK	: *Natural Language Toolkit*
SQL	: *Structured Query Language*
CRF	: *Conditional Random Field*

CHAPTER 1
INTRODUCTION

1.1 Overview

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment. The huge number of deaths caused by the novel pandemic COVID-19, which can affect anyone of any sex, age and socio-demographic status in the world, presents a serious threat for humanity and society. At this point, there are two types of citizens, those oblivious of this contagious disaster's danger that could be one of the causes of its spread, and those who show erratic or even turbulent behavior since fear and anxiety invades our surroundings because of confinement and panic of being affected. It was first reported in December 2019 in Wuhan City in China . Up to April 2020, more than 144,000 people have died globally from the COVID-19, while more than 2 millions infections have been confirmed in dozens of countries, according to the World Health Organization, as a result the COVID-19 is now declared a pandemic. In fact, while the number of people who are being treated for COVID-19 is increasing by the day, some citizens are not aware of the real threat of this outbreak which explains its quickly spread all over the world, however others, panicked and desperate, fell headlong into the trap of this grim and even worse committing suicide.

This project aims at proposing a medical chatbot that helps users get engaged during quarantine days. It creates awareness among the users about the risk factors of the COVID-19 and the precautions that must be taken. Through a series of queries using desicion tree algorithm the chatbot predicts if the users may have contracted COVID-19. It also analyse the emotional state of the user at the point of time through their chat replies and perform corresponding actions based on their current mood.

1.2 Objectives

The main objective of this project is to introduce an interactive chatbot that help users get aware of the symptoms of novel coronavirus and check if they may have contracted COVID-19 using Machine Learning. It also analyzes the mental state of the user.

1.3 Data Description

The data for this project is taken from an open source platform known as Kaggle and is available at https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge. The data-set as a whole is generally divided into two categories: training and test. As is the case with a usual machine learning problem, we would be training the model using training dataset and evaluating the performance with the test dataset.

1.4 Organization of the project

The report is organised as follow:

- **Chapter 1:Introduction** Gives an introduction "Chatbot facility for people who are being quarantined ".
- **Chapter 2:Literature Survey** Summarizes the various existing techniques that helps in achieving the desired result.
- **Chapter 3: Problem Statement** Discuses about the need for the proposed system
- **Chapter 4:Project Management** Contains the effective project management model to be used for the project.
- **Chapter 5:Proposed System** Describes the various steps involved to produce this project.
- **Chapter 5:System Requirements & Specification**Describes the various technologies needed for implementation.
- **Chapter 6:Conclusion** Concludes with the future scope of implementation.
- **References** Includes the references for the project.

CHAPTER 2
LITERATURE SURVEY

2.1 Chatbots in the fight against the COVID-19 pandemic

During the novel coronavirus (COVID-19) pandemic, institutions like the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) have begun utilizing chatbots to share information, suggest behavior, and offer emotional support.The CDC has named theirs "Clara".[7]Chatbots are software programs that talk with people through voice or text in their natural language. Some well-known examples include "Alexa" from Amazon, "Siri" from Apple, and "Cortana" from Microsoft. They often come pre-installed on smartphones or home-based smart speakers. In recent years, chatbot use for health-related purposes has increased considerably, from supporting clinicians with clinical interviews and diagnosis to aiding consumers in self-managing chronic conditions. While promising, the use of chatbots may pose safety risks. Chatbots have varied widely in their responses to questions about physical health, suicide, intimate partner violence, substance abuse, and other sensitive conversations. In one study, about a third (29%) of chatbot responses to health questions could have caused harm, and about half of those (16%) could have resulted in death if acted upon. The COVID-19 pandemic puts in stark relief the potential for chatbots to help save lives.

- **Special Features of Pandemics**:

  Pandemics have unique characteristics that make them amenable to tailored interventions deliverable via chatbots. In particular,pandemics differ from other natural disasters in three key ways. First, individual actions can significantly worsen outcomes in a pandemic, given that a single person may infect many others depending on their behavior. Second, the fear of infecting others, especially loved ones or healthcare workers, makes infectious diseases more insidious through disease-related stigma. As a result, people can feel personally responsible for bad outcomes during a pandemic and also hide symptoms from others.[5] Third, the physical gatherings typically used to connect with others in difficult times (e.g., family meals, community centers, sports, spiritual and religious events) are exactly what we are supposed to avoid during a pandemic, worsening the risk for future mental health problems. Chatbots have unique affordances, outlined below, which may mitigate short- and long- term disease burden during infectious disease pandemics

- **Information dissemination**

  During a pandemic, people do not know what to do. Doing too little (e.g., not following prophylactic measures) can increase everyone's risk of infection. Doing too much (e.g., going to the emergency room for mild symptoms) can overburden the healthcare system, wasting precious resources. Thus, reliable information sources are crucial to prevent a "misinfodemic": the spread of a disease facilitated by viral misinformation.[10] For instance, during the Zika outbreak in 2016, misleading posts spread faster and were more popular than accurate posts on the large social-media site, Facebook. Because chatbots provide a single answer to most questions, they are able to present concise information from credible sources, which may be less overwhelming than social media or web search engines' long list of results. This matters because false news spreads online both faster and further than accurate news18. Chatbots, in contrast to newspapers and online information sources, can often hear and respond in natural language, improving access for people who cannot read or have difficulty using the internet. They can be available any time of the day to answer questions with up-to-date information, and unlike human experts, can concurrently speak with millions of people at the same time in local languages and dialects.

- **Symptom monitoring**

  During a pandemic, both individuals and institutions want to know how and where infections are spreading. Individuals want to avoid getting sick, and institutions such as hospitals or local governments need data-informed policies to increase capacity (i.e., ordering more testing kits) and to plan social interventions (e.g., closing businesses). However, efforts to quickly and accurately gather population level infection rates are stymied by individuals' fear that disclosing symptoms may harm their professional and social lives. Chatbots may be uniquely well suited for symptom screening in a pandemic because people with stigmatized conditions often avoid seeking health care and education. Prior research suggests people are more willing to disclose sensitive personal symptom information to a chatbot than to a human.[4] This means that people may be more forthcoming with chatbots than other humans, providing timelier and more accurate personal triage and population-level infection rate estimates. Healthcare organizations, large corporations like Apple, Amazon, Facebook, Microsoft, and Tencent, governmental agencies like the CDC, and non-governmental organizations like the WHO have launched or helped develop COVID-19 focused chatbots on platforms available to billions of users, likely with the aim of increasing accessibility.

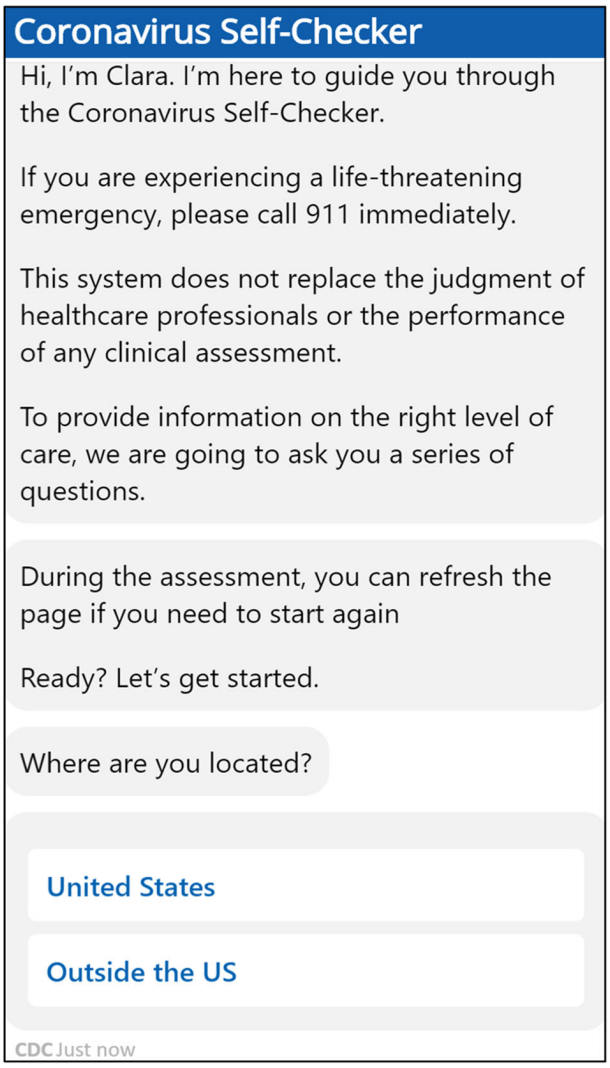

Figure 2.1: CDC coronavirus chatbot

- **Challenges**

Chatbots may be uniquely useful in a pandemic, but challenges in information dissemination, symptom monitoring, behavior change, and mental health support are worthy of attention. Providing reliable evidence-based information is critical in a pandemic and two issues have material impact: conflicting advice between global and local authorities, and misinformation18. Chatbot developers must decide whose voice to amplify and should provide reliable information from global sources like the WHO, while also coordinating with regional authorities. Both a feature and a challenge of chatbots is their ability to link users to third-party services (e.g., "skills") that then gather and share data with unknown or unexpected consequences. If deployed for symptom screening, which is currently happening for COVID-19, constitutional and regulatory boundaries are tested by sharing health-related information between companies and governments. This concern is

not theoretical, as both the United States and Israel have reportedly explored using digital contact tracing to understand infection vectors. Finally, although chatbots have demonstrated feasibility in behavior change and mental health treatment, they are untested in pandemics and have demonstrated limits in health crisis detection and response.

These challenges, if only addressed in real time during a crisis, may lead to erroneous outputs from a lack of testing. With more than a billion voice searches per month, any health-related mistakes, such as misidentifying key symptoms, would be amplified with extensive harmful repercussions. . Additionally, medical and public health experts must inform what chatbots say, and how they say it. Translating medical information into advice for the public requires expertise and evaluation to prevent unintended consequences. Without proper design and deployment, and ongoing monitoring, chatbots may confuse rather than help users.

2.2 Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19

One of the major challenges that India as a country faces is to cater to good quality and affordable healthcare to its growing population.[1] The World Health Report issued by WHO has ranked India's healthcare system at 112 out of 190 countries. This inaccessibility of healthcare facilities especially in rural India and the intricacy in accessing means of transport further causes patients to postpone their treatment, or opt for medical facilities that may be closer but at the same time are not cost-efficient and well-matched to their medical needs. To seek more efficient ways to provide timely medical care, access and quality treatment to the patient, the role of Telemedicine comes into play which connects patients with healthcare providers and healthcare information.

Due to the recent "COVID-19" pandemic, social distancing will stay in India for a long time, especially for patients with chronic diseases, thereby imposing a hindrance for the population to access healthcare facilities.

The data released by the National Health Mission, amid COVID-19 shows that there has been a fall in other acute illnesses being reported during the lockdown in India. This data indicates that a reduced hospitalization case indicates a lack of access to healthcare, rather than a lack of illness. In this alarming situation, telemedicine acts as a boon for people. By using conversational artificial intelligence, healthcare providers can diagnose and treat patients without the need for a personal visit, whilst promoting social distancing and reducing the risk

of COVID-19 transmission.

Tele-Health is the distribution of health-related services via electronic and telecommunication technologies. It enables long-distance patients to get care, advice, reminders, education, monitoring, and remote admissions from clinicians. A chatbot is a conversational agent that communicates with users using natural language. Though there exist some applications that serve as virtual healthcare consultants, none of them provides generic healthcare information, preventive measures, home remedies, and consultation for India-specific context with multi-lingual support. India being a country with a diverse population speaking different languages, access to healthcare at present has multiple barriers including language, lack of healthcare professionals, and lack of access to facilities in rural and remote areas and costs associated with medical consultation. Therefore, an application "Aapka Chikitsak" is developed to provide users healthcare consultation, counseling and information with multi-lingual support (for now, English and Hindi) to improve the healthcare and well-being of the growing population in India and continue provision of healthcare access at ease post the lockdown as well.

In recent years, serverless architectures (Functions-as-a-Service) are gaining traction as an alternative way of providing backend services without requiring a dedicated infrastructure. Serverless allows its users to deploy their stateless functions into platform infrastructures. This stateless behavior makes every invocation independent of the previous runs. For our application, Firebase Cloud Functions and Google Cloud Platform as our backend infrastructure is provided.[2]

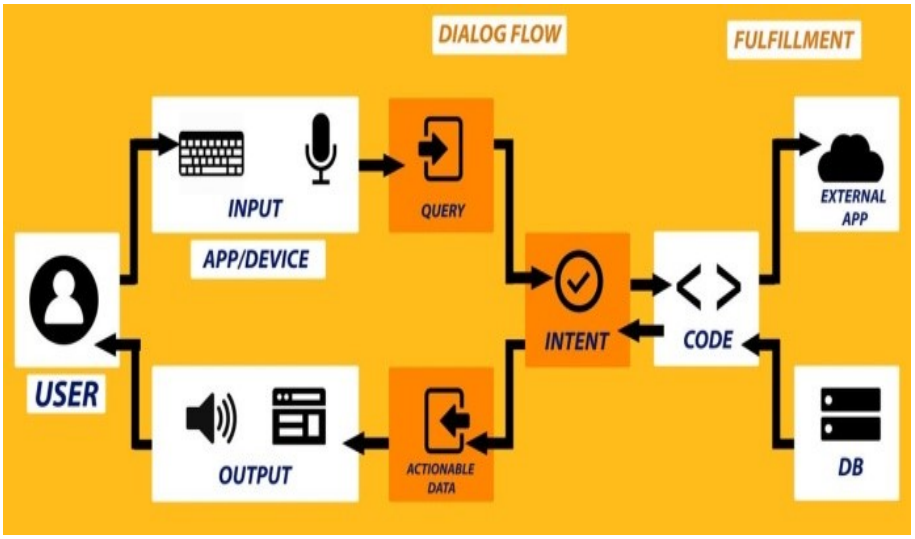

Figure 2.2: The Architecture of the Conversational Bot

The conversation of our bot has been framed and designed in a way to mimic human

behavior to develop a user-friendly chat system that allows them to feel at ease, overcoming the bias of machine interaction. While designing the conversational flows, multitudinous utterances that can be used by the users while interacting with the chat system is considered. Our application uses the Dialogflow Conversation API for creating an automated conversational chat system to hold a conversation with the user by understanding the natural language. Dialogflow employs Google's ML expertise, its products like Google Cloud Speech-to-Text and runs on GCP (Google Cloud Platform) that allows scaling of applications to hundreds of millions of users. When a user chats with our application, there can be a broad variety of utterances that they can utilize aiming at the same purpose. In particular, Dialogflow provides a high-level dialogue flow to identify user queries by mapping them to intents that have been trained with an extensive amount of training phrases pool. Training phrases are certain model phrases that users would use, referred to as end-user expressions. An Intent is a collection or a category of such related end-user expressions during one conversation turn. In our conversational design, it created 255 intents and each of them has been trained with multiple user utterances that were collected during user-testing. By recognizing the intent of the query, Dialogflow chooses the specific actions to fulfill the intent. Entities in Dialogflow are meaningful sequences of characters or lexemes. It is used to extract specific keywords in a user's query. Entities allow us to map the synonyms to its reference value. A typical voice-user interface (VUI) makes possible spoken human chat with computers via speech recognition to understand spoken words and answer questions, and specifically uses text to speech to create a reply. Speech Synthesis Markup Language (SSML) is used to make the voice experience more interactive and robust. Several points are kept while designing the voice user interface of our conversational system. Our bot has its personality so that it sounds more human-like, convincing, and user-friendly. The flow of the conversation consists of short and concise replies in layman language that aids in saving time and makes the interaction more human-like. It includes various health tips and suggestions to increase interactivity and encourage greater engagement. Our bot is devised with multilingual support which has aided us to target both rural and urban areas in India. It structures the VUI of our conversational bot in such a way, to give responses with the focus on having a doctor-patient-like conversation. Our bot has its naturalness as it duly greets users and is trained to give suggestions, health tips, location-based food recommendations according to the patient's concerns and queries.

- Providing information about the majority of prevalent diseases in India along with their possible symptoms and applicable preventive measures.
- Home remedies and quick remedies for common illnesses.
- Interactive counseling sessions for emotional support and expectant mothers.

- Local food recommendations and recommended diet depending upon the geographic location of the user.
- Vaccine administration information.

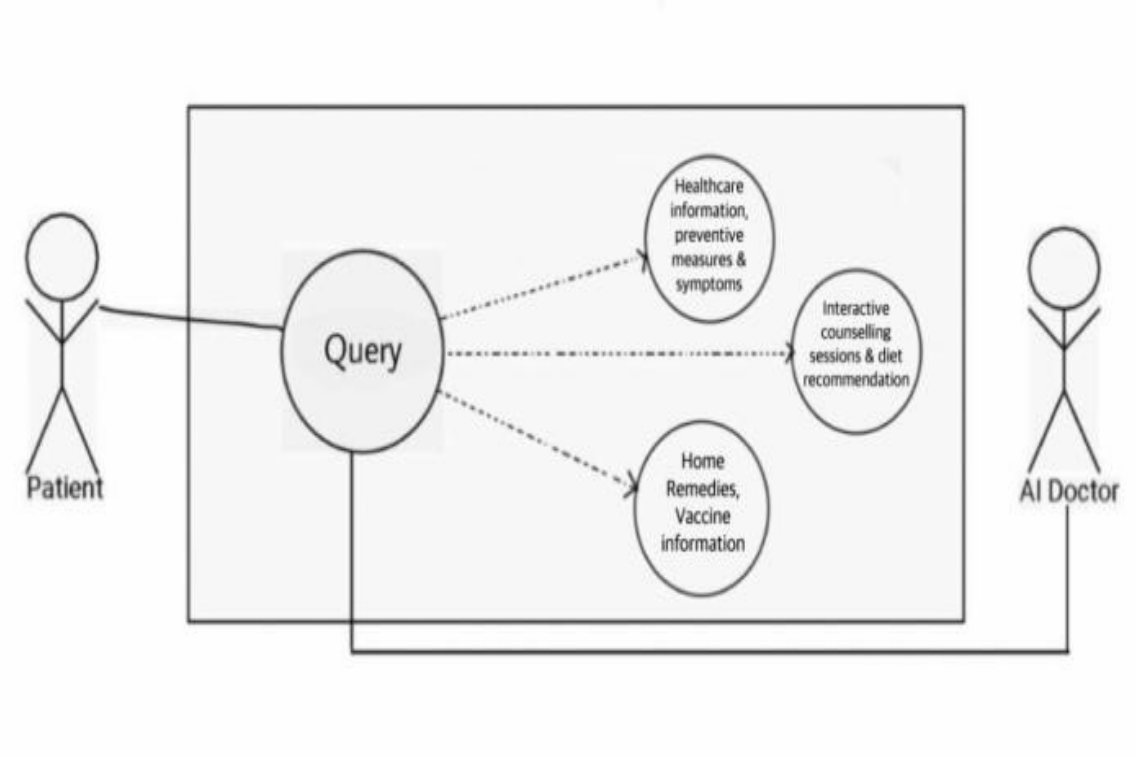

Figure 2.3: UML diagram showing the use cases of the Bot

2.3 Compassion Driven Conversational Chatbot Aimed for better Mental Health

In 2017, there were 197·3 million people with mental disorders in India, comprising 14·3% of the total population of the country. Mental disorders contributed 4·7% of the total DALYs in India in 2017, compared with 2·5% in 1990.[9] Mental illnesses are health conditions that involve change in thinking through one's behavior and emotion. These illnesses are associated with stress and/or problems due to ongoing work, financial, society or family activities. As reported latest in 2014, the number of mental health professionals in India was as low as "one in 100,000 people". The average suicide rate in India is 10.9 for every lakh people and the majority of people who commit suicide are below 44 years of age. Mental illness, also called mental health disorders, refers to a wide range of mental health conditions — disorders that affect your mood, thinking and behavior. Mental illness is treatable and the stigma around needs to be stopped. Many people suffering from mental illness do not wish to talk about it due to fear of being judged and treated differently. But mental illness is nothing to be ashamed of. Mental health problems include anxiety, depression, OCD, Panic attacks, bipolar disorder, DAD, loneliness, paranoia, PTSD, PMDD etc. and the list goes on. Bottomline is that a large number of population is getting affected due to mental health issues and this needs to be curbed right now and with lack of doctors and not many resources for people how do you take care of this problem.

In this regard there have been applications like Wysa chatbot which is one of a kind chatbot which learns every user's experience and builds reports accordingly. Also the Pacifica app also designed to target anxiety, stress through Cognitive Behavioural Therapy and mindfulness.[3] In this paper, the stigma around mental health and how the chatbot is designed to combat situations is discussed . The design and methodology has been discussed for how the chatbot will be implemented and what differentiates it from previously used chatbots in healthcare. Henceforth, the paper also discusses the architecture for the same and how the user responses are understood and then accordingly attended by the bot

2.3.1 Buddy : The 3AM Friend

The chatbot "Buddy" is a compassion-driven AI chatbot that works as your 3AM friend. It is an artificial intelligence based mobile chatbot app for initiating positive conversations, building mental resilience and focuses on improving mental well-being using a text based interface of its own. It has the ability of holding conversations just like any of your friend would and therefore helps in times when people often feel isolated. Instead of rule based or retrieval based , this chatbot is designed by generative based learning that there is no set of questions or rules that the chatbot will definitely ask , it learns from the older conversations and is capable of answering the conversations which it encounters very first time. This application drives the conversation towards a positive scope and tries to understand and assist the user in handling the situations of stress better. The application also helps them develop self-expression. The main benefit of the application is that it entirely free and available to you 24x7 so you always have someone to share your problems with.

The approach picked for this chatbot uses Deep Learning and Natural Language processing for generating and understanding responses , i.e generative approach.

- **Using Natural Language Understanding and Natural Language Generation**

  The very first task in hand is to understand the intent of the user for which the famous natural language understanding concept has been utilized. The job of NLU unit is to transform the user utterance to some semantic format that can be easily understood by the system which includes two very important tasks : Slot filling and Intent detection. The job of filling the slots and detecting the intent of user is seen as a sequence tagging problem. Just for this reason,the LSTM based recurrent neural network have been used for the implementation of NLU unit. Natural Language Generation is the process of generating text from a meaning representation. For tasks like machine translation, text summarization and dialog systems, NLG systems provide a much critical role. Chatbots that use the rule-based NLG , the outputs are predefined template statements for a given frame thus have limited response without adding anything of its own.

- **Generative Based Approach**

  The generative approach used here doesn't use any predefined templates or sentences rather uses deep learning. Therefore unlike the usual retrieval based where we translate from one language to another, here we translate from an input to an output. Seq2Seq model framework most suited for this type of chatbot eliminates the dialog problem faced in rule-based or retrieval-based approaches.

With the aid of deep-learning, generative based has yielded great results. The Sequence-to-Sequence model architecture has two RNNs with different sets of arguments or parameters. This particular approach consists of 2 RNNs, the first one is used to encode the input sequence , what we call as the encoder, and the second one is used to decode the encoded input sequence into the output or target sequence, known as the decoder. This idea has provided great results for text summarization and conversation based on questions and answers where the sequence of words matter and the system needs to remember the significance of ordering.

- **Data Processing**

  The data pre-processing is a key part of the entire process because making the data ready for modeling plays important role on how efficient the system will turn out to be. Reading the words, converting to index of each word, word to index dictionaries, filter too long and too short sequences, bag of words, array dealt with it, tokenizing with "End of sentence" and "Start of sentence" to make sentences more clear are all done in pre-processing phase. The data pre-processing has changed it from raw lines of conversations to zero-padded numpy arrays of indices, and further required dictionaries of word2index and unit index2word to keep counts. It is important to note here that the LSTM takes only batches of inputs and not a single input at a time. With a few helper functions, we gather random examples from dataset, in batches and feed to model. The most important class Seq2Seq that does all the training, evaluation, graph building, saving model takes several parameters as input like batch size, num-layers, emb-dim, learning rate, epochs etc.

  The LSTM cell that is the most important part of the entire model has a keep-prob placeholder which is used to control dropout rate and prevent overfitting. The basic LSTM cell used is wrapped with a Dropout Wrapper. The LSTM cells here are stacked LSTM architecture.Next we used a function embedded-rnn-seq2seq from seq2seq module, to create a seq2seq model, which does word embedding internally. For testing also, we use the same model that has almost the same parameters as training but with the feed previous switch enabled. Thereby the decoder has to use the output of previous step as input to its present timestep, while during training, the input to decoder is taken from real output sequence. The function sequence loss has been used to get the expression for loss. The next step is training with a dropout selected as 0.5. The data was split into training, testing and validation and further the results were analyzed.

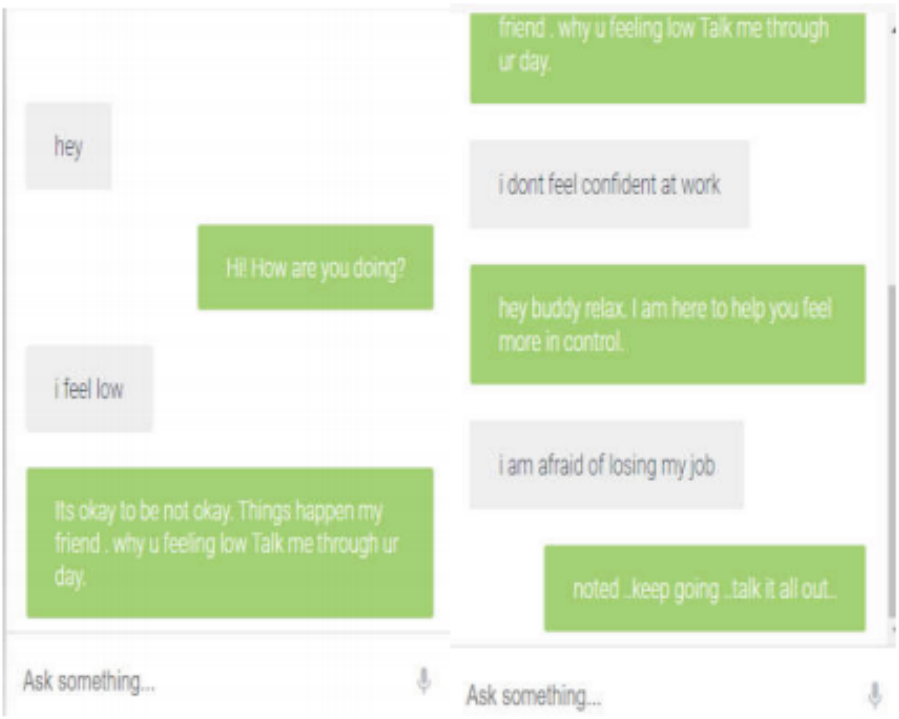

Figure 2.4: Sample images of how chatbot "Buddy" holds up conversation and interacts.

2.4 Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning

A prosperous society is when its entire people are healthy. It is important to maintain the health if one wishes to be happy. Only a healthy body can have a healthy mind and it has a positive impact on the performance of people. Nowadays, people are less aware of their health. In their busy life, they forget to take suitable measures to maintain their health and are less aware of their health status. Most people comprising the working section of the society claim that their hectic schedule gives them no time for periodic medical check-ups and that they disregard any uneasiness shown by their body until it is too severe. Hospitals are the most widely used means by which a sick person gets medical check-ups, disease diagnosis and treatment recommendation. This has been a practice by almost all the people over the world. People consider it as the most reliable means to check their health status.

The proposed system[6] creates an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get diagnosis. A medical chatbot is 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 machine learning algorithm employed here is K-nearest neighbor algorithm (KNN). This clearly shows that a medical chatbot can somewhat accurately diagnose patients with simple symptom analysis and a conversational approach done with the help of natural language processing.

The main objective of the proposed system is to have the importance of health in life reach out to people and encourage people to follow measures to maintain health by making the chatbot available to all. Chatbot and health have a history of working well together. It creates a good human-like conversational environment for interaction between the user and the system.

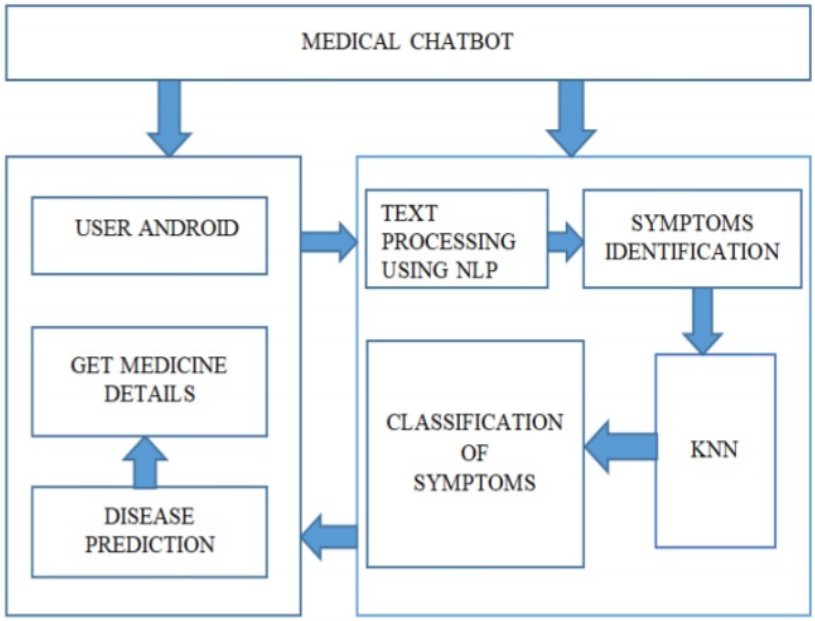

Figure 2.5: Working model of the proposed system

In this system, the user talks about their health and it is a great way for the users to regulate the healthy lifestyle. An important aspect of this system is that talking with a non-human entity provides a sense of security especially when it comes to mental health as it remains as a confidential meeting with the diagnosis being available only to the user. This system is meant to help and deliver immediate actions where humans cannot reach due to timing or budget as it is readily available and free of cost. It allows the user to have free medical check-up based on the symptoms where the user's health issue is easily identifiable.

The chatbot allows user to login to the system. User registers on chatbot application. They need to submit some personal details which will be confidential. User then interacts with the system and the words are recognized by the use of natural language processing and the system recognizes symptoms of the user. There is an admin who controls the chatbot application. The admin views the details of all the users and can even manually add, delete or update symptoms and diseases. The chatbot is trained on symptoms-disease dataset. From the symptoms identified by the user, KNN algorithm can predict the disease, depending on the dataset. The system recognizes the disease and finally recommends the suitable treatment needed for the same.

The chatbot helps to encourage patients to discuss about their medical issues and provides a suitable diagnosis and recommends treatment. Here the text processing is done using natural language processing. Initially the user enters their symptoms through text. It uses the machine learning algorithm KNN. KNN identifies the symptoms from the interaction with the user. KNN maps the symptoms to the particular disease and finally recommends the suitable treatment to the user.

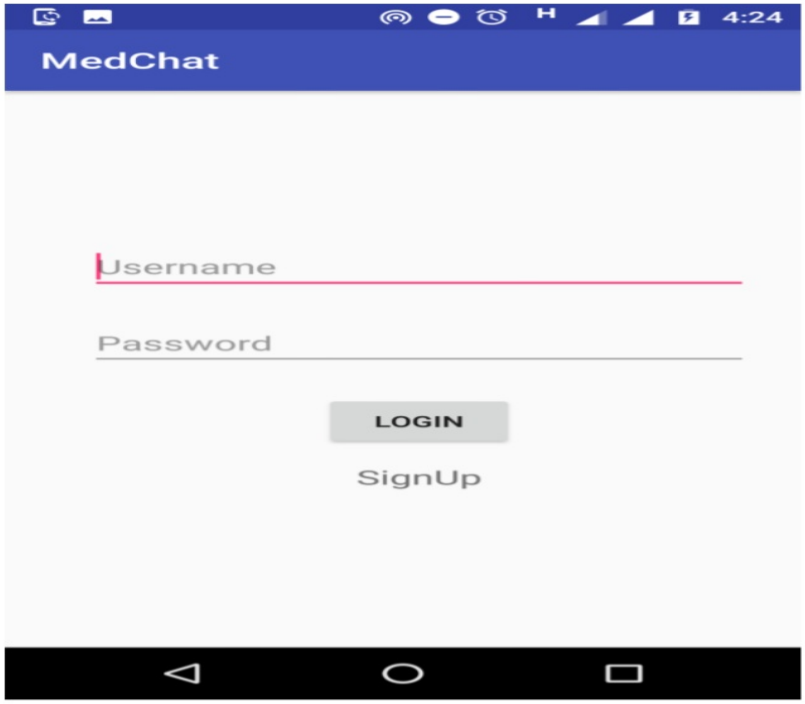

Figure 2.6: User login module

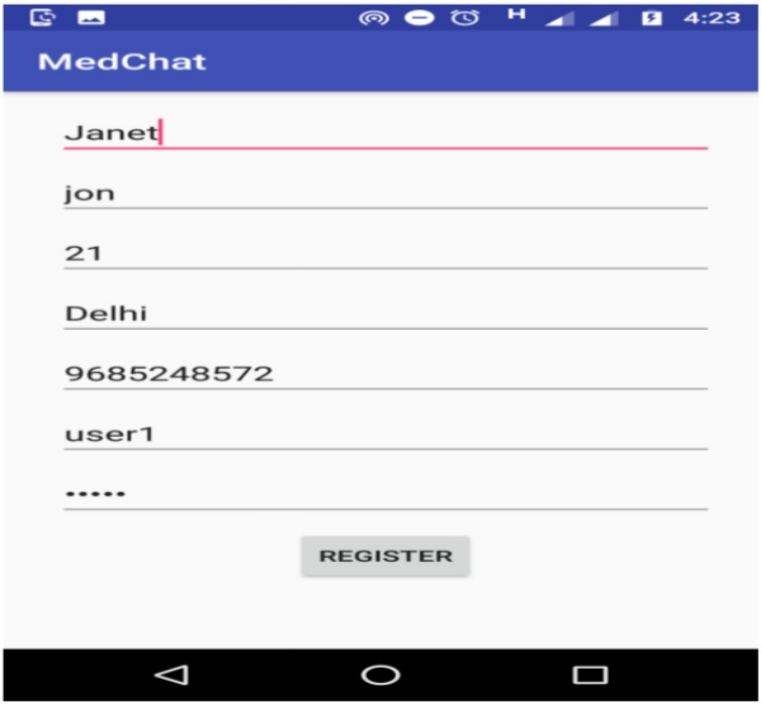

Figure 2.7: User register module

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. General 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 analyzing the symptoms of the different users, it finally predicts the disease to the user and provides with a link where details about the treatment is visible.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 doesnt 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.

2.5 Sentimental Analysis- Capturing favor ability using Natural Language Processing

A technique to detect favorable and unfavorable opinions toward specific subjects (such as organizations and their products) within large numbers of documents offers enormous opportunities for various applications. It would provide powerful functionality for competitive analysis, marketing analysis, and detection of unfavorable rumors for risk management.

For example, enormous sums are being spent on customer satisfaction surveys and their analysis. Yet, the effectiveness of such surveys is usually very limited in spite of the amount of money and effort spent on them, both because of the sample size limitations and the difficulties of making effective questionnaires. Thus there is a natural desire to detect and analyze favorability within online documents such as Web pages, chat rooms, and news articles, instead of making special surveys with questionnaires. Humans can easily recognize natural opinions among such online documents. In addition, it might be crucial to monitor such online documents, since they sometimes influence public opinion, and negative rumors circulating in online documents may cause critical problems for some organizations.

However, analysis of favorable and unfavorable opinions is a task requiring high intelligence and deep understanding of the textual context, drawing on common sense and domain knowledge as well as linguistic knowledge. The interpretation of opinions can be debatable even for humans. For example, when we tried to determine if each specific document was on balance favorable or unfavorable toward a subject after reading an entire group of such documents, we often found it difficult to reach a consensus, even for very small groups of evaluators. Therefore, we focused on finding local statements on sentiments rather than analyzing opinions on overall favorability. The existence of statements expressing sentiments is more reliable compared to the overall opinion.

In this paper[8], we discuss issues of sentiment analysis in consideration of related work and define the scope of our sentiment analysis in the next section. Then we present our approach, followed by experimental results. We also introduce applications based on our sentiment analysis.

Based on human perception experiments, Mel-Frequency analysis is employed to re-weight dimension of frequency and gain more perceptually-relevant representation of speech audio.

The essential issue in sentiment analysis is to identify how sentiments are expressed in texts and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject. Thus, sentiment analysis involves identification of

- Sentiment expressions,
- Polarity and strength of the expressions, and
- Their relationship to the subject

These elements are interrelated. For example, in the sentence, "XXX beats YYY", the expression "beats" denotes a positive sentiment toward XXX and a negative sentiment toward YYY. However, most of the related work on sentiment analysis to date has focused on identification of sentiment expressions and their polarities. Specifically, the focus items include the following:

- Features of expressions to be used for sentiment analysis such as collocations and adjectives
- Acquisition of sentiment expressions and their polarities from supervised corpora, in which favorability in each document is explicitly assigned manually, such as five stars in reviews, and unsupervised corpora, in which no clue on sentiment polarity is available except for the textual content, including the WWW

In all of this work, the level of natural language processing (NLP) was shallow. Except for stemming and analysis of part of speech (POS), they simply analyze co-occurrences of expressions within a short distance or patterns that are typically used for information extraction to analyze the relationships among expressions. Analysis of relationships based on distance obviously has limitations.

One major reason for the lack of focus on relationships between sentiment expressions and subjects may be due to their applications. Many of their applications aim to classify the whole document into positive or negative toward a subject of the document that is specified either explicitly or implicitly, and the subject of all of the sentiment expressions are assumed to be the same as the document subject. For example, the classification of a movie review into positive or negative assumes that all sentiment expressions in the review represent sentiments directly toward that movie, and expressions that violate this assumption (such as a negative comment about an actor even though the movie as a whole is considered to be excellent) confuse the judgment of the classification. On the contrary, by analyzing the relationships between

sentiment expressions and subjects, we can make in-depth analyses on what is favored and what is not. In this paper, we define the task of our sentiment analysis as to find sentiment expressions for a given subject and determine the polarity of the sentiments. In other words, it is to identify text fragments that denote a sentiment about a subject within documents rather than classifying each document as positive or negative towards the subject. In this task, the identification of semantic relationships between subjects and sentiment-related expressions is a key issue because the polarity of the sentiment may be entirely different depending on the relationships, as in the above example of "XXX beats YYY."

CHAPTER 3

PROBLEM STATEMENT

The project "QRT Chatbot - Chatbot facility for people in quarantine" aims to create a simple and efficient way for people to find out if they have a chance of being COVID-19 positive and also to provide a social outlet for people who are isolated due to quarantine. In contrast to existing chatbots, our motivation is to not only provide high detection rate of COVID-19 symptoms, but also to provide proper emotional support for people who are in need of it. To accomplish these tasks we make use of machine learning algorithms that can create a system that has the capability to perform symptom analysis as well as sentimental analysis. One of the main advantages of such a system built by machine learning is that the chatbot will work in an open domain, and can thus understand and respond to complex user inputs and queries, which allows us to have better user-chatbot interactions. Thus, we can reduce close contact between people in public by instead encouraging the use of online resources such as QRT Chatbot.

CHAPTER 4
PROJECT MANAGEMENT

4.1 Introduction

Project management is the discipline of planning, organizing, securing, managing, leading, and controlling resources to achieve specific goals. A project is a temporary endeavor with a defined beginning and end (usually time-constrained, and often constrained by funding or deliverables), undertaken to meet unique goals and objectives, typically to bring about beneficial change or added value. The temporary nature of projects stands in contrast with business as usual (or operations), which are repetitive, permanent, or semi-permanent functional activities to produce products or services. In practice, the management of these two systems is often quite different, and as such requires the development of distinct technical skills and management strategies.

In our project we are following the typical development phases of an engineering project

1. Initiation
2. Planning and Design
3. Execution and Construction
4. Monitoring and Controlling Systems
5. Completion

4.1.1 Initiation

The initiating processes determine the nature and scope of the project. The initiating stage should include a plan that encompasses the following areas :

1. Analysing the business needs/requirements in measurable goals
2. Reviewing of the current operations
3. Financial analysis of the costs and benefits including a budget
4. Stakeholder analysis, including users, and support personal for the project

5. Project charter including costs, tasks, deliverables, and schedule

4.1.2 Planing and design

After the initiation stage, the project is planned to an appropriate level of detail (see example of a flow-chart). The main purpose is to plan time, cost and resources adequately to estimate the work needed and to effectively manage risk during project execution. As with the initiation process, a failure to adequately plan greatly reduces the project's chances of successfully accomplishing its goals.

- Determining how to plan
- Developing the scope statement
- Selecting the planning team
- Identifying deliverables and creating the work breakdown structure
- Identifying the activities needed to complete those deliverables
- Developing the schedule
- Risk planning

4.1.3 Execution

Executing consists of the processes used to complete the work defined in the project plan to accomplish the project's requirements. The execution process involves coordinating people and resources, as well as integrating and performing the activities of the project in accordance with the project management plan. The deliverables are produced as outputs from the processes performed as defined in the project management plan and other frameworks that might be applicable to the type of project at hand.

4.1.4 Monitoring & controlling

Monitoring and controlling consists of those processes performed to observe project execution so that potential problems can be identified in a timely manner and corrective action can be taken, when necessary, to control the execution of the project. The key benefit is that project performance is observed and measured regularly to identify variances from the project management plan.

4.2 System Development Life Cycle

The Systems development life cycle (SDLC), or Software development process in systems engineering, information systems, and software engineering, is a process of creating or altering information systems, and the models and methodologies that people use to develop these systems. In software engineering, the SDLC concept underpins many kinds of software development methodologies. These methodologies form the framework for planning and controlling the creation of an information system.

The SDLC phases serve as a programmatic guide to project activity and provide a flexible but consistent way to conduct projects to a depth matching the scope of the project. Each of the SDLC phase objectives is described in this section with key deliverables, a description of recommended tasks, and a summary of related control objectives for effective management. The project manager must establish and monitor control objectives during each SDLC phase while executing projects. Control objectives help to provide a clear statement of the desired result or purpose and should be used throughout the entire SDLC process.

4.2.1 Spiral Model

We have used the Spiral model in our project. The Spiral model incorporates the best characteristics of both- waterfall and prototyping model. In addition, the Spiral model also contains a new component called Risk Analysis, which is not there in the waterfall and prototype model. In the Spiral model, the basic structure of the software product is developed first. After the basic structure is developed, new features such as user interface and data administration are added to the existing software product. This functionality of the Spiral model is similar to a spiral where the circles of the spiral increase in diameter. Each circle represents a more complete version of the software product. The spiral is a risk-reduction oriented model that breaks a software project up into main projects, each addressing one or major risks. After major risks have been addressed the spiral model terminates as a waterfall model. Spiral iteration involves six steps:

1. Determine objectives, alternatives and constraints.
2. Identify and resolve risks.
3. Evaluate alternatives.
4. Develop the deliverables for the iteration and verify that they are correct.
5. Plan the next iteration.

6. Commit to an approach for the next iteration.

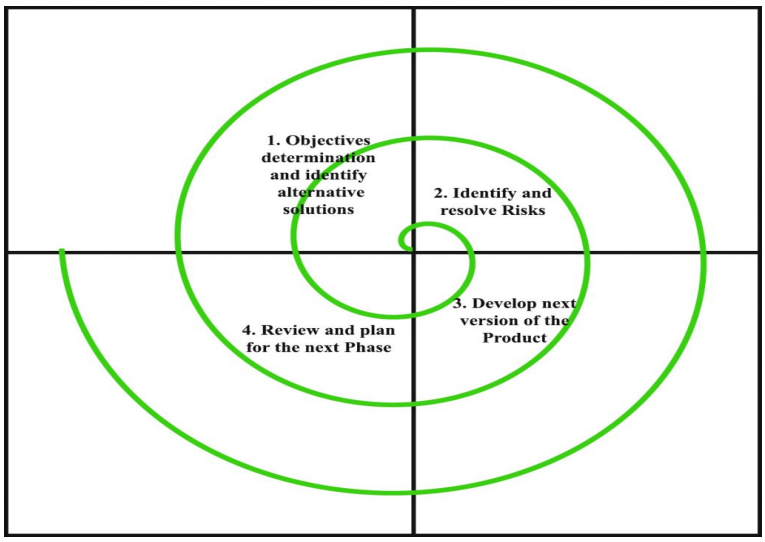

Figure 4.1: Spiral Model

CHAPTER 5
METHODOLOGY

5.1 System Requirements & Specifications

5.1.1 Windows 10

Windows 10 is a series of personal computer operating systems produced by Microsoft as part of its Windows NT family of operating systems. It is the successor to Windows 8.1 and was released to manufacturing on July 15, 2015, and to retail on July 29, 2015. Windows 10 receives new builds on an ongoing basis, which are available at no additional cost to users. Mainstream builds of Windows 10 are labeled version YYMM with YY representing the year and MM representing the month of release. For example, the latest mainstream build of Windows 10 is Version 1809. There are additional test builds of Windows 10 available to Windows Insiders. Devices in enterprise environments can receive these updates at a slower pace, or use long-term support milestones that only receive critical updates, such as security patches, over their ten-year lifespan of extended support.

5.1.2 Python 3.6.2

Python is a dynamic object-oriented programming language that can be used for many kinds of software development. It offers strong support for integration with other languages and tools, comes with extensive standard libraries, and can be learned in a few days. Many Python programmers report substantial productivity gains and feel the language encourages the development of higher quality, more maintainable code.

Python runs on Windows, Linux/Unix, Mac OS X, OS/2, Amiga, Palm Handhelds, and Nokia mobile phones. Python has also been ported to the Java and .NET virtual machines. Python is distributed under an OSI-approved open source license that makes it free to use, even for commercial products.

5.1.3 SCIKIT-learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. The library is built upon the SciPy (Scientific Python) that must be installed before you can use scikit-learn. This stack that includes:

1. NumPy: Base n-dimensional array package
2. SciPy: Fundamental library for scientific computing
3. Matplotlib: Comprehensive 2D/3D plotting
4. IPython: Enhanced interactive console
5. Sympy: Symbolic mathematics
6. Pandas: Data structures and analysis

Extensions or modules for SciPy are conventionally named SciKits. As such, the module provides learning algorithms and is named scikit-learn.

5.1.4 Pandas

In computer programming, 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. The name is derived from the term "panel data", in econometrics term for data sets that include observations over multiple periods for the same individuals. Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, statistics, analytics, etc

5.1.5 Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.Best of all, NLTK is a free, open source, community-driven project.

5.1.6 MySQL

MySQL is a freely available open source Relational Database Management System (RDBMS) that uses Structured Query Language (SQL). SQL is the most popular language for adding, accessing and managing content in a database. It is most noted for its quick processing, proven reliability, ease and flexibility of use.It is one of the open source databases used for storing Python web applications' data.

5.1.7 Dialogflow

Dialogflow is a natural language understanding platform that makes it easy to design and integrate a conversational user interface into your mobile app, web application, device, bot, interactive voice response system, and so on. Using Dialogflow, you can provide new and engaging ways for users to interact with your product.

Dialogflow can analyze multiple types of input from your customers, including text or audio inputs (like from a phone or voice recording). It can also respond to your customers in a couple of ways, either through text or with synthetic speech.

5.1.8 Android Studio

Android Studio is a new and fully integrated development environment, which has been recently launched by Google for the Android operating system. It has been designed to provide new tools for app development and to provide an alternative to Eclipse, currently the most widely used IDE.

Android Studio provides a unified environment where you can build apps for Android phones, tablets, Android Wear, Android TV, and Android Auto. Structured code modules allow you to divide your project into units of functionality that you can independently build, test, and debug

5.2 Proposed System

Modules

5.2.1 Data Acquisition Module

The Data Acquisition Module's mission is to collect the user's non-confidential data and create a dataset that contains these user's information:

- Location
- Symptoms (Fever, coughs, dyspnea, etc)
- Age
- Gender
- Status (Not infected/Suspected to be infected)
- Contact with Infected Person
- Chronic diseases (Alzheimer disease and dementia, Arthritis, Asthma, Heart disease, Cancer, Diabetes, etc)

When the user sends an input message, COVID-Chatbot must transform unstructured text to a structured representation composed of entities and intents which called the natural language processing (NLP), through several successive steps such as Tokenization , Part of Speech tagging (PoS tagging) , Lemmatization and Stemming , etc. Then we used the pre-trained word embedding model GloVe to transform text to vector. Next, we extracted entities by using Conditional Random Field (CRF) . And finally, intent classification is done using Support Vector Machines (SVM) classifier because it requires less training to guarantee confident intent classification.

5.2.2 Data Preprocessing Module

Deep learning is quickly becoming a powerful tool for solving complex modeling problems across a broad range of industries. An efficient model is developed through intensive training by providing a large number of datasets. These datasets contain a significant proportion of unwanted data (also called noise) in it. These data, if not removed leads to tremendous misclassification of the input data. This will ultimately degrade the performance and efficiency of the classifier. Thus, it is very important to remove unwanted noise from the dataset to

improve efficiency. Therefore, the step of data preprocessing plays a very important role in contributing to the accuracy of any training model. In this project, we mainly perform three steps in the preprocessing stage. They are denoising, sampling, and segmentation.

1. **Denoising:**

   This refers to the process of removing unwanted data from the input and extracting only the required signal/data from the mixture. This forms one of the most important steps in preprocessing.

2. **Sampling:**

   A sample is defined as a smaller set of data that is chosen and/or selected from a larger population by using a predefined selection method. These sets are known as sample points, sampling units or observations. Sampling is a process used in statistical analysis in which a predetermined number of observations (samples) are taken from a larger population.

5.2.3 Identification and Classification Module

This module identifies and classifies the data given. Two types of classifications are used, one is the binary classification for predicting whether the user is contracted with COVID, and the other one is the multiclass classification purpose, used for identifying the emotional state of the user. In binary classification, it identifies the symptoms and predicts the severity of disease. In multiclass classification, it classifies the sentiment polarity as happy, sad, or neutral.

ALGORITHM IDENTIFIED

Algorithm identified for classification is Support Vector Machine (SVM).

ALGORITHM USED: SVM

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

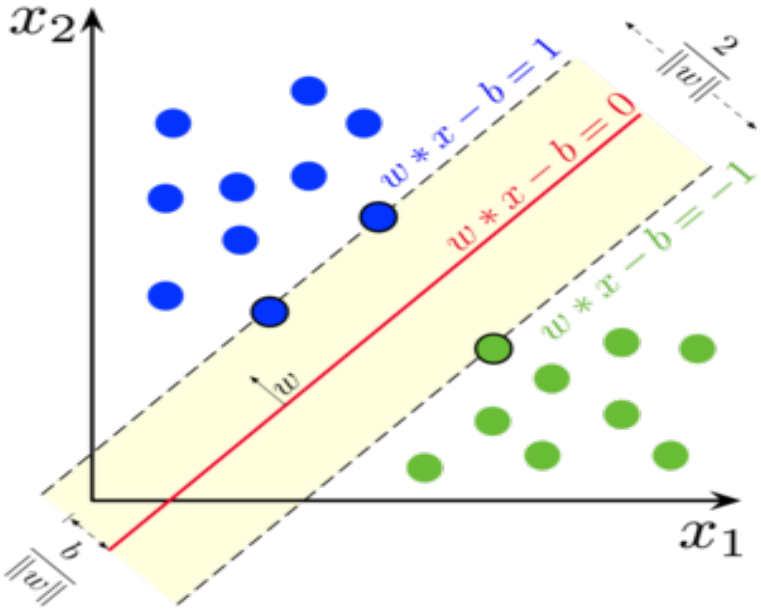

Figure 5.1: Support Vector Machine

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.

SVM can be of two types:

1. Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

2. Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

The advantages of support vector machines are:

- Effective in high dimensional spaces
- Still effective in cases where number of dimensions is greater than the number of samples
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

5.3 Data Flow Diagrams

5.3.1 Data Flow Diagram- Level 0

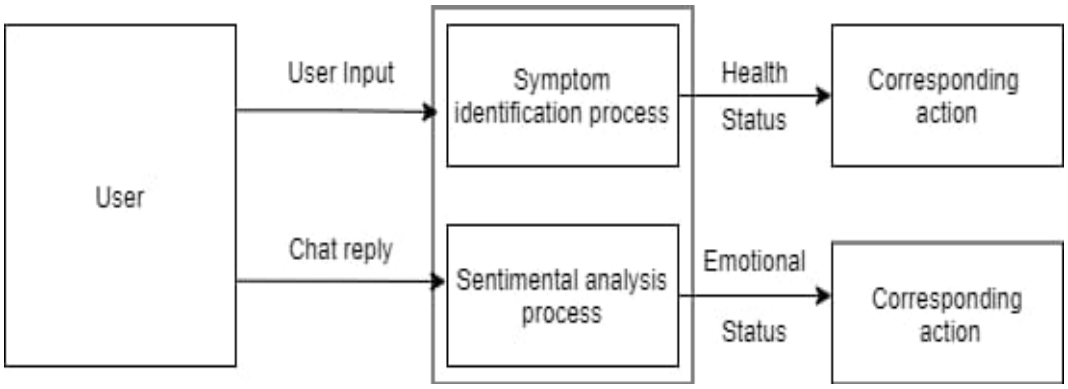

Figure 5.2: DFD- Level 0

5.3.2 Data Flow Diagram- Level 1

Module 1 - Symptom Identification Process

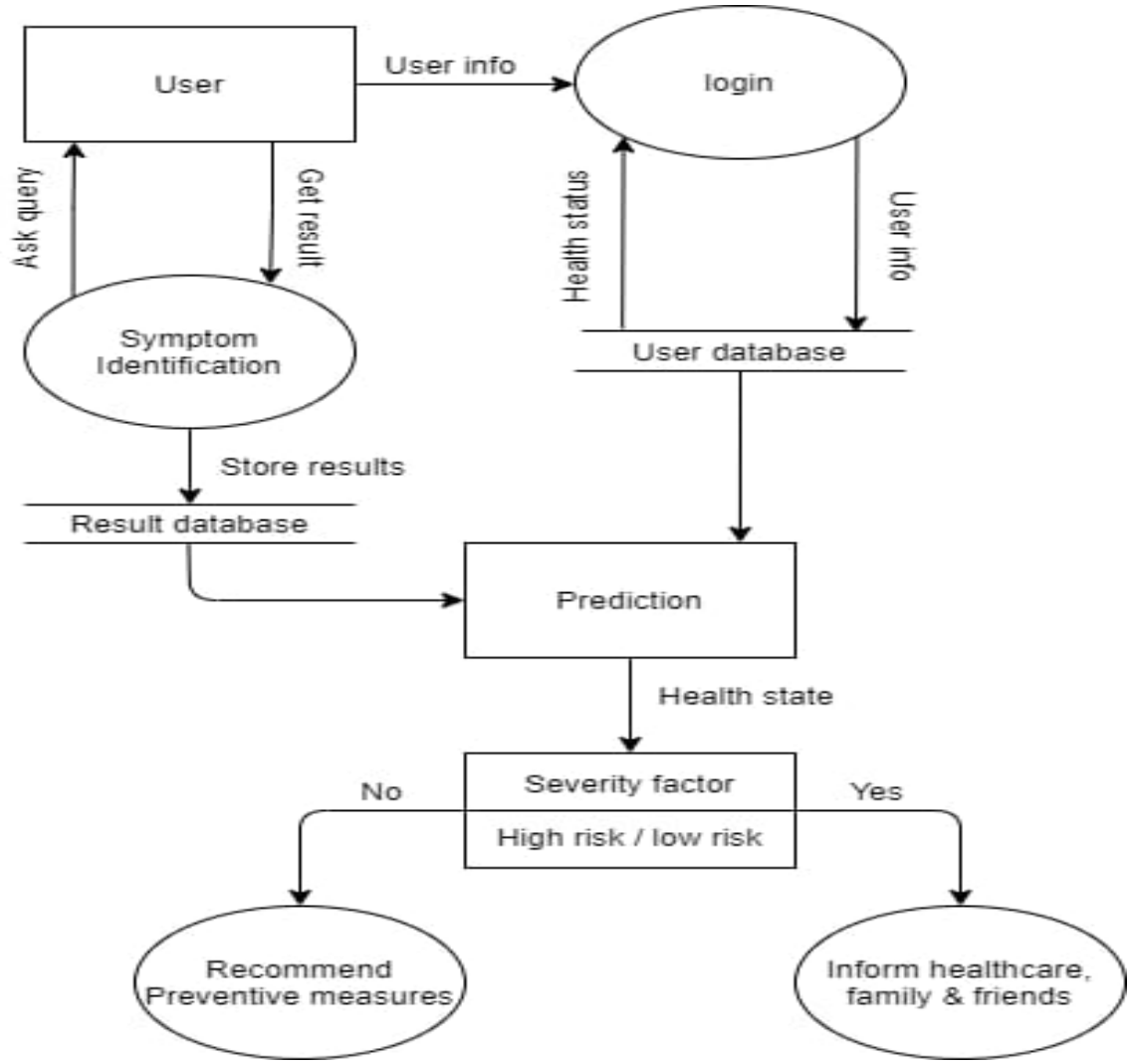

Figure 5.3: DFD- Level 1 mod 1

Module 2 - Sentimental Analysis

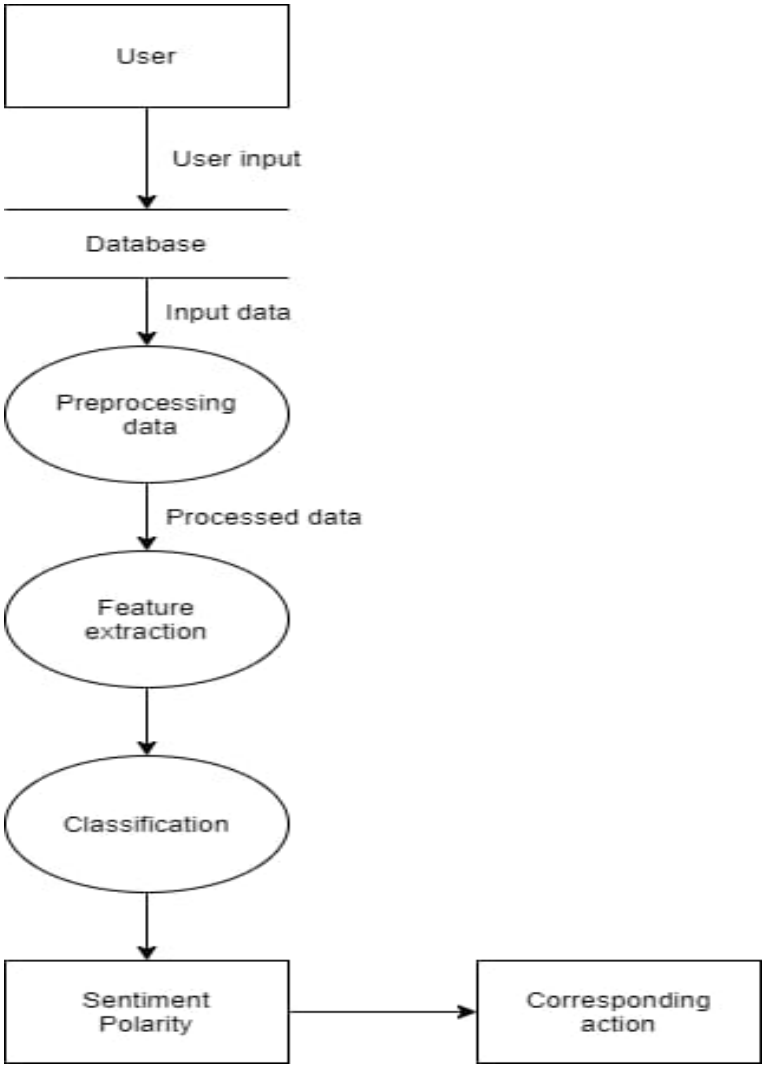

Figure 5.4: DFD- Level 1 mod 2

5.3.3 Data Flow Diagram- Level 2

Training

Figure 5.5: DFD- Level 2

Testing

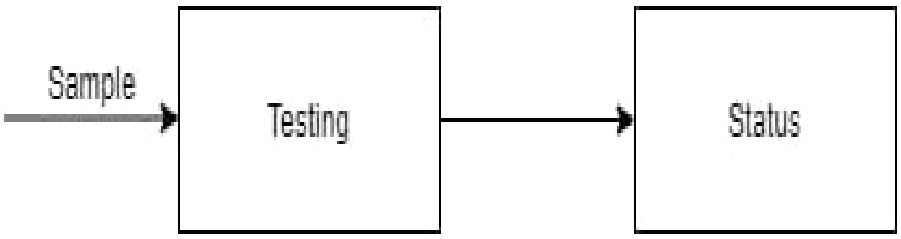

Figure 5.6: DFD- Level 2

Prediction

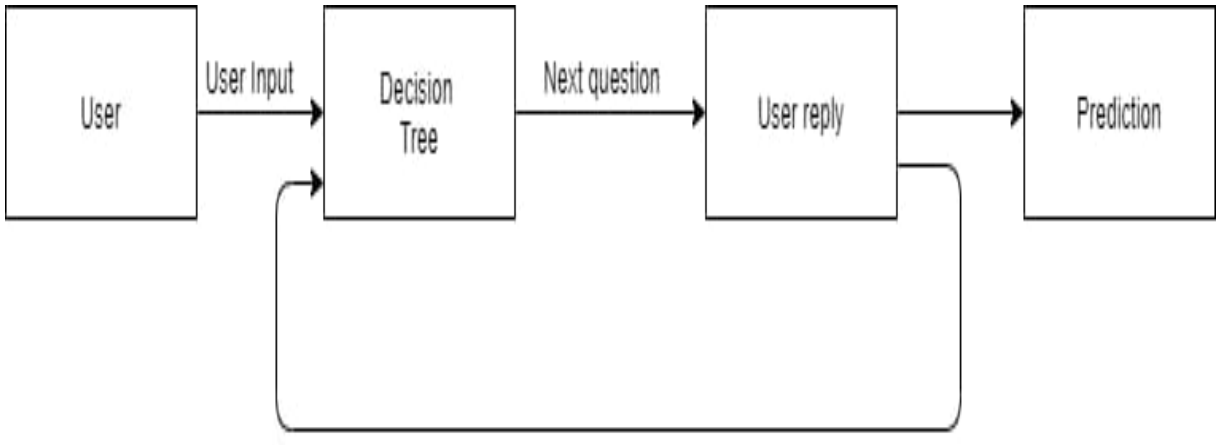

Figure 5.7: DFD- Level 2

5.4 Flow Chart

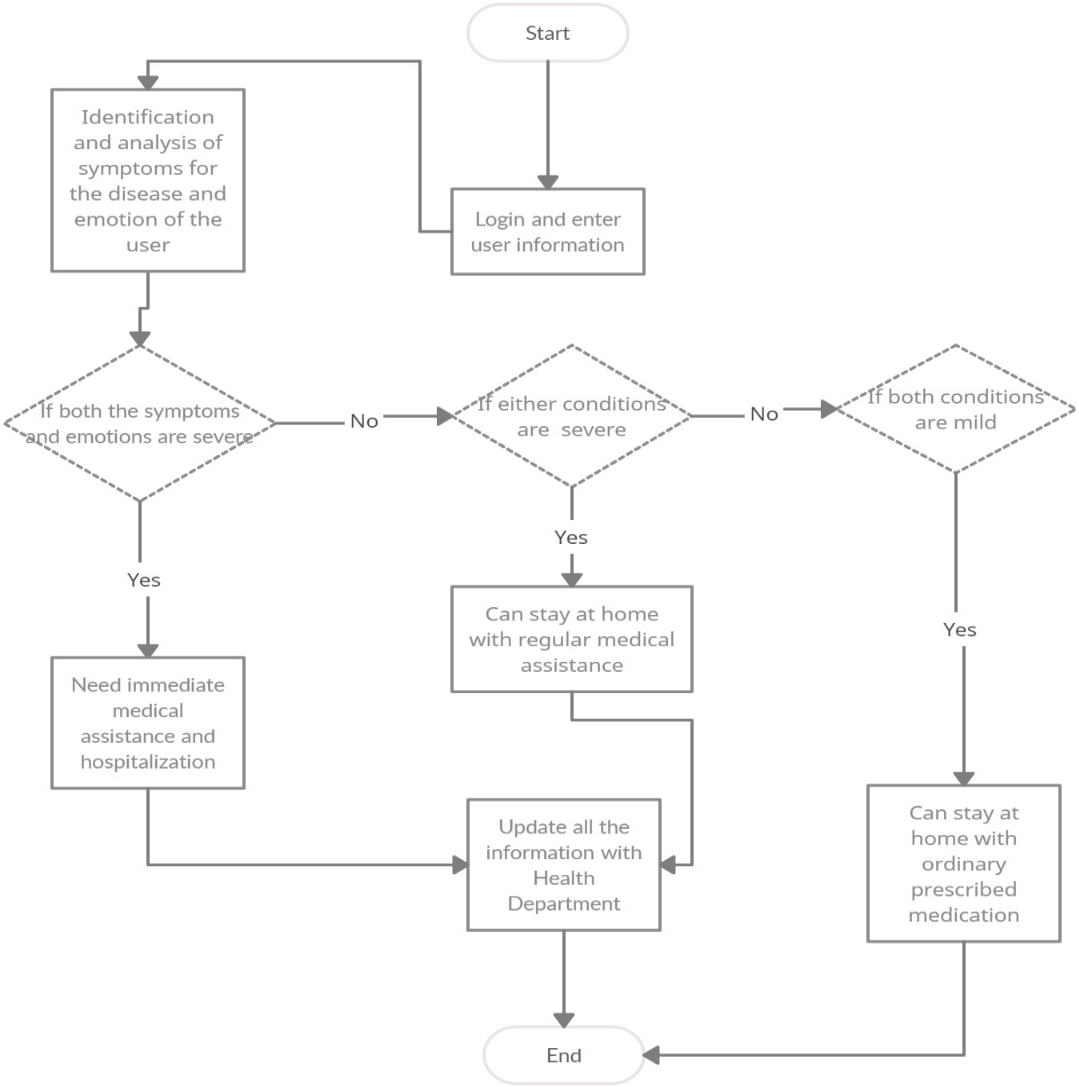

Figure 5.8: Flow Chart

5.5 Architecture

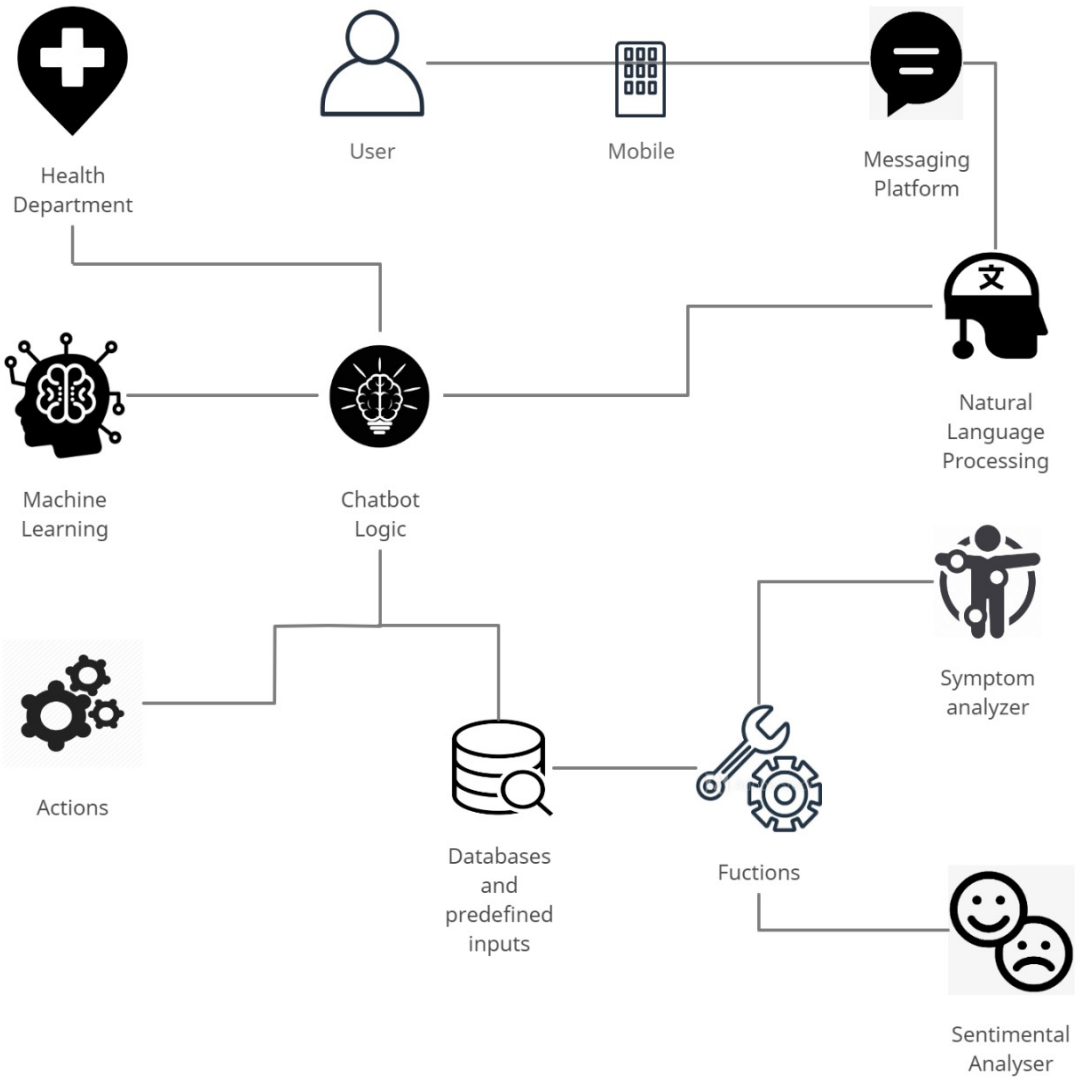

Figure 5.9: Architecture

CHAPTER 6

RESULTS

The proposed method uses machine learning algorithms to perform symptom identification process and sentimental analysis process. Here, the effectiveness of the model in symptom prediction and sentimental analysis using SVM are examined.The library used for implementing the model is Scikit-learn and NLTK.

The main goal of this project is to predict whether the user have been contracted COVID-19 and also to analyse the emotional state of the user.

CHAPTER 7
CONCLUSION AND FUTURE WORKS

Medical chatbot has a high impact on the health culture of the state. It has improved reliability and is less prone to human errors. The facts that the chatbot is free and can be accessed wherever the user is, be it their working environment, prompt the user to have it and use it. We develop a smart chatbot for COVID-19 Assistance using Machine Learning and Natural Language Processing during quarantine.

We have outlined the design of the proposed project, which aims to predict whether the user have been contracted COVID-19. It includes two major modules. The first one is Symptom Prediction which uses Support Vector Machine algorithm and analyses user replies to determine if high case of positive COVID-19. The second one is Sentimental Analysis which analyses the user replies to determine emotional state of the user. In the future, we can improve our sentiment analysis model and make it applicable to the human voice. A video call with a specialized doctor can also be made depending on the availability of the user rather than based on the availability of doctors.

CHAPTER 8

Appendices

Project GitHub Link : https://github.com/ShilpaSivadas24/FinalProject.git

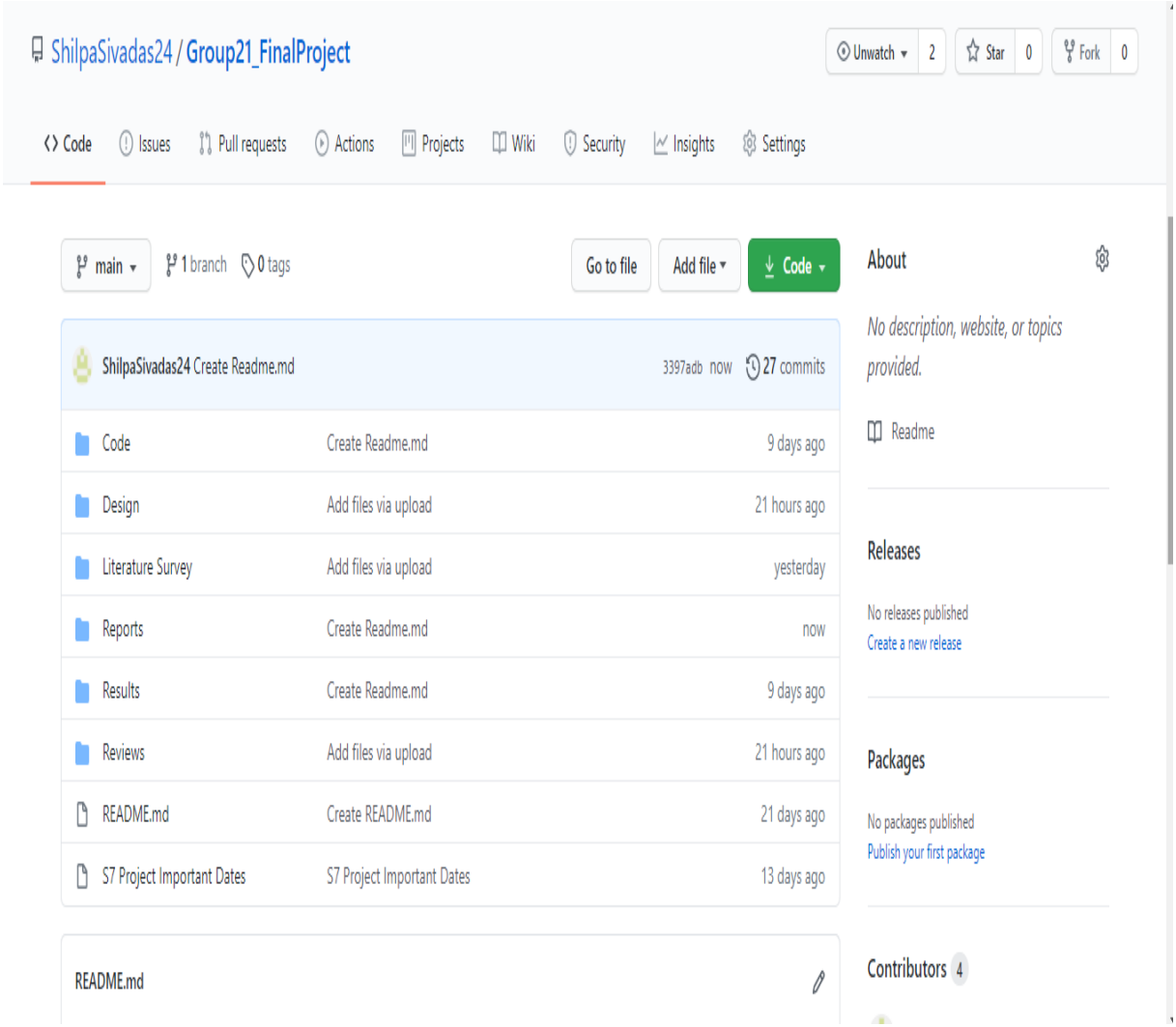

Figure 8.1: Project GitHub Link

REFERENCES

[1] Urmil Bharti, Deepali Bajaj, Hunar Batra, Shreya Lalit, Shweta Lalit, and Aayushi Gangwani. Medbot: Conversational artificial intelligence powered chatbot for delivering telehealth after covid-19. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 870–875. IEEE, 2020.

[2] Urmil Bharti, Deepali Bajaj, Payal Budhiraja, Meghna Juyal, Sushmita Baral, et al. Android based e-voting mobile app using google firebase as baas. In *International Conference on Sustainable Communication Networks and Application*, pages 231–241. Springer, 2019.

[3] Sunath Khadikar, Pawan Sharma, and Priyanka Paygude. Compassion driven conversational chatbot aimed for better mental health.

[4] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100, 2014.

[5] WW Mak, F Cheung, J Woo, D Lee, P Li, KS Chan, and CM Tam. A comparative study of the stigma associated with infectious diseases (sars, aids, tb). *Hong Kong medical journal= Xianggang yi xue za zhi*, 15:34, 2009.

[6] Rohit Binu Mathew, Sandra Varghese, Sera Elsa Joy, and Swanthana Susan Alex. Chatbot for disease prediction and treatment recommendation using machine learning. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 851–856. IEEE, 2019.

[7] Adam S Miner, Liliana Laranjo, and A Baki Kocaballi. Chatbots in the fight against the covid-19 pandemic. *npj Digital Medicine*, 3(1):1–4, 2020.

[8] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77, 2003.

[9] Rajesh Sagar, Rakhi Dandona, Gopalkrishna Gururaj, RS Dhaliwal, Aditya Singh, Alize Ferrari, Tarun Dua, Atreyi Ganguli, Mathew Varghese, Joy K Chakma, et al. The burden of mental disorders across the states of india: the global burden of disease study 1990–2017. *The Lancet Psychiatry*, 7(2):148–161, 2020.

[10] Dr Hannah Behrendt Sarah Merriam. Covid-19: how do we encourage the right behaviours during an epidemic : https://www.bi.team/blogs/covid-19-how-do-we-encourage-the-right-behaviours-during-an-epidemic/.