

MANGALORE INSTITUTE OF TECHNOLOGY AND ENGINEERING

(An ISO 9001:2015 Certified Institution)

BADAGA MIJAR, MOODABIDRI – 574225

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(NBA Accredited)



A

PROJECT SYNOPSIS ON

A PERSON HEALTH ASSISTANT CHATTERBOT

SUBMITTED BY

4MT17CS005	ABIGAL SHARAL MATHIAS
4MT17CS015	ANANYA P SHETTY
4MT17CS028	BINDIYA C M
4MT17CS042	HAIMASHREELAKSHMI

Project Guide

RAVINARAYANA B

ASSOCIATE PROFESSOR, CS&E

TABLE OF CONTENTS

TITLE.....	1
1. ABSTRACT.....	1
2. PROBLEM STATEMENT.....	1
3. OBJECTIVE AND SCOPE.....	1
4. PROJECT DESCRIPTION.....	2
5. METHODOLOGY.....	3
6. HARDWARE AND SOFTWARE SUPPORT.....	6
7. POSSIBLE OUTCOMES.....	6
8. REFERENCES.....	6

A PERSONAL HEALTH ASSISTANT CHATBOT

1. ABSTRACT

eHealth is the way to communicate with patients virtually through technology. With this record a personal Health Assistant Bot is an intelligent virtual assistant which is able to talk with patients in order to understand their symptoms. It suggests doctor, monitor treatments and health parameters. In case of pandemic situation or for old age people, it is difficult to visit a doctor and get themselves tested. Instead of this, one can use this to predict disease, select doctor and directly meet specialist nearby as suggested by our HealthAssistantBot. In this era of technology revolution, machine learning techniques can be used to process users symptoms and predict disease. This information is used to suggest nearest doctor who is specialized in that field who can give best treatment to users condition. Chatbot does natural-language based interactions. Chatbots are computer programs that interact with users using natural language. The chatbots store information in database to identify keywords from sentences and make decision for query and answers the questions. This project has four options: suggest doctor, symptom checker, treatments, medical dictionary.

2. PROBLEM STATEMENT

Design a personal health assistant chatbot that can be used by the patients to know about their disease and suggest proper doctor for treatment.

3. OBJECTIVE AND SCOPE OF THE PROJECT

Objectives:

- To design a chatbot to identify the diseases of patients for given symptoms.

- To introduce a strategy to detect users' health condition and clinical area by relying on the symptoms described in the profile
- To be able to suggest doctors and to monitor the treatments required
- To support patients by reminding them to take medicine on time

Scope:

Patients sitting in their houses, can get the information about their medical condition and get the proper solution for their problems.

4. PROJECT DESCRIPTION

A personal HeathAssistantBot consists of four sections symptom checker, suggest doctor, treatments, medical dictionary. The graphical user interface is provided to make interactions through dialog intuitive for final users. To achieve this, combination of textual elements and graphical components is used. The interactions with chatbot is guided by commands. By using such command, the user can activate specific functionalities of chatbot. "/start" is the initial command sent to chatbot to start conversation between user and chatbot, and to initialize all functionalities. In order to obtain list of doctors relevant for the user, suggest doctor button is used. In this case system asks user to describe his symptoms by writing text message and select one of clinical areas automatically identified by the system. Then user can give his province residence and get nearby doctor who can solve the problem. The second button symptom checker has symptom analysis feature that allows user to use chatbot as self-diagnosis tool that is result given by chatbot for symptoms presented by user will be description of diseases user has. The medical dictionary module provides user with medical

encyclopaedia. The users can input a medical term on which they want to obtain more information. The chatbot responds with short description and Wikipedia link. The treatments functionality regards the storage and management of various medical treatments of the user. The system asks user to provide name of the treatments, the dosage of the medicines, the time at which system should generate reminders, the frequency of treatments, the last day of treatment. List of commands supported by chatbot is given below:

Command	Functionality
/start	It is used for starting interactions with the chatbot
/menu	It is used for coming back to the first dialog screen
/suggestdoctor (Suggest Doctor)	It is used for opening the functionality of doctors recommendation
/symptomchecker (Symptom Checker)	It is used for opening functionality of disease prediction from user symptoms
/medicaldictionary (Medical Dictionary)	It is used for opening functionality of medical term description
/therapy (Treatments)	It is used for opening functionality for scheduling treatments

5. METHODOLOGY

The goal of Experiment is to evaluate the validity of the Symptom Checker module by varying the classification algorithm and the amount of data used for the training.

In this experiment, we want to test the following research hypotheses:

- H1: the performances of the symptom checker are affected by the classification strategy applied;
- H2: a larger dataset makes the classification model for the Symptom Checking task more accurate;

- H3: ensemble strategies can improve the performance of the classification model for the symptom checking task.

H1: In order to test hypothesis H1, we evaluated the use of the following algorithms: Naive Bayes, Logistic Regression, Random Forest, Multilayer Perceptron Network.

H2: we changed the number of patients for each disease k among the values: 100, 1000, 2500, 5000 obtaining datasets of corresponding size l of 21700, 217000, 542500 and 1085000 instances.

H3: In order to test hypothesis H3, we decided to define m models to simulate specialists into one specific clinical area.

- The classifier is trained on a synthetic patient dataset $P = p_1, p_2, \dots, p_n$ where each p_i is described by a set of symptoms $S = s_1, s_2, \dots, s_n$.
- Specifically, a one-hot representation was used to describe each user. Each symptom is represented by a cell of the patient vector p_i and the value 1 is inserted if this symptom is present in the patient, and 0 if it is not. P is used for training a classification model.
- Naive Bayes is a supervised learning algorithm suitable for solving multi-class classification problems. It is based on the use of conditional probability to determine the probabilities of model element. The peculiarity of the algorithm is the idea that all features are independent to one. The presence or absence of one feature does not affect the presence or absence of others.
- Logistic regression can be considered as a classification method within the family of supervised learning algorithms that are robust to noise. For multi-class classification problems, it is possible to use the multinomial logistic regression, that generalizes the classification problem.

Random Forest is an ensemble-type classifier, i.e. it is made up of a set of simpler classifiers. Specifically, it employs several decision trees, each capable of producing an output response when given an input example trees. Unlike the Naive Bayes and Logistic Regression models, Random Forest is able to manage datasets with high dimensionality and categorical features effectively.

- A Deep Multilayer Network (Multilayer Perceptron Network, MLP) consists of a layer of input neurons, each of which corresponds to an explanatory variable, one or more hidden layers, each

of which consists of a number of neurons, and an output layer, consisting of as many neurons as there are response variables.

- K-Means algorithm on the descriptions of the items, and into 24 different groups using the subdivision obtained by the manual association of each disease to the corresponding clinical area.

Results of hypothesis:

- We considered the results obtained by the Naive Bayes algorithm, our best performing model, and we compared them with the results obtained from the other classification models.
- Regarding hypothesis H2, (a larger dataset makes the classification model for the Symptom Checking task more accurate), it is clear that in machine learning, a large enough amount of data is required to allow the classifier to generalize well on the task.
- The F1 score shows us an increase in performance from 0.595 when the model is trained with only 100 patients per disease, to 0.615 when trained with 2500 examples per disease. A similar trend can be observed for each classifier.
- Finally, for hypothesis H3, ensembling strategies influence the performances of the classifier we can observe that best scores have been obtained with an ensemble of only five models.
- If each model of the ensemble does not know the diseases on which the other sectorial models are working, it will be not able to consider them during its reasoning process.
- The Symptom Checker module shows a qualitatively satisfactory behaviour from our point of view, especially for the identification of common and well-defined diseases.

In this experiment we tested the following research hypotheses:

- H1: The Symptom Checker module is able to detect a disease analysing the symptoms provided by the user;
- H2: The user feels satisfied by the services offered by the chatbot;
- H3: The chatbot interface is efficient and simple to use

The main contribution of this project is the definition of the Symptom Checker module, that identifies the patient's disease with a certain degree of accuracy, starting from a set of symptoms. This functionality is designed to assist the user in obtaining a set of automatic diagnoses that can be later discussed with the doctor.

6. HARDWARE AND SOFTWARE SUPPORT

Software: Scikit Learn, PyTorch, TensorFlow, Weka.

Hardware: 128 GB RAM, 2TB Hard disk, graphics card, GPU.

7. POSSIBLE OUTCOMES

- This system shows the effectiveness of the strategy in supporting the patients in taking care of their health
- It presents the description of the disease, supported by the confidence level of the prediction and a link to Wikipedia with more details about the disease
- It identifies the doctors suitable for treating the patients
- It describes the management of the various medical treatments for the user

8. REFERENCES

- [1] M. Polignano et al, Fedelucio Narducci, "A Personal Health Assistant for the Italian Language", Department of Computer Science, University of Bari Aldo Moro, 70125 Bari, Italy, Department of Electrical and Information Engineering, Politecnico di Bari, 70125 Bari, Italy