

A MINI-PROJECT REPORT
ON
“HEALTH CARE CHAT BOT”

Submitted by
Aryan Vetrekar (B-27)
Danish Shaikh (B-20)
Himanshu Singh (B-23)

Supervisor:
Prof. Varsha Wangikar



Department of Computer Engineering
K.C. College of Engineering and Management Studies And
Research, Thane (E)
University of Mumbai
2019-20

CERTIFICATE

Certified that the mini-project work entitled “**HEALTH CARE CHAT BOT**” is a bonafide
work carried out by

Aryan Vetrekar (B-27)

Danish Shaikh (B-20)

Himanshu Singh (B-23)

The report has been approved as it satisfies the academic requirements in respect of mini-project
work prescribed for the course.

.....
Mrs. Varsha Wangikar
Supervisor

.....
Mr. Mandar Ganjapurkar
Head of the Department

.....
Dr. Vilas Nitnaware
Principal

Department of Computer Engineering
K.C. College of Engineering and Management Studies And
Research, Thane (E)
University of Mumbai

2019-20

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of Students & Roll Nos :

Signature

Aryan Vetrekar (B-27)

Danish Shaikh (B-20)

Himanshu Singh (B-23)

Date:

ACKNOWLEDGEMENT

No project is ever complete without the guidance of those expert who have already traded this past before and hence become master of it and as a result, our leader. So, we would like to take this opportunity to take all those individuals how have helped us in visualizing this project.

We would take this opportunity to thank our project co-ordinator **Prof. Varsha Wangikar** and other staff for their guidance in selecting this project and also for providing timely assistant to our query and guidance of this project.

We are very grateful to our Head of the Department **Prof. Mandar Ganjapurkar** for extending his help directly and indirectly through various channels in our project work.

We extend our sincerity appreciation to our Principal **Dr. Vilas Nitnaware** for providing us the opportunity to implement our project.

We are really thankful to all our Professors from K.C. College of Engineering and Management Studies and Research for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual.

Thanking You.

Index

Sr.No.	Topic	Page No.
	Declaration	I
	Acknowledgement	II
	Index	III
	List of Figures	V
	List of Tables	VI
	Abstract	VII
1	Introduction 1.1 Introduction 1.2 Purpose and Scope 1.3 Problem Statement	 1 2
2	Project Analysis 2.1 Review of Literature 2.2 Project Timeline 2.3 Dataset details 2.4 Constraints & Assumptions (If Applicable) 2.5 Methodology Used	 3 4 5 6

3	Project Design 3.1 Block Diagram 3.2 Data Flow Diagram 3.3 Use Case Diagram 3.4 Sequence Diagram	.7 8 9 . .
4	Implementation 4.1 Project Implementation Technology 4.1.1 Hardware Requirement 4.1.2 Software Requirement 4.2 Experimental Setup 4.3 Coding (Max 5-7 pages) 4.4 Testing	10 11 12 13 14 15 16
5	Result 5.1 Snapshot of Result <ul style="list-style-type: none"> • Snapshot • Analysis of result 5.2 Comparative Study with Graph (if applicable)	17 18 19 20 21 22 23
6	Advantage and Disadvantages of Model 6.1 Advantages 6.2 Disadvantages	24

7	Conclusion & Future Scope 7.1 Conclusion 7.2 Future Scope	25
8	References	26

List of Figure

Figure No.	Name of Figure	Page No.
1.1	Block Diagram	7
	Data Flow Diagram	8
	Use Case Diagram	9
	Sequence Diagram	

List of Table

Table No.	Name of Table	Page No.
	Project Timeline	4

ABSTRACT

The main aim of project “**HEALTH CARE CHATBOT**” is to help you better visualize the presentation of mined data (information). It deals with all the health care issues which will really benefit stakeholders in the health care space.

1. Introduction:

1.1 Introduction:

Through chat bots one can communicate with text or voice interface and get reply through artificial intelligence. Typically, a chat bot will communicate with a real person. Chat bots are used in applications such as ecommerce customer service, call centers and Internet gaming. Chat bots are programs built to automatically engage with received messages.

Chat bots can be programmed to respond the same way each time, to respond differently to messages containing certain keywords and even to use machine learning to adapt their responses to fit the situation. A developing number of hospitals, nursing homes, and even private centers, presently utilize online Chat bots for human services on their sites. These bots connect with potential patients visiting the site, helping them discover specialists, booking their appointments, and getting them access to the correct treatment.

An ML model has to be created wherein we could give any text input and on the basis of training data it must analyze the symptoms. A Supervised Logistic Regression machine learning algorithm can be implemented to train the model with data sets containing various diseases CSV files. The goal is to compare outputs of various models and suggest the best model that can be used for symptoms in real-world inputs. Data set contains CSV file having all diseases compiled together. The logistic regression algorithm in ML allows us to process the data efficiently. The goal here is to model the underlying structure or distribution of the data in order to learn more from the training set.

In any case, the utilization of artificial intelligence in an industry where individuals' lives could be in question, still starts misgivings in individuals. It brings up issues about whether the task mentioned above ought to be assigned to human staff. This healthcare chat bot system will help hospitals to provide healthcare support online 24 x 7, it answers deep as well as general questions. It also helps to generate leads and automatically delivers the information of leads to sales. By asking the questions in series it helps patients by guiding what exactly he/she is looking for.

1.2 Purpose and Scope:

Almost everyone kept on hold while operators connect you to a customer care executive. On an average people spend around 7 minutes until they are assigned to a person. Gone are the frustrating days of waiting in a queue for the next available operative. They are replacing live chat and other forms of slower contact methods such as emails and phone calls. Since chat bots are basically virtual robots they never get tired and continue to obey your command. They will continue to operate every day throughout the year without requiring to take a break.

1.3 Problem Statement:

Through chat bots one can communicate with text or voice interface and get reply through artificial intelligence. Typically, a chat bot will communicate with a real person. Chat bots are used in applications such as ecommerce customer service, call centers and Internet gaming. Chat bots are programs built to automatically engage with received messages.

Chat bots can be programmed to respond the same way each time, to respond differently to messages containing certain keywords and even to use machine learning to adapt their responses to fit the situation. A developing number of hospitals, nursing homes, and even private centers, presently utilize online Chat bots for human services on their sites. These bots connect with potential patients visiting the site, helping them discover specialists, booking their appointments, and getting them access to the correct treatment.

An ML model has to be created wherein we could give any text input and on the basis of training data it must analyze the symptoms. A Supervised Logistic Regression machine learning algorithm can be implemented to train the model with data sets containing various diseases CSV files. The goal is to compare outputs of various models and suggest the best model that can be used for symptoms in real-world inputs. Data set contains CSV file having all diseases compiled together. The logistic regression algorithm in ML allows us to process the data efficiently. The goal here is to model the underlying structure or distribution of the data in order to learn more from the training set.

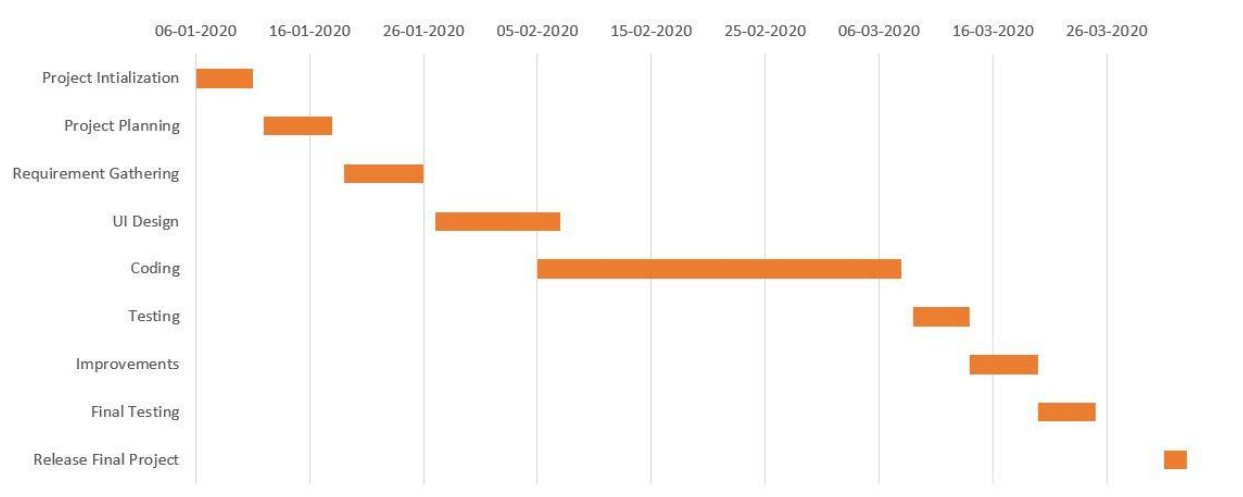
In any case, the utilization of artificial intelligence in an industry where individuals' lives could be in question, still starts misgivings in individuals. It brings up issues about whether the task mentioned above ought to be assigned to human staff. This healthcare chat bot system will help hospitals to provide healthcare support online 24 x 7, it answers deep as well as general questions. It also helps to generate leads and automatically delivers the information of leads to sales. By asking the questions in series it helps patients by guiding what exactly he/she is looking for.

2. Project Analysis:

2.1 Review of Literature:

The main purpose of the scheme is to build the language gap between the user and health providers by giving immediate replies to the Questions asked by the user. Today's people are more likely addicted to the internet but they are not concerned about their personal health. They avoid going to hospital for small problems which may become a major disease in future. Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant documents from the web. Many of the existing systems have some limitations such as there is no instant response given to the patients they have to wait for experts to acknowledge for a long time. Some of the processes may charge an amount to perform live chat or telephony communication with doctors online. The aim of this system is to replicate a person's discussion.

2.2 Project Timeline:



2.3 Dataset Details:

Dataset contains description of different types of diseases. There are different sets of different types of diseases. These sets consists of descriptions of a single disease with different doctors, hospitals, etc.

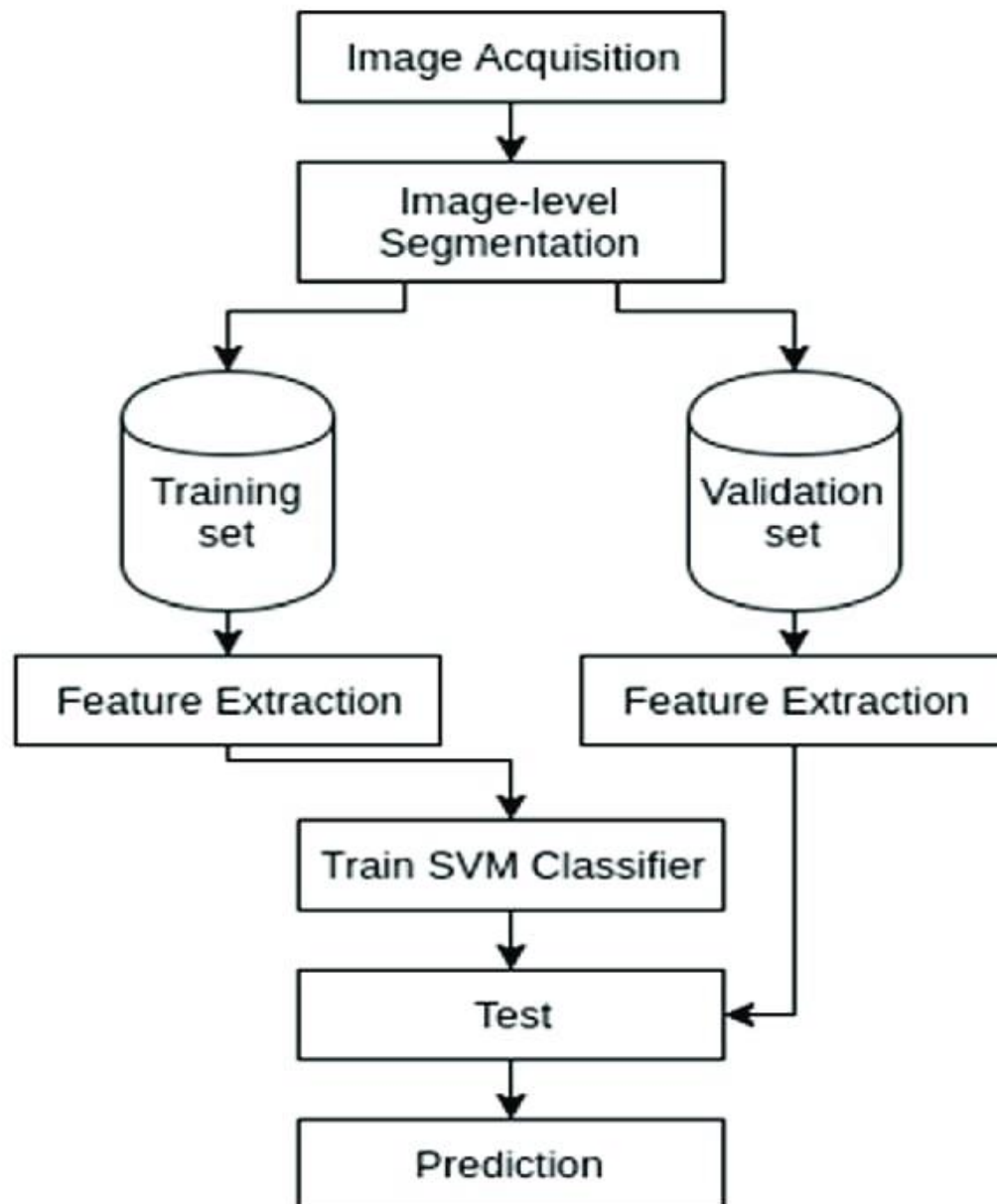
A dataset has been created by recording sequences from over 133 number of diseases and doctors and hospitals.

2.5 Methodology Used:

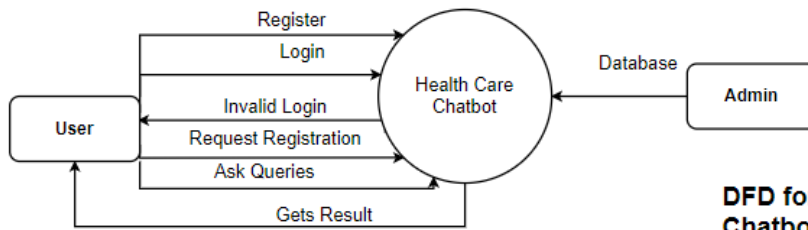
The Health-Care Chat Bot System should be written in Python, GUI links and a simple, accessible network API. The system must provide a capacity for parallel operation and system design should not introduce scalability issues with regard to the number of surface computers, tablets or displays connected at any one time. The end system should also allow for seamless recovery, without data loss, from individual device failure. There must be a strong audit chain with all system actions logged. While interfaces are worth noting that this system is likely to conform to what is available. With that in mind, the most adaptable and portable technologies should be used for the implementation. The system has criticality in so far as it is a live system. If the system is down, then customers must not notice, or notice that the system recovers quickly (seconds). The system must be reliable enough to run, crash and glitch free more or less indefinitely, or facilitate error recovery strong enough such that glitches are never revealed to its end-users.

3. Project Design:

3.1 Block Diagram:

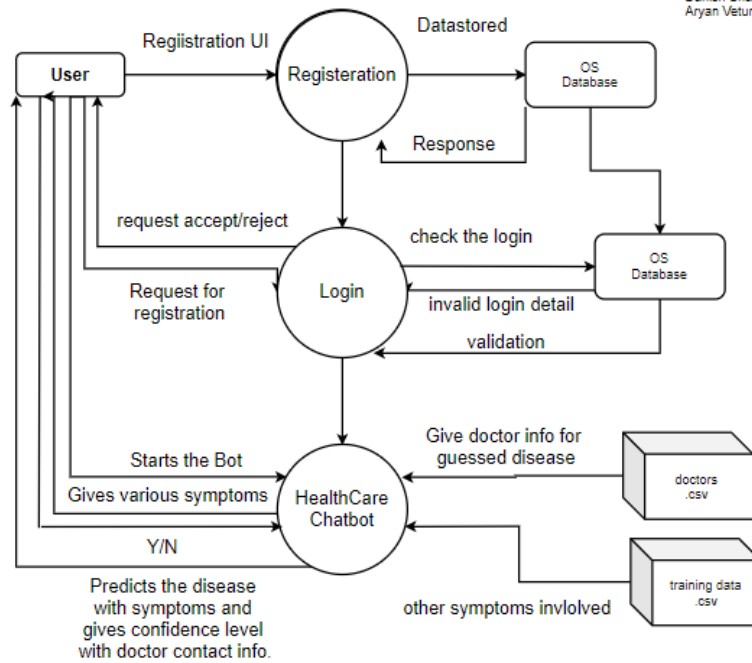


3.2 Data Flow Diagram:

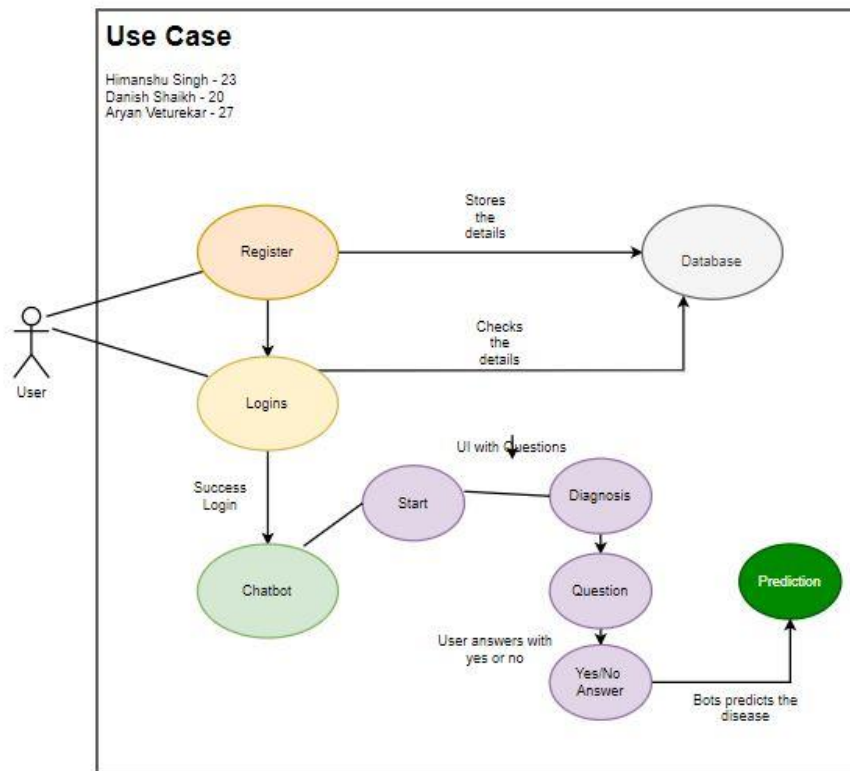


DFD for HealthCare Chatbot

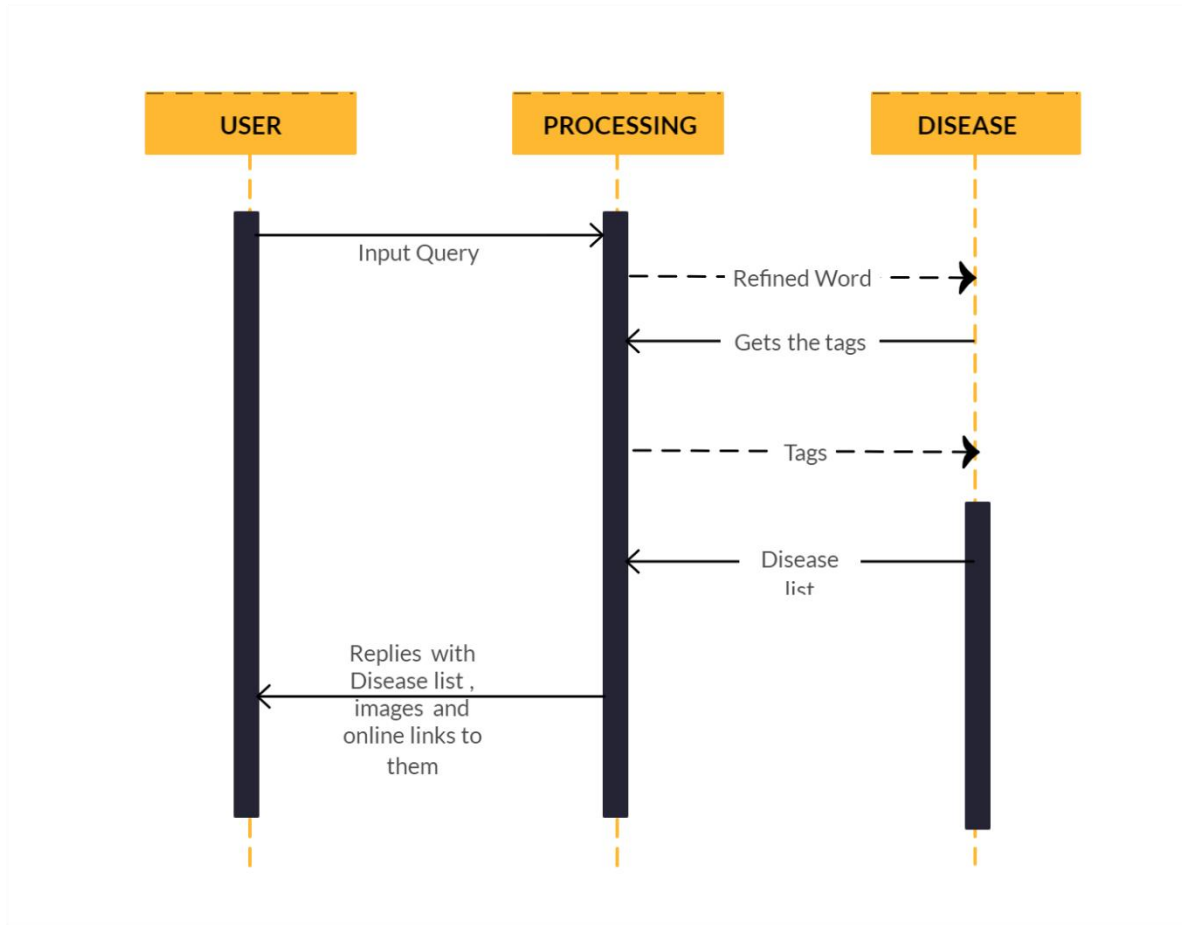
Himanshu Singh - 23
Danish Shaikh - 20
Aryan Vetrekar - 27



3.3 Use Case Diagram:



3.4 Sequence Diagram:



4. Implementation:

4.1 Project Implementation Technology:

In machine learning, **support-vector machines (SVMs, also support-vector networks)** are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

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.

4.1.1 Hardware Requirement:

In recent years, a great variety of hardware solutions for real-time TSR has been proposed. These include conventional (general purpose) computers, custom ASIC (application-specific integrated circuit) chips, field programmable gate arrays (FPGAs), digital sign processors (DSPs) and also graphic processing units

4.1.2 Software Requirements:

In a software-based solution running on a Linux or window system with a 2.4-GHz dual core CPU is presented.

4.2 Experimental Setup:

The main purpose of the scheme is to build the language gap between the user and health providers by giving immediate replies to the Questions asked by the user. Today's people are more likely addicted to the internet but they are not concerned about their personal health. They avoid going to hospital for small problems which may become a major disease in future. Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant documents from the web. Many of the existing systems have some limitations such as there is no instant response given to the patients they have to wait for experts to acknowledge for a long time. Some of the processes may charge an amount to perform live chat or telephony communication with doctors online. The aim of this system is to replicate a person's discussion.

4.3 Coding:

```
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

# Importing the dataset

training_dataset = pd.read_csv('Training.csv')

test_dataset = pd.read_csv('Testing.csv')

X = training_dataset.iloc[:, 0:132].values

#print(X)

y = training_dataset.iloc[:, -1].values

#print(y)

dimensionality_reduction = training_dataset.groupby(training_dataset['prognosis']).max()

#print(dimensionality_reduction)

from sklearn.preprocessing import LabelEncoder

labelencoder = LabelEncoder()

y = labelencoder.fit_transform(y)

#print(y)

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)

from sklearn.tree import DecisionTreeClassifier
```

```

classifier = DecisionTreeClassifier()
classifier.fit(X_train, y_train)

cols = training_dataset.columns
cols = cols[:-1]

importances = classifier.feature_importances_
indices = np.argsort(importances)[::-1]
features = cols

from sklearn.tree import _tree

def execute_bot():

    print("Please reply with yes/Yes or no/No for the following symptoms")
    def print_disease(node):
        #print(node)
        node = node[0]
        #print(len(node))
        val = node.nonzero()
        #print(val)
        disease = labelencoder.inverse_transform(val[0])
        return disease

    def tree_to_code(tree, feature_names):
        tree_ = tree.tree_
        #print(tree_)
        feature_name = [
            feature_names[i] if i != _tree.TREE_UNDEFINED else "undefined!"
            for i in tree_.feature

```

```

]

#print("def tree({}):".format(", ".join(feature_names)))

symptoms_present = []

def recurse(node, depth):

    indent = " " * depth

    if tree_.feature[node] != _tree.TREE_UNDEFINED:

        name = feature_name[node]

        threshold = tree_.threshold[node]

        print(name + " ?")

        ans = input()

        ans = ans.lower()

        if ans == 'yes':

            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])

        print( "You may have " + present_disease )

        print()

        red_cols = dimensionality_reduction.columns

        symptoms_given =

red_cols[dimensionality_reduction.loc[present_disease].values[0].nonzero()]

        print("symptoms present " + str(list(symptoms_present)))

```

```

print()
print("symptoms given " + str(list(symptoms_given)) )
print()
confidence_level = (1.0*len(symptoms_present))/len(symptoms_given)
print("confidence level is " + str(confidence_level))
print()
print("The model suggests:")
print()
row = doctors[doctors['disease'] == present_disease[0]]
print('Consult ', str(row['name'].values))
print()
print('Visit ', str(row['link'].values))
#print(present_disease[0])

```

```

recurse(0, 1)

```

```

tree_to_code(classifier,cols)

```

```

doc_dataset = pd.read_csv('doctors_dataset.csv', names = ['Name', 'Description'])

```

```

diseases = dimensionality_reduction.index

```

```

diseases = pd.DataFrame(diseases)

```

```

doctors = pd.DataFrame()

```

```

doctors['name'] = np.nan

```

```

doctors['link'] = np.nan

```



```
doctors['disease'] = np.nan
doctors['disease'] = diseases['prognosis']
doctors['name'] = doc_dataset['Name']
doctors['link'] = doc_dataset['Description']
record = doctors[doctors['disease'] == 'AIDS']
record['name']
record['link']
execute_bot()
```

4.4 Testing:

Without a well-thought testing effort, the project will undoubtedly fail overall and will impact the entire operational performance of the solution. With a poorly tested solution, the support and maintenance cost will escalate exponentially, and the reliability of the solution will be poor.

Therefore, project managers need to realize that the testing effort is a necessity, not merely as an ad hoc task that is the last hurdle before deployment.

The project manager should pay specific attention to developing a complete testing plan and schedule. At this stage, the project manager should have realized that this effort would have to be accommodated within the project budget, as many of the testing resources will be designing, testing, and validating the solution throughout the entire project life cycle—and this consumes work-hours and resources.

The testing effort begins at the initial project phase (i.e. preparing test plans) and continues throughout until the closure phase.

5. Result:

5.1 Snapshot of Result:

- **Snapshot**
- **Analysis of Result**

Question

slurred_speech ?

Question

Digonosis

You may have :['Hypoglycemia']
symptoms present: ['slurred_speech']
symptoms given: ['vomiting', 'fatigue', 'anxiety', 'sweating', 'headache', 'nausea', 'blurred_and_distorted_vision', 'excessive_hunger', 'drying_and_tingling_lips', 'slurred_speech', 'irritability', 'palpitations']
confidence level is: 0.08333333333333333
The model suggests:
Consult ['Dr. Jyoti Arora Monga']
Visit https://www.practo.com/delhi/doctor/dr-jyoti-arora-ayurveda?specialization=Ayurveda&practice_id=693424

No

Yes

Clear

Start

Question

slurred_speech ?

Question

Digonosis

Please Click on Yes or No for the Above symptoms in Question

No

Yes

Clear

Start

Register

Please enter details below

Username *

dan

Password *

Register

Account Login

Health Care Chatbot

Login

Register

Made by:

Aryan Veturkar

Himanshu Singh

Danish Shaikh

6. Advantage and Disadvantages of Model:

6.1 Advantages:

1. Omni-capable

- The chat bot converses seamlessly across multiple digital channels and retains data and context for a seamless experience. In best cases, even passing that information to a live agent if needed.

2. Free to Explore

- The chat bot can reach, consume, and process vast amounts of data— both structured and unstructured—to surface insights from any source - to gather relevant data to solve customer issues quickly.

3. Autonomous Reasoning

- The chat bot can perform complex reasoning without human intervention. For example, a great Service chatbot should be able to infer solutions based on relevant case histories.

4. Pre-Trained

- The chat bot is pre-trained to understand brand-specific or industry-specific knowledge and terms. Even better, it's pre-configured to resolve common customer requests of a particular industry.

5. Register/Log-in

- To access this chat bot and individual needs to register and then use the registration ID to log in to access the features.

6. User Interface

- A user friendly interface which is engaging and easy to access.

6.2 Disadvantages:

- **Complex Interface** – Chatbots are often seen to be complicated and require a lot of time to understand user's requirement. It is also the poor processing which is not able to filter results in time that can annoy people.
- **Inability to Understand** – Due to fixed programs, chatbots can be stuck if an unsaved query is presented in front of them. This can lead to customer dissatisfaction and result in loss. It is also the multiple messaging that can be taxing for users and deteriorate the overall experience on the website.
- **Time-Consuming** – Chatbots are installed with the motive to speed-up the response and improve customer interaction. However, due to limited data-availability and time required for self-updating, this process appears more time-taking and expensive. Therefore, in place of attending several customers at a time, chatbots appear confused about how to communicate with people.
- **Zero decision-making** – Chat bots are known for being infamous because of their inability to make decisions. A similar situation has landed big companies like Microsoft etc. in trouble when their chat bot went on making a racist rant. Therefore, it is critical to ensure proper programming of your chat bot to prevent any such incident which can hamper your brand.
- **Poor Memory** – Chat bots are not able to memorize the past conversation which forces the user to type the same thing again & again. This can be cumbersome for the customer and annoy them because of the effort required. Thus, it is important to be careful while designing chat bots and make sure that the program is able to comprehend user queries and respond accordingly.

7. Conclusion & Future Scope:

7.1 Conclusion:

Thus, we can conclude that this system giving the accurate result. As we are using large dataset which will ensures the better performance. Thus we build up a system which is useful for people to detect the disease by typing symptoms

7.2 Future Scope:

Chat bots are a thing of the future which is yet to uncover its potential but with its rising popularity and craze among companies, they are bound to stay here for long. Machine learning has changed the way companies were communicating with their customers. With new platforms to build various types of chat bots being introduced, it is of great excitement to witness the growth of a new domain in technology while surpassing the previous threshold.

8. References:

- <https://en.wikipedia.org/wiki/Chatbot> <https://en.wikipedia.org/wiki/Disease>
- <https://data-flair.training/blogs/python-chatbot-project/>
- <https://www.youtube.com/playlist?list=PLQVvva0QuDdc2k5dwtDTyT9aCja0on8j>