

# **Disease Prediction system**

Submitted in partial fulfillment of the requirements of the degree of

**BACHELOR OF COMPUTER ENGINEERING**

by

Bhavesh Pawar - 20102069

Sakshi Thakare - 20102142

Selvina Swarna - 20102194

Guide:

Prof. Bharti Khemani



Department of Computer Engineering

**A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE**

(2022-2023)



## A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

### CERTIFICATE

This is to certify that the Mini Project 2B entitled “**DISEASE PREDICTION SYSTEM**” is a bonafide work of “**Sakshi Thakare (20102142), Swarna Selvina (20102142), Bhavesh Pawar (20102163)**” submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering in Computer Engineering**.

---

Guide:  
Prof. Bharti Khemani

---

Project Coordinator:  
Prof. D.S. Khachane

---

Head of Department  
Prof. S.H. Malave



A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

## Project Report Approval for Mini Project-2B

This project report entitled “**DISEASE PREDICTION SYSTEM**” by ***Bhavesh Pawar, Selvina Swarna, Sakshi Thakare*** is approved for the partial fulfillment of the degree of ***Bachelor of Engineering in Computer Engineering, 2022-23.***

Examiner Name

Signature

1. \_\_\_\_\_

\_\_\_\_\_

2. \_\_\_\_\_

\_\_\_\_\_

Date:

Place:

## **Declaration**

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I 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.

-----  
Bhavesh Pawar – 20102069

-----  
Sakshi Thakare – 20102142

-----  
Selvina Swarna – 20102194

Date:

## **Abstract**

It is a system which provides the user the information and tricks to take care of the health system of the user and it provides how to search out the disease using this prediction. Now a day's health industry plays major role in curing the diseases of the patients so this is often also some quite help for the health industry to inform the user and also it's useful for the user just in case he/she doesn't want to travel to the hospital or the other clinics, so just by entering the symptoms and every one other useful information the user can get to grasp the disease he/she is affected by and also the health industry may also get enjoy this method by just asking the symptoms from the stoner and entering within the system and in only many seconds they'll tell the precise and over to some extent the accurate conditions. This Disease Prediction Using Machine Learning is totally through with the assistance of Machine Learning and Python programming language and also using the dataset that's available previously by the hospitals using that we are going to predict the diseases

**Keywords:**

**Health system, Symptoms, Machine learning , Healthcare , Data analysis ,Disease prediction.**

## **CONTENTS**

<b>Sr. No.</b>	<b>Chapter Name</b>	<b>Page No.</b>
<b>1</b>	<b>Introduction</b>	<b>8</b>
<b>2</b>	<b>Literature Survey</b>	<b>9</b>
<b>3</b>	<b>Problem Statement, Objective &amp; Scope</b>	<b>11</b>
<b>4</b>	<b>Proposed System</b>	<b>13</b>
<b>5</b>	<b>Project Plan</b>	<b>18</b>
<b>6</b>	<b>Experimental Setup</b>	<b>19</b>
<b>7</b>	<b>Implementation Details</b>	<b>20</b>
<b>8</b>	<b>Results</b>	<b>21</b>
<b>9</b>	<b>Conclusion</b>	<b>23</b>
<b>10</b>	<b>References</b>	<b>24</b>

## LIST OF FIGURES

<b>Sr. No.</b>	<b>Figure Name</b>	<b>Page No.</b>
<b>1</b>	<b>Architecture Diagram</b>	<b>14</b>
<b>2</b>	<b>Data Flow Diagram</b>	<b>15</b>
<b>3</b>	<b>Use Case Diagram</b>	<b>16</b>
<b>4</b>	<b>Activity Diagram</b>	<b>17</b>
<b>5</b>	<b>Gantt chart</b>	<b>18</b>

# **Chapter 1**

## **Introduction**

With the rise in number of patient and disease every year medical system is overloaded and with time have become overpriced in many countries. Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap. Prediction of disease by looking at the symptoms is an integral part of treatment. In our project we have tried accurately predict a disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%. Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interface to facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project.



## Chapter 2

### Literature Survey

1. Wiens, J., Shenoy, E. S. (2018). Machine Learning for Healthcare

This article describes the use of recurrent neural network models to predict heart failure onset, with an analysis of the model's performance and its potential clinical implications.

2. Rajkumar, A., Dean, J. (2019). Machine Learning in Medicine. The New England Journal of Medicine, 380(14), 1347-1358.

This review article discusses the use of machine learning in medicine, including disease prediction, and provides an analysis of the strengths and limitations of this approach.

3. Chen, X., Yan, C., Zhang, X., You, J., Huang, K. (2018). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 6, 6495-6505.

This paper gives information about chronic disease prediction system using ML algorithm.

Research Paper	ANALYSIS
1. Wiens, J., Shenoy, E. S. (2018). Machine Learning for Healthcare	This article describes the use of recurrent neural network models to predict heart failure onset, with an analysis of the model's performance and its potential clinical implications.
2. Rajkumar, A., Dean, J. (2019). Machine Learning in Medicine. The	This review article discusses the use of machine learning in medicine, including

Research Paper	ANALYSIS
New England Journal of Medicine, 380(14), 1347-1358.	disease prediction, and provides an analysis of the strengths and limitations of this approach.
3. Choi, E., Schuetz, A., Stewart, W. F., Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361	This conference paper presents an analysis of mortality prediction models in intensive care units using physiological data, highlighting the importance of feature selection and interpretability in developing accurate models
4. Chen, X., Yan, C., Zhang, X., You, J., Huang, K. (2018). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 6, 6495-6505.	This conference paper introduces a deep learning architecture called densely connected convolutional networks, which has been used in disease prediction tasks

## **Chapter 3**

### **Problem Statement, Objective & Scope**

#### **Problem Statement: -**

To implement disease prediction system where user can input symptoms of disease and gets the predicted output. Health information needs are also changing the information seeking behavior and can be observed around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses, and different treatments. If a recommendation system can be made for doctors and medicine while using review mining will save a lot of time. In this type of system, the user face problem in understanding the heterogeneous medical vocabulary as the users are laymen.

User is confused because a large amount of medical information on different mediums are available.

The idea behind recommender system is to adapt to cope with the special requirements of the health domain related with users.

#### **Objective: -**

- The aim of this project is to predict disease based on symptoms. The project is set up in such a way that the device takes the user's symptoms as input and generates an output
- This system gives a user-friendly environment and easy to use and implement.

**Scope: -**

- Machine learning technology offers a strong application forum in the medical industry for health disease prediction concerns based on user/patient experience.
- Using predictive analytics in healthcare can improve the quality of healthcare, collect more clinical data for personalized treatment, and successfully diagnose the medical condition of individual patient.

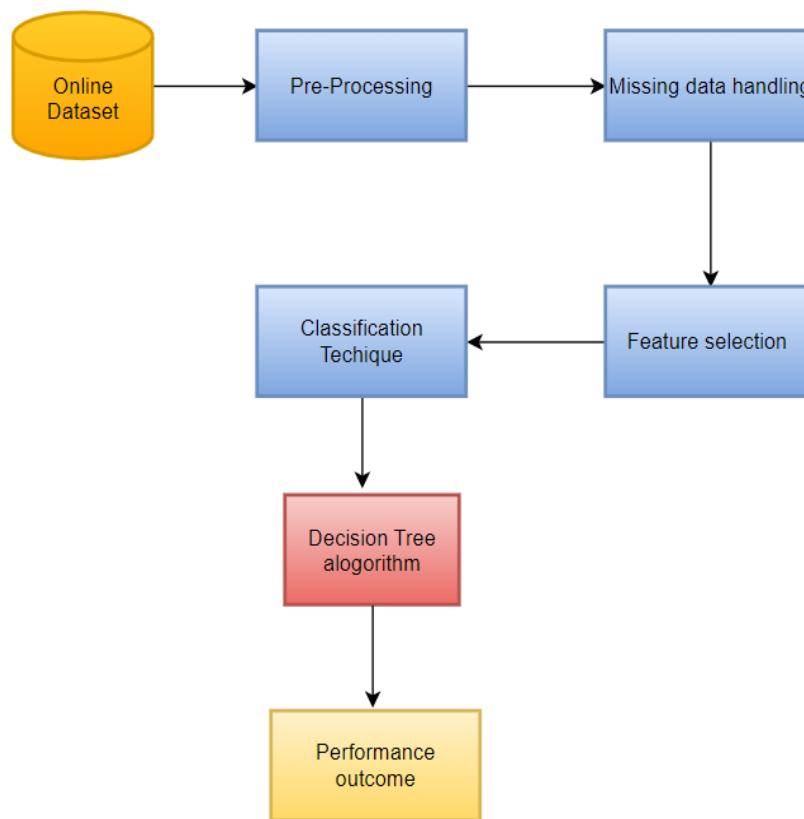
## **Chapter 4**

### **Proposed System Architecture**

- **Description about Proposed System:**

Health information needs are also changing the information seeking behavior and can be observed around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses, and different treatments. If a recommendation system can be made for doctors and medicine while using review mining will save a lot of time. In this type of system, the user face problem in understanding the heterogeneous medical vocabulary as the users are laymen. User is confused because a large amount of medical information on different mediums are available. The idea behind recommender system is to adapt to cope with the special requirements of the health domain related with users.

- **Architecture / Block Diagram**



#### **4.1 Architecture for Disease prediction system**

Architectural diagram is a graphical representation that shows the major components of a system.

It typically includes boxes or rectangles to represent the various components or modules of the system and lines or arrows to represent the interactions or dependencies between them.

- **Data Flow Diagram (Level 0, Level 1 & Level 2)**

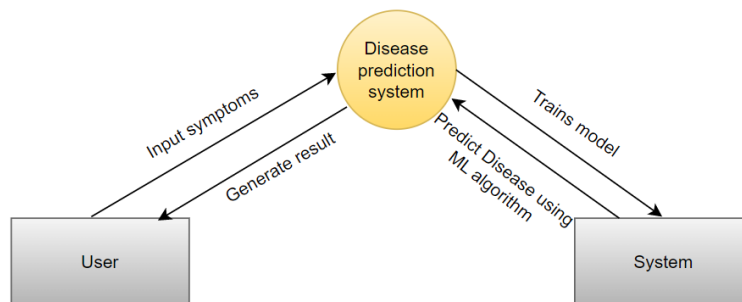
### Level 0



#### 4.2.1 DFD level 0 for Disease prediction system

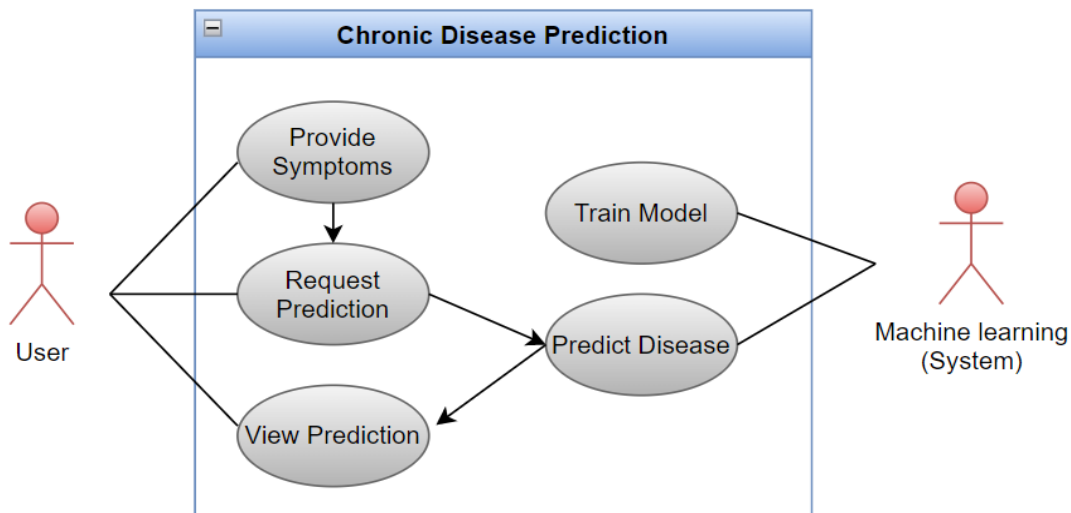
A Data Flow Diagram (DFD) is a graphical representation of the flow of data through system Level 0 DFD Describes the highest level of abstraction in a DFD hierarchy, and it provides a high-level view of the system's major components and the flow of data between them.

### Level 1



#### 4.2.2 DFD level 1 for Disease prediction system

- **Use Case Diagram**



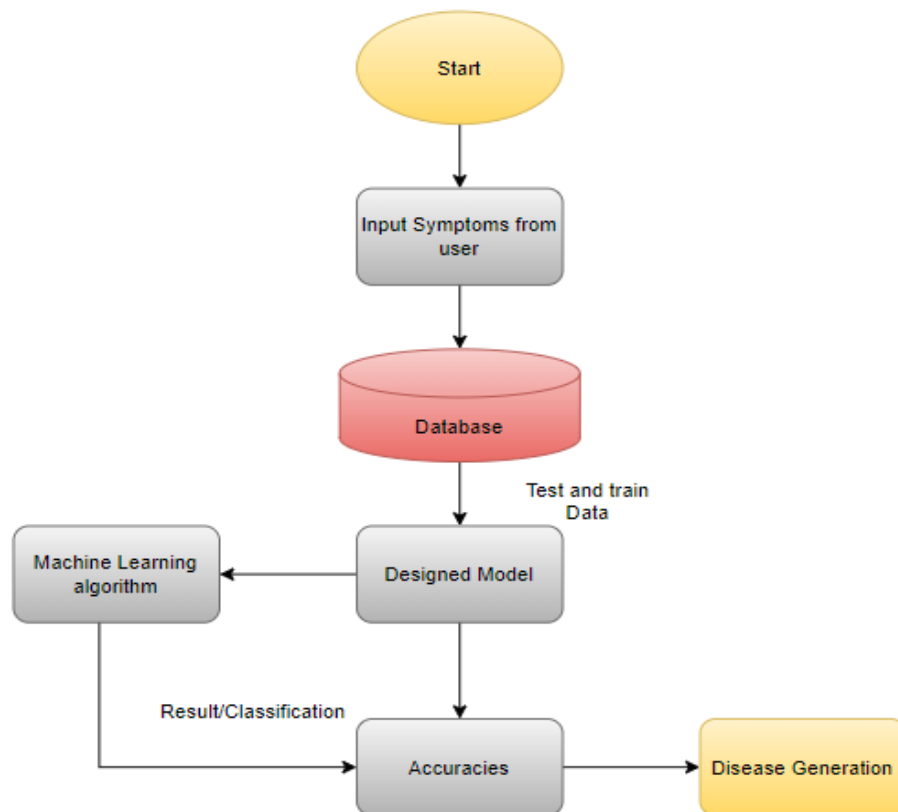
#### **4.3 Use Case Diagram for Disease prediction system**

A use case diagram is a graphical representation of the interactions between actors or and system or application. It is used to model the different ways in which users or system interact with the system



- **Activity Diagram**

---

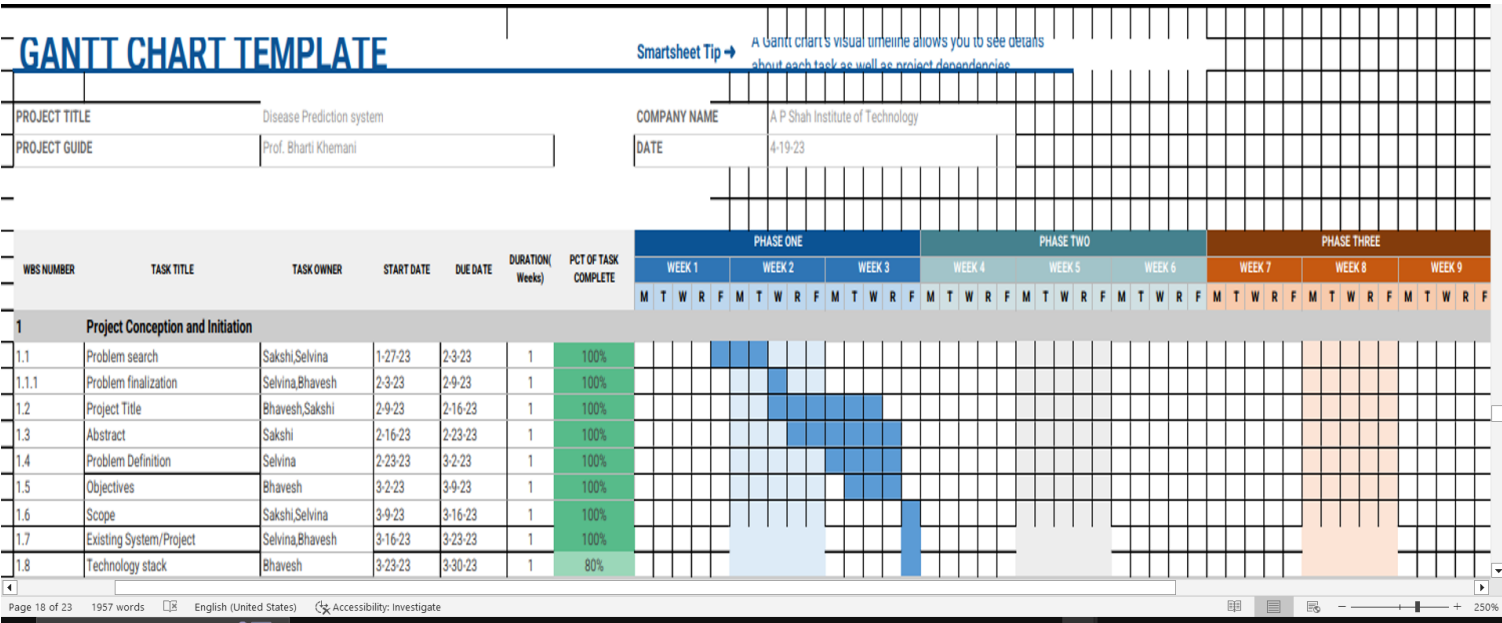


#### **4.4 Activity Diagram for Disease prediction system**

An activity diagram is a graphical representation of the flow of activities, actions and decision points in a system or process.

# Chapter 5

## Project Planning



### 5.1 Gantt chart

# **Chapter 6**

## **Experimental Setup**

- **Software Requirements: -**

- 1) Python3: -
  - i. Tkinter
  - ii. NumPy
  - iii. Pandas
- 2) Html
- 3) CSS

- **Hardware Requirements: -**

- 1) CPU: Basic 64-bit Windows 10 Laptop with i3 core processor.
- 2) RAM: 128 MB
- 3) Disk space: 256 MB
- 4) Processor: Minimum Pentium 2 266 MHz processor

## Chapter 7

### Implementation Details

In this project standard libraries for database analysis and model creation are used. The following are the libraries used in this project.

1. Tkinter: It is a standard GUI library of python. Python when combined with tkinter provides fast and easy way to create GUI Using tkinter we were able to create an interactive GUI for us model.
2. NumPy: NumPy is core library of scientific computing in python. It provides powerful tools to deal with various multi-dimensional arrays in python. It is a general-purpose array processing package.
3. Pandas: It is the most popular python library used for data analysis. It provides highly optimized performance with back-end source code purely written in C or python.
4. sklearn: Sklearn is an open-source python library with implements a huge range of machine-learning, pre-processing, cross-validation, and visualization algorithms.  
In this project we have used sklearn to get advantage of inbuilt classification

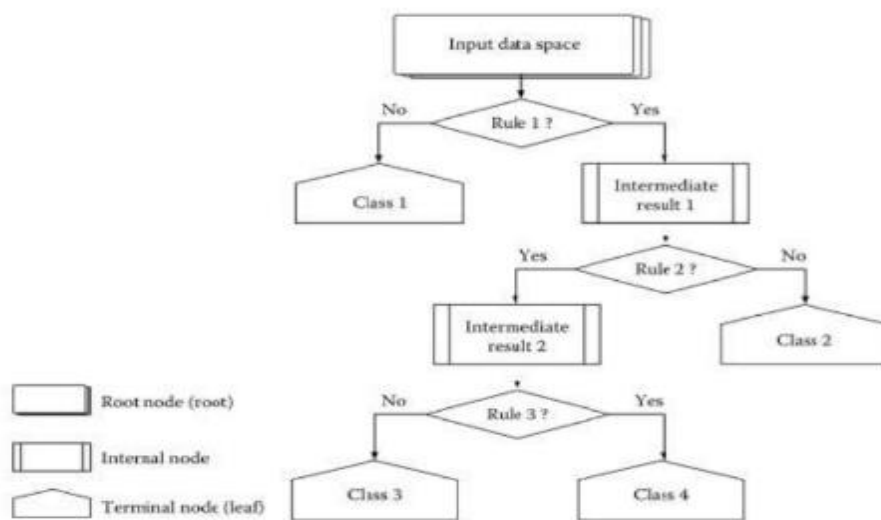
#### Model

Model present in our project to predict the disease

#### Decision tree:

Decision tree is classified as very effective and versatile classification technique. It is used in pattern recognition and classification for image Decision tree is the prediction method we have used in our project. It gives us an accuracy of ~95%. It helps identify the important risk of all factors associated with particular disease and provide a prediction of the likelihood of most of developing the disease. Decision trees can be very effective in disease prediction systems since they can handle complex relationships between multiple risk factors and can provide a clear, interpretable results that can help clinicians make informed decisions about diagnosis and treatment. It is also capable of engaging problems of higher dimensionality.

It mainly consists of three parts root, nodes and leaf. Roots consists of attribute which has most effect on the outcome, leaf tests for value of certain attribute and leaf gives out the output of tree. Decision tree is the only prediction method we have used in our project.



GUI:

GUI made for this project is a simple tkinker GUI consisting of labels, message box, button, text title and option menu

# Chapter 8

## Result

### GUI INTERFACE :

#### FOR USER 1:

Chronic Disease Prediction System

Chronic disease prediction system

Name of the Patient \*

User1

Predict

Symptom 1 \*

throat\_irritation

Symptom 2 \*

mild\_fever

Symptom 3

loss\_of\_smell

Symptom 4

Select Here

Symptom 5

Select Here

Reset Inputs

Exit System

Result

Disease predicted is Common Cold

#### FOR USER 2:

Chronic Disease Prediction System

Chronic disease prediction system

Name of the Patient \*

User2

Predict

Symptom 1 \*

obesity

Symptom 2 \*

watering\_from\_eyes

Symptom 3

yellowing\_of\_eyes

Symptom 4

foul\_smell\_of\_urine

Symptom 5

constipation

Reset Inputs

Exit System

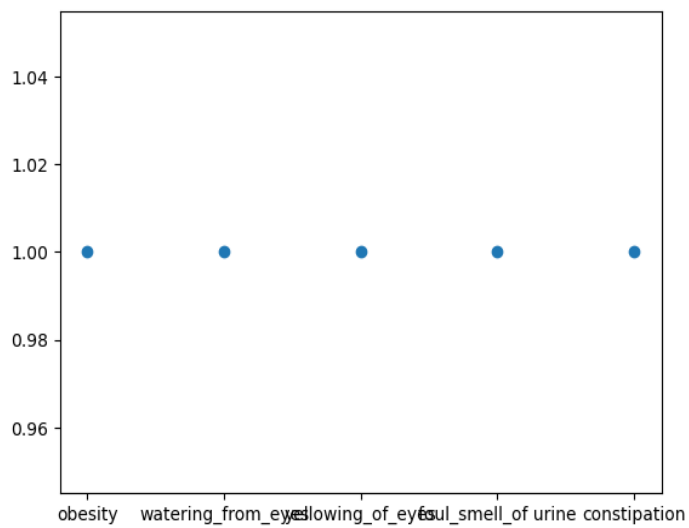
Result

Disease predicted is Chronic cholestasis

## ACCURACY GRAPH:

```
Decision Tree
Accuracy
0.9512195121951219
39
Confusion matrix
[[1 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
['obesity', 'watering_from_eyes', 'yellowing_of_eyes', 'foul_smell_of urine', 'constipation']
[1, 1, 1, 1, 1]
```

```
C:\Users\thaka\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but DecisionTreeClassifier was fitted with feature names
  warnings.warn(
```



## **Chapter 9**

### **Conclusion**

We set out to create a system which can predict disease on the basis of symptoms given to it. Such a system can decrease the rush at OPDs of hospitals and reduce the workload on medical staff. We were successful in creating such a system and use 4 different algorithms to do so. On an average we achieved accuracy of ~94%. Such a system can be largely reliable to do the job. Creating this system, we also added a way to store the data entered by the user in the database which can be used in future to help in creating better version of such system. Our system also has an easy-to-use interface. It also has various visual representation of data collected and results achieved



## **Chapter 10**

### **References**

- [1] Almutairi, B. M., Alhumaidi, R. M., Alkharaz, A. M., Abdullatif, A. Y., & Al-Ammar, M. H. (2020). Machine learning-based prediction models for diabetes mellitus: A systematic review. *Diabetes, metabolic syndrome and obesity: targets and therapy*, 13, 4131–4142.
- [2] Scherer, J., & Uhlmann, L. (2018). Machine learning for medical diagnosis: History, state of the art and perspective. *Computing*, 100(3), 207–254.
- [3] Rahman, M. M., Islam, M. R., Tarek, M. A., Hannan, M. A., & Begum, H. A. (2021). Chronic kidney disease prediction and classification using machine learning techniques: A systematic review. *Computer methods and programs in biomedicine*, 198, 105823.
- [4] Comparative Analysis of Machine Learning Algorithms for Disease Prediction" by D. D. Mau, K. M. K. Sung, and J. M. Kim. *Healthcare Informatics Research*, Vol. 26, No. 4, October 2020.
- [5] Predictive Modeling for Disease Diagnosis and Prognosis: A Review of the Literature" by A. U. Khan, M. A. Ullah, and S. Ali. *IEEE Reviews in Biomedical Engineering*, Vol. 14, 2021.