# DISEASE PREDICTION AND RECOMMENDATION APPLICATION

**A PROJECT REPORT**

*Submitted by*

**AMRITHA BARADE**
**(2017103055)**
**HARITHA POORNACHANDRAN**
**(2017103057)**

***in partial fulfilment for the award of the degree***
***of***

**BACHELOR OF ENGINEERING**
***in***
**COMPUTER SCIENCE AND ENGINEERING**

**COLLEGE OF ENGINEERING, GUINDY**

**ANNA UNIVERSITY:: CHENNAI 600 025**

**OCTOBER 2020**

# ANNA UNIVERSITY : CHENNAI 600 025

## BONAFIDE CERTIFICATE

Certified that this project report on "**DISEASE PREDICTION AND RECOMMENDATION APPLICATION**" is the bonafide work of **AMRITHA BARADE AND HARITHA POORNACHANDRAN** who carried out the project work under my supervision.

| | |
|---|---|
| **SIGNATURE** | **SIGNATURE** |
| Dr. S. Valli | Ms. K. Lalitha Devi |
| **HEAD OF THE DEPARTMENT** | **SUPERVISOR** |
| | Teaching Fellow |
| Department of Computer Science and Engineering | Department of Computer Science and Engineering |
| Anna University - Chennai, | Anna University - Chennai, |
| Chennai 600 025, India. | Chennai 600 025, India. |

# TABLE OF CONTENTS

| CHAPTER NO. | TITILE | PAGE NO. |
|---|---|---|
| | **ABSTRACT** | iii |
| | **LIST OF FIGURES** | iv |
| **1** | **INTRODUCTION** | |
| | 1.1 Project Introduction | 1 |
| | 1.2 Main Objective | 1 |
| **2** | **PROJECT DESIGN** | |
| | 2.1 Block Diagrams | 2 |
| | 2.2 Input and Output | |
| | 2.2.1 Overall Input | 2 |
| | 2.2.2 Overall Output | 2 |
| | 2.2.3 Module 1 | 3 |
| | 2.2.4 Module 2 | 4 |
| | 2.2.5 Module 3 | 5 |
| **3** | **BACKGROUND WORK** | 6 |
| **4** | **FINAL RESULT ANALYSIS** | |
| | 4.1 User Interface | 7 |
| | 4.2 Prediction Algorithms | 8 |
| | 4.3 Comparative Analysis | 9 |
| | 4.4 Result | 11 |
| | 4.5 Conclusion | 12 |
| | **REFERNCES** | 13 |

## ABSTRACT

This project involves the collection of a user's symptoms which may be applicable for a wide range of diseases and then, by use of prediction algorithms, determining the most probable affliction. In this project we have chosen Decision Tree algorithm, Random Forest classifier and Naive Bayes algorithm. Based on this result as well as personal information such as age, three levels of suggestion are to be provided to the patient. The suggestion would include immediate solutions and over the counter medication. It will also suggest an appropriate specialist to consult with if the condition worsens. The purpose of this application is to reduce the burden on hospitals by allowing them to conduct consultations, at least for benign diseases, virtually.

# LIST OF FIGURES

| Figure No. | Title | Page No. |
|---|---|---|
| 2.1.1 | Detailed Block Diagram | **2** |
| 2.2.3.1 | Module 1 – Block Diagram | **3** |
| 2.2.4.1 | Module 2 – Block Diagram | **4** |
| 2.2.5.1 | Module 3 – Block Diagram | **5** |
| 4.1.1 | User Interface | **8** |
| 4.3.1 | Comparison Of Accuracies Across Four Trials | **9** |
| 4.3.2 | Accuracy Calculation | **10** |
| 4.3.3 | Comparison Of Averages Across Four Trials | **10** |
| 4.4.1 | Console Output | **11** |
| 4.4.2 | Final Output | **12** |

# CHAPTER – 1
# INTRODUCTION

## 1.1. Project Introduction

Our project aims to fill the gap between various resources for looking up symptoms and the many advice blogs available online on how to deal with a particular condition. Most people tend to experience a few symptoms and after a quick Google search, start self-medicating. This can be dangerous. With the help of classification and predictive algorithms, we aim to build a functional prototype to not only analyze symptoms and inform the probability of disease but also give useful suggestions based on the input of each individual user.

## 1.2. Main Objective

With the increasing technological advancements in the medical field and the availability of surplus amounts of data online, we plan on building an application that uses patient's data to identify the afflictions. Then use this prediction to suggest/recommend what to do next. This is different from other predictive projects in that it also gives information on how to move on. It is useful as it reduces hysteria and gives peace of mind for patients who otherwise simply Google symptoms and assume, they have the worst.

# CHAPTER -2
# PROJECT DESIGN

## 2.1 Block Diagram:

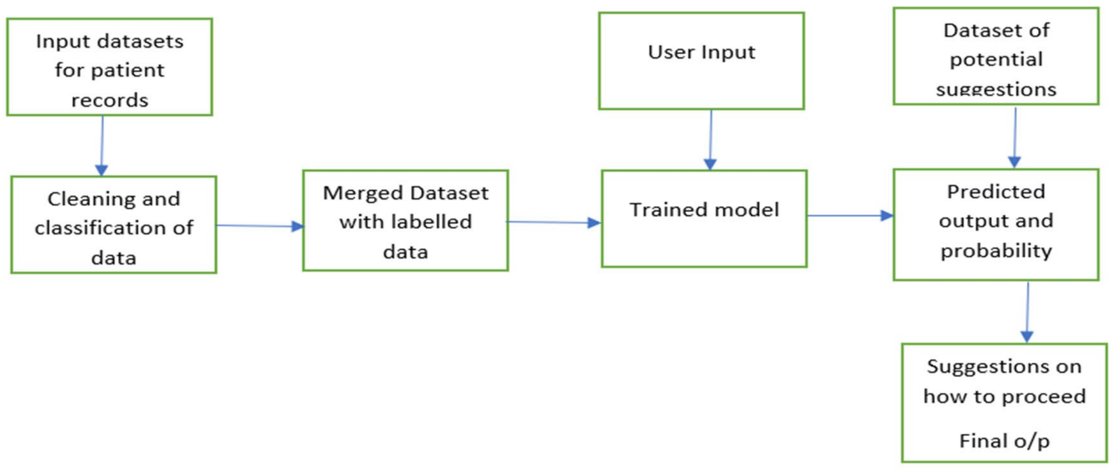

Input datasets for patient records
Cleaning and classification of data
Merged Dataset with labelled data
User Input
Trained model
Dataset of potential suggestions
Predicted output and probability
Suggestions on how to proceed
Final o/p

Figure 2.1.1 DETAILED BLOCK DIAGRAM

## 2.2 Input and Output

### 2.2.1 Overall Input:

We import a variety of datasets from resources like Kaggle, data world where we first cleanse and enrich the data using best qualitative and quantitative data cleansing techniques and then train the cleansed data using predictive models to make predictions and finally use a recommendation model to list out suggestions.

### 2.2.2 Overall Output:

Using the data cleaning techniques, we get clean datasets free from redundancies and anomalies. When training algorithms like Naive Bayes, Decision Trees etc..

are used on the enriched datasets, we get the probability and severity of the type of flu the user may be having and based on this prediction and the degree of severity, the recommendation algorithm gives suggestions on what the user must do.

### 2.2.3 Module 1:

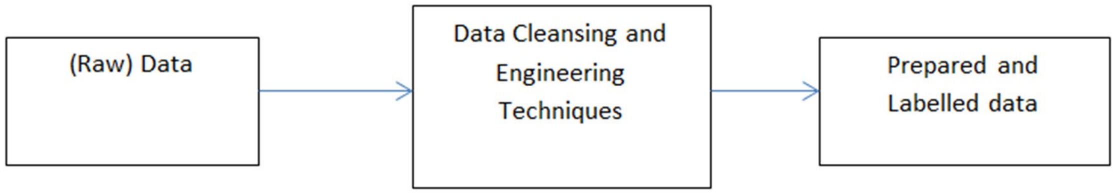

(Raw) Data
Data Cleansing and Engineering Techniques
Prepared and Labelled data

Figure 2.2.3.1 MODULE 1 – BLOCK DIAGRAM

Input:

We use a variety of datasets here to predict the type of disease and also give suggestions. Before we train and feed the data to the prediction and recommendation models, to perform all the intensive tasks of our application, we have to cleanse the error-prone dataset before feeding it to the algorithms. We use qualitative and quantitative data-cleansing techniques to classify the data errors. The first step is error identification i.e. to identify the anomalies and set up a quality plan for data cleansing.

Output:

After you perform data cleansing techniques like filling-out the missing and incomplete values, removing or deleting rows with missing values, fixing typographical errors and inconsistencies in the structure, reducing the data for proper data handling, we have error-free clean datasets which can be fed to the training algorithms to facilitate better decision making and automation tasks.

### 2.2.4 Module 2:

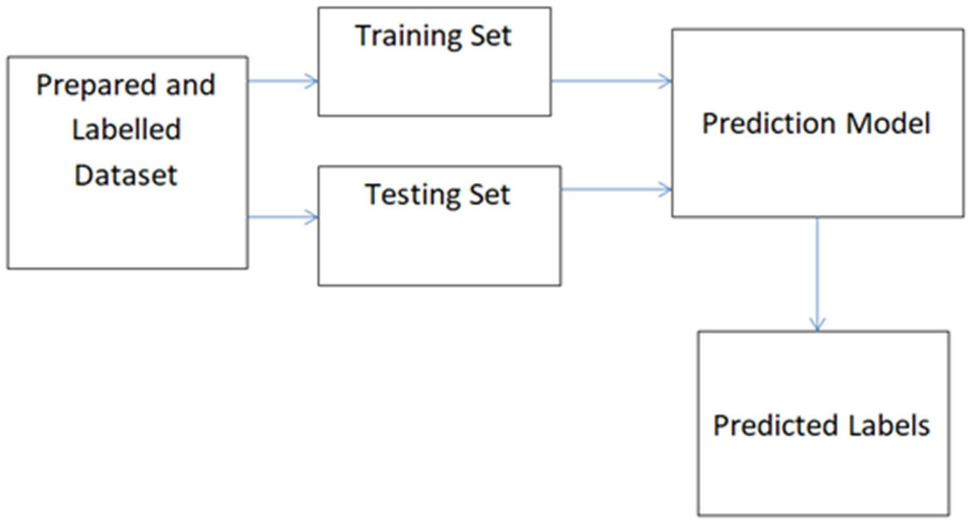


Figure 2.2.4.1 MODULE 2 – BLOCK DIAGRAM

Input:

The algorithms that you may use can be powerful, but without the relevant or right data training, your system may fail to yield ideal results. So we take our cleansed and enriched datasets and train them using predictive modelling approaches like Decision trees, Naive Bayes, Random Forest algorithms to make the right predictions based on the symptoms and other user details like age, height, weight, calculated BMI - obtained through the GUI.

Output:

Using statistics, the predictive models predict the disease a user is most probable to have based on the symptoms and other user-inputs given.

### 2.2.5 Module 3:

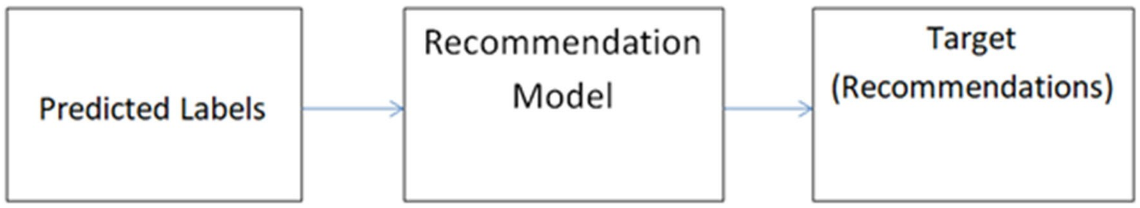


Figure 2.2.5.1 MODULE 3 – BLOCK DIAGRAM

Input:

Based on the output of the prediction model and other user-inputs, we use a content-based recommendation model to recommend a treatment based predicted diseases and the age group the prospect belongs to.

Output:

Using the recommendation model, the application will give you a set of suggestions like whether he/she has to be hospitalized immediately or if they can just take an over-the-counter tablet or an instant-relief medicine.

# CHAPTER-3
# BACKGROUND WORK

We were inspired to do this project due to the ambiguity surrounding the symptoms of COVID-19 in the initial stages of the pandemic. We found many online resources (WebMD, Healthline etc.) that only gave a list of possible diseases when a patient searches for any one symptom. They would recommend home remedies without any other information such as age or underlying conditions. This usually gave no one clear result, often inaccurate or extreme. This could lead to panicking the user with no clear direction to proceed. At most, these articles would suggest a blanket suggestion of "Seek medical help" for all afflictions.

When researching we came across a few papers that discussed using a set of symptoms to predict diseases using Machine Learning algorithms. We found one paper that discusses a more comprehensive prediction method based on multiple input.

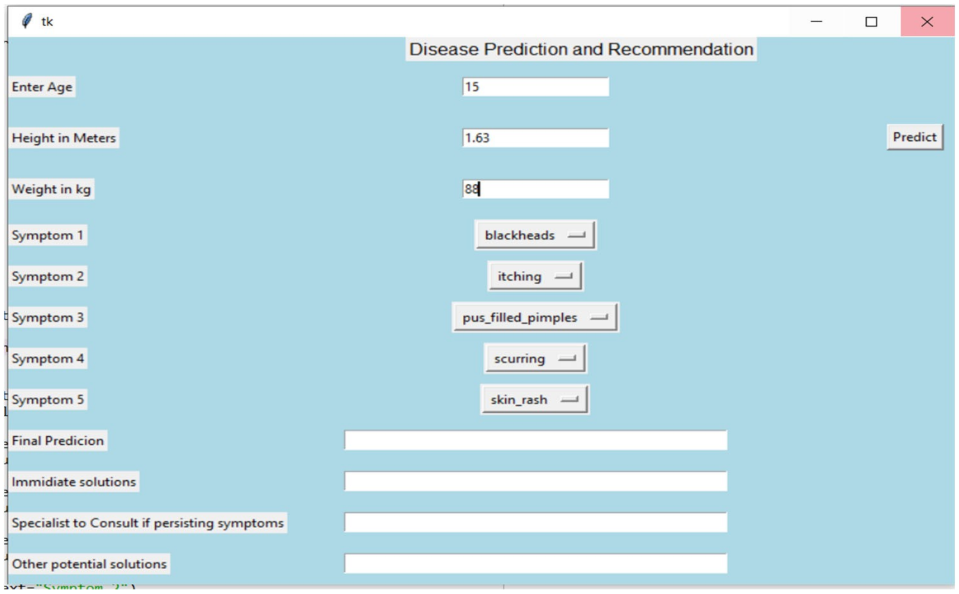
The addition of suggestions on how to proceed once the prediction is accomplished was a component we came up with on our own by discussing what more can be done to improve our application's usability.

# CHAPTER-4
# FINAL RESULT ANALYSIS

## 4.1. User Interface

Using data cleansing and engineering techniques, the prognosis-symptoms dataset was enriched using best practices and was tailor made to suit our application. The prepared and labelled dataset from the data engineering module was used to train the predictive module to find out the highest probable disease that the user might most likely have.

In the interface, the user enters his age using which the application categorizes the user under the right age group and prints it on the console, so that an accurate disease prediction and the right medical recommendation for the user can be done. The user then enters his height in meters and then weight in kilograms which is used for his BMI calculation. Using the BMI, the application checks if the user has obesity or not and the result is printed on the console. If the user has obesity it prints 1, if not then 0. Next, the user prints the top five symptoms he/she is having and then these inputs are fed to the prediction module to predict the affliction and then to the recommendation module, to get the right set of suggestions.



tk

Disease Prediction and Recommendation

Enter Age: 15

Height in Meters: 1.63

Weight in kg: 88

Symptom 1: blackheads

Symptom 2: itching

Symptom 3: pus_filled_pimples

Symptom 4: scurring

Symptom 5: skin_rash

Final Predicion

Immidiate solutions

Specialist to Consult if persisting symptoms

Other potential solutions

Figure 4.1.1 USER INTERFACE

## 4.2. Prediction Algorithms

In the prediction module, we used three different prediction algorithms - Decision Tree Algorithm, Random Forest Algorithm and the Naive Bayes Algorithm. On training, each prediction algorithm predicted a disease that the user might most likely have based on the symptoms given by the user. In order to display a highly accurate result so as to get the right medical recommendation, we took into account all the three diseases predicted by the prediction algorithms and we finally printed one disease - the most probable one, which had the highest number of occurrences and after executing the prediction module for different input samples, we could infer that, out of the three prediction algorithms, it was the **Random Forest Algorithm**, that gave out precise predictions with higher accuracy compared to the other two algorithms

## 4.3. Comparative Analysis

Our project differs from other similar applications by taking not one but three different Machine Learning Prediction Algorithms – Decision Tree, Random Forest, and Gaussian Naive Bayes. Most other similar projects utilize only one algorithm. However, by running all these three algorithms together we have observed a higher average accuracy than that with one algorithm at a time.

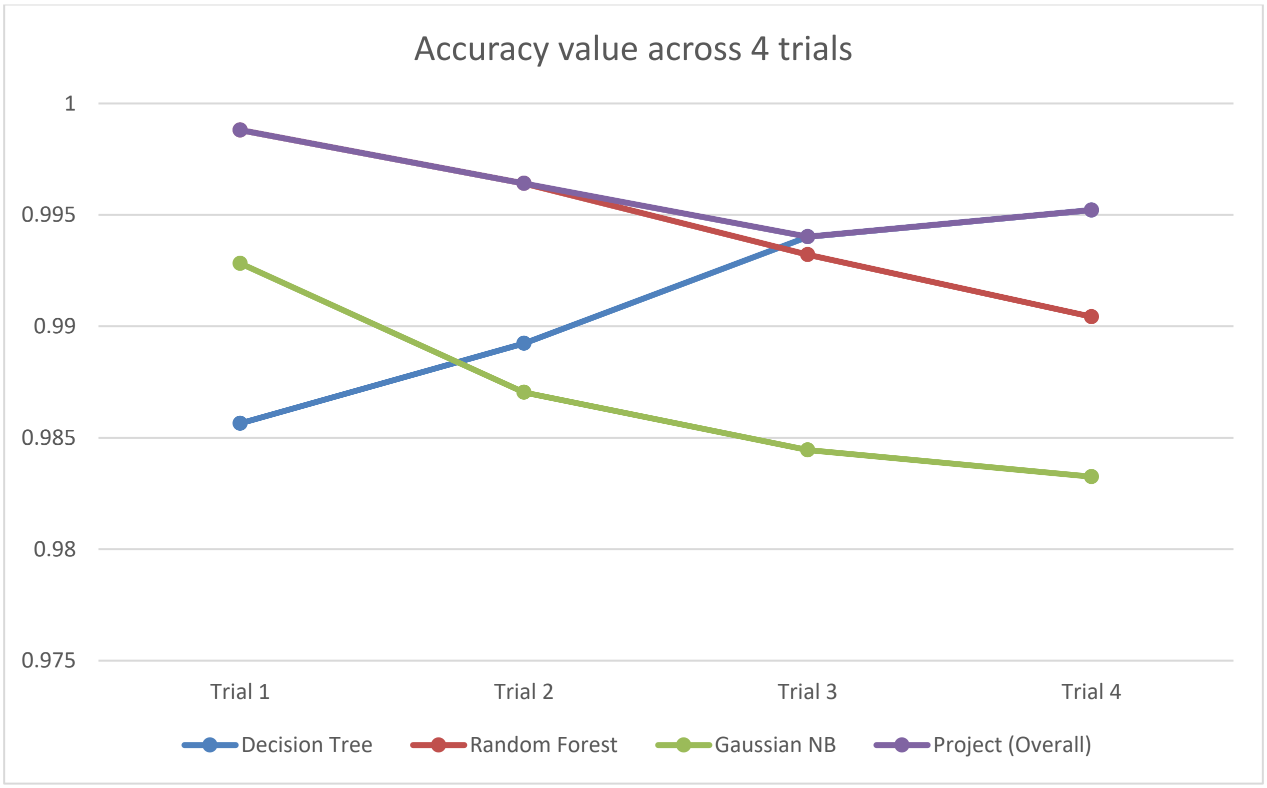
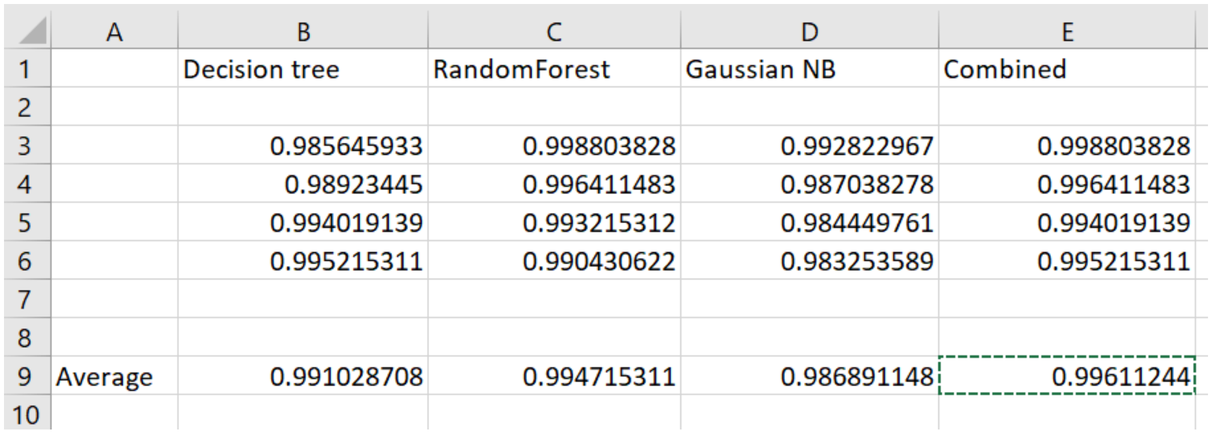


Figure 4.3.1 COMPARISON OF ACCURACIES ACROSS FOUR TRIALS

To compare the results, we ran the program and collected the accuracy of each individual algorithm as well as the overall results accuracy across four trials. The averages were then calculated and interpreted using graph.



| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | Decision tree | RandomForest | Gaussian NB | Combined |
| 2 | | | | | |
| 3 | | 0.985645933 | 0.998803828 | 0.992822967 | 0.998803828 |
| 4 | | 0.98923445 | 0.996411483 | 0.987038278 | 0.996411483 |
| 5 | | 0.994019139 | 0.993215312 | 0.984449761 | 0.994019139 |
| 6 | | 0.995215311 | 0.990430622 | 0.983253589 | 0.995215311 |
| 7 | | | | | |
| 8 | | | | | |
| 9 | Average | 0.991028708 | 0.994715311 | 0.986891148 | 0.99611244 |
| 10 | | | | | |

Figure 4.3.2 AVERAGE CALCULATION

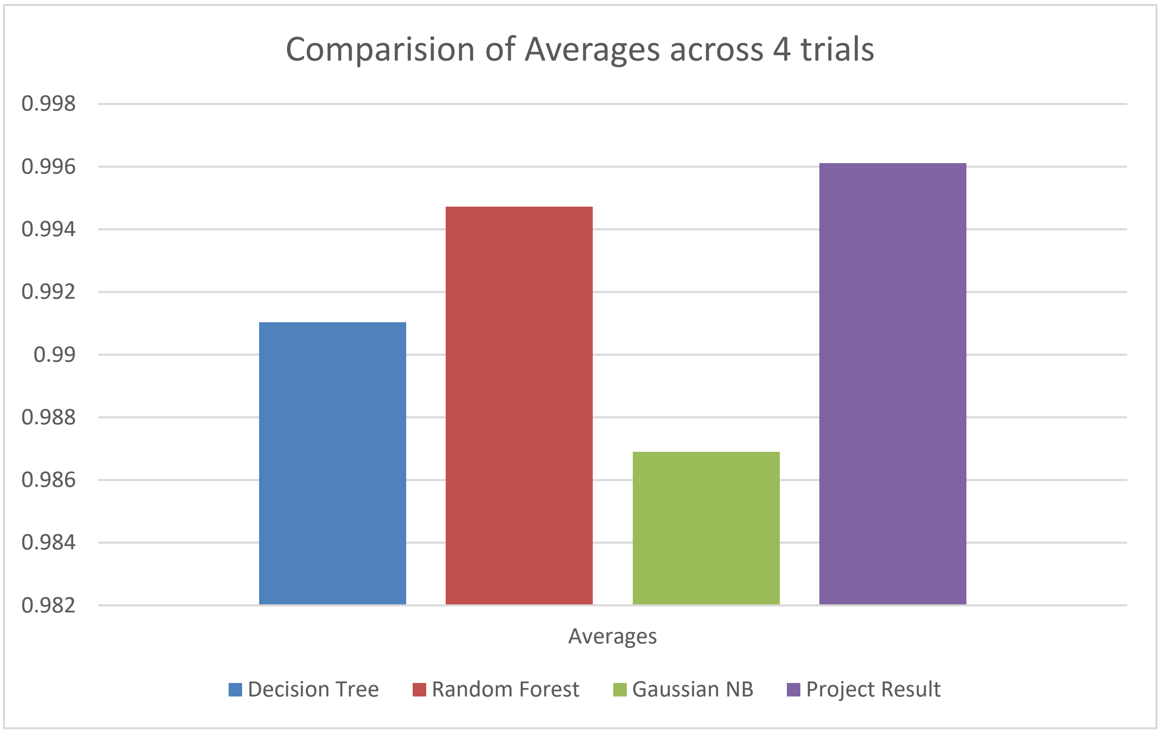


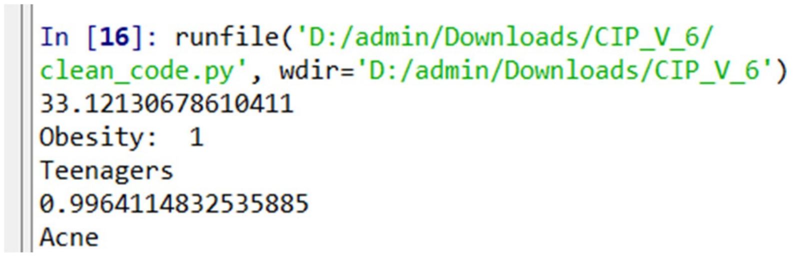
Figure 4.3.3 COMPARISON OF AVERAGES ACROSS FOUR TRIALS

Within the three algorithms **Random forest** showed the highest average accuracy followed by Decision tree algorithm. The project's overall accuracy shows higher value than any one algorithm taken individually. Hence our method of combining the algorithms provides more dependable results.

## 4.4. Result

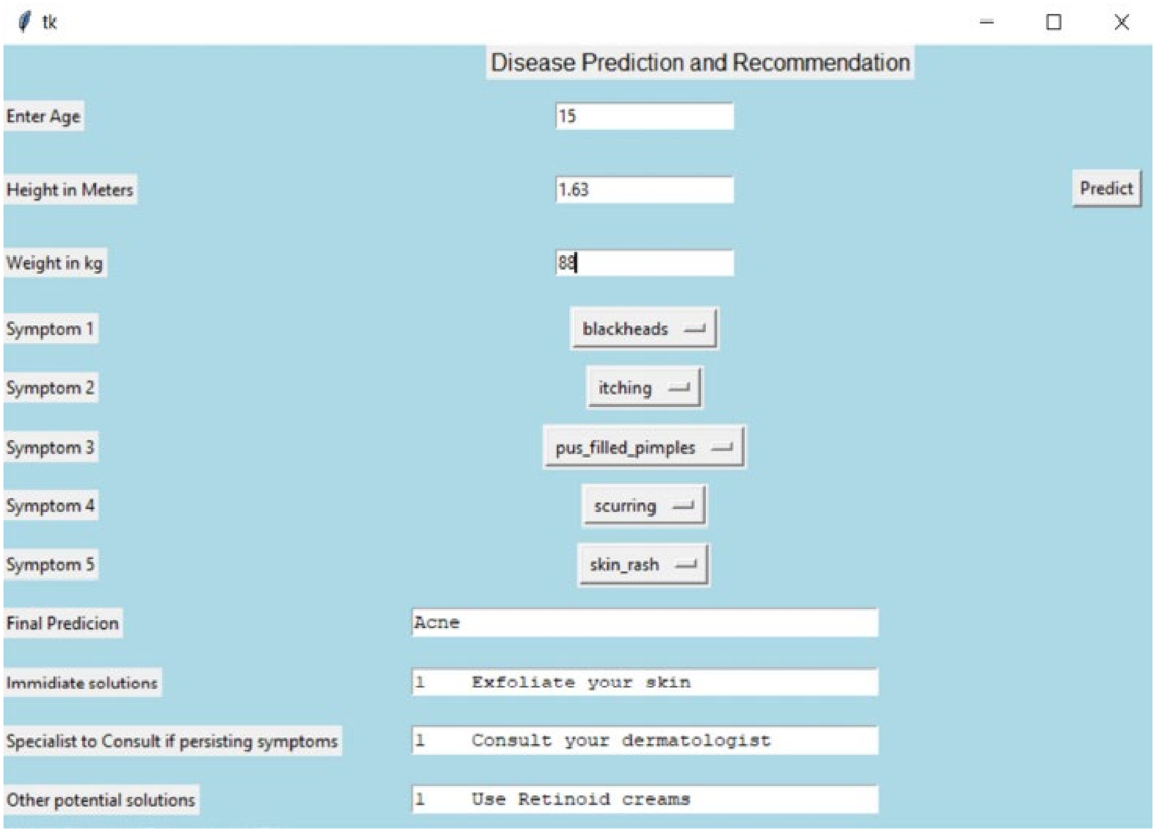
The predicted disease that the user might most likely have is displayed on the interface as the "Final Prediction" and accuracy of the predicted disease is printed on the console.

The output from the prediction module, the prognosis-recommendation dataset and the user-inputs given in the interface are pushed to the recommendation module where we use content-based recommendation. Based on the disease the user might most likely have and the age group he belongs to, a set of medical suggestions which includes what has to be done immediately, what specialist the user has to consult if the given symptoms persist and other potential suggestions are all printed on the interface.



```
In [16]: runfile('D:/admin/Downloads/CIP_V_6/
clean_code.py', wdir='D:/admin/Downloads/CIP_V_6')
33.12130678610411
Obesity:  1
Teenagers
0.9964114832535885
Acne
```

Figure 4.4.1 CONSOLE OUTPUT


tk

Disease Prediction and Recommendation

| | |
|---|---|
| Enter Age | 15 |
| Height in Meters | 1.63 |
| Weight in kg | 88 |
| Symptom 1 | blackheads |
| Symptom 2 | itching |
| Symptom 3 | pus_filled_pimples |
| Symptom 4 | scurring |
| Symptom 5 | skin_rash |
| Final Predicion | Acne |
| Immidiate solutions | 1 Exfoliate your skin |
| Specialist to Consult if persisting symptoms | 1 Consult your dermatologist |
| Other potential solutions | 1 Use Retinoid creams |

Figure 4.4.2 FINAL OUTPUT

## 4.5. Conclusion

By using the "Disease Prediction and Recommendation" application, it becomes easier for people to diagnose the type of disease they are having and also take the right medication to lessen the severity of the symptoms, rather than rushing to the hospital every time or taking the wrong Google-suggested medication without knowing what disease they might be having.

## REFERENCES:

1. https://ieeexplore.ieee.org/document/9002132
2. https://www.aaaai.org/conditions-and-treatments