Contents

Chapter 1

Introduction

1.1 Introduction

Accurate and early disease diagnosis plays a pivotal role in the healthcare industry that enables a timely treatment. However, diagnosing diseases can be complex, requiring a combination of medical knowledge and clinical expertise. Machine Learning (ML) has the potential to support doctors and clinicians by identifying patterns in patient data that may not be discernible to the human eye. The ML model will have a paramount advantage to save Doctors, as well as the patients time by focusing their attention on a particular disease. The model has been integrated with a GUI based form in Python, the symptoms also get stored in a database.

1.2 Motivation

The project was a great learning opportunity for us and was the first project in Python that we worked on. We were excited to begin our journey by applying our skills to the AI/ML domain for real-world use. Dealing with the different types of problems that exist in the current systems is what motivated us to curate a ML model.

Improving and tackling real-life issues motivated us even further. Disease prediction is one of the most crucial and challenging problems in the field of healthcare. Artificial Intelligence (AI) and Machine Learning (ML) techniques have shown promising results in healthcare, including disease prediction.

1.3 Problem Statement & Objectives

These days, healthcare has become very expensive, compelling a lot of patients to refrain from visiting doctors for small illnesses which can eventually lead to major consequences. To overcome this, we have developed a machine-learning model to predict the likelihood of a patient having a particular disease based on the symptoms collected via a GUI based form.

Objectives:

1. After surveying through multiple disease prediction services based on the web platform, we realized that the information asked for predictions may not be available with the patient and quick diagnosis may not be possible. This could lead major complications in the future. For eg. If the form asks for your hemoglobin level, lack of equipment can hamper the predictions in such cases. We aim to overcome this by asking for simple but important symptoms that are very noticeable and very important for prediction.

2. The current systems also give a very extensive report about patient's health condition which may not be a wise option in emergency cases. We plan on giving a short but informative summary of the condition of patient which can be comprehended by the patient easily.

1.4 Organization of the Report

This report consists of three chapters. The first chapter deals with introduction of the topic, problem statement, motivation behind the topic and objectives. The second chapter is the Literature Survey. It includes all the research work done related to this topic. All information related to study of existing systems as well as learning of new tools is mentioned in this chapter. The third chapter is about the proposed system which is used in this project. The block diagram, techniques used, hardware and software used screenshots of the project are presented in this chapter. All the documents related to development of this project are mentioned in references.

Chapter 2

Literature Survey

2.1 Survey of Existing System

1. Symptomate :- Symptomate is an intelligent tool for symptom checking made for you. Each time you go through it and add your symptoms, Symptomate adjusts the interview to your state. Since its creation in 2012, Symptomate has evolved to be the most accurate and accessible symptom analysis tool available to anyone in need. For this reason, it is available for your private use at any time, for free. It also respects your privacy and security, as using Symptomate is always anonymous.

2. Mediktor :- Mediktor is a technology developed by health professionals and based on scientific knowledge extracted from leading medical sources and supported by multiple clinical trials conducted internationally.

2.2 Limitation of the existing system

1. The accuracy of AI-based disease prediction models heavily depends on the quality and quantity of data. Incomplete or inaccurate data can negatively impact the accuracy of the model.
2. AI models may suffer from bias based on the training data, which can result in inaccurate predictions for certain groups of people, particularly underrepresented populations.
3. AI models may be good at predicting diseases in a specific population, but may not be able to generalize to other populations, which limits their applicability.
4. The use of patient data for training AI models raises privacy concerns, and there is a need to ensure that patient data is adequately protected.
5. The information required can be very complex to fetch and hence lead to incorrect predictions.

.

2.3 Mini Project Contribution

An AI-based disease prediction model can provide significant contributions to people in various ways:

1. Early detection: Early detection is crucial for effective disease management and treatment. An AI-based disease prediction model can help detect diseases at an early stage, which can increase the chances of successful treatment.

2. Personalized healthcare: By analyzing patient data, an AI-based disease prediction model can provide personalized healthcare recommendations, such as lifestyle changes, treatment plans, and preventive measures, tailored to individual patient needs.

3. Improved patient outcomes: AI-based disease prediction models can help healthcare providers make more informed decisions, leading to improved patient outcomes and better health outcomes overall.

4. Cost-effective healthcare: Early disease detection and personalized healthcare recommendations can reduce the cost of healthcare services by preventing expensive and unnecessary treatments.

5. Increased efficiency: An AI-based disease prediction model can help healthcare providers prioritize patients who are at higher risk of developing specific diseases, resulting in more efficient use of resources

Chapter 3

Proposed System

3.1 Introduction

The basic structural background of our model works 2 types of machine learning algorithms.

NAÏVE BAYES ALGORITHM

Naive Bayes algorithm is a classification technique based on Bayes' theorem with the assumption of independence between predictors. It is a probabilistic algorithm that calculates the probability of each class given the input variables and selects the class with the highest probability as the output.

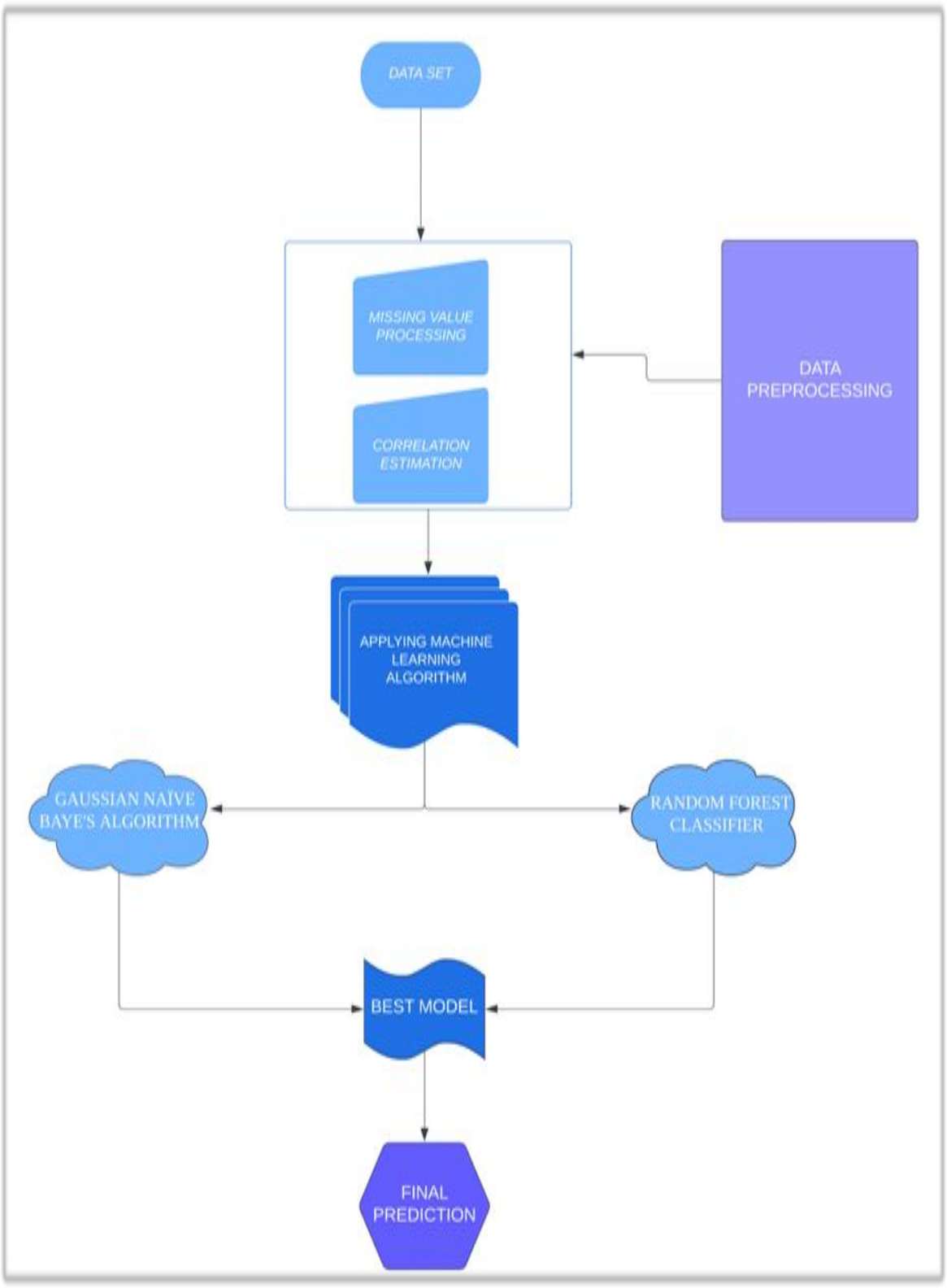
Naive Bayes assumes that all input variables are independent of each other, which is often not true in real-world scenarios. However, despite its simplifying assumptions, the Naive Bayes algorithm is widely used in text classification and spam filtering, among other applications.

RANDOM FOREST CLASSIFIER

Random Forest Classifier is a type of ensemble learning algorithm that combines multiple decision trees to make more accurate predictions. Each decision tree in the Random Forest is trained on a random subset of features and samples from the dataset to reduce overfitting.

The algorithm works by constructing a multitude of decision trees and then taking a majority vote to determine the final prediction. This makes the algorithm more robust to noise and outliers in the data, as it is less likely to overfit on the training data.

3.2 Architecture



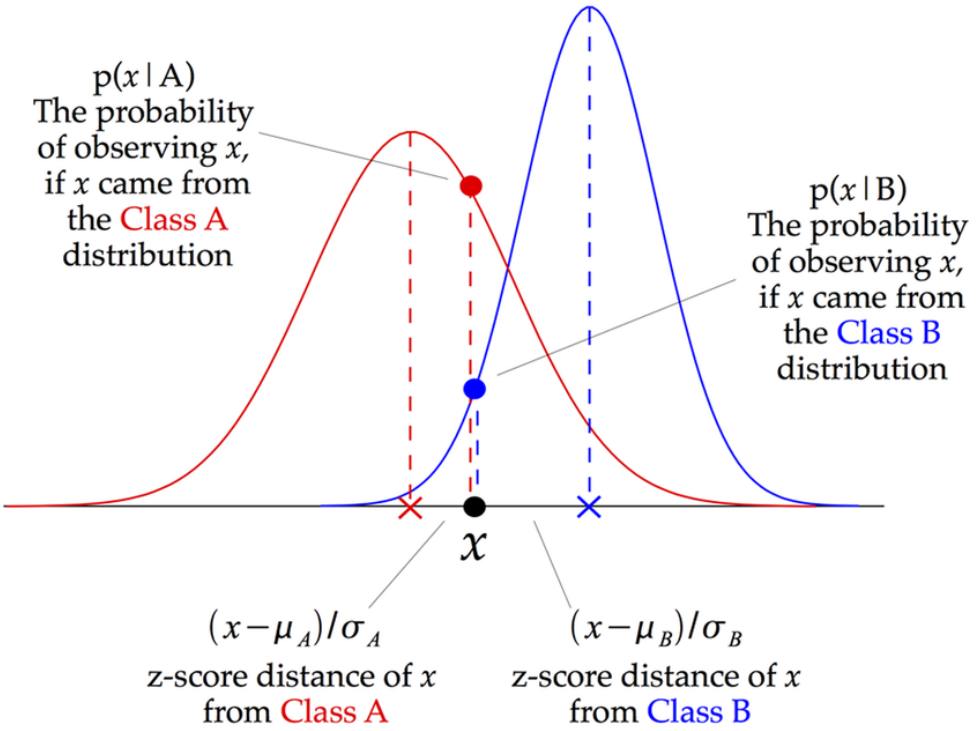
3.3 Algorithm and Process Design

NAÏVE BAYES ALGORITHM:

- A Naïve Bayes Classifier is a probabilistic machine learning model that's used for a classification task.
- The crux of the classifier is based on the Bayes Theorem.

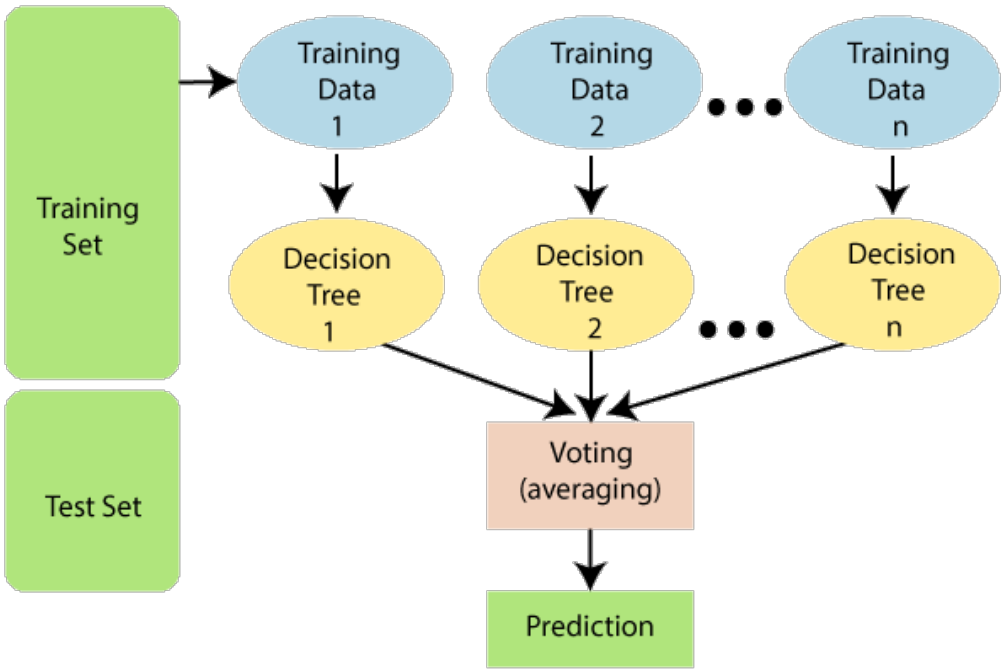
$$\mathbf{P}(\boldsymbol{y}|\boldsymbol{x_1 x_2}, \ldots\ldots \boldsymbol{x_n}) = \frac{[\{\mathbf{P}(\boldsymbol{x_1}|\boldsymbol{y}) * \boldsymbol{P}(\boldsymbol{x_2}|\boldsymbol{y}) \ldots . \mathbf{P}(\boldsymbol{x_n}|\boldsymbol{y})\} * \mathbf{P}(\boldsymbol{y})]}{\{\mathbf{P}(\boldsymbol{x_1})\mathbf{P}(\boldsymbol{x_2}) \ldots . \mathbf{P}(\boldsymbol{x_n})\}}$$

- Where, $\{\boldsymbol{x_1 x_2}, \ldots\ldots \boldsymbol{x_n}\}$ are set of Mutually independent features.
- The fundamental Naïve Bayes assumption is that each feature makes an :
  1.Independent
  2.Equal
  contribution to the outcome
- We have made use of the Gaussian Naïve Bayes Classifier with an assumption that all the features obey the Gaussian distribution plot so as to tackle the 'zero-frequency problem', indirectly contributing to the accuracy of our model.
-



RANDOM FOREST CLASSIFIER:

- It is a Supervised Machine Learning algorithm that is mainly used for classification and regression problems.
- It contains multiple decision trees working on a random subset of the given dataset that will individually predict the occurrences of each disease and based on majority voting, the appropriate disease will get predicted.
- It uses Bagging, an ensemble technique to randomly select the sub-data with replacement
- It reduces the probability of overfitting due to majority voting.



3.4 Details of Hardware & Software

HARDWARE:

- A device capable of hosting the chatbot after the compilation of Python code.
- Modern Operating System
- 4 GB RAM
- 5 GB free disk space

SOFTWARE

- Python IDE which supports the Tkinter module
- Jupyter,VScode

3.5 Results

Our Login screen will look as follows:

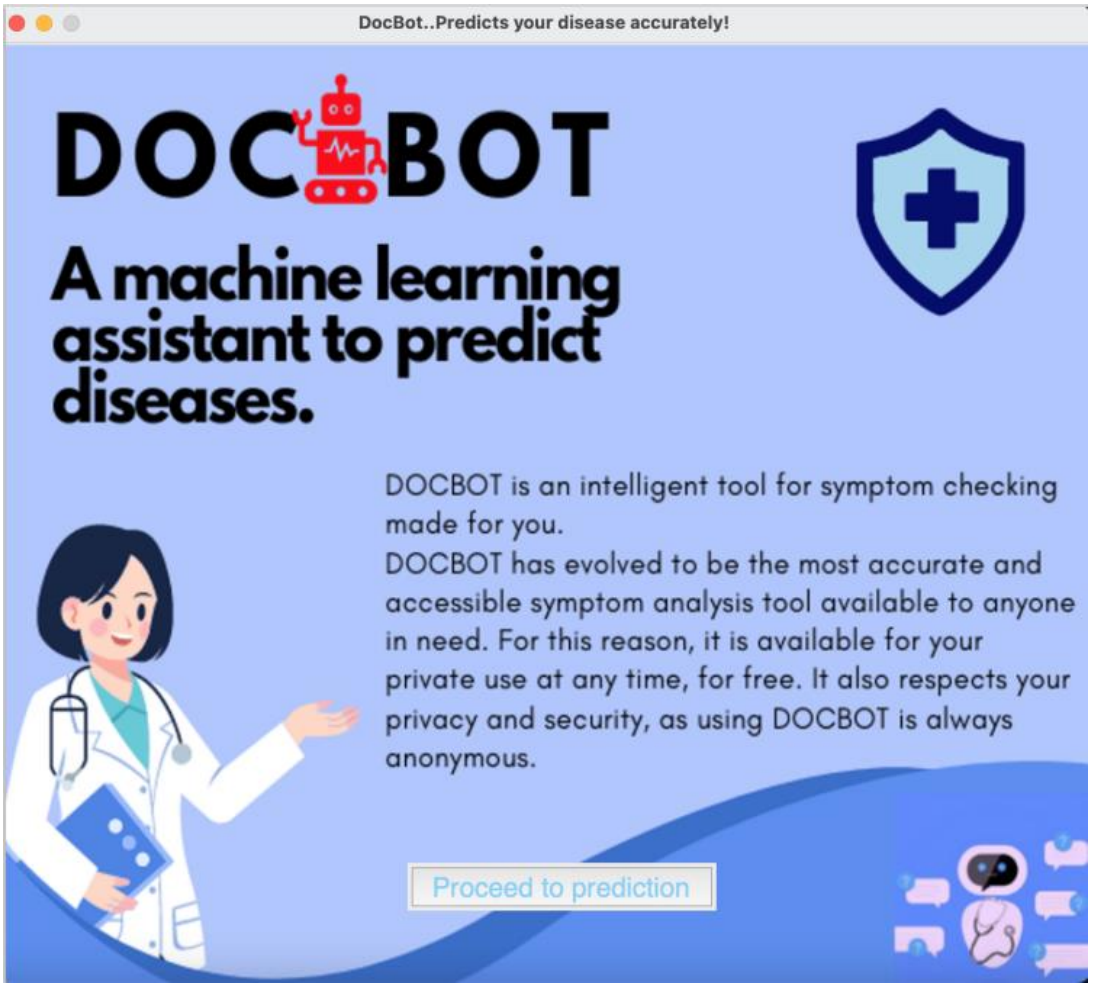


Fig no:- 2.1 First Window

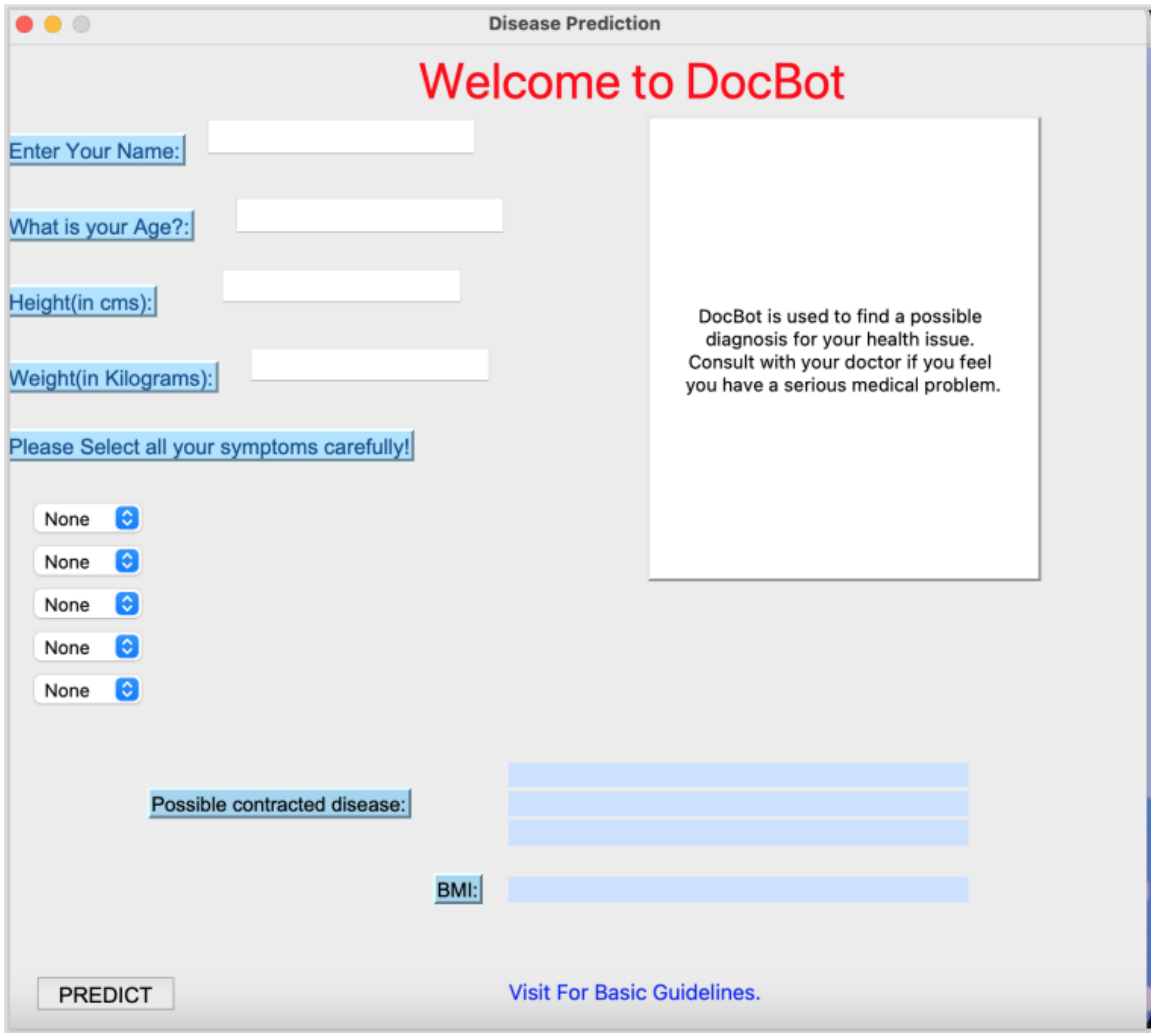


Fig no :- 2.2 Prediction Window

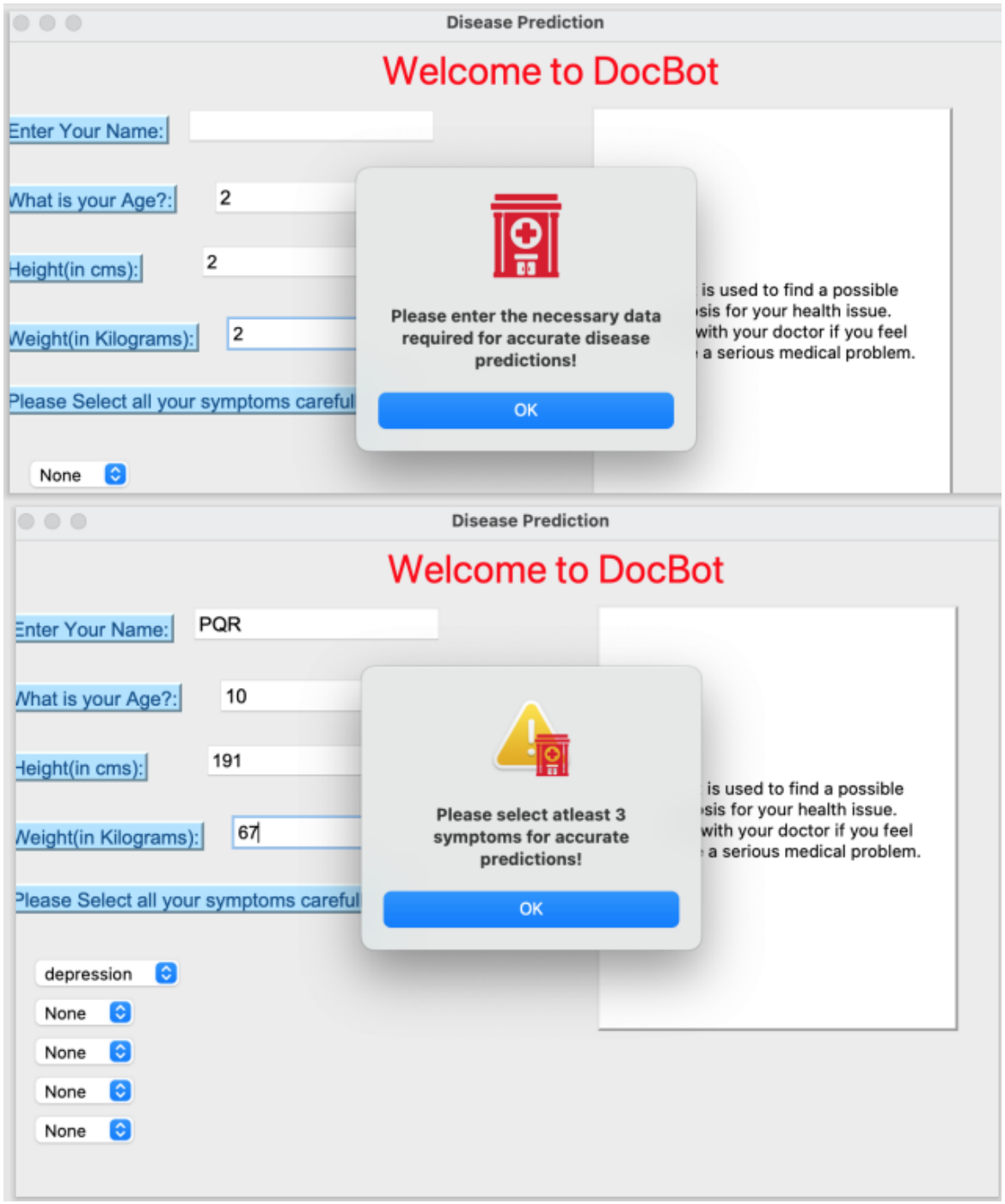
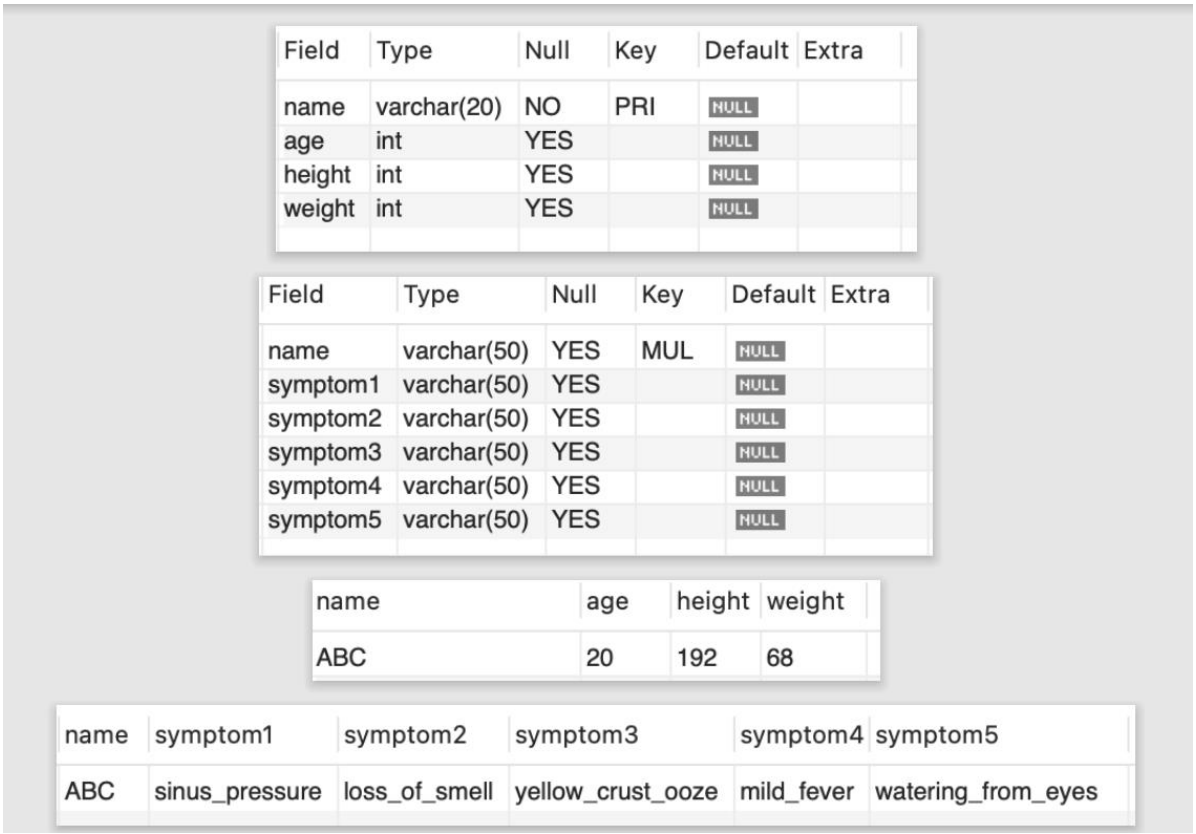


Fig no:2.3 Error popup

Fig no:2.4 Prediction Result



Field	Type	Null	Key	Default	Extra
name	varchar(20)	NO	PRI	NULL	
age	int	YES		NULL	
height	int	YES		NULL	
weight	int	YES		NULL	

Field	Type	Null	Key	Default	Extra
name	varchar(50)	YES	MUL	NULL	
symptom1	varchar(50)	YES		NULL	
symptom2	varchar(50)	YES		NULL	
symptom3	varchar(50)	YES		NULL	
symptom4	varchar(50)	YES		NULL	
symptom5	varchar(50)	YES		NULL	

name	age	height	weight
ABC	20	192	68

name	symptom1	symptom2	symptom3	symptom4	symptom5
ABC	sinus_pressure	loss_of_smell	yellow_crust_ooze	mild_fever	watering_from_eyes

Fig no:2.5 Database

3.7 Conclusion and Future Work

In conclusion, a disease prediction model using AI and ML techniques has the potential to revolutionize healthcare by improving disease detection, personalized healthcare, patient outcomes, and cost-effectiveness. While there are limitations to AI-based disease prediction models, such as data quality, bias, and lack of transparency, the benefits of these models are significant. By analyzing patient data, an AI-based disease prediction model can provide valuable insights that can inform healthcare providers' decision-making and ultimately improve patient health outcomes.

References

[1] https://symptomate.com/
[2] https://www.mediktor.com/
[3] https://www.semamticscholar.org
[4] https://ieeexplore.ieee.org
[5] https://www.kaggle.com/datasets/rabisingh/symptom- checker?select=Training.csv