VIT®
BHOPAL
www.vitbhopal.ac.in

# Project Exhibition 1

GROUP 17

# DISEASE PREDICTION FROM SYMPTOMS USING ML

# Presented By


**AYUSH GUPTA**

21BHI10055

GUIDE: DR. SWAGAT KUMAR SAMANTRAY

# Introduction

- The reality of modern day medicine is that there are many more patients than doctors who are available to treat them.
- With each and every passing day complex and new diseases are being noticed. This has lead  to extreme stresss on medical infrastructure available thus more number of people being deprived of quality healthcare.
- For people living in rural areas are completly dependent on government hospitals and clinics and the government allopathic doctor to patient ratio is 1:12000 according to National Health Profile 2021.

# Introduction

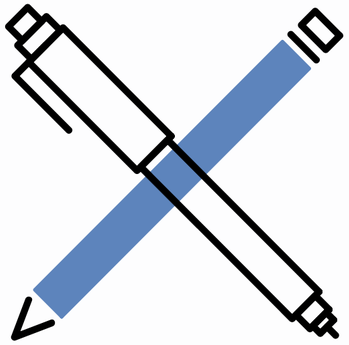
- The availability of physicians and nurses varies widely across the country with central, northern and north-eastern states being poorly served.
- Rural areas have an especially severe shortage of qualified health professionals.
- The main purpose of this system is to predict the disease and controlling the disease by accurate and trustworthy predictions.
- Depending upon the disease being diagnosed, a specialist doctor will be assigned.

# Our Vision

To provide a quick medical diagnosis to the patients in remote rural and inaccesible ares

To provide contactless healthcare services in postcovid times

To enhance quality of healthcare by increasing accessibility to everyone



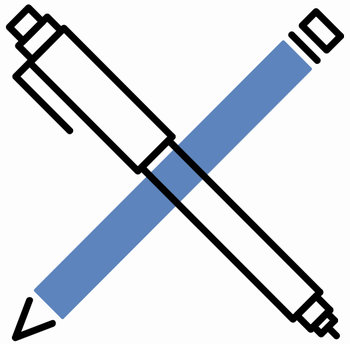
# Proposed Work

## Gathering the Data

The dataset consists of two CSV files one for training and one for testing. There is a total of 133 columns in the dataset out of which 132 columns represent the symptoms and the last column is the prognosis.

## Model Building

After gathering the data is used to train a machine learning model. We will be using this to train the Decision tree, KNN, Naive Bayes Classifier, and Random Forest Classifier.



# Proposed Work

After training the four models we will be predicting the disease for the input symptoms. This makes our overall prediction more robust and accurate.

At last, we will be defining a function that takes symptoms separated by commas as input, predicts the disease based on the symptoms by using the trained models, and returns the predictions.

# Workflow Design

New User

Register

Patient Detail Database

PATIENTS

Existing User

Enter User ID & password

Input Symptoms

Trained Model

Datasets

Machine Learning

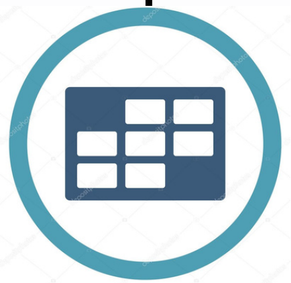
Doctor

Suggest Doctor

Show predicted disease

Patient Symptom Database

Doctor Detail Database



Datasets

Train data

Preprocessing and splitting the data

Train Data

Performing K fold Cross Validation for Model Selection

Training Decision tree

Training Naive Bayes Classifier

Training Random Forest Classifier

Test data

Computing Metrices on test data for Decision tree

Computing metrices on test data for Naive Bayes Classifier

Computing Metrices on test data for Random Forest Classifier

Validation Data

Predictions on validation dataset by Decision tree

Prediction on validation dataset by Naive Bayes Classifier

Prediction on Validation dataset by Random Forest Classifier

Final Prediction

# Software requirements



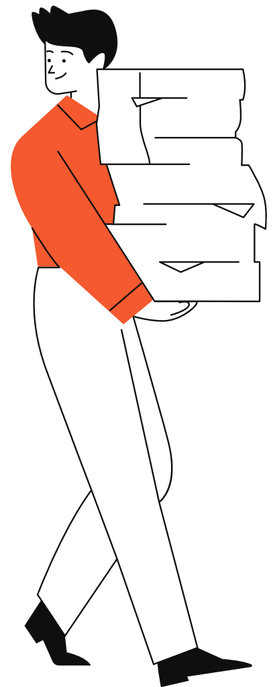


Visual Studio Code

# Libraries Used

**Pandas**

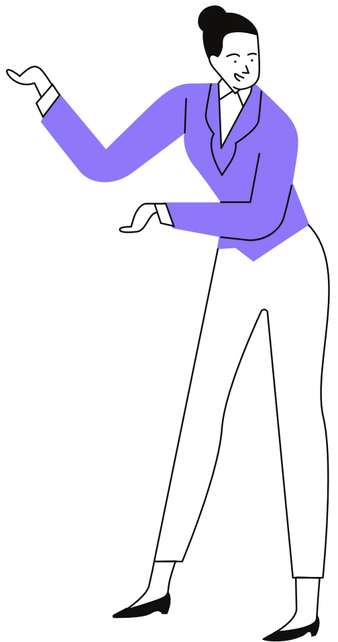
**Numpy**



**Sklearn**

**Tkinter**

**Matplotlib**



# Functions Used

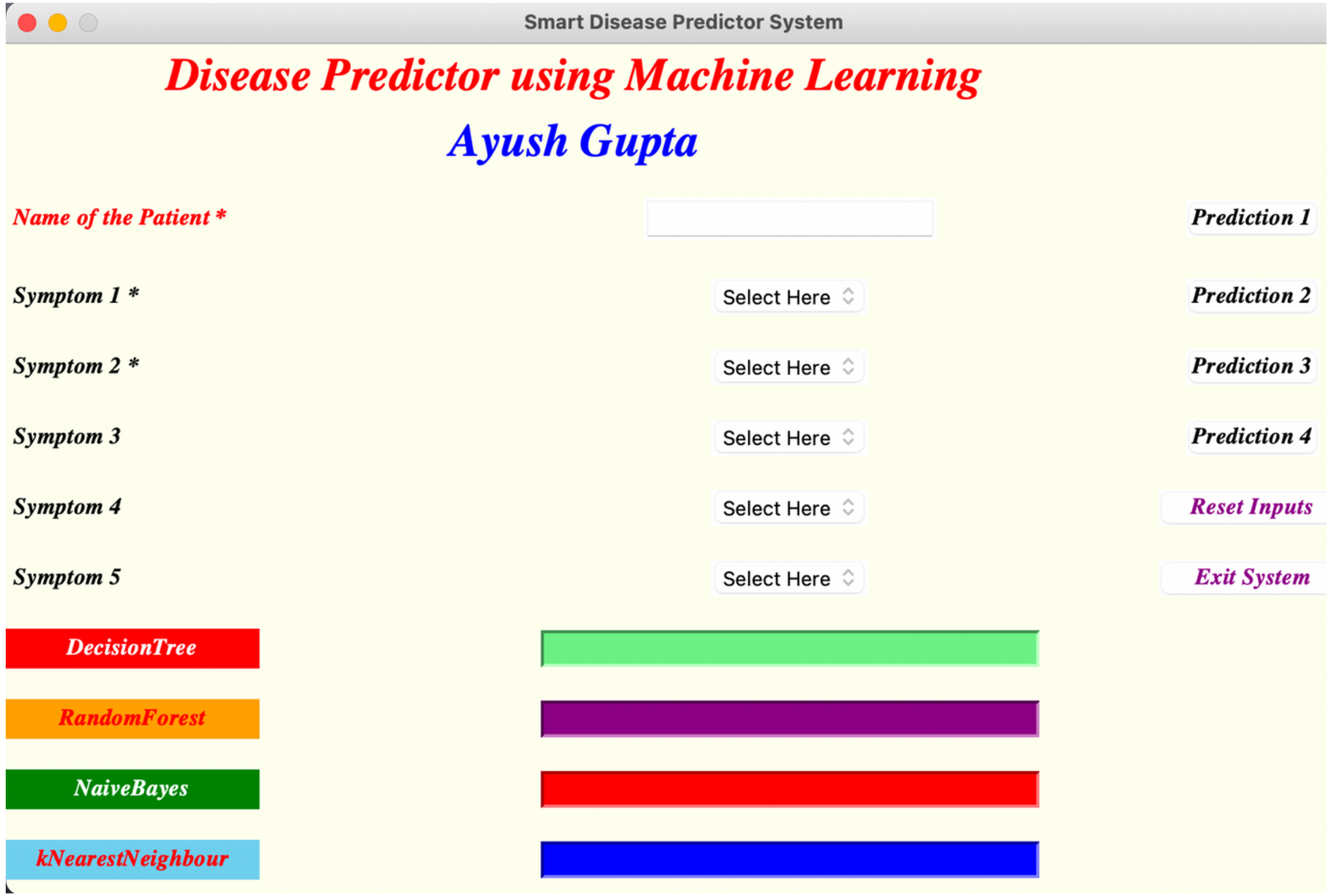
1. **plotScatterMatrix()**
Used for plotting scatter and density plots

2. **scatterplot()**
Scatterplot of prediction of diseases

3. **plotPercolumDistribution()**
Shows distribution graph of column data

4. **scatterinp()**
Scatterplot of symptoms entered

# GUI



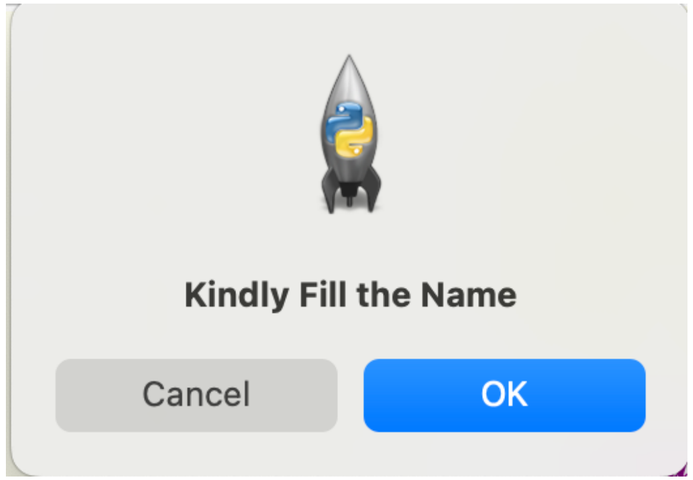

**Title, description and labels** are set for different sections.

**Option menu** <Select Here> is used to shows the list of different supported diseases.

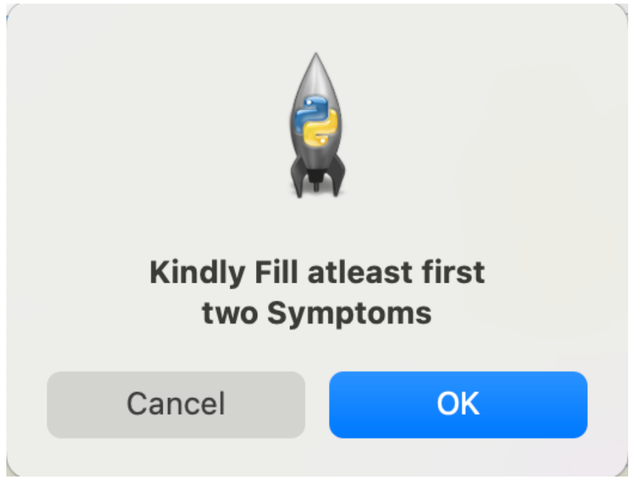
**Prediction buttons** are used to give functionalities and predict the outcome of models.

**Text is used** to show output of the prediction using blank space.

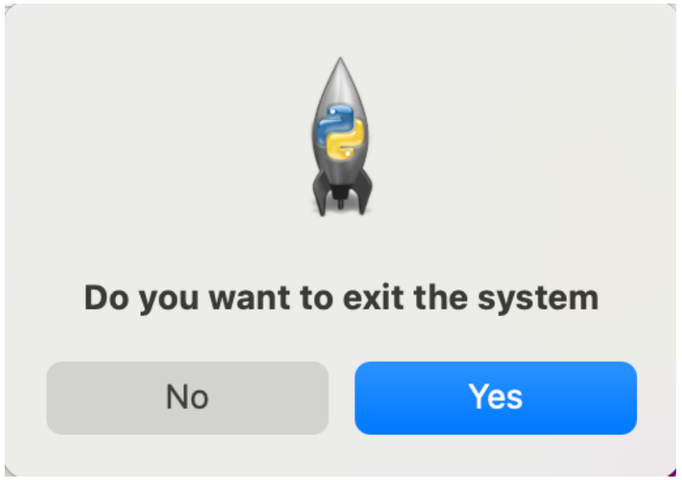
# GUI




If name is not entered, we ask to fill the name.




If atleast 2 symptoms are not entered, we ask to fill the same.




If you try to exit the application, a double confimation on the same.

# ML – Algorithms Used

DECISION TREE ALGORITHM

RANDOM FOREST ALGORITHM

K NEAREST NEIGHBOUR

NAIVE BAYES ALGORITHM

# DECISION TREE ALGORITHM

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. The decision tree learning is like as decision tree algorithm which uses maps input about an item to output of the item. The tree models with finite classes of output are called classification trees. In these tree structures leaves shows class labels and branches shows relation between attributes that the results in those class labels of the system.

Decision trees with continuous output classes are called regression trees.

# RANDOM FOREST ALGORITHM

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

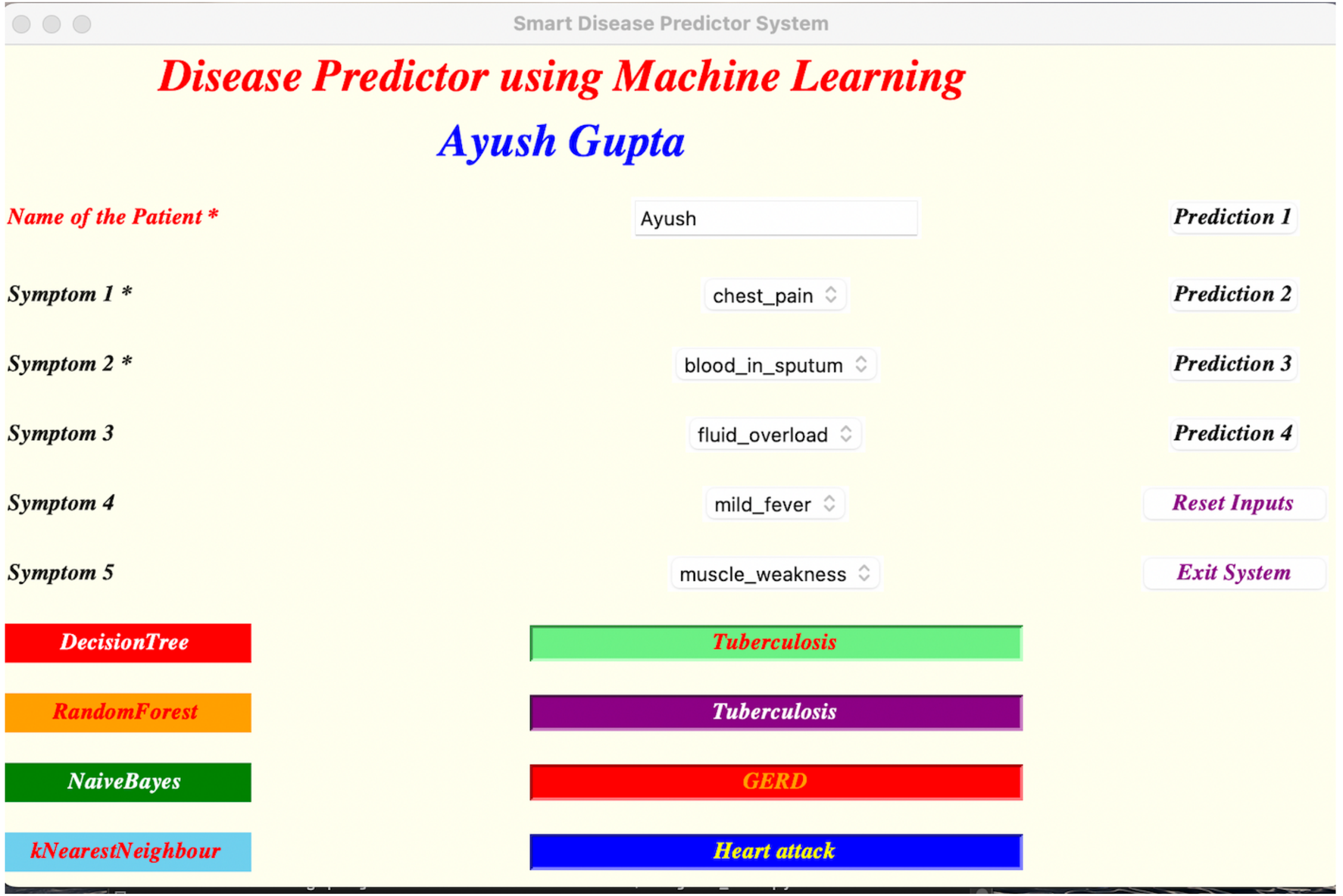
# K NEAREST NEIGHBOUR

KNN finds intense application in pattern recognition, data mining and intrusion detection.
It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data.

# NAIVE BAYES ALGORITHM

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. The supervised machine learning method of classification is represented by Naive Bayes algorithm. It uses a probabilistic model by determining probabilities of the outcomes/outputs. It is used in analytical and predictive problems. Naive
Bayes is robust to noise in input dataset.

DEMO


Smart Disease Predictor System
Disease Predictor using Machine Learning
Ayush Gupta
Name of the Patient *
Ayush
Prediction 1
Symptom 1 *
chest_pain
Prediction 2
Symptom 2 *
blood_in_sputum
Prediction 3
Symptom 3
fluid_overload
Prediction 4
Symptom 4
mild_fever
Reset Inputs
Symptom 5
muscle_weakness
Exit System
DecisionTree
Tuberculosis
RandomForest
Tuberculosis
NaiveBayes
GERD
kNearestNeighbour
Heart attack

# 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 used 4 different algorithms to do so. On an average we achieved accuracy of ~87%. 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 a better version of such a system. Our system also has an easy to use interface.

# RESOURCES USED

kaggle

# Thank you

Have a great day ahead.