



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

Deemed to be University

Declared as category 'A' University by MHRD, Govt. of India
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119, Tamil Nadu. India.



C.S.E / I.T
2017-2021

DISEASE PREDICTION USING NAIVE BAYES

Project Supervisor : Dr. B. Ankayarkanni, M.E., Ph.D.,

Team Members : Ms. Roop Chandrika Mallela

Reg. No : 37110648

Ms. Reddy Lakshmi Bhavani

Reg. No : 37110637

Presentation Outline

- Introduction
- Literature Review
- Objectives
- System Architecture
- Project Implementation
- Methodology
- Project Snapshots
- Results and Discussions
- Conclusion
- References

Introduction

- Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. The traditional way of diagnosis may not be sufficient in the case of a serious ailment.
- In this situation, where everything has turned virtual, the doctors and nurses are putting up maximum efforts to save people's lives even if they have to danger their own. There are also some remote villages which lack medical facilities.
- The dataset was processed in ML models Naive Bayes and Decision Tree. While processing the data, symptoms are given as input and the disease was received as an output.
- This project helps to get the idea about the disease of an individual based on the symptoms he/she have, and get the treatment easily by contacting the concern doctor.

Literature Review

AUTHOR	YEAR OF PUBLICATION	DESCRIPTIONS	PROS	CONS
K. Gomathi Kamaraj	2016	Multi Disease Prediction using Data Mining Techniques.	Evaluate the best classifier.	Concentrate on only finding the best classifier.
Sonam Nikhar	2018	Prediction of Heart Disease Using Machine Learning Algorithms.	Used combination of algorithms like Naïve Bayes and decision tree to get the best accuracy.	Naïve Bayes accuracy should be improved.
Benjamin Fredrick David. H	2020	Heart Disease Prediction Using Data Mining Techniques	Predict more exactly the occurrence of heart disease using data mining techniques	Research work can be made to produce an impact in the accuracy of the Decision Tree and Bayesian Classification.

Literature Review

AUTHOR	YEAR OF PUBLICATION	DESCRIPTIONS	PROS	CONS
Akash C. Jamgade	2019	Disease Prediction using Machine Learning.	Determine the major chronic diseases in a particular region and in particular community.	Data collection is difficult.
Ana-Maria Mihaela	2020	A review of Data Mining Techniques in medicine.	Predicts diseases like breast cancer, lung cancer and so on.	Time consuming.
Siddhika Arunachalam	2020	Cardiovascular Disease Prediction Model using Machine Learning Algorithms	Six classification algorithms are used after analyzing 14 attributes in the dataset.	Sometimes we get confused what algorithm to use.
23 February 2021		School of Computing		5

Existing and Proposed System

Existing System

- ❑ In the existing system the data set is typically small, for patients and diseases with specific conditions. These systems are mostly designed for the more colossal diseases such as Heart Disease, Cancer etc.

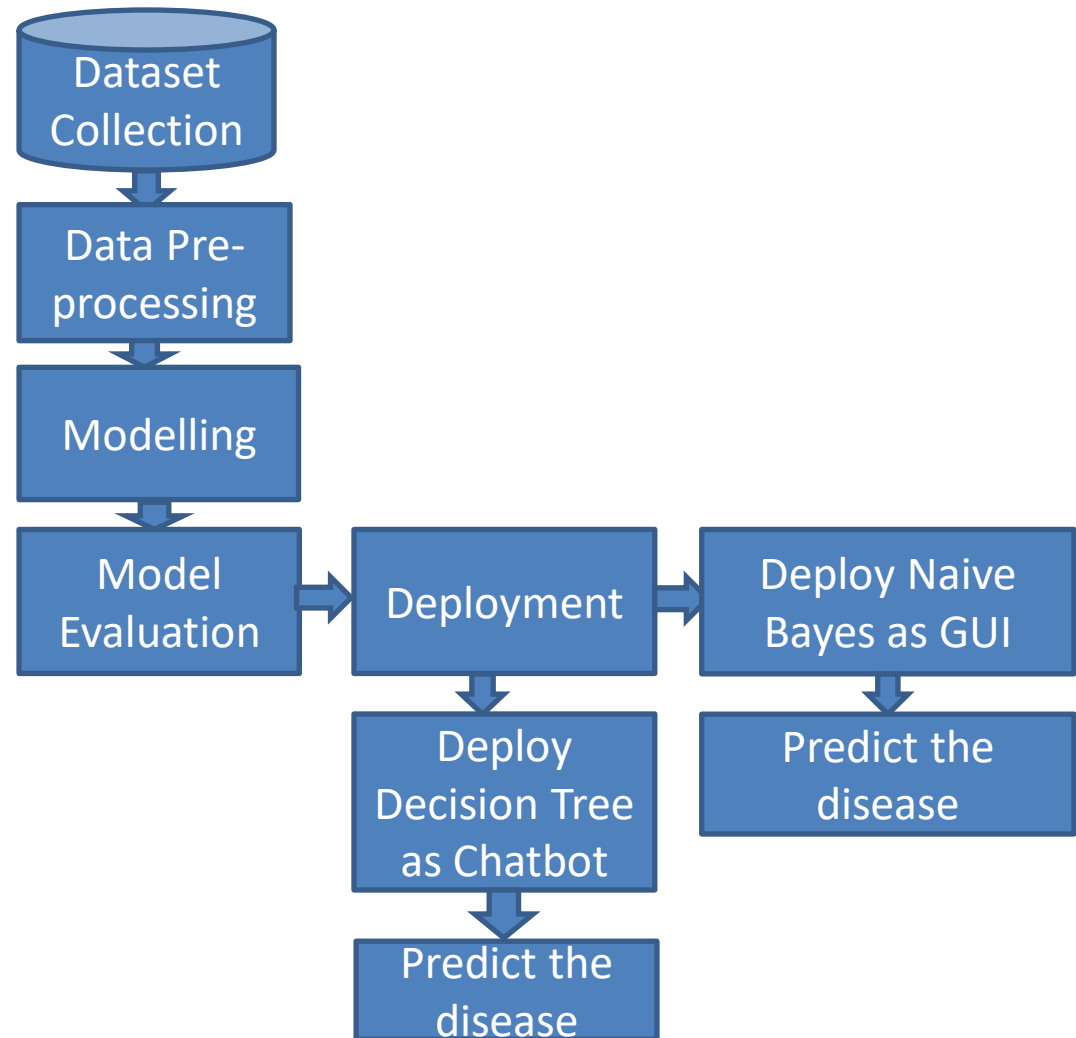
Proposed System

- ❑ We are proposing such a system which is a simple GUI using Tkinter by deploying Naïve Bayes model. Our aim with this system is to get the disease based on symptoms given by the user,
- ❑ And also, we are creating a chatbot using Decision Tree, which will help us in getting accurate predictions by taking into account the symptoms faced by an individual.

Objectives

- Developing a project based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method is the main objective of the project.
- We have designed a disease prediction system using ML algorithm (Naive Bayes), and also chatbot using decision tree which is used it to find the disease based on the symptoms.
- Based on the symptoms of an individual, the ML model gives the output, i.e., the disease that the individual might be suffering from.
- This project helps to get the idea about the disease of an individual based on the symptoms he/she have, and get the treatment easily by contacting the concern doctor.

System Architecture



Project Implementation

Software Requirements:

- ☐ Python
- ☐ Anaconda
- ☐ Jupyter Notebook

Hardware Requirements:

- ☐ Processor: Intel Core i5
- ☐ RAM: 8GB
- ☐ OS: Windows

The libraries used on this project are:

- ☐ Tkinter- For creating GUI based application.
- ☐ Sklearn- used for importing machine learning models, get accuracy, get confusion matrix.
- ☐ Pandas- Used to read the dataset.

Project Implementation

- ❑ Matplotlib- Used for data visualisation.
- ❑ Numpy- Used to change 2-dimensional array into contiguous flattened array.
- ❑ Pandas profiling- Used to get detailed report of the dataset.

Naive Bayes Classifier

It is a machine learning algorithm for classification problems and is based on Bayes' probability theorem. The primary use of this is to do text classification which involves high dimensional training data sets.

$$P(h|d) = \frac{P(d|h).P(h)}{P(d)}$$

$$P(d)$$

Where $P(h|d)$ is the probability of hypothesis h given the data d . This is called the posterior probability. $P(d|h)$ is the probability of data d given that the hypothesis h was true. $P(h)$ is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h . $P(d)$ is the probability of the data (regardless of the hypothesis).

Project Implementation

Decision Tree

Decision Trees are a type of Supervised Machine Learning where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

Following are the steps to do this project (use Jupyter Notebook):

1. Collect the dataset.
2. Import the necessary libraries.
3. Visualise the dataset.
4. Train the dataset using Naïve Bayes classifier and Decision Tree.
5. Test the model and find the accuracies of both.
6. Deploy the model- a) Naïve Bayes as GUI Interface using Tkinter
b) Chatbot using Decision Tree
7. Predict the disease based on the symptoms given by the user.

Methodology

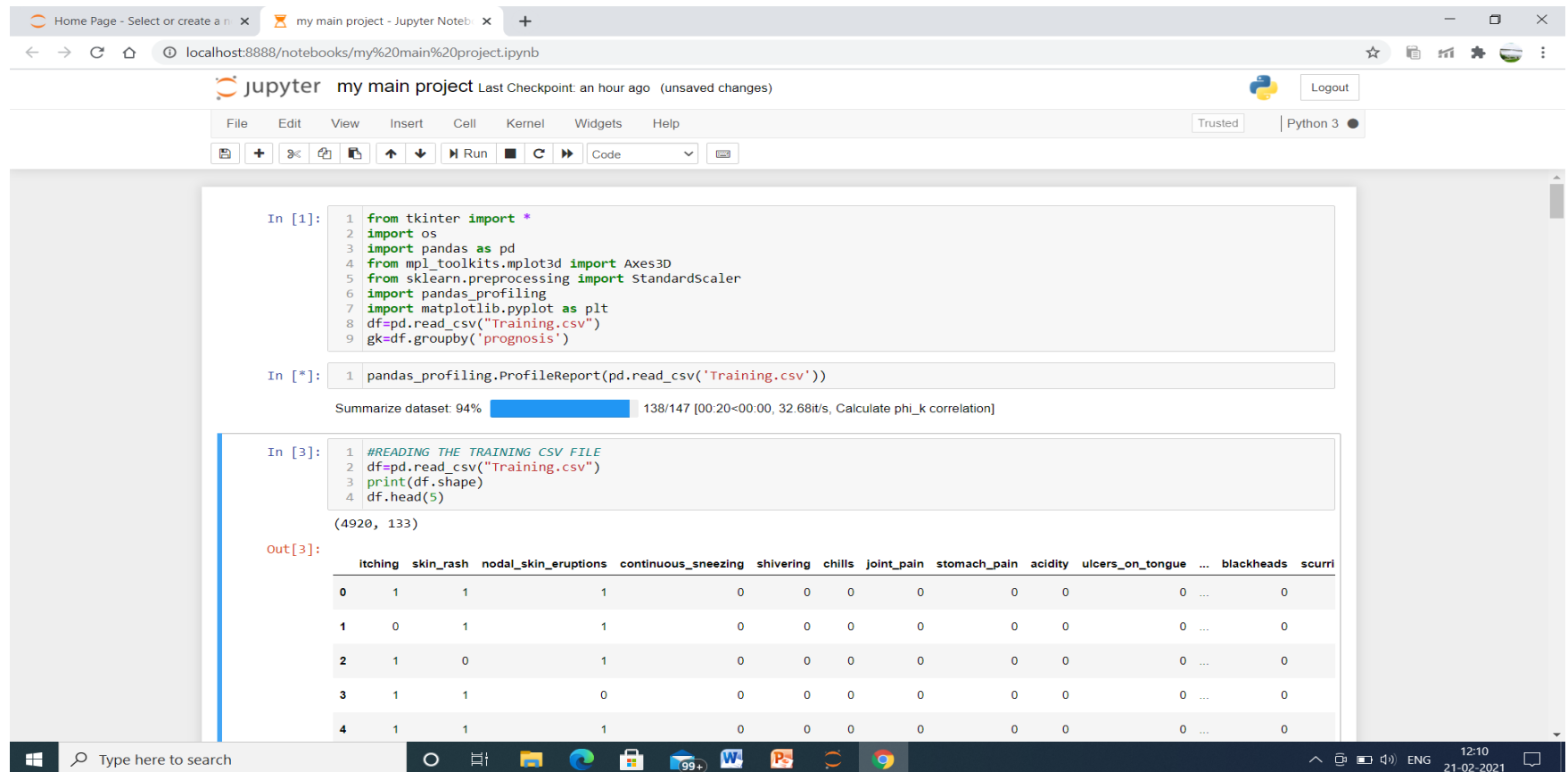
Project contains three parts:

1. DATASET COLLECTION.
2. TRAIN AND TEST THE MODEL.
3. DEPLOY THE MODELS.

- ❑ **Dataset Collection-** We had collected dataset from kaggle notebooks. The dataset contains the symptoms and the corresponding disease. It contains 4920 rows and 133 columns.
- ❑ **Train and Test the model-** We had used the Naïve Bayes Classifier as a model to train the dataset. After training, we had tested the model and found its accuracy.
- ❑ **Deploy the models-** Deployed Naïve bayes by creating interface to get the name, symptoms of an individual. By this, we will get the disease and accuracy of model as the output. We have also created a chatbot using Decision Tree which helps an individual to get the corresponding disease by checking whether he/she is being faced by the symptoms.

Project Snapshots

Importing Libraries



The screenshot shows a Jupyter Notebook interface with the following content:

In [1]:

```
1 from tkinter import *
2 import os
3 import pandas as pd
4 from mpl_toolkits.mplot3d import Axes3D
5 from sklearn.preprocessing import StandardScaler
6 import pandas_profiling
7 import matplotlib.pyplot as plt
8 df=pd.read_csv("Training.csv")
9 gk=df.groupby('prognosis')
```

In [*]:

```
1 pandas_profiling.ProfileReport(pd.read_csv('Training.csv'))
```

Summarize dataset: 94% ██████████ 138/147 [00:20<00:00, 32.68it/s, Calculate phi_k correlation]

In [3]:

```
1 #READING THE TRAINING CSV FILE
2 df=pd.read_csv("Training.csv")
3 print(df.shape)
4 df.head(5)
```

Out[3]:

```
(4920, 133)
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	blackheads	scurri
0	1	1	1	0	0	0	0	0	0	0	0 ...	0	
1	0	1	1	0	0	0	0	0	0	0	0 ...	0	
2	1	0	1	0	0	0	0	0	0	0	0 ...	0	
3	1	1	0	0	0	0	0	0	0	0	0 ...	0	
4	1	1	1	0	0	0	0	0	0	0	0 ...	0	

Project Snapshots

Dataset Visualisation

Home Page - Select or create a notebook | disease prediction second review python3 | Last Checkpoint: 6 hours ago (autosaved) | Logout

File Edit View Insert Cell Kernel Widgets Help | Trusted | Python 3

Pandas Profiling Report | Overview Variables Correlations Missing values Sample Duplicate rows

Overview

Overview Warnings 64 Reproduction

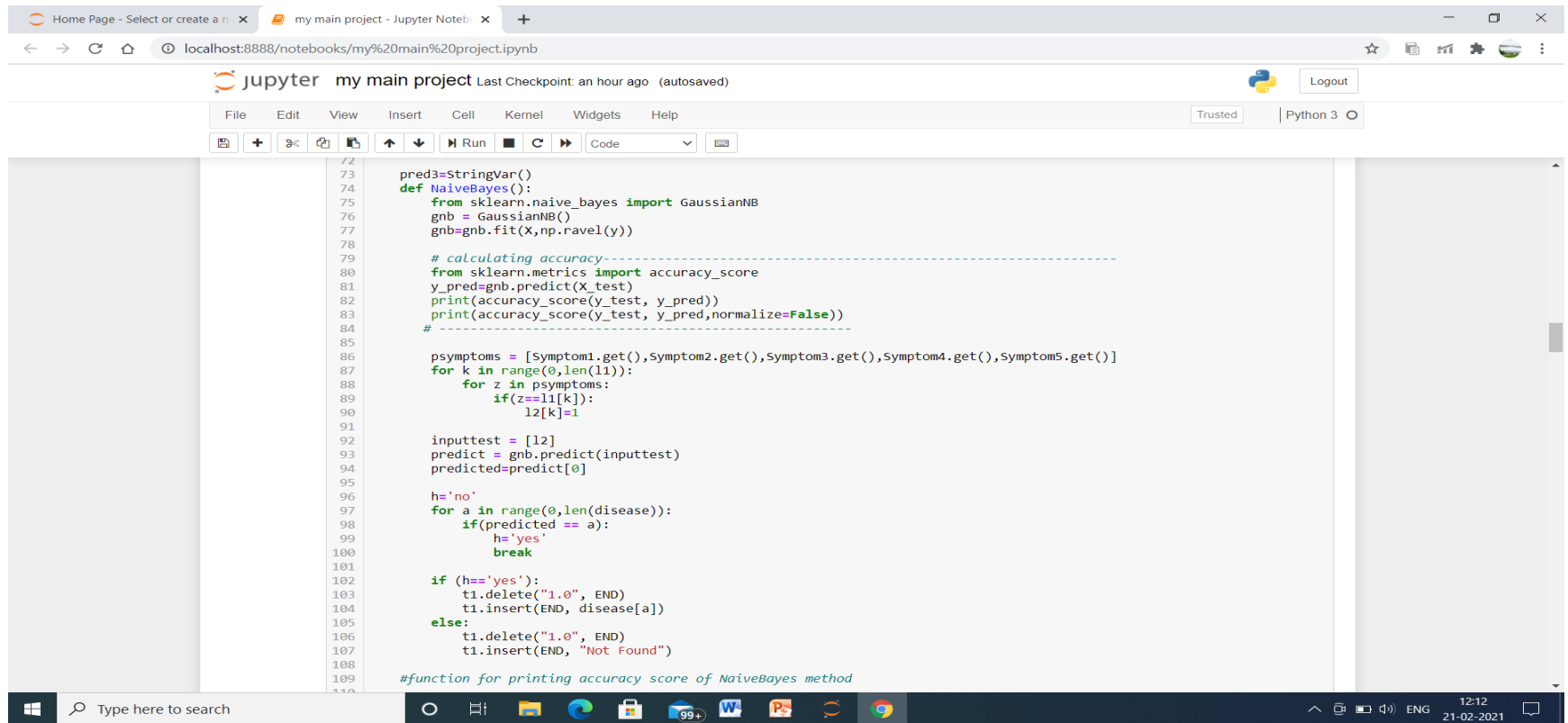
Dataset statistics	
Number of variables	133
Number of observations	4920
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	4616
Duplicate rows (%)	93.8%
Total size in memory	5.0 MIB
Average record size in memory	1.0 KIB

Variable types	
BOOL	132
CAT	1

Windows taskbar: Type here to search | 13:43 30-01-2021

Project Snapshots

Model Training and Testing

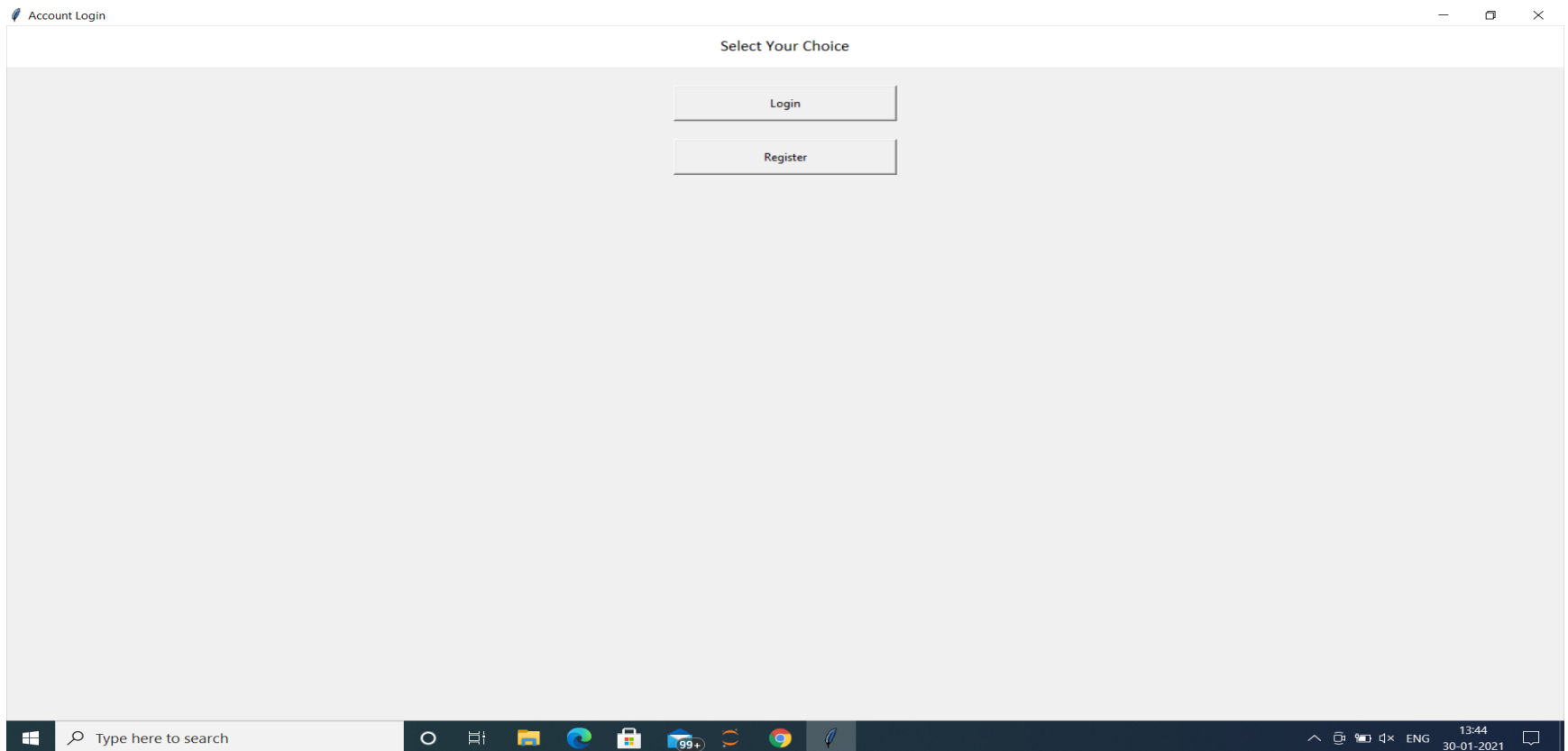


The screenshot displays a Jupyter Notebook interface in a web browser. The browser's address bar shows the URL `localhost:8888/notebooks/my%20main%20project.ipynb`. The Jupyter interface includes a top bar with the project name "my main project" and a "Logout" button. Below this is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. A toolbar contains icons for file operations, running cells, and code execution. The main area shows a Python script for a Naive Bayes model. The script defines a `NaiveBayes` class, imports `GaussianNB` from `sklearn.naive_bayes`, and uses `accuracy_score` from `sklearn.metrics`. It includes a loop to process symptoms and a function to print the accuracy score.

```
72
73 pred3=StringVar()
74 def NaiveBayes():
75     from sklearn.naive_bayes import GaussianNB
76     gnb = GaussianNB()
77     gnb=gnb.fit(X,np.ravel(y))
78
79     # calculating accuracy-----
80     from sklearn.metrics import accuracy_score
81     y_pred=gnb.predict(X_test)
82     print(accuracy_score(y_test, y_pred))
83     print(accuracy_score(y_test, y_pred,normalize=False))
84     # -----
85
86     psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
87     for k in range(0,len(l1)):
88         for z in psymptoms:
89             if(z==l1[k]):
90                 l2[k]=1
91
92     inputtest = [l2]
93     predict = gnb.predict(inputtest)
94     predicted=predict[0]
95
96     h='no'
97     for a in range(0,len(disease)):
98         if(predicted == a):
99             h='yes'
100             break
101
102     if (h=='yes'):
103         t1.delete("1.0", END)
104         t1.insert(END, disease[a])
105     else:
106         t1.delete("1.0", END)
107         t1.insert(END, "Not Found")
108
109     #function for printing accuracy score of NaiveBayes method
110
```

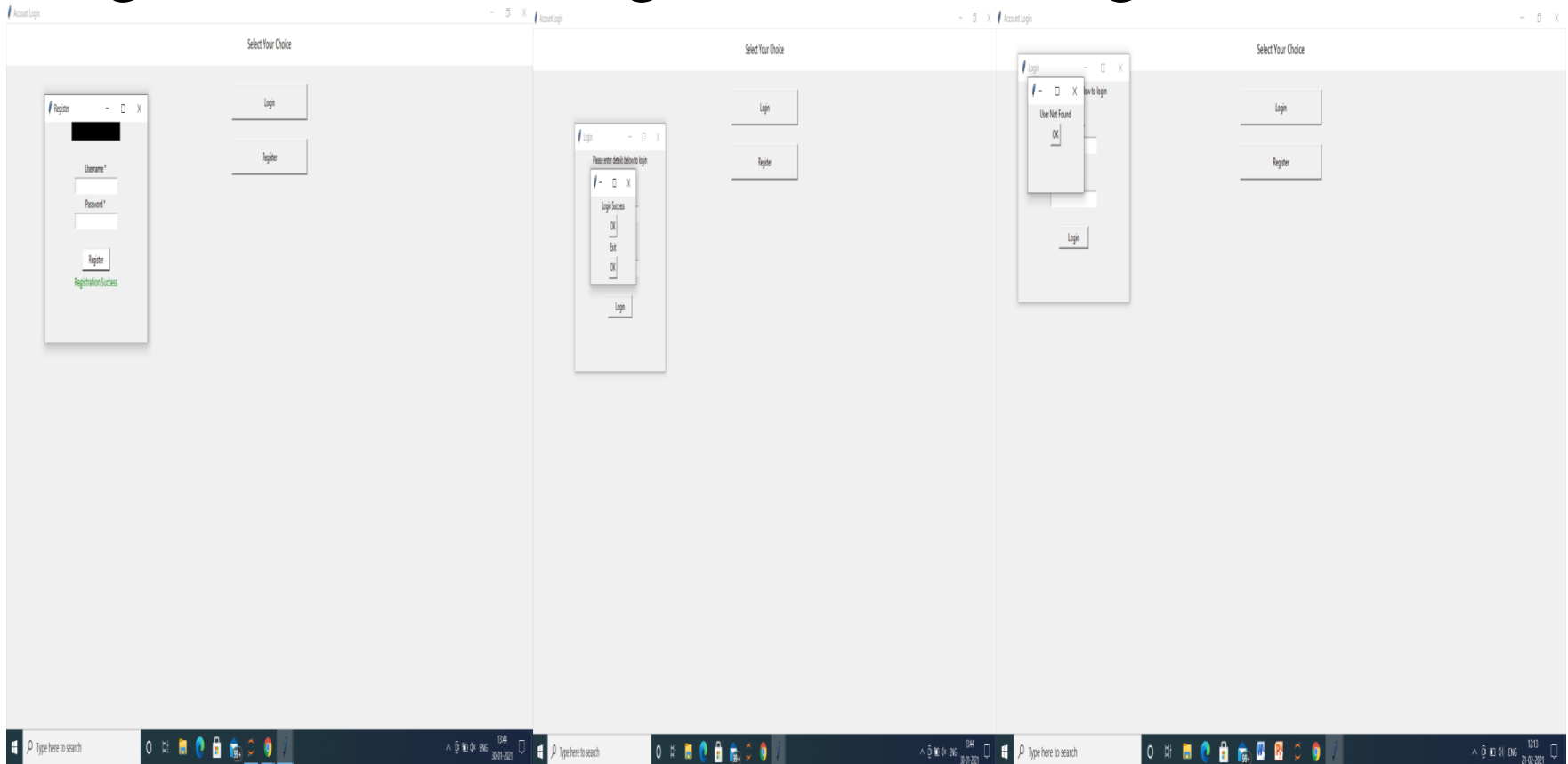
Project Snapshots

Account Registration Screen



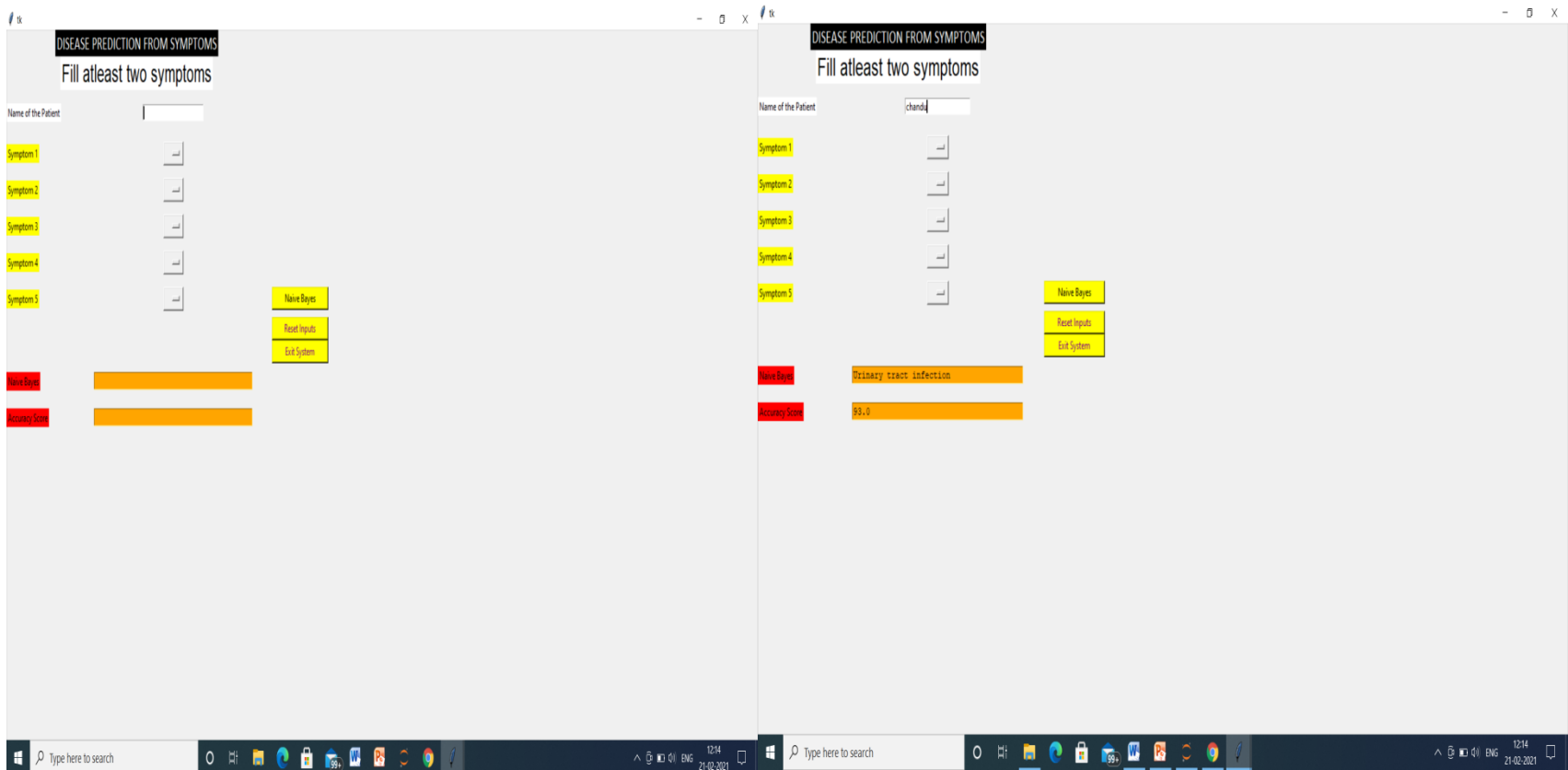
Project Snapshots

Registration and Login Success, Login invalid



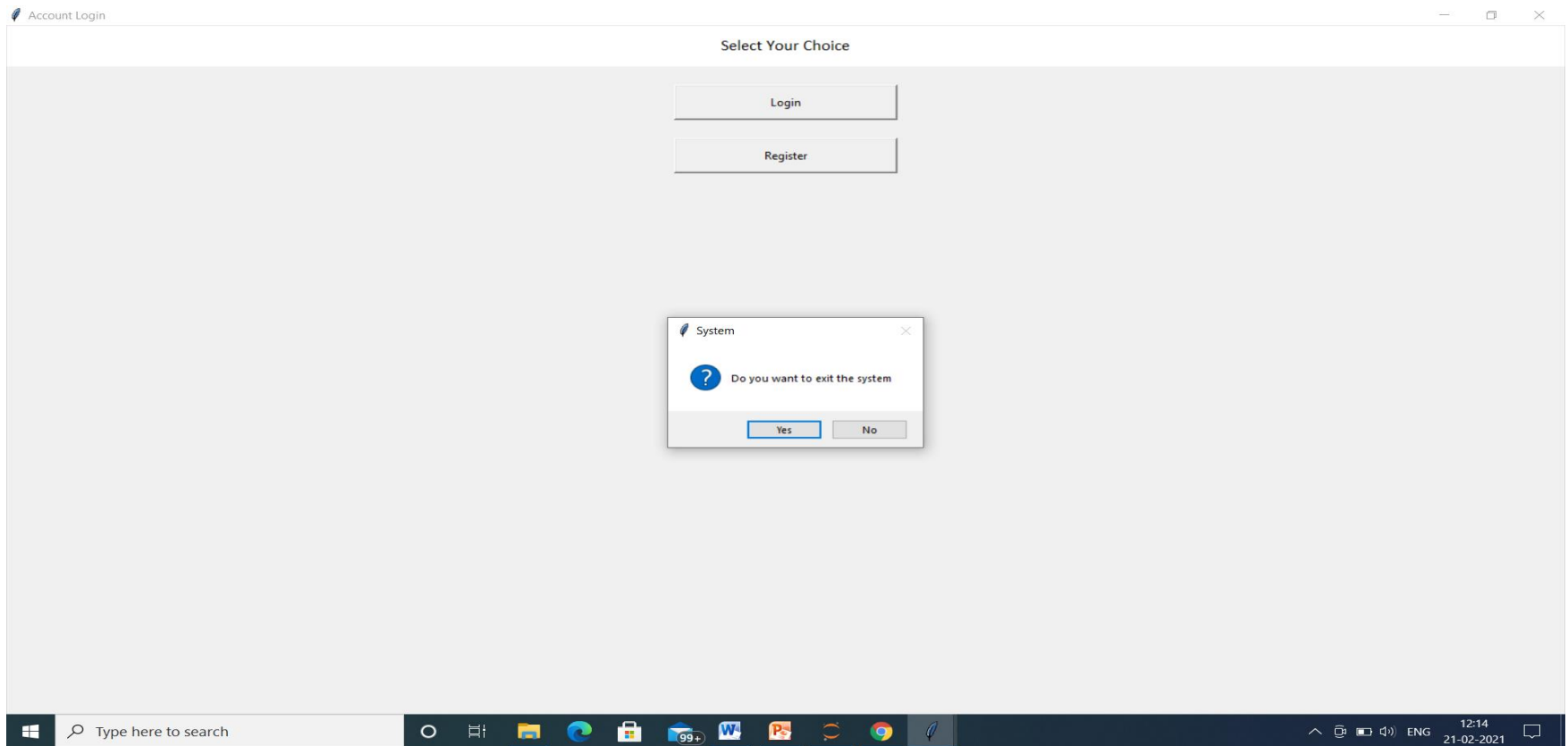
Project Snapshots

GUI Interface with and without input



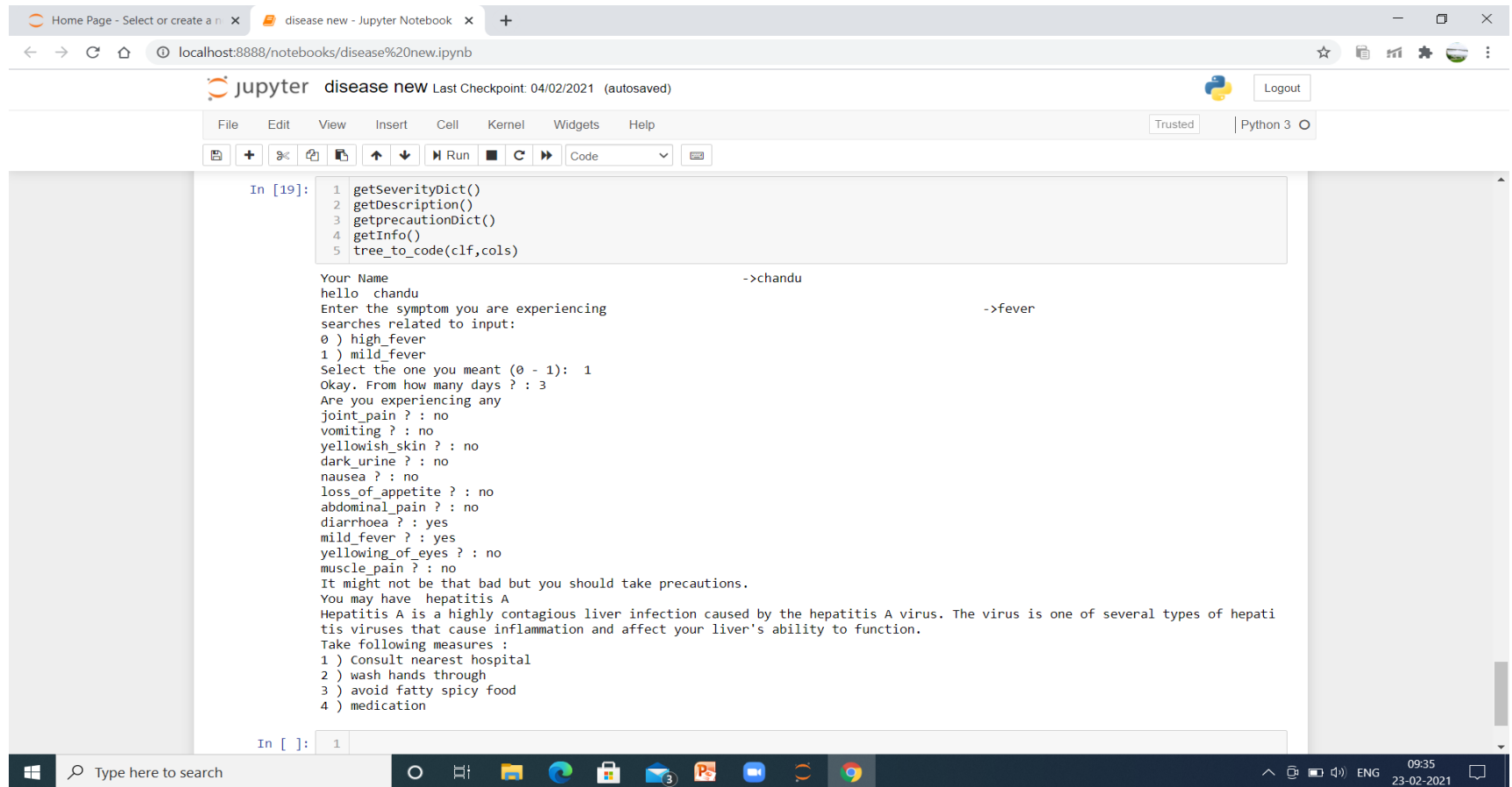
Project Snapshots

Exit System



Project Snapshots

Chatbot



The screenshot displays a Jupyter Notebook environment with the following components:

- Browser Tabs:** 'Home Page - Select or create a n...' and 'disease new - Jupyter Notebook'.
- Address Bar:** 'localhost:8888/notebooks/disease%20new.ipynb'.
- Page Header:** 'jupyter disease new Last Checkpoint: 04/02/2021 (autosaved)' and a 'Logout' button.
- Menu Bar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help.
- Toolbar:** Includes buttons for file operations, running, and a dropdown menu set to 'Code'.
- Code Cell:**

```
In [19]: 1 getSeverityDict()
2 getDescription()
3 getprecautionDict()
4 getInfo()
5 tree_to_code(clf,cols)
```
- Output:**

Your Name ->chandu
hello chandu
Enter the symptom you are experiencing ->fever
searches related to input:
0) high_fever
1) mild_fever
Select the one you meant (0 - 1): 1
Okay. From how many days ? : 3
Are you experiencing any
joint_pain ? : no
vomiting ? : no
yellowish_skin ? : no
dark_urine ? : no
nausea ? : no
loss_of_appetite ? : no
abdominal_pain ? : no
diarrhoea ? : yes
mild_fever ? : yes
yellowing_of_eyes ? : no
muscle_pain ? : no
It might not be that bad but you should take precautions.
You may have hepatitis A
Hepatitis A is a highly contagious liver infection caused by the hepatitis A virus. The virus is one of several types of hepatitis viruses that cause inflammation and affect your liver's ability to function.
Take following measures :
1) Consult nearest hospital
2) wash hands through
3) avoid fatty spicy food
4) medication
- Bottom Bar:** Windows taskbar with search, task view, and application icons, and a system tray showing '09:35 23-02-2021'.

Results and Discussion

- ❑ Doctors and medical professionals are always required in case of an emergency.
- ❑ The project presented the technique of predicting the disease based on the symptoms of an individual patient.
- ❑ In the current situation of COVID-19, where sufficient facilities and resources are unavailable, our prediction system can prove to be helpful and can be used in the diagnosis of a disease.
- ❑ This project would help in lowering the cost required in dealing with the disease and would also make the recovery process easy.

Conclusion

- ❑ The project presented the technique of predicting the disease based on the symptoms of an individual patient.
- ❑ Almost all the ML models gave good accuracy values but the most accurate one is selected and the disease given by it is considered as the disease of an individual.
- ❑ Once the disease is predicted, we could easily manage the medicine resources required for the treatment.
- ❑ Doctors and medical professionals are always required in case of an emergency. In the current situation of COVID-19, where sufficient facilities and resources are unavailable, our prediction system can prove to be helpful and can be used in the diagnosis of a disease.
- ❑ This project would help in lowering the cost required in dealing with the disease and would also make the recovery process easy.

References

- [1] Jaymin Patel, Prof. Tejal Upadhyay, Dr. Samir Patel “Heart disease prediction using Machine learning and Data Mining Technique” March 2016.
- [2] Disease Prediction Using Machine Learning Akash C. Jamgade, Prof. S. D. Zade (5 May, 2019).
- [3] Disease prediction from various symptoms using machine learning journal by Rinkal Keniya posted on 8 October, 2020.
- [4] Disease Prediction Using Machine Learning Over Big Data” Vinitha S, Sweetlin S, Vinusha H and Sajini S (2018).

References

- [5] Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning Reddy Prasad, Pidaparthi Anjali, S. Adil, N. Deepa (2019) .
- [6] Heart Disease Prediction Using Data Mining Techniques H. Benjamin Fredrick David and S. Antony Belcy (2018).
- [7] Multi Disease Prediction Using Data Mining Techniques”K. Gomathi , Dr. D. Shanmuga Priyaa (2017).
- [8] Prediction of Cardiovascular Disease using Machine Learning Algorithms Muktevi Srivenkatesh (3 Feb, 2020).

THANK YOU