VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama" Belagavi - 590018

A PROJECT REPORT ON

"CORONAVIRUS DETECTION USING MACHINE LEARNING"

Submitted in the partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

INFORMATION SCIENCE & ENGINEERING

Submitted By

CHANDANA P(1RR18IS009)

DEEPASHREE S RANNORE(1RR18IS011)

MAHESHWARI B(1RR18IS023)

Under the Guidance of

Dr. R. BALAKRISHNA

PROFESSOR AND DEAN – ISE, RRCE

Dept of ISE, RRCE

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

RAJARAJESWARI COLLEGE OF ENGINEERING

MYSORE ROAD, BANGALORE-560074

(An ISO 9001:2008 Certified Institute)

(2021-22)

RAJARAJESWARI COLLEGE OF ENGINEERING

MYSORE ROAD, BANGALORE-560074

(An ISO 9001:2008 Certified Institute)

(Affiliated to Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

CERTIFICATE

Certified that project work entitled

"CORONA VIRUS DETECTION USING MACHINE LEARNING"

Carried out by

CHANDANA P(1RR18IS009)

DEEPASHREE S RANNORE(1RR18IS011)

MAHESHWARI B(1RR18IS023)

The students of **"RajaRajeswari College of Engineering"** in partial fulfillment for the award of **Bachelor of Engineering in Information Science & Engineering** of the Visvesvaraya Technological University, Belagavi during the year **2021–2022.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Eighth semester.

…….....	……	……
Signature of guide	**Signature of HOD**	**Signature of Principal**
Dr. R. BALAKRISHNA Prof. &Dean Dept. of ISE RRCE, Bengaluru	**Dr. AMUTHARAJ. J** Prof. &HOD Dept. of ISE, RRCE, Bengaluru	**Dr. T. CHANDRASHEKAR** Principal, RRCE, Bengaluru

External Viva-Voce

Examiners: ***Signature:***

1. _____

2. _____

ABSTRACT

The Coronavirus Disease originated in Wuhan City, Hubei Province, China on December 31, 2019, with pneumonia of unknown variant, but drastically evolved into a pandemic. This disease was named as COVID-19 and the virus is called SARS CoV2. Most corona viruses affect the animals, but because of their zoonotic nature, it is possible that, it can also be transmitted to people. Severe Acute Respiratory Syndrome Corona virus (SARSCoV) and Severe Acute Respiratory Syndrome Corona virus (MERSCoV) are the main causes of severe respiratory disease and death in humans. Asymptotic patients are the vital source of spreading infection.

The method currently used for corona virus (COVID19) detecting is RTPCR. Chest X-rays, plays a crucial role in the early diagnosis and treatment of Covid-19 disease. Various AI techniques have been applied successfully in predicting the virus. Machine learning based models are implemented to identify the possible threat of Covid-19 all over the world. The researchers had made a statement that the combination of scientific photographic features and laboratory effects will help in early detection of the corona virus (COVID19).

ACKNOWLEDGEMENT

We would like to take this opportunity to thank a lot of eminent personalities, without whose constant encouragement, this endeavor of ours would not have become a reality. I consider myself privileged to express gratitude and respect towards all those who guided me throughout my work.

Firstly, we would like to thank the **VTU Belguam**, for having this project as part of its curriculum, which gave me the wonderful opportunity to work on my research and presentation abilities and **RAJARAJESWARI COLLEGE OF ENGINEERING** for providing me with such excellent facilities, without which, this project could not have acquired the shape, it has now done.

We would like to express our gratitude to, **Dr. T. Chandrashekar, Principal, RAJARAJESWARI COLLEGE OF ENGINEERING** for providing me a congenial environment and surroundings to work on.

We are very proud and thankful **to Dr. J. Amutharaj Professor and Head of the Department of Information Science and Engineering, RAJARAJESWARI COLLEGE OF ENGINEERING** for his valuable advice and ideas at various stage of this work.

We would like to express my sincere thanks to my Project **Coordinator Dr. R. Balakrishna Prof & Dean of Department of Information Science and Engineering, RAJARAJESWARI COLLEGE OF ENGINEERING** for his valuable knowledge on the subject and guidance on the completion of the project.

We extend our sincere thanks to the teaching and non-teaching staffs for their cooperation, all the kindness forwarded to me in carrying out this project work in college.

CHANDANA P(1RR18IS009)
DEEPASHREE S RANNORE(1RR18IS011)
MAHESHWARI B(1RR18IS023)

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Coronavirus disease which is called as the COVID-19 is a highly contagious viral illness which is caused by severe acute respiratory syndrome SARS-CoV-2. Coronavirus are considered to be the family of viruses that cause illness such as respiratory diseases or gastrointestinal diseases and can also cause death. The first cases of this predominantly respiratory viral illness were first reported in Wuhan, Hubei Province, China, in late December 2019. Thereafter when this spread across the world in a short span of time, compelled the World Health Organization (WHO) to declare the outbreak as a global pandemic on the March 11, 2020. Globally as of December 2021, there have been 280,119,931 confirmed cases of COVID-19, including 5,403,662 deaths, reported by WHO.

The COVID-19 is broadly spread by the dust particles and fomites while close unsafe touch between an infector and the infected person happens. The virus travels in respiratory droplets released into air when an infected person coughs, sneezes, talks or breathes near you and a person can be affected by inhaling these droplets. The Airborne distribution for COVID-19 has not been recorded as it is not known to be a significant transmission. Fast spreading has been seen in certain patients, and the active virus has been reported in a small number of clinical studies. Symptoms of COVID-19 are variable, often include fever, cough, headache, difficulty in breathing, loss of smell and taste and few more. By wearing masks in public, keeping distance from others and by often sanitizing themselves, we can reduce the chances of infection being spread. The WHO continues to encourage each and every individual to take care of their own health and also protect others by following all the guidelines generated.

1.1 EXISTING SYSTEM

Using existing systems to detect COVID19. Prediction of the symptoms of COVID is done by using various machine learning models such as Naïve Bayes Classifier, Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RT).

DRAWBACKS:

- The main challenge faced was, the requirement for the detection of COVID-19 cases with a standard symptom is not verified.
- The main drawback was limited number of samples, and it's not adequate to train and test the model with better accuracy.
- Some of the limitations are susceptible to selection bias among patients, and patients with very slight symptoms are excepted from the study.
- Failed to analyze the high dimensional datasets.
- It was found that the Polynomial Regression model produced the best accuracy of 90%. Still 90% accuracy is not sufficient for medical data analysis.

1.2 PROPOSED SYSTEM

Now-a-days the raising covid cases and its complexity in detecting it is posing a greater pressure on the health department with several thousands of samples for testing and timely report generation for them.

This problem can be automated using our proposed system with an efficient algorithm such as K Nearest Neighbors. Here we use [UI] user interface and take the input from users which are basically the symptoms they are facing (which may or may not be covid).

Later we clean the acquired data using some of the libraries from pandas and load these data from processing in the form of datasets to machine learning models. These datasets help to visualize the data clearly and process it accurately. After processing the predicted output is displayed.

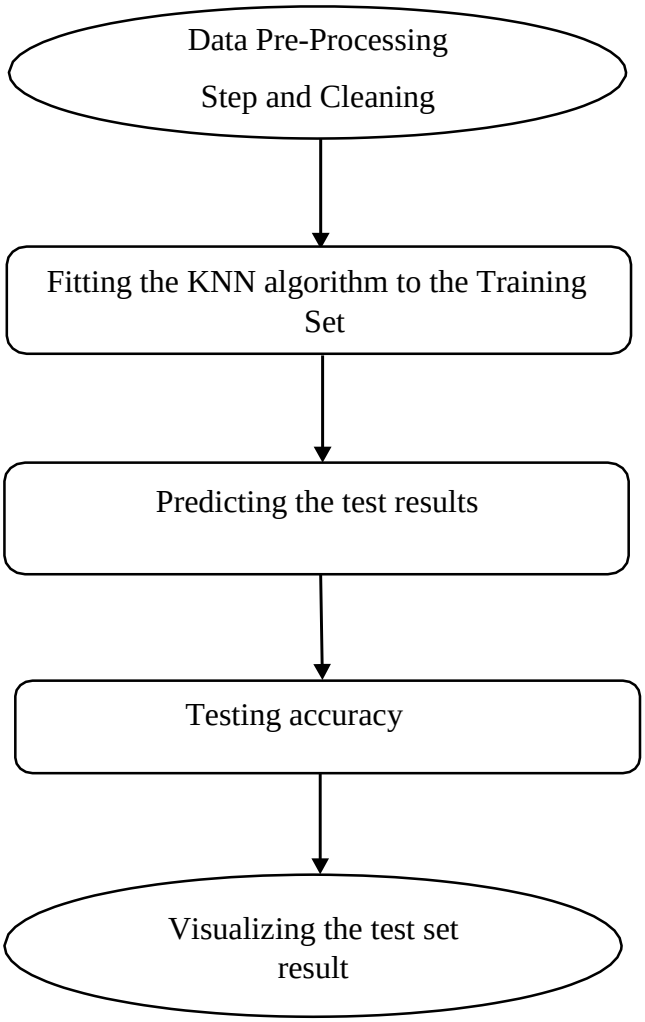


Figure 1.2.1: Proposed KNN System

ADVANTAGES:

- The standard symptoms of covid-19 detection cases are **verified effectively**.
- Increasing the **prediction accuracy to 90%+ efficiency** using K Nearest Neighbors Machine Learning technique.
- Based on the we can easily **detect multi-class prediction**.
- Accurate output prediction will be **useful and fruit-full for humanity** and to detect the virus in the early stage and safe lot of lives.
- K Nearest Neighbors performs **better compared** to other models like logistic regression and you need **less training data.**

1.3 MOTIVATION OF THE PROJECT

Machine Learning Techniques are very useful in tracing the COVID-19 cases by predicting the cases and diagnosing it with proper medications, developing alerting systems for generating social distance and few others for controlling the spread of this virus. There are few mechanisms that available to analyze and detect COVID-19 but give the result **with approximately 90% accuracy rate**. Even both the **Reverse Transcription Polymerase Chain Reaction (RT-PCR) and CT scan** for the COVID-19 identification are defective and **are found to have faults**.

Therefore, the **main motive** or the objective of this project is to **effectively improve the prediction of the coronavirus disease** with the **combination of various methods** and by designing a mechanism with the help of the **Machine Learning algorithms** such as the **K Nearest Neighbors**, where it detects whether the individual is suffering with COVID-19 or not at early stages itself and helping in predicting the severity of the COVID-19 positive patients as well.

1.4 OBJECTIVE OF THE PROJECT

The main objectives of the project are:

- The main objective of the project is detecting the COVID-19 symptoms in patients using **K Nearest Neighbors.**
- Increasing the prediction accuracy to **90%** efficiency using **K Nearest Neighbors ML technique.**
- The standard symptoms of covid-19 detection cases are verified effectively. Based on the value of we can easily detect multi-class prediction.

CHAPTER 2

LITERATURE SURVEY

Sl no.	Author's Name	Title of the Paper	Journal Name	Year of Publication	Advantages	Drawbacks
1	Rohini M, Naveena K R, Jothipriya G, Kameshwaran S, Jagadeeswari M	A Comparative Approach to predict Corona Virus using Machine Learning. **"Base Paper"**	International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)	2021	This work helped in the study of COVID-19 outbreak that has utilized many Machine Learning classification results that visualizing the origin of this disease and by performing predictions which helped to control the impact of this disease in future.	Could have provided better results by using several other Machine Learning models and techniques for finding estimates that helped the clinicians, medical and Governmental Organizations to look forward for real-time preparations.
2	Sugandh Bhatia, Jyoteesh Malhotra	Naïve Bayes Classifier for predicting the Novel Coronavirus. .	Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021)	2021	The objective was to design a mechanism for detecting and predicting the COVID-19 disease using Naïve Bayes classifier to decide and detect whether a person is suffering with COVID-19 or not.	It wasn't possible to find out both the cases of coronavirus, which are with the symptomatic or asymptotic as it is the only way to break the chain in order to control the spread of coronavirus.

Sl. no	Author's Name	Title of the Paper	Journal Name	Year of Publication	Advantages	Drawbacks
3	Chekuri Sri Sumanth, Rudra Kalyan Nayak	Machine Intelligent Techniques for COVID-19 Detection: A Critical Review and Analysis.	12th International Conference on Computing Communication and Networking Technologies (ICCCNT)	2021	The survey here is on the COVID 19 diagnosis with a chain of CT scan pictures obtained from the COVID-19 data sets by using ML algorithms like marine predator, simplified suspected infected recovered (SIR), image acquisition, and few other techniques of ML.	This work would have been very much useful if it was embedded with two or more models with proper compatibility to the dataset to have gotten even more better results while detecting COVID 19. This will help the human race too.
4	Hina Gull, Gomathi Krishna, May Issa Aldossary, Sardar Zafar Iqbal	Severity Prediction of COVID-19 Patients using Machine Learning Classification.	6th International Conference on Computer and Communications (ICCC)	2020	Support Vector Machine (SVM) was identified the better classification algorithm in order to predict the severity of the disease. The early detection of the COVID patients, support the concerned authorities to take better action in prior. And also this helped to isolate the affected people as early as possible.	The main challenge faced in this study was, the basic requirement for the detection of coronavirus cases with a standard symptom was not verified as these cases have been difficult to obtain in the present situation

Sl no.	Author's Name	Title of the Paper	Journal Name	Year of Publication	Advantages	Drawbacks
5	Akshay Kumar Siddhu, Dr.Ashok Kumar and Dr.Shakti Kundu	Detection of COVID-19 from Medical images and/or Symptoms of Patient using Machine Learning Approaches.	9th International Conference System Modeling and Advancement in Research Trends (SMART)	2020	It is mainly focused to select the best Deep Learning models in order to detect the segment of the lungs and to predict the COVID-19 patients using the Deep Learning techniques.	If in case there was the availability of more public databases, then better Deep Learning models could have been developed to detect and predict the COVID-19 accurately, thus helping to develop the best performing model.

CHAPTER 3

SYSTEM REQUIREMENTS AND SPECIFICATION

3.1 SYSTEM ANALYSIS

System analysis is nothing but a process which is responsible for grouping, interpreting and collecting the facts, identifying and solving the problems and further decomposing the system into its components. The main aim of the system analysis is to get to know about the system and its parts for the purpose of identifying its objectives and it is a technique which helps the system to work efficiently in order to accomplish their purpose. The main focus is basically on **Systems, Processes** and **Technology.**

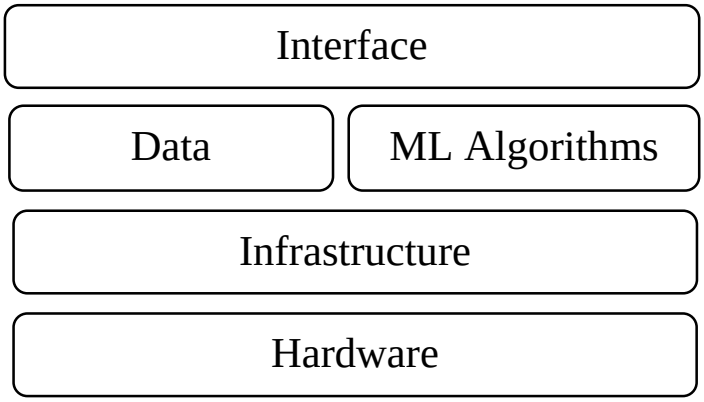


Figure 3.1.1: Schematic Diagram for ML Systems

3.2 FUNCTIONAL REQUIREMENTS

In order to enable the success of a system, the requirement analysis is a very important process. The functional requirements are those where there are demands put on by the end user based on the basic facilities that the system should provide or offer. This is a type of requirement where one can see the **final product directly**.

- **Software Interface**: An interactive UI is required for interaction between user and system designed.
- **Software Interface**: Python packages required for Modeling, Interfacing and Data extraction is needed.

3.3 NON-FUNCTIONAL REQUIREMENTS

A non-functional requirement is that where the quality attributes of the system must be satisfied in accordance with the project contract. The attributes like **usability** and **effectiveness** of the system entirely is essential in a non-functional requirement. There can be possibility where the satisfaction of the user needs fails when the system fails to meet non-functional requirements.

The non-functional requirement mainly deals with the issues like **flexibility, reliability, scalability, security** etc. It helps in the verification of performance of the software. For example, when the number of simultaneous users are greater than 10000, the site should load in 3 seconds.

- **Efficiency:** The Machine Learning algorithm prediction ability must be greater than 90% for accurate results.
- **Portability:** The system is designed to work on a computer system with low processing capacities and also independent OS.
- **Scalability:** The current system will be system dependent and hence any cloud platform can be used to scale up to any device by hosting the same system.
- **Robustness:** The data applied to the Machine Learning algorithm should train the model for all data vulnerabilities.

3.4 TOOLS AND TECHNILOGY REQUIRED

➢ HARDWARE REQUIREMENTS:

- Processor : AMD Ryzen 3 or more
- Random Access Memory(RAM) : 512 MB or more
- Processor Speed : 250Mhz to 830Mhz
- Hard Disk : 1GB Recommended

➢ SOFTWARE REQUIREMENTS:

- Operating system : Windows 10 or above
- Editor : Python 3.7 and above interpreter and compiler
- Python Packages : Tkinter, Scikit Learn, Pandas, Numpy and Matpotlib

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE:

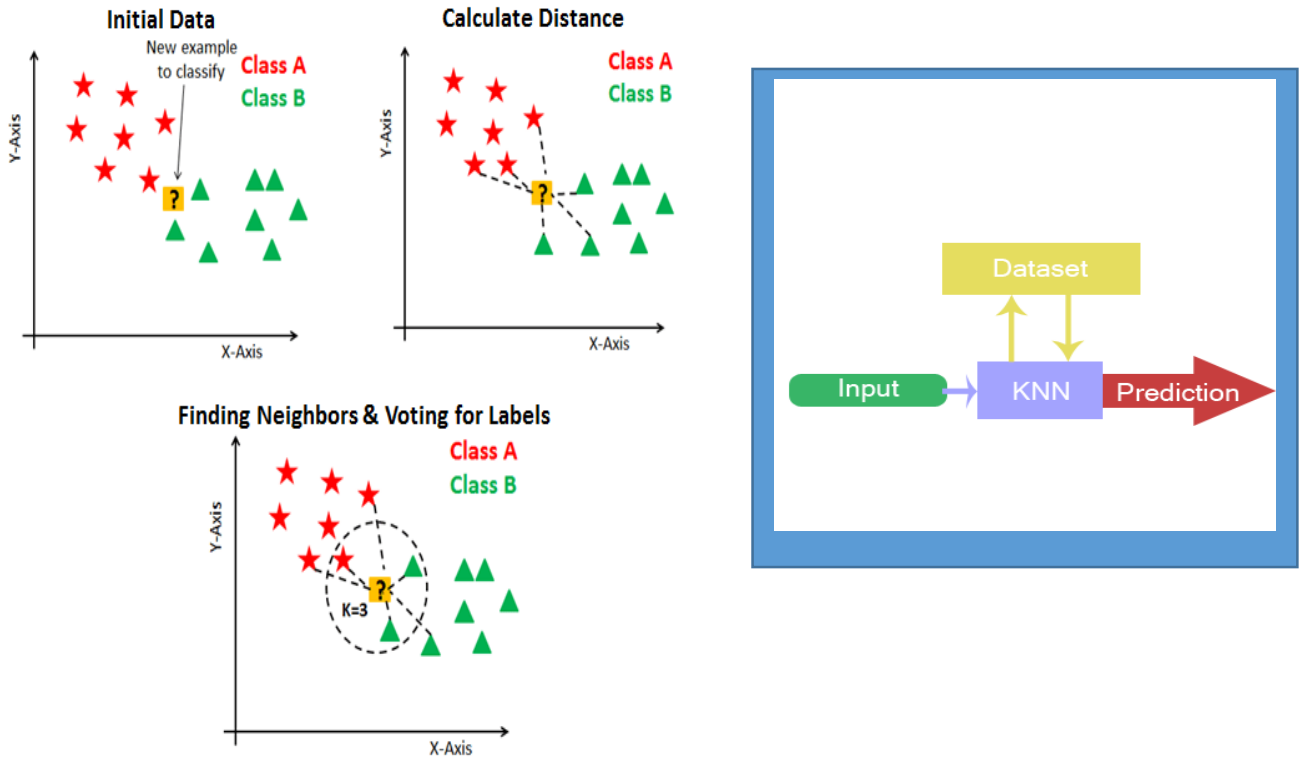


Figure 4.1.1: System Architecture

4.2 INPUT/OUTPUT DESIGN:

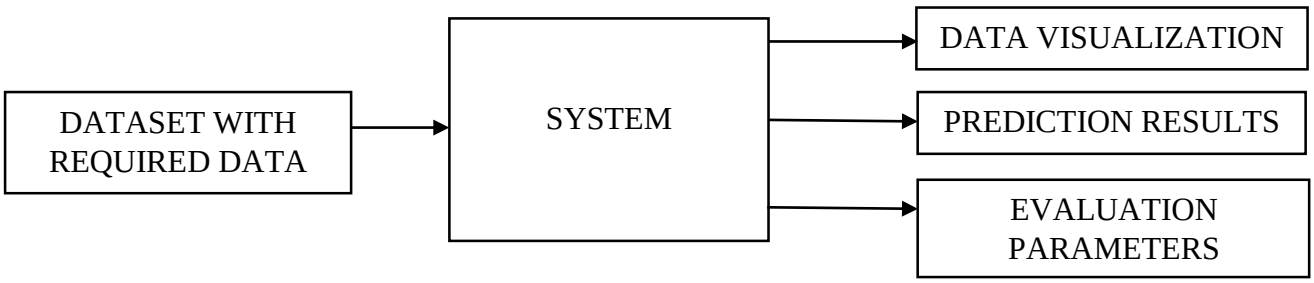


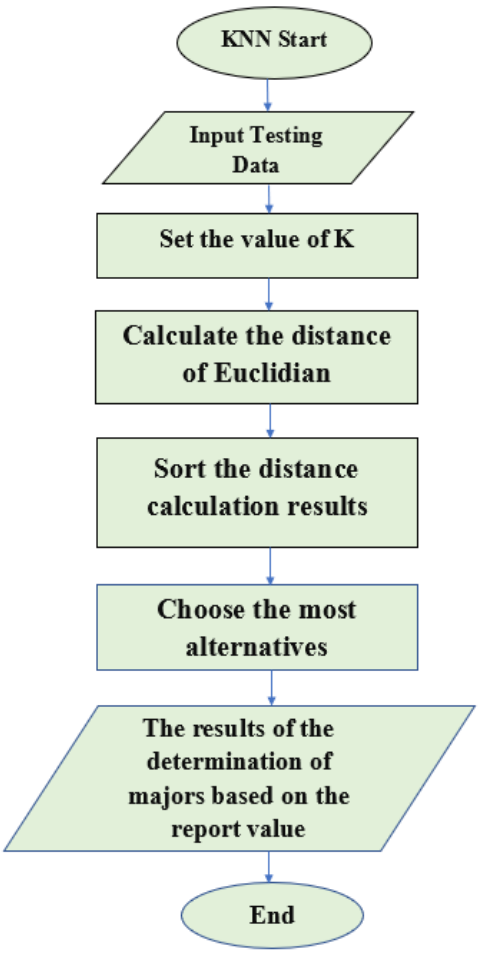
Figure 4.2.1: Input/ Output Design

4.3 ALGORITHM:

ALGORITHM TO BE IMPLEMENTED: K Nearest Neighbors for The Prediction Of COVID- 19

- **Step-1:** Start
- **Step-2:** Select the number K of the neighbor.
- **Step-3:** Calculate the Euclidean distance of **K number of neighbors.**
- **Step-4:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-5:** Among these k neighbors, count the number of the data points in each category.
- **Step-6:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-7:** End

FLOW CHART:



$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i\right)^2}$$

Figure 4.3.1: KNN Flow Chart

CHAPTER 5

SYSTEM IMPLEMENTATION

5.1 MODULE DESCRIPTION:

- **DATA ACQUISITION AND CLEANING:**

Data Acquisition means collecting the data from the relevant sources before it can be stored. Data Cleaning refers to identification and correction or the errors in the datasets that may negatively or wrongly impact a predictive model.

- **DATA VISUALIZATION:**

Data Visualization means graphically representing the data and information in a graphical or pictorial formats like maps, charts, graphs. Data Visualization enables the users to see and understand the pattern and trends in data and outliers.

- **DATA MODELLING:**

The data that are cleaned and processed has to be trained to recognize certain types of patterns. Here we train a model over datasets by, providing it an algorithm so that it can use to reason over and learn from those datasets.

- **TESTING:**

Testing is the process where the performance of a fully trained model is evaluated. The tested data provides a final and real-world check of an unseen data to confirm that the Machine Learning algorithm was trained productively.

- **COMPARISON AND MEASUREMENT:**

Here the objective is to compare the result of the testing with other Machine Learning algorithms and choose the best algorithm that suits the data. Hence, the performance of the algorithm is efficient.

1) LR:

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1.** Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems.**

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature. The independent variable should not have multi-collinearity

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

- The above equation is the final equation for Logistic Regression

2) LDA:

Linear Discriminant Analysis or **Normal Discriminant Analysis** or **Discriminant Function Analysis** is a dimensionality reduction technique that is commonly used for supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space. But Linear Discriminant Analysis fails when the mean of the distributions are shared, as it becomes impossible for LDA to find a new axis that makes both the classes linearly separable. In such cases, we use non-linear discriminant analysis.

3) CART:

The CART algorithm is a type of classification algorithm that is required to build a decision tree on the basis of Gini's impurity index. It is a basic machine learning algorithm and provides a wide variety of use cases. A statistician named Leo Breiman coined the phrase to describe Decision Tree algorithms that may be used for classification or regression predictive modeling issues.

CART is an umbrella word that refers to the following types of decision trees:

- **Classification Trees:** When the target variable is continuous, the tree is used to find the "class" into which the target variable is most likely to fall.
- **Regression trees:** These are used to forecast the value of a continuous variable.

4) NAÏVE BAYES CLASSIFIER:

- Basically, Naïve Bayes is a model of conditional probability.
- It is possible to classify an instance of given problem which is denoted by a vector X= (x1, …, xn) where n represents features of independent variables.
- Results were accumulated and when the test data was provided, we receive the probabilities for various classes on the basis of provided symptom details.

5) SVM:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Types of SVM

SVM can be of two types:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.
- The Gini Index, which is used to find the information gained by selecting an attribute. The Gini Index favours larger partitions and is calculated for i = 1 to the n, number of attributes:
- Gini = $1 - \sum(p_i)$ Here, p is the probability that a tuple in D belongs to the class C_i.

6) K NEAREST NEIGHBORS:

K- Nearest Neighbors algorithm is a non- parametric supervised learning method. It is used for classification and regression. In both cases, the input consists of the k closest training examples in a data set.

CHAPTER 6

SYSTEM TESTING

6.1 SYSTEM TESTING:

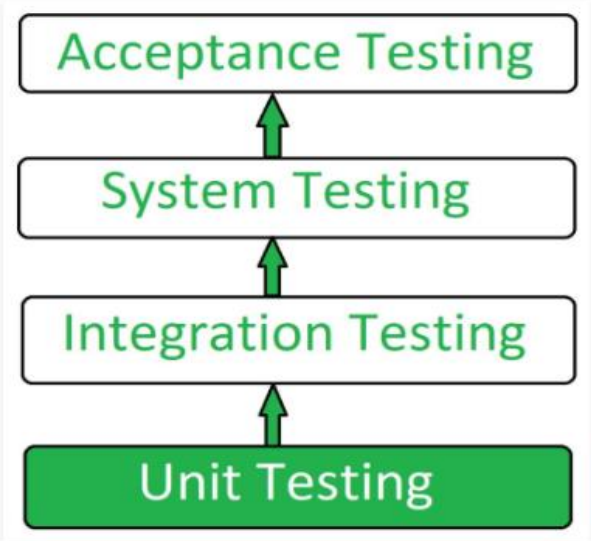


Figure 6.1.1: System Testing

6.2 UNIT TESTING:

Unit Testing is a software testing technique by which individual units of software (i.e., group of computer program modules, usage procedures and operating procedures) are tested to determine whether they are suitable for use or not.

It is a testing method or procedure using which each and every independent module are tested to determine if there are any issue by the developer himself. It is correlated with functional correctness of the independent modules.

The objectives of Unit Testing are:

- To isolate and test a section of code.
- To detect the correctness of code.
- To test all the function and procedure present.
- To fix bug early in development cycle and to reduce expenditure.
- To help the developers to understand the code base and enable them to make changes quickly.
- To facilitate for code reuse.

❖ **Advantages of Unit Testing:**

- Unit Testing allows developers to learn the functionality and gain a basic understanding of the unit API.
- It allows the programmer to refine code and make sure the module works properly.
- It enables to test parts of the project without waiting for others to complete.

6.3 INTEGRATION TESTING

The second level of the software testing process is Integration testing which comes after unit testing. In this, units or individual components of the software are tested in a group. The focus of the integration testing level is to expose defects at the time of interaction between integrated components or units.

➢ **Types of Integration Testing**:

Integration testing can be classified into two parts:

- **Incremental integration testing**
- **Non-incremental integration testing**

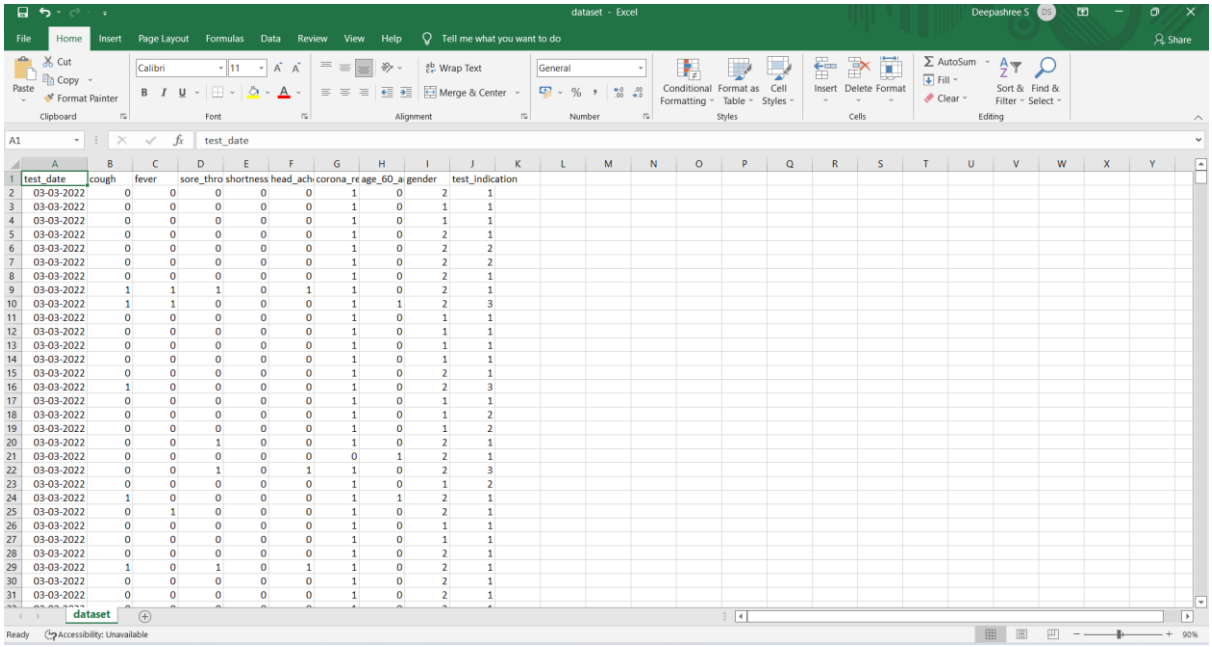
➢ **Advantages:**

Successful integration of modules.

- **Data integrity.**
- **User-based scenarios.**
- **Third-party testing.**

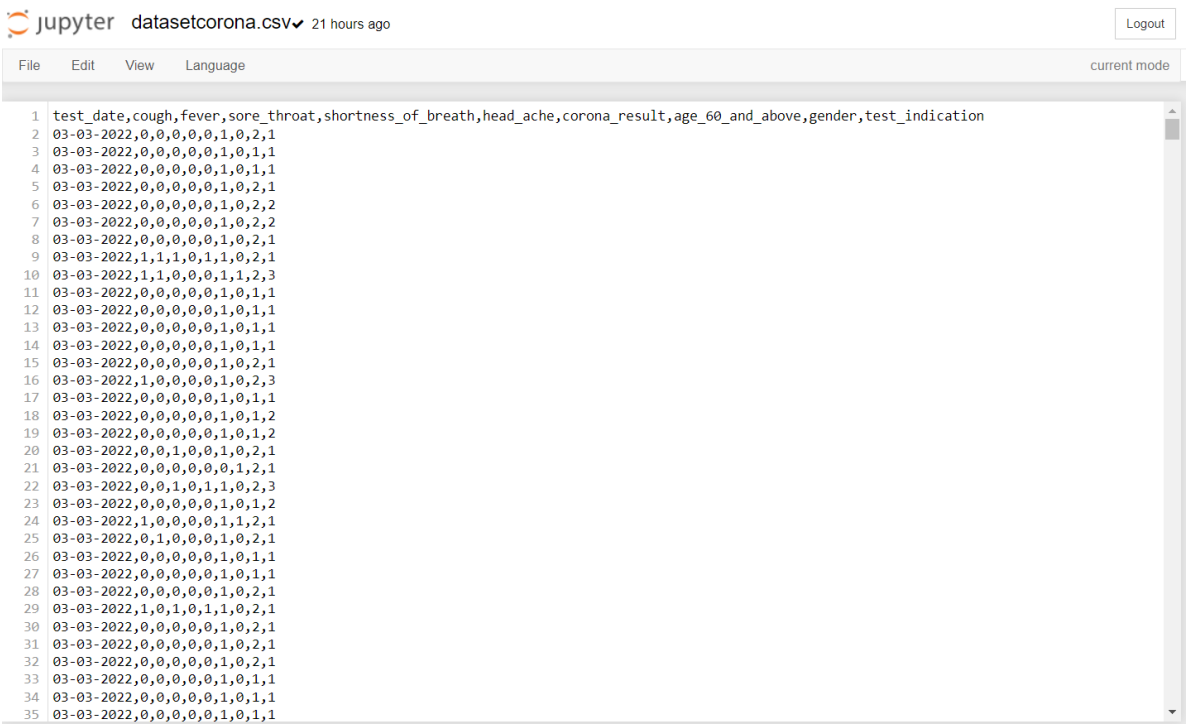
CHAPTER 7

SNAPSHOTS



test_date	cough	fever	sore_throat	shortness_of_breath	head_ache	corona_result	age_60_and_above	gender	test_indication
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	1	0	2	2
03-03-2022	0	0	0	0	0	1	0	2	2
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	1	1	1	0	1	1	0	2	1
03-03-2022	1	1	0	0	0	1	1	2	3
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	1	0	0	0	0	1	0	2	3
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	2
03-03-2022	0	0	0	0	0	1	0	1	2
03-03-2022	0	0	1	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	0	1	2	1
03-03-2022	0	0	1	0	1	1	0	2	3
03-03-2022	0	0	0	0	0	1	0	1	2
03-03-2022	1	0	0	0	0	1	1	2	1
03-03-2022	0	1	0	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	1	1
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	1	0	1	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	1	0	2	1
03-03-2022	0	0	0	0	0	1	0	2	1

Figure 7.1: Datasets Stored in CSV Format



```
test_date,cough,fever,sore_throat,shortness_of_breath,head_ache,corona_result,age_60_and_above,gender,test_indication
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,2,2
03-03-2022,0,0,0,0,0,1,0,2,2
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,1,1,1,0,1,1,0,2,1
03-03-2022,1,1,0,0,0,1,1,2,3
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,1,0,0,0,0,1,0,2,3
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,2
03-03-2022,0,0,0,0,0,1,0,1,2
03-03-2022,0,0,1,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,0,1,2,1
03-03-2022,0,0,1,0,1,1,0,2,3
03-03-2022,0,0,0,0,0,1,0,1,2
03-03-2022,1,0,0,0,0,1,1,2,1
03-03-2022,0,1,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,1,0,1,0,1,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,2,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
03-03-2022,0,0,0,0,0,1,0,1,1
```

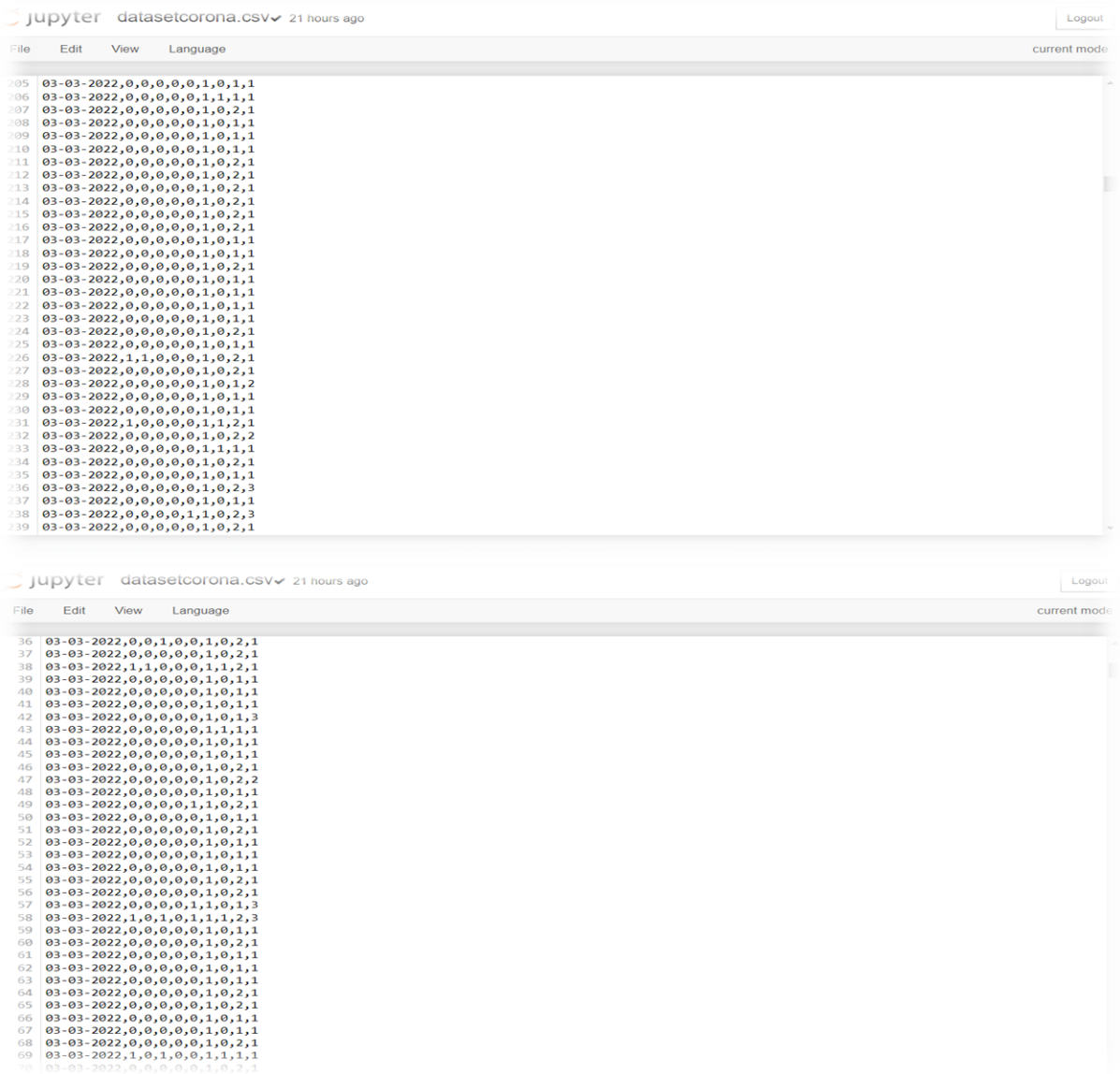


Figure 7.2: Datasets

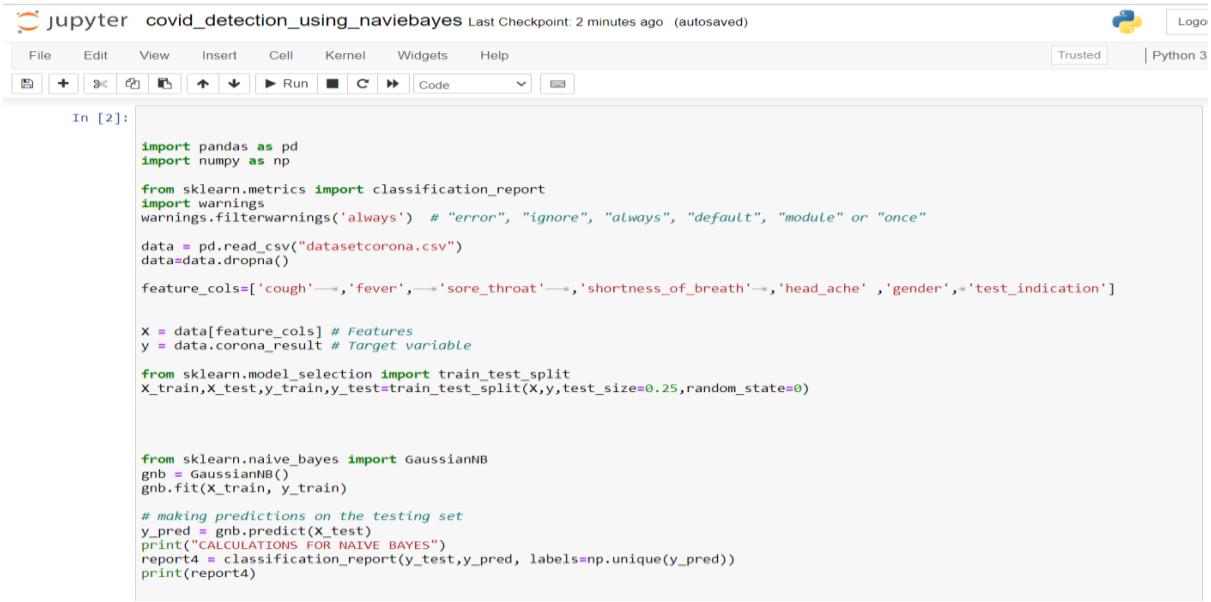


Figure 7.3: Import libraries & making prediction on testing set

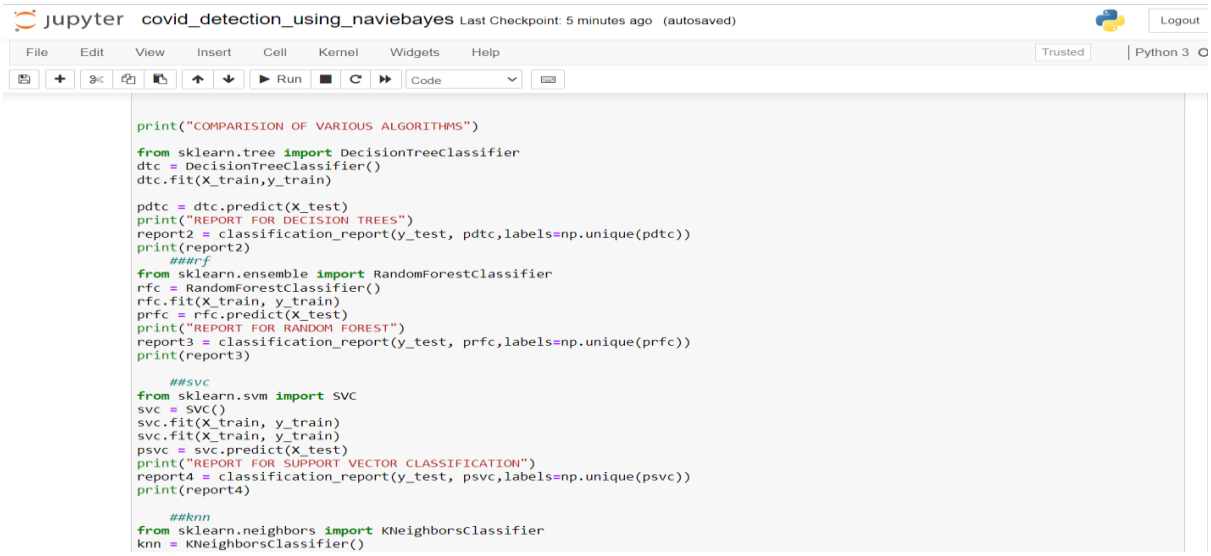


Figure 7.3.1: Comparison of various algorithms

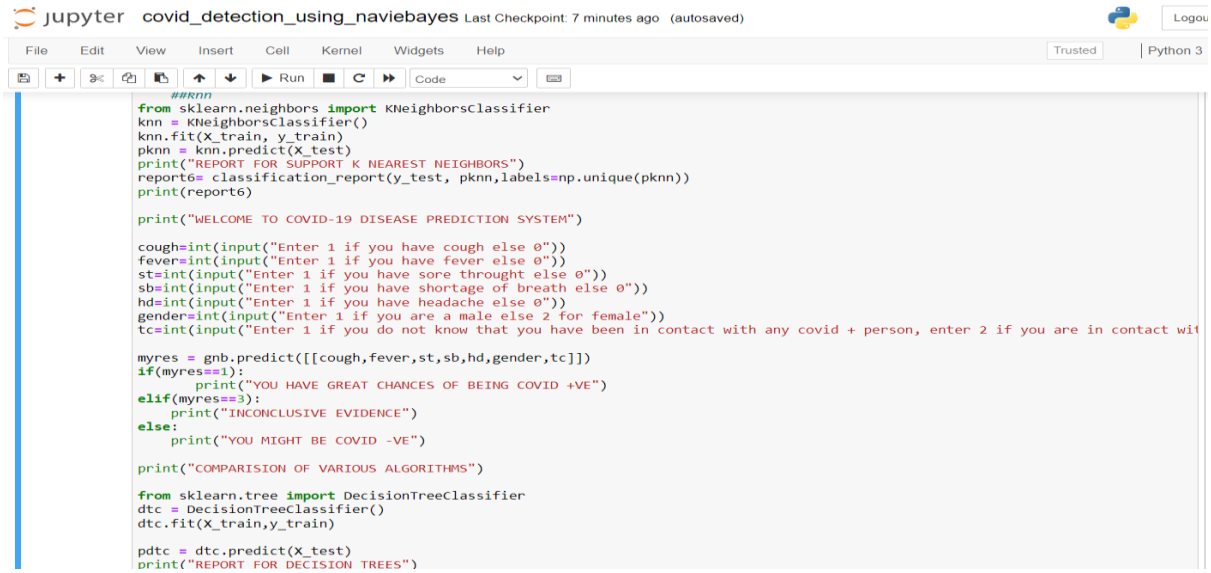


Figure 7.3.2: Taking inputs from users

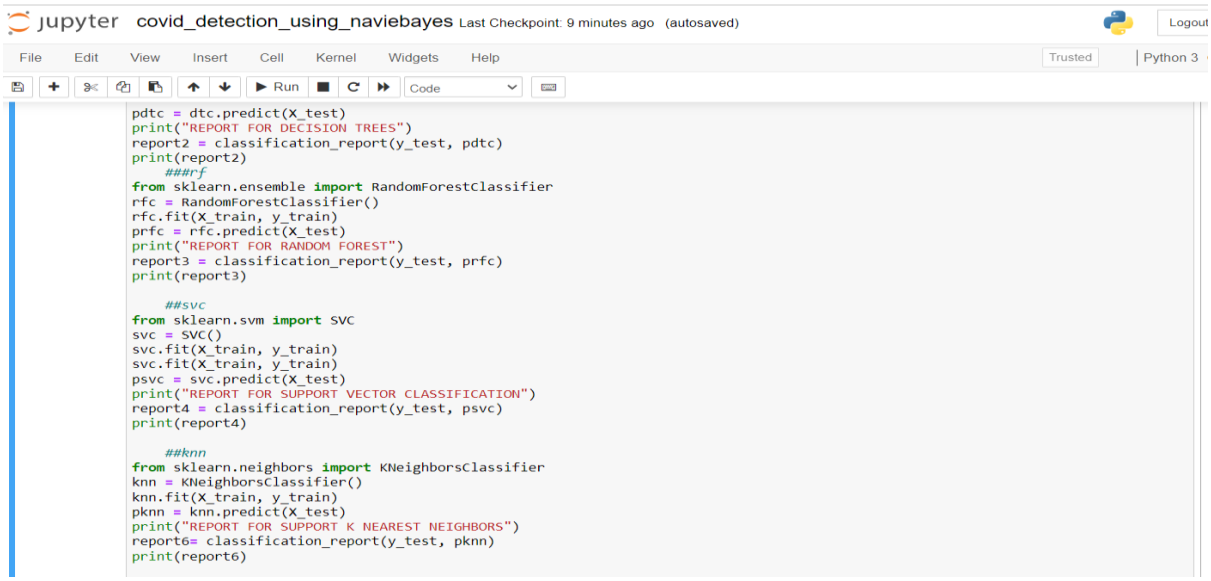


Figure 7.3.3: Calculating to prepare configuration for CV

jupyter covid_detection_using_naviebayes Last Checkpoint: 11 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

```
import pandas
import matplotlib.pyplot as plt
from sklearn import model_selection
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
# load dataset

# prepare configuration for cross validation test harness
seed = 7
# prepare models
models = []
models.append(('LR', LogisticRegression()))
models.append(('LDA', LinearDiscriminantAnalysis()))
models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier()))
models.append(('NB', GaussianNB()))
models.append(('SVM', SVC()))
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'
for name, model in models:
    kfold = model_selection.KFold(n_splits=10,shuffle=True,random_state=seed)
    cv_results = model_selection.cross_val_score(model, X, y, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
# boxplot algorithm comparison
```

7.3.4: Model preparation and evaluation

```
# boxplot algorithm comparison
fig = plt.figure()
fig.suptitle('Algorithm Comparison')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```

7.4: Boxplot algorithm comparison

```
CALCULATIONS FOR NAIVE BAYES
              precision    recall  f1-score   support

           0       0.40      1.00      0.58        57
           1       0.93      0.14      0.24       102
           3       0.00      0.00      0.00         2

    accuracy                           0.44       161
   macro avg       0.45      0.38      0.27       161
weighted avg       0.73      0.44      0.36       161

COMPARISION OF VARIOUS ALGORITHMS
REPORT FOR DECISION TREES
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        57
           1       0.62      0.96      0.76       102

   micro avg       0.61      0.62      0.61       159
   macro avg       0.31      0.48      0.38       159
weighted avg       0.40      0.62      0.49       159

REPORT FOR RANDOM FOREST
              precision    recall  f1-score   support

           1       0.63      1.00      0.78       102

   micro avg       0.63      1.00      0.78       102
   macro avg       0.63      1.00      0.78       102
weighted avg       0.63      1.00      0.78       102
```

Figure 7.5: Output comparing various algorithm

```
REPORT FOR SUPPORT VECTOR CLASSIFICATION
              precision    recall  f1-score   support

           1       0.63      1.00      0.78       102

   micro avg       0.63      1.00      0.78       102
   macro avg       0.63      1.00      0.78       102
weighted avg       0.63      1.00      0.78       102

REPORT FOR SUPPORT K NEAREST NEIGHBORS
              precision    recall  f1-score   support

           0       0.26      0.28      0.27        57
           1       0.57      0.55      0.56       102

   micro avg       0.45      0.45      0.45       159
   macro avg       0.41      0.41      0.41       159
weighted avg       0.46      0.45      0.45       159

WELCOME TO COVID-19 DISEASE PREDICTION SYSTEM

Enter 1 if you have cough else 0
```

Figure 7.5.1: Output on report for SVC & KNN

```
WELCOME TO COVID-19 DISEASE PREDICTION SYSTEM
Enter 1 if you have cough else 01
Enter 1 if you have fever else 00
Enter 1 if you have sore throught else 01
Enter 1 if you have shortage of breath else 00
Enter 1 if you have headache else 00
Enter 1 if you are a male else 2 for female2

Enter 1 if you do not know that you have been in contact with any covid + person, enter 2 if you are in contact with abroad per
son,enter 3 if you are direct contact of covid +ve person 1
```

Figure 7.5.2: User input details

```
YOU HAVE GREAT CHANCES OF BEING COVID +VE
COMPARISION OF VARIOUS ALGORITHMS
REPORT FOR DECISION TREES
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        57
           1       0.62      0.96      0.76       102
           3       0.00      0.00      0.00         2

    accuracy                           0.61       161
   macro avg       0.21      0.32      0.25       161
weighted avg       0.40      0.61      0.48       161

REPORT FOR RANDOM FOREST
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        57
           1       0.63      0.97      0.76       102
           3       0.00      0.00      0.00         2
```

Figure 7.5.3: Report for Decision Tree

```
           1       0.62      0.96      0.76       102
           3       0.00      0.00      0.00         2

    accuracy                           0.61       161
   macro avg       0.21      0.32      0.25       161
weighted avg       0.40      0.61      0.48       161

REPORT FOR RANDOM FOREST
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        57
           1       0.63      0.97      0.76       102
           3       0.00      0.00      0.00         2

    accuracy                           0.61       161
   macro avg       0.21      0.32      0.25       161
weighted avg       0.40      0.61      0.48       161

REPORT FOR SUPPORT VECTOR CLASSIFICATION
```

Figure 7.5.4: Report for Random Forest

	precision	recall	f1-score	support
0	0.00	0.00	0.00	57
1	0.63	1.00	0.78	102
3	0.00	0.00	0.00	2
accuracy			0.63	161
macro avg	0.21	0.33	0.26	161
weighted avg	0.40	0.63	0.49	161

REPORT FOR SUPPORT K NEAREST NEIGHBORS

	precision	recall	f1-score	support
0	0.26	0.28	0.27	57
1	0.57	0.55	0.56	102
3	0.00	0.00	0.00	2
accuracy			0.45	161
macro avg	0.27	0.28	0.28	161

Figure 7.5.5: Report for SVC

weighted avg	0.40	0.63	0.49	161

REPORT FOR SUPPORT K NEAREST NEIGHBORS

	precision	recall	f1-score	support
0	0.26	0.28	0.27	57
1	0.57	0.55	0.56	102
3	0.00	0.00	0.00	2
accuracy			0.45	161
macro avg	0.27	0.28	0.28	161
weighted avg	0.45	0.45	0.45	161

```
LR: 0.665745 (0.036537)
LDA: 0.659495 (0.031227)
KNN: 0.678053 (0.047075)
CART: 0.650144 (0.029682)
NB: 0.405938 (0.044483)
SVM: 0.665745 (0.036537)
```

Figure 7.5.6: Report for Support KNN

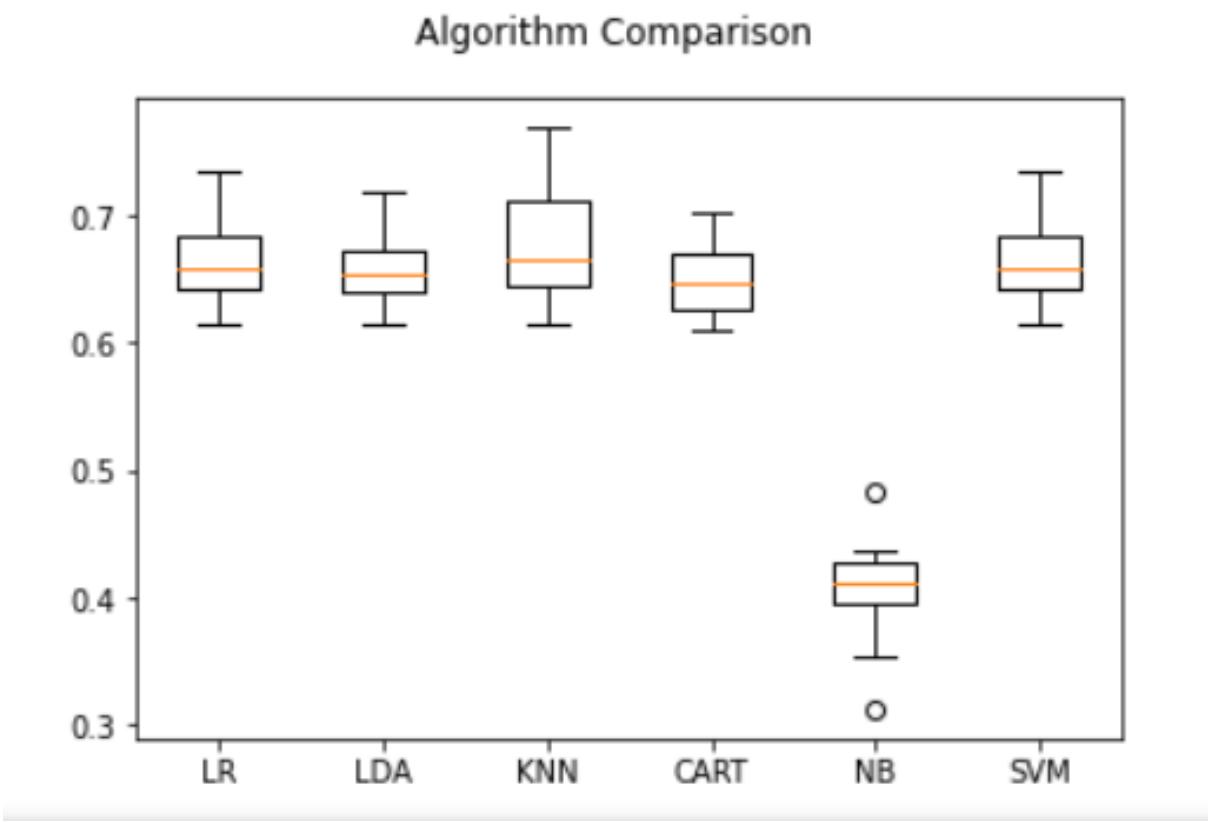


Figure 7.6: Accuracy of all Algorithms

CHAPTER 8

CONCLUSION

Coronavirus Pandemic now appears to be the most serious infected spreading disease like any other wide-spread diseases. Because of drastic rise in the number of COVID-19 cases during pandemic, it has caused severe problems in the healthcare sector to come up with appropriate and suitable treatment solutions. Machine Leaning methodologies are used as a different method for classification and prediction of virus. In our project the supervised Machine Learning classification model for COVID-19 virus developed using K Nearest Neighbors algorithm. The data mining techniques can be implemented with the K Nearest Neighbors algorithm which plays a prominent role in diagnosing the COVID-19 or coronavirus infection which is declared as pandemic by World Health Organization.

By implementing K Nearest Neighbors algorithm in our project we can conclude that,

- The standard symptoms of covid-19 detection cases are verified effectively.
- Increasing the prediction accuracy to 90%+ efficiency using K Nearest Neighbors Machine Learning technique.
- Based on the value we can easily detect multi-class prediction.
- K Nearest Neighbors performs better compared to other models like LR, LDA, CART, NB and SVM with less training data.
- Accurate output prediction will be useful and fruit-full for humanity and to detect the virus in the early stage and safe lot of lives.

REFERENCES:

1. Titled **"A Comparative Approach to predict Corona Virus using Machine Learning"** by **Chekuri Sri Sumanth, Rudra Kalyan Nayak** in the year **2021**. Journal name is **International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)**.
2. Titled **"Naïve Bayes Classifier for predicting the Novel Coronavirus Base Paper"** by **Sugandh Bhatia, Jyoteesh Malhotra** in the year **2021**. Journal name is **Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021)**.
3. Titled **"Machine Intelligent Techniques for COVID-19 Detection: A Critical Review and Analysis"** by **Chekuri Sri Sumanth, Rudra Kalyan Nayak** in the year **2021**. Journal name is **12th International Conference on Computing Communication and Networking Technologies (ICCCNT).**
4. Titled **"Severity Prediction of COVID-19 Patients using Machine Learning Classification"** by **Hina Gull, Gomathi Krishna, May Issa Aldossary, Sardar Zafar Iqbal** in the year **2020**. Journal name is **6th International Conference on Computer and Communications (ICCC).**
5. Titled **"Detection of COVID-19 from Medical images and/or Symptoms of Patient using Machine Learning Approaches"** by **Akshay Kumar Siddhu, Dr.Ashok Kumar and Dr.Shakti Kundu** in the year **2020**. Journal name is **9th International Conference System Modeling and Advancement in Research Trends (SMART)**.
6. Titled **"Journal of Infection and Public Health"** by **Harapan , Naoya Itoh, Amanda Yufika, Wira Winardi, Synat Keam, Haypheng Te, Dewi Megawati, Zinatul Hayati, Abram L. Wagner, Mudatsir** in the year **2019**.

   https://www.sciencedirect.com/science/article/pii/S1876034120304329
7. Titled **"Detection of COVID-19 from Medical Images and/ or Symptoms of Patient using Machine Learning Approaches"** by **Akshay Kumar Siddhu, Dr. Ashok Kumar and Dr. Shakthi Kundu** in the year **2020**.

   https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9336799
8. **Rajan Gupta, Ganrav Pandey, Poonam Chaudhary, Saibal K. Pal**, **"Machine Learning Models for Government to Predict COVID-19 Outbreak"**, Digital Government: Research and Practice, Vol. 1, N0. 4, Article 26, August **2020**.