

COVID-19 DATA VISUALIZATION AND PREDICTION

Report on Innovative Project Development (MCA-216)

by

Isha Tyagi (52112209)

Neha patidar (52122102)

Ankita (52112201)

Under the supervision of

Munish Bhatia

Assistant Professor



DEPARTMENT OF COMPUTER APPLICATIONS NATIONAL
INSTITUTE OF TECHNOLOGY KURUKSHETRA – 136119,
HARYANA (INDIA) June 7, 2023

DECLARATION

I hereby attest that the work displayed in "COVID-19 DATA VISUALIZATION AND PREDICTION" is an accurate representation of the work we did within January 2023 and May 2023, working under the direction of Munish Bhatiya, Assistant Professor, Department of Computer Applications.

I further assert that this project has not been turned in yet earlier for any other examination or evaluation purpose. All sources of information used in this report have been duly acknowledged.

Signature of Candidate

Isha Tyagi

Naha Patidar

Ankita

To the best of my knowledge, I hereby attest that the candidate's aforementioned assertion is true.

Signature of Supervisor

Date:7 June 2023

Munish Bhatiya,

Assistant Professor

ACKNOWLEDGEMENT

First of all, we would want to express our gratitude to the Divine for giving me the willpower and enthusiasm to successfully accomplish this study project. We sincerely thank Munish Bhatiya, an assistant professor in the computer applications department at National Institute of Technology in Kurukshetra, Haryana, for his ongoing support and encouragement in helping us better understand the research field. Throughout the whole process of writing our MCA dissertation, their advice and support were priceless. Notably, the detailed comments, helpful conversations, and beneficial activities involving our supervisors directly affected the outcome organization and standard of the MCA thesis-related work.

We want to express our appreciation to Sandeep Kumar Sood, the computer application department's head, of his invaluable leadership throughout the initial phase of doubt and confusion. We thank the teaching members and administrative personnel of the Computer Applications Department for their unflinching collaboration & unflagging support.

Without the unwavering support of my friends, I wouldn't have been able to finish my report. I would like to express my sincere thanks to my gradient for all of their blessings, love, and steadfast support. Finally, but not least, I want to express my gratitude to God for giving me the internal motivation, physical stamina, and discernment needed to complete this study project.

Isha Tyagi

Neha Patidar

Ankita

ABSTRACT

It is now recognised that a recent epidemic of acute atypical respiratory infections with an origin in Wuhan, China, a novel coronavirus of the family Coronaviridae, which causes the severe acute respiratory syndrome (SARS) coronavirus.. The coronavirus illness 19, also known as COVID-19, is a condition brought on by this virus that has spread alarmingly quickly over the world. On March 11, 2020, the WHO classified this disease to be a pandemic. This review provides an update on the etiology, clinical manifestation, and most recent management approaches for COVID-19. In particular for epidemiological monitoring, data visualization is a crucial tool for studying and sharing findings in medical research. The daily data visualization and analysis of COVID-19 are produced using the COVID19-Tracker app in a methodical manner. It automatically gathers daily information on COVID-19 diagnosed cases, hospitalisations to intensive care units, and death. 1) To assess short-term predictions and analyze data patterns. In order to enhance the app, we will soon add new data visualization and analytic apps that may aid in a better understanding of the covid-19 epidemic.

Contents

DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
List of Figures	vi
List of Tables	vii
1 INTRODUCTION	1
1.1 OBJECTIVE	1
1.2 SCOPE:	2
1.3 FEATURES:	2
2 Problem Definition	3
3 METHODOLOGY	4
3.1 Classification Algorithms	4
3.2 Polynomial Regression Model	5
3.3 Why Polynomial Regression:	5
4 PROPOSED MODEL	6
4.1 Dataset collection and data information	6
4.2 Preprocessing Input Data	6
5 Design of Model	7
6 ALGORITHM USED	8
6.1 Support Vector Machine (SVM)	8
6.2 Polynomial Regression Model	9

7	Technologies	10
7.1	Python.....	10
7.2	Jupyter Notebook.....	10
8	Coding and Results	12
8.1	Libraries.....	12
8.2	Loading Dataset.....	13
9	1st Paper	18
9.1	CONCLUSION AND FUTURE WORK.....	18
	Bibliography	18

List of Figures

5.1	The design of our project is shown. It includes data collection, data pre-processing, data training, model, COVID data given to the model After that it shows the resulting number of cases.	7
6.1	SVM(support vector machine)	8
6.2	Polynomial Regression Model	9
8.1	confirmed	14
8.2	Future Forecast	15
8.3	Support Vector Machine	15
8.4	Polynomial Regression	16
8.5	Polynomial Regression	17

List of Tables

7.1	Software Specifications	11
7.2	Hardware Specifications	11

Chapter 1 INTRODUCTION

1.1 OBJECTIVE

The primary goals of the Covid-19 Data Visualisation and Prediction model are to count the number of covid-19 instances and predict whether or not corona cases, which are a subset of covid-19 cases, would rise or fall. In late February 2020, there were the first verified cases of COVID-19.

- After the United States and Italy, Spain became the third most afflicted nation globally by the end of March and recorded the second-highest number of SARS-CoV-2 pandemic-related fatalities after Italy.
- Since March 16th, lockdown measures focused on flattening the epidemic curve have been in place in Spain, limiting social interaction, cutting back on public transport, and shutting all but the most critical enterprises to the nation's supply networks.
- This, however, has not been sufficient to stop the epidemic's upward trend. This is why a more stringent lockdown was proposed.

1.2 SCOPE:

Covid-19 Data Visualization and Prediction model help Government to understand how covid cases increased or decreased in different countries in the world

1.3 FEATURES:

- It helps a Government to develop an effective strategy for making a good plan to deal with covid cases.
- Covid-19 Data Visualization and Prediction model can also help a Government to understand how covid cases increased or decreased, what is important to a country, and what is not.

Chapter 2

Problem Definition

As we know that since last year march 2020 we are facing a epidemic all over the world and now a days it has became a global crisis for whole world and the world is not ready for a sudden growth of covid cases across the globe and if that happens than we can again face a global lockdown so before the number of death cases increases from covid-19 epidemic if we can predict the amount of death cases in near future from the ongoing current cases than by the help of that we can save many lives by taking proper precautions before hand and we will not have to face severe lockdown.

Chapter 3 METHODOLOGY

Machine learning, frequently abbreviated as ML, is a cutting-edge area of computer science that aims to teach computers to do tasks without being specifically programmed to do so. In order to build algorithms that make estimates upon data sets, machine learning employs a variety of methodologies. These are employed in mining data, a technique for finding patterns and trends in data sets when relationships were earlier unobserved. Computer vision, optimisation issues, search engines, and other applications all employ machine learning. These ideas have been used in Google's self-driving vehicle and in the recommendation systems on websites like Netflix and Amazon. Building a model from a training set is how machine learning algorithms operate. A data set used as input into an algorithm with proven valid outputs is referred to as a training set. The machine learning algorithm creates the model as it reads the training set, predicts the next input, checks its forecast against the actual output, and makes adjustments as necessary.

3.1 Classification Algorithms

For categorizing newly collected data with respect to the training data, the Classification method is a Supervised Learning approach. A programme identifies additional data points after training from the set of data or supplied information. into various classes or groups. such as, "Yes" or "No," "0" or "1," "Spam" or "Not Spam," "cat" or "dog," etc. Classes may also be referred to as goals, labels, or categories..

Logistic Regression: To calculate the probability of a target variable, the supervised learning classification technique logistic regression is utilized. There are only two possible classes since the aim or dependent variable is dichotomous. The dependent variable is, to put it simply, a binary variable, meaning that data is represented as either 1 (which represents success/yes) or 0 (which stands for failure/no). A logistic regression model predicts $P(Y=1)$ as a function of X mathematically. It is one of the most basic ML techniques and may be applied to a variety of classification problems, such as spam detection, diabetes prediction, cancer diagnosis, etc.

3.2 Polynomial Regression Model

In polynomial regression, a kind of linear regression, the relationship between the independent variable x and the dependent variable y is represented as an n th degree polynomial. Polynomial regression is used to fit a nonlinear relationship between the value of x and the corresponding conditional mean of y , denoted by the notation $E(y|x)$.

3.3 Why Polynomial Regression:

- Some correlations may be curvilinear, according to a researcher's hypothesis. Clearly, a polynomial term will be present in these types of circumstances.
- Checking the leftovers. In the center of a scatter plot of residuals (Y axis) on the predictor (X axis), patches of many positive residuals will form if we attempt to fit a linear model to curved data. Therefore, it is improper in this situation.
- CommonThe basis for working in multiple linear regression analysis is that each independent variable is independent of the others. This presumption is not true for polynomial regression models. Polynomial Regression Applications: Generally speaking, they are employed to characterize or describe nonlinear phenomena like:
 - Tissues' pace of growth.

- Development of the pandemic illness
- Distribution of carbon isotopes in lake sediments

Chapter 4 PROPOSED

MODEL

1. Data and package import: First, we import the xlsx (Excel spreadsheet) data file, followed by the packages we will need to conduct our analysis. In the folder (directory) where I provide Jupiter's notebook, I save the xlsx file for this example.
2. Data organization: We need to clean and organize this data so that we can provide more actionable insights since after importing the package and data, we will discover that the data is not as useful as that.

4.1 Dataset collection and data information

The covid-19 dataset, which largely consists of active cases and fatalities, is the one taken into consideration for use in the specific suggested task. The John Hopkins University GitHub Page is accessed at https://raw.githubusercontent.com/CSSEGISandData/COVID19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_active_cases_time_series.csv to receive an open source dataset.

4.2 Preparing The Input Data

The input dataset is subjected to a number of pre-processing procedures, including filling in scaling values, categorical data encoding, and missing values to the proper range. dividing up data for testing and analyzing After that, the pre-processed dataset are divided into a training and testing dataset depending on the given split ratio. The split ratio taken into account in the proposed study is 74:26, which implies 74% of the

dataset was used for the ensemble model's train and remaining 26% is used for testing.

Chapter 5 Design

of Model

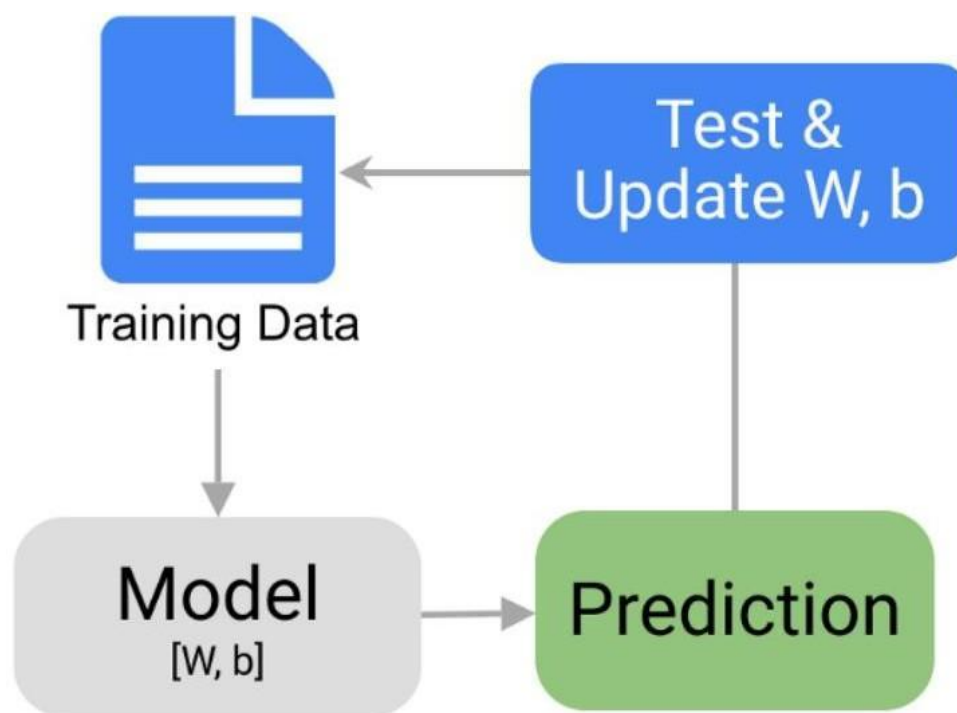


Figure 5.1: The design of our project is shown. It includes data collection, data pre-processing, data training, model, COVID data given to the model after that it shows the resulted number of cases.

Chapter 6 ALGORITHM

USED

6.1 Support Vector Machine (SVM)

Support vector machines (SVMs) and other kernelized models use the polynomial kernel as a kernel function to represent the similarity of vectors (training samples) in a feature space over polynomials of the original variables, enabling the learning of non-linear models.

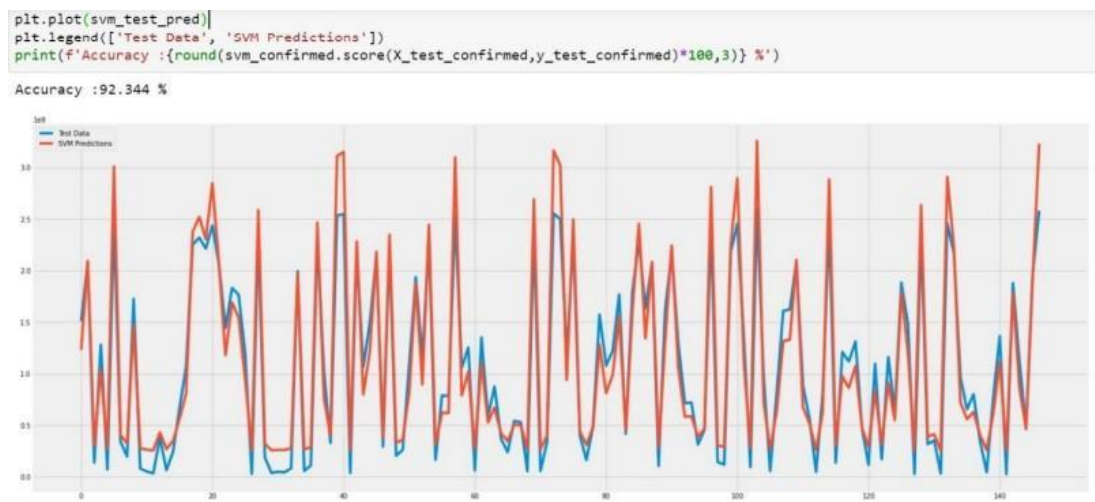


Figure 6.1: SVM(support vector machine)

6.2 Polynomial Regression Model

A kind of linear regression known as polynomial regression models the relationship between the independent variable x and the dependent variable y as an n th degree polynomial. By using polynomial regression, a nonlinear relationship between the value of x and the corresponding conditional mean of y , denoted by the notation $E(y|x)$, is fit.

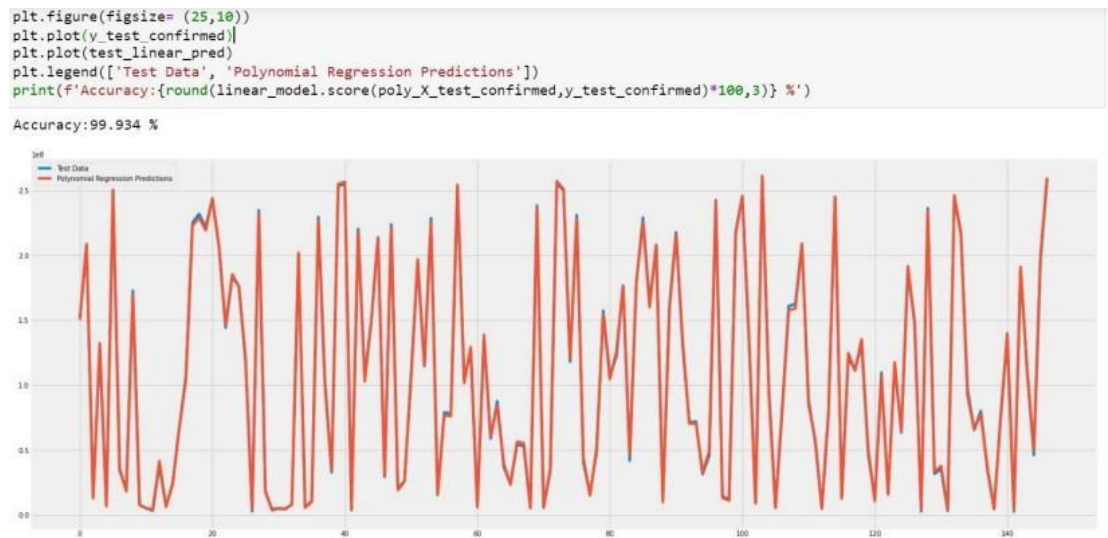


Figure 6.2: Polynomial Regression Model

Chapter 7

Technologies

7.1 Python

Python is an interpreted, general-purpose programming language. Guido van Rossum created Python, which was first made accessible in 1991. Its design philosophy makes excellent use of significant whitespace and lays a heavy focus on code readability. With its linguistic components and object-oriented architecture, programmers should find it easier to write logical, clear code for both small and large projects. Python has dynamic typing and garbage collection. The procedural, object-oriented, and functional programming paradigms are all supported. Python is commonly referred to as a "batteries included" language because of its substantial standard library.

7.2 Jupyter Notebook

Use the free and open-source Jupyter Notebook online tool to create and share documents that contain live code, equations, visualizations, and narrative prose. Examples of applications include data purification and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and many more.

Table 7.1: Software Specifications

ITEM	NEEDED(%)
Operating System	Window XP/7/8/10, UNIX/LINUX
Language	Python 3.8 or above
Platform	pycharm, Jupyter Notebook
Libraries	pycharm, Jupyter Notebook

Table 7.2: Hardware Specifications

ITEM	NEEDED(%)
Processor	Intel Pentium, core i3 Or Above
Monitor	Color Monitor
Web Server	At least 2gb RAM and Pentium 4 2000 MHz processor
Database Server	At least 1 GB and depends on the usages of application.

Chapter 8

Coding and Results

8.1 Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
import random
import math
import time
from sklearn import metrics

from sklearn.linear_model import LinearRegression, BayesianRidge
from sklearn.model_selection import RandomizedSearchCV
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.svm import SVR
from sklearn.metrics import confusion_matrix
from sklearn.metrics import mean_squared_error, mean_absolute_error, f1_score, accuracy_score
import datetime
import operator
```

```
plt.style.use('fivethirtyeight')
%matplotlib inline import
warnings
warnings.filterwarnings("ignore")
```

8.2 Loading Dataset

```
confirmed_df =
pd.read_csv('https : //raw.githubusercontent.com/CSSEGISandData
/COVID19/master/csvcovid19data/covid19 time series
/time_series_covid19_confirmed_global.csv')
deaths_df =
pd.read_csv('https : //raw.githubusercontent.com/CSSEGISandData/COVID19
/master/covid19data/covid19 time series/time_series_covid19_deaths_global.csv')
recoveries_df =
pd.read_csv('https : //raw.githubusercontent.com/CSSEGISandData/COVID19
/master/covid19data/covid19 time series/time_series_covid19_recovered_global.csv') latest_data =
pd.read_csv('https : //raw.githubusercontent.com/CSSEGISandData/COVID19
/master/covid19data/covid19 daily reports/08 - 01 - 2020.csv')
```

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	...	11/10/21	11/11/21	11/12/21	11/13/21	11/14/21	11/15/21	11/16/21
0	0	0	0	0	0	0	0	0	0	0	...	156414	156456	156487	156510	156552	156610	156649
1	0	0	0	0	0	0	0	0	0	0	...	190815	191440	192013	192000	193075	193269	193856
2	0	0	0	0	0	0	0	0	0	0	...	207509	207624	207764	207873	207970	208104	208245
3	0	0	0	0	0	0	0	0	0	0	...	15744	15744	15819	15819	15819	15907	15929
4	0	0	0	0	0	0	0	0	0	0	...	64815	64857	64875	64899	64913	64913	64940
...
275	0	2	2	2	2	2	2	2	2	2	...	882735	1000897	1009879	1018346	1026522	1035138	1045397
276	0	0	0	0	0	0	0	0	0	0	...	456186	456407	456632	456632	456632	457154	457390
277	0	0	0	0	0	0	0	0	0	0	...	9902	9907	9912	9918	9918	9936	9936
278	0	0	0	0	0	0	0	0	0	0	...	209939	209953	209963	209971	209983	209996	210008
279	0	0	0	0	0	0	0	0	0	0	...	133302	133329	133329	133393	133428	133438	133505

280 rows × 668 columns

deaths.head()

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	...	11/10/21	11/11/21	11/12/21	11/13/21	11/14/21	11/15/21	11/16/21	11/17/21
0	0	0	0	0	0	0	0	0	0	0	...	7291	7292	7292	7292	7293	7294	7295	7295
1	0	0	0	0	0	0	0	0	0	0	...	2975	2978	2983	2986	2990	2997	3004	3004
2	0	0	0	0	0	0	0	0	0	0	...	5960	5966	5971	5977	5984	5991	5997	5997
3	0	0	0	0	0	0	0	0	0	0	...	130	130	130	130	130	130	130	130
4	0	0	0	0	0	0	0	0	0	0	...	1723	1725	1726	1727	1727	1727	1727	1729

5 rows × 668 columns

recoveries.head()

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	...	11/10/21	11/11/21	11/12/21	11/13/21	11/14/21	11/15/21	11/16/21	11/17/21
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0

5 rows × 668 columns

confirmed.head()

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	...	11/10/21	11/11/21	11/12/21	11/13/21	11/14/21	11/15/21	11/16/21	11/17/21
0	0	0	0	0	0	0	0	0	0	0	...	156414	156456	156487	156510	156552	156610	156649	156649
1	0	0	0	0	0	0	0	0	0	0	...	190815	191440	192013	192000	193075	193269	193856	193856
2	0	0	0	0	0	0	0	0	0	0	...	207509	207624	207764	207873	207970	208104	208245	208245
3	0	0	0	0	0	0	0	0	0	0	...	15744	15744	15819	15819	15819	15907	15929	15929
4	0	0	0	0	0	0	0	0	0	0	...	64815	64857	64875	64899	64913	64913	64940	64940

5 rows × 668 columns

Figure 8.1: confirmed



Figure 8.2: Future Forecaste



Figure 8.3: Support Vector Machine

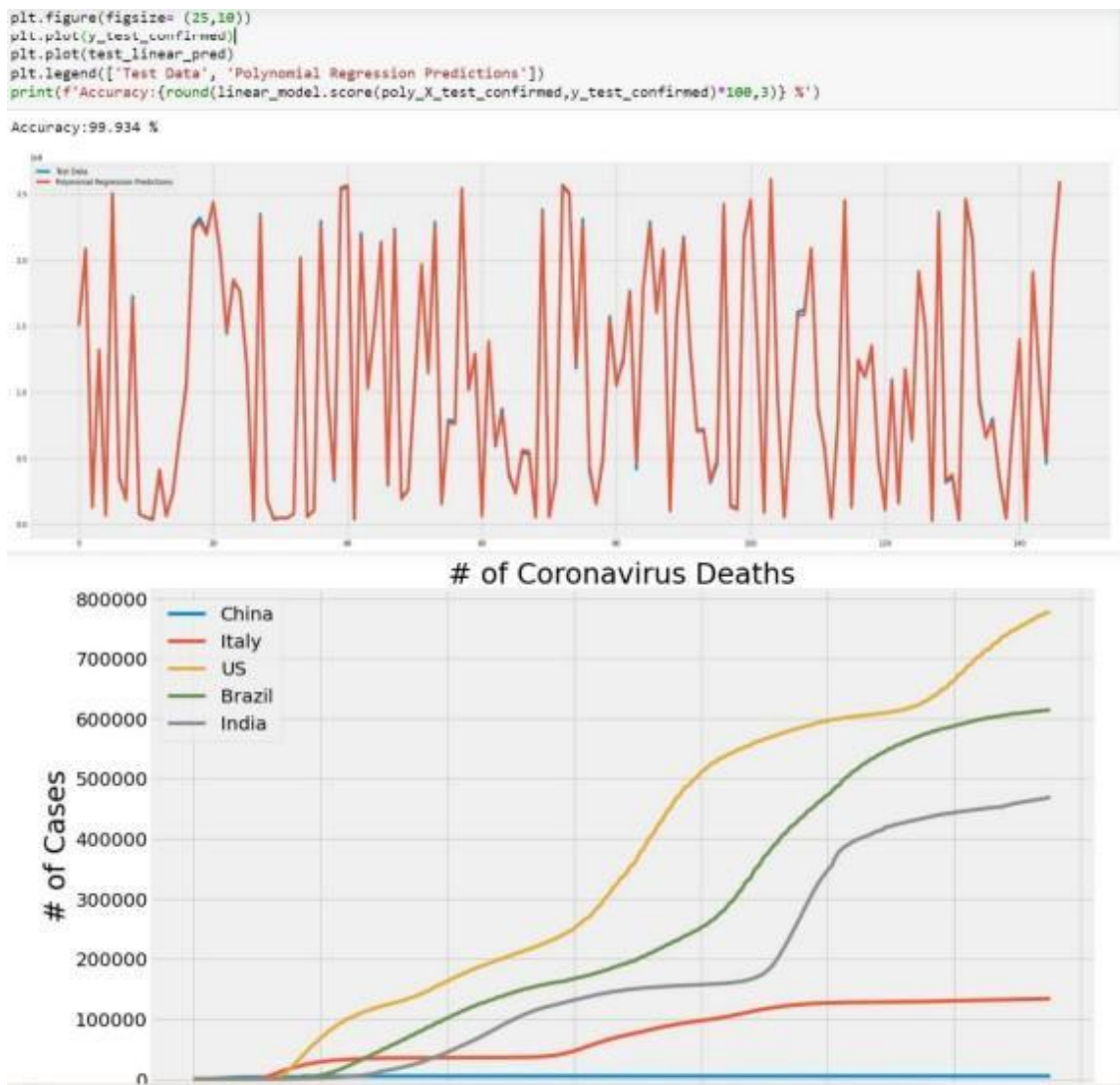


Figure 8.4: Polynomial Regression

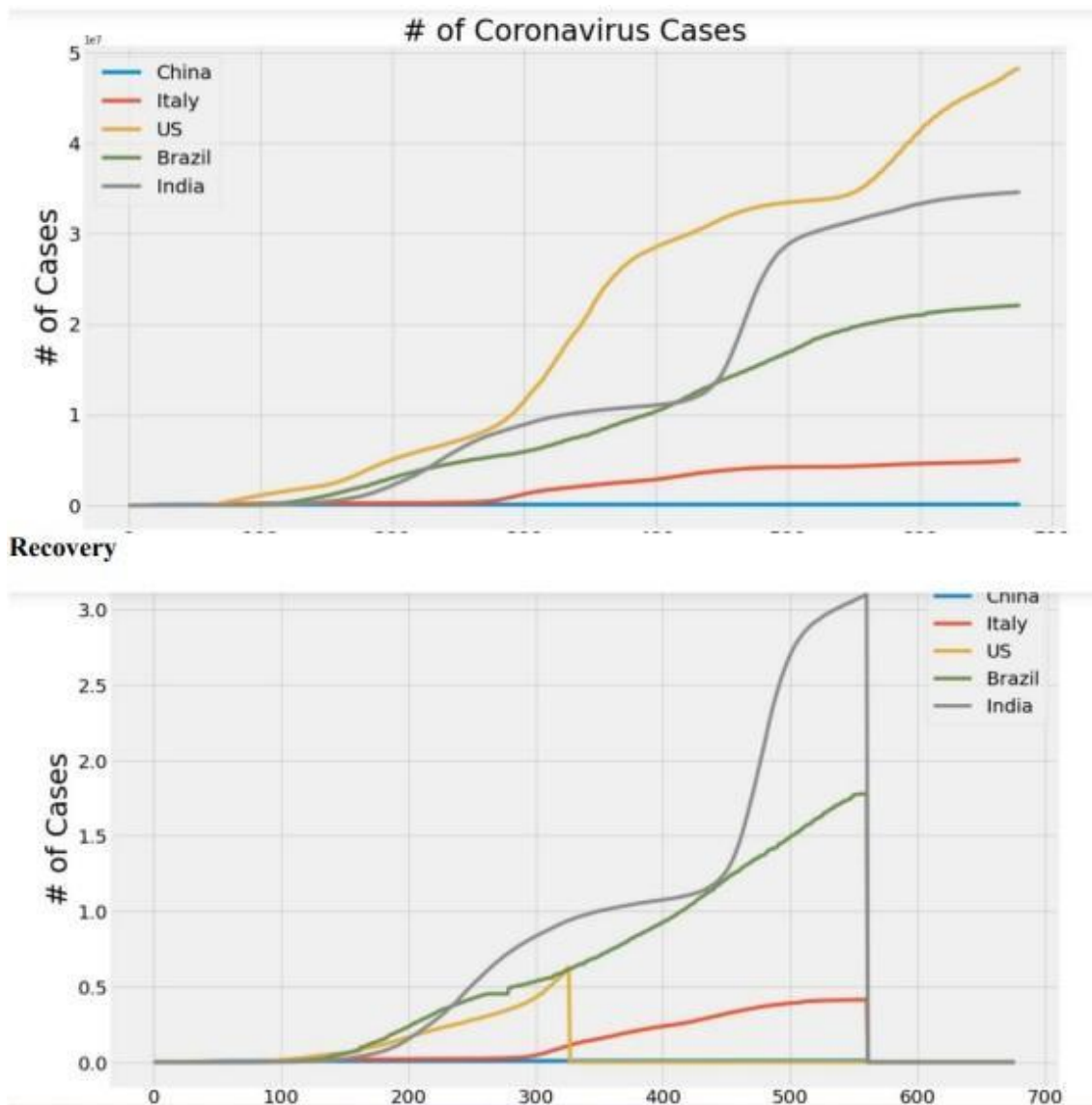


Figure 8.5: Polynomial Regression

Chapter 9 1st

Paper

9.1 CONCLUSION AND FUTUREWORK

Now-a-days it is very essential to have a future preparation of medical services to fight against any type of epidemic that may come into the future . Covid-19 model have been design to predict the number of cases of any type of epidemic and also its data visualization part will help government to analyze the data and to take proper steps so that we can overcome these types of epidemics in future. For the future, we can use this project for future implementation of various precautions. So that we can reduce the spreading of covid-19. Since ML project is all about data more data gives more accurate results. Since here the online available data is used for it. The real laboratories data can be used for more accurate results. In Future the this model can also be used for collecting the data all over. So that there is no limitation of data.. Deep Learning techniques can also be used for further research.

Bibliography

- [1] Santosh, K. C. "COVID-19 prediction models and unexploited data." *Journal of medical systems* 44, no. 9 (2020): 170.
- [2] Hassan, Ahmed Hassan Mohammed, Arfan Ali Mohammed Qasem, Walaa Faisal Mohammed Abdalla, and Omer H. Elhassan. "Visualization & prediction of COVID-19 future outbreak by using machine learning." *Int. J. Inf. Technol. Comput. Sci* 13, no. 3 (2021): 16-32.
- [3] Hamzah, F. Binti, C. Lau, Hafeez Nazri, Dominic Vincent Ligot, Guanhua Lee, Cheng Liang Tan, M. K. B. M. Shaib, Umami Hasanah Binti Zaidon, Adina Binti Abdullah, and Ming Hong Chung. "CoronaTracker: worldwide COVID- 19 outbreak data analysis and prediction." *Bull World Health Organ* 1, no. 32 (2020): 1-32.
- [4] Sen, Snigdha, B. K. Thejas, B. L. Pranitha, and I. Amrita. "Analysis, visualization and prediction of COVID-19 pandemic spread using machine learning." In *Innovations in Computer Science and Engineering: Proceedings of 8th ICICSE*, pp. 597-603. Springer Singapore, 2021.
- [5] Yu, Cheng-Sheng, Shy-Shin Chang, Tzu-Hao Chang, Jenny L. Wu, Yu-Jiun Lin, Hsiung-Fei Chien, and Ray-Jade Chen. "A COVID-19 pandemic artificial intelligence-based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study." *Journal of medical Internet research* 23, no. 5 (2021): e27806.
- [6] Jahmunah, Vicnesh, Vidya K. Sudarshan, Shu Lih Oh, Raj Gururajan, Rashmi Gururajan, Xujuan Zhou, Xiaohui Tao et al. "Future IoT tools for COVID-19

contact tracing and prediction: A review of the state-of-the-science.” International journal of imaging systems and technology 31, no. 2 (2021): 455-471.