

A
Seminar Report on

“REGRESSION ANALYSIS OF COVID-19”

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN
PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF

THIRD YEAR COMPUTER ENGINEERING

SUBMITTED BY

DESHPANDE APURV SANJAY

ROLLNO:C31161

UNDER THE GUIDANCE OF

Prof. V. M. GOSAVI



Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

STES SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, VADGAON

BK, PUNE - 411041

2022-23

AFFILIATED TO



Savitribai Phule Pune
University



CERTIFICATE

This is to certify that the seminar report entitled

" REGRESSION ANALYSIS OF COVID-19"

submitted by

DESHPANDE APURV SANJAY

ROLLNO:C31161

is a bonafide work carried out by the above student under the supervision of Prof. V. M. Gosavi and it is approved for the partial fulfilment of the requirement of Savitribai PhulePune University, Pune for the award of the degree of Bachelor of Engineering (Computer Engineering).

V. M. Gosavi
Seminar Guide

Mr. R.H. Borhade
Head of Department

Dr. A.V. Deshpande
Principal,
STES' S SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, VADGAON, PUNE.

Examiner Name & Sign:

Place:

Date:

ACKNOWLEDGEMENT

With due respect and gratitude, I take the opportunity to thank those who have Helped me directly and indirectly. I convey my sincere thanks to Mr. R.H. Borhade, HoD Computer Dept. and Prof. Varsha Gosavi for their help in selecting the seminar topic and support. I thank my seminar guide Prof. Varsha Gosavi for her guidance, timely Help and valuable suggestions without which this seminar would not have been. Possible. Her direction has always been encouraging as well as inspiring for me. Attempts have been made to minimise the errors in the report. I would also like to express my appreciation and thanks to all my friends who Knowingly or unknowingly have assisted and encouraged me throughout my hard Work.

ABSTRACT

The novel Corona Virus (Covid 19) is a global pandemic. It was first reported in 2019, and has exceedingly affected every characteristic of life. The outbreak of the SRS CoV2 virus, commonly referred to as the Covid-19 has posed the social, economic as well as the cultural lives of global citizens. The pandemic to this day, is still a menacing warning to public health. The main objective of this report to provide a better view of various machine learning regression algorithms are to be implemented in such real-world problems. Using many popular regression techniques like linear regression, polynomial regression and support vector machine algorithm, we try to analyse all the data available on the covid-19 impact, in India. Divergent trends and patterns are observed the spread of covid-19 outbreak in India and its different states. We use the statistical data obtained with help of datasets from Ministry of Health and Family Welfare of India. The existing data and information can be further analysed to predict number of patients in different states of India.

TABLE OF CONTENTS

LIST OF TABLES

NOMENCLATURE

CONTENTS

1. INTRODUCTION

1.1 About Covid-19

1.2 Effects of covid-19

1.3 Spread of covid-19 in India

1.4 Analysing spread of covid-19

2. LITERATURE SURVEY

3. PROJECT DESCRIPTION

3.1 Data acquisition

3.2 Model selection

3.3 Model implementation

4. ANALYTICAL/ EXPERIMENTAL WORK

5. BENEFITS, DRAWBACK AND APPLICATIONS

5.1 Benefits

5.2 Drawbacks

5.3 Applications

6. PLAGARISM REPORT

7. FUTURE PROSPECT

8. CONCLUSION

9. REFERENCES

1. INTRODUCTION:

1.1 About Covid-19

According to World health Organization, life threatening viruses such as the Covid-19 virus continue to pose a huge danger ahead of human lives. It is an infectious disease caused by the SARS-CoV-2 virus which travels in respiratory droplets released into the air when an infected person coughs, sneezes, talks, sings or breathes near you. A person may be infected if they inhale these droplets. You can also get coronavirus from close contact (touching, shaking hands) with an infected person and then touching your face. The WHO (World health Organization) declared the coronavirus disease as a global pandemic on March 11, 2020.

1.2 Effects of covid-19

The covid-19 virus is transmitted through direct contact with respiratory droplets of an infected person generated through coughing and sneezing, and touching surfaces contaminated with the virus. It may survive on surfaces for several hours, but simple disinfectants can kill it most of the time.

Most people infected with the virus experience mild to moderate respiratory illness, and recover without requiring special treatment. However, a small percentage of people become seriously ill and require medical attention. The most appalling effect of covid-19 is anyone can get sick with COVID-19 and become seriously ill or die at any age.

1.3 Spread of covid-19 in India

In India, the first case of covid-19 was reported in Kerala, and soon the virus spread in all parts of the country mostly due to imported cases from other countries. As the cases continued to rise continuously, the government was thus forced to impose a nationwide lockdown on March 25, 2020. During this period of lockdown, the cases kept increasing and peaked at 90,000 cases per day in mid-September. In March 2021, a second wave of covid-19 virus gripped the country. The situation was aggravated by a shortage of hospital beds, oxygen cylinders and medicines. The discovery of a new, more transmissible Omicron variant gave rise to the third wave in the country. However, the new variant caused symptoms which were less severe. The states of Maharashtra, Kerala and Delhi remain some of the worst-hit states in the country.

1.4 Analysing spread of covid-19

Ever since the world has been struck by the pandemic, there have been many reports and analytics of the covid-19 spread. But there are very limited studies on state-wise analysis of the outbreak. We know that India has abundant population and diversity, therefore all the states have much more population as compared to other parts of the world.

Taking into account the population of the country India, it becomes essential to utilize all the resources efficiently and effectively. Despite not having some of the best health care facilities, India has controlled the pandemic the best way possible. Hence, there has been a need to analyse the spread of covid-19 virus.

2. LITERATURE SURVEY

2.1 Literature Survey Table

Sr No	Paper Name	Purpose
1	Regression analysis of Covid-19 Spread in India and its different states.	To demonstrate the impact of spread of covid-19 across India and its different states.
2	Regression analysis of Covid-19 using machine learning algorithms	Provides a better understanding of how various machine learning algorithms work and be implemented in real-world situations.
3	Regression Analysis for COVID-19 Infections and Deaths Based on Food Access and Health Issues	Give a better understanding of the danger caused by covid-19 in the food industry.
4	Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction rate	Provide a brief analysis of the impact of covid-19 in the world and how it affected the overall reproduction rate across all the countries.

2.2 Methods:

Regression analysis takes a form of predictive modelling technique which investigates the relationship between variables which can be dependent or independent. It can also be termed as a statistical method that helps us to analyse and understand the relationship between two or more variables of interest. We use the relationship between various variables which can depend on each other to fit the lines of regression. This can in-turn be used for making many predictions.

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is widely used for predictive analysis.

Polynomial regression uses the relationship between two or more variables to find the best way to draw a line through the data points.

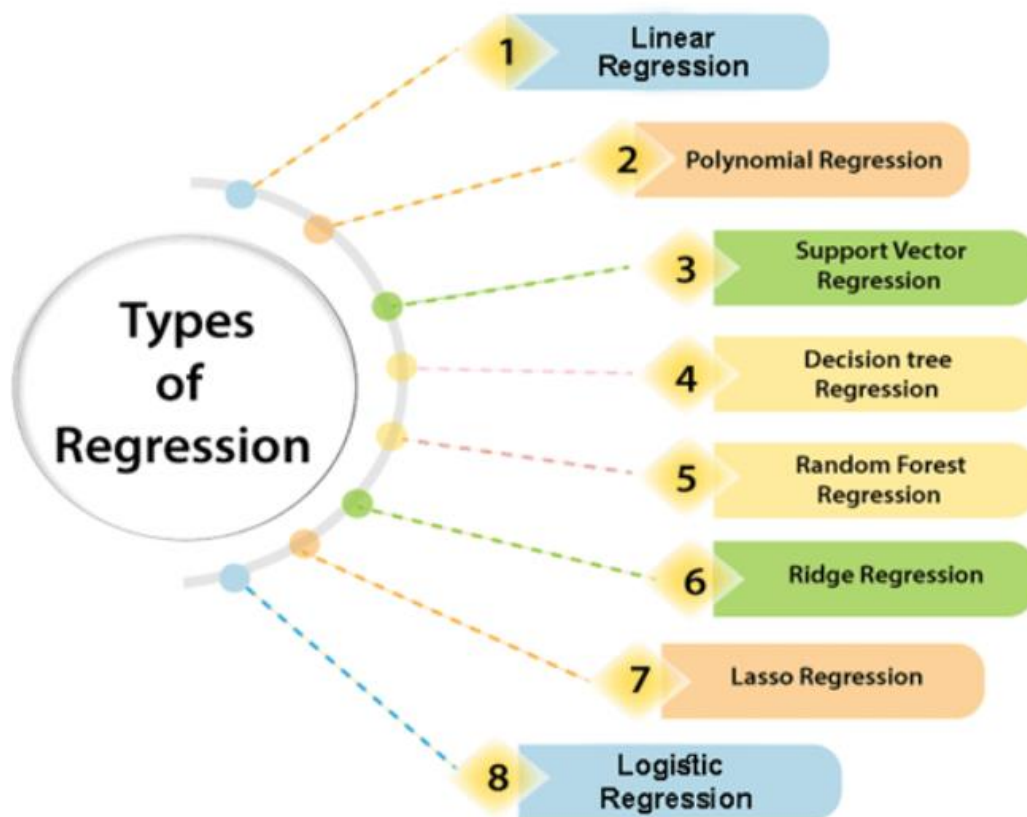


Figure 1: Types of Regression methods

2.3 State-wise spread of Covid-19 in India

The covid-19 disease has spread across 212 countries and territories around the world, with a total of more than 3 million confirmed cases. In the vast India, the disease was first detected on January 30, 2020, in Kerala. The total number of confirmed infected people is more than 4.46 crores to date across India.

After the first 3 cases from January 30 to February 3, 2020, there were no confirmed COVID-19 cases for about a month across India. The data is acquired from the datasets of the Ministry of Health and Family Welfare of India.

The COVID-19 cases appeared again from March 2, 2020. These cases are related to people who have been evacuated or have arrived from covid prone countries. From March 20, 2020, onwards, there is an exponential growth in the daily number of COVID-19 cases at the pan-India level.

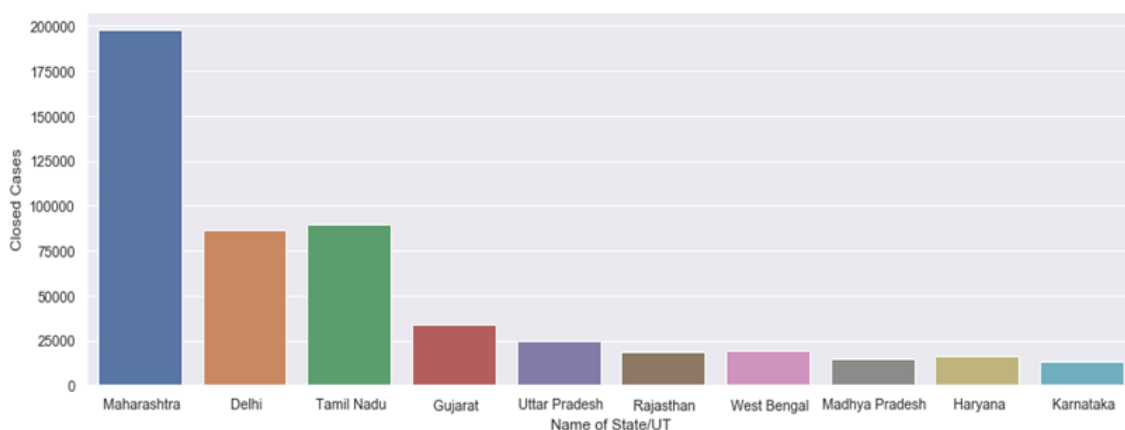


Figure 2 : Graph depicting state-wise closed cases

The above bar diagram depicts the state-wise closed cases of Covid-19 virus in India. Evidently, the state of Maharashtra was the most affected state.

	Name of State/UT	Active cases	Recovered cases	Deaths	Total Confirmed Cases
1	Maharashtra	61807	67706	6283	135796
2	Delhi	23820	36602	2233	62655
3	Tamil Nadu	27181	34112	794	62087
4	Gujarat	6232	19909	1684	27825
5	Uttar Pradesh	6152	11601	569	18322
6	Rajasthan	2966	11910	356	15232
7	West Bengal	5102	8687	569	14358
8	Madhya Pradesh	2342	9215	521	12078
9	Haryana	4940	5916	169	11025
10	Karnataka	3527	5730	142	9399

Figure 3: Comparison of most affected states

The above table demonstrates and compares the spread of the virus across the top 10 most affected states. It is evident that the states of Maharashtra, Delhi and Tamil Nadu were the most affected states, Maharashtra crossing the 1,00,000 mark the first.

3. PROJECT DESCRIPTION

There are 3 steps to study the regression analysis of Covid-19:

3.1 Data acquisition

We use the statistical data obtained with help of datasets from Ministry of Health and Family Welfare of India. The dataset available from the data repository for the “2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Centre for Systems Science and Engineering is parameterized dataset having relevant parameters such as Province/State, Country/Region, Latitude, Longitude and dates. Separate datasets have been used for Confirmed, Death, and Recovered cases along with the number of cases on each day.

The total dataset used in the study is obtained for 154 days i.e. from January 22, 2020, till June 24, 2020. The data obtained from these datasets were finally merged to obtain the parameterized dataset of the world from January 22, 2020, till June 24, 2020.

The reason behind choosing regression analysis for the current problem statement is the type of dataset. The dataset is a continuous dataset and regression analysis is best suited when a continuous dependent variable is wanted to predict from several independent variables. The coefficients of the dependent and independent variables in the regression equation determine the relationship between the dependent and independent variables. Since we have used a continuous dataset, our job of analysis of data is well put together.

3.2 Model selection

Model selection refers to the process of choosing the model that best generalizes the given data from datasets. This procedure is performed to select the most suitable model for the dataset. Part of the data available is used to train the predictive model and the remaining to test its accuracy. Once trained the predictive model is able to provide future forecasts of the variables under analysis.

3.3 Model implementation

A. Support Vector Machine Algorithms are supervised machine learning models that are associated with data classification and regression analysis. It constructs a hyperplane or set of hyperplanes in an N-dimensional space for classification or outlier detection. Below graph depicts implementation of Support vector machine algorithm.

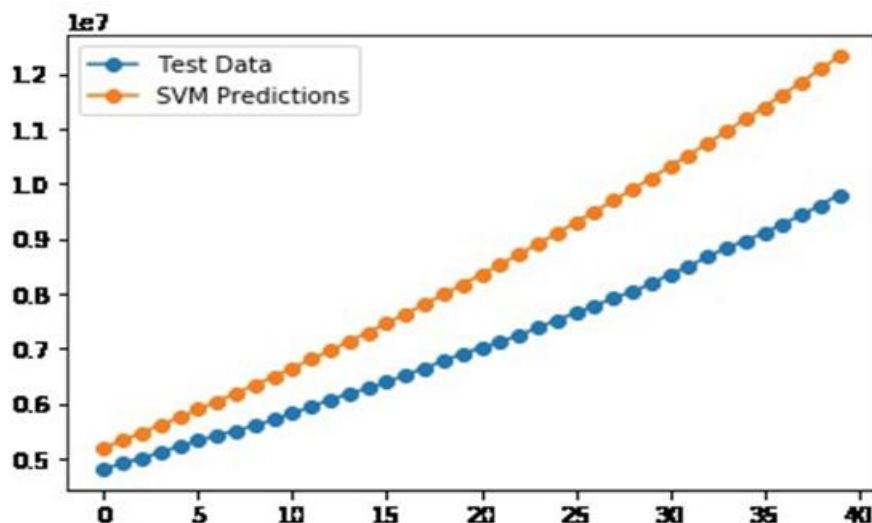


Figure 4 : Support vector machine

B. Polynomial Regression can be expressed as a special case of Linear Regression. In Linear Regression it works on continuous data is known and the two variables are correlated. But there can be a situation that the variables are corelated but the relation is not linear, so polynomial regression to fit a polynomial equation can be used to our dataset.

Polynomial Regression is a supervised Machine learning Algorithm that is trained based on prior data and then tested on another dataset to validate its accuracy.

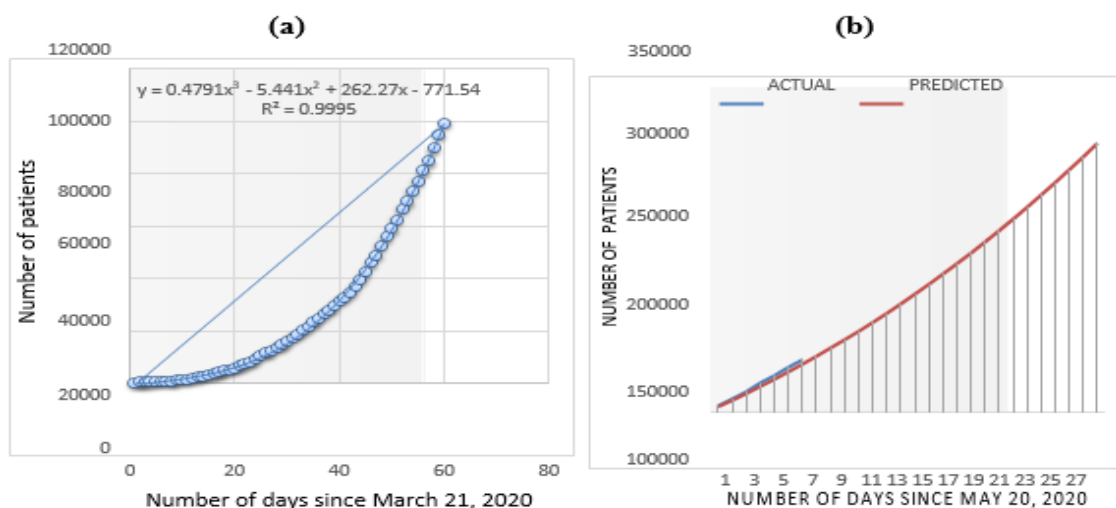


Figure 5 : Polynomial regression

In the above graphs, we consider the analysis of spread of covid-19 using polynomial regression model in India. It is evident in figure 5(a) that till 20th march 2020, there was not much rise in covid patients. But after 20th march 2020, we can see a substantial rise in number of cases.

C. Linear regression is the most widely used regression technique. It is used for finding out the relation between multiple variables that can be dependent or independent along a straight line.

The simple linear regression analysis of number of deaths is shown in Figure 6(a).

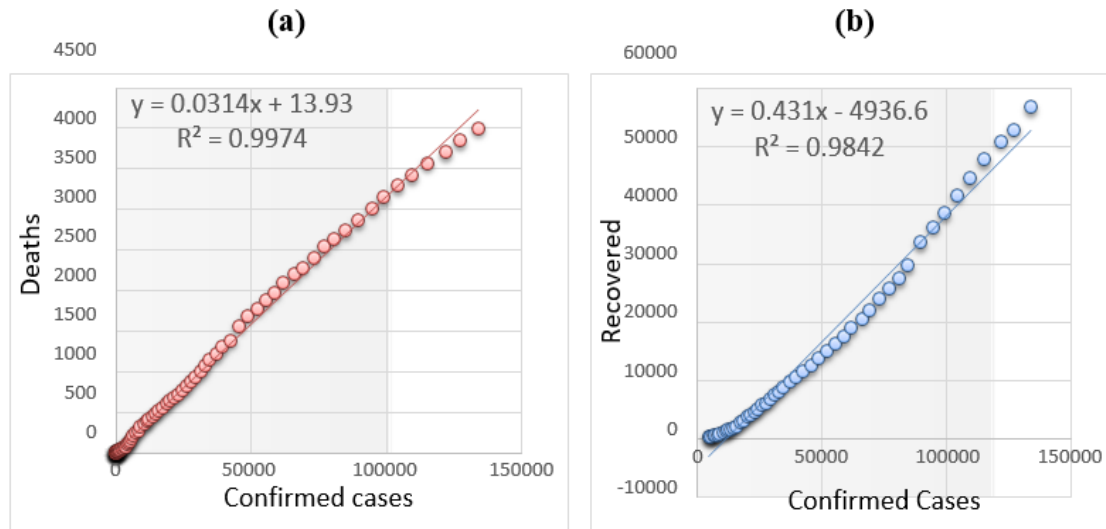


Figure 6 : Linear regression

The above diagrams help us to grasp the analysis of the total number of cases with the total number of deaths as well as recovered cases, in the states where the hit of covid-19 was abounded. Data is considered till 26th May 2020.

We observe that in figure 6(a), the graph of total cases vs deaths shows a very strong linear relation between the two. The simple linear regression analysis of the total cases vs recovered cases are shown in figure 6(b). It also shows a linear correlation between confirmed and total recovered cases.

4. ANALYTICAL WORK

India has the second most population in the whole world, it is appalling that the mortality rate in India the lowest other than any other country with more than 4 lakh cases. Mortality rate is a measure of the number of deaths in a particular population, scaled to the size of that population per unit of time.

The below graph shows that Delhi has the highest mortality rate, followed by Maharashtra.

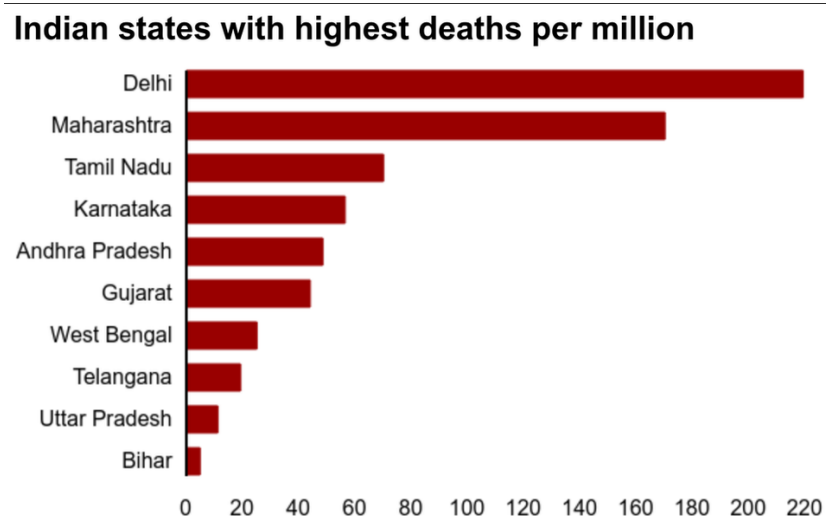


Figure 7 : Graph comparing death cases statewise

The following set of graphs help us to grasp the analysis of the total number of cases with the total number of deaths as well as recovered cases, in the states where the hit of covid-19 was abounded. Data is considered till 26th May 2020.

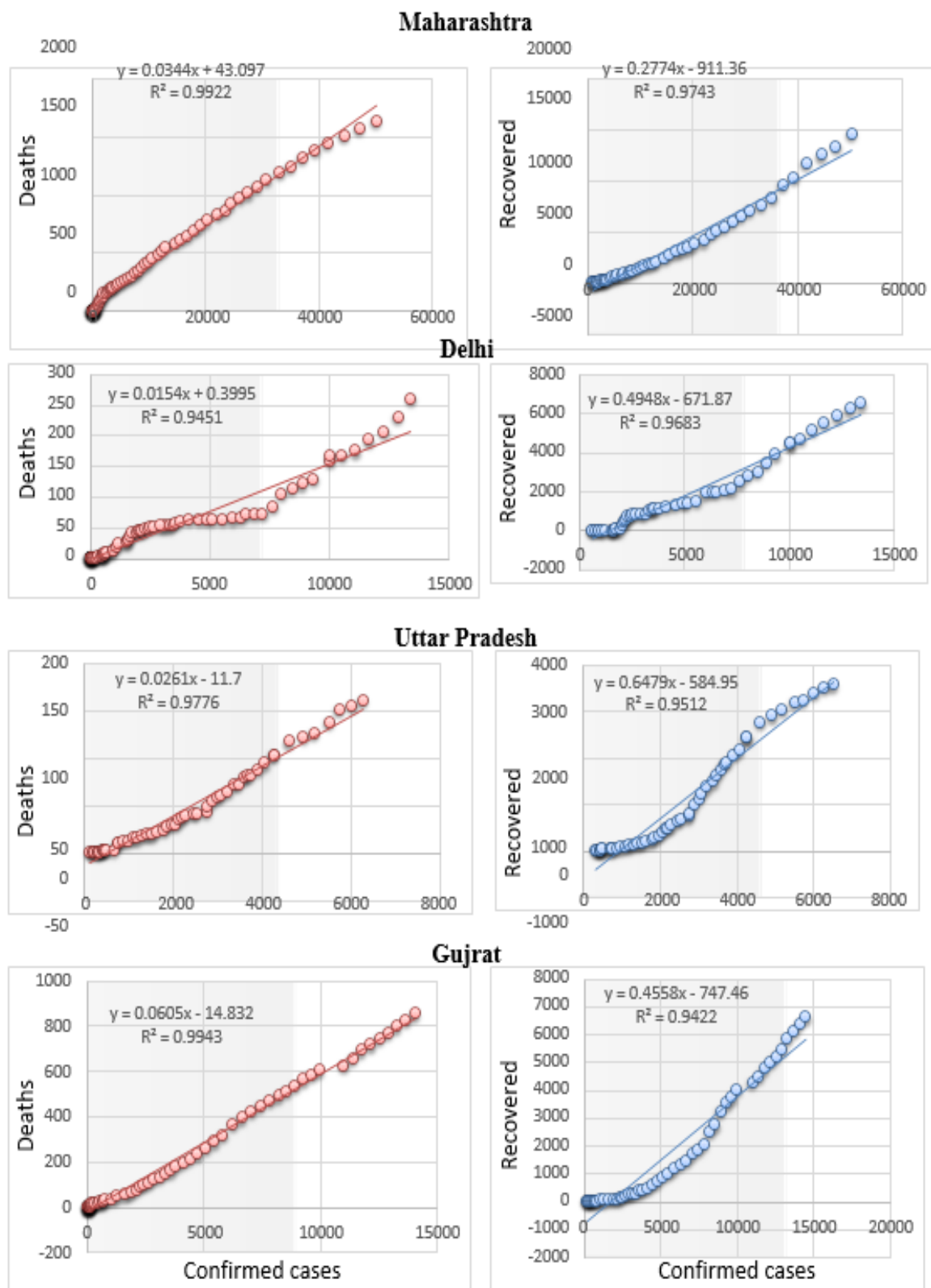


Figure 8 : Comparing highest affected states in India

Timeline of covid related events in India

Date	Event
30 Jan 2020	India's first novel coronavirus patient - a student studying at Wuhan University - was reported in Kerala's Thrissur district
3 Feb 2022	Kerala government declared coronavirus a state calamity after two more cases were reported in Alappuzha and Kasaragod district.
4 Feb 2020	India cancelled existing visas for Chinese and foreigners who had visited China in the last two weeks
12 Mar 2020	First death due to Covid-19 in India, of a 76-year-old man in Karnataka.
15 Mar 2020	100 cases reported in Maharashtra. Cases double every 5 days.
22 Mar 2020	50 days after the virus was first reported in India, a 14-hour voluntary lockdown called 'Janata Curfew' was observed in India.
25 Mar 2020	A nationwide lockdown was imposed till April 14, two days after the 'Janata Curfew'.
2 Apr 2020	129 new cases in Delhi and 74 new cases in Tamil Nadu.
5 Apr 2020	100 confirmed deaths were recorded.

14 Apr 2020	10,000 confirmed cases were recorded, as the nationwide lockdown was further extended till May 3.
29 Apr 2020	1000 confirmed deaths were recorded.
7 May 2020	50,000 confirmed cases were reported.
16 May 2020	India with 85,940 cases overtook China in terms of the total number of cases reported.
19 May 2020	India reported 100,000 cases.
8 Jun 2020	Phased reopening begins after 75 days of lockdown, as India records more than 2,50,000 COVID-19 cases
27 Jun 2020	Total number of covid-19 cases reported are 500,000.

5. BENEFITS, DRAWBACK AND APPLICATIONS

5.1 Benefits of regression analysis

- Making predictions and forecasts.
- Improving business efficiency.
- Supporting business decisions.
- Review how variables impact some phenomena.
- Analysing results and correcting errors.
- Finding new opportunities.
- Forecast future trends and optimize operations.

5.2 Drawbacks of regression analysis

- Assumptions: Your assumptions as a business owner will limit the data you see as significant enough to include in a regression model.
- Wrong questions: When your regression model answers a question that doesn't really impact your business, it won't result in needed change.
- Poor data: If you gather data that is too generalized, too specific or missing pertinent information, your regression model will be unreliable.
- Software limitations: Statistical analysis software is helpful but can be glitchy or not offer enough variable options to suit the specifics of your situation.
- Human error: If your accountant has a tendency to mistype data or forget to link the database to statistical analysis software, this can create inaccurate regression models.

5.3 Applications

- Market research : The brief analysis which is provided with regression has real-world applications in stock market, bonds, etc.
- Expert consultation : Many consultancy companies use well-built regression models for this purpose and have many other applications as well.
- Operation efficiency: Almost all multi-national companies use various regression models for optimization of their business project. They usually get a gut instinct and make well crafted predictions on real data.
- Trend forecasting : It a very useful way of looking at the past operations and market value. We determine the next possible trends from previous trends to predict about the condition of market or sales in the future.

7. FUTURE PROSPECT:

We have done the analysis of the impact of Covid -19 in India and its different states. This particular virus is not foreseeable, as are its various variants which we hear about every few months. It is very important to know beforehand, the overall information about the pandemic situation. Using the regression models, we choose a suitable model for prediction. A befitting model can be used to further improvise the current model and its observations. If some variants of the Covid-19 occur in the near future, we can use this model to predict the number of patients in India and also make use of the data analysis which would help us to gather more insights on the situation that we would face. We can anticipate the further spread of the covid-19 virus, it will help in taking the required actions by concerned authority to control the future spread of Covid-19.

8. CONCLUSION

To summarise, we carried out the regression analysis and investigated the spread of covid-19 in India and its states. Using the results that we have obtained from the analysis of spread of covid-19 across India and its different states, we see that Maharashtra, Delhi, Gujarat, Madhya Pradesh, Andhra Pradesh, Uttar Pradesh, and West Bengal are the most affected states category. Among the remaining states, we see that Tamil Nadu, Rajasthan, Punjab, and Bihar are in the moderate category, whereas Kerala, Haryana, Jammu and Kashmir, Karnataka, and Telangana are in the controlled category. We also depicted the results through many bar diagrams and graph plots.

9. REFERENCES:

- [1]Ministry of Health and Family Welfare of India.
<https://www.mohfw.gov.in/dashboard/index.php>
- [2]Official covid-19 India portal: <https://www.covid19india.org/>
- [3]IEEE paper by Ekta Gambhir, Ritika Jain, Alankrit Gupta, Uma Tomer
<https://ieeexplore.ieee.org/document/9215356>
- [4] Regression Analysis of COVID-19 Spread in India and its Different States
[https://www.researchgate.net/publication/341759652_Regression Analysis of COVID-19 Spread in India and its Different States](https://www.researchgate.net/publication/341759652_Regression_Analysis_of_COVID-19_Spread_in_India_and_its_Different_States)
- [5] Covid 19 pandemic analysis using regression
<https://www.medrxiv.org/content/10.1101/2020.10.08.20208991v1.full.pdf>
- [6] Performance evaluation of regression models for COVID-19
<https://www.sciencedirect.com/science/article/pii/S2090447921003385>

