**SRI SIDDHARTHA ACADEMY OF HIGHER EDUCATION**

***(Declared as Deemed to be University Under Section 3 of the UGC Act, 1956***

**Approved by AICTE, Accredited by NBA, NAAC ‘A’ Grade*)* AGALKOTE, TUMAKURU – 572107 KARNATAKA**



## Project Report On

**“Prediction of Covid-19 cases in Karnataka using**

**Linear Regression”**

**Submitted by**

**VIVEK V R (18CS117)**

**SHREYAS T M (18CS088)**

**THEJASWI K (18CS104)**

**VINOD S C (18CS114)**

**Under the guidance of**

**CHENNABASAVARAJU T P**

**Asst.Professor,Dept of CSE SSIT,Tumakuru**

**In partial fulfillment of**

## BACHELOR OF ENGINEERING IN

**COMPUTER SCIENCE AND ENGINEERING**

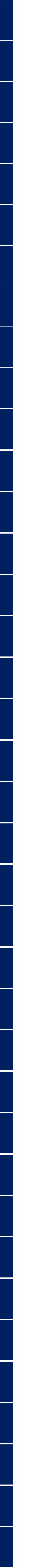
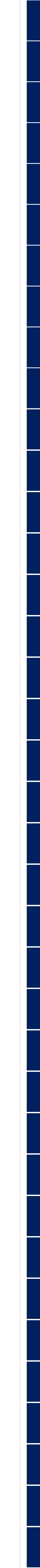
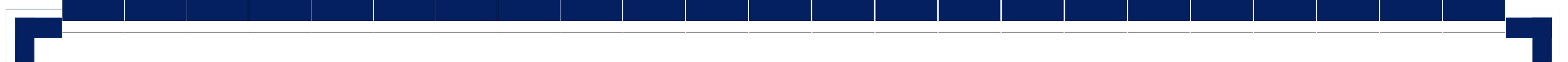
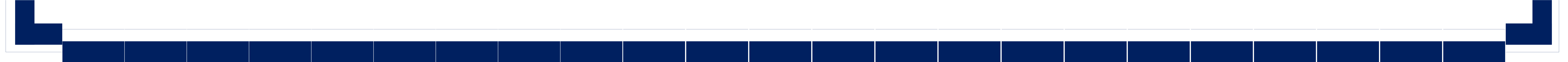
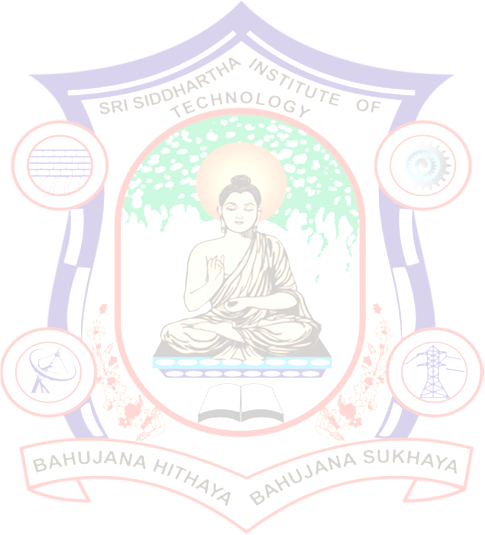
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

SRI SIDDHARTHA INSTITUTE OF TECHNOLOGY

**(A Constituent College of SRI SIDDHARTHA ACADEMY OF HIGHER EDUCATION)**

### MARALUR, TUMAKURU -572105,

**2020-21**



**SRI SIDDHARTHA INSTITUTE OF TECHNOLOGY**

**(A Constituent College of SRI SIDDHARTHA ACADEMY OF HIGHER EDUCATION Approved by AICTE, Accredited by NBA, NAAC ‘A’ Grade)**

**MARALUR, TUMAKURU -572105**

**DEPARTMENT OF COMPUTER SCIENCE ANDENGINEERING**

+

***CERTIFICATE***

*This is to certify that the project entitled* **“Prediction of covid-19 cases in Karnataka using Linear Regression”** *is a bonafide work carried out by* **SHREYAS T M, THEJASWI K, VIVEK V R** *and* **VINOD S C** *in the partial fulfillment for the award of the degree of Bachelor of Engineering in “COMPUTER SCIENCE AND ENGINEERING****”*** *during the academic year of 2020-2021.*

*It is certified that all of the corrections/suggestions indicated for the internal assessments have been incorporated in the report. The project report has been approved as it satisfies the academic requirements with respect to the project work prescribed for the award of the degree of Bachelor of Engineering in* ***Computer Science and Engineering.***

|  |  |  |
| --- | --- | --- |
| **Signature of the Guide** | **Signature of the HOD** | **Signature of the Principal** |
| **Chennabasavaraju T P** | **Dr. M Siddappa** | **Dr. M.S.Raviprakasha** |
| B.E., M.Tech., | B.E., MTech., Ph.D., MISTE |  |
| Asst. Professor,  Dept. of CSE, SSIT, Tumakuru. | Prof & HOD,  Dept. of CSE,  SSIT, Tumakuru. | Principal  SSIT, Tumakuru |

**Project Associates:**

**SHREYAS T M (18CS088)**

**VIVEK V R (18CS117)**

**THEJASWI K (18CS104)**

**VINOD S C (18CS114)**

**External Examiners:- Signature with date:-**

**1.**

**2**

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without mentioning the people who made it possible and their support had been a constant source of encouragement which crowned our efforts with success.

We are deeply indebted and would like to express my sincere thanks Dr. M S Raviprakasha, Principal, S.S.I.T, Tumakuru, for providing me an opportunity to carry out this project work.

We express our deep sense of gratitude to Dr. M SIDDAPPA HOD, Department of Computer Science and Engineering , S.S.I.T, Tumakuru, for his guidance, constant encouragement and wholehearted support.

Our special thanks to CHANNABASAVARAJU T P, Asst. Professor, Department of CSE S.S.I.T, for his guidance, constant encouragement and wholehearted support.

Finally we would like to express our sincere thanks to all the staff members of Computer Science Department, S.S.I.T. for their valuable guidance and support.

PROJECT ASSOCIATES **SHREYAS T M (18CS088)**

**VIVEK V R (18CS117)**

**THEJASWI K (18CS104)**

**VINOD S C (18CS114)**

**CONTENTS:- Page Number**

**1. Introduction………………………………………………………. 6 – 8**

**2.Literature Review…………………………………………………. 9**

**3.Objectives………………………………………………………….. 10**

**4. Methodology………………………………………………………. 10**

**5.Algorithm…………………………………………………………... 16 - 17**

5.1 Linear Regression

**6.Facilities required for proposed work…………………………… 18**

7.1 Software Requirements

7.2 Hardware Requirements

**8.Conclusion………………………………………………………… 18**

**9.Bibliography………………………………………………………. 18**

**ABSTRACT**

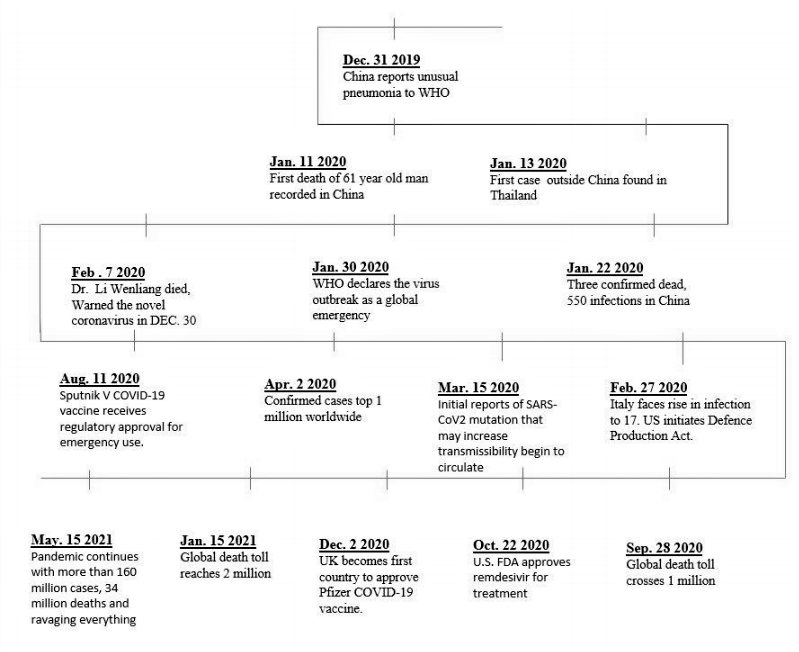
COVID-19 pandemic has affected the economy and changed the human way of life, disrupting everyone's mental, physical, and financial well-being. Many of the fastest-growing economies are strained owing to the severity and communicability of the epidemic. Because of the increasing diversity of cases and the resulting burden on healthcare practitioners and the government, therefore, predicting the number of infected COVID-19 cases which could be useful in planning the required hospital resources in the future. In this paper, we focussed on information-led methods of estimating the numbers of COVID-19 confirmed cases in the country and their implications in the future, using Linear Regression model . Hence, this work would help the decision makers to understand the upcoming of the pandemic trajectory in the country and take necessary actions for the effect of interventions.

Keywords: COVID-19, Coronavirus disease, Data Analysis, Visualization, Linear regressionmodel

**1.INTRODUCTION:-**

At the end of December 2019 in Wuhan, China, it was first reported that a human infection was caused by a novel coronavirus (nCov) or Wuhan virus or 2019-nCov . One of the biggest challenges of this epidemic is the human-to-human transition of nCov. The coronavirus (COVID-19) infected cases increase at an exponential rate worldwide. On 30 January 2020, the World Health Organization (WHO) issued a worldwide health emergency warning notice , labeling that 2019-nCoV is of urgent global concern. The disease and mortality rates for the COVID-19 are uncertain at the early stage especially for young ones and aged people. WHO has estimated the reproduction factor of nCov is 2.7. In demand to control the extensive and quick spread of the nCov, public health sectors took reliable preventative measures and imposed curfew or lockdown infested cities in China, United States, India, and other countries also. This is to limit the social distance between people and to avoid the broadcast of this novel virus via humans to humans.

The best way to prevent and slow down transmission is maintaining social distance. We have to protect our self and others from infection by washing our hands or using sanitizers and avoid touching face. The number of COVID-19 cases in India are 67,161 and the death toll is 2,212 by 11th MAY 2020, as per the Worldometer data. Worldwide 4,180305 people have been attacked by virus and the total number of deaths caused by disease now are 283,865. There are very less number of COVID-19 test kits available in hospitals which are not at all sufficient for the increasing cases. Artificial Intelligence is actually dominant tool in the fight against the COVID-19 crisis. AI has subdomain like Machine Learning .It helps in diagnose and predict COVID-19. ML Techniques are useful in tracking COVID cases and predicting.The aim of this project is to develop a prediction model to predict the covid cases.

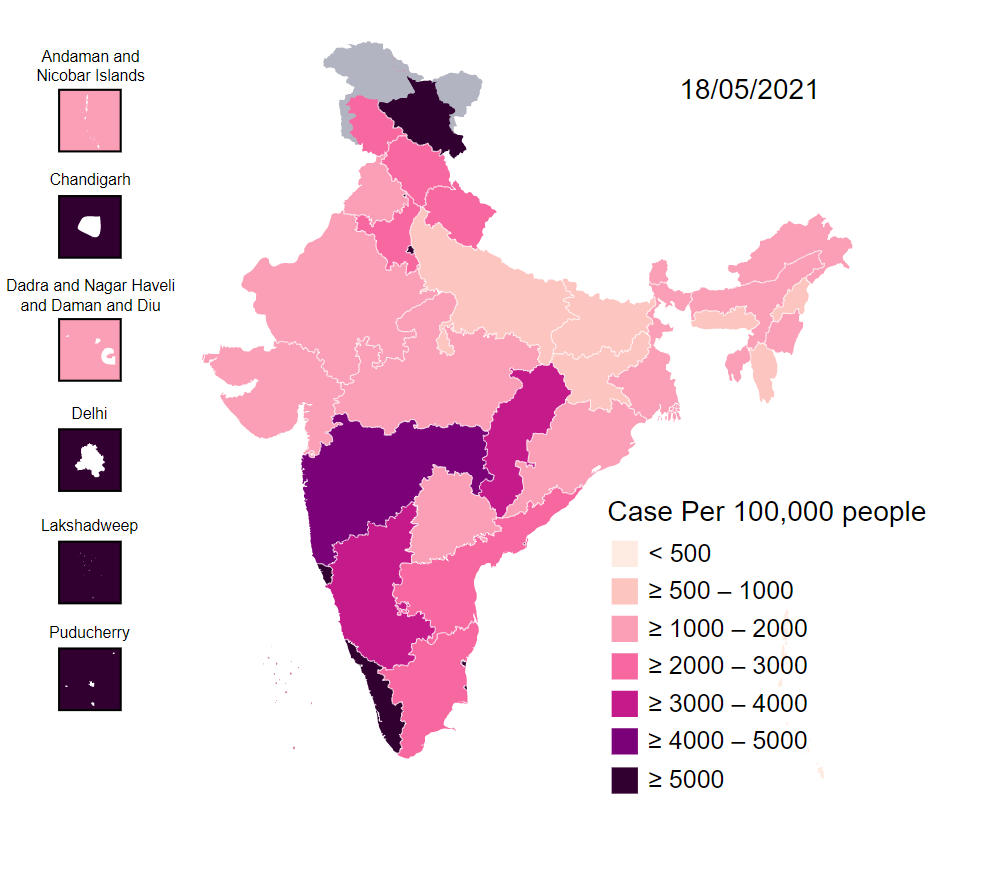


The prediction of covid-19 cases is playing an important role during the ongoing COVID-19 pandemic. The first case of a COVID-19 affected patient was reported in Wuhan, China. Within one month of the outbreak, the number of positive cases and deceased rose at an exponential rate. As the pandemic situation has taken over the world , the main goal of this project is to develop a machine learning model that could predict number of COVID cases of a state in future. To develop such a model, a literature study alongside an experiment is set to identify a suitable algorithm. To assess the features that impact the prediction model.

**ADVANTAGES:-**

Following are the benefits or advantages of predicting the covid cases:

* Prediction of COVID-19 by using Machine Learning could help increase the speed of disease identification resulting in reduced mortality rate.
* Analyzing the results obtained from experiments, linear regression was identified to perform better compared to other algorithms.



**2. LITERATURE REVIEW**:-

COVID-19 is a communicable disease that has been declared as a pandemic by the WHO. Moreover, there is no medicine or vaccine available to cure this infection as of now. Hence, the only way to protect oneself from this pandemic is to get protected from the contact of an infected person. With the ongoing pandemic threat, researchers started to study the future of COVID-19. Based on the predictions, resources can be prepared to treat people and minimize fatalities.

Since the beginning of COVID-19, various researchers have predicted its spreading trend for different countries and their states. linear regression is a technique that have been used for this study.Linear regression has been used to predict the number of confirmed cases in a state of a country.By this prediction we can be getting a knowledge to how get protection and preventing the disease by all means.

In this paper, a compartmental epidemiological model, named linear reggression model, has been proposed for the prediction of COVID-19 peak in India and their states with the highest number of total cases. Further, short-term predictions have also been computed using this model, and the results are compared with the real world cases.



**3.OBJECTIVES**

We aim to analyze data on the number of infected people in each Indian state (restricted to only those states with enough data for prediction) and predict the number of confirmed cases in that state. We hope that such state wise predictions would help the state governments better channelize their limited health care resources.

The main objectives are:

• Identifying the most suitable machine learning technique for prediction, to perform on clinical reports of patients.

• Preparing a machine learning model that could make accurate predictions of COVID-19 in patients.

• Identifying the features that affects the prediction of COVID-19 in patients.

**4.Methodology**

Data was collected in a CSV file and uploaded in Jupyter notebook and analysed with the Python 3.8.2 software.

Input: Total number of infected cases, active cases, recovery numbers.

Output: Predicted confirmed cases of a state.

In order to get a good predictive value data was analysed for all infected states within India.We use linear regression algorithm ,Regression analysis has been used to understand the effect of Covid-19 on various Indian states. India is a country where Initially Covid-19 infection was controlled through lockdown of entire country and later by restricted lockdown. As the number of cases increased, depending on clusters lockdown is imposed.

* As we are analyzing loads of data, we are trying to do it in a methodological approach by using the following steps
* First, we imported the COVID19 dataset and prepared it for analysis by dropping columns and aggregating rows.
* Choosing on and evaluating a good measure for our analysis.
* Visualizing our analysis results using Matplotlib,Seaborn.

**5. Project overview**

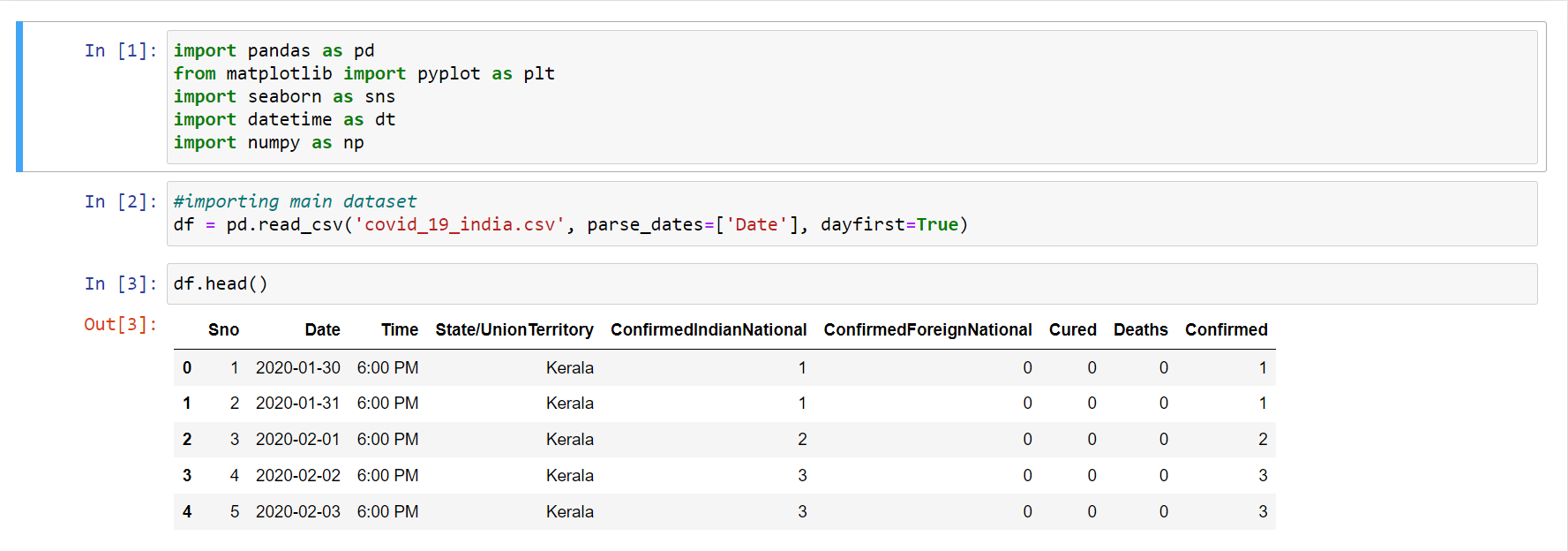


Figure 1: Importing libraries and reading dataset

LIBRARIES USED:

* Pandas - It is a python package that provides expressive data structures designed

to work with both relational and labelled data. It is an open source

python library that allows reading and writing data between data structures.

* Numpy - It is an open source python package for scientific computing. Numpy

also adds fast array processing capacities to python.

* Matplotlib - It is an open source python package used for making plots and

2D representations. It integrates with python to give effective and interactive

plots for visualization.

* Sklearn - It is an open source python machine learning library designed to work alongside Numpy. It features various machine learning algorithms for classification, clustering and regression.
* Seaborn –It is a library in python predominantly used for making statistical graphics and is a data visualization library built on top of matplotlib and closely integarated with pandas data structures in python.

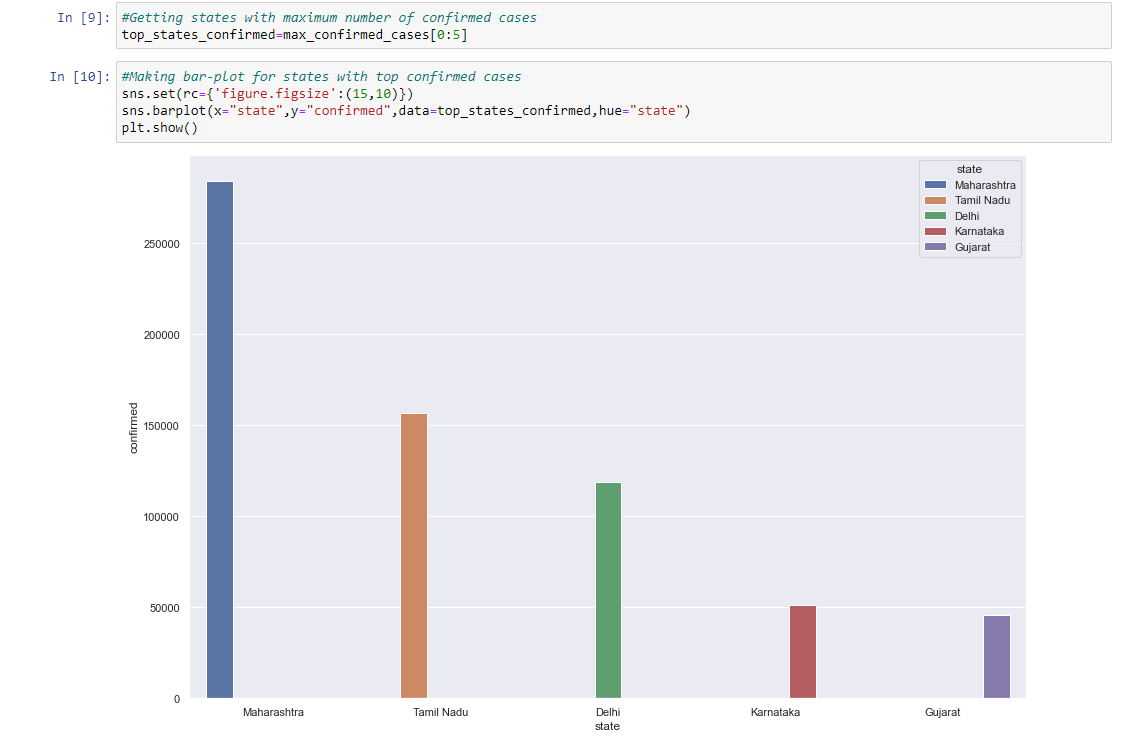


Figure 2: Plotting a bar graph for top 5 states with maximum number of confirmed cases



Figure 3: Plotting a bar graph for top 5 states with maximum number of death cases



Figure 4: Plotting a bar graph for top 5 states with maximum number of cured cases

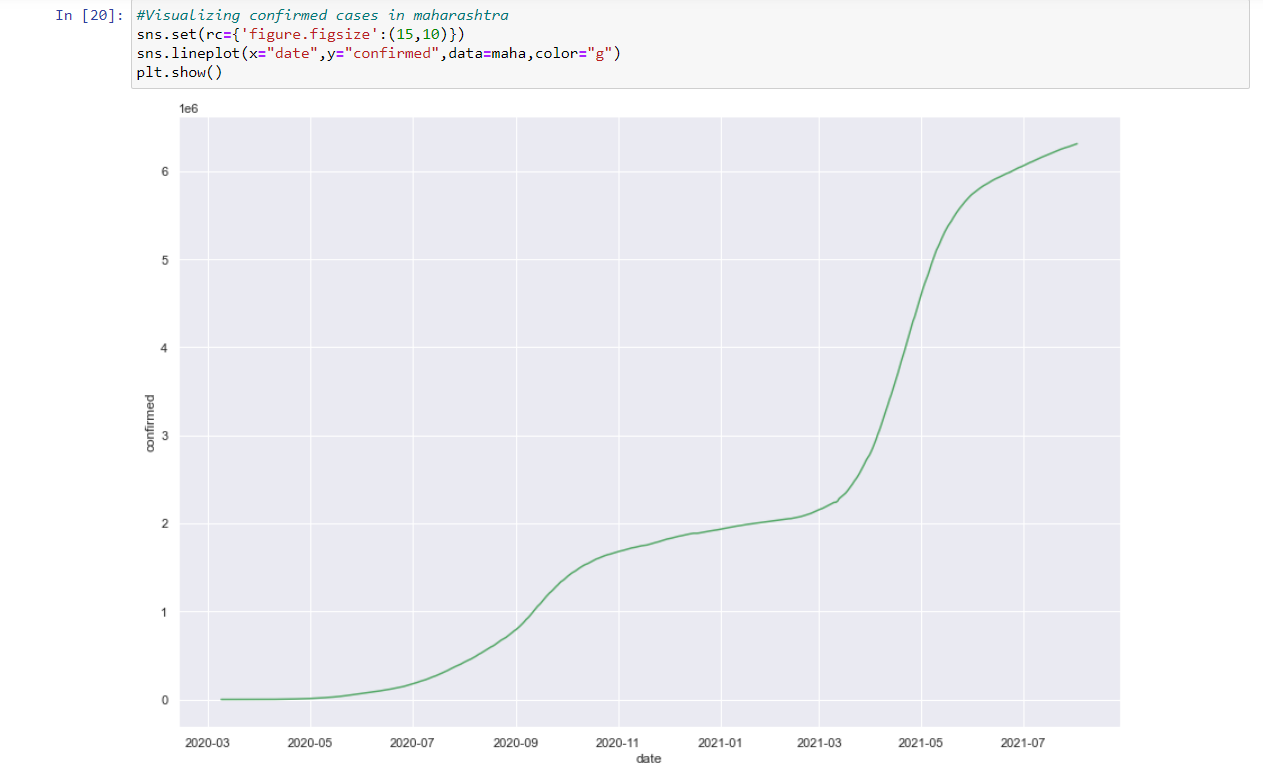


Figure 5: Plotting a line graph for confirmed cases in Maharashtra

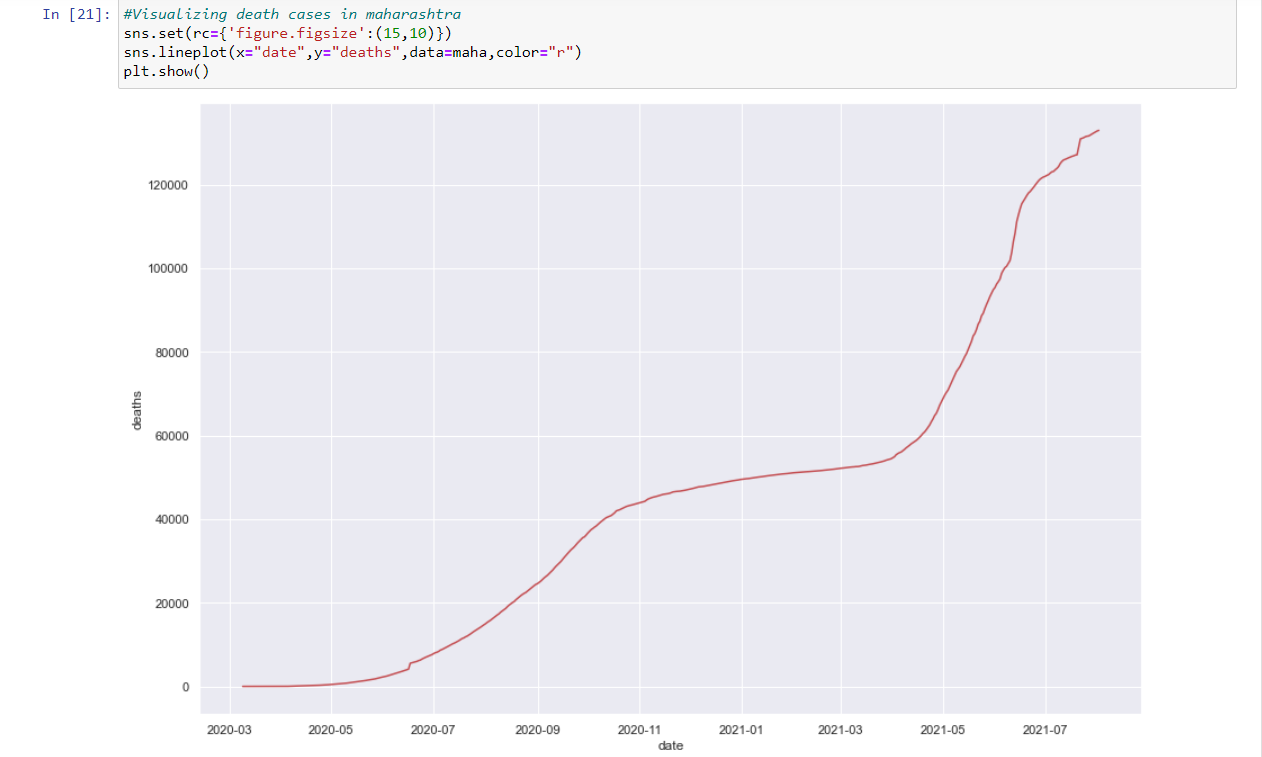


Figure 6: Plotting a line graph for death cases in Maharashtra

THE STATE-WISE TEST DETAILS OF ALL STATES IN INDIA

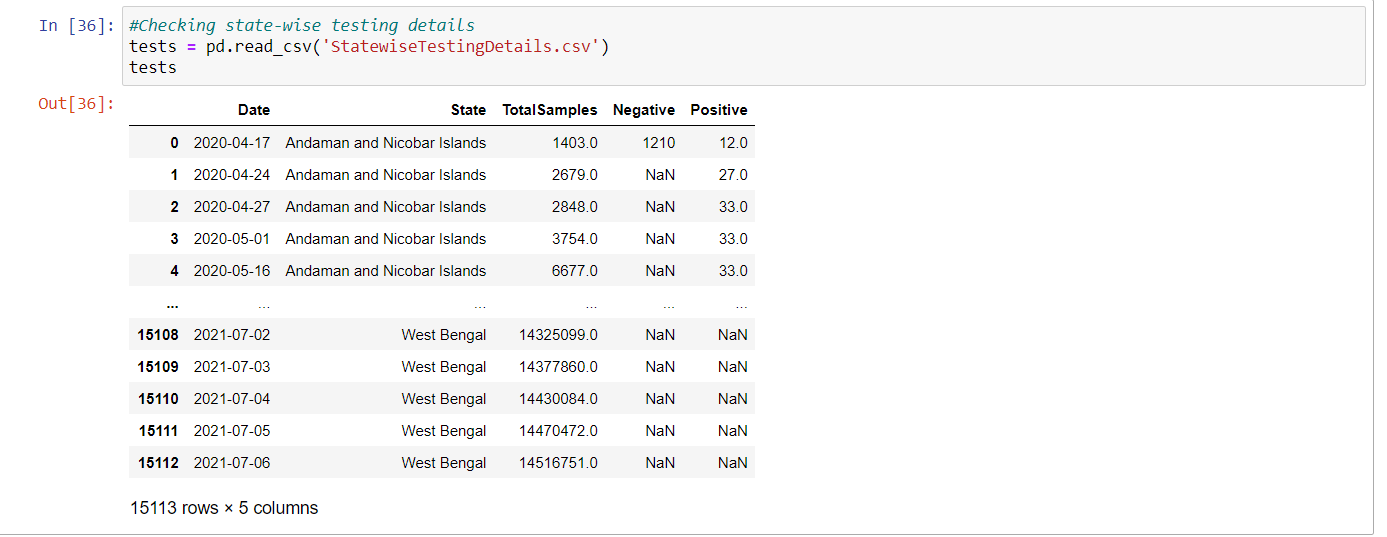


Figure 7: Reading the dataset of state-wise testing details of all the states in India

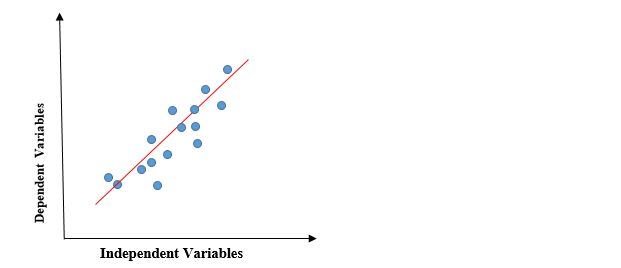


Figure 8: Plotting a bar graph for states with maximum number of test cases

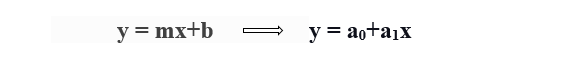
## 6. ALGORITHM USED

## 6.1Linear Regression

Linear regression is a quiet and simple statistical regression method used for predictive analysis and shows the relationship between the continuous variables. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression. If there is a single input variable (x), such linear regression is called **simple linear regression**. And if there is more than one input variable, such linear regression is called **multiple linear regression.** The linear regression model gives a sloped straight line describing the relationship within the variables.



The above graph presents the linear relationship between the dependent variable and independent variables. When the value of x (**independent variable**) increases, the value of y (**dependent variable**) is likewise increasing. The red line is referred to as the best fit straight line. Based on the given data points, we try to plot a line that models the points the best.

To calculate best-fit line linear regression uses a traditional slope-intercept form.y= Dependent Variable.  
  
x= Independent Variable.  
  
a0= intercept of the line.  
  
a1 = Linear regression coefficient.

### STEPS TO IMPLEMENT TO LINEAR REGRESSION

### Import some required libraries.

### Define the dataset.

### Plot the data points.

### Plot the regression line.

### Predicting the values.

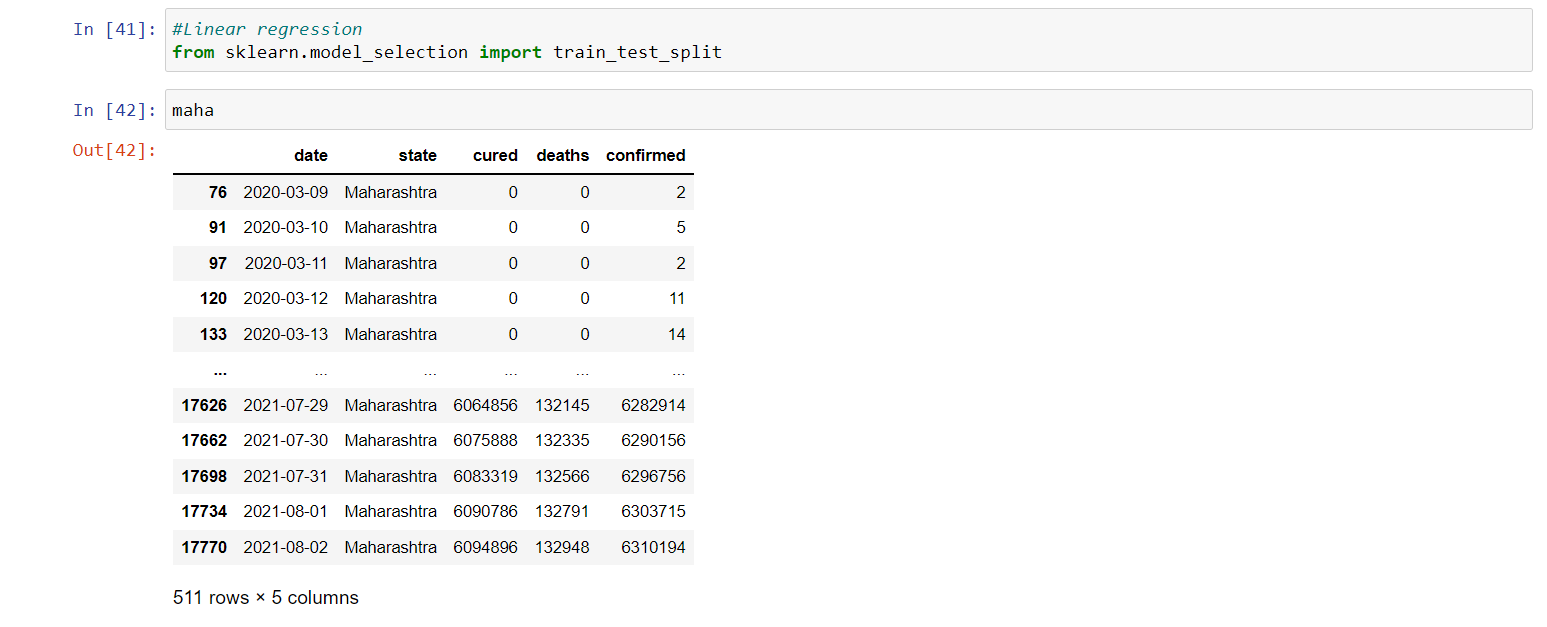


Figure 9: Implementing linear regression using sklearn model



Figure 10: predicting the confirmed cases using linear regression

**7.FACILITIES REQUIRED FOR PROPOSED WORK**

**7.1 SOFTWARE REQUIREMENTS**

* Jupyter Notebook or any Python Idle
* Pandas
* Numpy
* Matplotlib
* Seaborn
* Sklearn

**7.2 HARDWARE REQUIREMENTS**

* Any Operating System with latest version.
* i5 or i7 processor
* RAM 4 or 8GB

### 9.CONCLUSION

In this study, a linear regression model has been used for the COVID-19 predictions. A comparative study shows that it performs better than the other existing model and hence can be used for future prediction in better accuracy. In addition, we analyzed patterns with the confirmed, death, recovered cases. In the future, this study will be further continued with much more accurate algorithms and updated datasets.

**8.BIBLIOGRAPHY**

* **<https://www.kaggle.com/sudalairajkumar/covid19-in-india>**
* [**https://www.greatlearning.in/academy/learn-for-free/courses/predicting-covid-19-in-india-using-machine-learning**](https://www.greatlearning.in/academy/learn-for-free/courses/predicting-covid-19-in-india-using-machine-learning)
* **https://www.tutorialspoint.com/machine\_learning\_with\_python/regression\_algorithms\_linear\_regression.htm**