

## COVID-19 CASES PREDICTION FOR NEXT 30 DAYS USING MACHINE LEARNING

**P.BhagyaSri, D. JessyMercedes, H.ChaitanyaReddy, A.LasyaPriyaReddy,P.Y.M.NaveenKumar**

**Abstract:**This project deals with the analysis, or prediction of pacing COVID cases by using machine learning algorithms. The goal is to provide the users with the information about cases in and around with the help of a web application.so whenever someone needs to travel, they would know-how safe or unsafe the desired place could possibly be to make a way to dexterity.The spread of covid-19 in the whole world has put the humanity at risk. the resources of some of the largest economies are stressed out due to the large Infectivity and transmissibility of this disease. Due to the growing magnitude of number of cases and its subsequent stress on the administration and health professionals, sum prediction methods would be required to predict the number of cases in future.The highly infectious corona virus disease covid-19 was first detected in Wuhan, China in December 2019 and subsequently spread to 212 countries and Territories around the world, infecting millions of people.in India, a large country about 1.3 Billion people, the disease was first detected on January 30,2020, in a student returning from Wuhan .the total number of confirmed infections in India

**Keywords:**COVID-19, prognosis of cases, prediction.

\* Correspondence Author

P.BhagyaSri, Assistant Professor,Department of CSE,  
Usha Rama college of Engineering and Technology,  
Email:bejagam.bhagyasri@gmail.com,

D.Jessy Mercedes, Department of CSE,  
Usha Rama college of Engineering and Technology,  
Email:jessymercedesd@gmail.com

BH.Chaitanya Reddy,Department of CSE,  
Usha Rama college of Engineering and Technology,  
Email:chaitanyareddybhimavarapu@gmail.com

A.LasyaPriyaReddy, Department of CSE,  
Usha Rama college of Engineering and Technology,  
Email:priyalasya18@gmail.com

P.Y.M.NaveenKumar, Department of CSE,  
Usha Rama college of Engineering and Technology,  
Email:puttiyaswanth@gmail.com

India.

## 1. INTRODUCTION

The number of cases globally as well as in India is increasing at a very rapid rate. As is evident from the data, Maharashtra is the worst-affected state in terms of total cases, which accounts for about 21% of the cases in India. The next three worst-affected states/union territories are Andhra Pradesh, Tamil Nadu, and Karnataka having approximately 32% of the total cases, and the rest of Indian states/ union territories having another 47% cases. North-eastern states of India are much better, like Mizoram and Sikkim, each having less than 2000 cases so far. To understand the future spread of the pandemic and to devise management strategies, various models have been designed, which provide information regarding the time of attainment of infection peak, the number of infected cases, and the requirement of medical infrastructure to manage the spread. The proposed model utilizes the reported data of infections, recoveries, and deaths caused by COVID-19 to make predictions.

## 2. DESIGN

The Covid-19 cases prediction for next 30 days using machine learning consists of the following modules. Modules used are:

**Data Preprocessing:** It includes 6 steps. Data preprocessing is about taking the raw data and making it useful for the Machine Learning Model. The first step is to analyze the problem statement, next it is important to identify the input and output variables, further there is a necessity to identify the missing elements and convert the categorical data into non-categorical data then to perform train and test data.

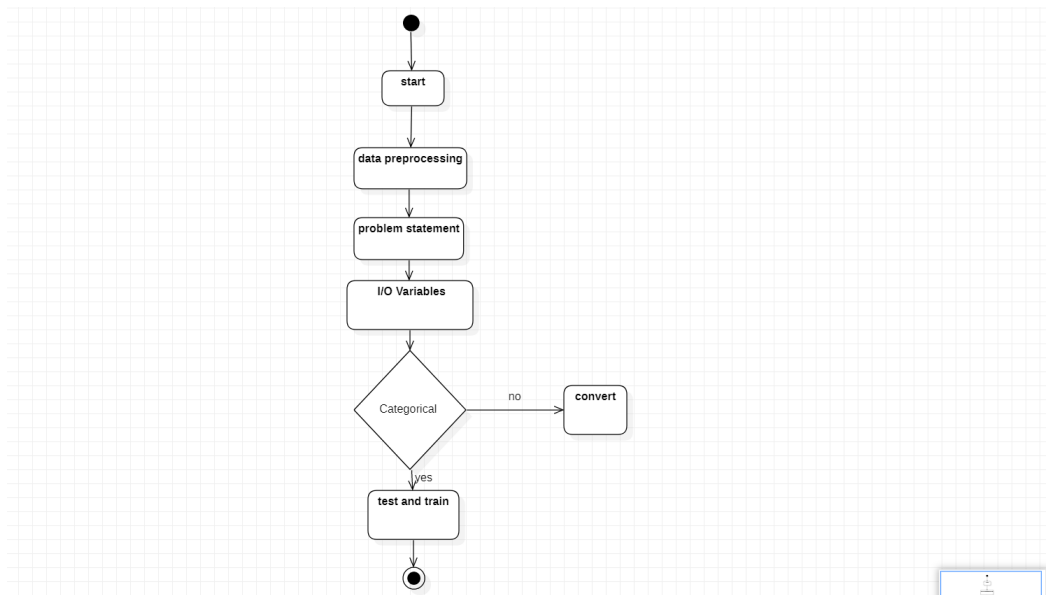
**WebPage:** Using the strongest languages like python, HTML, CSS, JS, PHP the webpage is built to demonstrate the covid cases for the next 30 days, we derive the plots whether it's a bar plot or a line plot. Whereas the webpage also includes the trending information about the new variants and vaccines available in India, and also the number of cases that purged throughout India. 4 states are represented, displaying both confirmed and predicted cases based on the linear regression model.

**Linear Regression:** In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. we have started our linear regression model because accuracy 88%.

Linear regression is an attractive model because the representation is so simple.

The representation is a linear equation that combines a specific set of input values ( $x$ ) the solution to which is the predicted output for that set of input values ( $y$ ). As such both the input values ( $x$ ) and the output value are numeric. The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta ( $B$ ). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

### 3. Figure

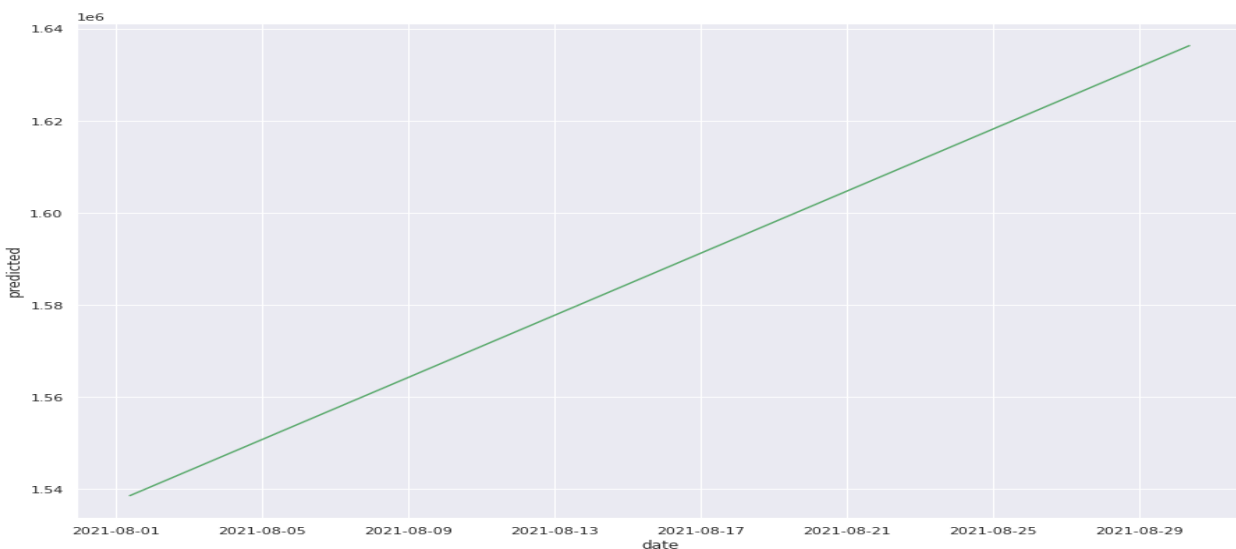


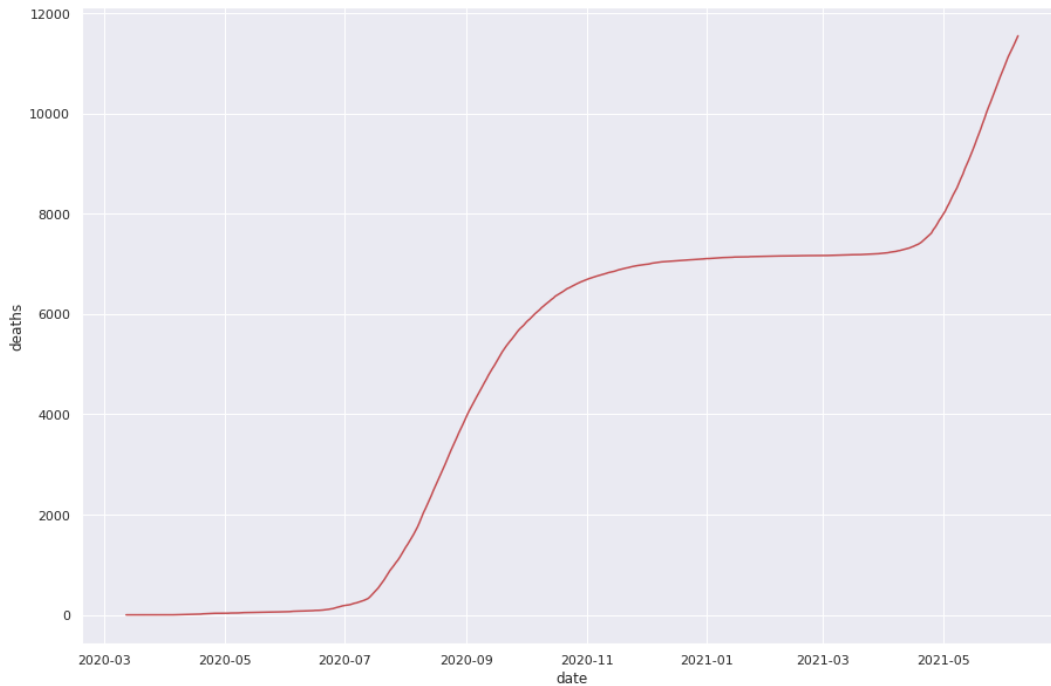
### 4. ANALYSIS

There will be analysis for every state in India, taking the independent variable as date and dependent as confirmed cases, the prediction using train and test data will be made by using the linear regression algorithm.

### 5. RESULTS

By taking the independent and dependent variable as an input and target variable. We train and test data by using train split, it helps in segregating the data into two subsets and then by using  $x\_pred$  and  $y\_pred$  it is used to predict the number of cases in any of the states. This ordinal encoding transform is available in the scikit-learn Python machine learning library via the Ordinal Encoder class. By default, it will assign integers to labels in the order that is observed in the data. If a specific order is desired, it can be specified via the “categories” argument as a list with the rank order of all expected labels.





## 6CONCLUSIONS

In this method we have discussed how the machine learning algorithm could help in providing the futuristic data by taking the current and historic data available. The dataset is considered since the pandemic outbreak has begun up to date and the dataset is taken in the format of csv, it is crucial to know the predicted cases in states of India as it worsens in the peak of year in the country. Linear model is applied to predict as it is stable, most of the linear models are stable in nature as they won't have much variance when the changes are made to the training data. Whereas the tree models are different at root so are they different at child's nodes, hence less accuracy.

## REFERENCES

- [1] Patrikar S., Poojary D., Basannar DR., Kunte R. Projections for COVID-19 and Evaluation of Epidemic Response: Strategies for India. Personal Communication.
- [2] Ranjan R. Predictions for COVID-19 outbreak in India using Epidemiological models
- [3] Chakraborty T., Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis.
- [4] Mazumder A., Arora M., Bharadiya V. Study of epidemiological characteristics and in-silico analysis of the effect of interventions in the SARS-CoV-2 epidemic in India.