

# Predictive Analysis and Forecasting of Covid-19 Outbreak

**PES Institute of Technology and Management, Shivamogga.  
Department of Information Science & Engineering.**

Ms. Akanksha J (akankshaj1999@gamil.com),  
Ms. Akshatha S M (akshuyash330@gmail.com),  
Ms.Rajeshwari M Hebballi (rajeshwarimmhh77@gmail.com)

Guide: Vinay S K (vinaysk@pestrust.edu.in)

***Keywords: COVID-19, Machine Learning, Analytics, Visualization, regression models, Forecasting.***

## INTRODUCTION

The novel coronavirus disease 2019 (COVID-19) pandemic caused by the SARS-CoV-2 continues to pose a critical and urgent threat to global health. The Outbreak of Novel CoronaVirus has created a disastrous situation in many countries around the world and caused millions of deaths. Prediction of the future trend of the disease can be useful for managing the outbreak. In our project the machine learning guided linear regression model Multivariate Regression Model has been used to predict the future cases. Both the models have been fitted into the dataset to deal with the total number of confirmed, recovered, and death cases. Several outbreak prediction models for COVID-19 are being used by officials around the world to make informed decisions and enforce relevant control measures. Among the standard models for COVID-19 global pandemic prediction, simple epidemiological and statistical models have received more attention by authorities, and these models are popular in the media. Due to a high level of uncertainty and lack of essential data, standard models have shown low accuracy for long-term prediction.

## OBJECTIVES

Since the start of the pandemic, the whole world is facing the problem of ever-increasing COVID-19 cases. Through the data analysis one can analyze how countries all over the world are doing in terms of controlling the pandemic and lowering the graph. Analyzing data leads to adaptation to the prevention model. Predictions are made with the dataset available, this helps to decide how far it is able to control the pandemic or up to what extent preventive measures can be taken. To achieve this, the objectives are broken down into different tasks;

- To Analyze and Visualize COVID-19Data.
- To Forecast the COVID-19 cases using Time Series Analysis.
- To build a model that will predict the next n day's Coronavirus cases.

## METHODOLOGY

Initially we have collected the COVID-19 Worldwide and India's dataset from the available resource. The worldwide dataset contains five columns like date, name of different countries, confirmed, recovered and death cases, similarly India's dataset contains same number of columns but the country column is replaced by name of different states of India.

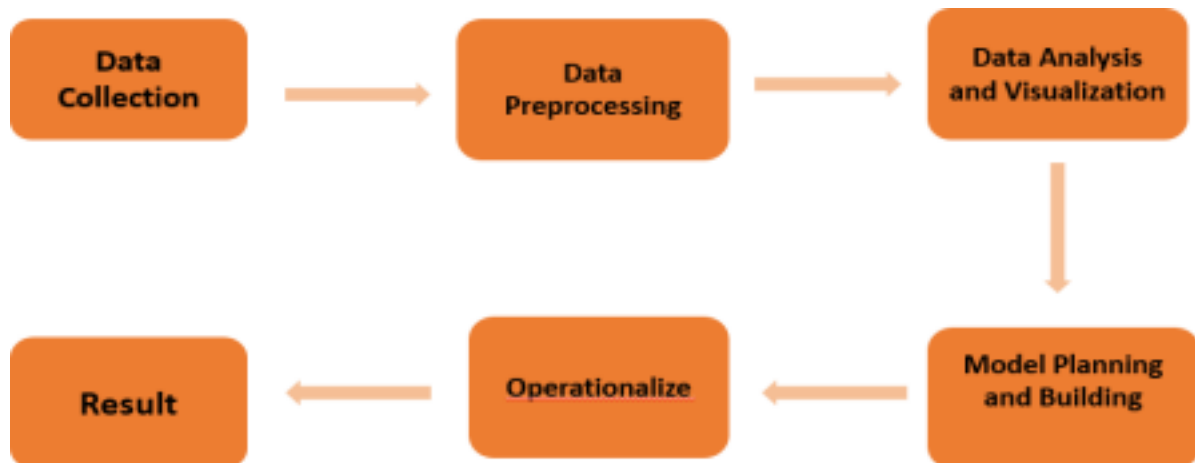


Fig: Methodology.

Further the collected data is preprocessed using various preprocessing techniques like data cleaning, data reduction and data transformation. The preprocessed data is analyzed and visualized and later fed into the Linear Regression Model. The data samples are then fed to the model as a training and testing set. Finally the trained model predicts the probable number of COVID-19 cases of upcoming days.

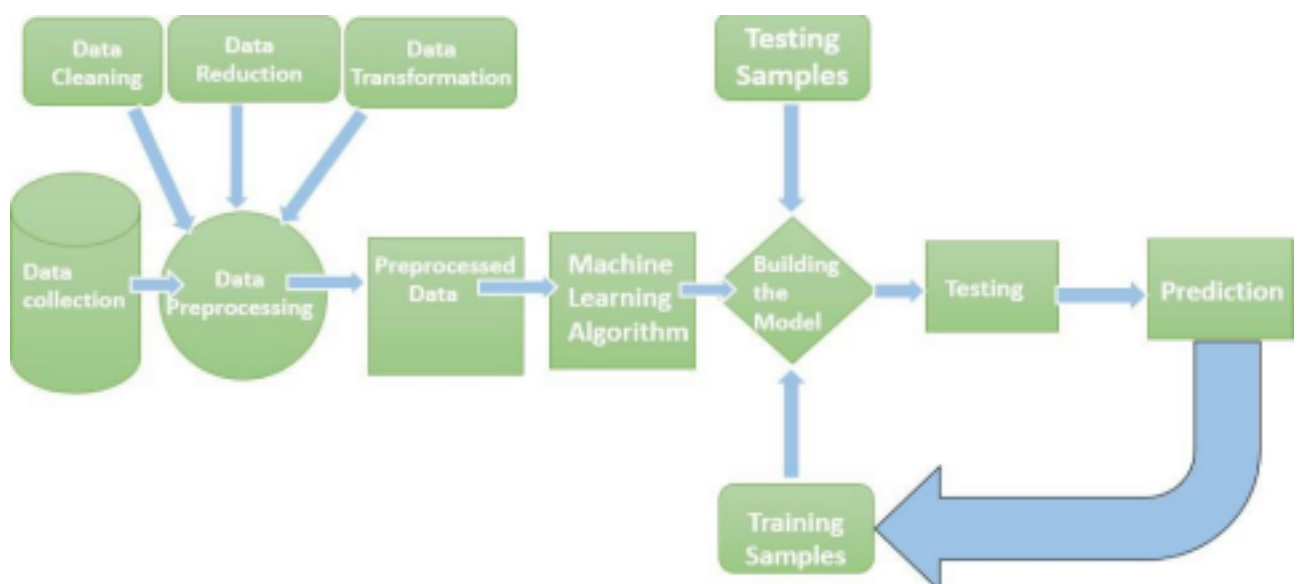


Fig: Architecture.

The requirement specification for

Our project includes software requirements like Windows 7 operating system, Google colab a web application and web browser. And hardware specification includes processor core i3, 4GB RAM and minimum of 3GB disk space.

## **RESULT AND CONCLUSION**

In this project, the analysis of COVID-19 data is performed and the pandemic spread is compared between different countries. Data Visualization techniques are applied to give a clear look on the trend of the data, such as how the virus is spreading, which countries are getting affected mostly and how different countries are recovering. On applying Linear Regression Model and Multivariate Regression Model the possible number of COVID-19 cases are obtained. Accuracy of the model has been calculated for both the models using cross validation function in the prophet model. Among Linear and Multivariate Regression Models, the linear model has performed well and has shown better accuracy than the multivariate model. By considering different or more features, a multivariate model can be improvised.

## **SCOPE FOR FUTURE WORK**

The global pandemic coronavirus 2 (SARS-CoV-2) has become the major issue of many nations. Advancement of accurate prediction models for the outbreak is essential to provide insights into the spread and consequences of this infectious disease. This project presents a comparative analysis of Machine Learning models to predict the COVID-19 outbreak. For the advancement of higher performance models for long-term prediction, future research should be devoted to comparative studies on various Machine Learning models for individual countries. Due to the fundamental differences between the outbreaks in various countries, advancement of global models with generalization ability would not be feasible. As observed and reported in many studies, it is unlikely that an individual outbreak will be replicated elsewhere. For future research, modeling the mortality rate would be of the utmost importance for nations to plan for new facilities. For future research, integration of the Machine Learning model is suggested to enhance the existing standard epidemiological models in terms of accuracy and longer lead time. There is a lot of scope for Machine Learning in Healthcare. For Future work, it is recommended to work on calibrated and ensemble methods that could resolve quirky problems faster with better outcomes than the existing algorithms. Healthcare prediction is an essential field for the future. A prediction system that could find the possibility of outbreak of novel diseases that could harm mankind through socio-economic and cultural factor consideration can be developed. In the future, the competence of the postulated method could be investigated and some other regression models or algorithms can be used and evaluated.