# PREDICTION OF COVID-19 CASES USING MACHINE LEARNING

**GOWSHIKA P[1], HARSHINI K[2], SUBASH R[3], MONISHA R[4], SANGEETHA K[5]**

*[1-3]UG Scholar, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.*
*[4]Assistant Professor, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.*

---

**Abstract:** *In 2019, a very serious pandemic virus COVID -19 started spreading, which has globally infected over 31 million people. Many people were infected by the virus few were recovered but the death rate is on peek. In India, the situation is still paints grim picture and it is also an alarming factor for the whole world. In such case, the complete understanding of the disease and its rate of spreading is vital one. This infectious growth is need to be analyzed to take the preventing measures for the people against this pandemic. If the rate of spreading is calculated in accurate manner, then the protection measures can be proposed easily by the government. To analyse the rate of spreading, Machine learning plays a vital role. Machine learning proved itself a prominent field to solve many complex real world problem. This project illustrates how the machine learning algorithm predict the number of upcoming cases with the help of historical data of Corona virus. These algorithms will predict the upcoming number of newly infected cases, recoveries and death. The performance of these algorithms will be analyzed based on their prediction accuracy and efficient approach will be proposed.*

***Keywords: COVID-19 cases, Machine Learning algorithms, Prediction***

## 1. Introduction

The machine learning is a subset of an artificial intelligence AI where It plays a major vital role in real world. Machine learning is being one of the top most technology, where the system automatically train itself to learn from the statistical data and can be forecasted. The severe acute respiratory syndrome corona virus also known as novel corona virus COVID-19 is an infectious and contagious disease found in Wuhan, Japan.The rate of spreading is increased gradually all over the world where the people affected and the world is stumbled by its spreading rate resulting to the complete shutdown. In order to overcome the scenario and to prevent ourselves from the spreading, the calculating rate should be known in accurate manner where machine learning plays a vital role through the algorithm. The several algorithms are used to measure the statistical data and to be trained where it shows its prediction method.

### 1.1 Methodology

The main objective of this paper is to select the best machine learning algorithm for forecasting the spread of corona virus infected cases with the help of statistical data of infected, active and the death rate in India. The best machine learning algorithm can be selected by its accuracy and time taken for forecasting. Its done by using WEKA software.

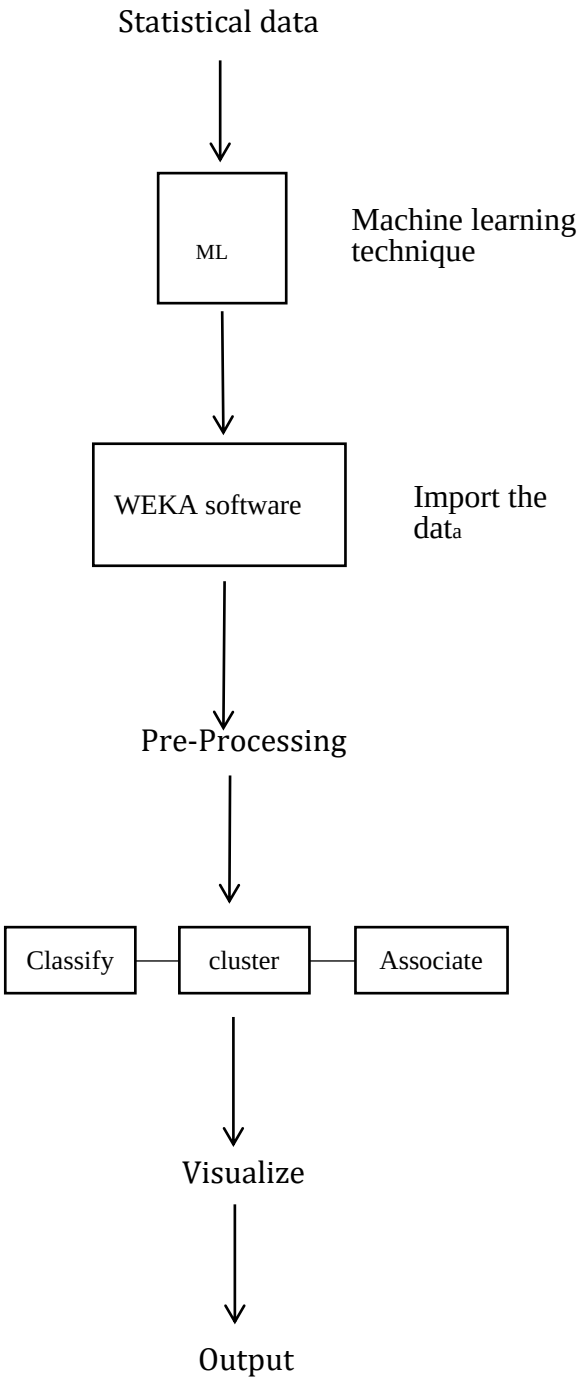
### 1.2 WEKA software

WEKA is a software, is environment for knowledge analysis developed at the university of Waikato, New Zealand. WEKA is a Graphical User Interface (GUI), which contains collection of machine learning algorithms especially for mining the data.

It is a tool for pre-processing, regression, clustering and association of the data. Initially the statistical data of Covid-19 cases are fed into the software for training the machine. By the training, the system can be able to decide the data and can be predicted. By the training, the

system can be able to decide the data and can be predicted. By training the machine with various algorithms it provides various accuracy with the different period of time to complete its process by choosing the different algorithms. Finally, the higher accuracy with the minimum amount of time taken to complete the process is chosen as the best machine learning algorithm for forecasting the Corona virus cases.

### 1.3 Block Diagram




## 2. Literature Survey

There are a few algorithms that focus on the trend analysis and forecasting for Indian region. The study papers on Indian region that presents long term and short term trend, respectively. These studies use time series data from World Health Organization (WHO) However network modelling and pattern mining get varied with different prediction models.

### 2.1 Network-Based Sir Model

Similarly, a network structure approach was used by one of the study to see whether any specific node clusters were getting formed. But only travel data nodes were considered by the authors to check which the prominent regions are affecting Indian traveller coming back to the India.

### 2.2 Curve Fitting

Apart from India, a few models are also available for other countries primarily for China, Italy and USA as the number of infected patients was high. Studies that worked on various mathematical models to determine the spread of the disease, predict the number of infected patients, commenting on the preparedness for each country in tackling COVID-19 spread and finding the patterns of flattening curve in different conditions. A lot of researches are still in preprint stage for the world level and are yet to be peer reviewed. With respect to the research activities conducted in the Indian region, the studies are yet to work on the impact of different policies working towards containment of the corona virus.

### 2.3 Generalized Logistic Model

Accuracy of results depends proportionally on the size of the sample data collected. In this study, modelling logistic growth models for COVID-19 pandemic in its most nascent early stages, inferring that the size of sample data available for data extrapolation is very small.

Hence, as the sample data size is very limited, saw differences between actual and predicted forecasts using this method.

Some external factors also might have resulted in misreporting of actual cases.

### 2.4 Sir Model

SIR model is one of the simplest and most fundamental of all epidemiological models. It is based upon calculating the proportion of the population in each of the three classes (susceptible, infected and recovered) and determining the rates of transition between these classes.

## 3. Proposed System

1. The project illustrates how the machine learning algorithms such as Linear regression, Gaussian processes, Support vector regression to predict the number of upcoming cases of covid-19 and their accuracy in prediction.
2. These algorithm will predict the newly infected cases, number of death and number of recoveries accurately in the future

### 3.1 Linear Regression

Linear regression a type of regression modelling is the most usable statistical technique for predictive analysis in machine learning. Each observation in linear regression depends on two Linear regression determines a linear relationship between these dependent and independent variables.

### 3.2 Gaussian process

It is a collection of random variables indexed by time or space, such that every finite linear combination of them is normally distributed and it is a joint distribution of all random variables. It measure the similarity between points to predict the value for an unseen points from training data.

### 3.3 Support Vector Regression

The regression problem is a generalization of the classification problem, in which the model returns a continuous-valued output, as opposed to an output from a finite set. In other words, a regression model estimates a continuous-valued multivariate function.

## 4. Conclusion

The machine learning algorithms such as Linear regression, Gaussian processes, Support vector regression gives accuracy in prediction, quickly and easy prediction of spreading virus so that the Protective measure can be taken earlier. Among these algorithms Support vector regression results in better accuracy.

## 5. References

1. Gupta, R., & Pal, S. K. (2020), “Trend Analysis and Forecasting of COVID-19 outbreak in India”. medRxiv. India, 2020.

2. Gupta, R., Pandey, G., Chaudhary, P., & Pal, S. K, “SEIR and Regression Model based COVID-19 outbreak predictions in India”. medRxiv. India, 2020.

3. Joe Hasell, Lucia Russo, “Prediction and forecast for COVID-19 Outbreak in India based on Enhanced Epidemiological Models”, Second International Conference on Inventive Research in Computing Applications, IEEE Publications, May 11,2020.

4. Minni Jain, Prajwal Kumar Bhati, Pranav Khataria, Rohit Kumar, “Modelling Logistic growth model for Covid-19 pandemic in India”, 5th Internation Conference, IEEE Publications, 2020. Gupta, R., & Pal, S. K. (2020), “Trend Analysis and Forecasting of COVID-19 outbreak in India”. medRxiv. India, 2020.