International Journal of
**INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING**
ISSN:2147-6799 www.ijisae.org Original Research Paper

# SARS-CoV-2 Future Forecasting Using Multi-Linear Regression Model

**Jinal Patel*[1], Safvan Vahora[2]**

**Submitted: 28/12/2021** **Accepted : 07/03/2022**

***Abstract:*** The 2019 pandemic in Wuhan, China caused a devastating global outbreak of the Coronavirus Disease (SARSCoV-2). Machine learning offers a number of prediction models for future events that are based on training and testing, including conventional machine learning and Deep Learning. This study shows that machine-learning models can anticipate the number of future SARS-CoV-2 patients that are currently seen as a possible risk to the human race. Supervised machine learning models like linear regression, vector support and regression tree are used for prediction. Data on the total cases and recovery cases are based on two types of predictions: new infections and recovery situations. The machine-learning regression model is used to generate the outcome. In this paper, we present prediction of future forecasting of Covid cases based on current situation by applying dataset of before and after pre-trial vaccine.

***Keywords:*** *Linear Regression, LASSO, Support Vector Machine (SVM), Random Forest, Exponential Smoothing and Regression Tree.*



## 1. Introduction

Machine learning is very useful to improve decision-making in future predictions. Machine learning has several future situation prediction models based on training and testing data sets, including supervised and unchecked machine learning algorithms. Number of prediction methods are in use to address prediction issues. The study shows that the number of new patients with SARS-CoV-2 is currently considered as potential human threats by the Machinery Learning Models. Support model vector machine is one of the supervised machine learning model, that has been used for prediction, such as linear regression. Two predictions, the number of cases newly infected and the number of recovery cases, provide data on the overall cases concerned. Two cases of datasets are used: before pre-tested vaccine and after pre-tested vaccine. This dataset predicted result is achieved through the linear regression [1,2,4,10] machine learning models.

Data science remains one of the best and most demanding career opportunities for skilled people. Successful data professionals today recognise the need for wide-ranging data analysis, data analysis and programming beyond traditional abilities. In addition, data scientists have to master the whole range of the life cycle of the sciences and have a degree of polyvalence and understanding to optimise returns at each stage of the process in order to detect important information for their organisations.

Figure 1 indicates how data mining works and how the various parts of data mining works like, capture, process, maintain, communicate and analyse.

Data and data science strategies of the SARS-CoV-2:

1. Clear data strategy development Research Article2
2. Identifying and building targeted cases of business use
3. Train the Applied Data Science Workforce

[1] *Government Engineering College, Modasa – 383315, India*
*OCID ID:0000-0002-2783-7770*
[2] *Government Engineering College, Modasa – 383315, India*
*ORCID ID:0000-0002-8051-8429*
** Corresponding Author Email: jinalpatel6411@gmail.com*

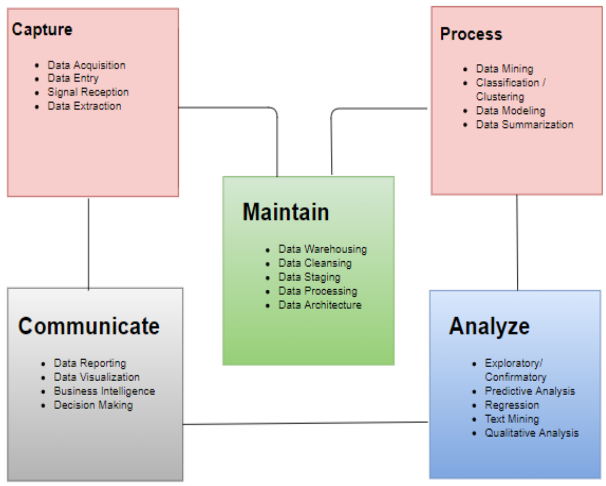


**Figure1.** Process of data mining

## 2. Literature Review / State-of-the art

IEEE COVID-19 member Future forecasting use of supervised machine learning models [1]. The advantage is that ES is the best performed on a limited time series, with Mehmood et. al. Model-based ML forecasting can help decision makers contain pandemics from COVID-19. The problem is real-time live forecasts.

Ardabili et al. presents outbreak Prediction for Covid-19 [2]. Here, advantage is that the long-term forecast capacity of Multi Layered Perceptron (MLP) & Adaptive Neuro Fuzzy Inference System (ANFIS) is large. Death rate modelling is the key issue in this model. Global models with broad capacity could not be promoted.

Pasupuleti et al. use Covid-19 Patient Health Prediction Boosted Random Forest Algorithms [3]. The advantage is to predict the most probable outcome of the patient. The problem is the construction of a pipeline combining C*R scanning models with demographic and medical data supporting mobile health.

Jyotir et. al. presented a covid 19 prediction model for the pandemic machine learning in India is a predictable Multi Layered Perceptron method [4]. The problem is the use of certain in-depth learning methods to better predict time series data.

Kalamkar, et al. presented a State-of-the-Art Survey (Covid-19) [5]. The major coverage in this research is that different data sets for spreading and replicating predictions should be analysed. The problem is that there is an alarming effect on this pandemic.

Pinter et al. predict pandemics for Hungary; hybrid learning on equipment [6]. The advantage is that in comparison with other Membership Function (MF) types, Gaussian Membership Function is the least error and highest precision. The downside is essential if the results are to be validated and prediction quality improved. Further development of a deep learning and a profoundly strengthening learning model is strongly suggested for a comparative study of various ML models for individual countries.

The prediction of mortality rates presented by Dhamodhara vadhani et al. for Covid19 in India using the statistical neural network models [7]. The benefit is to develop the hybrid SNN-NAR-NN model, which is suitable for short-term mortality prediction. The problem is the predicted model cannot be used for recurring time series of Deep Learning neural networks. For the Covid 19 pandemic model for the modulation of logistic farming in Indian, Jain et al. [7]. The advantage is that Generalized Linear Model (GLM) has a 0.99 trust interval of 95 percent.

In current medical interventions, Alzahrani et al. have the benefit to predict a pandemic Covid-19 in Saudi Arabia using an ARIMA prediction model [8]. The problem with this model is statistics for predicting daily deaths and rehabilitation.

Alboaneen et al. are the predetermined epidemiological outbreak of Corona-virus disease 2019 (Covid-19) in Saudi Arabia [9]. The Logistical Growth Model is the advantage over others. The problem is that no tests are carried out so far.

The Artificial neural network, the Fuzzy Time series and ARIMA models for research and project presented by the Mishra et al. for Covid-19 in India [10]. The key advantage of is that Auto Regressive Integrated Moving Average (ARIMA) and Fuzzy Time Series (FTS) models are more suitable for predicting the trajectory. Long-term trajectory of data is the problem.

The internet of things helped to reduce the rapid spread of Covid-19 with the Drone-bases approach. The benefits are efficient if the city is implemented in different states in a clustered way. The benefits are effectively achieved by the drone-based approach by Angurala et al. [11]. New recharging techniques is presented in order to overcome the problem.

For the Covid-19 time series analysis in India, Khan et al. proposed Auto Regressive Integrated Moving Average (ARIMA) & Nonlinear Auto Regressive (NAR) based prediction model [12]. The advantage of this proposed approach that both models are almost as efficient. The problem is to predict the scenario of unconfirmed, rejected and recovered cases of Covid-19.

Khan et al. proposed for the modelling and prediction of new cases, death and recovery by implementing the vector autoregression model of Pakistan, Covid-19 cases [13]. The benefit is a 10-day trust interval forecast of 95%. These results can be beneficial for policymakers, other health actors and other departments. The problem is that fewer data are available on Covid 19 because further data make the parameters of the model more stable and predictable.

For Covid-19's new adaptive deep-education model focusing on the strategy to reduce mortality proposed by Farooq et al. [14]. This model is a useful tool for policy makers, health workers and researchers all 3 over the world. Their use is a good tool. This model works effectively every day without any decrease in performance.

For Covid-19, outbreak predictions and forecasts in India based on enhanced epidemiological disease models are based on method proposed by Singh et al [15]. Predictions of the scenario in India are one of the benefits. Approximate mathematical modelling was used to model the covid outbreak in India. This is due to the fact that a huge number of asymptomatic patients are not checked and are not strictly seized. The issue could be that there is no real proof of cases.

The neural network was driven by a model proposed by Wieczork et al. for Covid-19 [16]. The advantage is that neural network predictors can lead to precision and better adaptation than other stochastic methods. ANN achieves maximum accuracy in far less iteration. The problem is to improve the network's efficiency.

The analysis of genetically modified pandemics in India presented by Salgotra et al. [17]. The advantage is highly reliable. Work on fewer data from time series and produce reliable results.

## 3. Proposed Methodology

A process of prediction of data follows several fundamental steps listed below to be computational:

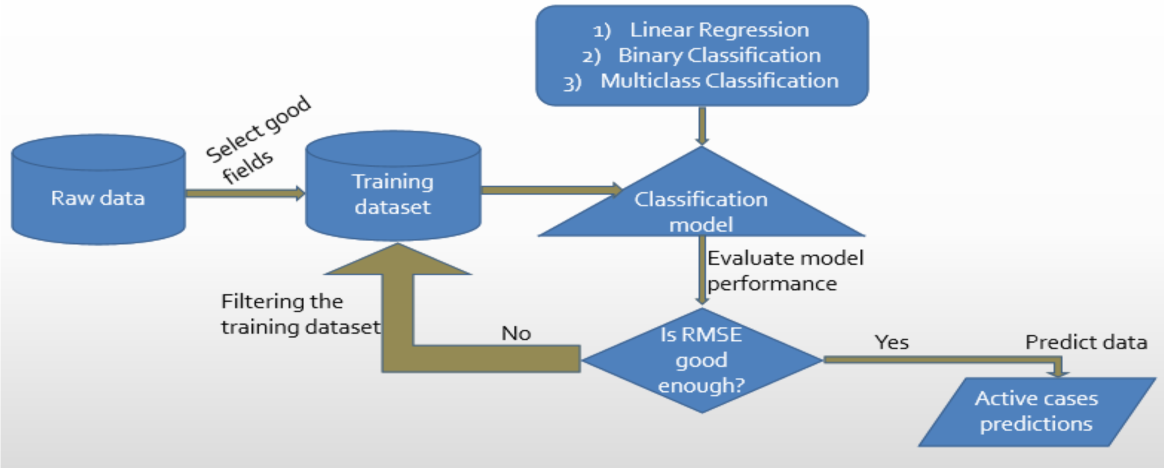


**Figure 2.** Proposed System Flow

1) Dataset used in current paper
2) Regression models
3) Evaluation of Parameters
4) Pre-processing

The details shown in figure 2 are summarised in the flow chart.

**3.1. Dataset used in current paper [13]**

The objective of this evaluation is to establish in future the distribution of COVID-19 with emphasis on the number, number of new occurrences and retrievals. The Team ESRI Living Atlas maintained the vault that the school provided for the visual dashboard Novel Coronavirus in 2019. In the regular time schedule development tables, the coordinator will be included including the number of advertised cases, transition and regeneration. The data tests of the records are shown in Tables I, II separately.

**Table.1.** Before pre-trial vaccination dataset of Gujarat

| Date | TotalSamples | Negative | Positive | year | month | day |
|---|---|---|---|---|---|---|
| 2020-05-01 | 68774.0 | 64053 | 4721 | 2020 | 5 | 1 |
| 2020-05-02 | 74116.0 | 69062 | 5054 | 2020 | 5 | 2 |
| 2020-05-03 | 80060.0 | 74632 | 5428 | 2020 | 5 | 3 |
| 2020-05-04 | 84648.0 | 78844 | 5804 | 2020 | 5 | 4 |
| 2020-05-05 | 89632.0 | 83387 | 6245 | 2020 | 5 | 5 |
| ... | ... | ... | ... | ... | ... | ... |
| 2020-10-18 | 5374249.0 | 4974315 | 386976 | 2020 | 10 | 18 |
| 2020-10-19 | 5426621.0 | 5022789 | 390747 | 2020 | 10 | 19 |
| 2020-10-20 | 5479536.0 | 5071766 | 394558 | 2020 | 10 | 20 |
| 2020-10-21 | 5532522.0 | 5120809 | 398373 | 2020 | 10 | 21 |
| 2020-10-22 | 5585445.0 | 5169794 | 402184 | 2020 | 10 | 22 |

175 rows × 6 columns

**Table.2.** After vaccination dataset of Gujarat

| Date | Confirmed | Recovered | Deceased | Tested | year | month | day |
|---|---|---|---|---|---|---|---|
| 2020-12-01 | 211257 | 192468 | 4004 | 7894467.0 | 2020 | 12 | 1 |
| 2020-12-02 | 212769 | 194038 | 4018 | 7963653.0 | 2020 | 12 | 2 |
| 2020-12-03 | 214309 | 195465 | 4031 | 8033388.0 | 2020 | 12 | 3 |
| 2020-12-04 | 215819 | 197092 | 4049 | 8102712.0 | 2020 | 12 | 4 |
| 2020-12-05 | 217333 | 198627 | 4064 | 8172380.0 | 2020 | 12 | 5 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2021-03-27 | 298596 | 283241 | 4484 | 13182148.0 | 2021 | 3 | 27 |
| 2021-03-28 | 300866 | 284846 | 4492 | 13263977.0 | 2021 | 3 | 28 |
| 2021-03-29 | 303118 | 286577 | 4500 | 13334810.0 | 2021 | 3 | 29 |
| 2021-03-30 | 305338 | 288565 | 4510 | 13406686.0 | 2021 | 3 | 30 |
| 2021-03-31 | 307698 | 290569 | 4519 | 13497177.0 | 2021 | 3 | 31 |

121 rows × 7 columns

**3.2. Pre-Processing**

The columns with dates, strings and numbers are included. In addition, categorical variables are found in the dataset. Since all data transmitted by the machine learning model in numerical form are necessary, the categorical variables have been labeled[4]. This assigns each categorical single value of the column to a number. The dataset consists of several missing values that cause an error when directly passed as an input. Thus, to the missing values we add "NA." Patient records containing both death columns and missing recovery values have been removed and patient records compiled into remaining records from main data in the test dataset. The data set also contains columns in the date format. Due to the lack of direct use of data columns, functional engineering was implemented [7].

**3.3. Regression Models**

**3.3.1. Linear regression [2,10]:**

The free features of the reverse illustration are used for a target class. Therefore, the relationship of free and security factors and decision making can be eliminated by this method. Direct reversal of such a reversal is the most commonly used authentic method for AI perceptive testing. The two characteristics depend on every discernment in a direct backlash; one is the penniless variable and the other is the free factor. Immediately backwards, these poor and free factors are linked immediately. The reverse check has been locked with two (x, y) factors. It shows how "y" is linked to x called backslide. Y=b0+b1x+e (1) E(y)= b0+b1x (2) Here, e is the blunder term of straight relapse. The terms "error" are used to account for the variability between x and y, b0 speaks to y-capture, b1 speaks to slant.

**3.3.2. LASSO [2]:**

Tether is an inverse model with the reverse shrinking process. In this context, declining indicates that the data test for central features has been assessed without precedent. The cut-off cycle improves and stabilises the LASSO, reducing the failure. Tie is regarded as a more appropriate model of multimedia. The discipline recorded is equivalent to the meaning of coefficients in this case, since the model regularises L1. The LASSO background therefore reduces the amount of features it uses. It normally rejects additional features with a regulatory procedure.

**3.3.3. SVM [3]:**

A coordinated AI system for demand is a vector support system (SVM). In the case of demands, or for different companies such as confirmation of unusual cases, SMV develops a hyperplane or hyperplanes in a high-dimensional area. The hyperplane, which is best prepared, inspires every class of information in order to make the best arrangement possible.

**3.3.4. Exponential Smoothing [2]:**

An exponential smoothing technique is advanced for the use in smooth time series results of exponential windows. When previous observations weigh the basic moving average evenly, exponential functions are used to assign an exponential reduction in weight over time. This is a clearly learned method that is easy to use in determining seasonality on the basis of consumer premises. Exponential smoothing is sometimes used to analyse time series results. Exponential smoothing is a multi-window window used to remove high frequency sonicity as low-pass filters for smooth signal handling. This approach was followed in the 19th century by the use in their turbulent experiments of the recurrent exponential windows and the use of the recursive moving medium by Kolmogorov and Zurbenko.

**3.3.5. Regression Tree:**

Each reactions approach has one variable (response) and at least one part (pointer). A number of returns are available. A standard approach for tree development allowing the floor to be a mixture of constant flexibility and stage flexibility. If each decision centre point in a tree is tested for the evaluation of a dataset, the decision tree is established. The expected yields are included in the final centre of the tree. The background tree may be considered a kind of decision tree, which should align activities rather than collect techniques with real value. The cycle known as the double division marks the backbone tree as each branch of the route advances, which means that it is a dark cycle that isolates the data into zones or separations. From the very start the entire training set is collected in a similar class (pre-coordinated records for the selection of tree game plans). The calculation then begins to divide the results into the two underlying sections or advances using each possible two divide. For the total number of square bottles the diagram divides a book into two distinct sections. For each new branch, the rules of the section apply. This cycle continues until the last centre point is specified by the customer. (If the entire square divergence between the centre image is nil, or not the base size, the centre point shall be considered as the ultimate element.)

## 4. Evaluation Parameters

### 4.1. $R^2$ SCORE [1,2]

R-squared ($\boldsymbol{R^2}$) score is a statistical measure used to evaluate the performance of regression models. The statistic shows the dependent variable's variance percentage that collectively determines the independent variable. It measures the relationship strength between the dependent variable and regression models on a convenient 0 – 100% scale. After training the regression model, we can check the goodness of-fit of trained models by using the $\boldsymbol{R^2}$ score. $\boldsymbol{R^2}$ score finds the scatteredness of data points around the regression line which can also be referred to as the coefficient of determination. Its score always between 0 and 100%. 0% score implies the response variable has no variability around its mean explained by the model, and 100% implies that the response variable has all the variability around its mean. The high $\boldsymbol{R^2}$ score shows the goodness of the trained model. $\boldsymbol{R^2}$ is a linear model that explains the percentage of variation independent variable. It can be found as:

$$R^2 = \frac{Variance\ explained\ by\ model}{Total\ Variance} \tag{1}$$

### 4.2. $R^2_{adjusted}$ SCORE [1]

The Adjusted R-squared ($\boldsymbol{R^2_{adjusted}}$) is a modified form of $\boldsymbol{R^2}$, which also like $\boldsymbol{R^2}$ shows how well the data points fit the curve. The primary difference between $\boldsymbol{R^2}$ and $\boldsymbol{R^2_{adjusted}}$ is that the later adjusts for the number of features in a prediction model. In the case of $\boldsymbol{R^2_{adjusted}}$, the increase in new features can lead to its increase if the newly added features are useful to the prediction model. However, if the newly added features are useless, its value will decrease. The $\boldsymbol{R^2_{adjusted}}$ can be defined as:

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{n-1}{n-(k+1)} \tag{2}$$

Here, n is the sample size and k is the number of independent variables in the regression equation.

### 4.3. MEAN SQUARE ERROR (MSE) [1,2]

Mean square error is another way to measure the performance of regression models. MSE takes the distance of data points from the regression line and squaring them. Squaring is necessary because it removes the negative sign from the value and gives more weight to larger differences. The smaller mean squared error shows the closer you are to finding the line of best fit. MSE can be calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

### 4.4. MEAN ABSOLUTE ERROR (MAE)[1,9,11]

The mean absolute error is the average magnitude of the errors in the set of model predictions. This is an average on test data between the model predictions and actual data where all individual differences have equal weight. Its matrix value range is from 0 to infinity and fewer score values show the goodness of learning models that's the reason it's also called negatively-oriented scores.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

### 4.5. ROOT MEAN SQUARE ERROR (RMSE) [1,2,6,7,9,11,13,18]

Root mean square error can be defined as the standard deviation of the prediction errors. Prediction errors also known as residuals is the distance from the best fit line and actual datapoints. RMSE is thus a measure of how concentrated the actual data points are around the best fit line. It is the error rate given by the square root of MSE given as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

## 5. Results and Analysis

**Table.3.** Input table for Gujarat data

| Date | Confirmed | Recovered | Deceased | Tested | year | month | day |
|---|---|---|---|---|---|---|---|
| **2020-12-01** | 211257 | 192468 | 4004 | 7894467.0 | 2020 | 12 | 1 |
| **2020-12-02** | 212769 | 194038 | 4018 | 7963653.0 | 2020 | 12 | 2 |
| **2020-12-03** | 214309 | 195465 | 4031 | 8033388.0 | 2020 | 12 | 3 |
| **2020-12-04** | 215819 | 197092 | 4049 | 8102712.0 | 2020 | 12 | 4 |
| **2020-12-05** | 217333 | 198627 | 4064 | 8172380.0 | 2020 | 12 | 5 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **2021-03-27** | 298596 | 283241 | 4484 | 13182148.0 | 2021 | 3 | 27 |
| **2021-03-28** | 300866 | 284846 | 4492 | 13263977.0 | 2021 | 3 | 28 |
| **2021-03-29** | 303118 | 286577 | 4500 | 13334810.0 | 2021 | 3 | 29 |
| **2021-03-30** | 305338 | 288565 | 4510 | 13406686.0 | 2021 | 3 | 30 |
| **2021-03-31** | 307698 | 290569 | 4519 | 13497177.0 | 2021 | 3 | 31 |

121 rows × 7 columns

As shown in table 3 we have Gujarat data of six months as an input and forecasted next one month data of Gujarat confirmed cases using method Regression Tree After Vaccine in figure 3.

As shown in table 3 we have Gujarat data of six months as an input and forecasted next one month data of Gujarat recover cases using method Regression Tree After Vaccine in figure 4.

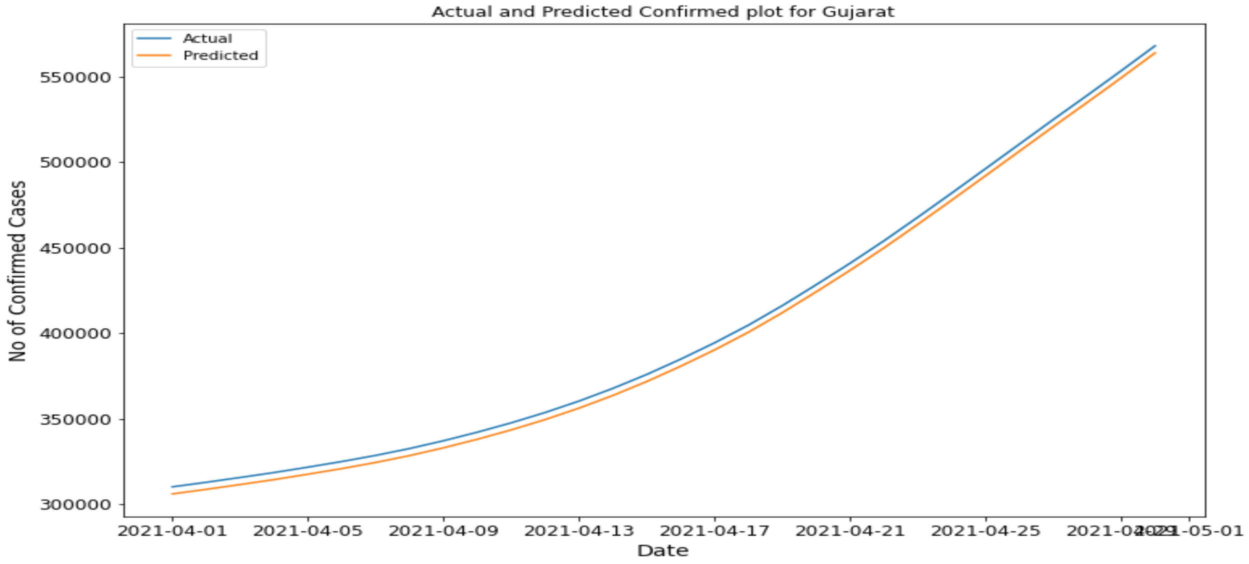


**Figure 3.** Confirmed case future forecasting for Gujarat

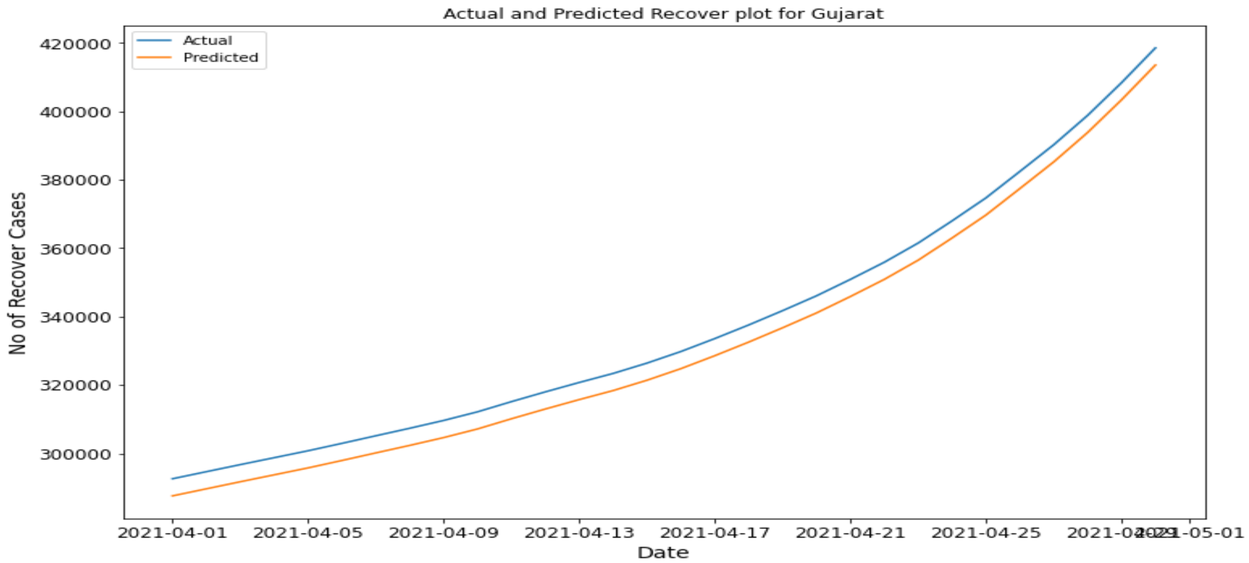


**Figure 4.** Recover case future forecasting for Gujarat

**Table.4.** Input table for Maharashtra data

| Date | Confirmed | Recovered | Deceased | Tested | year | month | day |
|---|---|---|---|---|---|---|---|
| **2020-12-01** | 1828826 | 1691412 | 47246 | 10915683.0 | 2020 | 12 | 1 |
| **2020-12-02** | 1832176 | 1695208 | 47357 | 10989496.0 | 2020 | 12 | 2 |
| **2020-12-03** | 1837358 | 1703274 | 47472 | 11059305.0 | 2020 | 12 | 3 |
| **2020-12-04** | 1842587 | 1710050 | 47599 | 11132231.0 | 2020 | 12 | 4 |
| **2020-12-05** | 1847509 | 1715884 | 47694 | 11205118.0 | 2020 | 12 | 5 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **2021-03-27** | 2673461 | 2314579 | 54073 | 19192750.0 | 2021 | 3 | 27 |
| **2021-03-28** | 2713875 | 2332453 | 54181 | 19358341.0 | 2021 | 3 | 28 |
| **2021-03-29** | 2745518 | 2353307 | 54283 | 19495189.0 | 2021 | 3 | 29 |
| **2021-03-30** | 2773436 | 2377127 | 54422 | 19625065.0 | 2021 | 3 | 30 |
| **2021-03-31** | 2812980 | 2400727 | 54649 | 19792143.0 | 2021 | 3 | 31 |

121 rows × 7 columns

As shown in table 4 we have Maharashtra data of six months as an input and forecasted next one month data of Maharashtra confirmed cases using method Regression Tree After Vaccine in figure 6. As shown in table 4 we have Maharashtra data of six months as an input and forecasted next one month data of Maharashtra confirmed cases using method Regression Tree After Vaccine in figure 6.

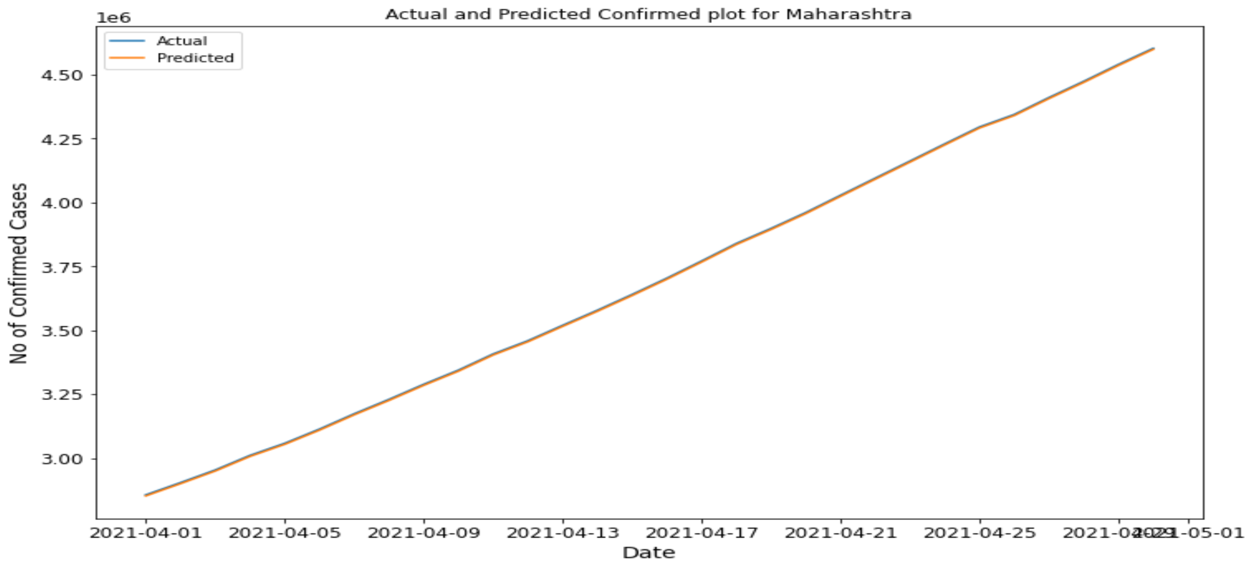


**Figure.5.** Confirmed case future forecasting for Maharashtra

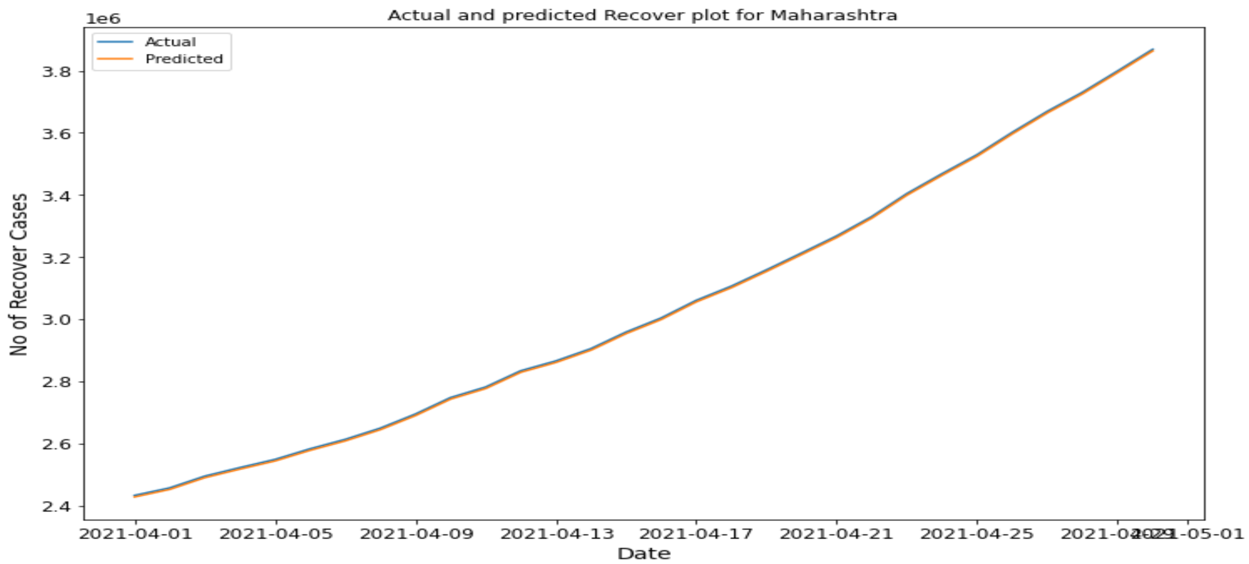


**Figure 6.** Recover case future forecasting for Maharashtra

**Table 5.** Efficiency of the models in terms of future forecasting for new infected confirm cases for Gujarat

| Model | | R2 Score | | R2 Adjusted | | MSE | | MAE | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper |
| **LR [1]** | | 0.83 | 0.44 | 0.79 | 35.02 | 1472986504.96 | 22741958074.38 | 30279.55 | 60315.90 | 38390.51 | 150804.37 |
| **LASSO [1]** | | 0.98 | 0.43 | 0.97 | 35.31 | 234489560.99 | 23258662039.84 | 11693.97 | 61264.04 | 15322.11 | 152507.91 |
| **SVM [1]** | | 0.59 | -0.13 | 0.47 | 69.34 | 5760890969.30 | 45729194950.92 | 60177.90 | 78212.57 | 75911.28 | 213843.86 |
| **Exponential Smoothing [1]** | | 0.98 | 0.98 | 0.97 | 1.03 | 283201302.2 | 492859077.46 | 8867.43 | 7811.07 | 16828.58 | 22200.43 |
| **Regression Tree (proposed)** | **Before Vaccination** | - | 0.99 | - | 1.0 | - | 1252161.0 | - | 1119.0 | - | 1119.0 |
| | **After Vaccination** | - | 0.99 | - | 1.0 | - | 17040384.0 | - | 4128.0 | - | 4128.0 |

**Table 6.** Efficiency of the models in terms of future forecasting for new infected recover cases for Gujarat

| Model | | R2 Score | | R2 Adjusted | | MSE | | MAE | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper | Dataset used in [1] | Dataset used in current paper |
| **LR** | | 0.39 | 0.99 | 0.21 | 1.50 | 480922814.51 | 48214606442.62 | 17016.08 | 100521.30 | 21929.95 | 219578.24 |
| **LASSO** | | 0.29 | 0.99 | 0.08 | 1.52 | 1462144344.82 | 49961940232.16 | 30705.27 | 103937.94 | 38237.99 | 223521.68 |
| **SVM** | | 0.24 | -0.21 | 0.02 | 74.06 | 13121148615.72 | 7335810641237.62 | 106739.82 | 1297790.42 | 114547.58 | 2708470.16 |
| **Exponential Smoothing (existing)** | | 0.99 | 0.99 | 0.99 | 1.00 | 5970634.07 | 1773127374.97 | 1827.85 | 32071.54 | 2443.48 | 42108.52 |
| **Regression Tree (proposed)** | **Before Vaccination** | - | 0.99 | - | 1.0 | - | 228614400.0 | - | 15120.0 | - | 15120.0 |
| | **After Vaccination** | - | 0.98 | - | 1.0 | - | 25080064.0 | - | 5008.0 | - | 5008.0 |

## 6. Conclusion

The findings show that the LR generates low results in all scenarios due to ups and downs in the data set values. A precise hyperplane between the data sets was very difficult to place. The previews for the next situation based on the existing scenario can be correct and helpful. The projections of the studies may also help the authorities to deal with the COVID-19 crisis from time to time. The prediction method was also examined with the most accurate and suitable methods for predit or forecasting Covid-19 cases including LASSO [F. Rustam et al., S.F. Ardabili], SMM [F. Rustam et al., C. Iwendi et al], LR [F. Rustam et al., S. F. Ardabili] and ES [F. Rustam et al.]. The results of the prediction methodology were also explored with vixen. It could be said that the Regression Tree model gives better R2 values for everyone in the comparative table. The regression tree model could be used for the future prediction of covid-19 cases.

## References


[1] Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," IEEE Access, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[2] F. Ardabili et al., "COVID-19 Outbreak Prediction with Machine Learning," SSRN Electron. J., 2020, doi: 10.2139/ssrn.3580188.

[3] Iwendi et al., "COVID-19 patient health prediction using boosted random forest algorithm," Front. Public Heal., vol. 8, no. July, pp. 1–9, 2020, doi: 10.3389/fpubh.2020.00357.

[4] Sujath, J. M. Chatterjee, and A. E. Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India," Stoch. Environ. Res. Risk Assess., vol. 34, no. 7, pp. 959– 972, 2020, doi: 10.1007/s00477-020-01827-8.

[5] R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, "Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art," SN Comput. Sci., vol. 1, no. 4, pp. 1–15, 2020, doi: 10.1007/s42979-020-00209-9.

[6] Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach," SSRN Electron. J., 2020, doi: 10.2139/ssrn.3590821.

[7] Dhamodharavadhani, R. Rathipriya, and J. M. Chatterjee, "COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models," Front. Public Heal., vol. 8, no. August, pp. 1–12, 2020, doi: 10.3389/fpubh.2020.00441.

[8] M. Jain, "Pandemic in India," no. Icces, pp. 784–789, 2020.

[9] I. Alzahrani, I. A. Aljamaan, and E. A. Al-Fakih, "Forecasting the spread of the COVID 19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions," J. Infect. Public Health, vol. 13, no. 7, pp. 914–919, 2020, doi: 10.1016/j.jiph.2020.06.001.

[10] Alboaneen, B. Pranggono, D. Alshammari, N. Alqahtani, and R. Alyaffer, "Predicting the epidemiological outbreak of the coronavirus disease 2019 (COVID-19) in Saudi Arabia," Int. J. Environ. Res. Public Health, vol. 17, no. 12, pp. 1–10, 2020, doi: 10.3390/ijerph17124568.

[11] Mishra et al., "Trajectory of COVID-19 Data in India: Investigation and Project Using Artificial Neural Network, Fuzzy Time Series and ARIMA Models," Annu. Res. Rev. Biol., vol. 35, no. 9, pp. 46–54, 2020, doi: 10.9734/arrb/2020/v35i930270.Covid-19 Future Forecasting using Supervised Multi-linear Regression 50

[12] Angurala, M. Bala, S. S. Bamber, R. Kaur, and P. Singh, "An internet of things assisted drone based approach to reduce rapid spread of COVID-19," J. Saf. Sci. Resil., vol. 1, no. 1, pp. 31–35, 2020, doi: 10.1016/j.jnlssr.2020.06.011.

[13] M. Khan and R. Gupta, "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India," J. Saf. Sci. Resil., vol. 1, no. 1, pp. 12–18, 2020, doi: 10.1016/j.jnlssr.2020.06.007.

[14] Khan, A. Saeed, and S. Ali, "Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan," Chaos, Solitons and Fractals, vol. 140, p. 110189, 2020, doi: 10.1016/j.chaos.2020.110189.

[15] Farooq and M. A. Bazaz, "A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies," Chaos, Solitons and Fractals, vol. 138, 2020, doi: 10.1016/j.chaos.2020.110148.

[16] Singh, P. Raj, R. Kumar, and R. Chaujar, "Prediction and forecast for COVID-19 Outbreak in India based on Enhanced Epidemiological Models," Proc. 2nd Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2020, pp. 93–97, 2020, doi: 10.1109/ICIRCA48905.2020.9183126.

[17] Wieczorek, J. Siłka, and M. Woźniak, "Neural network powered COVID-19 spread forecasting model," Chaos, Solitons and Fractals, vol. 140, 2020, doi: 10.1016/j.chaos.2020.110203.

[18] Salgotra, M. Gandomi, and A. H. Gandomi, "Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming," Chaos, Solitons and Fractals, vol. 138, 2020, doi: 10.1016/j.chaos.2020.109945.

[19] Johns Hopkins University Data Repository. Cssegisanddata. Accessed: Mar. 27, 2020. [Online]. Available: https://github.com/CSSEGISandData