A Comparative Study on Covid-19 Cases in Top 10 States/UTs of India in Using Machine Learning Models

Dr. Deena Babu Mandru[a], Dr. A. Ramasway Reddy[b]

[a]Assoc.Professor, Department of CSE, Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana-500100
[b]Professor, Department of CSE, Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana-500100

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Coronavirus is a dangerous sickness came from a new virus. It has been assumed as an overall pandemic and a very hard circumstance to control the COVID-19 epidemic in India and global and so needed some severe actions to control its rate of increment. This disease causes for cold, dry cough, high fever, sore throat and serious breathing problems. This paper presents analysis of confirmed, cured and deaths cases, age and gender based cases in top 10 States/UTs of India. We analyzed various trends and patterns from various state/UTs units, MHFW of India data sources (up to 16th November 2020). Now a day's plentiful models are proposed to predict covid-19 cases in India and world countries. The novel COVID-19 datasets are taken from Kaggle and GitHub repositories to analyze the epidemiological cases of the disease in top 10 states/UTs of India. We used various machine learning algorithms like Linear Regression, KNN Regressor, LASSO Regression, Elasticnet Regression and Decision Tree regressor to analyze the number of novel Coronavirus (COVID-19) reported cases in top 10 states/UTs of India. The model analyzes datasets containing the COVID-19 cases (confirmed, cured and death cases) up to 24th November, 2020 using ML models. From the results it is proven that Decision Tree and KNN regressor performs best in analyzing the number of confirmed cases and number of death cases. But for number of cured cases LASSO and linear regression models give the best accuracy results. Unfortunately, Elastic net produced poor accuracy results due to some changes in original datasets. Especially, this work analyzes the calculations based on the exactness rate on a test dataset.

Keywords: Regression, LASSO, COVID-19, Decision Tree, K-Neighbors, R2-score, mean absolute error

1. Introduction

Coronavirus (COVID-19) is a contagious disease caused by a recently exposed a new virus. The majority people who fall unwell with COVID-19 resolve incidence easygoing to modest symptoms and get well without particular treatment. Until 24th November, 2020, as per Ministry of Health and Family Welfare records, it says less than 40000 daily new cases are found and surprisingly number of active cases reducing significantly. As of now total number of active cases in India is less than 4.4 and positive cases lies in between 3.45% and 4%. Throughout India, government establishes 2134 testing centers (labs) and every day conducted more than one million (10 Lakhs) tests. Indian government conducted more than 13.3 crore (13, 36, 82,275) tests cumulatively. The Cumulative National Positivity Rate (CNPR) is 6.87%, (< 7%) where daily positive case rate is 3.45%. In top 10 States/UTs, it is recognized that the new recovered positive cases are 75.71% of the new recovered cases. In single day among Top 10 states Delhi, Kerala, and Maharashtra maximum number of single day recoveries are 7,216, 5,425, and 3,729 respectively. Overall new cases of India 77.04% cases are identified in these 10 States/UTs only. Within 24 hours fatality cases are 480. And total number of reported new deaths in top 10 States/UTs 73.54% of new deaths. Highest death cases in Delhi are 121, in West Bengal 47 and in Maharashtra 30.

Over decade machine learning provides itself as a majority area of study for solving a lot of exceptional complicated universal problems. The most important applications in real-time are High dimensional data clustering [1], NLP, Automobile Industries, Health care sectors weather condition modeling, robotics, and many more areas. Naturally machine learning depends on trial-and-error technique moderately contrary of standard algorithms, which follows the programming commands based on decision-making statements [2]. One of the majority important areas of ML is forecasting [3], in this study plentiful typical ML models have been used to conduct the future path of actions required in a lot of applications as well as weather conditions forecasting analysis, stock market analysis and also for disease diagnosis. A variety of regression and NN models have extensively applied in predicting the circumstances of patients in the prospect with an explicit disease [4]. Present numerous scholars who studied for the prediction of various kinds of diseases like coronary vein sickness [5], cardiovascular (heart and blood vessels) related sickness forecasting [6], and prediction of breast cancer [7] with help of machine learning models. Especially, the work in [8] is concerned on live status of prediction of confirmed cases of COVID-19 and in [9] COVID-19 epidemic as well as forecasting of the early response. However all of these prediction models are extremely supportive in assessment making to hold the current circumstances to direct early interventions to deal with these problems efficiently.

To support to the present human disaster, our effort in this work is to build model for analysis of COVID-19 cases in top 10 states of India, and analysis of gender-based effected cases and then analysis of age-based effected cases in India. The analysis is made for the following significant key points of the disease up to November 16,

2020: 1.Number of confirmed cases, 2.Number of death cases, and 3.Number of cured cases in top 10 States/UTs of India. For identifying COVID-19 cases in top 10 States/UTs of India, we applied regression and classification models [9] like linear regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), K Neighbors Classifier, Elastic net Regression and Decision Tree Classifier (DTC).

COVID-19 India dataset provided by Kaggle [11] are used to train the learning models. Initially the dataset files are preprocessed and then separated into training set (80%) and testing set (20%). By using evaluation metrics the performance is assessed in terms of significant measures like R-Squared score, Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Some of the key points of this study are listed here:

- KNN and Decision tree algorithms are provided 100% accuracy if the time-series data contains limited entries.
- For various class predictions accuracy of different algorithms seems to be better.
- To predict good accuracy of ML models sufficient quantity of data is needed

The remaining sections of the paper organized as, data and methods in which the complete details of the dataset and methods, Methodology, results, and final conclusion.

2. Materials And Methods

A. Datasets

Required datasets gathered in this study has been taken from the Kaggle[12] sources which are provided by the JHU. In this study we analysis the number of positive cases, death cases, and cured cases of COVID-19 cases in top 10 states of India. The datasets files_Age_roup_Details, covid_19_india,Hospital_Beds_India, and Individual_Details contains 9 records, 8486 records, 37 records, and 5356 records respectively. The files consists of age group data tables (peoples from 0 to 90 years), covid_19_india file contains confirmed, cured and death cases in various states/UTs including Indian and Foreign citizens statistics, Hospital_Beds file consists of number of beds available in different kinds of hospitals and individual_details file contains status of individual citizen of India. In below tables respective dataset samples are presented.

TABLE-1: % Of Age-Group Confirmed Cases In India.

	S no	Age Group	Total Cases	Percentage
0	**1**	**0-9**	**22**	**3.18%**
1	2	10-19	27	3.90%

TABLE-2: India Covid-19 Individual Patient Information**.**

Sno	Date	Time	State/Union Territory	Confirmed Indian National	Confirmed Foreign National	Cured	Deaths	Confirmed
1	30/01/20	6:00 PM	Kerala	1	0	0	0	1
2	31/01/20	6:00 PM	Kerala	1	0	0	0	1

TABLE-3: India Covid-19 State-Wise Cases (Cured, Deaths And Confirmed).

id	govt_id	diagnosed_date	age	gender	detected_city	detected_district	detected_state	nationality	current_status	status_change_date	notes
0	KL-TS-P1	30/01/2020	20	F	Thrissur	Thrissur	Kerala	India	Recovered	14/02/2020	Travelled from Wuhan

B. Machine Learning Algorithms

In general, machine learning techniques are used to predict results [10] from an unknown inputs when if we create a model. The designed algorithm takes input instances from datasets beside with the help of a comparable regressor to instruct the regression model. After taking input, the model produces the corresponding prediction based on the unidentified input dataset (or test dataset). To get efficient accuracy of a model, the learning model uses either regression techniques or classification techniques. From above all considerations in this study we used following ML models for analyzing COVID-19 overall cases in top 10 states of India/UTs.

- Linear Regression
- K Neighbors Regressor
- Lasso Regression
- Elasticnet Regression
- Decision Tree Regressor

1. Linear Regression

In machine learning, to perform analysis on predictions the main useful statistical method is linear regression. To predict the association among dependent and independent values linear regression is used. From [12], through regression technique with help of autonomous features[17] the target group is predicted [13]. In the study of linear regression mainly two values (a-independent, b- dependent) are important. In the equation shown below specifies what way a is associated to b,

$b = \beta 0 + \beta 1 a + \varepsilon$ ---------- (1)

or

$$E(b) = \beta 0 + \beta 1 a - - - - - - (2)$$

Here, ε is error rate which is used to report the inconsistency among (a,b). $\beta 0$ is b_intercept and $\beta 1 -$ slope. To obtain the best-fit regression, discover the most excellent value for $\beta 0$ and $\beta 1$. In linear regression best-fit value is obtained when differentiating among actual values and predicted values. Hence best-fit can be specified as

$$bestfit \frac{1}{n} = \sum_{i=1}^{n} (Pi - bi)^{\,2} - - - - (3)$$

$Cf = \frac{1}{n}\sum_{i=1}^{n}(Pi - bi)^{\,2}$---------- (4)

Here, Cf is a cost function, which is the RMS of the Pi and bi and n specifies number of data points.

2. LASSO Regression

LASSO regression is one type of linear regression model which is works on shrinkage [14].This shrinkage method makes LASSO better and additionally makes stable and reduced errors [15]. In multi-colinearity cases LASSO is considered as a best appropriate model. LASSO uses a regularization technique which means that the features that are not help out the regression outcome sufficient can be put to a extremely little value zero. The LASSO regression adds features one at a time and then it could not be added zero if the fresh attribute does not get better fit sufficiently. To minimize the LASSO regression an objective function is defined in the equation (5) as,

$$LASSO_{min} = \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{bi} - \sum \mathrm{aij}\beta\mathrm{j})2 + \epsilon \sum_{\mathrm{j=1}}^{\mathrm{m}} |\beta\mathrm{j}| - - - - - - (5)$$

Here LASS0min calculated as min (sum of squares) of residuals +ϵ |slope|, where, ϵ |slope| is penalty.

3. KNN Regression

In KNN, "nearest" specifies a distance measure. KNN regression is used to estimate the average of the geometric targets of the K-Nearest Neighbours and it also used to find an opposite distance weighted average of the K- neighbours [18]. It uses the same distance functions to calculate distance between two data points what KNN classification technique is used. The distance function used in K-NN regression is

$M^P(a^i, b^j) = (\sum_D |a_D^i - b_D^j|^{\,P})^{\frac{1}{P}}$---------- (6)

The weighted average distance of $a1, a2, .. ak$ is calculated as,

$Dist_{(Weight\_Avg)} = \sum_{i=1}^{k} \left(\frac{1}{D^k}\right) R^t / (\sum_{i=1}^{k} D^k)$ ----- (7)

It is Euclidean Distance if p=2, it is Manhattan Distance if p=1 and is Hamming distance if Boolean attributes.

4. ElasticNet Regression(ENR)

Elasticnet initial come out for an effect of analysis on LASSO, whose variable choice can be also dependent on data and so it is unbalanced. The resolution is to join the errors of ridge regression and LASSO [18] to get the best results. Main concentration of Elastic Net is to minimizing the loss function [20], is shown in the equation (8),

$$E_{min} = MSE(E) + r\alpha \sum_{i=1}^{n} |min_i| + \frac{1-r}{2} \alpha + \sum_{i=1}^{n} min_i^2 \text{---- (8)}$$

Where Elastic Net is Ridge Regression when r = 0, and if r = 1, it is Lasso Regression.

5. Decision Tree Regression

In this regression, the predictor space is dividing into the set of possible values for X1, X2 . . . , Xp into j separate and non-overlapped areas, A1, A2 . . . , Aj. For each observation that falls into the area Aj, produces the similar prediction [19], which is simply the mean of the response values for the training observations in Aj. Here the main objective is to locate areas A1, A2 . . . , Aj to minimize the RSS by,

$$R_{SS} = \sum_{j=1}^{J} \sum_{i \in A_j} (y_i - \hat{y}\, A_j)^2 \text{----- (9)}$$

Where, the training values inside the j^{th} area are represented by their mean response.

C. Evaluation Metrics

In this learning the performance of every model evaluated with help of the terms R-squared score, Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) respectively.

1. Mean Square Error (MSE)

MSE is another method to calculate the accuracy of regression methods [21]. The data points commencing the regression line are taken and then squaring them. The advantage of squaring is that it eliminates the negative symbol from the results. The smaller MSE shows that the result the line of best fit. MSE can be calculated by using below equation

$$MeanSquareError = \frac{1}{n} \sum_{k=1}^{n} (|y_k - \hat{y}_k|)^2 \text{------ (10)}$$

2. Root Mean Square Error (RMSE)

Root mean square error (RMSE) is the Standard Deviation (std) of the forecasting errors (residuals). Residuals are the gap from the best-fit line and real data points [20]. RMSE is the error rate specified by the square root of MSE which is given as follows.

$$\text{RootMeanSquareError} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (|y_k - \hat{y}_k|)^2} \text{----- (11)}$$

3. Mean Absolute Error (MAE)

In a model, the groups of predictions [22] are calculated through the average of amount of errors. When individual differences have equivalent value, mean absolute error is normal on test data among the model predictions and real data. MAE value ranges from 0 to infinity and only some results show the righteousness of learning methods that's the motive MAE also named negatively-oriented scores [23].

$$\text{Mean Absolute Error} = \frac{1}{n} \sum_{k=1}^{n} |y_k - \hat{y}_k| \text{------- (12)}$$

4. R-Squared Score

R-squared score used to assess the performance of regression techniques [24],[25]. It deals with the connection potency among the dependent value and models is suitable 0 to 100% range. 0% score says the reacted variable has no changeability in its mean, and 100% specifies that the reacted variable has the changeability in its mean. The integrity of fit of the trained models efficient can be finding using the R2 scores. R2 score of a model says the fraction of difference in free variable. It can be represented as

$$R^2 - \text{Score} = \frac{\text{Variance(Model)}}{\text{Total(Variance)}} \text{---------- (13)}$$

3. Methodology

In this study we predict and analyze about novel Coronavirus or COVID-19 cases in top 10 states of India. Also we predict age based and gender based COVID-19 cases. It is a very harmful situation to human life because daily confirmed cases are increased and good news is that the numbers of cured cases are also increased along with confirmed cases. This dangerous virus causes to deaths of thousands of people in India. For controlling epidemic condition, our work tried to make analysis on number of every day confirmed positive cases and the number of recovery cases and death cases in top 10 states/UTs of India.

The prediction analysis performed with help of five ML models that are suitable to this perspective. The dataset taken in this work consists of outline of percentage of age groups are affected, number of confirmed, cured and death cases of Indian national, Foreign national. In very beginning the dataset is preprocessed to discover the states/UTs of India statistics for the daily number of deaths, confirmed, and cured cases.

Covid-19 Cases by Gender and Age

The following figures 1 (a), (b), and (c) are showing the % of people affected by their gender (male-77% and female-23%); % of affected cases by age. Highly affected age group is 20-29 years (24.9%) and number of people is affected by Corona cases in India in one day of age 20-29 group (172 cases).

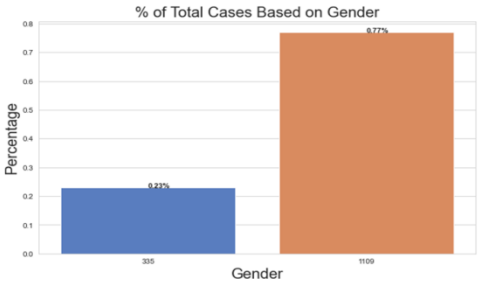


Figure. 1 (a)

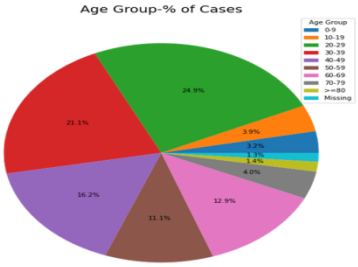


Figure. 1 (b)

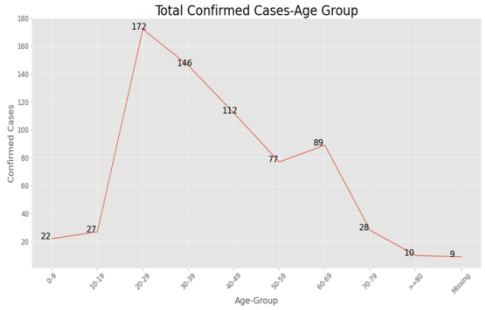


Figure. 1 (c)

Figure. 1(a) % of people affected by gender , (b) % of people affected by age-group and (c) Number of people affected by age-group

Figure 1(a), 1(b), and 1(c) shows the percentage of confirmed cases in India regarding age and gender respectively. Form Figure 1 (a), , it is observed that percentage of confirmed cases maximum for age groups 20-29, 30-49 and 40-49 years respectively. From figure 1 (b), the % confirmed cases when compare to female (23%), it is mostly high for male (77%) category. From Figure 1(c), it shows that number of confirmed cases for age group 20-29 is 172 which is high and for age group 30-39 is 146 is second highest and soon.

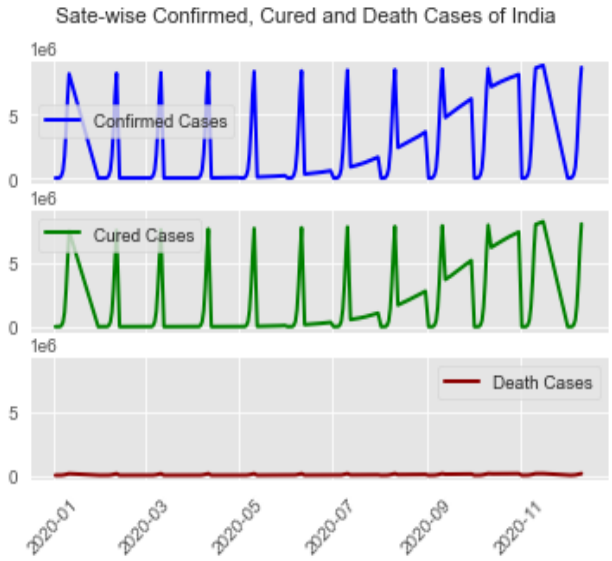
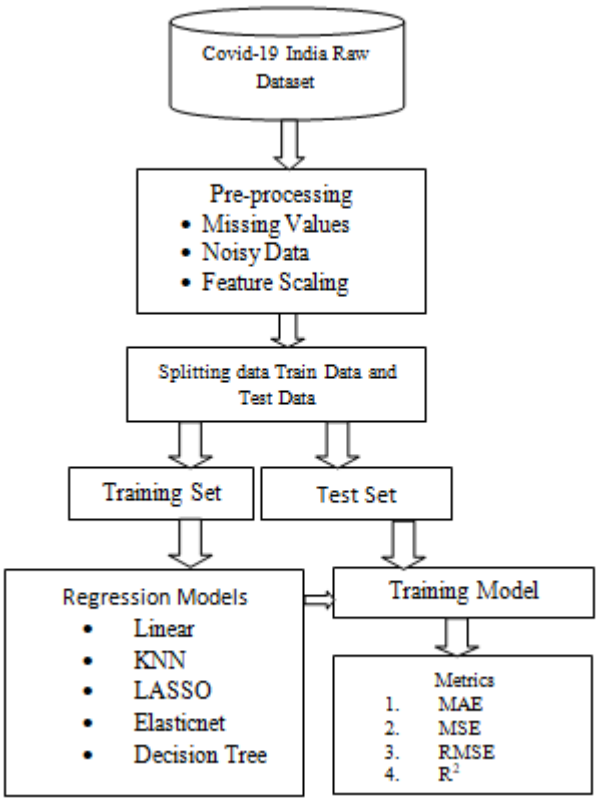


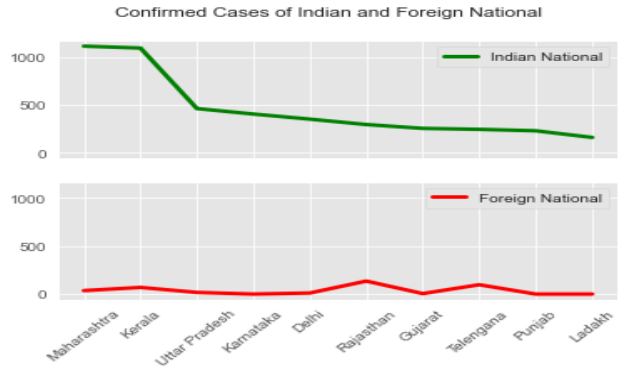
Figure-4: Monthly confirmed, cured and death cases

In figure 4, it shows that the number of confirmed, cured and death cases from January, 2020 to November, 2020 bi-monthly reports. It is observed that numbers of cured cases per confirmed cases are same approximately, but numbers of death cases are very low. Covid-19 Cases: Indian and Foreign Nationals

Table-4 shows a sample data record of cases in State/UTs of India where as from Figure-4, it is observed that the number of confirmed cases are somewhat same as cured cases. And number of death cases is very low comparing to confirmed cases. From table the first case in India is held in Kerala state on 30/01/2020.



Figuare-4: Covid-19 India confirmed cases Indian and Fariegn Nationals



no	Date	Time	State/UT	Confirmed Indian National	Confirmed Foreign National	Cured	Deaths	Confirmed
1	30/01/20	6:00 PM	Kerala	1	0	0	0	1

TABLE-5: Sample Record Of Cases For Indian And Foreign National

The main aim of this study is to try to find the number of people that can be caused by Coronavirus with respect to infected cases, deaths and number of predictable cured cases in the top 10 states/UTs of India. The data used in this work consists of data about the daily cases details of the infected, the number of cured, and the number of deaths cases due to COVID-19 in states/UTs of India. The number of citizens who can be faced by the COVID-19 deadly disease in various states/UTs of the country is not well-known. To predict number of cases (confirmed, cured, and death) we apply different machine learning algorithms such as LR, LASSO, KNN, Elasticnet and Decision Tree are applied on data.

To train the models, first we preprocess the data and then the divide dataset into training set (80%) and testing set (20%). Models have been educated on the confirmed, cured, and death cases patterns. The learning techniques are used *R*2-score, MSE, RMSE, and MAE to evaluate the model accuracy and then required results have been reported. The projected method used in this study is shown in figure.4.

4. Results And Discussion

From figure 5, Maharashtra (28.28% and 26.70%), Tamil Nadu (13.31% and 14.11 %) and Andhra Pradesh (13.31% and 14.01%) are in top 3 positions among Indian states with respect to number of confirmed cases and cured cases. But coming to dead cases the top 3 states are Maharashtra (46.47%), Tamil Nadu (11.92%) and Karnataka (10.20%).

Figure-5: Top 10 states of India with confirmed / cured/ death Cases

A. Model Performance for Confirmed Cases

This learning does prediction on confirmed cases and from the results KNN and Decision Tree regressor produces same accuracy (100%) comparing with other models. LR and LASSO achieve same accuracy (99%) and achieve approximately the similar R2 score. In assessment, Elastic net performs somewhat less accuracy (91%) comparing with other models. The outcomes are represented in Table 6.

TABLE-6. Models performance results for confirmed cases.

S. No	Model	R2-Score	MEAN	MAE	MSE	RMSE
1	Linear	0.99	8297.43	542.27	228433324.54	15114.01
2	KNN	1.00	585.97	62.67	1604060.97	1266.52
3	Lasso	0.99	8296.87	542.24	228432237.87	15113.97
4	Elastic Net	0.91	31505.14	1798.84	2452047616.73	49518.15
5	Decision Tree	1.00	5056.85	388.08	61388206.53	7835.06

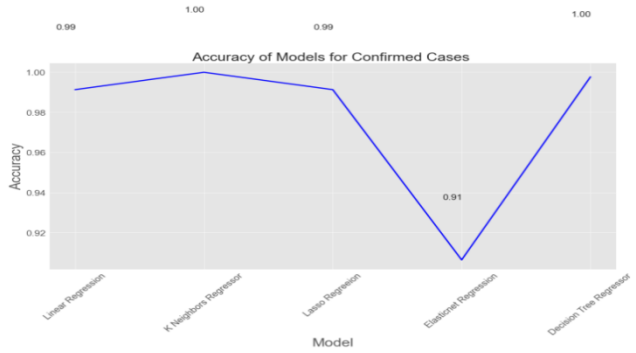


Figure 6. Accuracy of confirmed cases for different models

In figure 6 the accuracy of LR, LASSO, KNN, Decision tree and Elastic net models respectively. In that graph it says that the number of confirmed cases prediction accuracy is high for decision tree and KNN Regressor and Elasticnet produces poor accuracy results (91%).

B. Model Performance for Cured Cases

This learning does predictions on cured cases and from the results Linear Regression and LASSO produces high accuracy (99%) comparing with other models. KNN regressor achieves the accuracy (97%). In assessment, Elastic net performs somewhat less accuracy (88%) comparing with other models. The outcomes are represented in Table 7.

Table-7 Models Performance For Cured Cases.

S. No	Model	R2-Score	MEAN	MAE	MSE	RMSE
1	Linear	0.99	7537.01	540.65	228433324.54	14594.79
2	KNN	0.97	7536.19	540.60	623337130.28	24966.72
3	LASSO	0.99	38375.68	2132.39	213011500.49	14594.91
4	Elastic net	0.88	27618.35	1053.65	2421355081.91	49207.27
5	Decision Tree	0.92	1836.98	281.75	1513595189.29	38904.95

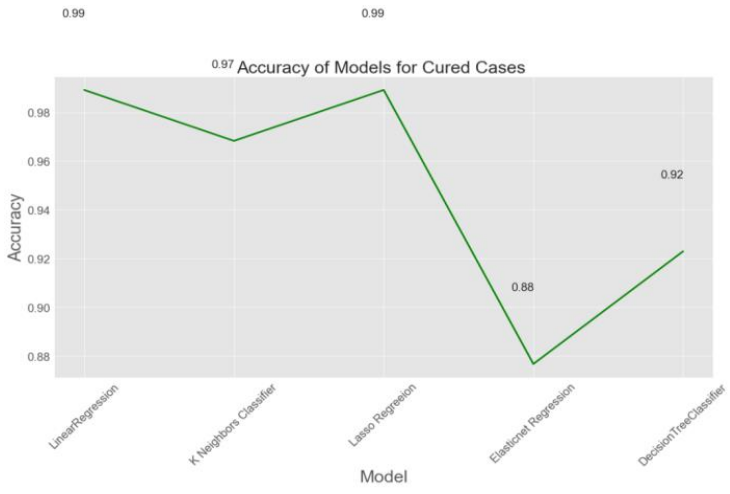


Figure 8. Accuracy of different models for cured cases

In figure 8 the accuracy of LR, LASSO, KNN, Decision tree and Elastic net models respectively. In that graph it says KNN Classifier accuracy (100% high for the number of cured cases and Elasticnet produces poor accuracy results (88%).

C. Model Performance for Death Cases

This learning does predictions on death cases and from the results KNN produces high accuracy (100%) comparing with other models. LR and LASSO achieve same accuracy (99%) and achieve approximately the

similar R2 score. In assessment, Elastic net performs somewhat less accuracy (90%) comparing with other models. The outcomes are represented in Table 6.

TABLE 8. Models Performance For Death Cases

S. No	Model	R2-Score	MEAN	MAE	MSE	RMSE
1	Linear	0.81	779.47	45.17	1958852.88	1399.59
2	KNN	0.98	82.43	17.58	203676.04	451.30
3	Lasso n	0.81	778.83	49.12	1960534.51	1400.19
4	Elasticnet	0.76	643.02	29.17	2425565.76	1557.42
5	Decision Tree r	1.00	28.85	12.57	4258.23	65.26

In figure 9 the accuracy of LR, LASSO, KNN, Decision tree and Elastic net models are represented. In that graph it says that KNN Classifier gives high accuracy (98%) comparing with the remaining models for the number of death cases and Elasticnet produces poor accuracy results (76%).

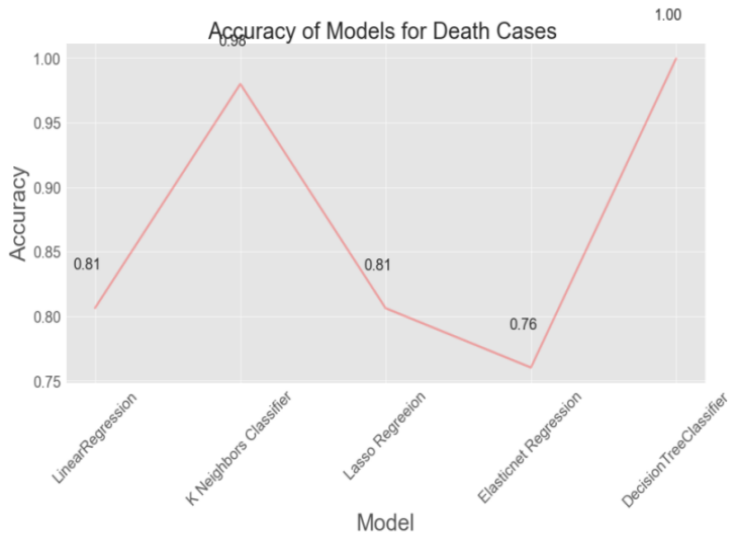


Figure 9. Accuracy of different models for death cases

4. Conclusion

Covid-19 epidemic can put in danger situation with massive instability disaster throughout the universe. Some researchers and public sectors people all over the India and world have unpleasant that the epidemic can persuade a huge number of the humankind [24], [25]. Hence this work proposed a prediction model based on Machine Learning algorithms for predicting the threat of COVID-19 eruption in top 10 sates/UTs of India. The model analyzes datasets containing the COVID-19 cases (confirmed, cured and death cases) up to 24th November, 2020 using ML models. From the results it is proven that Decision Tree and KNN regressor performs best in analyzing the number of confirmed and death cases. But for number of cured cases LASSO and linear regression models give the best accuracy results. Unfortunately, Elastic net produced poor accuracy results due to some changes in original datasets. Finally, we are concluding that those models predictions are based on present situation are accurate which may be supporting to know the future condition. In future we extend this study to find the prediction framework with the help of newly modified models for forecasting. Especially, this work analyzes the calculations based on the exactness rate on a test dataset.

References

1. Deena Babu Mandru and Y.K. Sundara Krishna,"Multi View Cluster Approach to Explore Multi Objective Attributes based on Similarity Measure for High Dimensional Data" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 15 (2018) pp. 12289-12295.
2. V.Assimakopoulos, E.Spiliotis, and S.Makridakis, ``Statistical and machine learning forecasting methods: Concerns and ways forward," PLoSONE, vol. 13, no. 3, Mar. 2018, Art. no. e0194889.

3. Y.A. Le Borgne, S. B. Taieb, and, G. Bontempi, ``Machine learning strategies for time series forecasting," in Proc. Eur. Bus. Intell. Summer School. Berlin, Germany: Springer, 2012, pp. 62_77.
4. T. A. Reichert, F. E. Harrell Jr, D. B. Matchar, and, K. L. Lee, ``Regression models for prognostic prediction: Advantages, problems, and suggested solutions," Cancer Treat. Rep., vol. 69, no. 10, pp. 1071_1077, 1985.
5. L. Labree, S. P. Azen, and P. Lapuerta, ``Use of neural networks in predicting the risk of coronary artery disease," Comput. Biomed. Res., vol. 28, no. 1, pp. 38_52, Feb. 1995.
6. K. M. Anderson, W. B. Kannel, P. M. Odell, and P. W. Wilson, ``Cardiovascular disease risk pro_les," Amer. heart J., vol. 121, no. 1, pp. 293_298, 1991.
7. H. A. Moatassime, H. Asri, H. Mousannif, and T. Noel, ``Using machine learning algorithms for breast cancer risk prediction and diagnosis," Pro- cedia Comput. Sci., vol. 83, pp. 1064_1069, Jan. 2016.
8. S. Makridakis and F. Petropoulos, ``Forecasting the novel Coronavirus COVID-19," PLoS ONE, vol. 15, no. 3, Mar. 2020, Art. no. e0231236.
    A. Pesenti , G. Grasselli, and M. Cecconi, ``Critical care utilization for the COVID-19 outbreak in lombardy, italy: Early experience and forecast duringan emergency response," JAMA, vol. 323, no. 16, p. 1545, Apr. 2020.
9. Deena Babu Mandru and Y.K. Sundara Krishna,'Enhanced Cluster Ensemble Approach Using Multiple Attributes in Unreliable Categorical Data', International Journal of Psychosocial Rehabilitation, Volume-3, issue-1, p.254-263, 2019.
10. https://www.kaggle.com/sudalairajkumar/covid19-in-india? select = covid_19_india.csv
11. https://www.kaggle.com/ankitranjann/individualdetails.csv
12. M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, ``Analytics defined," in Information Security Analytics, M. R. M. Talabis,R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, Eds. Boston, MA, USA: Syngress, 2015, pp. 1_12. [Online].
13. Furqan Rustam Aijaz Ahmad Reshi , (Member, Ieee), Arif Mehmood, Saleem Ullah , Byung-Won On, Waqar Aslam , (Member, Ieee), And Gyu Sang Choi ,``COVID-19 Future Forecasting Using Supervised Machine Learning Models," J. Eval. Clin.Pract., vol. 14, no. 2, pp. 275_280, Apr. 2008.
14. R. Tibshi rani, ``Regression shrinkage and selection via the lasso," J. Roy. Stat. Soc., Ser. B, Methodol., vol. 58, no. 1, pp. 267_288, Jan. 1996.
15. R. Tibshi rani, ``Regression shrinkage and selection via the lasso," J. Roy. Stat. Soc., Ser. B, Methodol., vol. 58, no. 1, pp. 267_288, Jan. 1996.
16. Y.K. Sundara Krishna Deena Babu Mandru,' Enhanced Feature Selection Clustering Algorithm for Attribute Similarity in High Dimensional Data' , 2018, International Journal of Engineering & Technology, Volume-7,Pages 688-693
17. Zizhen Yao,Allen Institute for Brain Science, Walter L Ruzzo,University of Washington Seattle, "A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data", 20 March 2006, BMC Bioinformatics 2006, 7(Suppl 1):S11
18. Engin Pekel 'Estimation of soil moisture using decision tree regression, Theoretical and Applied Climatology volume 139, pages1111–1119(2020)
19. Geoffroy MOURET, Jean-Jules BRAULT , Vahid PARTOVINIA, "Generalized Elastic Net Regression", JSM 2013 - Section on Statistical Learning and Data Mining, , pp. 3457–3464
20. R. Kaundal, A. S. Kapoor, and G. P. Raghava, ``Machine learning techniques in disease forecasting: A case study on rice blast prediction," BMC Bioinf., vol. 7, no. 1, p. 485, 2006.
21. K. Matsuura and C. Willmott ,``Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Res., vol. 30, no. 1, pp. 79_82, 2005.
22. J. Lupón, H. K. Gaggin, M. de Antonio, M. Domingo, A. Galán, E. Zamora, J. Vila, J. Peñafiel, A. Urrutia, E. Ferrer, N. Vallejo, J. L. Januzzi, and A. Bayes-Genis, ''Biomarker-assist score for reverse remodeling prediction in heart failure: The ST2-R2 score,'' Int. J. Cardiol., vol. 184, pp. 337–343, Apr. 2015.
23. J.-H. Han and S.-Y. Chi, ''Consideration of manufacturing data to apply machine learning methods for predictive manufacturing,'' in Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN), Jul. 2016, pp. 109–113.
24. N. C. Mediaite. Harvard Professor Sounds Alarm on 'Likely' Coronavirus Pandemic: 40% to 70% of World Could be Infected This Year. Accessed: Feb. 18, 2020. [Online]. Available: https://www.mediaite.com/news/ harvard-professor-sounds-alarm-on-likely-%coronavirus-pandemic-40-to-70-of-world-could-be-infected-this-year/

25. "BBC. Coronavirus: Up to 70% of Germany Could Become Infected,"Merkel, Accessed: Mar. 15, 2020. [Online]. Available:https://www.bbc.com/news/world-us-canada-51835856