iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.46133

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII August 2022- Available at www.ijraset.com

Covid Future Prediction Using Machine Learning

Sai Sri Bhavya Lakkakula[1], Sai Naga Durga Maddukuri[2], Shivani Neelam[3], Dr. M. Ramasubramanian[4]
[1, 2, 3]*I,II,III,IV year B.Tech,* [4]*Professor, Department of Computer Science and Engineering Sridevi Womens Engineering College Hyderabad*

Abstract: ***The Machine Learning models have long been used in numerous software application domains which demanded the identification and prioritization of adverse factors for a threat. Several prediction methods of machine learning domain are being popularly used to handle forecasting or soothsaying problems.Machine learning (ML) based forecasting methods have proved their significance to anticipate in perioperative outcomes to improve the decision making on the future course of actions. This study substantially demonstrates the capability of Machine Learning models to forecast the number of forthcoming patients,death cases and also recovered cases of COVID-19 which is presently considered as a implicit trouble to humanity. In particular, four standard forecasting supervised machine learning models, such as linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and exponential smoothing (ES) have been used in this study to forecast the dangerous factors of COVID-19 that we have previously discussed. Three types of predictions are made by each of the models, such as the number of recently infected cases, the number of death, and the number of recoveries in the coming 10 days. The results have been proved that the exponential smoothing (ES) performs exceptionally well among all the used models followed by LR and LASSO which performs better in forecasting the new confirmed cases, death rate as well as recovery rate, while SVM performs inadequately in all the prediction scenarios given the available dataset that has been collected from various resources.***

Keywords: Linear Regression (LR); Least Absolute Shrinkage and Selection Operator (LASSO); Support Vector Machine (SVM); Exponential Smoothing (ES).

I. INTRODUCTION

Machine learning has gained its significance it many of the real time software application and also in many areas.Machine learning (ML) is a sub-domain of artificial intelligence (AI) that allows software operations, websites to become more accurate in predicting outcomes without being explicitly programmed to do so. We just need to input the data and it depends on algorithms, it might use historical data, raw data, structured data,unshaped data or else semi structured data as input to predict new output values. The type of the data that it takes as input purely depends upon the source from which the data is being collected. So we need to preprocess the data using data mining algorithms and methods in order to predict the most accurate outcomes. Here preprocess includes the cleaning of data such as removing redundant values, empty values, errors etc. so by preprocessing the data the time for performing the task decreases and efficiency in the output increases. pivotal point is that the training data percentage should be more than testing data, as the data it get trained more the output will be that accurate.

A. What Makes Machine Learning Important In Current Trend

Machine learning is important because it gives enterprises,organizations a view of trends in customer behavior and business operational patterns, as well as supports the development of new productsin the market. Many of today's leading companies, such as Facebook, twitter,instagram,Google and Uber, make machine learning a central core part of their operations.these operations can include tuning the customer needs, going throuhg the customer reviews about the products etc. Machine learning has become a significant competitive differentiator for many companies. It also gives us useful insights about the customer interests in e-commerce websites like amazon, flipkart etc. it also gains its importance in iris scanning, sentimental analysis, object recognition etc.

B. Different Types Of Machine Learning

Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches where each approach has its significance:

©IJRASET: All Rights are Reserved

The type of algorithm data scientists choose to use depends on what type of data they want to predict and also the type of data that we have collected.

1) *Supervised learning:* Here, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Here every data will be labeled and the algorithms get the clear picture of the objects in the environment.Both the input and the output of the algorithm is specified.
2) *Unsupervised learning:* This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through tons of data sets looking for any meaningful connection between the data values or domain values. It makes the prediction by categorizing the attributes, characteristics of the objects and forms the different clusters based on the attributes. through the clustering technique process gets much easier than before.
3) *Reinforcement learning:* Here in case of reinforcement learning an agent is put into an environment and it learns by itself without being explicitly programmed to do so. it generally learns through awards and rewards it gains by performing the certain actions in the environment. Same methodology is being followed in algorithms too.

Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative clues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way. It's the best learning method. Here it learns by itself through rewards it gets through the actions that it performs.

II. PROPOSED SYSTEM

Here, we proposed a prediction machine for COVID-19. Predictions of 3 key primary variables are revamped the following 10 days.

1) The range of newly shown cases.
2) Number of deaths.
3) Number of restores.

This prediction hassle became considered a regression hassle on this examine, so this examine is primarily based totally on cutting-edge supervised ML regression models including linear regression (LR), least absolute shrinkage and selection operator (LASSO), Support vector Machine (SVM) and exponential smoothing (ES).

The training version became educated the usage of the COVID-19 affected person data dataset furnished via way of means of Johns Hopkins. The dataset became preprocessed and cut up into subsets: the training set (85%) and the testing set (15%). Performance checks encompass key R-squared scores (R2-score), adjusted R-squared scores (R2-adjusted), suggest squared error (MSE), suggest absolute error (MAE), and suggest squared error (RMSE). It was carried out with a metric.

III. ARCHITECHTURE

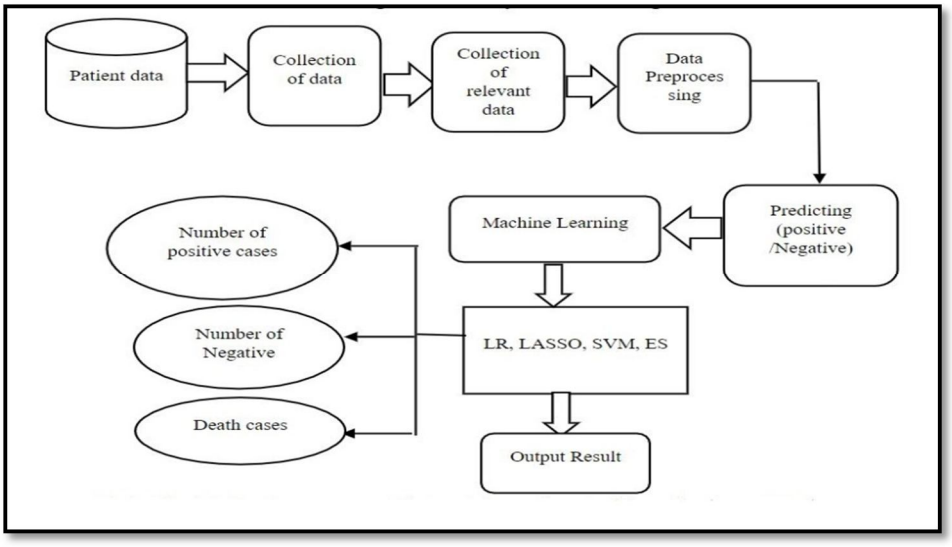


Figure1Architecture of the proposed system

1) Initially we collect the patients data from various sources such as governamentsites, hospitals, kangle datasets etc., the data that we collected will be enormos and it leads to huge amount of testing data. As we know more the testing data the result will be more accurate. The data that we have accumulated can be structured, semi-structured, unstructured, alphanumeric etc.,
2) All the data collected will be stored in databases, datawarehouses,servers etc., in order to ease the retrieval,updation,deletion,insertion etc., tasks.
3) The data that we have collected might also contain the fields that are not required for our project so we collect only project relevent data.
4) Then we will preprocess the relevant data.It includes -

a) *Data Cleaning:* The data that we have collected might have missing data,redunduncy etc., so we clean the data in such a way that it leads the results to be error free.
b) *Data Transformation:* It is a technique used to convert the raw data into a suitable format. The algorithm that we use might need the data in the specified format so in this step we convert the data into required format.
c) *Data Reduction:* Here we will truncate the data by removing the garbage values of the fields present in the data.

5) Below are four unsupervised machine learning algorithms that will be used in project .

a) Linear Regression
b) LASSO Regression
c) Support Vector Machine
d) Exponential Smoothing

By using the above four algorithms we predict the below three variables which helps us in covid prediction.

- Daily death rate
- New cases
- Recovery Cases

IV. IMPLEMENTATION

1) Linear Regression: Linear regression evaluation is used to expect the cost of a variable primarily based totally at the cost of any other variable. The variable you need to expect is referred to as the established variable. The variable you're using to expect the other variable's cost is referred to as the independent variable.

Y=a1+b1m

a1 and *b*1 are given by the following formulas:

$$a1(intercept) = \frac{\sum n \sum m^2 - \sum m \sum mn}{(\sum m^2) - (\sum m)^2}$$

$$b1(slope) = \frac{l\sum mn - (\sum m)(\sum n)}{l\sum m^2 - (\sum m)^2}$$

Where,

m and *y* are two variables on the regression line.

*b*1= Slope of the line.

a1 = *n*-intercept of the line.

m = Values of the first data set.

n= Values of the second data set.

2) *LASSO Regression:* Lasso regression is a form of linear regression that makes use of shrinkage. Shrinkage is in which information values are shriveled in the direction of a primary factor, just like the meanPerforms L1 regularization, i.e. provides penalty equal to absolute price of the magnitude of coefficients.
Minimization objective = LS Obj + α * (sum of absolute price of coefficients)
3) *Support Vector Machine:* The SVM set of rules aims to create the great line or choice boundary which could segregate n-dimensional space into training so that we can easily position the new information point in the appropriate class. This great selection boundary is referred to as a hyperplane.
4) *Exponential Smoothing:* Exponential smoothing is a rule of thumb approach for smoothing time series data the use of the exponential window function.

Whereas in the simple moving average the past observations are weighted equally, exponential features are used to assign exponentially lowering weights over time.

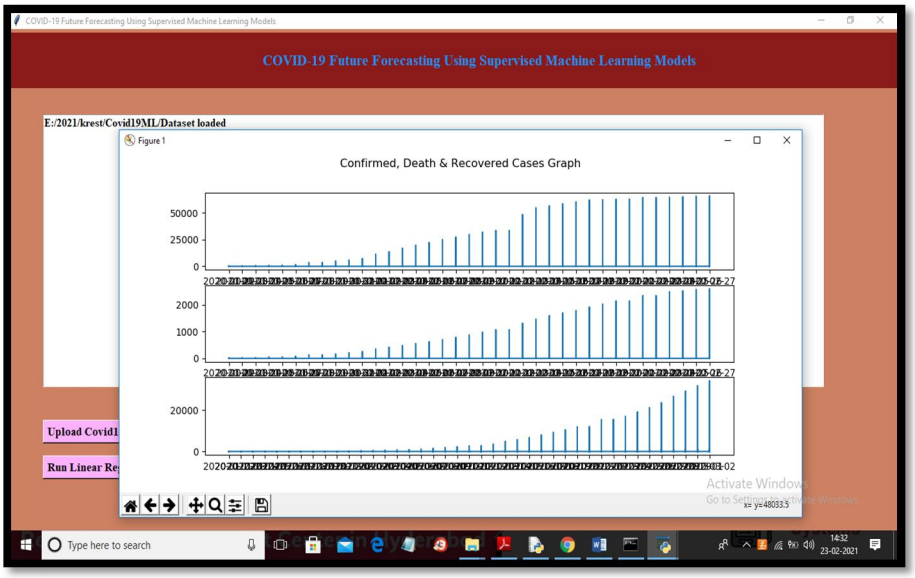
$$pt+1 = \alpha st + (1-\alpha)\, pt$$

Where pt = The smoothed value of time series at time t

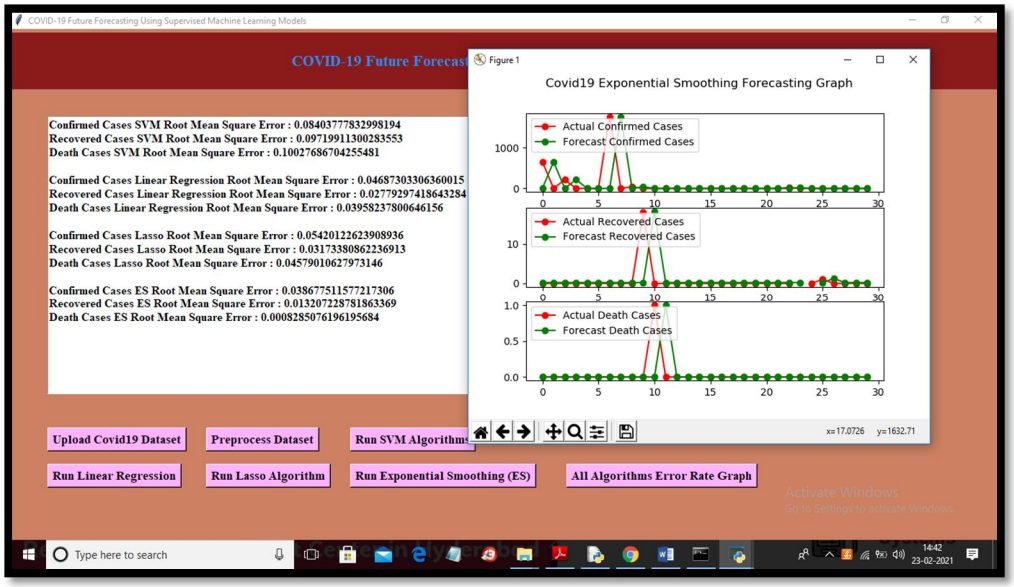
st =Actual value of time series at time

α =Smoothing constant at time t

V. RESULTS



In above screen dataset stacked and afterward we are plotting number of affirmed, recuperated and demise cases in the above chart. In above chart x-hub addresses date and y-pivot addresses number of cases on that date. In the above chart dataset size is colossal so date values get impacted and presently close above diagram and afterward click on 'Preprocess Dataset' button to read dataset and to extract details from it.



we can see ES error rate and its forecast values has close difference with actual values so ES also working well and we can see compare to all algorithms ES is having less error rate so its better compare to other algorithms.

Now click on 'All Algorithms Error Rate Graph' button to get above graph.

VI. CONCLUSION

The trickiness of the COVID-19 pandemic can touch off a gigantic worldwide emergency. A few specialists and government organizations all through the world have fears that the pandemic can influence an enormous extent of the total populace. In this review, a ML-based expectation framework has been proposed for anticipating the gamble of COVID-19 episode universally. The framework examinations dataset containing the day-wise real past information and makes expectations for impending days utilizing AI calculations.

The aftereffects of the review demonstrate that ES performs best in the ongoing guaging area given the nature and size of the dataset. LR and LASSO additionally perform well for guaging somewhat to anticipate demise rate and affirm cases. As per the aftereffects of these two models, the passing rates will increment in impending days, and recuperations rate will be dialed back. SVM produces unfortunate outcomes in all situations as a result of the high points and low points in the dataset values. It was truly challenging to put an exact hyper plane between the given upsides of the dataset. Generally we presume that model forecasts as per the ongoing situation are right which might be useful to grasp what is going on.

The review gauges consequently can likewise be of extraordinary assistance for the specialists to make ideal moves and go with choices to contain the COVID-19 emergency. This study will be improved constantly later on course, next we intend to investigate the expectation procedure utilizing the refreshed dataset and utilize the most reliable and suitable ML techniques for determining. Continuous live guaging will be one of the essential concentrations in our future work.

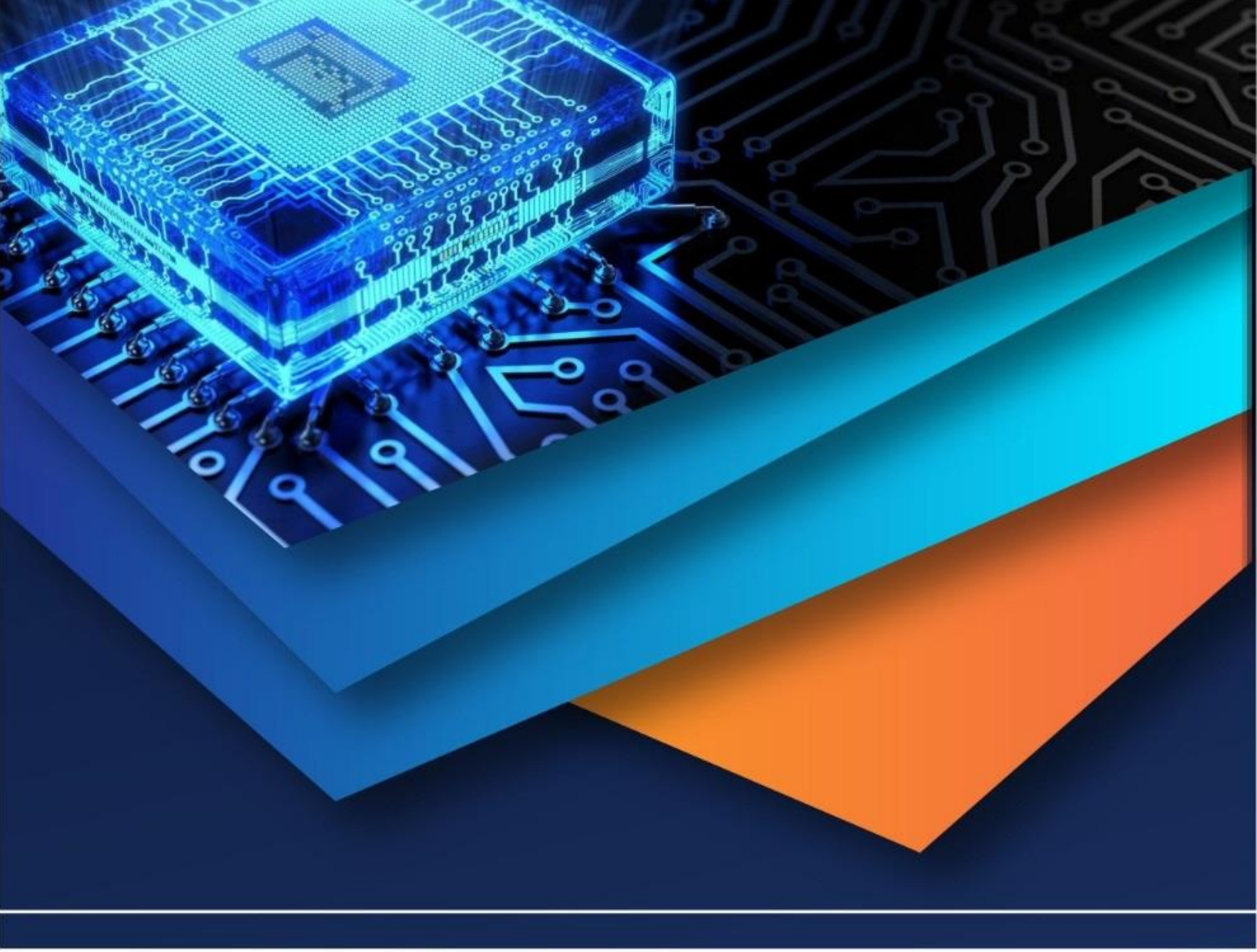
VII. ACKNOWLEDGMENT

The authors would like to thank the anonymous authors for their time and effort reviewing this paper.

REFERENCES

[1] "Statistical and machine learning forecasting methods: Concerns and ways forward," by S. Makridakis, E. Spiliotis, and V. Assimakopoulos.
[2] WHO database named 'Naming the coronavirus disease (covid-19) and the virus that causes it'.
[3] "Biomarker- assist score for reverse remodeling prediction in heart failure: the st2-r2 score," International journal of cardiology, vol. 184, pp. 337–343, 2020 by J. Lupón, H. K. Gaggin, M. de Antonio, M. Domingo, A. Galán, E. Zamora, J. Vila, J. Peñafiel, A. Urrutia, E. Ferrer et al.
[4] "Consideration of manufacturing data to apply machine learning methods for predictive manufacturing,"which has being written in the year 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN) by J.-H. Han and S.-Y. Chi.
[5] "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," Climate research, vol. 30, no. 1, pp. 79–82, 2020, authors of the above journal are C. J. Willmott and K. Matsuura
[6] "Machine learning tech-niques in disease forecasting: a case study on rice blast prediction," by R. Kaundal, A. S. Kapoor, and G. P. Raghava.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)