Proceedings of the Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
IEEE Xplore Part Number: CFP21OSV-ART; ISBN: 978-1-6654-2642-8

2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 978-1-6654-2642-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I-SMAC52330.2021.9640960

Prediction of the COVID-19 pandemic with Machine Learning Models

1st P. Lakshmi Sruthi
MTech Student
Computer Science department
GRIET, Hyderabad, Telangana, India.
plsruthi@gmail.com

2nd Dr. K. Butchi Raju
Professor
Computer Science department
GRIET, Hyderabad, Telangana, India.
Raju_katari@yahoo.co.in

***Abstract*— The latest destructive outbreak, Corona virus (2019), is rapidly sweeping the globe. Not only are economies deteriorating, but countries' entire strengths and confidence are as well. Machine learning forecasting strategies have demonstrated their importance to anticipate in outcomes of the perioperative period to improve the future decision-making actions. The machine learning algorithms have long been used in several applications which require the detection of adverse factors for a threat. Forecasting techniques are essential for producing accurate results. This study shows the ability to predict the number of cases affected by COVID-19 as potential risk to mankind. In this analysis, four prediction algorithms have been used which are linear regression (LR), Exponential Smoothing (ES), least absolute shrinkage and selection operator (LASSO) and support vector machine (SVM). Each of these models has three different kinds of predictions, such as the newly infected patients, death cases and the recovery cases in the next ten days. These approaches are better used to forecast the covid-19 pandemic, as shown by the findings of analysis. The ES, that is effective in forecasting new corona cases, death cases and recovery cases.**

Keywords—*Pandemic, COVID-19, Corona virus, R2 Score Adjusted, Exponential Smoothing, Machine Learning Supervised.*

I. INTRODUCTION

In the last decade, machine learning (ML) has become a popular area of study for a variety of complicated and advanced challenges. In contrast to the traditional algorithms, which use decision statements such as if-else, ML Algorithms/Techniques frequently learn by trial - and - error. One of the most essential components of machine learning is forecasting. A range of widely used ML algorithms have been used to influence future activities in this subject [1].

The global impact of the novel (corona virus) COVID-19 needs precise patient predictions along with mortality and recovery rates research. Prediction, on either way, needs very huge portion of historical information. At the similar period, no prediction can be made with certainty because the future rarely repeats itself. This study puts out the timeline for a practical forecasting activity with major consequences for planning and decision-making, as well as objective estimates for confirmed COVID-19 instances.

The following are some of the *study's significant findings:*
• When the period dataset includes a small quantity of data, ES works well.
• Distinct machine learning algorithms appear better at predicting different classes.
• Most machine learning algorithms require a large quantity of data to forecast the future; as the dataset grows larger, the model's efficiency increases.
• For decision-makers battling pandemics like COVID-19, ML model-based prediction could be quite valuable.

The rest of overall work is broken down into 6 pieces. The introduction is offered in Part I, followed by a description of the dataset and methods used in this work in Part II. Part III presents the proposed approach, Part IV discusses the results, and Part V summarises the work and provides the conclusion.

A. Covid 2019

On February 11, 2020, the novel disease has been labeled "(SARS-CoV-2) Severe Acute Respiratory Syndrome Corona Virus 2" by ICTV, and "Corona Virus"(Covid-19) by the (World Health Organization) WHO. This pathogen was given this name because it is genetically linked to the corona virus that caused the SARS outbreak in 2003. Although the two viruses are related, they are not similar. Corona viral infections (COVID-19) began in Wuhan, China, and also have progressed quickly throughout the country [2].

Prolonged body interaction, respiratory secretions, and contacting exposed items are the most common ways for the pathogen to circulate. One of most difficult element of its transmission is that an individual might be affected by the disease for days despite experiencing suffering. Nearly many areas are affected by it due to the causes of its spread and the danger it poses. Every country has declared full/ partial curfews.

With the commitment and cooperation of G20 countries, WHO's worldwide aims are to help every country vaccinate at least 10% of its population by the end of this month, at least 40% by the end of the year, and 70% of the world's population by the middle of next year [3]. While adapting to their new human hosts, SARS-CoV-2, like some of the other RNA virus, is susceptible to biological transformation with the formation of mutations over time, resulting in mutation variants with different features than their parental types. Several SARS-CoV-2 variations have now been identified during the pandemic, however only a fraction are classified variants of concerns by the WHO due to their worldwide public health impact.

B. *Covid-19 Varinats and Vaccination*

The world is in the grip of a COVID pandemic (2019). While the World Health Organization (WHO) and its partners fight to contain the epidemic, advise on critical treatments, and distribute critical healthcare supplies to individuals in need, they're also working to develop and deploy safe and effective vaccines.

Every year, vaccines save thousands of lives.COVID-19 vaccinations are effective and safe in preventing individuals from becoming dead or dying (In addition to social distancing, maintaining mask and sanitizing). Gamaleya: Sputnik V, Bharat Biotech: Covaxin and covisheild vaccines are used mostly in India. WHO has approved 7 vaccines for use, type of vaccine details can be found at "https://covid19.trackvaccines.org/types-of-vaccines/".'NRVV' stands for Non Replicating Viral Vector, RNA, and full form stands for ribonucleic acid [4].

Table I-1 WHO approved vaccines.

Vaccine	Codenamed	Referred also as	Type	Approved
Moderna	mRNA-1273	Spikevax	RNA	71 countries
Pfizer/BioNTech	BNT162b2	Tozinameran, Comirnaty	RNA	98 countries
Johnson & Johnson	Ad26.COV2.S, Ad26COVS1, JNJ-78436735	Janssen	NRVV	63 countries
Oxford/AstraZeneca	AZD1222	Vaxzevria	NRVV	121 countries
Oxford/AstraZeneca formulation(serum institute of India)	AZD1222	Covishield	NRVV	45 countries
Sinopharm (Beijing)	BBIBP-CorV	Covilo	Inactivated	64 countries
Sinovac	CoronaVac	__	inactivated	40 countries

According to the WHO's most current epidemiological update, four SARS-CoV-2 VOCs (Variants of Concerns) have been discovered since the beginning of the pandemic as of June 22, 2021 [13]:

1) Alpha [B.1.1.7]: In late December 2020, the United Kingdom (UK) reported the first variant of concern.

2) Beta [B.1.351]: first December 2020, South Africa reported the first version of concern.

3) Gamma [P.1]: In early January 2021, Brazil reported the first variant of concern.

4) Delta [B.1.617.2]: In December 2020, India reported the first variant of concern.

C. *Related Work:*

In the academic literature, machine learning (ML) methods have been offered as alternatives to statistical methods for time series forecasting. The goal of this research is to assess such performance over a variety of forecasting horizons using a subset of 1045 monthly data series from the M3 Competition. When we compared the post-sample accuracy of eight prominent ML algorithms to that of eight classic statistical methods, we discovered that the former consistently outperformed the latter across all accuracy measures and forecasting horizons. The study describes the findings, explains why ML models are less accurate than statistical models, and suggests some possible next steps. The empirical findings of our study highlight the necessity for objective and unbiased techniques to evaluate the effectiveness of forecasting methodologies, which can be accomplished through large, international events that allow for meaningful comparisons and conclusions [1].

The goal of this research paper is to provide a complete overview of the epidemiology, pathogenesis, patient characteristics, etiology, tools of diagnostics, and the most recent new medications for COVID-19 treatment. This paper also includes a brief description of the many SARS-CoV-2 mutations and the efficacy of several existing vaccines for COVID-19 and its mutations prevention [13].

In this article, researchers used time series models to create statistical projections for confirmed COVID-19 cases and examined the trend of recovered cases. Exponential smoothing model with multiplication error and multiplication trend components was employed in the methodology. In the case of significant, negative tilted estimates, persistent prediction deviations should be linked to modifications in observable trends but also the demand for extra measures and actions [16].

Scholars used currently accessible datasets to make predictions and used the optimal machine learning model for each dataset [5, 6, 7, 8, and 9]. In order to evaluate transmission levels in Italy and China, the Weibull-equation, Hill-equation, and Logistic-equation were used to develop a model [10]. This study uses statistical research to explore the impact of ecological elements upon that spreading disease COVID -2019. 3 ecological specifications were included in this model, wind speed, maximum temperature and relative humidity. COVID-19 spreading seems to have no association with temperature or weather rate, according to the findings. The study [11] suggested a model that employed patient records as well as contained a hybrid model, gradient boost trees, and logistic regression. The results of such abovementioned programs can assist throughout the design of administration plans for the execution of therapies to help minimize the disease transmission. The findings of research are represented in Table 2.

Table 1 Covid-19 forecasting on Machine Learning Models.

Work ref.	Studied regions	Data source	Parameters	Remark
[5]	China	Minimal data source	Feedback Corrections of models	Predicting Expected statistics of Corona Virus
[6]	China	Chinese Center for Disease Control and Prevention	The expenses of therapy and isolation, the expected number of candidates, and the number of identified COVID-19 cases	Guideline towards choosing a decision
[7]	China	WHO	Death toll on a daily basis	Prediction of death toll
[8]	102 countries	WHO	The quantity of treatment and the time needed to begin the therapy.	The effect of citizen prevention strategies on the wide impact of a disease
[9]	China	2003 SARS Data	Death Toll	Prediction of death toll
[10]	European Countries and China	WHO	Degree of transmission	Forecasting the degree of transmission
[11]	Global Data	International Classification of Diseases	Preexisting health issues(Blood Pressure & Sugar)	Identify the people who are the most vulnerable.

The study's main aim to improve the inter - operability of machine learning algorithms with Internet of things (IOT) when engaging among the community as well as their environments to minimize COVID (2019). The study looks at various architectures for generating, capturing, storing, and analysing the data using ml techniques. These algorithms can help recognize, inhibit, and track COVID-19 spread in smart areas, along with give researchers a better knowledge about the infection. Additionally, analyses mostly on application of machine learning at health care facilities to support mostly in combat over COVID (2019) were featured in the paper. The study takes a close look at the key components that go into merging machine learning with other AI-based solutions. The part of this analysis contains a broad outline into the aspects needed to integrate ml algorithms with alternative Intelligence systems. Smart cities' technology as well as information technology infrastructure creates a wide range of data types. The *statistical data (1)* frequently contains daily statistics such as the amount of confirmed rates, positive rates, deaths, and healed rates. The *epidemiological data (2)* includes predominantly every diagnostic medical report for prescription meds, clinical studies, the personal health background, and the person's sensitivity to different prescription drugs, among other things. The data generated by smart city sensors and cameras in *real-time surveillance (3)*. High Temperature is often early COVID-19 indications to be detected [21].

The multilayer perceptron (MLP) is a feed-forward artificial neural network (ANN) composed of layers of neurons that is fully coupled. MLP is also known for developing high-quality models with less training time than more advanced techniques. Hyper parameters (for example, the learning rate for training a neural network) are parameters that define the architecture of an ANN model. For a slightly elevated model, the max parametric choices should be accurate. To discover the ideal hyper - parameter combo, the grid search procedure has been developed. A time-series data source is converted into a regression data source to train a multilayer perceptron MLP of artificial neural network (ANN). The purpose of training is to develop a global model that contains as many patients as possible from all locations in each time unit. The robustness of the deceased patient model is strong, whereas that of the confirmed patient model is average, and when recovered patient scenario is weak. The MLP is the most accurate machine learning approach, followed by the Bayesian Neural Network (BNN) and Gaussian Processes (GP). The alternative approaches' sMAPE is not in the single digits, signifying a major variation in accuracy. [22].

Using COVID (2019) lung x-rays and the (HOG) histogram oriented gradients feature based methodology, researchers have developed an efficient classification technique for reliably detecting COVID (2019) virus strains. It achieves high outcomes via utilising precise COVID (2019) new disease classification relating to medical scans. Further, the efficiency of the Convolution Neural Network (CNN) classification method for healthcare images has also evaluated using several edge based artificial networks. The efficiency of final classification using decreases as the amount of classes in the trained network increases. Finally, a 10 fold cross-validation analysis with confusion metrics was performed to detect various conditions such as lung infection. Transfer learning obtained a third classification performance of 85 percent, which included healthy, COVID (2019) positive, and Lung Inflammation [23].

To estimate the overall number of COVID (2019) cases at the county division throughout the United States, researchers created the neural recurrent models built on Long Short Memory LSTM. Our algorithm takes the demography of the counties, while also prior daily sociocultural interaction and COVID (2019) records, and estimates the total number of COVID-19 instances in 2 weeks. When tested on the

timeframe beginning August 1, 2020, till January 22, 2021, this analysis generated a positive correlation between actual and predicted values [24].

II. METHODS & MATERIALS

A. Data Source

This purpose of this study aim here is to predict how COVID19 will spread in the future, with an emphasis on the number of newly confirmed patients, mortality, and recoveries. Data for this inquiry came from the GitHub Repository of the (CSSE) 'Center for Systems Science & Engineering', (JHU) 'Johns Hopkins University '[12]. The ESRI Living Atlas Team assisted the university in making the repository accessible for the 2019 New Corona virus graphical dashboards. On the Git-Hub Repository, Data source records may be located in the (csse covid 20 19 time - series data) Supervised Machine Learning Models.

B. Supervised ML Models

When given an unexpected inputs occurrence, then supervised machine learning model is developed to generate a forecast. When develop the regression model, the training process takes a dataset with incoming occurrences and their matching regressor. After that, the training classifier provides a forecast using the unexpected source information or testing data source [14]. For the building of prediction models, this learning method might utilize regression algorithms and classification methods. Therefore, research using COVID (2019) prediction, 4 regression algorithms were applied.

1. Linear Regression: In machine learning, linear regression is the most often used statistical procedure for forecasting. A linear association between the dependent and relationship between the independent variable is determined through linear regression. 2 values 'x' and 'y' are employed in linear correlation analyses. Mathematical equations below represent the y-x connection, popularly called as regression.

$$y = \beta 0 + \beta 1x + \varepsilon \quad (1)$$

$$E(y) = \beta 0 + \beta 1x \quad (2)$$

Here, ε represents error term (variability between y and x), β0 is the y-intercept and ß1 is the slope.

For the goal of model training of the linear regression in the framework of machines research, a class label is defined in the input data set. The goal is to determine the best 0 (intercept) and 1 (coefficient) values to obtain optimal regression line. To ensure that this minimization solution is shown, the differential between the real values and the estimated values should be as little as possible [19]:

$$Minimize \frac{1}{n} \sum_{i=0}^{n} (predi - yi)\ 2 \quad (3)$$

$$g = \frac{1}{n} \sum_{i=1}^{n} (predi - yi)\ 2 \quad (4)$$

Cost-function (g) is the root mean squared error of the estimated value of' 'y' (predi) and real value of 'y' (yi), and n represents set of sample units.

2. Least Absolute Shrinkage & Selecton Operator:'LASSO' is a sort of shrinkage-based regression model. Data variables are shrunk towards a centralized location, such as the mean, in shrinkage. Simple, sparse modelling are encouraged by the lasso approach which having few parameters. As a result of the shrinking process, LASSO becomes better and more stable, as well as reducing error .This regression is ideal for modelling techniques with a lot of multi – co linearity (when two or more independent variables are highly correlated with one another in a regression model) as well as when we wish to automate elements of the modelling selection process, such as selection of variable or parameter removal. Lasso is a type of linear regression in which the model is penalised for the sum of the weights' actual values. Ridge goes a bit further and penalising the model for the weights' sum of squared values. Weights were divided more equally in groups. Lasso generally gives sparse weights and also most zeros due to L1 Regularization (15). During training, the objective function is changed to:

$$\sum_{i=1}^{n} (yi - \sum_{j} xij\beta j)\text{^}2 + alpha \sum_{j=1}^{p} |\beta j| \quad (5)$$

Alpha (the coefficient) term refers to penalize weights.

3. Exponential Smoothing: Forecasting is done using data from prior periods in the exponential smoothing family of approaches. As time passes, the influence of previous data observations diminishes exponentially. As a result, the weight allocated to various lag values decreases exponentially. 'ES' is a simple-to-use, reliable temporal periodic prediction technique using univariate data [16], [17]. The prediction for the present time (Ft) in Exponential Smoothing is as follows:

$$F_t = \beta A_{t-1} + (1 - \beta)\ F_{t-1} \quad (6)$$

Here $0 \leq \beta \leq 1$, A_{t-1} represents actual value of the preceding period in the time-series, and F_{t-1} represents predicted value of the prior forecast, smoothing cost is used.

4. Support Vector Machine: SVM is a supervised ML technique that can be used to solve classification and regression problems (mainly). The value of each feature represents the value of a given position in the SVM algorithm, and each data item is displayed as a point in n-dimensional geometry (n being the quantity of attributes). Afterwards we locate the hyper-plane that best separates the 2 groups to complete classification. The coordinates of a single observation are what vectors were. This SVM classification algorithm is a frontier which effectively separates this 2 classes (hyper-line or plane).When we have a large data set, it does not perform well because the needed training time is longer.

C. Evaluation Parameters

Time-series prediction models can be assessed using the following commonly used accuracy measurement functions:

- MSE: Mean Square Error and the lower the MSE the better is the performance.

$$\frac{1}{n} \sum_{i=1}^{n} (yi - yi\text{^})\ 2 \quad (7)$$

- RMSE: square root version of MSE

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - yi\hat{}) 2} \quad (8)$$

- MAE: Mean Absolute Error is the difference between predicted and original value

$$\frac{1}{n}\sum_{i=1}^{n}|yi - yi\hat{}| \quad (9)$$

Here y_i and $y_i\hat{}$ represent the real and estimated values, respectively.

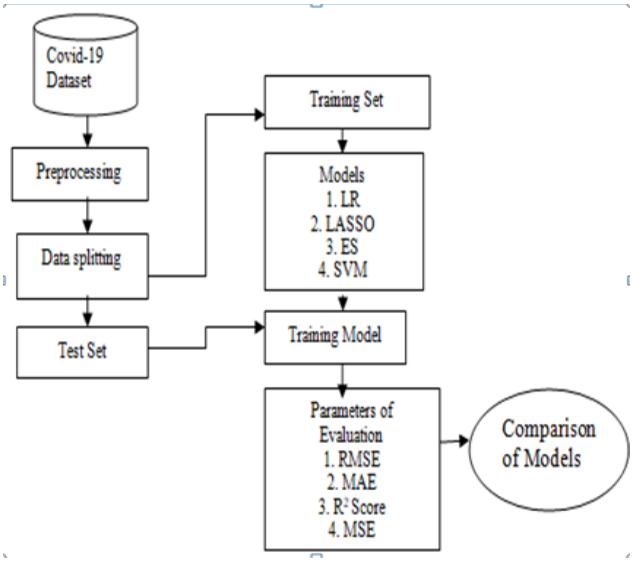
- R2Score: The degree of performance of a regression model is represented by R-squared, a statistical measure. For r-square, 1 is the optimal value. The nearer the r - squared value is to 1, the higher the model fits.

$$\frac{\text{Varianceexplainedbymodel}}{\text{Totalvariance}} \quad (10)$$

III. METHODOLOGY

The dataset was separated into two subsets following the preliminary data pre-processing: a training set for training the models *(85%)* and a testing set for testing the models *(15%)*.

Figure 1 Proposed Workflow



In this work, learning models such as LR, ES, LASSO, and SVM were applied. The new confirmed patients, recovery cases, and death rates were used to train the models [18]. The learning models were then assessed using essential metrics such the RMSE, MAE, R2 score, and MSE, and the findings were published. Figure 1 depicts detailed workflow.

Table III-1 Sample Dataset

Province/ State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20
Queensland	Australia	-27.46	153.02	0	0	0
South Australia	Australia	-34.92	138.60	0	0	0
Tasmania	Australia	-42.88	147.32	0	0	0
Victoria	Australia	-37.81	144.96	0	0	0
Western Australia	Australia	-31.95	115.86	0	0	0

Attributes in Data Source contains State, Country, Latitude, Longitude and contains series of dates. Dataset having the 3 folders (Recovery, death and Confirmed Cases)

IV. *RESULTS & DISCUSSION*

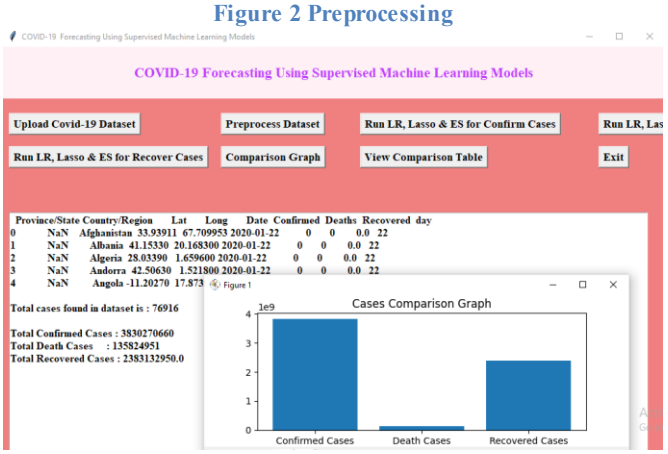
Using ML approaches, this work seeks to construct effective technique towards predicting the future amount of persons infected by COVID (2019). The study's dataset provides day updates on the amount of newly infected patients, the rate of recoveries, and the rate of COVID-19-related mortality around the world. The world is in an unpleasant scenario as the death count and reported illnesses keep on rising. The amount of people that could be infected with COVID (2019) in various parts of the globe is unknown. For the following 10 days, the goal of this study is to estimate the number of people who will be affected in addition to newly infected cases and fatalities, as well as the number of expected recovery rates has all been predicted using 4 ML models: LR, ES, LASSO, , and SVM.

Among all 4 algorithms Exponential Smoothing is giving best performance result and here SVM is taking much longer time for execution, so implementing 3 algorithms. In the dataset we have 3 files for death, recover and confirm cases.

A. *Dataset (From January 2020 to November 2020)*

The first predicted timeframe (from 22/01/20 to 03/11/20) is used to display the results. Click on 'Upload Covid-19 Dataset' button and upload entire dataset folder which contains 3 files such as confirm, recover and death. Select 'dataset' folder and then click on 'Select Folder' to load dataset then dataset is loaded and now click on 'Preprocess Dataset' button to read data and then replace missing values with 0 and clean the dataset.

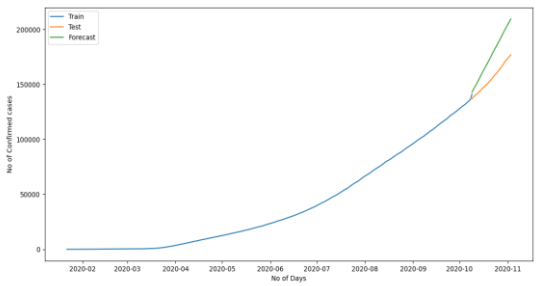
Figure 2 Preprocessing



In the screen we can see dataset is cleaned and here displaying few records from dataset and then displaying total death, recover and confirm cases and case comparison graph. Total cases found in this dataset are 76916 and total confirm

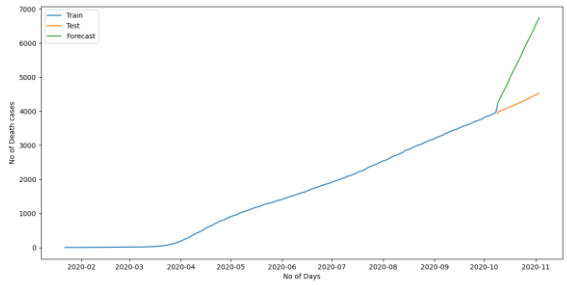
cases, Death cases and recovery cases are shown. First 5 records of the dataset are displayed in above screen.

Figure 3 Confirm cases forecasting using ES



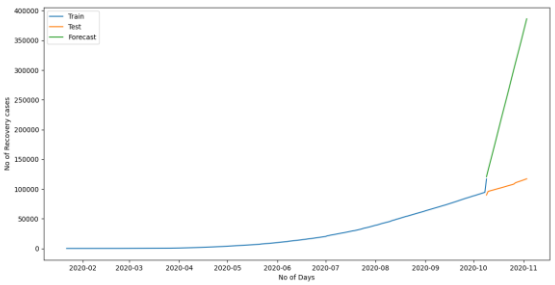
Click 'Run LR, LASSO & ES' for confirmed cases button to get above graph. Graph displays forecasting where blue line refers to train data and yellow line refers to test data and this both train and test data we got from dataset and green line refers to forecast/predicted values for next 10 days. In above graph x-axis represents days and y-axis represents number of confirmed cases.

Figure 4 Death cases forecasting using ES



Click 'Run LR, LASSO & ES' for death cases button to get above graph. In above graph *x-axis* represents days and *y-axis* represents number of death cases.

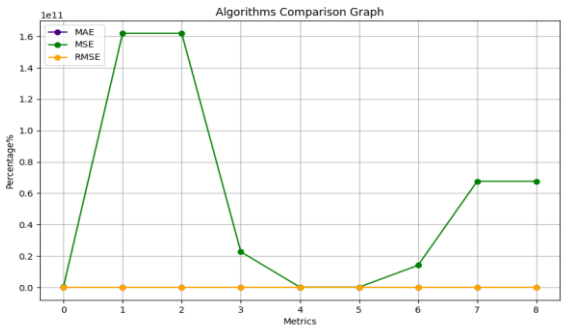
Figure 5 Recovery Cases forecasting using ES



Click 'Run LR, LASSO & ES' for Recovery cases button to get above graph. In above graph *x-axis* represents days and *y-axis* represents number of recovery cases.

Exponential Smoothing gives good performance among LR & Lasso. In above graph, green line represents recover forecast values. Now click on 'Comparison Graph' button to get below graph.

Figure 6 Comparison Graph



In above graph MAE, MSE is more for Linear Regression and Lasso but it got reduce for exponential smoothing. Now close above graph and then click on 'View Comparison Table' button to get below table.

Table IV-1 Comparison Table

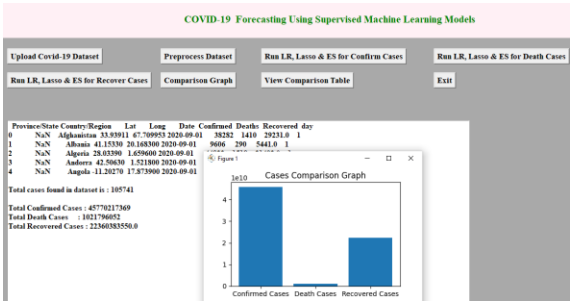
Algorithm Name	R2 Score	MSE	MAE	RMSE
Linear Regression	1.299	67649264400	54018	260094
Lasso	1.299	67649263727	54018	260095
Exponential Smoothing	95.61	14081501933	99360	118665

In above table, We can see MAE, MSE, RMSE and R2Sqaure values for each algorithm and R2square value of Exponential has got highest values compare to other 2 algorithms.

4.2 Updated Dataset (From September 2020 to September 2021)

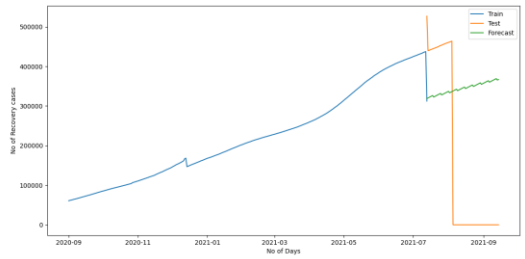
The Second predicted timeframe (from 01/09/20 to 14/09/21) is used to display the results in updated dataset.

Figure 7 Preprocessing for Updated Dataset



Total cases found in this dataset are 105741 and total confirm cases, Death cases and recovery cases are shown. First 5 records of the dataset are displayed in above screen.

Figure 8 Recovery Cases forecasting using ES for Updated Dataset



Click 'Run LR, LASSO & ES' for death cases button to get above graph. In above graph *x-axis* represents days and *y-axis* represents number of recovery cases.

Table IV-2 Comparison Table for updated Dataset

Algorithm Name	R2 Score	MSE	MAE	RMSE
Linear Regression	0.491	1451919935971	334613	1204956
Lasso	0.491	1451919930574	334613	1204956
Exponential Smoothing	81.16	160880475284	399641	401099

For the updated Dataset, Also R2square value of Exponential has got highest values compare to other 2 algorithms.

Our research and projections were based on the belief that the data was correct. Here, R2 Score of ES for the updated dataset got less due to probable data flaws and inadequate of confirmed instances. As in previous Dataset, cases are increased exponentially so R2 Score of ES got 95.6%.

V. CONCLUSION

In this paper, Machine Learning based forecasting approach that assessing the risk of a universal COVID (2019) outbreak is suggested. This approach evaluates a data source including day wise real historical data and generates predictions for the coming days using ML techniques. The study's results show that ES performed the best in the existing prediction field, given the kind and size of the data Source. To a certain degree, LASSO and LR are also good at predicting number of deaths and confirming patients. The results of these two models predict that death rates in the upcoming period, prices will climb, although healing levels will stall. SVM produces unsatisfying outcomes in all circumstances due to the volatility in the dataset values. Creating an exact hyper-plane between the dataset's given values proved difficult. Generally, we find that model projections based on existing conditions are correct, and that they will prove informative in predicting future events. The study's forecasts can thus be extremely useful in assisting authorities in taking appropriate actions and making decisions in order to contain the COVID-19 catastrophe. Many of the papers included in this report were preprints, which indicate they were not subjected to formal assessment. Given COVID-19's rapid worldwide spreading, a rigorous comparative survey is critically needed for humanity.

This research will be improved over time; next, we want to investigate for prediction, use the most accurate and appropriate machine learning approaches while using the updated dataset. Online prediction will be a major focus of our future efforts. For predicting, use the most accurate and appropriate machine learning approaches while using the updated dataset. Online prediction will be a major focus of our future efforts.

Conflict of interest: We rely on the ideas and perspectives of several authors for this study. This study provides a thorough review of the literature. As a result, there is no potential for a conflict of interest.

References

[1] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, 2018 "Statistical and machine learning forecasting methods: Concerns and ways forward," PloS one, vol. 13, no. 3,

[2] WHO. Naming the coronavirus disease (covid-19) and the virus that causes it. [Online]. Available:https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-guidance/naming-the-coronavirus-disease- (covid-2019)-and-the-virus-that-causes-it

[3] WHO Director-General's opening remarks at G20 Health Ministers Meeting - 5 September 2021 Rome, Italy [online]https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-g20-health-ministers-meeting---5-september-2021

[4] 7 Vaccines Approved for Use by WHO [online]https://covid19.trackvaccines.org/agency/who/

[5] Fong SJ, Li G, Dey N, Crespo RG Herrera-Viedma E. (2020) "Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak". arXiv preprint arXiv:2003.10776.

[6] Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E. (2020); "Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction". Appl Soft Comput. 106282.

[7] Batista M. 2020.03.11, "Estimation of the final size of the second phase of the coronavirus COVID-19 epidemic by the logistic model".20024901; doi: https://doi.org/10.1101/2020.03.11.20024901

[8] Hu Z, Ge Q, Li S, Jin L, Xiong M. (2020), "Evaluating the effect of public health intervention on the global-wide spread trajectory of Covid-19". medRxiv..

[9] Jia L, Li K, Jiang Y, Guo X. (2020), "Prediction and analysis of coronavirus disease 2019". arXiv preprint https://arXiv:2003.05447.

[10] Kumar J, Hembram KPSS, (2020), "Epidemiological study of novel coronavirus (COVID-19)". arXiv preprint https://arXiv:2003.11376

[11] DeCaprio D, Gartner J, Burgess T, Kothari S, Sayed S. "Building a COVID-19 vulnerability index". arXiv preprint https://arXiv:2003.07347.

[12] J. H. U. data repository. Cssegisanddata. [Online]. Available: https://github.com/CSSEGISandData

[13] Marco Cascella; Michael Rajnik; Abdul Aleem; Scott C. Dulebohn; Raffaela Di Napoli. 2020, "Features, Evaluation, and Treatment of Coronavirus" (COVID-19)https://www.ncbi.nlm.nih.gov/books/NBK554776/

[14] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, Eds. Boston: Syngress, "Chapter 1 - analytics defined," in Information Security Analytics, pp. 1 – 12.[Online]Available: http://www.sciencedirect.com/science/article/pii/B9780128002070000010

[15] R. Tibshirani, 2015 "Regression shrinkage and selection via the lasso" ,1996 Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288.

[16] F. Petropoulos and S. Makridakis, 2020, "Forecasting the novel coronavirus covid-19", Plos one, vol. 15, no. 3, p. e0231236.

[17] E. Cadenas, O. A. Jaramillo, and W. Rivera, 2010 "Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method" , Renewable Energy, vol. 35, no. 5, pp. 925–930.

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

[19] Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA. 2021, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset." SN Comput Sci.;2(1):11. doi: 10.1007/s42979-020-00394-7. Epub 2020 Nov 27. PMID: 33263111; PMCID: PMC7694891.

[20] Gitanjali R. Shinde, Asmita B. Kalamkar, Parikshit N. Mahalle, Nilanjan Dey, Jyotismita Chaki & Aboul Ella Hassanien , (2020): "Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art." SN computer science vol. 1, 4 197. doi:10.1007/s42979-020-00209-9

[21] Ezugwu, Absalom & Abaker, Ibrahim & Oyelade, Olaide & Chiroma, Haruna & Al-Garadi, Mohammed & Abdullahi, Idris & Otegbeye, Olumuyiwa & Shukla, Amit & Almutari, Mubarak, (2020) "A Novel Smart City Based Framework on Perspectives for application of Machine Learning in combatting COVID-19",: doi: 10.1101/2020.05.18.20105577.

[22] Zlatan Car, Sandi Baressi Šegota, Nikola Anđelić, Ivan Lorencin, Vedran Mrzljak, 2020, "Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron", Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 5714714, 10 pages. https://doi.org/10.1155/2020/5714714

[23] Chen, Joy Iong-Zong "*Design of Accurate Classification of COVID-19 Disease in X- Ray Images Using Deep Learning Approach*" Journal of ISMAC 3, no. 02 (2021): 132- 148.

[24] Murtadha D. Hssayeni, Arjuna Chala, Roger Dev, Lili Xu, Jesse Shaw, Borko Furht & Behnaz Ghoraani "*The forecast of COVID-19 spread risk at the country level*" Journal of Big Data 8, no. 99 (2021).