International Journal of Innovative Research in Engineering
Volume 3, Issue 3 (May-June 2022), PP: 411-416.
www.theijire.com
ISSN No: 2582-8746

Covid-19 Future Forecasting Using Supervised Machine Learning

Dheepin VM[1], Naveenkumar M[2], Sangameshwar S[3]
[1,2,3]ISE, College/ Kumaraguru college of technology,Tamilnadu, India.

How to cite this paper:
Dheepin VM[1], Naveenkumar M[2], Sangameshwar S[3], "Covid-19 Future Forecasting Using Supervised Machine Learning", IJIRE-V3I03-411-416.

Abstract: *Machine-learning (ML) algorithms have demonstrated their use in predicting perioperative outcomes in order to improve decision-making in future activities. ML models have long been used to identify and prioritise negative threat indicators in a variety of operational sectors. To deal with forecasting issues, a few predictive methods are extensively utilised. The power of the ML model to estimate the number of following patients who will be afflicted by COVID-19, which is currently regarded a severe threat to humanity, is demonstrated in this work. This study used four conventional prediction models to detect risk factors: linear regression (LR), total reduction and operator selection (LASSO), vector support (SVM), and exponential fluctuation (ES). COVID-19 projections have been completed. Each model was given three types of predictions: the number of new viral cases, the number of fatalities, and the number of patients who would be treated in the next 10 days. The outcomes of the study lead to a promising application of these strategies in the current COVID-19 outbreak. The results show that among all the most successful tools for forecasting new confirmed cases, death rate, and recovery rate, ES performs best, followed by LR and LASSO, while SVM incorporates available data sets and performs in all predicted scenarios.*

Key Word: *Machine Learning; Linear regression; LASSO; Support vector machine; Exponential Smoothing.*

I.INTRODUCTION

Over the last decade, machine learning (ML) has proven to be a popular subject of study by tackling many of the world's most complicated problems. Healthcare, autonomous vehicles (AV), business applications, natural language processing (NLP), intelligent robotics, games, weather modelling, voice, and image processing are just a few of the application fields. Learning ML algorithms is typically based on a manner of trial and error that differs greatly from common principles, such as following editing instructions based on decision statements as if they were not. The use of several common ML algorithms in this field to lead the future course of action required in many areas of the app, including weather forecasting, disease forecasting, stock market prediction, and disease forecasting, is one of the most important areas of ML forecasting. The neural network and various retrospective models have extensive functionality in predicting future patient states with a certain disease. Many studies are being undertaken to use machine learning techniques to forecast various diseases such as coronary artery disease, heart disease, and breast cancer speculation. The research [8] focused on COVID-19 outbreaks and early response, and the study concentrated on live prediction of COVID-19 certified cases. These predictive algorithms can be quite useful in making decisions about how to deal with the current condition and guiding early intervention to better control these diseases.

The goal of this work is to develop a preliminary prediction model for the transmission of the novel coronavirus, also known as SARS-CoV-2 and designated by the World Health Organization as COVID-19 (WHO). COVID-19 is currently the most serious hazard to human health on the planet. The virus was originally discovered in late 2019 in Wuhan, China, where numerous people got pneumonia-like symptoms. It has a number of impacts on the human body, including acute respiratory syndrome and multiple organ failure, all of which can result in death in a short period of time. Hundreds of thousands of people are infected worldwide every day, resulting in thousands of fatalities. Every day, tens of thousands of young individuals from all over the world are reported to be optimistic. Close physical contact, breathing droplets, or touching a contaminated place are the most common ways for the virus to spread. The fact that a person might be infected for days without showing any symptoms is the most dangerous aspect in its spread. Because of the causes of its spread and the dangers it poses, nearly every country has declared partial or complete closures in all impacted states and towns. Vaccines and medicines for the disease are now being developed by medical researchers all around the world. Because there are no licenced medications to kill the virus at this time, governments around the world are focused on ways to prevent the virus from spreading. "Information" on all aspects of COVID-19 is regarded particularly crucial in all safety measures.

Many researchers are exploring numerous epidemics and creating results to aid mankind in order to contribute to this area of knowledge. Our goal in this study is to develop the COVID-19 forecast system, which will help to alleviate the current humanitarian crisis. Over the next 10 days, three major illness developments will determine predictability: 1) A significant number of new cases have been confirmed. 2) The number of deaths 3) the total number of inmates. The study is based on various high-quality ML-controlled models such as linear regression (LR), total reduction and operator selection (LASSO), support vector machine (SVM), and exponential smoothing because this predictive challenge was regarded a retrospective problem in this study (ES). The COVID-19 patient statistics information given by Johns Hopkins is used to train study models. The database has been pre-processed and separated into two sub-sets: the training set (85% of entries) and the test set (the remaining 15% of records) (15 percent records). R points (R2 points), R2adjusted, RSE, MSE, and MAE were all used to evaluate performance (RMSE).
The following are the study's primary conclusions:
• When a timeline data collection has extremely minimal input, ES works well.
• Different machine learning methods appear to do better in certain class estimates.
• Most machine learning algorithms require a substantial amount of data to forecast the future, and model performance improves as data size grows.
• For decision-makers trying to contain epidemics like COVID-19, an ML model-based forecast can be quite useful.

II.RESEARCH BACKGROUND

S. Makridakis created a project titled Statistical and machine learning forecasting methods: Concerns and Future Directions in 2018. The study describes the findings, explains why ML models are less accurate than statistical models, and recommends possible next steps. Our study's evidence-based findings highlight the necessity for fair and meaningful methods of evaluating the success of predictable approaches, which may be accomplished through large, open competitions that allow for acceptable comparisons and straightforward conclusions. [1]

G. Bontempi introduced Machine-literacy methodologies for time series forecasting in 2012. The presentation of local learning approaches as an effective tool for dealing with temporal data, as well as the formalisation of one-step forecasting issues as supervised learning tasks, and the role of the forecasting strategy when moving from one-step to multiple-step forecasting are the three aspects of this paper. [2]

Regression models for prognostic prediction: Advantages issues and possible solutions was established by F. E. Harrell in 1985. Multiple regression models can be used to predict the outcome of individuals with a wide range of illnesses. When model hypotheticals are thoroughly examined; steps are taken (e.g., choosing another model or transubstantiating the data) when hypotheticals are violated; and the system of model expression has no effect on overfitting the data, retrogression models can make more accurate predictions than other styles such as position and recursive partitioning. [3]

Use of neural networks in prognosticating the threat of coronary roadway complaint was published by P. Lapuerta in 1995. The networks were built using a technique that allowed literacy from cleaned compliances. Unyoking the data into separate training and testing sets, assessing the neural network strategy's performance on the testing sets, and comparing scores with those obtained from Cox retrogression models generated on the same training data were used to do cross-validation. The network architecture provided a useful method for predicting difficulties in a clinical trial with varying follow-up intervals. [4]

K. M. Anderson created cardiovascular disease risk profiles in 1991. The parametric model utilised was shown to have a number of advantages over ordinary regression models. It can provide predictions for varied durations of time, unlike logistic regression, and probabilities can be represented in a more basic manner than the Cox proportional hazards model. [5]

H. Asri experimented with a new method in 2016. Machine learning algorithms are being used to detect and diagnose breast cancer risk. On the Wisconsin Bone Cancer (original) datasets, a performance comparison of different machine literacy methods Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB), and k Nearest Neighbors (k-NN) is undertaken. The major goal is to evaluate the validity of each algorithm's classification accuracy, precision, sensitivity, and specificity in terms of efficiency and effectiveness. [6]

F. Petropoulos proposed a mechanism for forecasting the novel coronavirus COVID-19 in the year 2020. This document details the timing of a live forecasting experiment with huge ramifications for planning and decision-making, as well as objective forecasts for COVID-19 confirmed instances. [7]

The Epidemiological Characteristics of a 2019 Novel Coronavirus Diseases (Covid-19) Outbreak in China were released in 2020 by the Novel Coronavirus Pneumonia Emergency Response Epidemiology (C. P. E. R. E. Novel) team. The Chinese Infectious Disease Information System was used to collect all COVID-19 cases reported through February 11, 2020. The following were among the analyses: 1) patient characteristics; 2) age distributions and sex ratios; 3) case fatality and mortality rates computation; 4) geo-temporal study of viral propagation; 5) epidemiological curve design; and 6) subgroup analysis. [8]

L. van der Hoek proposed the discovery of a novel human Coronavirus in 2004. Lethal coronavirus 229E (HCoV-229E), HCoV-OC43, and severe acute respiratory pattern (SARS)-associated coronavirus are the three known mortal coronaviruses (SARS-CoV). They present the finding of a fourth human coronavirus, HCoV-NL63, utilising a novel virus discovery method. A 7-month-old child with bronchiolitis and conjunctivitis was found to have the virus. The entire genome sequence suggests that this virus is a new group 1 coronavirus, not a recombinant. [9]

Prediction of breast cancer and lymph node metastatic status with tumour markers using logistic regression models was

created by H.-L. Hwa in 2008. Early detection of bone cancer can reduce complaint mortality. The researchers wanted to see how effective serum indicators were at detecting primary breast cancer and lymph node metastatic status. [10]

Retrogression loss and selection via the lariat were first introduced by R. Tibshirani in 1996. They present a new method for linear model estimation in this paper. The'lasso'reduces the residual sum of places if the total absolute value of the sections is less than a constant. As a result of the nature of this restriction, it tends to yield some portions that are exactly 0 and so produces interpretable models. According to their simulations, the lasso has some of the advantages of both subset selection and ridge regression. [11]

A. E. Hoerl and R. W. Kennard presented ridge regression as a biassed approximation for nonorthogonal issues in 1970. If the prediction vectors are not orthogonal, parameter estimates based on lowest residual sum of squares have a significant risk of being unsatisfactory, if not inaccurate, in multiple regression. An estimating approach based on adding modest positive amounts to the slant of X ′X is proposed. The crest trace, a system for displaying the benefits of nonorthogonality in two confines, is introduced. It's also demonstrated how to use X ′ X to get biassed estimations with smaller mean square error. [12]

In 2013, X. F. Du developed a support vector machine-based demand forecasting system for perishable farm products. On the basis of day absolute error, relative mean error, and FP, numerical studies show that forecasting systems based on SVMs and fuzzy theory outperform the radial basis function neural network. The variational range of free parameters and the effects of the parameters on prediction performance are examined in this article because there is no organised technique to choose the free parameters of SVMs. The use of SVMs forecasting system in perishable agricultural product demand forecasting is advantageous, according to the findings of the experiments. [13]

In 2010, E. Cadenas made changes to his paper "Analysis and forecasting of wind velocity in chetumal quintana roo using the single exponential smoothing method." In the first section of the paper, conventional and robust measures were used to conduct a statistical analysis of the time series. The single exponential smoothing method was also used to forecast the last day of observations (SES). For a value of 0.9, the findings showed that this technique provided very good data accuracy. Finally, the SES approach was compared to the artificial neural network (ANN) method, with the former outperforming the latter. [14]

In 2016, J.-H. Han published Manufacturing Data Consideration for Machine Learning Methods in Predictive Manufacturing. The real attempts to deploy smart factories have increased in response to recent developments in the internet of things and big data. To create a smart factory, you must first adopt a predictive manufacturing system. To use machine learning methods, we must first analyse the features of the data and then select the best appropriate approach based on those characteristics. As a result, this paper examines the characteristics of manufacturing data and contrasts various applications of machine learning techniques. [15]

R. Kaundal, A.S. Kapoor, and G. P. Raghava published a case study on rice blast prediction using Machine Learning techniques in disease forecasting in 2006. As predictor factors, six significant weather variables were chosen. A five-fold cross validation approach was used to generate and validate two series of models (cross-location and cross-year). Overall, our SVM-based prediction approach will open new doors in the field of plant disease forecasting for a variety of crops. [16]

C. Willmott and K. Matsuura discovered in 2005 that the mean absolute error (MAE) has advantages over the root mean square error (RMSE) in evaluating average model performance. The root-mean-square error (RMSE) and the mean absolute error (MAE) are compared in terms of their ability to reflect average model-performance error. The data suggest that MAE is a more natural measure of average error, and that it is unambiguous (unlike RMSE). As a result, MAE should be used in dimensionalized assessments and inter-comparisons of average model-performance error. [17]

The gamma distribution-based EMOS model for probabilistic quantitative precipitation forecasting was censored and shifted by S. Baran and D. Nemoda in 2016. All major weather prediction centres now offer forecast ensembles of diverse weather quantities derived from several runs of numerical weather prediction models with varying initial circumstances and model parametrizations. According to the findings, the suggested CSG EMOS model surpasses the GEV EMOS approach in terms of probabilistic and point forecast calibration, and outperforms the raw ensemble and the BMA model in terms of predictive skill. [18]

In 2020, Y. Grushka-Cockayne and V. R. R. Jose will compete in the m4 competition by combining prediction intervals. In addition to point estimates, the 2018 M4 Forecasting Competition was the first M Competition to elicit prediction intervals. The median and the interior trimmed average are found to be robust aggregators for the prediction interval submissions throughout all 1,00,000 times series, while averaging interval endpoints retains its practical appeal as being simple to implement and performing well when data sets are huge. [19]

Critical care utilisation was invented by G. Grasselli, A. Pesenti, and M. Cecconi in 2020 for the COVID-19 outbreak in lombardy, Italy: During an emergency reaction, early experience and forecast are crucial. Based on this experience, it appears that only an ICU network can deliver the first surge reaction necessary to ensure that every patient in need of an ICU bed receives one. If your health-care system isn't already part of a coordinated emergency network, you should start now. [20]

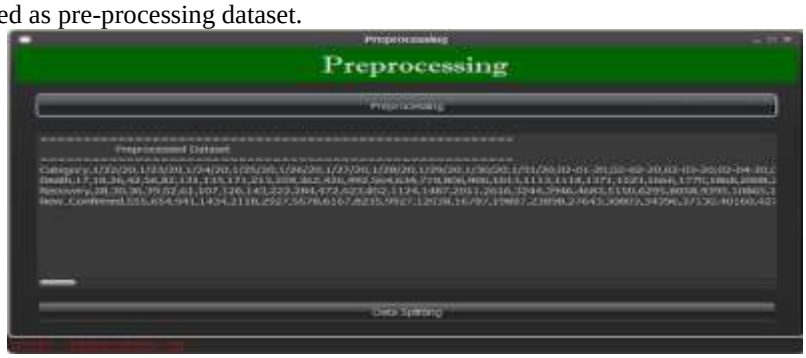
III.PROPOSED WORK

Due to the automatic extraction of relevant features from the training samples, feeding the activation from the previous time step as input for the current time step, and network self-connections, machine learning approaches showed to be beneficial for prediction. In our method, we propose LR, LASSO, SVM, and ES and compare these algorithms with the predicting of death

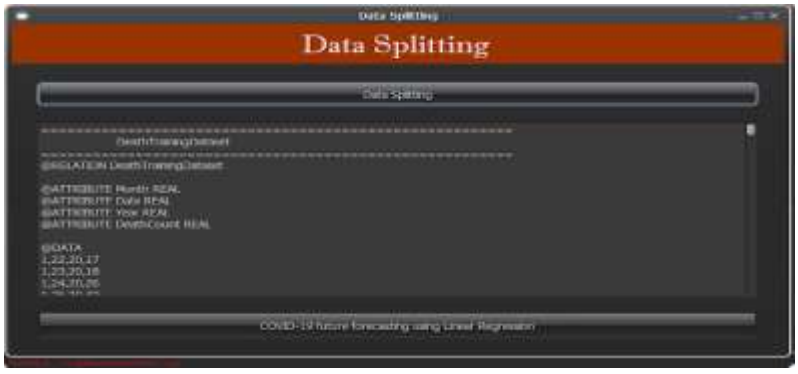
rate recovery rate and new confirmed case prior to 10 days, based on the findings of the model analysis.

IV.RESULTANT CONCLUSION

The daily time series summary tables in the data folder include the number of confirmed cases, deaths, and recoveries. All data comes from the daily case report, and data is updated once a day. The dataset loaded into the home page is split into different strings that has death rate, recovery and confirmed cases. Each string is split into corresponding string arrays after splitting into each line. After that, each array will be split into individual words, trimmed for null values. The same is for all string arrays. This is saved as pre-processing dataset.



This step involves splitting dataset into training and testing datasets. The training set is given 70% of input dataset and remaining 30% is testing dataset.

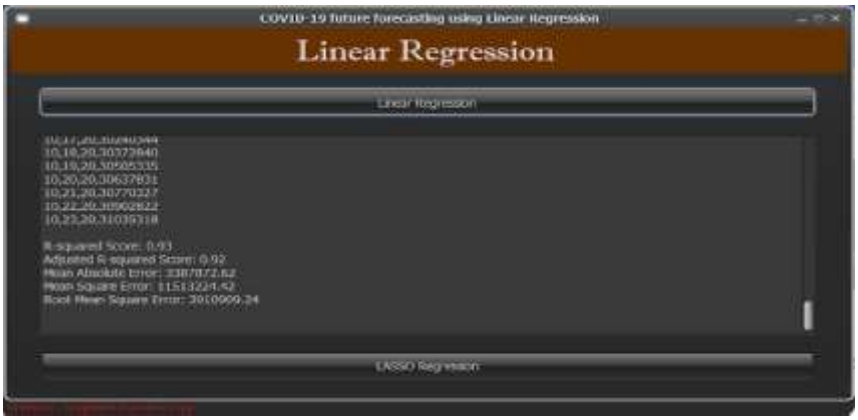


We take the training dataset and use it to train the Linear / Lasso model and we will be evaluating the model by calculating R-Squared Score, Adjusted R-squared Score, Mean Absolute Error, Mean Square Error, Root Mean Square Error. We follow this process for death, recoveries and confirmed cases. The model thus trained after removing the error is tested for accuracy using the test dataset. The cross-validation accuracies are displayed in the application using the following metrics.

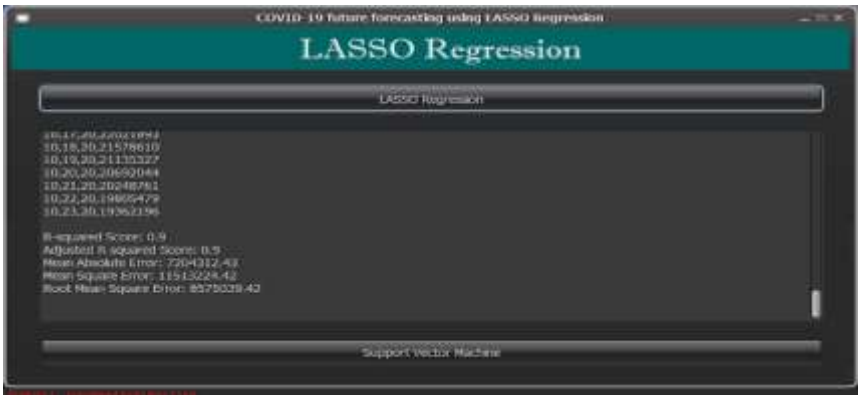
- R-Squared Score,
- Adjusted R-squared Score,
- Mean Absolute Error,
- Mean Square Error,
- Root Mean Square Error.

We have attached the resultant shows in the below images in algorithm wise.

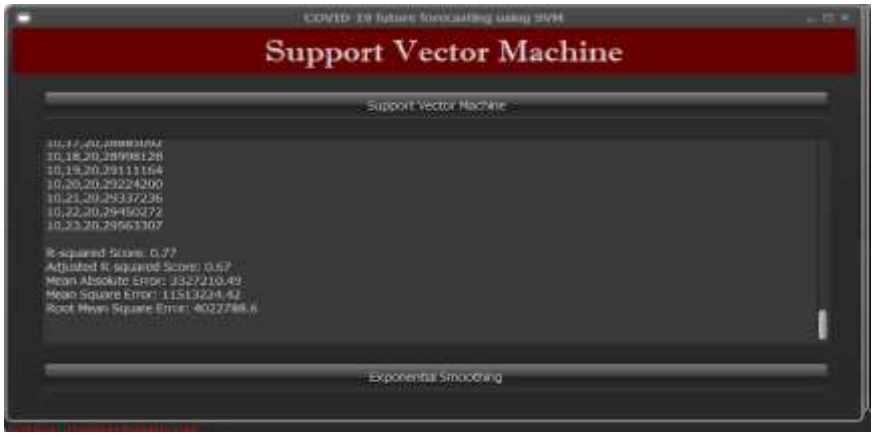
LR FRAME:



LASSO FRAME:



SVM FRAME:



V.CONCLUSION

The COVID-19 pandemic's precariousness has the potential to spark a catastrophic worldwide calamity. Some scientists and government agencies around the world are concerned that the epidemic will infect a big section of the global population. In this paper, an ML-based prediction system for estimating the worldwide risk of COVID-19 breakout is suggested. The system uses machine learning algorithms to analyse a dataset including day-by-day actual past data and create predictions for the following days. Given the type and amount of the dataset, the study's findings show that ES performs best in the current forecasting domain. LR and LASSO are also good at projecting mortality rates and confirming cases to some extent. The results of these two models predict that death rates will rise in the following days, while recovery rates will slow. Because of the ups and downs in the dataset values, SVM delivers unsatisfactory results in all cases. It was quite tough to create an accurate hyperplane between the dataset's given values. We conclude that model predictions based on the current scenario are valid, which may be useful in predicting future events. The study's predictions can thus be immensely beneficial in supporting authorities in taking the necessary procedures and making decisions to contain the COVID-19 disaster. This research will be improved over time; next, we want to investigate prediction methodology utilising the updated dataset and employ the most accurate and relevant machine learning methods for forecasting. One of the key focuses of our future work will be real-time live forecasting.

VI.ACKNOWLEDGMENT

We cordially thank our institution for their support and our lecturers for their constant insights and guidance for the survey.

References

[1] *S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and Future Directions," PLoS ONE, vol. 13, no. 3, Mar. 2018.*

[2] *G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine learning techniques for time series forecasting," in Proceedings of the European Business Intelligence Summer School, pp. 62-77, 2012.*

[3] *"Regression models for prognostic prediction: Benefits, Problems, and Suggestions," Cancer Treat. Rep., vol. 69, no. 10, pp. 1071-1077, 1985.*

[4] *P. Lapuerta,S.P. Azen andL. Labree,"Use of neural networks in prognosticating the threat of coronary roadway complaint", Comput. Biomed.Res.,vol. 28,no. 1,pp. 38-52,Feb. 1995.*

[5] *"Cardiovascular-disease risk profiles," Amer. heart J., vol. 121, no. 1, pp. 293-298, 1991.*

[6] *H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Computer Science, vol. 83, no. 1, January 2016, pp. 1064-1069.*

[7] Vaticinating the new coronavirus COVID-19, F. Petropoulos andS. Makridakis, PLoS ONE, vol. 15, no. 3, Mar. 2020.
[8] C.P.E.R.E. Novel,"The epidemiological characteristics of a 2019 new coronavirus complaint (Covid-19) outbreak in China,"Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi,vol. 41,no. 2,pp. 145, 2020.
[9] L. van der Hoek,K. Pyrc,M.F. Jebbink,W. Vermeulen-Oost,R.J. Berkhout,K.C. Wolthers, and others,"Identification of a new mortal Coronavirus,"NatureMed.,vol. 10,no. 4,pp. 368-373, 2004.
[10] H.-L. Hwa,W.-H. Kuo,L.-Y. Chang,M.-Y. Wang,T.-H. Tung,K.-J. Chang, and others," Vaticination of bone cancer and lymph knot metastatic status with tumour labels using logistic retrogression models,"J. Eval. Clin.Pract., vol. 14, no. 2, pp. 275-280, Apr. 2008.
[11] R. Tibshirani," Retrogression loss and selection using the lariat,"J. Roy. Stat. Soc. Ser. BMethodol., vol. 58, no. 1, Jan. 1996, pp. 267-288.
[12] R.W. Kennard andA.E. Hoerl,"Ridge retrogression Poisoned estimate for nonorthogonal problems,"Technometrics,vol. 12,no. 1,pp. 55-67, February 1970.
[13] X.F. Du,S.C.H. Leung,J.L. Zhang, andK.K. Lai," Demand soothsaying of perishable ranch products using support vector machine,"International Journal of System Lores,vol. 44,no. 3,pp. 556-567, March 2013.
[14] E. Cadenas,O.A. Jaramillo, andW. Rivera," Analysis and soothsaying of wind haste in chetumal quintana trees using the single exponential smoothing system,"Renewable Energy,vol. 35,no. 5, May 2010,pp. 925-930.
[15] J.-H. Han and S.-Y. Chi," Considering manufacturing data to use machine literacy styles for prophetic product,"in Proceedings of the 8th International Conference on Ubiquitous Unborn Networks (ICUFN), pp. 109-113, July 2016.
[16] R.Kaundal,A.S. Kapoor,andG.P. Raghava," Machine knowledge ways in complaint auguring A case study on rice blast prophecy,"BMCBioinf.,vol. 7,no. 1,pp. 485, 2006.
[17] "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in quantifying average model performance," ClimateRes., vol. 30, no. 1, pp. 79-82, 2005."
[18] S.Baran and D.Nemoda,"A probabilistic quantitative rush auguring model predicated on a shifted gamma distribution,"Environmetrics, vol. 27,no. 5,pp. 280-292, August 2016.
[19] Y. Grushka-Cockayne andV.R.R. Jose," Combining vaticination intervals in the m4 competition,"Int.J. Soothsaying,vol. 36,no. 1,pp. 178-185,Jan. 2020.
[20] G. Grasselli,A. Pesenti andM. Cecconi," Critical care application for the COVID-19 outbreak in Lombardy Italy Early experience and cast during an exigency response", JAMA,vol. 323,no. 16,pp. 1545,Apr.2020.