International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Using Supervised Machine Learning models, COVID-19 Future Forecasting

[1]DEEPA S R, [2]PRAVEEN K S

[1]Student, Department of Master of Computer Application, East West Institute of Technology, Bangalore, Karnataka, India.

[2]Assoc.Professor Department of Master of Computer Application, East West Institute of Technology, Bangalore, Karnataka, India.

OUTLINE

In order to improve decision-making regarding upcoming courses of action, AI (ML)-based assessment systems have shown their value in predicting perioperative outcomes. It is recognised that certain expectation tactics are frequently employed to address estimating issues. This work shows that ML models can estimate the amount of COVID-19 patients who will be impacted in the term, which is currently thought to be a potential danger to mankind. The underpinning variables of COVID19 were calculated in this research using four conventional prediction models, including Linear Relapsing (LR), Minimum Least Compression on Selection Supervisor (LASSO), Multilayer Perceptron (SVM), and Exceptional Levelling (ES). Three different forecast categories, like B, are produced by each of the algorithms. the amount of new infected patients, admissions, and recovered during the course of the following 10 days. The review's findings indicate that applying these tactics to the COVID19 pandemic's current state is a promising strategy. Results indicate that SVM performs badly in all expectation scenarios based on the information data set, with ES performing best between all models utilised, following both LASSO and LR are two function well here in evaluating new documented cases, clearing rates, and buffering capacity. Superior flattening technique, advance estimation, updated R2 score, and controlled Intelligent systems are recorded in TERMS for COVID19.

I. PREAMBLE

Through the resolution of multiple incredibly exciting and difficult real-world problems, deep learning has established itself as a formidable field of concentration during the past ten years. Practically every aspect of today's reality was covered in terms of application areas, including: B. Healthcare, AVs, business applications, NLP, smart robots, gaming, contextual demonstration, speech and media management. Following programming principles in words of endpoints of choice like ifelse, understanding Machine learning computations is frequently based on an exploration technique that is extremely opposed to ordinary calculations [1]. One of the most key aspects of deep leaning is anticipating [2], where a number of common ML computations have been utilised to power the expected flow of work necessary in many application domains, such as predicting weather, diagnosing diseases, gauging the health of populations, and identifying illnesses. Advertisement. Different models of brain networks and relapses have broad applicability for predicting future states of sufferers with a given disease [3]. Numerous studies have been done on applying AI techniques to anticipate various illnesses, including coronary heart disease [4], cardiometabolic prediction [5], and tumour prognosis [6]. In particular, the studies [7] and [8] concentrate on the in vivo evaluation of verified COVID19 cases as well as the early detection and reaction to the COVID19 epidemic. Decision making on how to handle the actual conditions and directing earlier intervention to manage those illnesses can both benefit greatly from these expectation frames. By developing an early numerical model, this study aims to better understand how the newly discovered Covid, also called as SARSCoV2 and designated COVID19 by the Health Organisation (WHO) [9], will spread. The coronavirus is increasingly posing a serious threat to all human life on Earth. Even by close the virus in 2019 had been discovered in Wuhan, a Chinese city, where several people had experienced complications like pneumonia. It has a variety of impact on the body, such as inter failure and acute severe respiratory disease, both of which can quickly result in death [11]. This pandemic has afflicted millions of individuals worldwide, and thousands of deaths are expected every day. Numerous newcomers from different countries are continually viewed favourably. The three basic ways that an infection spreads are through close contact, respiratory droplets, and touching contaminated surfaces. The fact that this person can carry the virus for a prolonged period without exhibiting any symptoms makes it the hardest to spread. Nearly all nations have implemented light or strong blockades in all impacted cities because to the factors that contributed to its spread and the risk involved. Currently, it is necessary for medical experts all around the world to discover an effective vaccination and treatment for the illness. Legislators in every nation are concentrating on protective measures that can curb the spread of the sickness because there is currently no licenced treatment to treat it. Nearly all of COVID19's safety precautions, aside from "being courteous," are regarded as essential. Numerous researchers are researching the various facets of the outbreak and providing the findings to benefit humanity in order to add to this component of knowledge in. In this study, we try to support an estimating framework for COVID19 to help with the current human catastrophe. The prognosis for the three primary infection-causing factors for the upcoming 10 days is available: 1) The quantity of newly discovered cases. 2) The total death toll three) The quantity of recoveries. The review is dependent on some requirements of the skill ML fallback models, like: Svm Classifier (SVM) and great smoothing, as this decision problem was regarded a backup issue in this evaluation (ES). The Johns Hopkins COVID19 patient dataset was used to develop the learning models. Pre-processing was done on the data set, and it was split into two subgroups: training set (which made up 85% of the large datasets) and test set (15 percent of the data sets). Meaningful metrics such as the R2 score, modified R2 score, mean squared error, mean absolute error, and mean squared error were performed for the exposure assessment (RMSE). The following list of significant findings from this study:

• Whenever the series data dataset has incredibly tiny portions, ES performs well.

• For various class expectations, certain ML calculations appear to perform better.

• Since the model exhibits advance as the data set expands, the majority of ML computations need a decent amount of data to predict the future.

• For executives, proper evaluation on Machine learning can be very helpful in containing pandemics like COVID19.

Six segments make up the remaining text of the paper. The presentation is presented in Section I, and the information set and review procedures are described in Section II. The technique is presented in Section III, the results are shown in Section IV, and the work is summarised and the conclusion is presented in Section V.

Tools AND Technologies

A. Set of data

By concentrating on the amount of new secure incidents, the quantity of fatalities, both the amount recovery rates, this study seeks to assess the future spreading of COVID19. The Johns Hopkins Institute for Computational Science and Engineering's GitHub vault was utilised to store the dataset for the review [12]. The university essentially made the store accessible for the

Coronavirus 2019 graphical display with help from the Living Atlas team at ESRI.

The organizer's dataset files can be found in the Wikipedia file csse covid 19 time series. The organiser has charts with weekly data set summary information, such as the total number of reported cases, fatalities, and recoveries. Every type of knowledge is taken from the report each day, and each time the information is updated, it takes one day. Columns 1, 2, and 3 present, correspondingly, information examples from the files.

Bench 1. Time-series of coronavirus patient cases occurring all around the world.

Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
Northern Territory	Australia	-12.46	130.84	0	0	...	0
Diamond Princess	Canada	0.000	0.000	0	0	...	1
NaN	Algeria	28.03	1.65	0	0	...	19

Bench 2. Worldwide time series of newly confirmed coronavirus cases.

Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
NaN	Afghan	33.00	65.00	0	0	...	74
Victoria	Australia	-37. 81	144. 96	0	0	...	411
NaN	Algeria	28.03	1.65	0	0	...	264

B. Directed

MODELS FOR MACHINE LEARNING

When confronted with an ambiguous information event, a controlled learning model attempts to establish an expectation. As a result, in this learning method, the learning computation uses a collection of input data that includes comparisons of the input cases.

Bench 3. Time Sequence of coronavirus recovery cases worldwide. the relapse model by using a regressor.

Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
Colombia	Canada	49. 28	-123. 1	0	0	...	4
Victoria	Australia	-37. 81	144. 96	0	0	...	70
NaN	Algeria	28.03	1.65	0	0	...	65

The constructed paradigm then produces a forecast for the test data set or unexpected input data [13]. To enhance predictive models, this learning strategy might use fallback techniques and categorization computations. In this COVID19 future estimation study, four fallback models were employed:

- Linear Regression
- LASSO Regression
- Support Vector Machine
- Exponential Smoothing

1) LINEAR REGRESSION

An subjective classifier is based on free reflexes in recurrence visualisation [14]. As a result, this approach can be used to forecast as well as determine the link among two or more variables. The most practical measurable technique for predictive modeling in deep learning is linear regression, a kind of fallback demonstration. Both the variable of interest and the free component serve as the foundation for each direct relapse observation. The relationship between such dependant and free elements is directly determined by a direct relapse. Directly test for recovery involves the interaction of two factors (x, y). The fallback connection between y and x is demonstrated in the following condition. y= 0+1x+1 or something equivalent E(y) = 0 + 1 x (2), where is the linear regression's error term. A To create the models x technique as the data preparatory database and address the class names contained in the information dataset, incorporate the concept of explicit backup into the AI setup. Finding the optimal characteristics for 0 (block) and 1 (coefficient) in order to obtain the best backup line is the objective of the Artificial Intelligence computation at this phase. This reduction problem can be approached as follows: Since the gap between the real characteristics and the predicted value must be as little as possible to achieve the optimal fit: Bound (prediyi) 1 n X I = 1 2 (3) g=1nXi=1 (prediyi) 2 (4) Here, denotes the number of information, and g is referred to as the output capacity. It is calculated by calculating the root mean square of the anticipated y value (consisting) and the calculated value of y (yi).

2) LASSO REGRESSION

is a recurrence model that employs contraction and the direct recurrence approach [15]. The compression in this setting refers to the shrinking of the outrageous focus-related advantages of an information test. As a result, the contraction cycle enhances and stabilises LASSO while simultaneously lowering error [16]. For multicollinearity cases, LASSO is thought to be a better model. This situation is comparable to the size of the coefficients because the model implements the L1 regularisation and the cost it contains. In terms of the quantity of functions you employ, that's how LASSO streamlines the fallback. To automatically punish additional elements, it employs a regularisation technique. In other words, objects that are insufficiently capable of supporting backup possibilities can be set to a lower limit, perhaps zero. Each of the given markers is used in a standard multivariate regression, and a coefficient of determination is given to each. The LASSO regression, in contrast, tries to add them separately. If the new component is not successful in beating the penalty term by adding this include, the significance cannot be included as anything. Because of this, regularization's value lies in its ability to decide for us by using the penalty idea for more highlights. But since loop takes away the variables after their quality are the same as zero in this regularisation situation, the models with few variables become lean. That suggests that the LASSO backup undermines the intention to confine. p X j=1 j n X i=1 (yiX j xijj)2+ (5) sets the coefficient, which may be seen as min(summ), where the slope is the penalty term and the quadratic residuals are +.

3) SUPPORT VECTOR MACHINE

A regulated ML computation known as a (SVM) is employed for both backup and categorization [17], [18]. SVM analysis is a nonparametric method that requires a variety of mathematical abilities. The part skill set modifies the data inputs into the desired structure. SVM uses a direct function to manage fallback issues. Therefore, it transforms the input sequence (x) into a dimensional space known as the feature space when tackling non-direct regression issues (z). This chart is done by non-direct planning methods after applying a direct fallback to the space. Use a bivariate classification model (xn) with N percepts and yn as the set of known responses to implement the concept in a ML environment. f(x)=x0 + b is a representation of the direct capacitance (6) Finding the result of f(x) with (0) as negligible standard characteristics is the next step after making it as flat as feasible. The issue then conforms to the capability of minimization as follows: J()= 1 2 0 (7) with the additional restriction that none of the residual values exceed, as in the subsequent restriction: "n:yn" "(x0 n + b")" (8)

Smoothing Exponentially

With the known methods of exponential smoothing, the forecast is based on information from previous periods. The influence of prior information perceptions rots away dramatically as they mature. In this way, the weighting of different sag qualities is reduced arithmetically. ES is an exceptionally simple, robust time series evaluation strategy, especially for univariate data [7], [19]. Ft = At1+(1)Ft1 is the formula for the duration value (Ft) in ES (9) Here is the flattening cost, where 0 1, At1, and Ft1, respectively, represent the past time frame's actual cost value and calibration value in the time series.

Assessment BOUNDS

In this analysis, we evaluate each learning model's performance using the Rsquared Score (R2), Adjustable Occur in the course (R2adjusted), Mean Square Error (MSE), Mean Absolute Percentage error (MAE), and Root - mean - square Error (MAE).

1) SCORE FOR R-SQUARED (R2)

The score serves as an objective metric for evaluating how relapse patterns are presented [20], [21]. The indicator displays the degree to which the confidence variable, which affects the independent variable overall, fluctuates. It uses a sensible scale of 0 to 100 percent to quantify the degree of correlation here between assurance factor and the recurrence models. After developing the repetition pattern, we can use the R2 score to assess the produced models' goodness of fit. The safety coefficient, also known as the spread of the relevant data around the backup line, is determined by the R2 score. Normally, your score falls between 0 and 100 percent. A score of 0% means the reaction contains no variation around its mean, which supports the model's validity. A score of 100% means the answer has all variation around its mean. The good nature of the created model is evident from the excellent R2 score. The diverse array factor level is explained by R2, a direct model. usually finds as: R2 is the variation that the model can account for overall (10)

ADJUSTED R-SQUARED SCORE

A modified form of R2 called r Value squared (adjusted R2), like R2, indicates how well the data approaches the curve. R2 and modified R2 differ primarily in that the latter measures the amount of peaks in a data characteristics. If the recently introduced features are helpful for the prediction model, the adjusted R2 may increase as a result of the inclusion of additional features. But the additional features will lose value if they don't make sense. Following is a definition of the adjusted R2: R2 adjusted = 1 (R2) n 1 (k + 1) (11) Here, k is the output variable that affect the likelihood of recidivism, and the sample size is indicated.

2) MEAN ABSOLUTE ERROR (MAE)

The average level of error in the model expectations' disposition is shown by the absolute error [22], [23]. All singular differences are given equal weight in this testing data mean between the predicted results and the actual data. Fewer score values demonstrate the integrity of learning techniques, therefore one also speaks about scores that are unfavourably located [24]. Its matrix values vary from 0 to infinity.

MAE =

1 n

n X j=1

|yj− ˆ yj| (12)

3) MEAN SQUARE ERROR (MSE)

Another metric for evaluating the effectiveness of fallback models is the mean square error [22]. MSE decrypts the information from the fallback line. Estimating is important because it takes away the value's negative sign and gives bigger differences more weight. The lowest squared error demonstrates that you are getting closer to identifying the line of greatest fit. MSE = 1 n X i=1 (yi yi) is a formula for calculating the MSE. ROOT MEAN SQUARE ERROR: 2 (13) (RMSE) The forecast error's standard deviation can be used to define the mean square deviation. The distance comparing the best-fit line to the actual data values are referred to as a forecast error or residual. The concentration of real information approached all around line of best fit is thus measured by RMSE. It is the mistake margin brought forth around the square base of the MSE shown below. X i=1 (yi yi)2 R - square =v u t 1 n (14) III. Philosophy This evaluation is of the COVID19 predictions, a new product from Covid. A potential contemporary threat to human survival has been brought to light by COVID19.

It results in a significant number of fatalities, and mortality rates are rising globally every day. This study seeks to predict the passing rate, the number of infections verified each day, and the amount of recoveries during the following 10 days in order to supplement the control of the pandemic scenario.

Four ML moves that were appropriate for this particular case were used to accomplish the determination. Summary tables containing daily time series, show that the number of cases reported, fatalities, and recoveries in the most recent days because the pandemic started, are included in the data set utilised for the evaluation. To establish worldwide estimates of the monthly number of reported cases and recoveries, the data set was first pre-processed for this review.

Bench 4. Test of data from the total number of cases time-series.

Category	Province /State	Country /Region	Lat	Long	1/22/20	...	3/27/20
Death	Victoria	Australia	-12.4	130.84	0	...	0
	Nan	Canada	0.000	0.000	0	...	1
	NaN	Algeria	28.03	1.65	0	...	19
Recovery	Colombia	Canada	49. 28	-123. 1	0	...	4
	Victoria	Australia	-37. 8	144. 96	0	...	70
	NaN	Algeria	28.03	1.65	0	...	65
New	NaN	Afghan	33.00	65.00	0	...	74
Confirmed	Victoria	Australia	-37. 8	144. 96	0	...	411
	NaN	Algeria	28.03	1.65	0	...	264

Bench 5. Information about all deaths on a daily basis.

Day 1 deaths	Day 2 deaths	...	Day 66 deaths
0	4	...	20

Bench 6. Information on the daily recovery rate test.

Day 1 recoveries	Day 2 recoveries	...	Day 66 recoveries
0	6	...	139

Bench 7. Day by day, all new confirmed cases and test data are complete.

Day 1 new cases	Day 2 new cases	...	Day 66 new cases
0	21	...	749

The data set was split into two subsets after the internet and information pre-processing step: a 56-day training period to create the prototypes and a collection of tests (10 days). In this review, educating models like SVM, LR, LASSO, and ES were applied. The designs for recovery and death were created using recent proven examples at the time. The learning models were then evaluated and published in the results after being compared to relevant measurable variables, such as B. R2score, R2adjusted score MSE, RMSE, and MAE. Figure 1 depicts the proposed strategy utilised in the review as a block diagram.

IV. CONCLUSION AND RESULTS

The goal of this project is to establish a framework for estimating the amount of COVID19 instances using AI techniques in the future. Data from regular updates on the number of innovations instances of infection the quantity of recoveries and the total number of COVID19 fatalities worldwide.

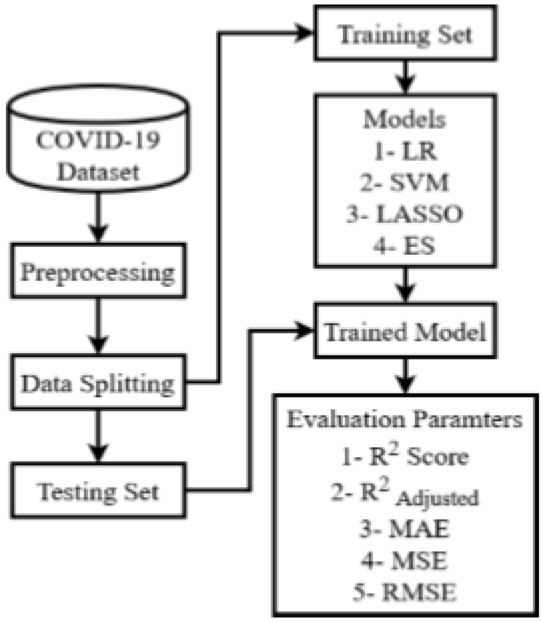

FIGURE 1. Proposed work process.

That is precisely what is occurring globally as the mortality rate and cases reported rise steadily. The number of persons who could be impacted by the COVID19 epidemic in various nations throughout the world is not very high. The goal of this study is to estimate the number of persons who may be affected by additional cases of infection, fatalities, and anticipated recovery over the course of the next 10 days. The number of new tainted cases, the number of passages, and the number of recoveries were all predicted using four AI models: LR, LASSO, SVM, and ES.

A. Future Mortality Rate Forecasting

The rework meets the mortality rate assumptions, and the results show that ES outperforms all other models. Similar performance and results are

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

obtained by LR and LASSO. R2 rating. Svm works worse in every circumstance, in correlation. Table 8 displays the outcomes.

Model	R^2 Score	$R^2_{Adjusted}$	MSE	MAE	RMSE
LR	0.96	0.95	840240.11	723.11	916.64
LASSO	0.85	0.81	3244066.79	1430.29	1801.12
SVM	0.53	0.39	16016210.98	3129.74	4002.02
ES	0.98	0.97	662228.72	406.08	813.77

Bench 8. models the execution of future death rate determination.

The performance of the LR, LASSO, SVM, and ES models are displayed separately in charts in Figures 2, 3, 4, and 5. The charts of all data indicate a growth in the approval rating over the next few days, which is a very concerning development. The ongoing death rate depicted on the graph in Figure 14 accurately corresponds to what the models predicted.

B. FORECASTING OF NEW Tainted Confirmed Future Cases

The number of COVID19 confirmed cases is rising daily. The prediction outcomes of the models employed in this investigation are displayed in Table 9. Performance-wise, ES and LASSO are at the top of the list, followed by LR and SVM.

Bench 9. Performance of the models for predicting new confirmed cases of infection in the future.

Interval	Dataset Size (Number of Days)	Dates (From 22 Jan 2020) To	LASSO Performance	LR Performance	SVM Performance	ES Performance
1.	26	16 Feb 2020	Very poor	Very poor	Very poor	Best
2.	41	2 Mar 2020	Very poor	Very poor	Very poor	Best
3.	56	17 Mar 2020	Poor	Good	Very poor	Best
4.	66	27 Mar 2020	Better	Best	Well improved	Best

quite poor across the board. Figures 6, 7, 8, and 9's graphs display the predictions made by the learning models.

C. PROJECTIONS FOR THE FUTURE Survival Rate

In the future projection of the recovery rate, the ES once more outperforms all other models. By nature, ES performs best, followed by LR, LASSO, and SVM in that order, with all other models doing poorly.

V. SUMMARY

The COVID19 outbreak vulnerability could lead to a severe worldwide calamity. The outbreak might potentially affect a significant section of the global population, according to certain researchers and government organisations [26, 27]. An ML-based forecasting system put forth in this study to forecast the likelihood of a COVID19 outbreak on a global scale. In order to estimate what will happen over the next several days, the system analyzes data set with the actual daily data from the past. Given the type and breadth of the data set, the study's findings demonstrate that ES performs better in the present forecast range. In some ways, LR and LASSO are also useful for projecting mortality and confirming cases. The outcomes of these two approaches predict that death rates will rise and returns will fall during the next few days. Due to the fluctuations in dataset values, SVM performs badly in every circumstance. Placing a precise hyperplane beneath the dataset's values proved to be exceedingly challenging. In general, we draw the conclusion that the model's projections for the present situation are accurate, which can be useful for interpreting the future situation. For the administration to take prompt action to make choices to contain the COVID19 issue, the study's predictions can also be of tremendous use. We intend to investigate the prediction approach utilising the revised data set and employ the most precise and relevant ML methods for forecasting as this work advances. One of the key areas of our future work will be real-time live forecasts.

BIBLIOGRAPHY

[1] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, ‘‘Statistical and machinelearningforecastingmethods:Concernsand waysforward,’’PLoS ONE, vol. 13, no. 3, Mar. 2018, Art. no. e0194889.

[2] G.Bontempi,S.B.Taieb,andY.-A.LeBorgne,‘‘Machinelearningstrategiesfortimese riesforecasting,’’inProc.Eur. Bus.Intell.Summer School. Berlin, Germany: Springer, 2012, pp. 62–77.

[3] F. E. Harrell Jr, K. L. Lee, D. B. Matchar, and T. A. Reichert, ‘‘Regression models for prognostic prediction: Advantages, problems, and suggested solutions,’’ Cancer Treat. Rep., vol. 69, no. 10, pp. 1071–1077, 1985.

[4] P.Lapuerta,S.P.Azen,andL.Labree,‘‘Useofneuralne tworksinpredicting the risk of coronary artery disease,’’ Comput. Biomed. Res., vol. 28, no. 1, pp. 38–52, Feb. 1995.

[5] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel, ‘‘Cardiovasculardiseaseriskprofiles,’’Amer.heartJ., 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 Comput. Sci., vol. 83, pp. 1064–1069, Jan. 2016.

[7] F. Petropoulos and S. Makridakis, ‘‘Forecasting the novel coronavirus COVID-19,’’ PLoS ONE, vol. 15, no. 3, Mar. 2020, Art. no. e0231236.

[8] G. Grasselli, A. Pesenti, and M. Cecconi, ‘‘Critical care utilization for the COVID-19outbreakinlombardy,italy:Earlyexperienceandfo recastduring an emergency response,’’ JAMA, vol. 323, no. 16, p. 1545, Apr. 2020.

[9] WHO. Naming the Coronavirus Disease (Covid-19) and the Virus That Causes it. Accessed: Apr. 1, 2020. [Online]. Available: https:// www.who.int/emergencies/diseases/novel-coronavirus-2019/technic% alguidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virusthat-cau%ses-it

[10] C. P. E. R. E. Novel, ‘‘The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (Covid-19) in China,’’ Zhonghua Liu Xing Bing Xue Za Zhi= Zhonghua Liuxingbingxue Zazhi, vol. 41, no. 2, p. 145, 2020.

[11] L. van der Hoek, K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout, ‘‘Identification of a new human Coronavirus,’’ Nature Med., vol. 10, no. 4, pp. 368–373, 2004.

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

[13] 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]. Available: http://www. sciencedirect.com/science/article/pii/B9780128002070000010

[14] H.-L.Hwa,W.-H.Kuo,L.-Y.Chang,M.-Y.Wang,T.-H.Tung,K.-J.Chang, and F.-J. Hsieh, ‘‘Prediction of breast cancer and lymph node metastatic statuswithtumourmarkersusinglogisticregressionm odels,’’J.Eval.Clin. Pract., vol. 14, no. 2, pp. 275–280, Apr. 2008. [15] R. Tibshirani, ‘‘Regression shrinkage and selection via the lasso,’’ J. Roy. Stat. Soc., Ser. B, Methodol., vol. 58, no. 1, pp. 267–288, Jan. 1996.

[16] A. E. Hoerl and R. W. Kennard, ‘‘Ridge regression: Biased estimation for nonorthogonal problems,’’ Technometrics, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[17] X.F.Du,S.C.H.Leung,J.L.Zhang,andK.K.Lai,''Demand forecasting of perishable farm products using support vector machine,'' Int. J. Syst. Sci., vol. 44, no. 3, pp. 556–567, Mar. 2013.

[18] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, ''Tweets classification on the base of sentiments for US airline companies,'' Entropy, vol. 21, no. 11, p. 1078, Nov. 2019.

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

[20] J.Lupón,H.K.Gaggin,M.deAntonio,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.

[21] 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.

[22] C. Willmott and K. Matsuura, ''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.

[23] R. Kaundal, A. S. Kapoor, and G. P. Raghava, ''Machine learning techniques in 51835856

disease forecasting: A case study on rice blast prediction,'' BMC Bioinf., vol. 7, no. 1, p. 485, 2006.

[24] S. Baran and D. Nemoda, ''Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting,'' Environmetrics, vol. 27, no. 5, pp. 280–292, Aug. 2016.

[25] Y. Grushka-Cockayne and V. R. R. Jose, ''Combining prediction intervals in the m4 competition,'' Int. J. Forecasting, vol. 36, no. 1, pp. 178–185, Jan. 2020.