Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Performance evaluation of regression models for COVID-19: A statistical and predictive perspective

Mohammad Ayoub Khan a,⇑, Rijwan Khan b, Fahad Algarni a, Indrajeet Kumar c, Akshika Choudhary b, Aditi Srivastava b

a College of Computing and Information Technology, University of Bisha, Bisha 67714, Saudi Arabia
b ABES Institute of Technology, Ghaziabad, U.P., India
c Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India

ABSTRACT

Research is very important in the pandemic situation of COVID-19 to deliver a speedy solution to this problem. COVID-19 has presented governments, corporations and ordinary citizens around the world with technology playing an essential role to tackle the crisis. Moderate and flexible innovation arrangements that can speed up progress towards giving critical well-being ability are proved hourly. Knowledge with the aid of creativity must be obtained, accepted and analysed in a short time frame. In this example, the machine learning model has a major role to play in predicting the number of next positive COVID-19 cases to come. For government departments to take effective and strengthened future COVID-19 planning and innovation. The ongoing global pandemic of COVID-19 has been non-linear and dynamic. Due to the especially perplexing nature of the COVID-19 episode and its diversity from country to country, this study recommends machine learning as a convincing means to demonstrate flare-up. In this linear regression, polynomial regression, ridge regression, polynomial ridge regression, support vector regression models, the COVID-19 data set from multiple on-line tools have been evaluated. During the work process comprehensive experiments were performed and each test was evaluated with the parameters mean square error (MSE), medium absolute error (MAE), root mean square error (RMSE) and R2 score. This study also offers a path for future research using regression models based on machine learning. Precise validation and data analysis can contribute to strategies for healing and disease prevention at an early stage. A systematic comprehensive strategy is a new philosophy in which statistical data for government agencies and community can be forecast.

© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Coronavirus or COVID-19 originated in Wuhan, China in the month of December 2019. Till date (12 July 2021), there is no avowed human immunization for combating it. COVID-19 engendering is speedy when people are in vicinity [1]. Coronavirus would be the most unnerving situation faced by the whole world since the end of the second world war. Even more unnerving as the enemy is invisible. COVID-19 has affected our lives to a great extent. Work, economy, education and almost everything has come to stand still. People nowadays only focus on what is essential to live. Government and other organizations have had to take unprecedented steps to stem the trail of destruction. The reasons for its spread and thinking about its threat, practically all the nations have proclaimed either partial or complete lockdowns all through the influenced districts and regions. Since there is no endorsed medicine till now for slaughtering the infection so the legislatures of all nations are concentrating on the precautionary measures which can stop the spread [2]. Since there is no embraced solution till now for dispatching this contamination, theauthorities of all countries are paying attention to the careful steps that can minimize the spread. This virus causes respiratory tract infections that can range from mild to fatal. This infection

https://doi.org/10.1016/j.asej.2021.08.016
2090-4479/© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Mohammad Ayoub Khan, R. Khan, F. Algarni et al., Performance evaluation of regression models for COVID-19: A statistical and predictive perspective, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2021.08.016
can influence individuals of any age yet older people or individuals with previous ailments are more inclined or vulnerable to it. The most testing part of its spread is that an individual can have the infection for a long time without indicating side effects [3–6]. Utilization of mask, sanitizer, customary hand washing and cleanliness is the most ideal approach to keep from this illness. People are urged to believe Staying homes safer lives. COVID-19 will reshape our reality.

The World Health Organization (WHO) on February 11, 2020 came forward stating this virus as a pandemic outbreak naming it as COVID-19 [21] further stating that the virus had first taken place in China, making it to different countries. USA, Brazil, and India have been the most affected countries where the number of cases of this pandemic COVID-19 multiplied rapidly on the daily basis.

Machine learning plays a major role in better understanding and examining COVID-19 crisis as it identifies the patterns in data and uses them to automatically make predictions or decisions. In the medicinal services area, Machine Learning can be seen as an asset that has the extension to process colossal datasets past the capacity of human personalities and the derived bits of knowledge help doctors in arranging and giving consideration to get acceptable treatment. On the basis of this concept, this work has been carried out for analysis of COVID-19 cases and prediction of upcoming new positive cases [18]. In this model authors are going to look at the development of Coronavirus affirmed, demise and recouped instances of whole world and comparison of five countries that have been vigorously tainted done with rest of world. Investigation and predictions are done using Linear Regression, Ridge Regression, Polynomial ridge Regression [5], polynomial Regression and Support Vector Regression (SVM). The proposed expectation model guarantees that it follows the authentic outcome with respect to this pandemic circumstance so that tremendous financial misfortune, network spread, measure of social separation of individuals can be identifiers and furthermore precise choice can be taken likewise [7]. This strategy will ensure the administration to yield preventive evaluations reliant on our next work for foreseeing the presence of this infection in future.

The remainder of the paper comprises five sections. Section 2 present the Related Work, Section 3 presents the Materials and Methods and Staging Analysis Parameter used for investigation, Section 4 presents the evaluation with comparison table and Section 5 sums up the paper and presents the conclusion with references.

2. Related work

Since the time the rise of COVID-19 and its resulting spread across landmasses overwhelming both progressed and creating countries, there has been a great deal of exploration papers distributions on different parts of COVID-19. So, in various research paper analysis is done on vaccination, drug therapy and also on the prediction of future infected, recovered and death cases [1,2]. Using various forecasting techniques

L. Jia et al., 2020 [12] did the prediction and analysis of COVID-19 through various models. The differential equation, classical differential equation, historical arrangement expectation model is utilized in the project. They also used an Internet-based infectious disease prediction model. They used various models which came under mathematical models such as the Logistic model, Bertalanffy model [12] and Gompertz model [12]. Model evaluation was done using regression coefficient (R2). Fitting and analysis of SARS and COVID-19 was done on two basis/categories: number of affirmed cases and number of passing toll. Different prediction graphs were obtained under these categories. According to this paper they concluded that this epidemic will be over presumably in the last of April but their estimation is not so accurate [12]. F. Rustam et al. (2020) introduced future forecasting on Covid-19 using various machine learning models. They use four regression models namely LR, LASSO, SVM and Exponential Smoothing. They predicted their results on the basis of Evaluation Parameters used in their models. According to their results ES model is best for predicting infected cases and recovered cases and LR model best predict the death rate [9].

S. Dutta et al., 2020 [20] introduced in their paper the prediction models for confirming COVID-19 cases. They used three models in their paper namely LSTM (Long Short-Term Memory) model, GRU (Gated Recurrent Unit) model and combined LSTM-GRU Model [20]. Deep Learning Neural Network is employed. The consolidated LSTM-GRU based RNN model gives a relatively better outcomes as far as forecast of affirmed, delivered, negative, demise cases on the information [20]. Similarly, Narinder et al., 2020 [19] in this paper mirrored that because the eruption of the COVID-19 has turned into pandemic, the analysis of medicine knowledge is required to arrange the society to combat this challenge. The four stages of COVID-19 are described in this paper. The regression is trained and cross-validated on real time knowledge victimization the quantity of confirmed, recovered, and death cases. The results were displayed through graphs that pictured COVID-19 worldwide epidemic analysis victimization of machine learning and deep learning techniques/algorithms [19].

The work carried by Sobhani and Sareeta in 2020 [10] proposed a model in which they have done data visualization after that they used various forecasting machine learning techniques. During this paper a comparison is formed between the expansion of Covid-19 confirmed deaths and recovered cases of Asian countries to differ-ent major countries that have conjointly been heavily infected. The various predictions are aforethought on a bar chart. Gender distribution is mirrored through a chart that reveals males are a lot of possible to be diagnosed with Covid-19. Age wise distribution is additionally shown through pie charts. They concluded that Sigmoid model is best among all the models they used [10]. Binti Hamzah et al., [8] and his fellow mates in year 2020 made corona tracker an Online stage that provides most up-to-date what’s a lot of, solid news advancement, even as insights and investigation on COVID-19. They have done continuous information inquiry and imagined on their site and at that point questioned information is utilized for SEIR modelling. On the basis of day by day observations they utilized the SEIR model to figure Covid-19 episodes inside and Outside of China. Similarly, they have done investigation on the Queried News [8].

L. Wynants et al., (2020) proposed a model in which they have inspected all the papers identified with covid-19 pandemic. They recognized that proposed schemes ineffectively detailed and at high danger of predisposition, raising worry that their expectations could be untrustworthy once applied day by day [4]. R. Suja et al., in 2020 used three models for their predictions and did visualization on India. Using graph plotting, they try to figure out the future trend of this epidemic. From apprehended qualities and coordination cases from dataset information they concluded that MLP strategy is giving acceptable forecast outcomes than that of the LR and VAR technique [11]. L. Peng et al., (2020) using SEIR model they predict in light of the open information of National Wellbeing Commission of China from January twentieth to February sixteenth, 2020, they dependably gauge key plague boundaries and make forecasts in the curvature period and conceivable consumption time of 24 regions in Territory and 16 regions in Hubei region [17]. Li Cuilian et al., in 2020 proposed a model to assess the forecast estimation of the Internet search information from online web indexes and web-based social networking for the COVID-19 episode in China [16]. Vinay and Lei Zhang (2020) reflected in their
paper the approximate finishing purpose of this epidemic in Canada and in the world. They have developed a forecasting model using deep learning techniques. The dataset required for this purpose was collected from a University and Canadian fitness management. Long short-term memory model was displayed to forecast COVID-19 and based on the graphs depicted by LSTM they predicted that this pandemic will end by June 2020. Despite the fact that their model accomplished better execution when contrasted and other anticipating models [15].

Dong Je et al., (2020) introduced a model that clarifies the high-risk factors of coronavirus. They collected the clinical data of all the patients that were admitted to Fuyang Second People’s Hospital in China between 20th January and 22nd February 2020. They divided around 208 patients under two categories- Stable group and Progressive group. Univariate and multivariate analysis depicted the autonomous high-hazard factors for COVID-19 movement. They concluded that the use of the CALL score model depicted the autonomous high-hazard factors for COVID-19 group and Progressive group. Univariate and multivariate analysis depicted the autonomous high-hazard factors for COVID-19 movement. They concluded that the use of the CALL score model can be an effective resource to combat this challenge [14]. Weston C. Roda et al., (2020) summarized that after comparing SEIR and SIR models that it is not necessary that a complex model such as SEIR is more effective than a simple model SIR. They used Bayesian framework, Markov chain Monte Carlo algorithms and Akaike Information Criterion (AIC). Parameters in the SIR model and Parameters in the SEIR model were shown through tables. Comparisons were made through graphs [13]. C. Sohrabi et al., 2020 [21] tells that on 31st December 2019 few cases of coronavirus were identified in China. WHO global health emergency, global reaction, reported UK cases and British response, transmission, viral spread, preventing, diagnosing, treating, forecasting and containment methods for this virus have been addressed by the authors.

The authors also compared the diagnostic criteria for the CDC against the WHO based on symptoms and travel. In the paper the authors have also described the retrospective from response to COVID-19.

R. Khan et al., 2020 [42] investigate sentiment analysis techniques for the analysis of twitter covid-19 data. They find that it has been seen by pre-handling the information utilizing the regex and coach has been a viable answer for getting out the unpredictability of the applied calculation also the information instead of straightforwardly applying on the crude information itself. By utilizing the prepared model and further utilizing it with the classifier end up being a superior route for grouping as it decreased the time period and the size of the information outline diminishing the time unpredictability associated with the cycle. Discussing the examination, it has been seen that the number for tweets shared by the dynamic clients has been consistently more prominent when contrasted with different assessments. It implies with respect to the pandemic greatest number of individuals thought and took the choices made by the administration or the nearby experts in a positive manner. While the quantity of tainted and demised individuals continued expanding it didn’t influence the psychological quality of the populace. For the 3-month investigation with respect to the Indian sub-landmass the variety among the positive, negative and nonpartisan estimations stayed consistent with the quantity of expanding cases step by step.

Gozes, Ophir, et al, 2020 [43], underlying examination, which is right now being extended to a bigger populace, exhibited that quickly created AI-based picture investigation can accomplish high precision in identification of Coronavirus just as measurement and following of sickness trouble.

Shi, Feng, et al, 2020 [44] discussed that for COVID-19 pandemic clinical imaging techniques has a significant function in battling against COVID-19. In their research authors examines how AI gives sheltered, precise and proficient imaging arrangements in Coronavirus applications. The X-ray and CT scan techniques have been utilized to exhibits the adequacy of Man-made intelligence engaged clinical imaging. It is significant that imaging just gives incomplete data about patients. This is imperative to join imaging information with clinical indications and lab assessment results to support the screening, location furthermore, finding of COVID-19. For this situation, we trust AI will exhibit its capacity in combining data from this multi-source information, for performing precise and productive conclusion, investigation and development.

Yan, Carol H., et al, 2020 [45], In ambulatory people with flu-like side effects, chemosensory brokenness was unequivocally connected with COVID-19 contamination and ought to be viewed as when screening manifestations. Most will recoup chemosensory work inside weeks, resembling goal of other illness related manifestations.

Sun, Yinxiaohe, et al, 2020, [46] discusses that clinical and laboratory data can be rapidly established for individuals who are exposed to COVID-19 and allow PCR testing and containment efforts to be made a priority. For prediction models, the primary laboratory test results were important.

Oliveiros, Barbara, et al, 2020 [47] discussed that COVID-19 progression rate is projected to be slower with spring and summer. However, the two variables make up at most 18% of the transition, with the other 82% connected to other factors, such as containment steps, general health policies, population density, transport, cultural matters, etc. Furthermore, the direct effect is small: if the temperature rises by 20 °C for example, the average time doubling is predicted to increase by 1.8 days in the best-case scenario.

Chen, Xiaofeng, et al, 2020 [48] discussed that pneumonia patients with and without COVID-19 can be differentiated on the basis of CT imaging and clinical symptoms alone. A model consisting of semantic and clinical radiological features performs very well in diagnosing COVID-19.

Dryhurst, Sarah, et al.[49] said that the risk of COVID-19 around the world clearly show the consistent connection between COVID-19 perceptions of risk and various experience- and socio-cultural factors across countries. At the same time, we notice that cultural differences in risk perception must be addressed.

2.1. Motivation

As we know, data science helps us to clarify and understand the data that has been accumulated so far. Data science helps one to simulate and imagine the patterns of how coronavirus spreads including the number of patients reported daily with coronavirus or possible infected cases, i.e. it is growing or more than before and so on. Data analysis helps one to gain some valuable insight into the data. Machine Learning lets us make predictions and lets us create models that map the real world and that will take some data and predict the future. So, in our model, we first try to imagine data and then forecast future data and also try to find the best fit regression model that could help us with future predictions.

3. Materials and methods

3.1. Dataset preparation

In this examination paper we have utilized the overall details of coronavirus from January 22, 2020, to July12, 2021, was assembled from the online resources like Kaggle, Weka 3.8.4 and Orange [11]. The datasets give us the quantity of affirmed cases, recouped cases, and passing cases everywhere throughout the world. The datasets are accessible in time arrangement position with death, month and year so fleeting parts are not neglected.
We isolated our datasets into a preparing set (85%) on which our model readied and a testing set (15%) to test the exhibition of our framework.

3.2. Experimental work flow chart

The experimental work flow diagram of the work is shown in Fig. 1. From the Fig. 1, it has been found that the complete work divided into two sections. Section one is used for data visualization and section two is used for future data prediction. For data visualization, progress bar, recovery rate and mortality rate are shown. In Section 2, linear, polynomial, ridge, ridge polynomial and support vector machine (SVM) having polynomial kernel function is used for future case prediction. Each model is evaluated using MSE, MAE, RMSE and R2_SCORE parameters.

3.2.1. Data visualization

3.2.1.1. Worldwide visualization framework. In this work demonstrate the week after week progress of various sort of cases in world which incorporates affirmed cases, recouped cases and demise cases. From this it has been inferred that the pace of increment in affirmed cases is a lot higher than the recovered and demise rate cases. Passing rate is low in contrast with other two cases though recouped rate is moderate. At that point we have indicated the everyday increment in affirmed cases, recouped cases and demise cases. The curve given in Fig. 2 shows the week number vs number of cases line graph shows the weekly increase of different types of cases in worldwide. From our prediction confirmed cases are increasing with a high rate and recovery rate is average as compared to number of active cases. Also, death rate is less as compared to number of active cases.
3.3. Visualization framework observations

After the completion of visualization experiments, the following observations have been drawn for India cases:

(a) It took 529 days in India to succeed in 30837222.0 Confirmed Cases.
(b) It took 528 days in Italy to succeed in range of Confirmed cases to India.
(c) It took 440 days in US to succeed in range of Confirmed cases to India.
(d) It took 527 days in Kingdom of Spain to succeed in range of Confirmed cases to India.
(e) It took 537 days in China to succeed in range of Confirmed cases to India.

3.4. Mortality and recovery rate visualization

Mortality rate is a proportion of the quantity of deaths in a specific populace, scaled to the size of that populace, per unit of time. Recovery rate is the degree to which head and accumulated enthusiasm on defaulted obligation can be recouped, communicated as a level of assumed worth. The monthly progress of different types of cases in world is shown in Fig. 3.

From the Fig. 3, it has been found that the mean and median of average mortality is reported as 3.413 and 2.622, respectively. Similarly, the mean and median of average recovery rate is evaluated as 52.20 and 56.61, respectively.

3.5. Comparative analysis

In this section a comparative analysis has been done according to mortality and recovery rate of top 15 countries. Fig. 4 shows the list of top 15 countries having high mortality rate and low mortality rate. According to this, Yemen has the highest mortality rate and Bhutan get the lowest position in the list of mortality rate country.

Fig. 5 shows the top 15 countries according to high recovery rate and low recovery rate. In this figure, it is noticed that Djibouti has highest recovery rate and Sweden has lowest recovery rate among the country list.

According to statistical analysis done by the author’s, the following key facts have been observed:

(a) Top 15 Countries according to High Mortality Rate are Yemen, Peru, Mexico, Sudan, Syria, Egypt, Somalia, Taiwan, Bosania and Herzegovina, Ecuador, China, Bulgaria, Afghanistan, Tanzania and Bolivia. Yemen has the highest mortality rate all over the world.
(b) Top 15 Countries according Low Mortality Rate are Bahrain, Mongolia, Entrea, Cyprus, Seychelles, Iceland, Vietnam, UAE, Maldives, Qatar, Timor-Leste, Burundi, Laos, Singapore and Bhutan. Bhutan has the lowest mortality rate all over the world.
(c) Top 15 Countries according to the High Recovery Rate are Singapore, Qatar, Bahrain, Gabon, Iceland, Israel, Djibouti, Cote d’Ivoire, Burkina Faso, Belarus, West Bank and Gaza, San Marino, Diamond Princess, Peru and Albania. Singapore has the highest recovery rate all over the world.
(d) Top 15 Countries according Low Recovery Rate are Vietnam, Greece, Fiji, Burundi, Norway, Iceland, Thailand, France, Spain, Netherlands, United Kingdom, Serbia, Sweden, Belgium and US. US has the lowest recovery rate all over the world.

3.6. Growth factor

Growth factor is a proportion of how rapidly the quantity of new cases is rising or falling and the basic thing to recall is that we need to keep it fewer than one. Critically, one thing that has changed since the last time the development factor was over one is that the quantity of new cases every day is low. It is simply the factor by which an amount duplicates itself after some time. Fig. 6 shows the growth factor of different types of cases worldwide.

3.7. Doubling rate

Multiplying of cases is additionally a significant factor of COVID-19. Perceiving the doubling rate can make the government and individuals aware about the severity of the disease and hence enhance the preventive measures. The following table shows the doubling rate of coronavirus. Table 1 shows the doubling rate of coronavirus infected cases.

The above Figs. 7 which compares the confirmed, recovered and death cases of India, China, UK, Italy, US and rest of the world. We can infer from the graph that confirmed, recovered and death cases of the rest world is higher than the other five countries.

(a) The blue line represents the confirmed, death and recovered cases of India. It is observed that confirmed cases and death cases- in India increased exponentially after March 2021. Recovery rate was maximum in July 2021.
(b) The orange line represents the confirmed, death and recovered cases of Mainland China. It is observed that it maintains a constant rate.
(c) The green line represents the confirmed, death and recovered cases of Italy. It is observed that the confirmed cases are constant but the recovered and death cases have increased after March 2021.
(d) The red line represents the confirmed, death and recovered cases of US. It is observed that the confirmed cases have increased exponentially whereas the death cases and recovered cases have increased after January 2021.
(e) The purple line represents the confirmed, death and recovered cases of UK. It can be inferred that the death cases have increased gradually after November 2020.

The brown line shows the confirmed, death and recovered cases of rest of the world. It is observed that confirmed, death and recovered cases have increased exponentially with time.
The above Figs. 7a and 7b which compares the confirmed, recovered and death cases of India, China, UK, Italy, US and rest of the world. We can infer from the graph that confirmed, recovered and death cases of the rest world is higher than the other five countries.

(a) The blue line represents the confirmed, death and recovered cases of India. It is observed that confirmed cases and death cases- in India increased exponentially after March 2021. Recovery rate was maximum in July 2021.
(b) The orange line represents the confirmed, death and recovered cases of Mainland China. It is observed that it maintains a constant rate.
(c) The green line represents the confirmed, death and recovered cases of Italy. It is observed that the confirmed cases
are constant but the recovered and death cases have increased after March 2021.

(d) The red line represents the confirmed, death and recovered cases of US. It is observed that the confirmed cases have increased exponentially whereas the death cases and recovered cases have increased after January 2021.

(e) The purple line represents the confirmed, death and recovered cases of UK. It can be inferred that the death cases have increased gradually after November 2020.

(f) The brown line shows the confirmed, death and recovered cases of rest of the world. It is observed that confirmed, death and recovered cases have increased exponentially with time.

The above graph shown in Fig. 8, presents the mortality rate and recovery rate comparison of 5 countries with the rest of the world. Mortality rate of U.K has increased in the timestamp whereas India has least mortality rate. Recovery rate of mainland China is highest and U.K has least recovery rate.

(a) Blue line represents the mortality and recovery rate of India. It is observed that the mortality rate increased from mid of April 2020 till mid of September 2020 and then it started decreasing. The recovery rate decreased from mid of March to mid of May and then started increasing gradually.

(b) Orange line represents the mortality and recovery rate of Mainland China. It is observed that the mortality rate increased steeply after May 2020. The recovery rate is represented by normal yield curve.

(c) Green line represents the mortality and recovery rate of Italy. It is observed that mortality rate increased after April
2020 and then decreased in mid of November 2020 and recovery rate increased after November 2020 in Italy.

(d) Red line represents the mortality and recovery rate of US. It is observed that mortality rate increased gradually from March till mid of April 2020.

(e) Purple line represents the mortality rate and recovery rate of UK. It is observed that mortality rate has decreased after November 2020 and recovery rate is constant after mid of December 2020.

(f) Brown line represents the mortality rate and recovery rate of rest of the world. It is observed that mortality rate has decreased after September 2020 and recovery rate has increased after September 2020.

3.7.1. Regression models

In this work unsupervised machine learning algorithm is used for the prediction of future cases of COVID-19 in terms of confirmed cases, recovered cases and death cases. For this study, linear regression (LR), polynomial Regression (PR), Ridge and Lasso regression (RLR), Ridge Regression (RR) and Support vector machine (SVM) regression models are used. The brief description of each model is given here.

3.7.2. Linear regression (LR)

It is an immediate method to manage showing the association between a scalar response (or ward variable) and at any rate one instructive factor (or free variables). It was the essential kind of backslide examination to be thought completely, and to be used broadly in feasible applications. This is because models which depend legitimately upon their dark limits are less difficult to fit than models which are non-straightforwardly related to their limits and considering the way that the real properties of the resulting estimators are easier to choose. One of the information digging methods utilized for expectation undertakings is Linear Regression [22–25]. In an issue with one indicator, this procedure attempts to locate the best line to fit. The formula to calculate linear regression is given in Eq. (1).

\[ \hat{y} = b_0 + b_1 x \]  

where, \( b_0 \) is a constant, \( b_1 \) is the regression coefficient; \( x \) is the independent variable and \( \hat{y} \) is the predicted value of dependent variable.
3.7.3. Polynomial regression

In estimations, it is a kind of relapse examination wherein the association between the free factor $x$ and the poor variable $y$ is shown as a farthest breaking point polynomial in $x$. It fits a nonlinear association between the estimation of $x$ and the relating unforeseen mean of $y$, implied $E(y|x)$. Even though polynomial regression fits a nonlinear model to the data, as a quantifiable estimation issue it is immediate, as in the relapse work $E(y|x)$ is straight in the dark limits that are assessed from the data [26–28]. Subsequently, polynomial relapse is seen as an uncommon case of various direct relapses. It can be calculated according to Eq. (2).

\[
\hat{y} = q_1 + q_2 x_1^2 + q_3 x_1^3 + \ldots + q_n x_1^n \quad (2)
\]

where $\hat{y}$ is the predicted value of dependent variable $x_1, x_1^2, x_1^3, \ldots x_1^n$ are independent variables and $q_1, q_2, q_3, \ldots q_n$ are all coefficients.

3.7.4. Ridge and Lasso regression

For some, machine learning issues with an enormous number of highlights or a low number of perceptions, a direct model tends to overfit and variable determination is dubious. Models that utilize shrinkage, for example, Lasso and Ridge can improve the forecast precision as they diminish the estimation difference while giving an interpretable last model. Ridge and Lasso expand on the direct model; however, their basic idiosyncrasy is regularization. The objective of these techniques is to improve the mistake work with the goal that it depends not just on the whole of the squared contrasts yet in addition on the relapse coefficients [29–31].

The primary thing in the advancement of such a framework is the right determination of the regularization boundary. Compared to Linear Regression, Ridge and Lasso models are progressively impervious to anomalies and the spread of information. Generally speaking, their fundamental reason for existing is to forestall overfitting. The primary distinction between Ridge relapse and Lasso is the means by which they dole out a punishment term to the coefficients.

3.7.5. Ridge regression

The Ridge regression is a method which is specific to multilinear regression information which is multicollinearity in nature. Ridge Regression is one of the most head regularization methodologies which isn’t used by various people due to the multifaceted science behind it [31–33]. On the off chance that you have a general thought regarding the idea of different relapse, it’s not that hard to investigate the science behind Ridge relapse in $r$. Relapse is the equivalent, what makes regularization diverse is that the way how the model coefficients are resolved. The equation for this technique is given in Eq. (3).

\[
\text{cost}(a) = \frac{1}{2b} \sum_{p=1}^{b} (z_p - \hat{z}_p)^2 + \lambda \sum_{q=1}^{n} a_q^2 \quad (3)
\]

This sets the coefficient that can be said as min (whole of square residuals + $\lambda \|\text{slope}\|_1$), where $\lambda \|\text{slope}\|_1$ is punishment term.

3.7.6. Support vector regression

They are astounding yet versatile controlled AI figuring which are used both for course of action and relapse. SVMs have their exceptional strategy for execution when diverged from other AI estimations. As of late, they are incredibly notable because of their

---

Table 1

| S. N | No. of cases | Days Since first Case | Number of days for doubling |
|------|--------------|----------------------|-----------------------------|
| 1    | 5,000        | 5 days               | 5 days                      |
| 2    | 10,000       | 9 days               | 4 days                      |
| 3    | 20,000       | 12 days              | 3 days                      |
| 4    | 40,000       | 17 days              | 5 days                      |
| 5    | 80,000       | 33 days              | 16 days                     |
| 6    | 160,000      | 52 days              | 19 days                     |
| 7    | 320,000      | 59 days              | 7 days                      |
| 8    | 640,000      | 65 days              | 6 days                      |
| 9    | 1,280,000    | 74 days              | 9 days                      |
| 10   | 2,560,000    | 90 days              | 16 days                     |
| 11   | 5,120,000    | 120 days             | 30 days                     |
| 12   | 10,240,000   | 158 days             | 38 days                     |
| 13   | 20,480,000   | 202 days             | 44 days                     |
| 14   | 40,960,000   | 272 days             | 70 days                     |
| 15   | 81,920,000   | 341 days             | 69 days                     |
| 16   | 163,840,000  | 481 days             | 140 days                    |

Comparison of India, China, UK, Italy, US and rest of the world.
Fig. 7a. Comparison of 5 countries with rest of world on the basis of confirmed and recovered cases.

Fig. 7b. Comparison of 5 countries with rest of world on the basis of death cases.
ability to manage different relentless and obvious components. A SVM model is basically a depiction of different classes in a hyperplane in multidimensional space [34,35]. The hyperplane will be delivered in an iterative manner by SVM with the objective that the goof can be constrained. The goal of SVM is to parcel the datasets into classes to find the biggest immaterial hyperplane (MMH).

**RBFkernel:** The RBF Kernel is additionally called the Gaussian part. There’s a boundless number of measurements within the feature space since it can be extended by the Taylor Arrangement. Within the arrange underneath, the $\gamma$ parameter characterizes how much impact a single preparing illustration has. The larger it is, the closer other examples must be to be influenced [36]. It may be a general-purpose part; utilized when there’s no earlier information around the data. Mathematically, it is represented as given in Eq. (4).

$$M(a, a') = \exp\left(-\frac{||a-a'||^2}{2\sigma^2}\right)$$  \hspace{1cm} (4)

$||a-a'||^2$ denotes the squared Euclidean distance between two feature vectors. $\sigma$ is a free parameter.

**Polynomial Kernel:** The polynomial kernel looks not as it were at the given highlights of input tests to decide their similitude [37], but too combinations of these. With n unique highlights and k degrees of polynomial, the polynomial bit yields $n^k$ extended highlights where d is the degree of polynomial. It is prevalent in image processing. Mathematically, it is represented as given in Eq. (5).

Fig. 8. Comparison of 5 countries with rest of world on the basis of Mortality rate and Recovery rate.
\[ a(t_i - t_c) = (t_i - t_d + 1)^k \]  

where \( k \) is the degree of polynomial.

**Sigmoid Kernel:** The Hyperbolic Tangent Kernel is additionally known as the Sigmoid kernel and as the Multilayer Perceptron (MLP) kernel \([41]\). The Sigmoid kernel comes from the Neural Systems field, where the bipolar sigmoid work is frequently utilized as an enactment function for artificial neurons. Mathematically, it is represented as given in Eq. (6).

\[ h(a, b) = \tan h(\alpha a^b + i) \]  

### 3.7.7. Support vector regression

Support Vector Regression is like Linear Regression in that the condition of the line is \( y = wx + b \) in SVR, this straight line is alluded to as a hyperplane. The data centres around either side of the hyperplane that are closest to the hyperplane are called Support Vectors which are used to plot the breaking point line.

Not at all like other Regression models that endeavour to confine the bumble between the verifiable and foreseen worth, the SVR endeavours to fit the best line inside an edge regard (Distance among hyperplane and limit line), \( a \). Consequently, we can say that SVR model attempts fulfill the following condition:

\[-a < y < (y - wx + b) < a\]

It utilized the focuses with this limit to anticipate the worth.

### 3.8. Staging analysis parameters

**RMSE:** It is the standard deviation of the residuals. Residuals are a proportion of how a long way from the relapse line information focuses is. RMSE is an extent of how spread out these residuals is. Figuratively speaking, it uncovers to you how centred the data focuses is. RMSE is an extent of the idea of an estimator. It is reliably non-negative, and characteristics increasingly like zero are better.

\[ \text{RMSE} = \sqrt{\frac{1}{k} \sum_{m=1}^{k} (a_m - \hat{a}_m)^2} \]  

where \( k \) is the number of observations, \( a_m \) is the observed value, and \( \hat{a}_m \) is the predicted value.

**R2 SCORE:** It is a factual proportion of how close the information is to the fitted relapse line. It is otherwise called the coefficient of assurance, or the coefficient of various assurances for numerous relapses. \( R^2 \) is always between 0 and 100%.

\[ R^2 \text{ Score} = 1 - \frac{\sum (t_i - \hat{t}_i)^2}{(t_i - \bar{t}_i)^2} \]  

where, \( t_i \) is the actual cumulative confirmed cases, \( \hat{t}_i \) is the predicted cumulative confirmed cases, \( \bar{t}_i \) is the average of the actual cumulative confirmed cases.

**Mean Absolute Error:** It measures the normal size of the errors in a lot of forecasts, without thinking about their heading \([38]\). It’s the normal over the test of the outright contrasts among expectation and genuine perception where every individual distinction has equivalent weight. The obtained value of MAE is calculated according to Eq. (9).

\[ \text{MAE} = \frac{1}{p} \sum_{q=1}^{p} |r_q - \hat{r}| \]  

where, \( p \) represents the number of errors, \( |r_q - \hat{r}| \) denotes the absolute errors.

**Mean Squared Error:** MSE of an estimator is the normal squared contrast between the assessed qualities and the real worth. The MSE is an extent of the idea of an estimator. It is reliably non-negative, and characteristics increasingly like zero are better. MSE is calculated according to Eq. (10).

\[ \text{MSE} = \frac{1}{z} \sum_{i=1}^{z} (y_i - \hat{y}_i)^2 \]  

where, \( z \) is the number of data points, \( y_i \) represent the observed values, and \( \hat{y}_i \) represent predicted value.

### 4. Experiments and results

#### 4.1. Dataset preparation and Experiment

To make our data perfect with sklearn design, we made another section called “Days since” which tracks the quantity of days since the initial date. We have taken four evaluation parameters namely Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), \( r^2 \)-score for identification of the best regression model which predicts the best about future cases, recovered cases and death cases. The brief description of experiment carried out for the work is given in Table 2.

This investigation endeavours to build up a framework for the future gauging of the quantity of cases influenced by COVID-19 utilizing AI strategies. The dataset utilized for the investigation contains data about the day by day reports of the number of recently contaminated cases, the quantity of recoverations, and the quantity of passing due to COVID-19 around the world. We have utilized five AI models LR, polynomial, Ridge, polynomial ridge and SVM to foresee the quantity of recently affected cases, the quantity of passing, and the quantity of recoverations. We additionally attempt to anticipate the best model for gauging affected cases, recuperated cases and the passing cases.

#### 4.2. Results analysis

**Experiment 1:** The investigation performs forecasts on infected cases and concurring to results polynomial ridge performs better among all the models; SVM additionally performs well and polynomial relapse performs normal score. In correlation, LR and Ridge performs most exceedingly terrible in this circumstance and accomplish nearly the same \( R^2 \)-score. Be that as it may, contrasting the current confirmed cases statistics and our models’ forecasts, the polynomial ridge expectation is following the patterns which are near the genuine circumstance.

Results of the extensive experiments have been reported in Table 3 for active cases.

Fig. 9 shows the performance of all regression models on the prediction of confirmed cases on the basis of the testing set of data.

### Table 2

| Experiment No. | Description |
|----------------|-------------|
| Experiment 1   | Experiment carried out for active cases using five different regression models. |
| Experiment 2   | Experiment carried out for recovered cases using five different regression models. |
| Experiment 3   | Experiment carried out for death cases using five different regression models. |

Mohammad Ayoub Khan, R. Khan, F. Algarni et al. Ain Shams Engineering Journal xxx (xxxx) xxx
From the figure we can visualize that the infected cases tend to increase day by day which is an alarming sign for all the infected countries.

(a) Black dotted line represents the actual confirmed cases on the basis of testing set.
(b) Yellow line represents the LR and Red line represents the ridge regression and the prediction of both the models are very much similar and overlap each other.
(c) Purple line represents polynomial regression predicted confirmed cases on testing set.
(d) Green line represents polynomial ridge regression predicted confirmed cases on testing set.
(e) Blue represents the SVM (kernel = poly) predicted confirmed cases on testing set.

Fig. 10 shows the performance of all the models on the basis of trained data. From this we have examine that polynomial ridge, SVM and polynomial predicts good result and following the pattern which are near to actual confirmed cases. LR and ridge predict similar result and does not follow the pattern of actual cases.

(a) Here, black line represents actual future cases and grey line represents the training cases.
(b) Yellow line represents the LR and Brown line represents the ridge regression and the future prediction of both the models are very much similar and overlap each other.
(c) Purple line represents the prediction of confirmed cases using polynomial regression.
(d) Green line represents the prediction of confirmed cases using polynomial ridge regression.
(e) Blue line represents the prediction of confirmed cases using SVM.

**Experiment 2:** The investigation performs forecasts on Recovered cases and concurring to results polynomial ridge performs better among all the models; polynomial Regression additionally performs well and SVM performs normal score. In correlation, LR and Ridge performs most exceedingly terrible in this circumstance and accomplish nearly the same R2_score.

Table 4 depicts the result of performance of all models and polynomial ridge gives best fit values of MSE, MAE, RMSE and R2_score.

Fig. 11 shows the performance of LR, polynomial, Ridge, polynomial ridge and SVM in the form of graph. Be that as it may, contrasting the current recuperation insights and our models’ expectations, the polynomial ridge forecast is following the patterns which are near the genuine circumstance.

(a) Black dotted line represents the actual recovered cases on the basis of testing set.
(b) Yellow line represents LR and red represents ridge regression having similar pattern of prediction of recovered cases.
(c) Brown line represents polynomial regression recovered cases prediction on testing set.

(d) Green line represents polynomial ridge prediction of recovered cases on the basis of testing set.
(e) Blue line represents SVM prediction of recovered cases on the basis of testing set.

From Fig. 12 we observe that polynomial and polynomial ridge following the pattern which is almost similar to actual recovered cases and SVM gives average performance but the pattern of linear and ridge regression is almost different from actual recovered cases.

(a) Here, Crimson line represents actual recovered cases and Black line represents the training cases.
(b) Yellow line represents the LR and Brown line represents the ridge regression and the prediction of recovered cases of both the models are very much similar and overlap each other.
(c) Purple line represents the polynomial prediction of recovered cases.
(d) Green line represents the polynomial ridge prediction of recovered cases.
(e) Blue line represents the SVM prediction of recovered cases.

**Experiment 3:** In death rate future estimation of the polynomial ridge performs better among the various models. Every single other model performs ineffectively, the request for execution from best to most noticeably awful is polynomial ridge is best trailed by LR, Ridge, polynomial and SVM because of the nature of accessible time-arrangement information. From our observation we also observed that no model can predict the best results for death rate prediction.

Table 5 depicts the result of performance of all models and linear and ridge gives best fit values of MSE, MAE, RMSE and R2_score.

Fig. 13 shows the performance of Different Regression Models on Prediction of Death cases on the testing set. From this we conclude that not a single regression model gives us best result.

(a) Black dotted line represents the actual death cases on the basis of testing set.
(b) Blue line represents LR and red represents ridge regression having similar pattern of prediction of death cases.
(c) Yellow line represents polynomial regression death cases prediction on testing set.
(d) Green line represents polynomial ridge prediction of death cases on the basis of testing set.
(e) Blue line represents SVM prediction of death cases on the basis of testing set.

From Fig. 14 we concluded that till now we cannot any model which follows the pattern of actual death cases. Only polynomial ridge model comes out with better result.

(a) Here, Red line represents actual death cases and Black line represents the training cases.

**Table 3** Obtained results for Confirmed cases using different regression model.

| Model Name          | MSE       | MAE       | RMSE     | R2_score  |
|---------------------|-----------|-----------|----------|-----------|
| Linear Regression   | 32303716  | 56457125  | 56836358| −113.4    |
| Polynomial Regression| 17953305 | 37454555  | 42371353| −21.0     |
| Ridge               | 32336474  | 56485978  | 56865169| −113.5    |
| Polynomial ridge    | 73738332  | 24617677  | 27154802| −125.5    |
| SVM                 | 36452186  | 56956658  | 60735646| −2.6      |

Mohammad Ayoub Khan, R. Khan, F. Algarni et al. Ain Shams Engineering Journal xxx (xxxx) xxx
(b) Yellow line represents the LR and Brown line represents the ridge regression and the prediction of death cases of both the models are very much similar and overlap each other.

(c) Purple line represents the polynomial prediction of death cases.

(d) Green line represents the polynomial ridge prediction of death cases.

(e) Blue line represents the SVM prediction of death cases.

Linear and Ridge Regression models’ results are very similar to each other and they predict very similar r2_score value in all the three cases.

4.3. Comparative analysis with state of art

Table 6 represent the comparative analysis of the work carried out for this study with previously published work.
Finally, comparative analysis of the present with previously published work [9–11] is performed and the major findings of the study is given as:

a. In the study, visualization of Active Cases, closed cases (recovered and death) all over the world has been reported.
b. The study also performs the visualization of weekly and daily progress of different types of cases in the world.
c. The work also highlights the country wise confirmed cases, Death cases, Mortality rate, Recovery Rate.
d. Comparison of India, China, US, UK, Italy has been done with rest of the world.
e. The study has also visualized the growth factor of different cases World-wide.
f. Analysis on India and comparison with other counties to find out how much time other countries took to reach a certain number of confirmed cases in comparison to India has been done.
g. Prediction of the time taken for doubling the number of confirmed cases has also been done.
h. Future prediction using five regression models and analysis on the premise of Mean square Error, Mean Absolute Error has been done.

i. From the comparison Model it can be inferred that no alternative has done as several things as through this model.

Table 6
Comparative analysis of the work carried out for this study with previously published work.

| Parameters                              | F. Rustam et al., 2020, [9] | S. Sengupta et al., 2020, [10] | R. Sujath et al., 2020, [11] | Present study |
|-----------------------------------------|-----------------------------|--------------------------------|-------------------------------|---------------|
| World Data Visualization on the basis of Different Factors | NO                          | Yes                            | They do only for India        | YES           |
| Comparison of 5 countries with rest of the world | NO                          | NO                             | NO                            | YES           |
| Growth Factor of different type of Cases Worldwide | NO                          | YES                            | NO                            | YES           |
| Data analysis of India and Comparison With other countries | NO                          | Done but not Compare with Other Countries | Done but not Compare with Other Countries | YES           |
| Time Taken for Doubling of Infected Cases | NO                          | Only with four models          | Only with three models         | ONLY with three models | YES           |
| Future Prediction using 5 Models        | Only with four models       | Only with three models          |                               |               |

Fig. 13. Performance of Different Regression Models on Prediction of Death cases on the testing set.

Fig. 14. Performance of Different Regression Models on Prediction of death cases.
5. Conclusion

This examination introduced current patterns of COVID-19 occurrence from twenty-second February 2020 to 2nd September 2020 as envisioned in our task. The quickly expanding number of latest COVID-19 cases day by day worldwide has placed an overwhelming weight on clinical assets in nations with enormous flare-ups. Therefore, prediction of future confirmed cases became necessary. The size of information accessible is gigantic associate degree gathered knowledge and obtaining an intriguing example out of the cumulated information could be a tough trip. With the common info regarding confirmed, recovered and ending across Republic of India for over the time term helps in anticipating and deciding the not thus distant future. Our outcomes imply that there is a sure comprehensiveness in the time development of Covid-19.

This proposes a nation that turns into the venue of a scourge flood can be regarded, in any event in first guess, as a very much mixed synthetic reactor, where various populace associate as indicated by mass-activity like principles with little association with topographical variances.

In light of itemized examination of the overall information, we gauge a few key boundaries for COVID-19, similar to the inactive time, the isolate time and the compelling propagation number in a moderately dependable manner, and foresee the inflection point, conceivable completion time and final absolute tainted cases for everywhere throughout the world. We have done prediction by using 5 regression models so that we can conclude which model is best and on the basis of that model we are able to tell the rate of increase in number of infected, recovered and death cases in future. The experiments have been completed with best MSE value as 73738324668782.8, MAE as 24617677.8, RMSE as 27154802.9 and R2_score as −125.5 using Polynomial Ridge model for confirmed cases. For recovered cases best MSE value as 1524067870205389.0, MAE value as 38118611.9, RMSE as 39039311.8 and R2_score as −197.1 using linear regression model. Also, for Death cases best MSE value as 1359199481041.9, MAE value as 1106760.1, RMSE value as 1165847.0 and R2_score as −2.5 using SVM model. From our investigation we concluded that polynomial ridge is the best model for predicting confirmed cases, linear regression model for recovered cases and SVM regression model for death cases.

Funding

There is no funding available for this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Petropoulos F, Makridakis S, Brauerstein LA. Forecasting the novel coronavirus COVID-19. PLoS ONE 2020;15(3):e0231236. doi: https://doi.org/10.1371/journal.pone.0231236
[2] Al-qaness MAA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. J Clin Med 2020;9(3):674. doi: https://doi.org/10.3390/jcm9030674
[3] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 2020;134:109761. doi: https://doi.org/10.1016/j.chaos.2020.109761
[4] Sperri M, Grant SW, Peek N. Prediction models for diagnosis and prognosis in Covid-19. BMJ 2020;m1464. doi: https://doi.org/10.1136/bmj.m1464
[5] Gupta R, Pandey G, Chaudhary P, Pal SK. SEIR and regression model based COVID-19 outbreak predictions in India; 2020. https://doi.org/10.1016/j.ijbiomac.2020.04.09525
[6] Perc M, GórskaMiksińska N, Slavíček M, Štoter A. Forecasting COVID-19. Front Phys 2020;8:doi: https://doi.org/10.3389/fphy.2020.00127
[7] Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020;26(7):1037–40. doi: https://doi.org/10.1038/s41591-020-0921-0
[8] Hamzah FAB. CoronaTracker: world-wide COVID-19 outbreak data analysis and prediction; 2020. https://doi.org/10.2471/BLT.20.255695
[9] Rustam F, Reshi AA, Mehmoond A, Ullah S, On B-W, Aslam W, et al. COVID-19 future forecasting using supervised machine learning models. IEEE Access 2020;8:101489–99. doi: https://doi.org/10.1109/ACCESS.2020.3009711
[10] Sengupta S, Muge R, Sharma G. Covid-19 pandemic data analysis and forecasting using machine learning algorithms; 2020. https://doi.org/10.1016/j.ejccse.2020.06.25.1040004
[11] Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Env Res Risk Assess 2020;34(7):959–72. doi: https://doi.org/10.1007/s00477-020-01827-8
[12] Predictive analysis and Coronavirus Disease; 2019. Lin Jia, Kewen Li, Yu Jiang, Xin Guo, Ting zhao. https://arxiv.org/abs/2003.05447v2
[13] Roda VC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 2020;5:271–81. doi: https://doi.org/10.1016/j.idm.2020.01.001
[14] JD. Prediction for progression risk in patients with COVID-19 pneumonia: the COVID-19 cohort study. Clin Infect Dis 2020; doi: https://doi.org/10.1093/cid/ciaa604
[15] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020;135:109864. doi: https://doi.org/10.1016/j.chaos.2020.109864
[16] Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance 2020;25(10). doi: https://doi.org/10.2807/1560-7917.ES.2020.25.10.2004414
[17] Peng L, Yang W, Zhang D, Zheng C, Hong L. Epidemiic analysis of COVID-19 in China by dynamical modelling; 2020. https://doi.org/10.1101/2020.02.16.20023465
[18] Ivanov D. Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (COVID-19- SARS-CoV-2) case. Transport Res Part E: Logist Transport Rev 2020;136:101922. doi: https://doi.org/10.1016/j.tre.2020.101922
[19] NS, Soobhadra SK. Agarwal S. COVID-19 epidemiic analysis using machine learning and deep learning algorithms; 2020. https://doi.org/10.1101/2020.04.08.20057679.
[20] Bandyopadhyay SK, Dutta S. Machine learning approach for confirmation of COVID-19 cases; positive, negative, death and release; 2020. https://doi.org/10.1016/j.epidem.2020.03.023
[21] Sohrabi C, Alsafi Z, O’Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020;76:71–6. doi: https://doi.org/10.1016/j.ijsu.2020.02.034
[22] Schleich M, Otteau D, Cucanu RD. Learning linear regression models over factorized joins. Proceedings of the 2016 international conference on management of data, 2016.
[23] Groeneboom P, Hendrickx K. Current status linear regression. Ann Statist 2018;46(4):1415–44.
[24] Cheng XL, et al. Polynomial regression as an alternative to neural nets, arXiv preprint arXiv:1806.06850; 2018.
[25] Aguirre-Urreta Miguel I, MilkoRönkköö, Hu Jiang. Polynomial regression and measurement error: implications for information systems research. In: ACM SIGMS database: the DATABASE for advances in information systems vol. 51, no. 3; 2020. p. 55–80.
[26] Li H, Yamamoto S. Polynomial regression-based model-free predictive control for nonlinear systems. 2016 55th annual conference of the society of instrument and control engineers of Japan (SICE), 2016.
[27] Yang X, Wen R. Ridge and lasso regression models for cross-version defect prediction. IEEE Trans Reliab 2018;67(3):885–96.
[28] Melkumova LE, Shatskikh SY. Comparing Ridge and LASSO estimators for data analysis. Procedia Eng 2017;201:746–55.
[29] Pereira Jose Manuel, Basto Mario, Ferreira Amelia, da Silva. The logistic lasso and ridge regression in predicting corporate failure. Procedia Econ Finance 2016;39:634–41.
[30] Chowdhury Agniva, Yang Jasen, Drinas Petros. An iterative, sketching-based framework for ridge regression. In: International conference on machine learning; 2018.
[31] Kilbia BM, Banik Shripa. Some ridge regression estimators and their performances. J Modern Appl Statist Methods 2016;15(1):206–38.
[32] Santamaria-Ronful G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. Renew Energy 2016;85:790–809.
Cheng Kai, Lu Zhenzhou, Zhou Yicheng, Shi Yan, Wei Yuhao. Global sensitivity analysis using support vector regression. Appl Math Model 2017;49:587–98.

Gopi Arepalli Peda et al. Classification of tweets data based on polarity using improved RBF kernel of SVM. Int J Inform Technol 2020:1–16.

Bin Lin, Shaobo, Zeng Jinshan. Fast learning with polynomial kernels. IEEE Trans Cybern 2019;49(10):3780–92.

de Myttenaere Arnaud, Golden Boris, Le Grand Bénédicte, Rossi Fabrice. Mean absolute percentage error for regression models. Neurocomputing 2016;192:38–48.

Chai Tianfeng, Draxler Roland R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci Model Dev 2014;7(3):1247–50.

Chai Tianfeng, Draxler Roland R. Root mean square error (RMSE) or mean absolute error (MAE)? GMDD 2014;7(1):1525–34.

Camps-Valls G, Martin-Guerrero JD, Rojo-Alvarez JL, Soria-Olivas E. Fuzzy sigmoid kernel for support vector classifiers. Neurocomputing 2004;62:501–6.

Khan R, Shrivastava P, Kapoor A, Tiwari A, Mittal A. Social media analysis with AI: sentiment analysis techniques for the analysis of twitter covid-19 data. J Critic Rev 2020;7(9):2761–74.

Gozes Ophir et al., Rapid ai development cycle for the coronavirus (covid-19) pandemic: initial results for automated detection & patient monitoring using deep learning and image analysis, arXiv preprint arXiv:2003.05037; 2020.

Shi Feng et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev Biomed Eng 2020.

Van Carol H et al. Association of chemosensory dysfunction and Covid-19 in patients presenting with influenza-like symptoms. Int Forum Allergy Rhinol 2020.

Sun Yinxiaohe et al. Epidemiological and clinical predictors of COVID-19. Clin Infect Dis 2020;71(15):786–92.

Oliveiros Barbosa et al. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases, medRxiv; 2020.

Chen Xiaofeng, Tang Yanyan, Mo Yongkang, Li Shengkai, Lin Daiying, Yang Zhijian, et al. (COVID-19) based on radiological semantic and clinical features: a multi-center study. Eur Radiol 2020;30(9):4893–902.

Dr. Ayoub Khan (Senior Member, IEEE) received a Ph. D (Electrical Engg.) from Jamia Millia Islamia, New Delhi, India and a Master of Technology (Computer Science and Engineering) from Guru Gobind Singh Indraprastha, New Delhi, India. He worked with many leading organizations like C-DAC (Ministry of IT and Communications), Noida, Sharda University. Presently, he is an Associate Professor at the University of Bisha, Saudi Arabia with interests in the Internet of Things, RFID, wireless sensor networks, ad hoc networks, smart cities, industrial IoT, signal processing, NFC, routing in network-on-chip, real time and embedded systems. He has more than 15 years of experience in his research areas. He has published many research papers and books in reputed journals and international conferences. He contributes to the research community by undertaking various volunteer activities in the capacity of editor for many journals and as a conference chair.

Dr. Rijwan Khan has obtained Ph. D in Computer Engg. From Jamia Millia Islamia, New Delhi, India. He is presently work as head of department & professor at ABES Institute of Technology (Affiliated to AKTU), Ghaziabad in the Department of Computer Science & Engineering. He has more than 50 publications in his area of research.

Dr. Indrajeet Kumar is working as assistant professor with graphic era hill university, Dehradun. His research interest includes IoT, Ad hoc network, data analytics etc.

Akshika Choudhary is a scholar at ABES IT Engg. College, Ghaziabad with research interest in IoT, Ad hoc network, data analytics etc.

Aditi Srivastava is a scholar at ABES IT Engg. College, Ghaziabad with research interest in data analytics and machine learning.