International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Prediction of Peak Time of Corona Virus Using Machine-Learning Approach

Manisha Thakre and Kirti Jain

Department of Computer Science & Engineering, SORT
Peoples University Bhopal, Madhya Pradesh, INDIA
[1] ugcmanishamam@gmail.com
[2]Kirtijain9@gmail.com

A B S T R A C T

COVID- 19 epidemics have a detrimental impact on the nation's economy as well as people's quality of life. On January 30, 2020, the World Health Organization declared it a worldwide health emergency. (WHO). By April 28, 2020, more than 3 million individuals had caught this virus, and there was no vaccine to prevent it. The WHO imposed certain safety restrictions, but they were only precautionary measures. In the fight against this pandemic, information technology, with a focus on fields such as data science and machine learning, might be beneficial. It is critical to have early warning systems in place that can estimate how much a disease would hurt society and allow the government to take proper action without severely affecting the country's economy. We present a review approaches for forecasting forthcoming instances based on previous data of the published papers in related to this work. Using machine-learning techniques, two methodologies are described: one for estimating the chance of developing the virus and the other for projecting the number of positive cases. Several algorithms were examined, and the focuses on the one that generated the most accurate results. The two classifiers were random forest and extra tree classifiers, both of which had an accuracy of greater than 90%. The additional tree classifier is the more accurate of the two. These results may be used by various governmental agencies to conduct remedial actions. It may be easier to tackle COVID-19 if infectious disease forecasting technologies are available.

Keywords: COVID-19, Regression model.

1. Introduction

COVID-19 is no longer simply a name. Tens of thousands of people have been affected by a dreadful, all-encompassing sickness that has swept the globe. Wuhan City, China, was the source in December 2019 [1]. COVID-19 began to spread from person to person when no one was aware of the virus; it has since steadily spread to practically every country and has now become a pandemic. COVID-19 is the acronym for coronavirus disease. SARS-CoV-2, also known as 2019-nCoV, is a disease caused by a new coronavirus (nCoV) discovered in 2019. Despite the fact that its official name was COVID-19, the International Committee on Taxonomy of Viruses called this virus SARS-CoV-2 since it shared symptoms with the virus that caused the SARS pandemic in 2003. The World Health Organization (WHO) named the virus COVID-19 to prevent misunderstanding with other viruses because it had not before been detected in humans and had significantly afflicted them this time.

Early on, the COVID-19 epidemic in Wuhan City, Hubei Province, China was only recognised as a case of respiratory illness. On December 31, 2019, China notified the WHO of this respiratory ailment [4]. The WHO declared COVID-19, a global health emergency, on January 30, 2020. According to WHO records, COVID-19 was declared a worldwide pandemic on March 11, 2020, one year after H1N1 was declared a global pandemic in 2009.

* *Corresponding author.*
E-mail address: ugcmanishamam@gmail.com

COVID-19 was chosen by the WHO to minimise stigmatising connotations with the virus's origins in terms of communities, geography, or animals [5]. Coronaviruses are defined by the WHO and other health organisations as a category of viruses with symptoms ranging from the common cold to more serious infections. nCoV, on the other hand, is an unique viral kind that has never been discovered in humans. As COVID-19 cases increased substantially, this respiratory condition was immediately recognised by countries all over the world. COVID-19 has infected an increasing number of people since it was found in China. Since it was declared a pandemic, the WHO has issued guidelines for all countries on how to tell if someone has the disease, how to avoid getting sick from it, what precautions to take, when to visit the hospital, the severity of an infection, and symptoms of this virus after a thorough examination of infected individuals.

To avoid public panic, the WHO is constantly disseminating information about this virus to individuals in various countries. The WHO did not warn against travel early in COVID-19. Strict guidelines included keeping a safe distance from sick persons, routinely washing your hands, and covering your mouth if you coughed or had a cold. However, travel history eventually became one of the primary COVID-19 indicators, and as a result, screenings of all visitors from other countries, particularly those from contaminated regions, were undertaken on a regular basis. According to the WHO, the symptomatic period for this virus was 14 days, thus anyone travelling from another nation was recommended to stay at home alone during that time. If a person demonstrated any indications of sickness, they were sent to the hospital for treatment.

The modelling approach considers three distinct groups of well-mixed populations: those susceptible to infection (class S), those infected (class I), and the eliminated population. It is predicated on the idea that infectious illnesses spread through human contact (class R is devoted to those who have recovered, developed immunity, been isolated or passed away). Furthermore, the infection from class I is likely to spread to class S, with the chance of transmission being inversely linked to the number of contacts [2].

The proportion of people in class S grows over time as a time series, which is typically computed using the following fundamental differential equation:

$$\frac{dS}{dt} = -\alpha SI$$

Eqn. **Error! No text of specified style in document.**.1

When I represents the infected population and S represents the susceptible population, both given as fractions, the differential equation's daily reproduction rate regulates the number of susceptible infectious contacts [6]. The value of the time-series generated by the differential equation slowly declines. Initially, it is assumed that the epidemic is in its early stages and that there are just a few persons in class I. As a result of this, the rise becomes linear, and class I may be computed as follows:

$$\frac{dI}{dt} = \alpha SI - \beta I$$

Eqn. **Error! No text of specified style in document.**.2

Where keeps track of the daily rate of new infections by calculating the number of affected persons who can spread the disease. Furthermore, the class R, which represents people who are immune to infection, is determined as follows:

$$\frac{dR}{dt} = \beta I$$

Eqn.1.3

The outbreak exponential growth can be calculated using Eq. 1.3's unrestricted conditions for the excluded group:

$$I(t) \approx exp((\alpha -$$

Eqn. **Error! No text of specified style in document.**.3

Eq. 1.4 has been used to simulate a wide range of infectious disease outbreaks. However, due of the strong limitations enforced by authorities, the susceptibility to infection has been significantly affected for the COVID-19 outbreak forecast. For example, the SIR model cannot provide favourable results in China, Italy, France, Hungary, or Spain because individuals deliberately entered quarantine and limited their social participation. The model has shown some relative accuracy for countries whose containment attempts have been delayed, such as the United States [7].

Another issue is the short lead-time of classical epidemiological models. The median success rate of the outbreak forecast gives helpful data for evaluating the models' efficiency. The following formula may be used to get the median prediction factor:

$$f = \frac{prediction}{True\ value}$$

Eqn. **Error! No text of specified style in document.**.4

With longer lead periods, the model's accuracy drops [8]. Overall, for conventional epidemiological models to be effective and trustworthy, social interactions must be stable across time (i.e., no changes to interventions or control measures), and they must also be well-known for class R in order to compute Eq. 1.3. Several unique models were developed that commonly used data from social media or call data records (CDR) to acquire information on class R, and the results were promising. However, an examination of COVID-19 behaviour in a variety of countries indicates a high amount of complexity and ambiguity.

To achieve reliable findings, epidemiological models must be adapted to the local environment based on insights about susceptibility to infection due to changes in public health interventions, as well as the various stages in the SIR/SEIR model. As a result, the robustness and generalizability of traditional models are severely limited. As a result, it is critical to develop accurate models with high generalisation capabilities that can be scaled to anticipate both regional and global pandemics.

Because of the intricacy and large-scale nature of the challenge in constructing epidemiological models, machine learning (ML) has recently received interest for generating outbreak prediction models. The purpose of machine learning approaches is to develop models that are more generalizable and trustworthy in their predictions over longer time periods. Although machine learning techniques have been used to simulate earlier pandemics (such as Ebola, Cholera, swine fever, H1N1 influenza, dengue fever, Zika, and oyster norovirus), there is a scarcity of peer-reviewed studies related to COVID-19. Table 1.1 lists notable machine learning algorithms for outbreak prediction. These machine learning approaches are limited to the classification and regression tree, neural network, Naive Bayes, Bayesian network, and random forest techniques (CART). Although machine learning (ML) has long been recognised as a fundamental tool for modelling weather forecasts and natural catastrophes, its application in modelling epidemics is still in its early stages. More study is needed on advanced ML approaches like as hybrids and ensembles. The goal of this study is to look at how machine learning may be utilised to model the COVID-19 pandemic.

The purpose of this study is to investigate the accuracy and generalizability of the suggested ML models for varied lead-times.

Two critical areas of research that machine learning can fill are highlighted by modern machine learning approaches for epidemic prediction modelling. The capacity to predict the pandemic using time series, followed by the creation of SIR and SEIR models. Given the inadequacies of the present SIR and SEIR, machine learning may surely assist.

Table Error! No text of specified style in document.**.1 Notable ML methods for outbreak prediction**

Authors	**Journal**	**Outbreak infection**	**Machine learning**
[1]	Geospatial Health	Dengue fever	Neural Network
[2]	Environment International	Oyster norovirus	Neural Network
[9]	Global Ecology and Biogeography	H1N1 flu	Neural Network
[15]	Transboundary and Emerging Diseases	Swine fever	Random Forest
[22]	BMC Research Notes	Influenza	Random Forest

This study contributes to the advancement of COVID-19's time series prediction. An early benchmarking is presented to demonstrate the potential of machine learning for further research. According to the article, genuine originality in outbreak prediction may be achieved by merging machine learning with SEIR models.

2. INCUBATIONPERIOD OF COVID-19

The incubation period is the time between catching the virus and the start of symptoms. This virus may incubate in the human body for 2 to 14 days, according to the WHO. According to the Centers for Disease Control and Prevention (CDC), the virus's mild symptoms worsen after the first five days.

More recent data on patients, on the other hand, demonstrated that the virus began to evolve and, after multiple negative tests, unexpectedly provided a positive result, increasing the incubation duration from 14 to 20 or 28 days. Those with a strong immune system tested positive for the virus while not showing any symptoms, according to sources. Because the symptoms of this virus do not appear in persons with strong immune systems, becoming infected by someone who has the virus but a strong immune system is unavoidable.

3. COUNTRIES MOST AFFECTED BY COVID-19

A number of countries are affected by the COVID-19 virus. Thousands of individuals have died as a result of the coronavirus in some of the worst-affected areas [22].

- On December 31, 2019, China announced the discovery of its first positive case in Wuhan, China, the genesis of COVID-19. In China, a COVID-19 update showed 82,816 cases, of which 4632 died, 77,346 were recovered, and 838 were still active.
- Rome got the first good case report from Italy on January 31, 2020. Of the 192,994 patients there, 25,969 died, 60,498 recovered completely, and 106,527 remained active.
- Spain announced its first positive case on January 31, 2020. On January 19, 2020, the United States reported its first case. There were 227,759 cases, 22,902 of which died, 95,708 recovered, and 105,149 remained active.
- Of the 928,364 patients, 52,356 perished and 110,490 were rescued. It is a substantial sum.
- Germany reported its first instance on January 27, 2020. Out of the 155,407 cases, 5802 died, 109,800 were recovered, and 39,805 were still active.
- France reported its first instance on January 24, 2020. There were 22,245 deaths, 43,493 recoveries, and 94,090 ongoing cases among the 159,828 patients.

- Iran reported its first incidence on February 19, 2020. There were 89,328 incidents, of which 5650 died, 68,193 were recovered, and 15,485 were still open.
- The United Kingdom reported its first case on January 29, 2020. Turkey reported their first incidence on March 10, 2020, and there were 148,377 cases in total, with 20,319 fatalities.
- 2600 individuals died and 21,737 people recovered from the 104,912 cases.
- Aside from these countries, Kerala, India, reported the first incidence on January 30, 2020. There were 24,942 cases, 5498 of which were recovered and 780 of which died.

4. MACHINE LEARNING

According to Arthur Samuel (1959), machine learning (ML) is the field of study that enables computers to learn without being explicitly programmed. As a result, we may define ML as the field of computer science in which self-learning machines can be developed.

Simply said, learning happens as a result of earlier work observations, such as examples, or teaching. This method is looking for patterns in data and utilising examples to help make decisions. The primary objective of machine learning (ML) is to teach computers to learn on their own, without human intervention, and to adjust their behaviour accordingly.

The ML technique is depicted in Figure 1.1 [5]. The model is trained on past data before being evaluated on new data and used to generate predictions. A portion of the available historical data is utilised to evaluate the trained ML model's performance (which is not present during training). This is commonly referred to as the validation procedure. In this approach, the ML model is evaluated using performance measures such as accuracy. Accuracy evaluates how well the ML model works on unknown data in terms of the ratio of correctly predicted features to all features available for prediction.

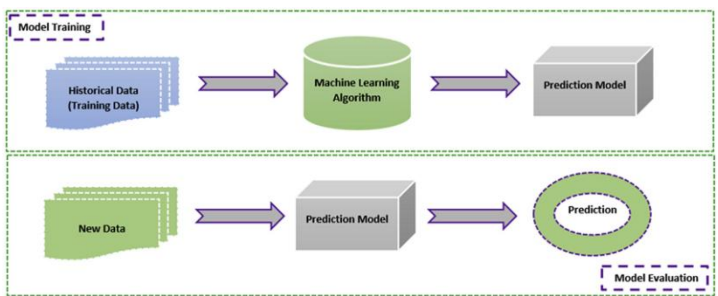

Figure Error! No text of specified style in document.**.1Machine learning process [5]**

SOME MACHINE LEARNING METHODS

Machine learning algorithms are classified into two types: supervised and unsupervised learning.

1. supervised ML algorithms are a type of ML approach that may be used to generate new data from labelled data and predict future events or labels based on what has previously been learnt. A supervisor (labels) is there to instruct or correct while learning in this method. Following the known training set, the learning process is utilised to anticipate the output values for this initial analysis. The output of the learning system may be compared to the actual output; if inconsistencies are discovered, they can be corrected and the model updated accordingly.
2. Unsupervised machine learning algorithms: These do not have a supervisor to instruct or correct them. This type of learning method is used when there is unlabeled or unclassified data available to train the system. The system does not describe the desired result, but it is capable of inferring (making rules from) datasets and uncovering latent patterns from unlabeled data by examining the data.
3. Semi-supervised machine learning approaches come under the categories of supervised learning and unsupervised learning. As a result, for training purposes, this type of learning algorithm employs both unlabeled and labelled data, frequently a small amount of labelled data and a large number of unlabeled data. This method is used to improve the precision of learning.
4. Reinforcement is a form of learning approach that provides rewards or penalties based on the task that the system completes. If the system fails to execute a job that we've trained it to complete, it will be penalised; if it completes the task correctly, it will be rewarded. It normally functions on a scale of 0 to 1, with 0 representing punishment and 1 denoting reward.
5. It is founded on the premise that if we train a dog or a bird to perform a task and it does it precisely as we ask, we should reward it with food or praise that it loves. This is the prize. If it failed to complete the assignment appropriately, we may scold it as a kind of punishment.

USE OF MACHINE LEARNING IN COVID-19

Machine learning is used in a range of industries, including medical, to predict disease and its progression. In medicine, prompt diagnosis and treatment are critical for positive outcomes. A therapy with a high error rate might result in several fatalities. As a result, scientists have begun to use artificial intelligence in medicine. It is a matter of life and death for the researchers to choose the suitable technology, making the procedure difficult.

ML achieved strides in the medical field with this endeavour. ML techniques are used to understand, analyse, and forecast the outcomes of large datasets. These ML algorithms were used to classify data into therapy groups and to identify illness signs and symptoms. With the use of ML, hospitals can cure infectious illnesses while also maintaining administrative procedures.

ML methods have previously been utilised to treat ailments such as cancer, pneumonia, diabetes, Parkinson disease, arthritis, neuromuscular disorders, and many more; they produce outcomes in prediction and forecasting that are more than 90% accurate. The COVID-19 pandemic illness is a terrible virus that has claimed countless lives throughout the world. There is no treatment for this virus. ML techniques have been used to predict if individuals are infected with the virus based on symptoms given by the WHO and CDC.

ML is also used to diagnose the sickness based on x-ray images. For example, chest images can be used to determine whether a patient has COVID-19 infection. Furthermore, ML can measure social distance; by using this strategy, we can defend ourselves from COVID-19.

DIFFERENT TECHNIQUES FOR PREDICTION AND FORECASTING

A number of ML methods are used to predict and anticipate future occurrences. Some ML techniques used for prediction include support vector machines, linear regression, logistic regression, naive Bayes, decision trees (random forest and ETC), K-nearest neighbour, and neural networks (multilayer perceptron).

Moving average, simple exponential smoothing, Holt's linear trend model, Holt-Winters model, seasonal autoregressive integrated moving average exogenous model (SARIMAX), and autoregressive integrated moving average model are some ML techniques used to forecast future events, similar to the naive approach (ARIMA).

Each approach is employed differently and has unique characteristics depending on the accuracy outcomes. During the model evaluation step, the model with the best accuracy is chosen for prediction or forecasting. When testing model performance, we selected to employ the ETC for symptom-based COVID-19 prediction and the ARIMA forecasting model to estimate the number of confirmed COVID-19 cases in India since these methods had the best accuracy scores of all the classifier and forecasting strategies we investigated.

Figure 1.2 depicts a flowchart of the ML technique, which demonstrates how data is collected and pre-processed, as well as how it is divided into a training dataset and a test dataset for training and performance evaluation.

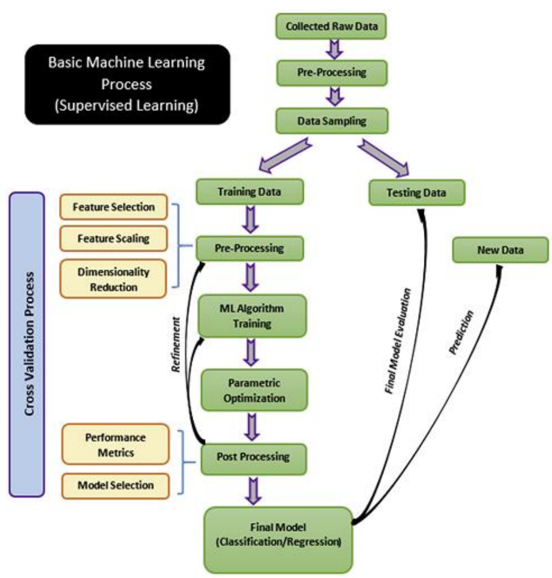

Figure Error! No text of specified style in document.**.2 Machine learning (ML) process flowchart [5]**

Machine learning may be divided into two categories: supervised and unsupervised learning. There are several alternative algorithms.

SUPERVISED LEARNING

An algorithm learns the link between given inputs and provided outputs by using training data and human feedback. For example, a practitioner might utilize marketing expenses and weather forecasts as input data to forecast can sales. Table 1.2 shows some supervised learning methods.

Table Error! No text of specified style in document.**.2 Some supervised learning algorithm**

Algorithm	**Description**	**Type**
Linear regression	Finds a technique to link each characteristic to the output in order to anticipate future values.	Regression
Logistic regression	A linear regression extension that is useful for categorization problems. 3 is a binary output variable.	Classification
Decision tree	Classification or regression model with high interpretability that divides data-feature values into branches at decision nodes	Regression Classification
Naive Bayes	The Bayesian method is a classification technique that employs the Bayesian theorem. The theorem updates an event's previous knowledge with the independent probability of each characteristic that potentially influence the occurrence	Regression Classification
Support vector	For classification, the Support Vector Machine, or SVM, is commonly employed. The	Regression (not very

machine	SVM method finds a hyperplane that divides the classes best. It works well when combined with a non-linear solution.	common) Classification
Random forest	The technique is based on a decision tree, which greatly improves accuracy. Random forest builds numerous basic decision trees and using the'majority vote' approach to determine the label to return. The final forecast for the classification job will be the one with the most votes, but the final prediction for the regression task will be the average prediction of all the trees.	Regression Classification
Ada Boost	A classification or regression strategy that use a number of models to make a judgement but ranks them according to their accuracy in predicting the outcome.	Regression Classification
Gradient-boosting trees	Gradient-boosting trees are a cutting-edge classification and regression approach. It focuses on the preceding trees' mistakes and attempts to fix them.	Regression Classification

When the output data is known, supervised learning may be used. New data will be predicted by the programme.

There are two categories of supervised learning:

- Classification task
- Regression task

Classification

Assume you want to forecast a customer's gender for an ad. You will begin collecting data from your client database on their height, weight, employment, pay, purchase basket, and so on. You know each of your customers' gender; it can only be male or female. The classifier's goal will be to assign a likelihood of being male or female (i.e., the label) based on the information (i.e., features you have collected). When the model has trained to distinguish whether a person is male or female, you may use new data to generate a forecast. For example, suppose you recently received fresh information from an unknown consumer and want to know whether he or she is male or female. If the classifier predicts male = 70%, it signifies that the algorithm is certain that this consumer is 70% male and 30% female.

The label might belong to two or more categories. The Machine Learning example above has only two classes, however if a classifier is required to predict an item, it has hundreds of classes (e.g., glass, table, shoes, etc. each object represents a class)

Regression

The job is a regression when the output is a continuous value. For example, a financial analyst may be required to anticipate the value of a stock based on a variety of factors such as equity, prior stock performance, and a macroeconomics index. The algorithm will be taught to estimate stock prices with the least amount of inaccuracy feasible.

UNSUPERVISED LEARNING

As method in unsupervised learning investigates input data without being assigned an explicit output variable (e.g., explores customer demographic data to identify patterns). Table 1.3 shows some unsupervised learning methods.

When you don't know how to categorise the data and want the algorithm to detect patterns and classify it for you, you may utilise it.

Table Error! No text of specified style in document.**.3 Some unsupervised learning algorithm**

Algorithm Name	**Description**	**Type**
K-means clustering	Puts data into k groups, each of which contains data with comparable qualities (as determined by the model, not in advance by humans)	Clustering
Gaussian mixture model	A generalisation of k-means clustering that allows for greater flexibility in group size and shape (clusters)	Clustering
Hierarchical clustering	To construct a categorization system, splits groups along a hierarchical tree. Customers with Cluster loyalty cards can utilise this.	Clustering
Recommender system	Assist in the definition of relevant facts for producing a suggestion.	Clustering
PCA/T-SNE	Typically used to reduce the dimensionality of data. The methods condense the quantity of characteristics to three or four vectors with the biggest variances.	Dimension Reduction

5. METHODOLOGY

Linear Regression (LR) is a simple strategy for showing the link between a dependent variable and at least one independent variable by rearranging the data as a four regression model. LR was the most focused and commonly utilised kind of regression analysis in relevant applications. By fitting a straight condition to base data, LR reveals the link between two variables. One variable is seen to be independent, whereas the other is thought to be dependent. Polynomial Features (PR), Support Vector Regression (SVR), and Decision Tree Regressor (DTR).

Taking data and modeling data as following steps:

1. Data Collection (http://api.covid19india.org/csv/latest/raw_data1.csv) in csv format or any other source
2. Data cleaning and data transformation

3. Data modeling
4. Validation
5. Prediction

6. MACHINE LEARNING

Machine Learning is regarded as a subset of AI. ML techniques teach the computer to forecast or make sensible judgments without explicit preparation by using patterns or "training samples" in data (processed information). Machine learning (ML) is the scientific study of algorithms and statistical models used by computer systems to perform tasks using patterns and inference rather than explicit instructions. Time-series data, which are data sequences collected over time, can be input into machine learning algorithms. This type of data demonstrates the progression of a phenomena through time. Let Xt be a time-series vector, T being the collection of all evenly spaced time points and xt being the outbreak at time point t. In order to appropriately train ML systems, we designed two scenarios, which are given in Table 1.4.

Table 1.4 indicates that scenario 1 forecasts the pandemic on day t using data from three weeks, but scenario 2 forecasts the outbreak on day t using data from five days. The Machine Learning approaches were used to these two cases.

Table 1.4 Input and output variables for training ML methods by time-series data

	Inputs	**Input number**	**Output**
Scenario 1	xt-1, xt-7, xt-14, and xt-21	Four inputs	xt (outbreak)
Scenario 2	xt-1, xt-2, xt-3, xt-4, and xt-5	Five inputs	xt (outbreak)

MULTI-LAYERED PERCEPTRON (MLP)

ANN is a concept influenced by the biological nervous system, which processes information similarly to the brain. The main component of this approach is the novel structure of the information processing system. The system is made up of many highly linked processing units known as neurons, which work together to discover a solution. ANNs, like humans, learn through imitation. Throughout the learning process, the neural network is programmed to perform specific tasks such as pattern recognition and data classification. Synaptic connections between neurons govern learning in biological systems. This approach is also used by neural networks. Through the analysis of experimental data, ANNs learn by passing knowledge or a law underlying the data to the network structure. In essence, the most important feature of such a smart system is its ability to learn. Because it is more versatile and easier to plan, a learning system can better respond to new challenges and changes in processes. A data structure that can simulate a neuron is constructed for ANNs using programming abilities. This data structure is referred to as a "node." To train the network connecting these nodes, this structure employs an instructional technique. In this neural network or memory, each edge (or link between nodes) has a weight, and each node has two active states (on or off) and one inactive state (off or 0). Positive weights awaken the next dormant node, and negative weights block or deactivate the next connected node (if active). In the ANN design, the input bp for neural cell c comes from cell p preceding it. The weight of input bp in respect to cell c is provided by wpc, and the sum of the weights of the inputs after multiplication is given by a_c:

$$a_c = \sum w_{pc} \qquad \text{Eqn. 1.6}$$

A non-linear function Θ_c is applied to a_c. accordingly,b_c can be calculated as Eqn. 1.7:

$$b_c = \theta_c(\qquad \text{Eqn. 1.7}$$

The bcn is the output of c to n, and wcn is the weight of the bcn. W is the set of all the weights in the neural network. The neural network's output for input x and output y is Hw (x). It is critical to learn these weights in order to reduce the error values between y and hw (x). In other words, the goal is to lower the cost function Q (W), as shown in Eqn. 1.8.

$$Q(W) = \frac{1}{-}\sum_{i=1}^{n}(y_i - o_i)^2 \qquad \text{Eqn. 1.8}$$

In the current work, the epidemic was predicted using one of the MLP, a common form of ANN. MLP was trained using a dataset applicable to both cases. Inner neurons of 8, 12, and 16 were examined to achieve the best response during network training. The RMSE and correlation coefficient were used to evaluate outcomes in order to reduce the cost function's value.

EVALUATION CRITERIA

The correlation coefficient and the root mean square error (RMSE) were utilised in the evaluation. By comparing the goal and output values, these statistics generate a score as an assessment of the efficacy and precision of the suggested methods. The following two equations Eqn. 1.9 and Eqn. 1.10 depict the assessment criteria.

$$\text{Correlation coefficient} = \frac{N\sum(AP)-\sum(A)\sum(P)}{\sqrt{[N\sum A^2-(\sum A)^2][N\sum P^2-(\sum P)^2]}} \qquad \text{Eqn.1.9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\Sigma(A\cdot P)^2} \qquad \text{Eqn.1.10}$$

7. ALGORITHMS

Regression algorithms

Regression algorithms are members of the Supervised Machine Learning algorithm family, which is a subset of machine learning algorithms. One of the key characteristics of supervised learning algorithms is their ability to forecast the value of fresh data by modelling dependencies and interactions between the intended output and input variables. Regression algorithms forecast output values based on input attributes from the system's data. The standard process is that the algorithm creates a model based on the properties of training data and then uses the model to predict the value of fresh data.

The Simple Linear Regression model is a statistical tool that allows users to summarize and investigate correlations between two continuous (quantitative) variables. Linear regression is a linear model in which a linear connection is assumed between the input variables (x) and a single output variable (y). In this case, y may be determined using a linear combination of the input variables (x). A basic linear regression approach is used when there is just one input variable (x). Numerous linear regression is the approach used when there are multiple input variables.

Linear regression

Linear regression is one of the most basic and widely used Machine Learning methods. It is a statistical approach for performing predictive analysis. Linear regression forecasts continuous/real or quantitative variables such as sales, salary, age, product price, and so on.

The linear regression technique demonstrates a linear connection between a dependent (y) variable and one or more independent (y) variables, thus the name. Because linear regression demonstrates a linear connection, it determines how the value of the dependent variable is c Linear regression is used to investigate the linear connection between a dependent variable (Y) and one or more independent variables (X). This linearity in the connection between the dependent and independent variables is a model assumption.

Linear regression has a linear model parameter connection, and adding polynomial terms to the model effectively allows the model to find non-linear trends. Polynomial regression extends the linear model by adding new predictors, which are created by increasing each of the original predictors to a specific power. For example, a cubic regression employs three variables as predictors: X, X2, and X3. This strategy makes it easier to create a non-linear fit to data.

In most cases, the input characteristics for a predictive modelling job react unexpectedly and nonlinearly. A learning algorithm can detect and simulate these behaviours. Another approach is to create new features that expose these interactions and test whether or not they improve model performance.

Polynomial features

Polynomial features are formed by exponentiating existing features or variables. Consider this: if a dataset had only one input feature, X, a polynomial feature would be the addition of a new feature (column) with values derived by squaring the values in X, e.g. X^2. This technique may be performed for each variable in the dataset, resulting in a changed version of each feature.

Polynomial property Transformation is a sort of feature engineering in which new input features are created based on current features. The degree of the polynomial is used to limit the amount of characteristics that may be added. When a degree of 3 is employed, two additional variables are added for each input variable. A modest degree, such as 2 or 3, is usually utilised. When the value exceeds 3 or 4, the polynomial curve becomes too flexible and can take on some quite deformed forms.

Support vector regression

Support Vector Machine may also be utilised as a regression approach while retaining all of the algorithm's fundamental characteristics (maximal margin). With a few minor exceptions, the Support Vector Regression (SVR) utilises the same classification concepts as the SVM. For starters, because the output is a real number, it becomes extremely difficult to forecast the information at hand, which has an endless number of possibilities. In the case of regression, a margin of tolerance (epsilon) is set in order to approximate the SVM that the issue would have already requested. However, there is another aspect to consider, and that is that the algorithm is more difficult. The essential notion, however, is always the same: to minimise error by individualising the hyperplane to maximise the margin while keeping in mind that some mistake is acceptable.

Decision tree regression

One of the most widely used and practical ways to supervised learning is the Decision Tree. It may be used to tackle both Regression and Classification problems, with the latter being more practical.

It is a classifier with three types of nodes in a tree structure. The Root Node is the starting node that represents the full sample and may be subsequently divided into nodes. Interior Nodes represent data set characteristics, whereas branches represent decision rules. Finally, the outcome is represented by the Leaf Nodes. This method is quite beneficial for dealing with decision-making issues. A specific data point is traversed through the entire tree by answering True/False questions until it reaches the leaf node. The final prediction is the average of the dependent variable's value at that particular leaf node. The Tree is able to forecast a suitable value for the data point after several rounds.

8. Result

As seen that most of the writer used the regression model or neural network approach and platform used as MATLAB of Python for the prediction of covid-19 peak time

9. CONCLUSION

With data analytics and data mining, information and communication technologies aid in decision-making based on historical data. The amount of data accessible is massive, and extracting information and creating an attractive pattern from the accumulated data is a difficult undertaking. The peak time can be predicted using any of the four regression methods. As can be shown, the Decision Tree Regressor model is the superior of these four-regression models.

REFERENCES

Backer JA, Klinkenberg D, Wallinga J. "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan," China, January 2020, pp. 20-28.

Bohdan M.P. "Machine-Learning Models for Sales Time Series Forecast: MDPI," Vol. 04. 2019.

Cai, Zhuhua, et al. "A comparison of platforms for implementing and running very large scale machine learning algorithms," Proceedings of the 2014 ACM SIGMOD international conference on Management of data, 2014.

Covid, C. D. C., and Response Team. "Severe outcomes among patients with coronavirus disease 2019 (COVID-19) United States," MMWR Morb Mortal Wkly 2020, Vol. 69, no.12, pp. 343-346.

Deepak Painuli, Divya Mishra, Suyash Bhardwaj and Mayank Aggarwal, "Forecast and prediction of COVID-19 using machine learning," Data Science for COVID-19, 2021.

Dong, Ensheng, Hongru Du, and Lauren Gardner. "An interactive web-based dashboard to track COVID-19 in real time," The Lancet infectious diseases 2020, pp. 533-534.

Hyndman R.J., Koehler A.B., Snyder R.D., Grose S. "A state space framework for automatic forecasting using exponential smoothing methods," Int. J. Forecast 2002, Vol.18, Issue 03, pp. 439-454.

Jain, Abhishek. "Predictive Models for Ebola using Machine Learning Algorithm," Florida Atlantic University, 2017.

Kalipe G., Gautham V., Behera R.K. "Predicting Malarial Outbreak Using Machine Learning and Deep Learning Approach: A Review and analysis," International Conference on Information Technology (ICIT) 2018, pp. 33-38.

Kane M.J., Price N., Scotch M., et al. "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks," BMC Bioinf 2014, Vol. 15, Issue 01, pp. 1-9.

Kenta Terada, Yutaka Watanobe, "Code Completion for Programming Education based on Recurrent Neural Network," Computational Intelligence and Applications (IWCIA) IEEE 11th International Workshop on 2019, pp. 109-114.

Liu, Chenghao, et al. "Online arima algorithms for time series prediction," 2016.

Makridakis S., Wakefield A., Kirkham R. "Predicting medical risks and appreciating uncertainty," Foresight Int. J. Appl. Forecast 2019, Vol. 52, pp. 28-35.

Punn N.S., Sonbhadra S.K., Agarwal S. "COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms," 2020.

Santosh K.C. "AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data," 2020, Vol. 44, Issue 93.

Sharmila S., Dharuman C., Venkatesan P. "Disease classification using machine learning algorithms-a comparative study," Int. J. Pure Appl. Math. 2017, Vol. 114, Issue 06, pp.1-10.

Shirsath S.S. "Disease prediction using machine learning over big data," Int. J. Innov. Res. Sci. 2018, vol. 07, Issue 06, pp. 6752-6757.

Siami-Namini, Sima, and Akbar SiamiNamin. "Forecasting economics and financial time series: ARIMA vs. LSTM," 2018.

Siami-Namini, Sima, NedaTavakoli, and Akbar SiamiNamin. "A comparison of ARIMA and LSTM in forecasting time series," 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018.

Sreeja S., Bhavya L., Swamynath S., Dhanuja R. "Chest x-ray pneumonia prediction using machine learning algorithms," Int. J. Res. Appl. Sci. Eng. Technol. 2019, Vol. 07, Issue 04, pp. 3227-3230.

Tiwari U.K. "Role of machine learning to predict the outbreak of covid-19 in India," J. Xi'an Univ. Archit. Technology 2020, Vol.12, Issue 4, pp. 2663-2669.

Tuli S., Tuli S., Tuli R., Gill S.S. "Predicting the Growth and Trend of COVID-19 Pandemic Using Machine Learning and Cloud Computing," Internet of Things 2020.

Xingjian, S. H. I., et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Advances in neural information processing systems 2015.