International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

# ANTICIPATION OF COVID -19 VIA ECOLOGICAL VARIABLES THROUGH A HYPOTHETICAL APPROACH

**Sri Rakesh Kancharla[1]**
*Department of Computer Science & Engineering*
*Sasi Institute of Technology and Engineering*
*Tadepalligudem-534301, India.*

**Jaya Priyadarshini Bayye[2]**
Department of Computer Science & Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem-534301, India.

**Madhu Vasanth Bendukuri[3]**
*Department of Computer Science & Engineering*
*Sasi Institute of Technology and Engineering*
*Tadepalligudem-534301, India.*

**Bala Venkata Lakshmi Narayana Kanisetti[4]**
*Department of Computer Science & Engineering*
*Sasi Institute of Technology & Engineering*
*Tadepalligudem-534301, India.*

**Kiran Kumar Kankipati[5]**
*Department of Computer Science & Engineering*
*Sasi Institute of Technology & Engineering*
*Tadepalligudem-534301, India.*

---

**ABSTRACT:** Numerous outbreak prediction models are being used by our officials around the world to implement appropriate control actions. Authorities have paid more attention to simple epidemiological and statistical models among the standard methods for COVID-19 global pandemic prediction. Due to a high level of uncertainty and lack of required data, standard models have shown low accuracy for long-term prediction. Although the study includes several attempts to address this issue, the essential generalization and robustness abilities of existing models need to be improved yet.

The purpose of this paper is to construct a simple average aggregated machine learning approach to forecast the number, size, and length of COVID cases extent and wind-up period throughout India. Covid-19 sickness can be accurately anticipated, according to the findings. According to the prognosis, COVID-19 could be transmitted by water and air ecological factors, necessitating the use of preventative measures. Putting in place the preventive measures can effectively manage the spread of COVID-19 and also the demises rate can be reduced eventually in India.

**Keywords**—Support Vector Regression, Linear Regression, Artificial Neural Network, COVID-19, Epidemic, Machine Learning,

**INTRODUCTION:**

Coronaviruses are a diverse group of pathogens, only a subset of which causes disease in humans, with the remainder causing disease in animals and nature. Animal coronaviruses will occasionally transmit to individuals and mostly transmit to human beings. In recent years, zoonotic coronaviruses have emerged in the Middle East, causing humanoid outbreaks such as coronavirus disease 2019(COVID-19), Severe Acute Respiratory Syndrome(SARS). The human disease occurs mostly as lung infection. The disease's clinical spectrum ranges from no symptoms to minor breathing problems to serious health concerns, the latter of which is growing increasingly common.

Transmission of such kind of disease is a dynamic mechanism of transmission that happens within the crowd. Frameworks for this method can be built to potentially examine and evaluate disease propagation mechanisms, allowing for accurate prediction of future infectious disease effects. As a result, the study and evaluation of predictive models for infectious diseases has emerged as a hot topic in science in order to track and limit the impact of infectious diseases. The coronavirus disease(COVID-19) has been reported to have infected more than 2 million people, with more than 132,000 confirmed deaths worldwide. The recent global COVID-19 pandemic has exhibited a nonlinear and complex nature making it difficult to predict the future intensity of disease.

The study demonstrates a simple average aggregated scheme, with the aggregated system established by aggregating three regression methods: Support Vector Regression, Linear Regression and Artificial Neural Network. The figures of COVID-19 instances were used as variables in the construction of an aggregated approach. The dataset was gathered and collected from

Statistica.com between the time period January to April 2021 gathered monthly. The study predicted the values from the previous COVID-19 incidents and values of environmental variables such as water and air.

**RELATED WORK**

The authentic datasets of COVID-19 have been collected from https://www.mygov.in/ and https://www.pharmaceuticaltechnology.com. The dataset is publicly available on cases from India from the first case index on January 30 2021. The datasets gathered were in a monthly form that is January 2021 to April 2021. Table 1 gives the scenario of COVID-19 incidents in India from January 2021 to April 2021. As at 25th April 2021 COVID-19 dataset includes accumulated 408,658 total samples; confirmed cases of 24,942; recovered cases of 5,209; 779 death cases and 1 migration.

Table 1: Datasets for COVID-19 Cases in India

| Total samples | confirmed | Active cases | Recovered | deaths | Migrated |
|---|---|---|---|---|---|
| 408,658 | 24,942 | 18,953 | 5,209 | 779 | 1 |

**METHODOLOGY:**

**Machine Learning Methods:**

There are several methods used to perform machine learning tasks. Machine learning approaches need some certain algorithmic approaches. According to Dataquest, 2020, there are three types of machine learning algorithms and they are:

- Supervised learning algorithms deals with classification, regression.
- Unsupervised learning algorithms such as Association, Clustering, Dimensional reduction.
- Reinforcement learning.

**Artificial Neural Network (ANN)**

Artificial neural networks (ANN) are used to model nonlinear situations and predict output values for given input parameters based on training values. Multi-layer perceptions (MLPs) are the most commonly used ANNs in the estimate problem, and they use a single tiered feed-forward network. This method is categorized by a system of three layers. The nodes in several tiers are identified as altering fundamentals. The outcome of the method is calculated employing the subsequent mathematical expression given here

$$y_t=\alpha_0 \quad +\sum_{j=1}^{q} \alpha_j \, g(\beta_{oj} \quad +\sum_{i=0}^{p} \beta_{ij} \quad y_{t-i}) \quad +\varepsilon_{t,\forall_t} \quad (1)$$

At this point $y_{t-i}$ ( where i = 1, 2, ……p) are the p entered and $y_t$ is the result. The digits p, q are the figures entered and concealed nodes. $\alpha_j$(j= 0, 1, 2, ……q) and $\beta_{ij}$(i = 0, 1, 2, ……p); j = 0, 1, 2,……q) are the formulation weightiness and $\epsilon_t$is arbitrary tremor; $\alpha_0$and $\beta_{oj}$ are the prejudice

terms. Generally, the logistics sigmoid function g($\alpha$)= $\frac{1}{-1+e^{-x}}$ is employed as non-linear stimulation function.

**Linear Regression (LR):**

Linear regression analysis is a statistical technique for predicting the value of one variable based on the value of another. The dependent variable is the variable you want to forecast. The independent variable is the one you're using to forecast the value of the other variable.

Linear regression is a statistical approach for illustrating the relationship between scalar or dependent variables and other self-determining factors. The circumstance of the descriptive variables is termed as a simple linear regression while in the circumstance of greater than single descriptive variables the procedure is referred to as multiple linear regressions. The mathematical formulation of the linear regression is as specified beneath.

$$y = a + b\,(x) \quad (2)$$

Where x and y are two variables of the regression line, b= Grade of the contour and a=Y interrupt of the contour.

**Support Vector Machines (SVM):**

Support vector machines (SVMs) are a class of supervised learning methods for classification, regression, and detection of outliers. The following are some of the benefits of support vector machines: In high-dimensional spaces, it works well. When the number of dimensions exceeds the number of samples, the method is still effective.

Support vector machines are a sort of supervised learning technique that may be used to handle problems involving classification and regression. When applied to dual categorization problems, the main impression of SVM is to identify an established threshold level that perfectly splits the two presumed groups of tutoring models. In

circumstances where data points are not linearly divisible, a lenient border manic level classifier is created as shown below

X∈ $R^n$:$w^T$x +b=0,where w∈ $R^n$ ,b∈ $R$ (3)

Minimize: K(w,∈ )=1/2 $||w||^2$ + c($\sum_{i=1}^{N} \in_i$ (4)

subject to :$y_i(w^T x_i$+b) >=1-$\in_i$ $\forall_i$ =1,2,...N $\in_i$ >=0 (5)

**Support Vector Machine Regression (SVMR):**

To employ SVMR, as in the categorization difficulty this belongs to a few clusters, for instance, A1 and A2 in the circumstance of regression and support vector machine at this point is the actual count and additional variables are equivalent as for the categorization glitches. Regression techniques are one of the prevailing methods for the prediction of specific datasets. Here, the authors formulated a simple mean aggregated method by combining three popular regression models and predicted the sum of COVID-19 in India. Authors such as and have employed the regression method predominantly to communicable datasets and as well as hypotheses an ensemble model typically with three estimation approaches.

**Description of Postulated Aggregated Method**

There is a key challenge in the study of prediction predominantly in the relation of predicting an occurrence of a specific illness as it had been stated earlier. There are several techniques obtainable in previous research work for the study of forecasting and yet there exists one problem or the other with the performance of the separate technique. To conquer this restraint, an aggregated method was postulated. In the study, a method was designated as let X be the group of self-determining variables where (i =1, 2, 3, ...). The actual data group of a sequence is demarcated as X= [x1, x2, ...]raise to power T being its prediction attained from the $i^t$ scheme.

**Algorithm**

1. Employing the independent variables, this revert them to discover the estimation by

employing SVR, LR, and ANN.

2.Derive a mean in each case by using entire three methods.

3.Calculate the predicted figures by the weighted aggregated method

4. Equate the distinct prediction error with a simple mean aggregated method.

**RESULT**

This section describes the objective of finding the equivalence between the figures of COVID-19 cases with environmental variables such as water (sewage overflow) and air (air streams). In this study, ecological variables play an important role in the spread of COVID-19 diseases. To accomplish this objective, a statistical P-value test is applied to determine the equivalence between the figure of COVID-19 cases. After examination by the P -value test. It was concluded that water is a prominent ecological variable for the incidence of COVID-19 incident in India. Many authors have found the equivalence between infectious diseases and ecological variables by P-value test. The application of the P-value test accomplishes that ecological variables like water and air had a positive equivalence with the occurrence of several cases for the period since this disease started. The statistical implication was measured p<0.05. The null hypothesis is found to be false in this investigation. The equivalence of COVID-19 cases with water and air is substantial table 2.

Table 2. Equivalent Values of COVID-19 with Ecological Variables

| Months | P value (air) | P value (water) | Coefficient of equivalence(air) | Coefficient of equivalence (water) |
|---|---|---|---|---|
| January | 0.064 | 0.013 | 0.67 | 0.75 |
| February | 0.078 | 0.013 | 0.58 | 0.63 |
| March | 0.082 | 0.019 | 0.55 | 0.86 |
| April | 0.093 | 0.019 | 0.55 | 0.86 |

Tables 3-6 reveal the prediction precision measured in terms of root mean square error, mean scaled error, and mean absolute error.

- RMSE: Root Mean Square Error
- MSE: Mean Scaled Error
- MAPE: Mean Absolute Percentage Error

Table 3. Support Vector Regression

| Datasets | RMSE | MSE | MAPE |
|---|---|---|---|
| COVID-19 | 265.7468934 | 67480.4537 | 123.5689343 |

Table 4. Linear Regression

| Datasets | RMSE | MSE | MAPE |
|---|---|---|---|
| COVID-19 | 259.46783 | 56777.2316 | 164.674563 |

Table 5. Neural Network

| Datasets | RMSE | MSE | MAPE |
|---|---|---|---|
| COVID - 19 | 246.7356982 | 64321.37210 | 161.467839 |

Table 6. Postulated Aggregated Method

| Datasets | RMSE | MSE | MAPE |
|---|---|---|---|
| COVID - 19 | 12.245913 | 167.456832 | 7.678 |

**DISCUSSIONS**

The core intention of this investigation is to formulate a simple mean aggregated method for the estimation of COVID-19 disease in India. In the second stage of formulating the method, the study intention is to decrease the prediction errors of COVID-19 in India and these errors include RMSE, MAE, and MAPE. In this study, the formulation of aggregated methods illustrates a substantial enhancement in the prediction of the COVID-19 disease in India.

If the datasets are reliable and no subsequent outbreaks occur, the above methods predicted that the COVID-19 outbreak in India could end by May. The aggregated methods permit entering the interferences information as well as examining the influence of interferences on the extent of the disease outburst and the ending period of the COVID-19. the aggregated methods (SVR, NN, and LR) provide real-time predicting instruments used for shaping and tracking COVID-19 disease in India.

**CONCLUSION**

The aggregated methods (SVR, NN, and LR) provide real-time predicting instruments used for shaping and tracking COVID-19 disease in India, reckoning the COVID-19 disease, obtaining COVID-19 disease affect, predicting the extent of the pandemic together with supporting government and health staffs to constitute strategy and competent verdicts towards the eradication of the COVID-19 diseases.

The integration of the prediction from different methods substantially decreases prediction errors and also makes available advanced precision. A few decades back, several researchers, studies have suggested several statistical methods. The study postulated a simple-mean aggregated method for the prediction of COVID-19 disease in India.

**FUTURE ENHANCEMENT:**

In the future the competence of the postulated method could be as well as be investigated and some other regression models or algorithms can be used.

**REFERENCES:**

1 .BMJBest Practice, 2021. Overview of corona viruses.

https://bestpractice.bmj.com/topics/en-us/3000165

2. Smith JD, MacDougall CC, Johnstone J, Copes RA, Schwartz B, Garber GE.

Effectiveness of N95 respirators versus surgical masks in protecting health care workers from acute respiratory infection: a systematic review and meta-analysis.

3. CDC. 2019 Novel Coronavirus, Wuhan, China: 2019 Novel Coronavirus (2019-nCoV) in the U.S. Centers for Disease Control and Prevention (CDC). Available at

https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html March 18, 2021; Accessed: March 19, 2021.

4.Wuhan Virus: What Clinicians Need to Know. A Medical News Available at https://www.medscape.com/viewarticle/924268 January 27, 2021; Accessed: January 27, 2021.

5. Severe results Among Patients with Coronavirus Disease 2019 (COVID-19) —

United States, February 12–March 16, 2021. Dataquest, 2020. Retrieved on 24th April 2020 from https://community.dataquest.io/c/social/covid19

6. Adhikari, R. & Agrawal, R.K. (2012). Forecasting strong seasonal time series with artificial neural networks. Journal of Scientific and Industrial Research 10: 657-666.

7. Science, C. (2003). Dimensionality Reduction for Indexing Time Series Based on the Minimum Distance Work 711: 697-711.

8. Yang, S.H., Liang, C.K. & Hsieh, C.Y. (2005). Watermarking MPEG-4 2D mesh animation with time-series analysis. Journal of Information Science and Engineering 21: 341-359.

9. Barrow, D.K., Crone, S.F. & Kourentzes, N. (2010). An evaluation of neural network ensembles and model selection for time series prediction. The 2010 International Joint Conference on Neural Networks (IJCNN), 1–8.

10. Shanthi, S. & Kumar, D. (2012). Prediction of blood glucose concentration ahead of time with feature based neural network. Malaysian Journal of Computer Science 25: 136-148.

11. Shaikh, M.S. & Dote, Y. (1999). An Application of Linear Regression Method and Fir Network for Fault 12: 57-63.