JOURNAL OF MULTIDIMENSIONAL RESEARCH & REVIEW

Website: http://www.jmrr.org **Vol.2, Iss.2 Pages:42-49**

PREDICTION OF COVID-19 PANDEMIC USING LINEAR REGRESSION ALGORITHM

[1]M S Simrin Fathima, [2]V Srividhya

1,2Dept of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamilnadu, India - 641043

Article Received: April 2021 Published: July 2021

Abstract

The globe is dealing with the Corona virus (COVID-19) pandemic, which began in the Gregorian calendar month of 2019 in the Chinese metropolis of Shanghai.As of March 9, 2020, the world's active cases totaled 11.71 lakh, with a cumulative death count of 25.99 lakh.The COVID-19 pandemic has an impact on the economy and changed people's views on life, affecting everyone's emotional, physical, and financial well-being.The first case of the COVID-19 pandemic in India was registered on January 30, 2020, and it emerged in China. As a result, predicting the long-term pattern of illness in a number of countries is often useful for handling natural disasters.Many information-driven works have been completed for the prediction of COVID-19 cases, and these works make use of past knowledge options to predict the future.The various types of COVID-19 connected problems have been addressed in this paper using a machine learning (ML) Linear regression model. The dataset has been fitted with a linear regression model to predict the full range of confirmed cases, recovered cases, deceased cases in India, which aids in predicting and stating the near future.

Keywords: Pandemic, Covid-19, Machine learning (ML), linear regression

I INTRODUCTION

The pandemic Corona Virus Disease (COVID-19) has spread across the globe, and every nation is doing everything possible to contain it.Various pathogens cause infectious diseases, which can be spread from person to person, animal to animal, or person to animal. They will be transmitted in a variety of ways, so the transmission speed will be high. Some nations, such as Taiwan, South Korea, Vietnam, New Zealand, Germany, France, and others, have successfully controlled the outbreak, but many others, such as India, the United States, and Brazil, are still fighting the pandemic. After a lockdown that lasted until May 31, 2020, the government of India started a phased unlocking of the country because it was difficult for a developing country like India to endure the financial losses of a prolonged lockdown. Throughout these unlock processes, the Government of India has issued guidance to the general public as well as various organizations on how to work.

Various types of COVID-19-related problems, such as the total number of confirmed cases, the number of cured cases, and the total number of deceased cases in India, have been discussed in this paper with the help of a machine learning model. Furthermore, the proposed scheme aims to estimate the total number of confirmed cases, cured cases, and deceased cases. The propagation of infection has been addressed, followed by a discussion of the linear regression model used in the proposed analysis. Various types of information of different states, such as data for confirmed cases, cured cases, and death cases, are collected in the proposed work. By analyzing COVID-19 results, this paper aims to predict confirmed cases, cured cases, and deceased cases in India.

There are five sections of the paper. The definition of introduction is discussed in Section I. The related work of prediction is presented in Section II. Section III explains the methodology flow diagram. The results and discussion are listed in Section IV. The last section of the paper is the conclusion and future study work directions are discussed in Section V.

II RELATED WORKS

According to Li Q et al., a new Coronavirus (COVID-19) pandemic began in Wuhan, China in December 2019 and has since spread throughout the world. COVID-19 has a variety of effects on public health, according to the authors, since there is no early detection and a person infected with COVID-19 can infect many people in a short period of time [1].

The clinical research by Liu K et al's. on COVID-19 patients of various ages compared clinical data and concluded that the morality of elderly patients is higher than that of teen and middle-aged patients. They also observed that in terms of PSI score IV and V, older patients have higher proportions than teen and middle-aged patients [2].

Cobb JS et al. discovered that COVID-19 has an impact on the accumulated growth pattern of shelter in place (SIP) orders in the United States. They analyzed the cumulative growth pattern of COVID-19 using statistical analysis and a random forest method. They came to the conclusion that the COVID-19 growth pattern has been effectively reduced as a result of the US government's SIP order, with countries with higher densities benefiting more [3].

To forecast the COVID-19 events, S.J. Fong et al. describe a hybrid approach for time series analysis. The hybrid model is a mix of composite Monte-Carlo (CMC) and GROOM methodology [4].

Using a deep learning neural network and a set of fuzzy rules, the CMC result was improved. The proposed model has been experimentally shown to have better results for forecasting possibilities on extremely unpredictable and small data sets [5].

Tanujit Chakraborty et al. used the COVID-19 data collection from France, Canada, South Korea, India, and the United Kingdom to conduct risk analysis and real-time forecasting. There were two stages to the research. For real-time forecasting, the first step covers a hybrid ARIMA [6] and Wavelet-based forecasting model [7]. They used the regression tree model to find the dependency parameter for mor in the second step [8].

Wei Yu et al. found that the SVM is the best classifier for evaluating ambiguous data. They conducted an experiment on a data collection of diabetes and pre-diabetes in the US population [9].

III METHODOLGY

This work comprises of four phases. The first phase of the work is to import the dataset. The second phase of the work is exploration and visualization. Applying the machine learning algorithm, linear regression is the third phase of the work. The fourth phase of work is to predict the result. The following fig.1 shows the overall methodology diagram.

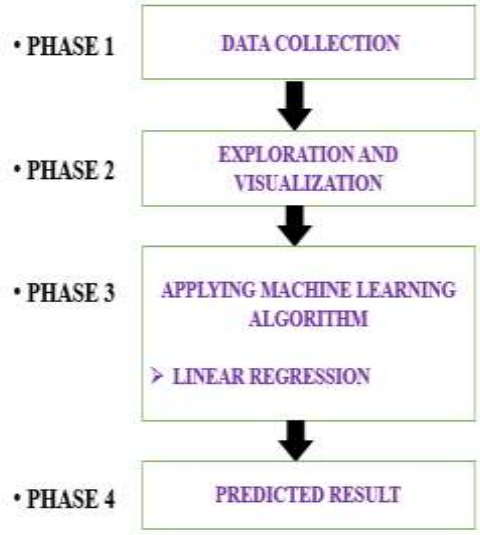

Fig 1 Methodology diagram.

(i) DATA COLLECTION

The Covid-19 prediction dataset have been extracted from repository. The dataset consist of 5 attributes whichcovers Cured cases, deceased cases, confirmedcases from date April 1, 2020 to May 29, 2020 and 100 instances.

(ii) EXPLORATION ANDVISUALIZATION

The key aspect of data analysis is data discovery. It is difficult to construct accurate predictive models without first investing a considerable amount of time understanding the data and its designs. Data discovery takes up a significant amount of time in the data science creation process, which also includes data cleaning and preprocessing. The graphical representation of data is known as data visualization. Visualizations were seen.

(iii) MACHINE LEARNING ALGORITHM

Linear Regression could also be a supervised machine learning algorithm with a continuous and unbroken slope as the expected output. It's used to predict values over an infinite range. It's an instantaneous demonstration of the interaction between a variable and a minimum of one independent variable due to impact. This prediction model utilizes linear regression, which is

a subset of regression model. Linear regression (LR) is a statistical technique that demonstrates the relation between a variable and a minimum of one experimental variable in real time.Linear Regression was the first and most detailed statistical approach to be analyzed in detail and widely used in practical applications. By fitting a straight condition dependent knowledge to two variables, linear regression identifies the relationship between both. For the train and test results, a linear regression model was first fitted. The expected linear value is increasingly being represented in array values.

(iv) PREDICTED RESULT

Finally, using a linear regression model, the predicted accuracy of each case of data in the current dataset of confirmed cases, recovered cases, and death cases is seen. The prediction was made using the python library sklearn.metrics and a formula.

IV RESULTS AND DISCUSSION

The linear regression covid-19 data has a continuous range and with maximum predicted values. The prediction is done on the current dataset of covid19 in India by using machine learning model.

The Fig. 2 shows the accuracy of the prediction of confirmed cases in current data of covid-19 in India using a linear regression model, with a prediction accuracy of 99.2%.

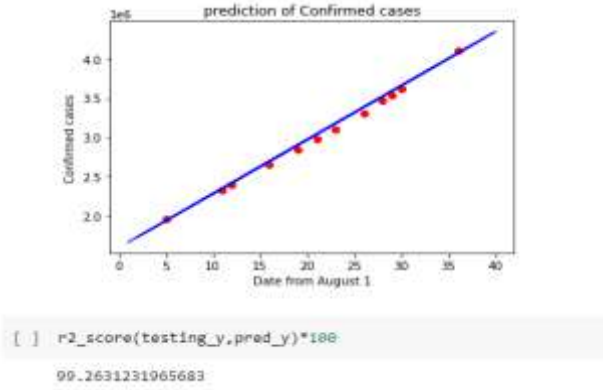

Fig 2 prediction of confirmed cases of covid-19 in India.

Fig. 3 shows the accuracy of the prediction of recovered cases in current data of covid-19 in India using a linear regression model, with a prediction accuracy of 99.6%.

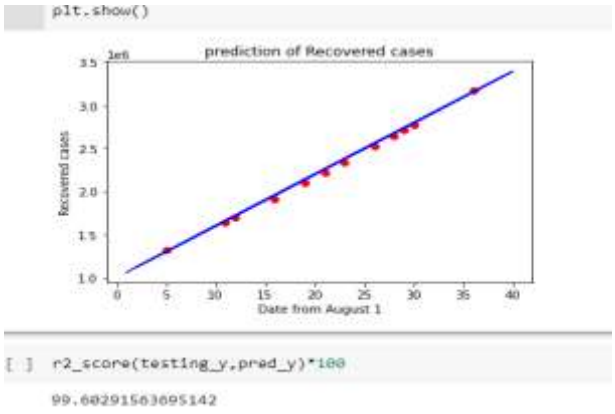

Fig 3 prediction of recovered cases of covid-19 in India.

Fig. 4 shows the accuracy of the prediction of deceased cases in current data of covid-19 in India using a linear regression model, with a prediction accuracy of 99.9%.

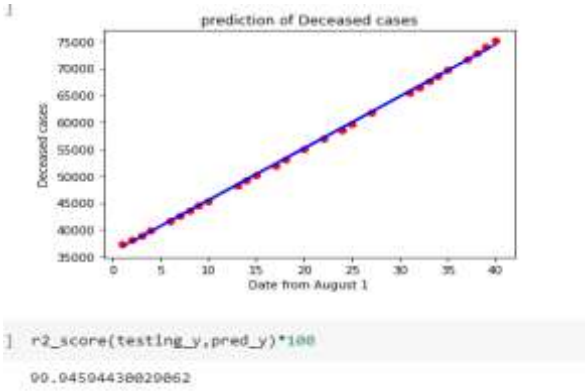

Fig 4 prediction of deceased cases of covid-19 in India.

V CONCLUSION

COVID-19 outbreaks have impacted negatively on many countries. COVID-19, on the other hand, has a recovery rate of over 88 percent in India. In this paper, Linear Regression is used to predict COVID-19 confirmed cases, recovered cases, and death cases using Machine

Learning models.As a result, the proposed scheme correctly predicts the result in all cases.As a conclusion, it is possible to conclude that our model could be applied to other COVID-19 parameters in any state or region.In the future, it will concentrate on developing various Machine Learning and Deep Learning-based models to improve efficiency in the fight against COVID-19 and other pandemics that may arise. In the direction, the goal is to solve the problem of finding an optimum partitioning point so that the error is minimized and it can focus on more Machine Learning models to better predict COVID-19 development.

REFERENCES

[1] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine (2020).

[2] K. Liu, Y. Chen, R. Lin, K. Han, Clinical features of covid-19 in elderly patients: A comparison with young and middle-aged patients, Journal of Infection (2020).

[3] J. Cobb, M. Seale, Examining the effect of social distancing on the compound growth rate of sars-cov-2 at the county level (united states) using statistical analyses and a random forest machine learning model, Public Health (2020).

[4] S. J. Fong, G. Li, N. Dey, R. G. Crespo, E. Herrera-Viedma, Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction, Applied Soft Computing (2020) 106282.

[5] A. Mubayi, C. K. Zaleta, M. Martcheva, C. Castillo-Ch´avez, A costbased comparison of quarantine strategies for new emerging diseases, Mathematical Biosciences & Engineering 7 (3) (2010) 687.

[6] C. Chatfield, H. Xing, The analysis of time series: an introduction with R, CRC press, 2019.

[7] D. B. Percival, A. T. Walden, Wavelet methods for time series analysis, Vol. 4, Cambridge university press, 2000.

[8] T. Chakraborty, I. Ghosh, Real-time forecasts and risk assessment of novel coronavirus (covid-19) cases: A data-driven analysis, Chaos, Solitons & Fractals (2020) 109850.

[9] W. Yu, T. Liu, R. Valdez, M. Gwinn, M. J. Khoury, Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes, BMC medical informatics and decision making 10 (1) (2010) 16.

[10]https://www.visualcapitalist.com/history-of-pandemics-deadliest/

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

[12] Mitchell, Tom M. "Machine learning and data mining." Communications of the ACM 42.11 (1999): 3036.

[13] Klaus Greff, Rupesh K. Srivastava, et.al " LSTM: A Search Space Odyssey " Transactions On Neural Networks And Learning Systems, arXiv:1503.04069v2 [cs.NE] 4 Oct 2017

[14] Yadav RS (2020) Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. Int J Inf Technol 1–10

[15] COVID-19 in India, dataset on novel corona virus disease in India.https://github.com/anshukumari2001/COVID19-CASES-PREDICTION-INDIA

[16] Verma Gaurav, Kumar Rahul, Khare Vijay (2019) Regression based FPGA power estimation tool (FPETool) for embedded multiplier block. Int J Inf Technol 11(4):795–798

[17] Gaurav Pandeya , Poonam Chaudharya , Rajan Guptab , Saibal Palc, SEIR and Regression Model based COVID-19 outbreak predictions in India

Cite this article:

M S Simrin Fathima, V Srividhya, "Prediction of Covid-19 Pandemic Using Linear Regression Algorithm", Journal of Multidimensional Research and Review (JMRR), Vol.2, Iss.2, pp.42-49, 2021.