IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907067 |

Predictive Analysis on Covid-19 by Using Forecasting Time Series

Dr.B. Sai Jyothi[1], P. Niharika[2], R. Sai Chaitanya[3], P.Darani[4], Sk.Farhina Tabassum[5]

Professor, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[1]

UG Student, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[2,3,4,5]

ABSTRACT: The role of mathematical modelling in predicting spread of an epidemic is of vital importance. The main purpose of present study is to develop and apply a computational tool for predicting evolution of different epidemiological variables for COVID-19 in India.It is the utmost importance to identify the future infected cases and the spread of virus rate for advance preparation in the healthcare services to avoid deaths.Accurately forecasting spread of COVID-19 is an analytical problem.So, we use day level information of COVID-19 spread for cumulative cases from whole world and 10 mostly affected countries. The most essential part is to minimize the spread of the virus by monitoring, tracking, and estimating the outbreak. To predict the spread of COVID-19 we made use of the Machine learning techniques like ML Prediction and Forecasting with Time Series Analysis with the available COVID-19 data set.In this paper we are predicting the COVID-19 confirmed, death, recovery cases and mortality rate of different countries based on the available data set. In this paper the algorithms[7] used are linear regression and SVM methods to predict the spread of COVID-19.The result is in the form of time series where one can able to know the spread of COVID-19 cases in different countries across the world.

KEYWORDS*:* COVID-19, Prediction, Machine Learning, Linear Regression, SVM

I. INTRODUCTION

Our paper COVID-19 Time Series Prediction is all about predicting the spread and containment of COVID-19. Coronavirus disease is a transmissible disease occurred by severe acute respiratory syndrome. Now a days forecasting the spread of COVID-19 is a challenge of utmost importance. The approach used to this problem is Linear/rectilinear regression model and Support vector regression and random forest models. So, using these methods one can know the spread of infectious diseases like covid-19 and also other parameters like deaths, active cases, recovery can be known. In this task, the Support Vector Regression (SVR) model can be used to solve the various sorts of COVID-19 related challenges. Random forests are used as it is a neural network method for classification, regression and for other works/performances. The proposed method will be fitted into the dataset containing the total number of COVID19 positive cases, and the number of recoveries and other parameters for different countries. These tasks can help a country/region to know the spread of the virus, facilitate/aware people, start mitigations. It'll also help that region/country to be prepared for what will happen within the future, which can help in saving lives.

II. LITERATURE SURVEY

2.1. Covid-19 Future Forecasting using Supervised Machine Learning Models

Machine learning (ML) helps to preoperative outcomes to improve the decision making on the future course of actions based forecasting mechanisms have proved their significance anticipation. The forecasting of the number of upcoming patients affected by COVID-19 is done by ML models. Standard forecasting models like least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), linear regression (LR), and exponential smoothing (ES) have been used to forecast the threatening factors of COVID-19. The ES performs best in the current forecasting domain is from the results of the study prove that given the nature and size of the dataset. LR and LASSO also perform well for forecasting to some extent to predict death rate and confirm cases. [1]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907067 |

2.2. Analyzing Recovery from Pandemics by Learning Theory:

The Case Of Covid-19 Predicting the recovery time from infectious diseases outbreaks is made in this study. The approach of theory of learning from errors is used here, by imposing countermeasures such as medical treatment, isolation, social distancing etc. specifically to adapt the control of the virus spread by reducing infection rates when these are effective, the infection rate, after reaching a peak, declines following what is the Universal Recovery Curve. The risk of an infectious disease is less when the infection rate slows down, called 'flattening the curve''. Once the rate peaks the rate, then, should decrease due to successful countermeasures. [2]

2.3. Data Mining and Analysis of Scientific Research Data Records on Covid-19

Mortality, Immunity, and Vaccine Development - In the First Wave of the Covid-19 Pandemic The Covid-19 mortality data, immunity details, and vaccine are obtained through data mining of scientific literature records from the Web of Science Core Collection, using. The analysis is compared with records on all Covid-19 research topics combined and the individual records are analyzed in isolation. Including the Web of Science and R Studio's Bibliometrix package, data mining tool the data records are analyzed with commutable statistical methods. From historical analysis of scientific data records on viruses, pandemics and mortality, it is identified that Chinese universities have not been leading on these topics historically. We identified a few clusters, containing references to exercise, inflammation, smoking, obesity and many additional factors. [3]

2.4. A machine learning forecasting model for COVID-19 pandemic in India

Linear regression, Multilayer perceptron and Vector auto regression are used for the prediction about COVID-19 cases. The data is collected from Kaggle and the regression models are applied to predict the future. Here the forecasting is done only on the COVID- 19 cases in India. By seeing the predicted values and matching with cases from John Hopkins University11 data we can conclude that the MLP method is giving good prediction results than that of the LR and VAR method using WEKA and Orange. The correctness of the model can be increased by introducing some more attributes.[9]

2.5. COVID-19 Outbreak Prediction with Machine

A comparative analysis of Machine learning and soft computing models to predict the covid-19 outbreak. The two Machine learning models are Multilayer perceptron (MLP) and Adaptive neuro fuzzy inference system (ANFIS). Data for 5 countries including Italy, Germany, Iran, USA, and China on total cases over 30 days is collected for analysis and predicted the future[10]

III. PROPOSED SYSTEM

This paper proposes system to predict the spread of COVID-19 cases for different countries using machine learning techniques like machine learning prediction and forecasting with Time series[8] analysis with available COVID-19 dataset. For the future prediction process, three data sets are taken which contains confirmed, recovered and death cases. The data which is available is used for training the system to predict the future cases. Mortality rate for future cases will be displayed for the taken forecast dates using linear regression.By using Linear Regression and Support vector regression the graph is plotted for number of corona virus cases over time. Then produces the result in the form of Time series where one can able to know the spread of COVID-19 cases in different countries across the world. It is used to evaluate the regression problems accuracy by using mean square error(MSE) and mean absolute error(MAE).

3.1. PROCESS

Prediction and Analysis of covid-19 cases by monitoring, Tracking and Estimating the Outbreak is the main aim of the project. In this paper, the first step is to extract the datasets from Kaggle[4] website. The extraction of data done by using the pandas library. Linear regression and support vector regression are used for prediction and the representation of data in the form of graphs.Data Visualization and graph plotting are done by using the matplotlib and seaborn libraries. The graph is plotted between number of cases and time(number of days). Here the start date and number of days are taken as input and performs prediction process on it. The result is shown between confirmed and prediction.Random Forest[6] is capable of performing both Classification and Regression tasks.The graph is plotted between the confirmed

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907067 |

and predicted cases by using Linear Regression and SVR[5]. So, it proves that SVR is better for prediction and it manages to fit the best line within a threshold of values.The output is in the form of time series by using Datetime library.

3.2.RESULTS

Fig.1 shows the dates which are to be predicted**.**
The prediction result for future cases will be given for the dates(future_forecast_dates).
The future_forecast_dates are based on starting date and number of days.

```
future_forecast_dates
```

```
'12/07/2020',
'13/07/2020',
'14/07/2020',
'15/07/2020',
'16/07/2020',
'17/07/2020',
'18/07/2020',
'19/07/2020',
'20/07/2020',
'21/07/2020',
'22/07/2020',
'23/07/2020',
'24/07/2020',
'25/07/2020',
'26/07/2020',
'27/07/2020',
```

Fig:1.Future Forecasting dates

Fig 2 shows the number of future cases for fig:1:

```
svm_pred
```

```
array([  12088.66869538,   12089.42707294,   12014.7357159 ,
         12029.14488964,   12057.20485954,   12103.46589105,
         12172.47824948,   12268.79220826,   12396.95880869,
         12561.52594059,   12767.04626076,   13018.06923478,
         13319.14512818,   13674.82420628,   14089.65673439,
         14568.19297783,   15114.98320191,   15734.57767289,
         16431.52665783,   17210.38041489,   18075.68921845,
         19032.08332825,   20083.87301056,   21235.84853547,
         22492.48015778,   23858.31815522,   25337.9127826 ,
         26935.81431   ,   28656.57380751,   30504.73912995,
         32484.86295582,   34601.49473755,   36859.18474762,
         39262.48325531,   41815.94851545,   44524.10688573,
         47391.53238862,   50422.76749604,   53622.36245148,
         56994.86748688,   60544.83287995,   64276.88886261,
         68195.34575837,   72304.99376862,   76610.38321686,
         81115.82433583,   85826.10741608,   90745.70263093,
         95879.1603946 ,  101231.030878  ,  106805.86437489,
        112608.21111481,  118642.62139313,  124913.64550202,
        131425.83361237,  138183.73607819,  145191.98319245,
        152454.88512604,  159977.23217194,  167763.49462311,
        175818.22277253,  184145.96679108,  192751.27703278,
        201638.70379059,  210812.79711335,  220278.10753815,
        230039.18517487,  240100.58031647,  250466.84325593,
```

Fig:2.Prediction using SVM

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907067 |

Fig:3. is taken for days vs number of cases. The graph is drawn for confirmed cases and prediction. The blue line represent the confirmed cases over a period of time. The yellow line represent the prediction of confirmed cases using linear regression model. From this graph we can say that confirmed and predicted vary with a lot of difference. We observe that the cases are increasing but using linear regression the cases are less than the actual. This indicate that linear regression is not best suitable regression algorithm for prediction.

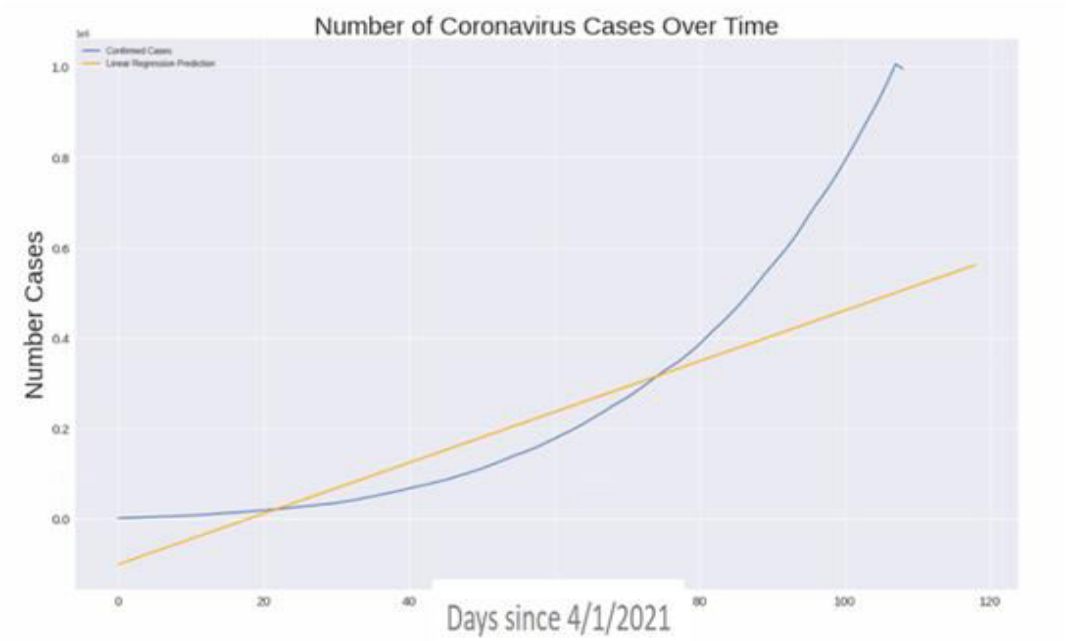

Fig:3.graph is drawn for confirmed cases and prediction.

Fig:4.shows the future predicted cases for 10 days i.e., 10 is taken as input for number of days.

```
linear regression future prediction
[[511265.94846977]
 [516888.83265108]
 [522494.12483239]
 [528108.2170137 ]
 [533722.30919501]
 [539336.40137632]
 [544950.49355763]
 [550564.58573895]
 [556178.67792026]
 [561792.77010157]]
```

Fig:4. Future Prediction using Linear Regression

IV. CONCLUSION AND FUTURE ENHANCEMENT

With the quickly changing circumstance including COVID-19, it is basic to have exact and powerful techniques for foreseeing present movement. Forecasting is a method that uses historical data as inputs and estimates its effect in the future. In our project we proposed an approach to predict time-series data. We then applied this approach to predict the new daily reports of COVID-19 cases in multiple states. The COVID-19 confirmed, death and recovered cases of different states are forecasted with the available data set using machine learning algorithms. The algorithms used in predicting the COVID-19 cases are linear regression, random forest and support vector regression. These are the basic regression models used for forecasting.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907067 |

FUTURE SCOPE:

In future it very well may be applied to anticipate the spread of infection as well as other related boundaries. These include medical and food supply shortages and demands, subgroup infections and immunity and many more. Related to spread of virus we can include the count of male and female, which age group are being affected more and their per-centage and many more. With the help of new and advanced machine learning algorithms like dictionary learning this can be implemented which require less data set unlike other algorithms which require large data sets. This method does not loose interpretability and also works with universal data types.

REFERENCES

[1]. Furqan Rustam , Aijaz Ahmad Reshi , Arif Mehmood ,Saleem Ullah , Byung-Won On, Waqar Aslam and Gyu Sang Choi may.2020 COVID-19 Future Forecasting Using Supervised Machine Learning Models" IEEE ACCESS pp 101489-101499

[2] .Romney B Duffey and Enrico Zio may.2020 Analysing Recovery From Pandemics by Learning Theory: The Case of CoVid-19 vol.8 IEEE Engineering in Medicine and Biology Society Section pp 110789-110795

[3]. PetarRadanliev, David De Roure and Rob Walton jun.2020 Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development - In the first wave of the Covid-19 pandemic vol.3 Science direct pp 1121-1132,

[4]. Dataset from kaggle: https://www.kaggle.com/datasets

[5].For Prediction techniques: https://towardsdatascience.com/machine-learning-basics- support-vector-regression-660306ac5226

[6]. .https://www.javatpoint.com/machine-learning-random-forest-algorithm

[7].Pandemic Analyzer For Efficient Prediction Of COVID-19 In India Using Machine Learning Algorithms **OWK MRUDULA, A MARY SOWJANYA** **PublishedIn :*European Journal of Molecular & Clinical Medicine*, *2020.***

[8]. Time Series Analysis of COVID-19 Data to Study the Effect of Lockdown and Unlock in India by Saswat Singh, Chandreyee Chowdhury, Ayan Kumar Panja & Sarmistha Neogy.
Published in: *Journal of The Institution of Engineers (India): Series B* (2021)

[9]. **A machine learning forecasting model for COVID-19 pandemic in India**
R. Sujath, Jyotir Moy Chatterjee & Aboul Ella Hassanien
Stochastic Environmental Research and Risk Assessment

[10].COVID-19 Outbreak Prediction with Machine LearningbySinaF.Ardabali, Amir Mosavi, Pedram Ghamisi,Filip Ferdinand, Annamaria R.Varkonyi-Koczy,Uwe Reuter, Timon Rabczuk,Peter.M.Atkinson.
Algorithms **2020**, *13*(10), 249; **https://doi.org/10.3390/a13100249**
Received: 8 September 2020 / Revised: 26 September 2020 / Accepted: 27 September 2020 / Published: 1 October 2020
(This article belongs to the Special Issue **Feature Papers in Evolutionary Algorithms and Machine Learning**)

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details