Proceedings of the International Conference on Future of Education, Vol. 5, Issue. 1, 2022, pp.59-68

ISSN 2630-7413 online
DOI: https://doi.org/10.17501/26307413.2022.5105

CONTROLLING THE SPREAD OF COVID-19 IN NIGERIA EDUCATIONAL INSTITUTIONS USING SUPERVISED REGRESSION BASED POLYNOMIAL MODEL OF DEGREE 4

Nze CV[1] and Oguike OE[2*]

[1]*Department of Computer Science and Mathematics, Faculty of Natural and Environmental Sciences, Godfrey Okoye University, Nigeria*

[2]*Department of Computer Science, Faculty of Physical Sciences, University of Nigeria, Nsukka, Nigeria*

Abstract: Coronavirus is a highly infectious disease, caused by the new Corona virus (COVID-19) and can spread from person to person through sneezing and coughing droplets. It has signs and symptoms, similar to the common cold but is dangerous and if not reported early and managed by health workers it can cause severe illness in humans and can lead to death. The disease has affected every sector of the economy, including the education sector, as a result, the education sector has devised new methods of delivering teaching and learning in the post COVID-19 era. Therefore, the need to control the spread of the virus cannot be overemphasized. This paper uses Object Oriented Methodology to analyze, design a machine learning system, which was implemented using a python interpreter and a python Integrated Development Environment (IDE) called pycharm. The implementation uses Polynomial Regression model, which is a supervised regression based machine learning model that can be used to control the spread of the virus by predicting the number of incidence of COVID-19 infections and deaths in Nigeria. Sample data dating from 23rd December 2021 to 8th May 2022 was used, which gave a high accuracy of 93% in COVID-19 confirmed cases and 100% accuracy in death cases. They both predicted low trend of COVID-19 death cases and confirmed cases in Nigeria in the next one month. This prediction will be very useful to health and government authorities in Nigeria, which have the responsibility of controlling the spread of the virus in Nigeria. It will help them to know how to adopt the various COVID-19 protocols in Nigeria.

Keywords: polynomial regression, COVID-19, supervised machine learning, python IDE

Introduction

As of June 10, 2020, there have been over 7.2 million cases of COVID-19 in the world and more than 411,000 deaths. In Africa, confirmed cases of 202,864 and 5539 number of deaths within a number of African countries have been reported. On April 25, 2021, Nigeria through NCDC (Nigeria Center for Disease Control) has reported 164,719 confirmed cases and 2,062death cases due to COVID-19. These reports came from data of COVID-19 confirm cases and death collected from each state in Nigeria. As the COVID-19 continues to spread in the world, countries or the world at large will need

*Corresponding Author's Email: *osondu.oguike@unn.edu.ng*

to take steps, which will aim at controlling the spread of the virus. One of measures that will help to control the spread of the virus is the non-pharmaceutical solutions, which includes: washing of hands, social distancing, wearing of face mask, partial or total lock down etc. However, because some of these COVID-19 protocols have negative economic implications, there is need to know when and how to apply them at different times. In order to do these, data of COVID-19 incidence and death will be used to train the polynomial regression model, which can predict the number of incidences of COVID-19 infections and deaths at any time of the year. The result of this prediction will help to determine the appropriate COVID-19 protocol to apply at any time. This study implements the polynomial regression model, which uses the daily incidence of COVID-19 infections and deaths to train the polynomial regression model with the aim of predicting the number of incidences of COVID-19 infections and death at any time in Nigeria. The prediction will help government and regional health organization know where and how to apply COVID-19 protocols, such as washing of hands, social distancing, putting on nose mask, partial or total lock down etc. By using polynomial regression machine learning algorithm, one will be able to make better prediction, because polynomial regression model is the ideal model for a pandemic situation, like COVID-19, due to the rise and fall nature of the pandemic. Even though the problem of fitting polynomial regression is similar to the one of fitting multiple regression, polynomial regression has special feature (Delhi-, 2004). The polynomial regression model is a special case of a multiple variable linear regression model (*Sample Data :*, 2009). We fit a polynomial to smooth out fluctuation in the data caused by random or uncontrollable error, not because it is thought to represent the relationship (Delhi-, 2004). The Polynomial Regression procedure is designed to construct a statistical model describing the impact of a single quantitative factor X on a dependent variable Y (*Sample Data :*, 2009). A polynomial model involving X and powers of X is fit to the data (*Sample Data :*, 2009). Tests are run to determine the proper order of the polynomial. The fitted model may be plotted with confidence limits and/or prediction limits. Residuals may also be plotted and influential observations identified (*Sample Data :*, 2009). By using polynomial regression algorithm we can predict the number of incidence of COVID-19 infections and death.

Statement of the Problem

The traditional approach of delivering teaching in most educational institutions in Nigeria requires students and lecturers to gather in designated places, like classrooms and lecture halls. Therefore, there is the need to adopt various COVID-19 protocols in order to control the spread of COVID-19 in educational institutions in Nigeria. In order to do this, there is need to track or keep records of number of incidence of infections and death caused by COVID-19. The main problem this research will solve is to help control the spread of COVID-19 in educational institutions in Nigeria, this will be achieved by predicting the number of incidence of COVID-19 infections and deaths in Nigeria educational institutions in particular, which will enable government and regional health organization know where and how to apply COVID-19 protocol such as washing of hands, social distancing, putting on nose mask, partial or total lock down etc. Therefore, this paper is expected to solve the following problems:

- Negative economic effect of COVID-19 due to misinformed adoption of COVID-19 protocols in Nigeria and at a particular time.
- Inappropriate allocation of relevant resources for the management and treatment of COVID-19 in Nigeria, at a particular time.

- General negative effect of COVID-19 in Nigeria educational sector due to misinformation about the spread of the virus.

Review of Related Literature

Various research works have been done on machine learning, mostly, research work on the use of machine learning to make prediction on COVID-19 virus. And some of these works include:

VitaliyYakovyna and Natalya Shakhovska (2021) in a research article, Modelling and predicting the spread of COVID-19 cases depending on restriction policy based on mined recommendation rules, developed models for the spread of COVID-19 depending on restrictive measures taken by governments around the world, and to identify signs that have high prognostic power to predict pandemic behavior and its consequences depending on government action. The dynamics of migration between countries in clusters, and their relationship with the number of confirmed cases and the percentage of deaths caused by COVID-19, were also studied on the example of Poland, Italy and Germany.

M. RubaiyatHossainMondal, SubratoBharati and PrajoyPodder (2021) provides a systematic review of the application of Artificial Intelligence (AI) in the form of Machine Learning (ML) and Deep Learning (DL) techniques in fighting against the effects of novel coronavirus disease (COVID-19). The systematic review was performed following PRISMA guidelines. The review reveals that ARIMA models of different orders, PR, RIDGE, SVR, logistic models, and the hybrid wavelet ARIMA model effectively predict the number of confirmed cases. It is further found that multilayer perceptron, SVM, random forest, XGBoost, etc. were useful in classifying COVID-19 patients from normal people. During the research in one particular case, random forest achieves a classification accuracy of 95.95%. In DL-based COVID-19 diagnosis, there are several stages, including image preprocessing, segmentation, feature extraction and classification.

On the other hand, Hari Singh and Seema Bawa (2021), used different machine learning regression models: Li-MuLi-Poly, applied linear regression (LR), multi-linear regression (MLR) and polynomial regression (PR) techniques to propose a model Li-MuLi-Poly.The model predicts COVID-19 deaths happening in the United States of America. The experiment was carried out on machine learning model, and minimum mean square error model, and maximum likelihood ratio model were used to evaluate the models. The best fitting model was selected according to the measures of mean square error, adjusted mean square error, mean square error, root mean square error (RMSE) and maximum likelihood ratio, and the statistical t-test was used to verify the results. Data sets were analyzed, cleaned up and debated before being applied to the proposed regression model. The correlation of the selected independent parameters was determined by the heat map and the Carl Pearson correlation matrix.

In a similar manner, Hafiz FarooqAhmad , Huda Khaloofi , Zahra Azhar, AbdulelahAlgosaibi and JamilHussain (2021), used an improved COVID-19 forecasting by Infectious Disease Modelling Using Machine Learning. This paper highlights the mechanisms of the coronavirus forecasts, namely, the ML and DL approaches. The united employment of informal datasets regarding the circulation and transmission of COVID-19 in Saudi Arabia, the Kingdom of Bahrain, Kuwait, and the UAE.

Additionally, the following algorithms were validated, and the optimal model was selected based on error rate in predictions: linear regression, PR, SVR, SIR, LSTM, and Bi-LSTM. During time of research the proposed models was easy to perform and provide reliable conclusions in real time. The DL models demonstrated crucial enhancements when dealing with time-series data in various applications. The DL methods promise time-series forecasting, such as the self-regulating comprehension of temporal dependence, and the administration of temporal complexes Bi-LSTM outperforms among DL models, and its forecast values entirely overlap significant numbers of actual cases. Therefore, it performs well with the highest accuracy and is most recommended for disease forecasting (Ahmad et al., 2021). Furthermore, Jayakumar Kaliappan1, Kathiravan Srinivasan1, SaeedMianQaisar, KarpagamSundararajan, Chuan-Yu Chang and Suganthan C (2021), evaluated the performance of Regression Models for the Prediction of the COVID-19 reproduction rate. In addition, Muhammad Naeem, Jian Yu, Muhammad Aamir, Sajjad Ahmad Khan, Olayinka Adeleye and Zardad Khan (2021), performed comparative analysis of machine learning approaches to analyze and predict the COVID-19 outbreak. The result of the study showed that ANN performed better than the other models (RF, SVM, and KNN)

Meanwhile, Meenu Gupta, Rachna Jain, SimrannArora, Akash Gupta, Mazhar Javed Awan, Gopal Chaudhar and Haitham Nobanee (2021), presented COVID-19 outbreak analysis and prediction, a case of Indian States vs. Union Territories. The work carried out comparative analysis of the prediction of an increase in the number of recovered and death cases in different states and union territories in India in the near future, as predicting these cases would help in estimating and arranging beds, ventilators, and other healthcare equipments on time and save many lives with proper facilities.

However, Sirage Zeynu Ahmed (2020), uses ML to analyze and forecast the outbreak of COVID-19 in Ethiopia, In this analysis, ML algorithms were used to predict the effects of coronavirus outbreaks in Ethiopia, which was believed may help the Ethiopian policy makers and governments to take inclusive and necessary action. The study used Support Vector Machine (SVM) and Polynomial Regression (PR) models to forecast the spread of pandemic trends in Ethiopia. Similarly, Mazharul Islam Leon, MdIfrahamIqbal, SayedMehediAzim, and Khondaker A. Mamun, (2020), analyzed and predicted Coronavirus infections and deaths in Bangladesh using Machine Learning Algorithms.

On the other hand, Gaurav Pandey, PoonamChaudhary, Rajan Gupta, Saibal Pal (2020) used regression model to predict the COVID-19 outbreak in India. The objective of the research was focused on finding the rate of spread of the disease in India, They developed a mathematical model called, SEIR (Susceptible, Exposed, Infectious, Recovered), which evaluated the spread of disease and Prediction of COVID-19 outbreak using SEIR and Regression models. Mohamed Lounis, B. Malavika (Year) forecasted COVID-19 cases in Algeria using Logistic Growth and Polynomial Regression models. The results of this research showed that the polynomial regression model is more adaptable for COVID-19 forecasting in Algeria. Thus, they used this model to predict the short-term future cases until January 19, 2020. The number of cases for this date was projected to reach 387,673 cases. This model could help the Algerian government in adapting the best strategies against the COVID-19 epidemic (Lounis & Babu, 2021)

In a similar manner, Ramjeet Singh Yadav (2020), used ML methods to perform data analysis of COVID-19 in India. In this paper, six regression analysis-based machine learning models for prediction of the COVID-2019 outbreak datasets of India were proposed. These models basically

regression analysis-based exponential, quadratic third degree, fourth degree, fifth degree and sixth degree polynomials. These models also predicted the outbreak of the COVID-19 in India for the next 7 days.

Materials and Methods

Time series data provided by NCDC (Nigeria Center for Disease Control) official web site was used for the empirical result analysis. The time period of data is from 23/12/2021(week 52) to 08/03/2022 (week 18). The data includes confirmed cases, death cases and recovered cases of all state in Nigeria but for the sake of this research work, we are interested in confirmed cases and death cases, which will serve as the training data for this research. The degree of polynomial regression used in the algorithm for the machine learning prediction is degree 4. F For the sake of providing a polynomial regression mathematical equation of our sample data, a polynomial regression of degree four (Y = A + BX + B2X2 + B3X3 + B4X4) was used since this is the fourth wave of COVID-19 in Nigeria.

Polynomial Regression

For this work, we used polynomial regression to make our predictions. The choice of polynomial regression is because it is the ideal model for the prediction of number of incidence of cases of pandemic like COVID-19, due to the polynomial nature of falling and rising. Polynomial regression is a special case of multiple regressions, with only one independent variables Xi. one-variable polynomial regression model can be expressed as shown in Equation (1).

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_k x_i^k + e_i \text{ , for } i = 1, 2, \ldots, n \quad (1)$$

Where k is the degree of the polynomial. The degree of the polynomial is the order of the model. Effectively, this is the same as having a multiple model with X1= X , X2= X2, X3= X3, etc. The polynomial regression model can be evaluated using the following metrics, the mean squared error MSE, Mean Absolute Percentage Error, MAPE (Ostertagová, 2014). Other metrics are the R-squared R2 (coefficient of determination) of the multiple regression is similar to the simple regression (Ostertagová, 2014). The value of R2 is always between zero and one, $0 \leq R2 \leq 1$. An R2 value of 0.9 or above is very good, a value above 0.8 is good, and a value of 0.6 or above may be satisfactory in some applications, although we must be aware of the fact that, in such cases, errors in prediction may be relatively high. When the R2 value is 0.5 or below, the regression explains only 50 % or less of the variation in the data; therefore, prediction may be poor. (Ostertagová, 2014). The graph representation of polynomial regression model is shown in Figure 1.

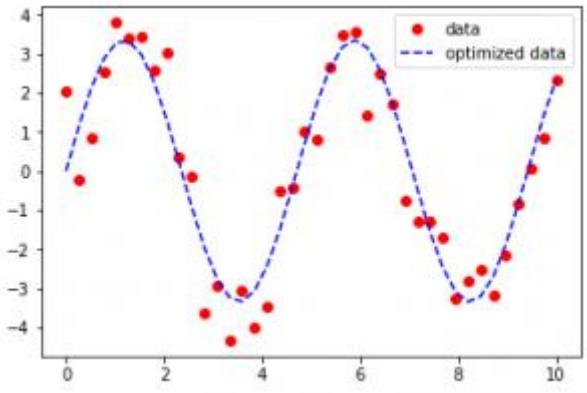


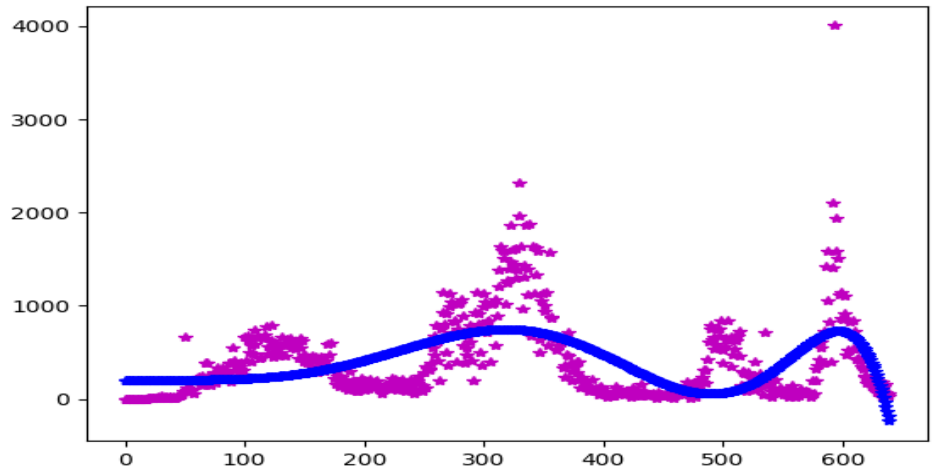
Figure 1. Graph Representation of Polynomial Regression Model

Dataset

The dataset used for this research was gotten from NCDC (Nigeria Center of Disease Control) web site. It contains data of Nigeria COVID-19 situation report starting from 14th March 2020. But for this research work, time series data from 23/12/2021(week 52) to 08/03/2022 (week 18) were used as the sample data to make the prediction. This data set is available at https://ncdc.gov.ng/diseases/sitreps/?cat=14&name=An%20update%20of%20COVID-19%20outbreak%20in%20Nigeria. The following are the attributes of the dataset, Serial Number, Date, Confirmed, Recoveries and Deaths. The attributes, Confirmed, Recoveries and Deaths are numerical values of the number of incidence of confirmed cases, recoveries and deaths.

Results and Discussion

The sample data was collected and stored in an excel spread sheet, saved as a CSV file, (comma-separated values). After that the data was loaded into the polynomial regression model, using the pandas programming library then it goes through the data preprocessing stage, then to the training stage and test stage. The result in Figure 2, below was obtained using polynomial regression of degree 4, for the weeks against the number of confirmed cases



.

Figure 2. Graph representations of polynomial regression line on confirmed cases sample data

The figure above is the graph representation of the polynomial regression, which trace the increase and decrease of the sample data (Covid-19 confirmed cases). The purple color represent data while the blue color represents a polynomial line. The degree of the polynomial regression determines the curvature of the line. From the graph we can see four different waves which indicate the number of times there has been increase and decrease in the number of COVID-19 confirmed cases in Nigeria. The circles in the image are used to indicate raise and fall of the waves as the day passes. The light blue circle are used to indicate an increase in COVID-19 confirmed cases in Nigeria as the day goes by while the red circle indicates a decrease in COVID-19 confirmed cases in Nigeria as the day goes by. The graph shows that the first wave increase happened after 100 days of COVID-19 discovery in Nigeria and was less than 1000 confirmed cases of COVID-19, while the second wave was after 300 days but higher than the first wave and it shows that more than 2000 confirmed cases were discovered. The third wave shows a decrease in confirmed cases after 500 days which is less than 1000 confirmed cases of COVID19 and the fourth wave, which has the highest number of confirmed COVID-19 cases has more than 4000 cases of COVID-19 confirmed cases after 600 days and currently just as the graph indicates the trend is going down . This graph shows that there has not been a constant raise and fall of COVID-19 confirmed cases in Nigeria but rather fluctuate (increases and decreases as the day passes). The result in Figure 3, below was obtained using polynomial regression of degree 4, for the weeks against the number of death cases.

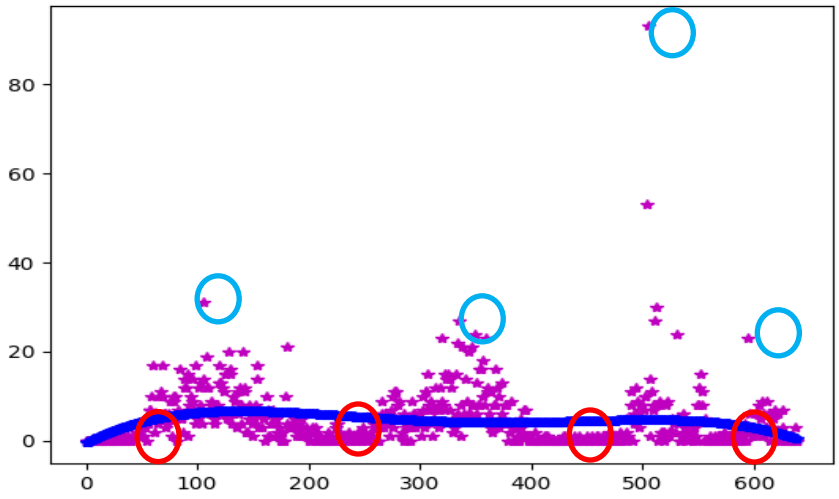


Figure 3 Graph representations of polynomial regression line on death cases sample data

The figure above is the graph representation of the polynomial regression line, which trace the increase and decrease of the sample data (Covid-19 death cases).The purple color represents data while the blue color represents a polynomial line or slope. From the graph we can see four different waves which indicate the number of times there has been increase and decrease in the number of COVID-19 death cases in Nigeria. The circles in the image are used to indicate raise and fall of the waves as the day passes. The light blue circle are used to indicate an increase in COVID-19 death cases in Nigeria as the day goes by while the red circle indicates a decrease in COVID-19 death cases in Nigeria as the day goes by. The graph shows that the first wave increase happened after 100 days of COVID-19 discovery in Nigeria which was 30 death cases of COVID-19 while the second wave was after 300 days but shows a decrease in COVID-19 death cases compare to the first wave and it shows that up to 26 death cases were discovered on that particular day. The third which has the highest number of death cases of COVID-19 after 500 days has more than 80 death cases of COVID-19 and the fourth wave which has the lowest number of death COVID-19 cases has more than 23 cases of COVID-19 death cases after 600 days and currently just as the graph indicates the trend is going down . This graph also shows that there has not been a constant raise or fall in COVID-19 death cases in Nigeria but rather fluctuate (increases and decreases the day passes). Figure 4 shows the result of the predicted death cases of COVID-19.

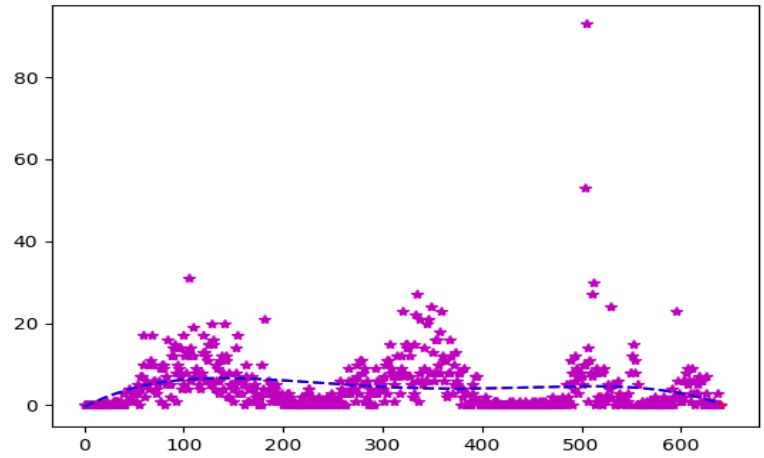


Figure 4 Graph representations of death cases sample data prediction after seven days (one week)

Figure 4, above shows a possible prediction of the Covid-19 death in Nigeria after seven day using the provided sample data. The prediction shows that there will be a slight decrease or no Covid-19 death after seven days. This prediction will be very useful in knowing how and when to impose the various COVID-19 protocols in the educational institutions in Nigeria. Example, whenever there is new wave of COVID-19, i.e. increase number of infections and deaths, lockdown can be imposed so that teaching and learning will continue with the online approach. Whenever the curve flattens, lockdown can be lifted so that normal traditional teaching will continue.

Conclusion

In this study, we have been able to use polynomial regression to predict the number of incidences of COVID-19 infections and deaths in Nigeria, using the live dataset generated by Nigerian Center for Disease Control (NCDC) for the following period, between 23rd December, 2021 and 8th March 2022. The result that this study has obtained will be very useful to the Presidential Task Force on COVID-19. With the result, they will know when and how to apply the various COVID-19 protocols.

Declaration of Interest Statement

The authors declare that they have no conflict of interests.

References

Ahmad, H. F., Khaloofi, H., Azhar, Z., & Algosaibi, A. (2021). applied sciences An Improved COVID-19 Forecasting by Infectious Disease Modelling Using Machine Learning.

Analysis, R. (n.d.). Chapter 12 Polynomial Regression Models. 1–12.

Delhi-, I. A. S. R. I. L. A. N. (n.d.). Polynomial regression.

Juliansyah Noor. (2019). ANALYSIS AND FORECASTING THE OUTBREAK OF COVID-19 IN ETHIOPIA USING MACHINE LEARNING Sirage. Journal of Chemical Information and Modeling, 53(9), 1689–1699.

Gupta, M., Jain, R., Arora, S., Gupta, A., & Awan, M. J. (2021). AI-Enabled COVID-19 Outbreak Analysis and Prediction : Indian States vs . Union Territories. https://doi.org/10.32604/cmc.2021.014221

Kaliappan, J., Srinivasan, K., & Qaisar, S. M. (2021). Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate. 9(September), 1–12. https://doi.org/10.3389/fpubh.2021.729795

Leon, M. I. (2020). Analysing and Predicting Coronavirus Infections and Deaths in Bangladesh Using Machine Learning Algorithms. January 2021. https://doi.org/10.2139/ssrn.3671978

Lounis, M., & Babu, M. (2021). Forecasting COVID-19 Cases in Algeria using Logistic Growth and Polynomial Forecasting COVID-19 Cases in Algeria using Logistic Growth and Polynomial Regression Models. June. https://doi.org/10.2991/dsahmj.k.210630.001

Magoulas, G. D., Prentza, A., Technical, N., & Athens, G.-. (1999). MACHINE LEARNING IN MEDICAL APPLICATIONS George D. Magoulas 1 and Andriana Prentza 2 1. 1999.

Mondal, M. R. H., Bharati, S., & Podder, P. (n.d.). Diagnosis of COVID-19 Using Machine Learning and Deep Learning : A Review. 1–17.

Naeem, M., Yu, J., Aamir, M., & Ahmad, S. (2021). Comparative analysis of machine learning approaches to analyze and predict the COVID-19 outbreak. 2019. https://doi.org/10.7717/peerj-cs.746

Ostertagová, E. (2014). Modelling Using Polynomial Regression. May. https://doi.org/10.1016/j.proeng.2012.09.545

Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2019). SEIR and Regression Model based COVID-19 outbreak predictions in India. 1–10. Sample Data : (2009). 1–15.

Singh, H., & Bawa, S. (2021). Predicting COVID - 19 statistics using machine learning regression model : Li - MuLi - Poly. Multimedia Systems, 0123456789. https://doi.org/10.1007/s00530-021-00798-2

Yadav, R. S. (2020). Data analysis of COVID-2019 epidemic using machine learning methods : a case study of India. *International Journal of Information Technology*, 12(4), 1321–1330. https://doi.org/10.1007/s41870-020-00484-y

Yakovyna, V., & Shakhovska, N. (2021). Modelling and predicting the spread of COVID-19 cases depending on restriction policy based on mined recommendation rules. 18(March), 2789–2812. https://doi.org/10.3934/mbe.2021142

Shalabh (2020) Polynoomial regression model Available onlinehttp://home.iitk.ac.in/~shalab/regression/Chapter12-Regression-PolynomialRegression.pdf