**Data Analysis and prediction of COVID-19 Dataset**

**Aaditya Pradipbhai Parekh Devanshi Shah Mayank Chauhan**

Masters of Applied Computing Masters of Applied Computing Masters of Applied Computing

University of Windsor University of Windsor University of Windsor

[Parekh23@uwindsor.ca](mailto:Parekh23@uwindsor.ca) [shah641@uwindsor.ca](mailto:shah641@uwindsor.ca) [chauha15@uwindsor.ca](mailto:chauha15@uwindsor.ca)

**Kshitiz Choudhary**

Masters of Applied Computing

University of Windsor

[Choudh22@uwindsor.ca](mailto:Choudh22@uwindsor.ca)

**Abstract:**

The likelihood of an outbreak of the CORONA virus has recently been researched, although there aren't many multiple model comparisons or uncertainty prediction analyses in the published literature. In order to find out a better way to predict and forecast the epidemic situation, this paper utilized machine learning and a series of regressions, including Bayesion ridge Regression, polynomial regression, Gradiant Boosting Prediction, and Support Vector Regression models. In addition, in order to improve the accuracy, we use the Hyper Parameter Tuning to optimize the result of the models. Investigating the Corona Virus Disease 2019 (COVID-19) forecast of confirmed cases of the next 10 days is the goal of this project. This paper tried to find out the model that fits the training data by calculating the mean square error (MSE) o [Mean Absolute Error](https://www.sciencedirect.com/topics/engineering/mean-absolute-error) (MAE) to evaluate the models. In this analysis, data including attributes such as location wise confirmed, deceased, recovered COVID-19, longitude and latitude were collected worldwide.

**Introduction:**

In December 2019, a cluster of infected patients with unknown cause was reported to WHO (World Health Organization) by Wuhan, China. A brand new virus which has not been discovered was named as novel Coronavirus in January 2020. COVID-19 was declared as a pandemic by WHO [1]. As of 30 April 2020, more than three million confirmed cases of COVID-19 were reported globally with 217,769 deaths [2].

Since the beginning, Covid-19 has had a firm hold on people’s lives and some countries have been struggling with an increase in infections [3]. With the help of data analysis, one can examine how different countries are performing in terms of controlling the outbreak of novel Coronavirus [5]. Predictions are made using the dataset provided to the nation assisting them in determining how much they can control the pandemic or how much they should direct preventive measures [6]. The objective of the project is to provide data analysis of Covid-19. Different situations, such as the most affected countries by the virus, have been analyzed through plotting of data.

Through this project, a step is taken to help citizens in understanding the spread and visualize cases in their nation [4]. To display and analyze the growth of cases and recovery graph, analysis of data from multiple countries is merged. This project also provides information on how prepared and well-advanced a nation is at controlling the spread. To analyze how much a country is being impacted by and recovering from each day, comparison graphs have also been displayed[7].

Machine learning plays a major role in better understanding and examining COVID-19 crisis as it identifies the patterns in data and uses them to automatically make predictions or decisions. In this model authors are going to look at the development of Coronavirus affirmed. Investigation and predictions are done using Bayesion ridge Regression, polynomial Regression, Gradiant Boosting Prediction and Support Vector Regression (SVM). With the help of all these models we are predicting total number of confirmed cases worldwide for the next 10 days.

**Related work:**

There are a lot of research papers published that are related to the covid-19. Some of them to name can be research work related to vaccines or other details related to covid-19. Also there are some papers on the medical drugs that can help to recover. Rajan Gupta and Saibal Kumar Pal, in their research paper 'Trend analysis and forecasting of covid-19 break in India' used exploratory data analysis to report the situation in the time period of January to March in India. They use time series forecasting methods to predict the future trends.

In this paper it is shown that there have been different modelling approaches presented by various researchers for different countries. Models for China [8,9], Italy [10], France [11], India [13] and USA [12]. But all this models are Country wise. The scope of this study is limited to building forecasting models for Indian region and uses time series based forecasting methods which are easy to build and easy to understand in these kind of critical conditions. The study does not include forecasting for any other nation suffering from COVID-19 outbreak [14]. Our purposed model is to do the visualization of covid-19 data world wise. In the paper the researchers identifies following finding based corona virus (COVID-19) datasets - Till 2nd April 2020 Total Cases in India, Date Wise Total Corona cases in India, India's Map with State wise data of Total Cases, Deaths and Cure[15]

It is important to detect the epidemic pattern for planning the early action to combating COVID-19. Machine Learning (ML) can help us to infer useful knowledge from past massive epidemic data. Various mathematical and statistical forecasting tools were applied to generate short-term and long-term forecasts. Among numerous approaches, Regression models have emerged as a promising solution and have proved its competence in this regard.  Multiple machine learning tools such as multiple regression analysis, progressive partial derivative linear regression model , and a hybrid approach of the auto regressive integrated moving average model and wavelet-based forecasting model have also been applied to make these predictions[17,18,19,20,21,22,23]. Several researchers have also tried their hand in establishing the role of regressions as to predict the number in healthcare.

# For instance [Apurbalal Senapati](https://link.springer.com/article/10.1007/s41870-020-00552-3#auth-Apurbalal-Senapati), [Amitava Nag](https://link.springer.com/article/10.1007/s41870-020-00552-3#auth-Amitava-Nag), [Arunendu Mondal](https://link.springer.com/article/10.1007/s41870-020-00552-3#auth-Arunendu-Mondal) & [Soumen Maji](https://link.springer.com/article/10.1007/s41870-020-00552-3" \l "auth-Soumen-Maji)  in their paper ‘A novel framework for COVID-19 case prediction through piecewise regression in India’ [24] uses the machine learning (ML)-guided linear regression model in which they predict the positive cases in the next 45 days with the help of piecewise regression analysis. Authors in Chakraborty et. al. [[25](https://link.springer.com/article/10.1007/s41324-021-00379-5#ref-CR17)] undertook the problem of current number of COVID-19 cases. Additionally, authors in [[25](https://link.springer.com/article/10.1007/s41324-021-00379-5#ref-CR17)] also focused on the spread of novel virus in different countries. Authors employed Wavelet-based forecasting model to predict the count of COVID-19 cases for a period of 10 days in various countries like Canada, France, India, South Korea and UK.

# F. Rustam et al. (2020) introduced future forecasting on Covid-19 using various machine learning models. They use four regression models namely LR, LASSO, [SVM](https://www.sciencedirect.com/topics/engineering/support-vector-machine) and Exponential Smoothing. They predicted their results on the basis of Evaluation Parameters used in their models. According to their results ES model is best for predicting infected cases and recovered cases and LR model best predict the death rate [[16]](https://www.sciencedirect.com/science/article/pii/S2090447921003385" \l "b0045).

# Authors in[26] implemented the SMA , Auto ARIMA Model , Holts linear model to predict the mortality rate of the covid-19 data from johns Hopkins. Similarly, author Benvenuto et. al. used the most conventional ARIMA forecasting model to predict the mortality rate of COVID-19 using data from Johns Hopkins epidemiological of the predominance [27]. Our model will be predicting the number of confirmed cases worldwide for the next 10 days. Additionally, authors in [2[9](https://link.springer.com/article/10.1007/s41324-021-00379-5#ref-CR25)] suggested a prognostic forecasting model with respect to three indices to predict the mortality rate and associated risks. They further suggested a clinical path to recognize different cases based on its severity that will aid doctors to timely diagnose and identify so as to minimize the mortality rate.

# Further, Authors in [28] uses the polynomial regression model is used to model the future scenario of any pandemic in terms of the number of deaths and the number of cases using R programming software prevalence in Pakistan, India, Afghanistan, Iran, Italy, and China. The work gave the forecasting there is a drastically increasing trend of COVID-19 confirmed cases in Italy depicting an average proportion of urn:x-wiley:01466615:media:jmv27506:jmv27506-math-0124, urn:x-wiley:01466615:media:jmv27506:jmv27506-math-0125 (Iran), and urn:x-wiley:01466615:media:jmv27506:jmv27506-math-0126 (India) relative to China. In this paper they are using the R language software for the prediction, in ours we are also using the polynomial regression worldwide.

# The limitation of this study is that they are using the polynomial regressions on only some of the countries and not the worldwide and also they are predicting the number of deaths using the R programming software. In our research we find out that in most of the studies they are predicting the mortality rate for a single country or worldwide and not many predicting the number of the cases in the future. That is also one of the reason in our model we are predicting the total number of cases worldwide for the next 10 days.

# Further, authors in [30] uses the Multiple linear regression model is proposed for prediction of Active cases in COVID-19 daily data for one country India. In this they predict the total value of active cases in India and predict the number of cases state wise. The implementation of various ML forecasting models in COVID-19 is also simulated by authors in [[31](https://link.springer.com/article/10.1007/s41324-021-00379-5#ref-CR21)] by implementing different models like ARIMA, CUBIST, RF, RIDGE, SVR and the stacking-ensemble method for time series analysis and prediction. The scope of these studies is limited to one country only and also in these studies they are forecasting the data using some different models.

# Authors in [[32](https://link.springer.com/article/10.1007/s41324-021-00379-5#ref-CR22)] adopted the SEIR and regression model to predict the COVID-19 cases. The efficacy of the model is evaluated by Root Mean Square Logarithmic Error (RMSLE) and reproduction number (R0)(R0). Similar line of research is also carried out by authors in [33] who utilizes a series of regressions, including linear regression, multi-regression, and Lasso regression models. This paper tried to find out the model that fits the training data by calculating the mean square error (MSE) or coefficient of determination (R}} 2 ) to evaluate the models. They are using these models to analyze the data from United States state wise and of the full country.

# Proposed Model:

# The dataset pertaining to COVID-19 is available at various sources. We use the data collected and distributed by Johns Hopkins COVID-19 Github data repository[34] , as input and outputs the predicted number of confirmed cases in the futurewhich provides an overview of COVID-19 cases (confirmed, deaths, and recovered) for countries around the world. The data on the site is updated daily. It comprises two main components: i) The data analysis and the visualization and ii) the prediction component consisting of four different models (i.e., Bayesion ridge Regression, polynomial Regression, Gradiant Boosting Prediction and Support Vector Regression (SVR). that takes the pre-processed data to generate the predictions.

# C:\Users\HP\Downloads\WhatsApp Image 2022-07-24 at 8.32.25 PM.jpeg

# Fig. Framework of Purposed Study

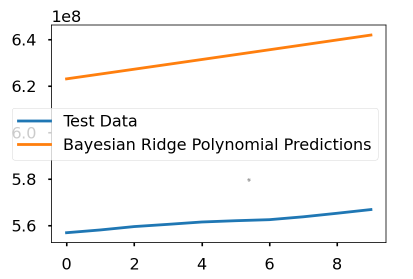
# Bayesion ridge Regression:

# Bayesian regression[35] allows a natural mechanism to survive insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates. The output or response ‘y’ is assumed to drawn from a probability distribution rather than estimated as a single value.

# C:\Users\HP\Downloads\123456.drawio.png

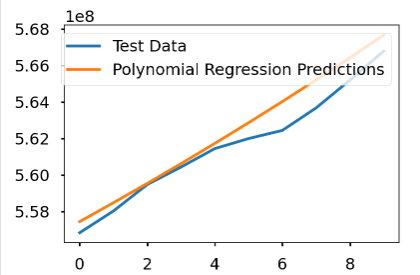
# Fig.Experimental Work Flowof the Model.

# One of the most useful type of Bayesian regression is Bayesian Ridge regression which estimates a probabilistic model of the regression problem. This resulting model is called Bayesian Ridge Regression and in scikit learn sklearn.linear\_model. BeyesianRidge module is used for Bayesian Ridge Regression.



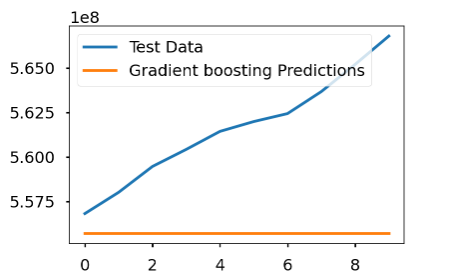
**Polynomial Regression:**

In [statistics](https://en.wikipedia.org/wiki/Statistics), polynomial regression[36] is a form of [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) in which the relationship between the [independent variable](https://en.wikipedia.org/wiki/Independent_variable) x and the [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) y is modelled as an nth degree [polynomial](https://en.wikipedia.org/wiki/Polynomial) in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding [conditional mean](https://en.wikipedia.org/wiki/Conditional_expectation) of y, denoted E(y |x). Polynomial Regression is a special case of Linear Regression where we fit the **polynomial equation** on the data with a curvilinear relationship between the dependent and independent variables. The relationship between the dependent variable and any independent variable is linear or curvilinear (specifically polynomial). We observe between the actual value and the best fit line, which we predicted and it seems that the actual value has some kind of curve in the graph and our line is no where near to cutting the mean of the points. This where polynomial Regression comes to the play, it predicts the best fit line that follows the pattern(curve) of the data



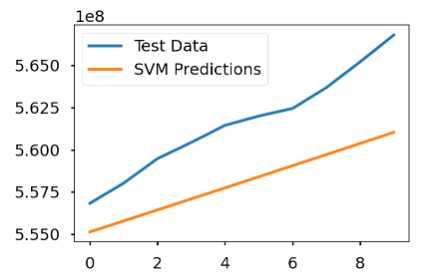
**Gradiant Boosting Regression :**

Gradient boosting[37] is a [machine learning](https://en.wikipedia.org/wiki/Machine_learning) technique used in [regression](https://en.wikipedia.org/wiki/Regression_(machine_learning)) and [classification](https://en.wikipedia.org/wiki/Classification_(machine_learning)) tasks, among others. It gives a prediction model in the form of an [ensemble](https://en.wikipedia.org/wiki/Ensemble_learning) of weak prediction models, which are typically [decision trees](https://en.wikipedia.org/wiki/Decision_tree_learning). When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms [random forest](https://en.wikipedia.org/wiki/Random_forest).



**Support Vector Regression :**

A SVM model is basically a depiction of different classes in a [hyperplane](https://www.sciencedirect.com/topics/engineering/hyperplanes" \o "Learn more about hyperplane from ScienceDirect's AI-generated Topic Pages) in multidimensional space. The hyperplane will be delivered in an iterative manner by SVM with the objective that the goof can be constrained. The goal of SVM is to parcel the datasets into classes to find the biggest immaterial hyperplane (MMH). The RBF Kernel is additionally called the Gaussian part. There'sa boundless number of measurements within the feature space since it can be extended by the Taylor Arrangement. Support Vector Regression[38] is like Linear Regression in that the condition of the line is y = wx + b in SVR, this straight line is alluded to as a hyperplane. The data centres around either side of the hyperplane that are closest to the hyperplane are called Support Vectors which are used to plot the breaking point line.



**Evaluation methods :**

[Mean Absolute Error](https://www.sciencedirect.com/topics/engineering/mean-absolute-error)*:* MAE indicates the average of absolute difference of predicted and actual data value, thus serving as an efficient parameter to refer effectiveness of prediction model. The mathematical expression for MAE is as follows:

MAE=∑Ni=1|x^i−xi|N

[Mean Squared Error](https://www.sciencedirect.com/topics/engineering/mean-squared-error)*:* MSE refers to the square error and is calculated as follows:

MSE=1N∑i=1N(x^i−xi)2

# Results:

# Data Visualization

The main objective of this data visualization is to communicate information simply and efficiently using various graphical representations as it has various conceivable functions in the domain. Visualization is a helpful medium for examining, understanding, and transmitting the information. Python is considered one of the best programming languages for handling data visualization because of its vast and active scientific computing community, as well as its numerous libraries that provide greater flexibility. It may also manage the specific parts of the graphs that are formed and make those requirements code reproducible. Python is also excellent at dealing with data and can manage massive volumes of data without crashing. For data visualization in python, we will use different libraries such as Matplotlib, Pandas, NumPy, Plotly and Folium.

The graphs are represented using interactive world maps along with time lapse. Cases density animation using heat maps represents the severity in different parts of the world. The product features/functions that we are planning to implement to accomplish the task are cases over the time with area plot as well as world map, worldwide cases on Folium maps, to display confirmed cases with choropleth maps (cases over time), confirmed and death cases with static ColorMap and deaths as well as recoveries per hundred cases.

There will be a scatter plot for death versus confirmed cases (cases are on log10 scale). Graphical representations using bar plot and line plot for confirmed, death, new cases, recovered cases for different countries with date. Growth rate of cases is shown after every hundred, thousand, ten-thousand, and hundred-thousand cases.

The representation also shows the tree map analysis of the number of confirmed and death cases, Gantt chart for first and last case report time of different countries, graphical representation of rise of confirmed cases of each country. Also, the visualization displays the comparison of COVID-19 pandemic and other similar epidemics such as SARS, EBOLA, MERS and H1N1. These comparisons are based on confirmed cases, death cases and the mortality rate.

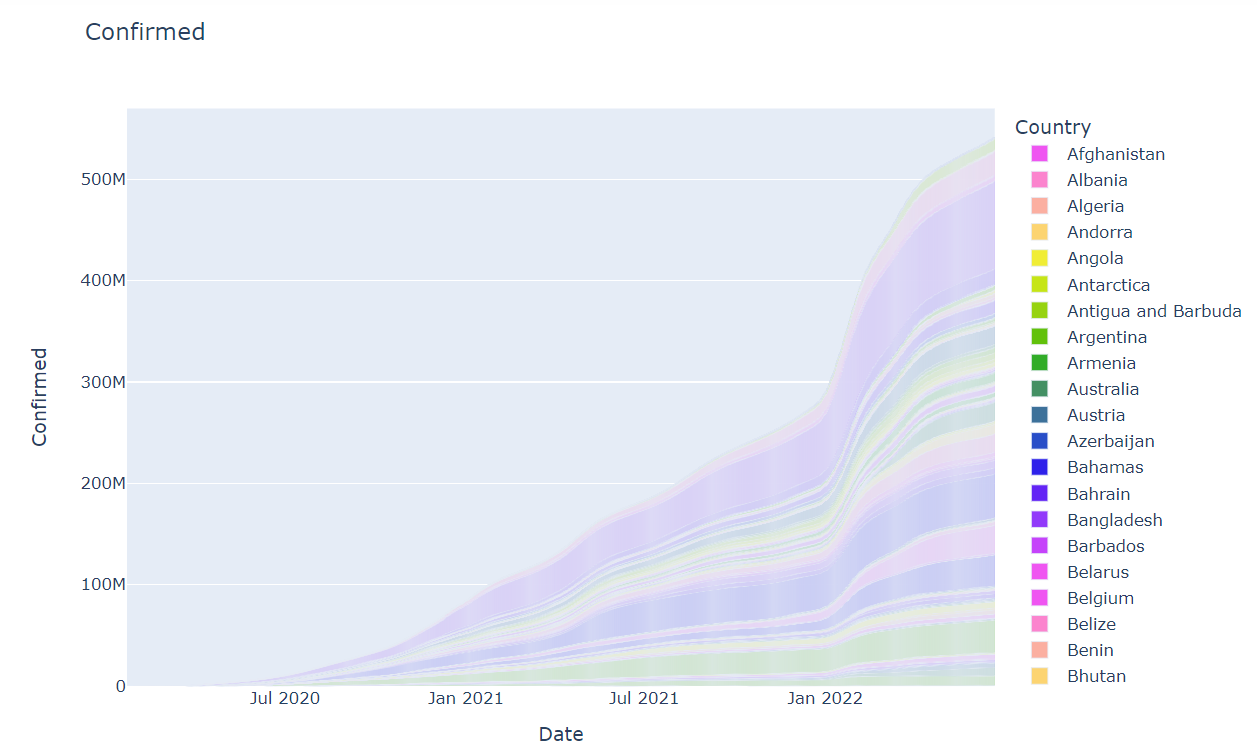


Fig. Confirmed Cases on Bar Plot



Fig. Deaths vs Confirmed Cases (Cases are on log10scale)

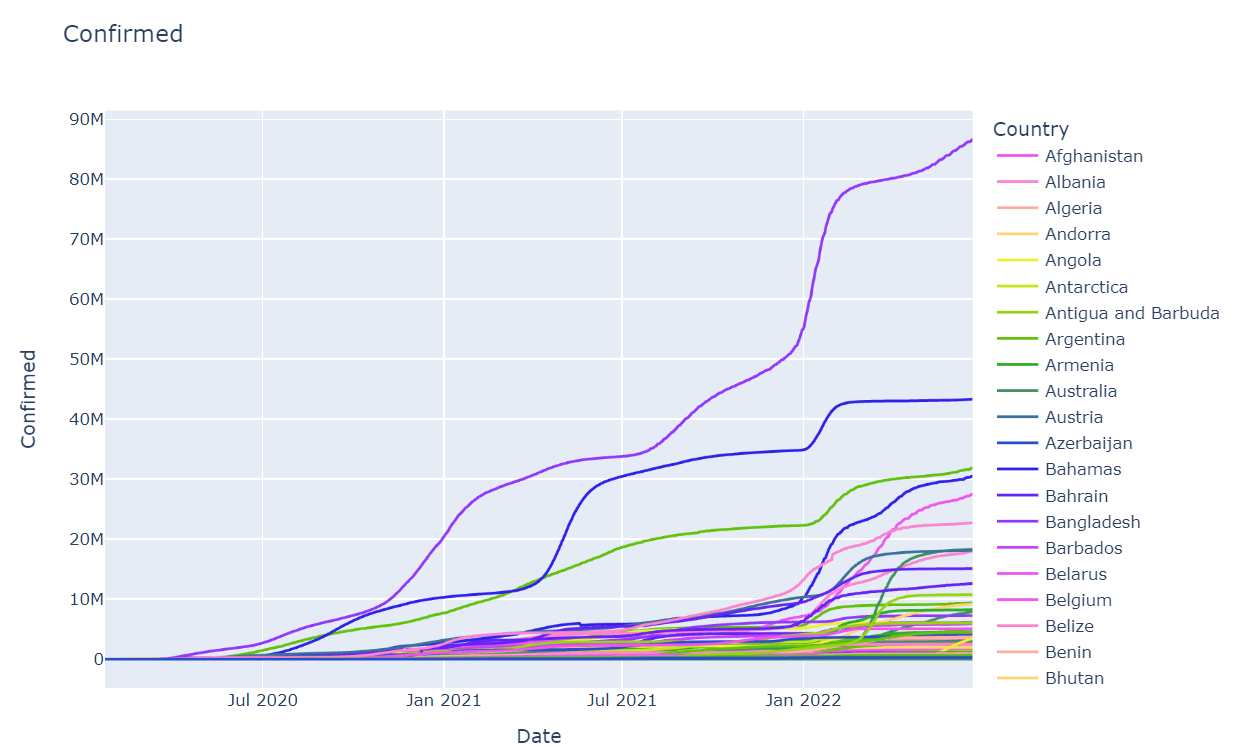


Fig. Confirmed Cases on Line Plot

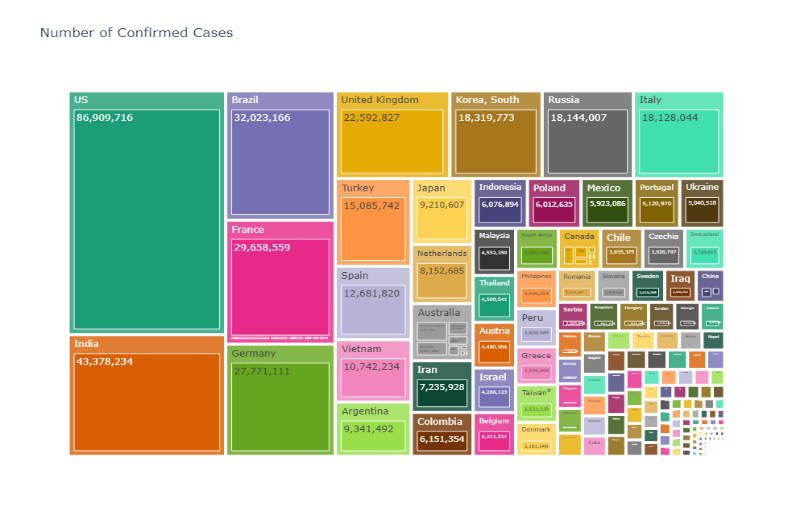


Fig. Confirmed Cases on Tree Graph

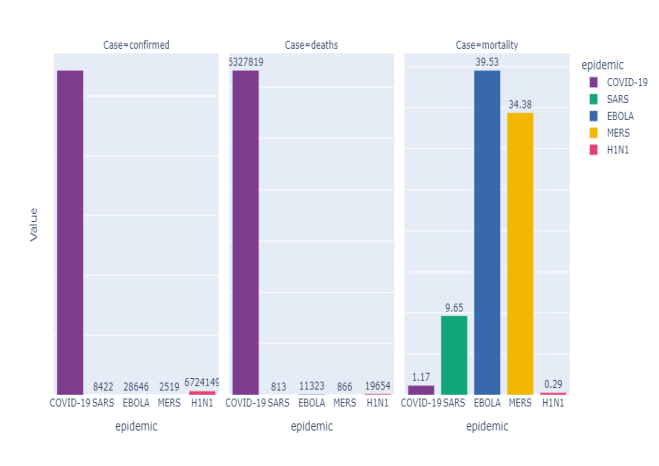


Fig. COVID-19 with other Epidemics

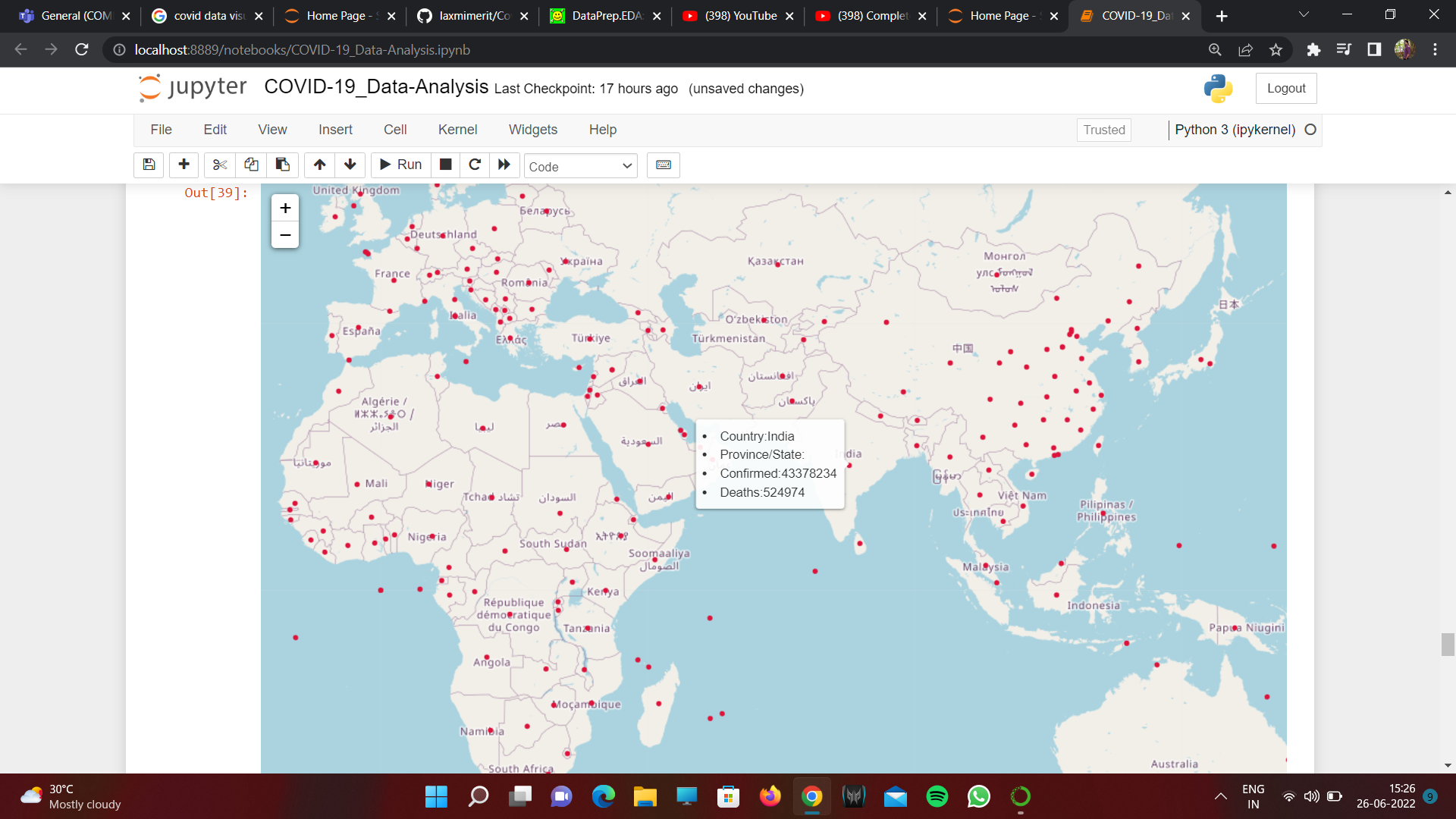


Fig. Worldwide cases on Folium Map

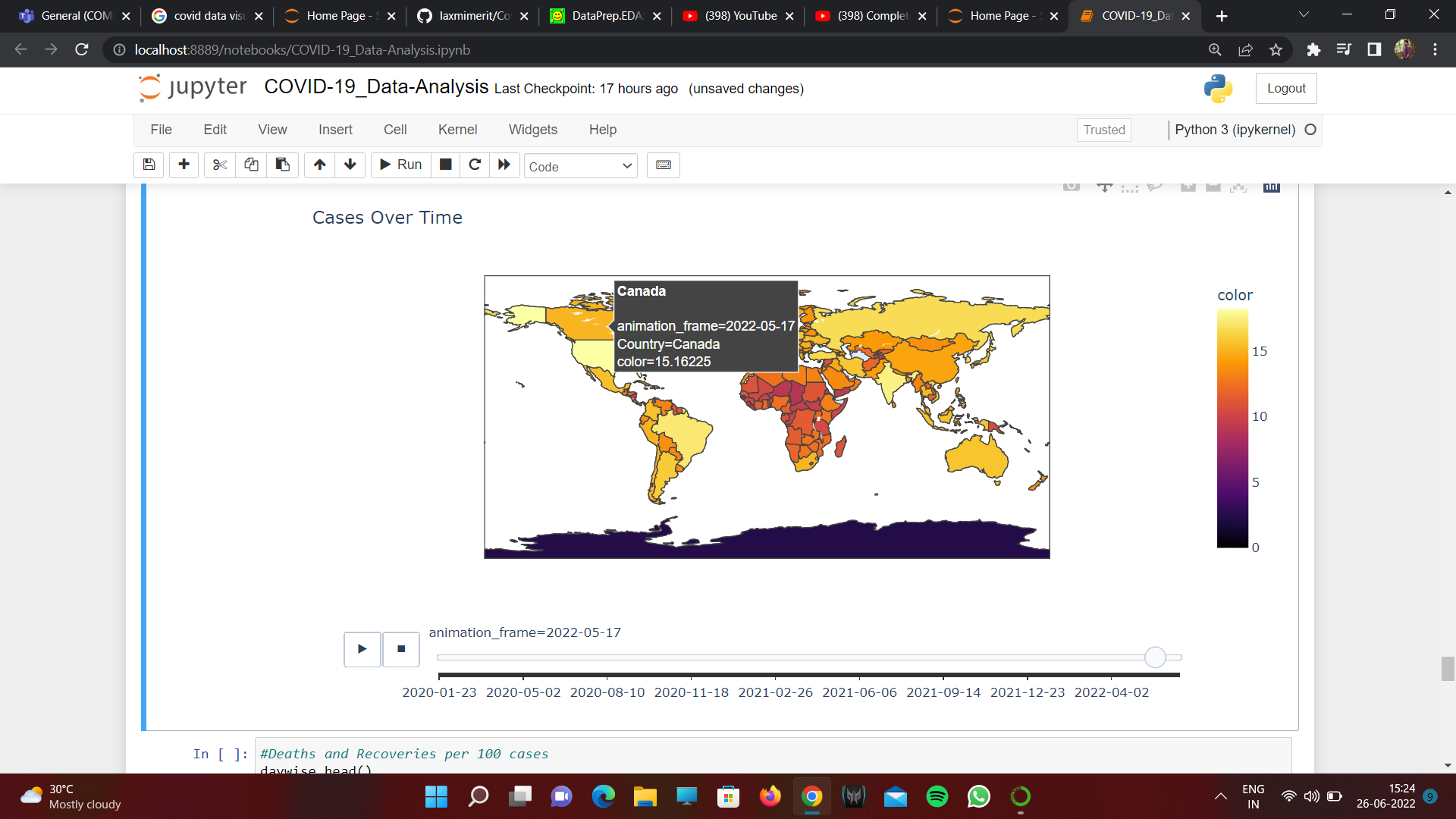
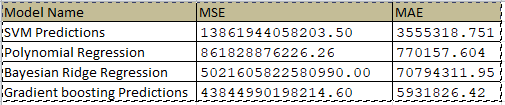


Fig. Confirmed Cases with Choropleth Map

# Data Prediction

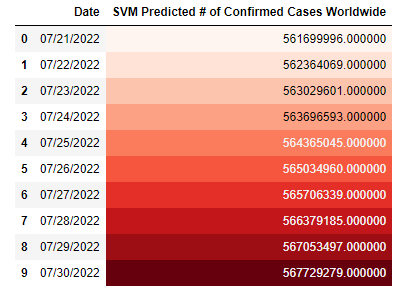
# The data set are first spilt into eighty percent training set and twenty percent testing set and then all the four Regression model is performed to train the 80% set. Here the number of daily active cases is predicted based on daily positive cases. The investigation performs forecasts on infected cases and concurring to results polynomial ridge performs better among all the models. We can see from the graph taken for all the models that the polynomial ridge line is more near to the actual test data.

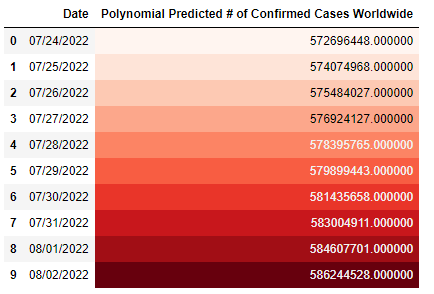
# For all the models we have calculated the MRE and MAE which is the total deviation or error from the actual data and the predicted data from our models per day as per the predictions.

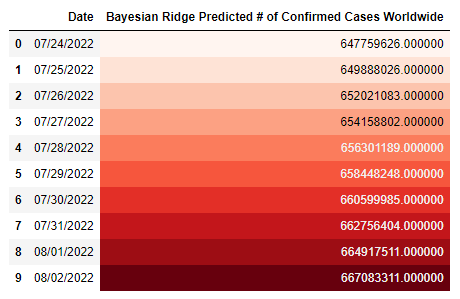


The table shows the Obtained results for Confirmed cases using different regression model. A lower MRE value indicates that data are scattered close to the regression line. From the table also we can see the polynomial regression has gives best fit values of MSE, MAE.

We also predicted the total number of confirmed cases using all these models. The models uses the training data







# Limitations and Challenges:

# Our method is straightforward and only uses historical data to forecast the future. We propose that the following factors i) the gap in testing capabilities of the various countries ii) the severity of the lockdown and the social distance measures iii) the veracity of the number of cases being reported by some countries are responsible for the differences in prediction performance for the various countries.

# There might be a time in future when the dataset used is not updated regularly and hence the project might not give accurate analysis and predictions. Also, Regression models do not automatically take care of nonlinearity. The user needs to imagine the kind of additional terms that might be needed to be added to the regression model to improve its fit. Sometimes SVM model and Bayesian Ridge model might not perform well when there is large dataset because the required training time is higher.

# Conclusion and Future works:

In order to find out a better way to predict the COVID-19 epidemic situation, we adopted four different machine learning methods to process the epidemic data and predict the future predictions of next 10 days. In order to improve the accuracy of the prediction results, we concentrated on MAE/MSE. We employed the algorithm polynomial regression, SVM Model, Bayesian ridge model, and gradian boosting model to plot all the graphs and to make it more optimize. To find out the model fit the training data best. In addition, with dividing the data into three different temporal periods, above mentioned models are selected to fit the data for different portions to make the result more accurate. With calculating MSEs and MAEs, the result is obtained. From our investigation we concluded that polynomial ridge is the best model for predicting confirmed cases

According to the results, there is valuable information to help people, especially the government employees, medical professionals, and businessmen. Through the forecast, people can make rational and feasible plans in order to encourage the restoration of the normal life and serious-affected economies.

Considering the strong variability of the COVID-19, the spread of the virus has shown different characteristics in different periods. Therefore, future research should take more data related to the COVID19 into consideration. At the same time, apart from objective data, people’s attitudes towards the COVID19, subjective emotions, government measures and media reports all have an impact on the spread of COVID-19 to a certain extent.

In future, our aim is to solve the problem to find an optimal partitioning point such that the error becomes minimum.  Others Machine learning algorithms such as BPNN (Back Propagation Neural Network) can effectively quantify these factors, which can also become the focus of future research. In future, we plan to design an interactive website where people can easily view the predictions for the COVID-19 confirmed cases. Additionally, we also plan to extend our models and make them work at the regional/county levels to assist with local governance.

**References:**

[1] Kwok Kin On, Wong Valerie Wing Yu, Wei Wan In, Wong Samuel Yeung Shan, Tang Julian Wei-Tze,"Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong, 13 February 2020",Euro Surveill,2020;25(16):pii=2000155,https://doi.org/10.2807/1560-7917.ES.2020.25.16.2000155

[2] "Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update",https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[3] Hema S. Gopalan, Anoop Misra,"COVID-19 pandemic and challenges for socio-economic issues, healthcare and National Health Programs in India,Diabetes & Metabolic Syndrome: Clinical Research & Reviews",Volume 14, Issue 5,2020,Pages 757-759,ISSN 1871-4021,https://doi.org/10.1016/j.dsx.2020.05.04

[4] Dmitry Ivanov,"Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case,Transportation Research Part E: Logistics and Transportation Review",Volume 136,2020,ISSN 1366-5545,https://doi.org/10.1016/j.tre.2020.101922

[5] Li Cuilian, Chen Li Jia, Chen Xueyu, Zhang Mingzhi, Pang Chi Pui, Chen Haoyu,"Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China", 2020. Euro Surveill. 2020;25(10):pii=2000199. https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000199

[6] A. Raoofi, A. Takian, A. Akbari Sari, A. Olyaeemanesh, H. Haghighi et al., “COVID-19 pandemic and comparative health policy learning in Iran,” Archives of Iranian Medicine, vol. 23, no. 4, pp. 220–234, 2020.

[7] Fanelli D, Piazza F (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 134:109761 [8] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman,… & G. Chowell, “Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020,” Infectious Disease Modelling, 5, 256-263, 2020.

[9] Q. Li, & W. Feng, “Trend and forecasting of the COVID-19 outbreak in China,” arXiv:2002.05866, 2020.

[10] D. Fanelli, & F. Piazza, “Analysis and forecast of COVID-19 spreading in China, Italy and France,” Chaos, Solitons & Fractals,134, 109761, 2020.

[11]C. Massonnaud, J. Roux, & P. Crépey, “COVID-19: Forecasting short term hospital needs in France,” medRxiv, 2020.

[12] P. Liu, P. Beeler, & R. K. Chakrabarty, “COVID-19 Progression Timeline and Effectiveness of Response-to-Spread Interventions across the United States,” medRxiv, 2020.

[13] B. S. Pujari, & S. M. Shekatkar, “Multi-city modeling of epidemics using spatial networks: Application to 2019-nCov (COVID-19) coronavirus in India,” medRxiv, 2020.

[14] R. Gupta, & S. K. Pal, “Trend Analysis and Forecasting of COVID-19 outbreak in India,” medRxiv, 2020.

[15] K. Pandey, & R. Panchal, “A Study of Real World Data Visualization of COVID-19 dataset usingPython,” International Journal of Management and Humanities 4(8):104-107, April 2020.

[16] Seven things countries have done right in the fight against COVID-19. GAVI, The Vaccine Alliance website; Published June 22, 2020. https://www.gavi.org/vaccineswork/seven-things-countries-have-done-right-fight-against-covid-19. Accessed 27 May 2022.

[17] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X (2020) Early transmission dynamics in Wuhan, China, of novel corona virus infected pneumonia. N Engl J Med 382:1199–1207,Article Google Scholar

[18] Wu JT, Leung K, Leung GM (2020) Now casting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395(10225):689–697 Article Google Scholar

[19] Fanelli D, Piazza F (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 134:109761

MathSciNet Article Google Scholar

[20]Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 20:553–558 Article Google Scholar

[21]Zhuang Z, Zhao S, Lin Q, Cao P, Lou Y, Yang L, He D (2020) Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: a modelling analysis based on overseas cases and air travel data. Int J Infect Dis 94:29–31Article Google Scholar

[22]Ghosal S, Sengupta S, Majumder M, Sinha B (2020) Prediction of the number of deaths in India due to SARS-CoV-2 at 56 weeks. Diabetes Metabo Syndr Clin Res Rev 14:311315. https://doi.org/10.1016/j.dsx.2020.03.017 Article Google Scholar

[23]Kavadi MDP, Patan R, Ramachandran M, Gandomi AH (2020) Partial derivative nonlinear global pandemic machine learning prediction of covid 19. Chaos Solitons Fractals 5:110056. https://doi.org/10.1016/j.chaos.2020.110056 MathSciNet Article Google Scholar

[24] Senapati, A., Nag, A., Mondal, A. et al,"A novel framework for COVID-19 case prediction through piecewise regression in India" Int. j. inf. tecnol. 13, 41–48 (2021). https://doi.org/10.1007/s41870-020-00552-3

[25] Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons & Fractals. https://doi.org/10.1016/j.chaos.2020.109850.

[26] Satpathy, S., Mangla, M., Sharma, N. et al."Predicting mortality rate and associated risks in COVID-19 patients"Spat. Inf. Res. 29, 455–464 (2021). https://doi.org/10.1007/s41324-021-00379-5

[27] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020)"Application of the ARIMA model on the COVID-2019 epidemic dataset" Data in Brief, 29, 105340. https://doi.org/10.1016/j.dib.2020.105340.

[28] Khan, N, Arshad, A, Azam, M, AL-Marshadi, AH, Aslam, M,"Modelling and Forecasting the Total Number of Cases and Deaths Due to Pandemic" J Med Virol. 2022; 94: 1592- 1605. doi:10.1002/jmv.27506

[29] Yan, Li & Zhang, Hai-Tao & Xiao, Yang & Wang, Maolin & Sun, Chuan & Liang, Jing & Li, Shusheng & Zhang, Mingyang & Guo, Yuqi & Xiao, Ying & Tang, Xiuchuan & Cao, Haosen & Tan, Xi & Huang, Niannian & Jiao, Bo & Luo, Ailin & Cao, Zhiguo & Xu, Hui & Yuan, Ye. (2020),"A machine learning-based model for survival prediction in patients with severe COVID-19 infection" https://doi.org/10.1101/2020.02.27.20028027.

[30] Smita Rath, Alakananda Tripathy, Alok Ranjan Tripathy,"Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model",Volume 14, Issue 5,2020,Pages 1467-1474,ISSN 1871-4021,https://doi.org/10.1016/j.dsx.2020.07.045

[31] Mhdm, R., Silva, R. G., Mariani, V. C., & Coelho, L. S. (2020),"Short-term forecasting COVID-19 cumulative conrmed cases: perspectives for Brazil" Chaos, Solitons & Fractals., 135(2020), 109853. https://doi.org/10.1016/j.chaos.2020.109853.

[32] G. Pandey, P. Chaudhary, R. Gupta and S. Pal,"SEIR and Regression Model based COVID-19 outbreak predictions in India", arXiv preprint,2020

[33] Z. Li, S. Yang and J. Wu, "The Prediction of the Spread of COVID-19 using Regression Models," 2020 International Conference on Public Health and Data Science (ICPHDS), 2020, pp. 247-252, doi: 10.1109/ICPHDS51617.2020.00055.

[34] Johns hopkins COVID-19 github data repository. https://github.com/CSSEGISandData/ COVID-19, 2020.

[35] Scikit Learn - Bayesian Ridge Regression, https://www.tutorialspoint.com/scikit\_learn/scikit\_learn\_bayesian\_ridge\_regression.htm

[36] Polynomial regression, https://en.wikipedia.org/wiki/Polynomial\_regression

[37] Gradient boosting, https://en.wikipedia.org/wiki/Gradient\_boosting

[38] sklearn.svm.SVR, https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html