

Review Article on Covid-19 Outbreak Prediction Using Machine Learning Algorithm

Surya Gouni Neha¹, Bala Sai Allagadda²

^{1,2}Research Scholar, Malla Reddy College of Engineering and Technology, Medak

Abstract - Towards Data Science is a medium distribution essentially dependent on the investigation of information science and AI. We are not wellbeing experts or disease transmission specialists, and the assessments of this article ought not be deciphered as expert wellbeing exhortation. Nonetheless, this article will be center around how AI can be utilized to anticipate the spread of the pandemic. This paper is utilized to decide information investigation on COVID 19 by AI procedures. the primary methods are addressing the accompanying piece of this paper. This paper also represents the regression model analysis for determine the patient records-based data analysis for COVID 19 DATASET.

Index Terms – Machine Learning Techniques, COVID 19 data set, regression techniques, classification algorithms, etc.

I. INTRODUCTION

Our general public is in the time of fantastic endeavors to battle upon the spread of this dangerous condition as far as foundation, money, business, fabricating, and a few different assets. Man-made consciousness (AI) analysts fortify their capability in creating numerical ideal models for exploring this pandemic utilizing cross country circulated information. This article means to apply the AI models all the while with the estimate of expected reachability of the COVID-19 over the countries by utilizing the continuous information.

II. MACHINE LEARNING

A typical meaning of machine learnedness is: "A PC program is said to gain for a fact E as for some class of undertakings T and execution measure P if its exhibition at errands in T, as estimated by P, improves with experience E." Basically, machine

savviness is the capacity of a PC to gain for a fact. Experience is typically given as information. Taking a gander at this information, the PC can discover conditions in the information that are excessively unpredictable for a human to shape. Machine learnedness can be utilized to uncover a secret class structure in an unstructured information, or it tends to be utilized to discover conditions in an organized information to make forecasts.

III. COVID 19 DATA SET ANALYTICS CHALLENGES IN EDUCATION

The audit of the writing uncovered the LA challenges about information following, information assortment, and information examination, an association with intelligence sciences, knowledge climate advancement, arising innovation, and moral concerns in regard to lawful and security issues. This information comprises of day-by-day case reports and day by daytime arrangement synopsis tables. In the investigation, we have chosen time-arrangement outline tables in CSV design having three tables for affirmed, passing, and recuperated instances of COVID-19 with six properties. For instance, area/state, country/locale, last update, affirmed, demise, and recuperated cases. The CSV information are accessible in GitHub archives.

Covid 19 spreads are sorted into four phases. The primary stage begins with the cases recorded for individuals who made a trip to or from influenced nations or urban communities, while in the subsequent stage, cases are accounted for territorially among family, companions, and gatherings who came into contact with the individual coming from the influenced nations. Subsequently, the influenced individuals are recognizable. Then, the third stage

causes the condition seriously as the tainted individual gets imperceptible and straightens across the people who neither have any movement records nor came regarding the influenced individual. This condition obliges prompt lockdown the country over to lessen the social contacts between people to quantify the development of the infection. At last, stage four beginnings when the transmission converts to endemic and wild. China is the principal country that felt under stage four of the COVID-19 transmission, while the majority of the created nations are currently in this phase of the transmission and bearing a further number of pestilences and misfortunes contrasted with China.

AI calculations assume a fundamental part in the pandemic examination and anticipating. Besides, AI procedures help to uncover the plague designs. Accordingly, a quick reaction may be set up to forestall the spread of the infection (Kalipe, Gautham and Behera, 2018¹; Singh, Singh and Bhatia, 2018²). Also, AI models are used to perceive aggregate conduct along with the expectation of the normal spread of the COVID-19 across the general public by utilizing the ongoing information.

IV.COVID 19 DATA ANALYTICS USING TOOL AND TECHNIQUES

The challenges faced in processing Big Data technologies are overcome by using various techniques. The most popular techniques used in educational data mining is listed below.

Regression – Regression is used in predicting values of a dependant variable by estimating the relationship among variables using statistical analysis.

Nearest Neighbor – In this technique the values are predicted based on the predicted values of the records that are nearest to the record than needs to be predicted.

Clustering – Clustering involves grouping of records that are similar by identifying the distance between them in an n-dimensional space where n is the number of variables.

Classification – Classification is the identification of the category/class to which a value belongs to, based on previously categorized values.

Open-Source Tools:

Several Open-source tools exist which help in taming Big Data [9] some of the top tools are listed below.

Mongo DB is a cross platform document-oriented database mgmt. system. It uses JSON like documents instead of a table-based architecture.

Hadoop is a framework that allows distributed processing of big datasets across clusters of networked computers using simple programming models.

V.MACHINE LEARNING TECHNIQUES: COVID 19 DATA ANALYSIS SYSTEMS

V.I. REGRESSION MODELS

Regression models involve the following variables:

- The unknown parameters, denoted as B, which may represent a scalar or vector.
- The independent variables, denoted as X.
- The dependent variable, denoted as Y.

In various fields of application, different terminologies are used in place of dependent and independent variables.

A regression model relates Y to a function of X & B.

$$Y=f(X, B)$$

The approximation is usually formalized as $E(Y/X) = f(X, B)$. To carry out regression analysis, the form of the function f must be specified. Sometimes the form of this function is based on acquaintance about the relation between Y & X that does not rely on data. If no such acquaintance is available, a flexible or convenient form of f is chosen.

Assume now that the vector of unknown parameters B is of length k. In order to perform a regression analysis, the user must provide information about the dependent variable Y:

Formula of applied regression analysis:

$$Y \text{ dependent variable} = F. (x \text{ independent variable}, \theta) + e.$$

$$Y = \text{COVID ANALYSIS} = F.X1\text{PATIENT RECORDS} + X2.\text{POSITIVE PERCENTILE} + X3.\text{DEATH RATIO} + X4.\text{RECOVER PATIENT}$$

RECORDS+X5.ADMIITED HOSPITALIAZE RATO
OF PATIENT++_____ -Qn.

VI.PREDICTION AND ANALYSIS

Coronavirus spread has conducted the society under the edge of loss in social lives. Additionally, it is crucial to investigate the transmission growth ahead and predict the future occurrences of the transmission. In concurrent, state-of-the-art mathematical models are chosen based on machine learning for a computational process to predict the spread of the virus, for instance:

- Support Vector Regression⁵ (SVR)
- Polynomial Regression⁶ (PR)
- Deep Learning regression models
- It is also involving:
- Artificial Neural Network⁷ (ANN)
- Recurrent Neural Networks⁸ (RNN) using Long Short-Term Memory⁹ (LSTM) cells.

Machine learning and deep learning strategies are performed using the python library to predict the total number of confirmed, recovered, and death cases extensively. This prediction will allow undertaking specific determinations based on transmission growth, such as expanding the lockdown phase, performing the sanitation plan, and providing daily support and supplies.

COVID DATA ANALYSIS: predicted by regression model. This model represents the data analysis by simple linear multiple regression. the overall percentile ratio is used to determine the covid 19. Data set records and analysis determined.

VII.CONCLUSION

This article intended to employ the machine learning models for pandemic analysis through a dataset from Johns Hopkins. In conclusion, the method of Polynomial Regression (PR) generated a minimum Root Mean Square Error (RMSE) amount over other methods in projecting the COVID-19 transmission. However, if the spread mimics the prognosticated trend of the PR model, then it would lead to extensive loss of lives as it presents the incredible growth of the transmission globally. As perceived in China, the

increased case of COVID-19 can be degraded by lessening the number of sensitive individuals from infected people. This new normal is obtainable by becoming unsocial and supporting the lockdown regulation with control.

REFERENCE

- [1] Zumla A, Hui DS, Azhar EI, Memish ZA, Maeurer M (2020) Reducing mortality from 2019-nCoV: host-directed therapies should be an option. *Lancet* 395(10224):e35–e36. [https://doi.org/10.1016/S0140-6736\(20\)30305-6.9-4393-y](https://doi.org/10.1016/S0140-6736(20)30305-6.9-4393-y)CrossRefGoogle Scholar
- [2] Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)Google Scholar
- [3] Guan W, Ni Z, Hu Y et al (2020) Clinical characteristics of coronavirus disease 2019 in China. *New Engl J Med*. <https://doi.org/10.1056/NEJMoa2002032>Google Scholar
- [4] Shao Y, Wu J (2020) IDM editorial statement on the 2019-nCoV. *Infect Dis Model* 5:233–234Google Scholar
- [5] Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, Mizukami Y, Tadono T (2019) Spatiotemporal dengue fever hotspots associated with climatic factors in taiwan including outbreak predictions based on machine-learning. *Geospatial Health* 14:183–194. <https://doi.org/10.4081/gh.2019.771>CrossRefGoogle Scholar
- [6] Liang R, Lu Y, Qu X, Su Q, Li C, Xia S, Liu Y, Zhang Q, Cao X, Chen Q et al (2020) Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transboundary Emerg Dis* 67:935–946. <https://doi.org/10.1111/tbed.13424>CrossRefGoogle Scholar
- [7] Tapak L, Hamidi O, Fathian M, Karami M (2019) Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC*

- Res Notes 12. <https://doi.org/10.1186/s13104-019-4393-y>
- [8] Raja DB, Mallol R, Ting CY, Kamaludin F, Ahmad R, Ismail S, Jayaraj VJ, Sundram BM (2019) Artificial intelligence model as predictor for dengue outbreaks. *Malays J Public Health Med* 19:103–108 [CrossRef](#) [Google Scholar](#)
- [9] Iqbal N, Islam M (2019) Machine learning for dengue outbreak prediction: a performance evaluation of different prominent classifiers. *Informatica* 43:363–371. <https://doi.org/10.31449/inf.v43i1.1548> [CrossRef](#) [Google Scholar](#)
- [10] Koike F, Morimoto N (2018) Supervised forecasting of the range expansion of novel non-indigenous organisms: Alien pest organisms and the 2009 H1N1 flu pandemic. *Global Ecol Biogeogr* 27:991–1000. <https://doi.org/10.1111/geb.12754> [CrossRef](#) [Google Scholar](#)
- [11] Chenar SS, Deng Z (2018) Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ Int* 111:212–223. <https://doi.org/10.1016/j.envint.2017.11.032> [CrossRef](#) [Google Scholar](#)
- [12] Chenar SS, Deng Z (2018) Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res* 128:20–37. <https://doi.org/10.1016/j.watres.2017.10.032> [CrossRef](#) [Google Scholar](#)
- [13] Riou J, Althaus CL (2020) Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro surveillance* 25(4) [Google Scholar](#) [Informed Technology](#). IOS Press, Amsterdam, The Netherlands, pp. 467-474.
- [14] Mostafa Sabbaghi (2006) “Using Feature Selection and Unsupervised Clustering to Identify Affective Expressions in Educational Games”, in *Proc. Of The Intelligent Tutoring Systems Workshop on Motivational and Affective Issues*. pp. 21-28.