

Use of Data Science during Worldwide Pandemic : An Efficient and Reliable System to Support Decision Making in Different Sectors

Tithi Choudhary¹, Muskan Shaikh¹, Shreya Khedekar¹, Richa Sirwani¹, Dr. Meenakshi A. Thalor²

¹Information Technology, B.E, AISSMS's Institute of Information Technology, Pune, Maharashtra, India

²Information Technology, HOD, AISSMS's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT

Article Info

Volume 8, Issue 2

Page Number : 17-27

Publication Issue

March-April-2021

Article History

Accepted : 01 March 2021

Published : 04 March 2021

Any calamities or emergency situations have created drastic and everlasting impacts on mankind since history. Such situations need to be managed in an efficient and effective manner. There are countries of the world where the use of evolving technology is not a part of the management system. In this paper using a combination of technologies and tools we will provide a better alternative solution to the intuitive decisions thus making systems more reliable. Data Mining, Data Analytics, Data Visualization and Decision making have key applications demonstrated in this paper. Upon the choice of User Selections, this system will look into the number of deaths and recoveries that have occurred everyday from the start of spread of the COVID-19. Based on this information, the goal will be to analyze and formulate a death and recovery analysis, which will in turn help us in understanding the effect of corona virus on different work sectors, which in our case will be the Business sector of the country. The system will also focus on forecasting the GDP based on the past datasets of the country. Different visualization methods will be shown according to the Automated Machine learning system that will generate statistical depiction from the inflow of dynamic data. These info graphic visuals will provide a better understanding to the user as to how the situation will affect her own sector/domain as well as current stance of the GDP and enhance the overall User Experience (UX) of the user.

Keywords : Data Mining, Data Analytics, Data Visualization and Decision Making, Automated Machine Learning, Prediction, UI, UX.

I. INTRODUCTION

The world has been greatly affected by any disasters, pandemics, epidemics, wars, calamities etc. These

situations have created major disturbances in the life of people and have created life changing impacts.

The aim is to develop a web application data intelligently and thus minimize the negative effects of the situation by educating the user of the application

with info-graphic details according to the sector the user is a part of. The system provides an insight into the real-time details of the pandemic, in our case COVID -19 and thus shows the tendency of the positive outcomes on making the smartest calculated decision relatively.

Data from the most relevant times of covid-19 spread is analyzed and predictions are used for pattern detection which makes the system learn from the past and ensure the machine learning for the crisis under consideration leads to trusted outputs maximizing the benefits of disaster management. The system works with dynamic data fed to the application at runtime supported by Automated Machine Learning.

The sections III, IV, V, and VI i.e., Proposed Work, Implementation, Methodology and Experimental Analysis of this paper provide useful details and focuses on all necessary and relevant information on the system development and execution process.

II. LITERATURE SURVEY

Authors Gunther Schuh et al. (2019) [37] describes existing approaches focusing on structuring the Data Mining domain so as to provide decision making to choose suitable Data Mining methods for the user.

Authors Jaya Bajpai et al. (2016) [5] explains the problems in Data Cleaning process and presents data cleaning approaches involving different phases such as Data Analysis, Definition of Transformation Work Flow and Mapping Rules, Verification, Transformation, Back Flow Process of Cleaned Data. The tools used for the process of transformation and Data Cleaning have been mentioned.

Authors M. J., Price et al. (2014) [13] describes the use of time series models such as ARIMA and Random Forest for prediction. A comparative study is presented which resulted in the fact that Prospective

Random Forest model outperformed the ARIMA model over a specified time period.

Authors Zhang, L et al. (2020) [15] proposes a method for time series forecasting known as LSTM networks. Long-short-term-memory (LSTM) also known as RNN with blocks, are found to be efficient algorithms to build a time series sequential model. LSTM helps in processing information with the help of activation function and the output comes to be between 0 or 1.

Authors H. R. Bhapkar et al. (2020) [6] proposed a new method for understanding the pattern of recovery and death rates by including the recovery time period.

Authors Mahipal Jadeja et al. (2015) [29] present the use of Tree Maps in Data Visualization and transition from traditional representation methods to Tree-Maps. Tree-Maps are widely used to express a variety of nested as well as hierarchical data and data structures. These maps prove to be very useful in finding trends in large data sets.

Along with the Tree Maps, other Data Visualization Techniques such as Scatterplot, Sparkline, Pie Chart, Gauge, Heat Map, Histogram, Box Plot etc. are also used to visualize the data by matching the data to right information.

III. PROPOSED WORK

The effect of pandemic has been negative on the economy across the world. During the Spanish influenza in 1918, the consequences on the business sector in the US and Europe were devastating. The overall economic loss incurred to the US business market was huge and many businesses had to shut down. The similar effects can be seen in 2020's covid-19 pandemic. The total loss in the global economy is predicted up to \$1Tn. The graph below shows the losses suffered by the share market in various

countries between January 10 2020 and March 16 2020. Fig.1.

Stock Markets Feel The Heat

US, European, Hong Kong and Indian stock markets suffered sharp losses between January 10 and March 16

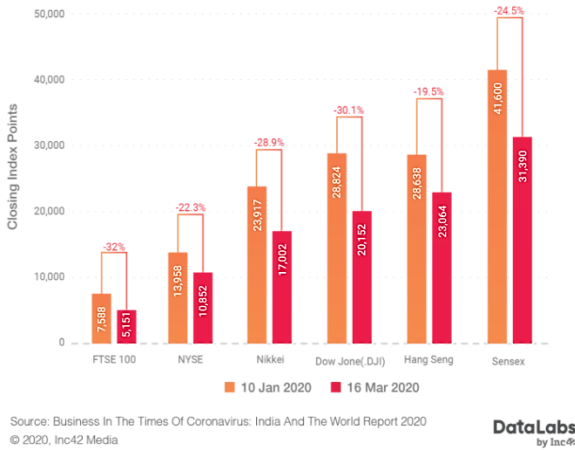


Fig.1.Share Market losses of various countries

This shows how much impact this pandemic has on the economy of countries. The most affected sector is the business sector. According to an article by BBC, there are certain countries that are trying their best to survive this economic crisis (Denmark, Singapore, The US, Rwanda and New Zealand).

The system proposed in this paper, will help various sectors to reduce loss and be resilient to any such situation. The data generated during this pandemic are brought together to get insights and propose better ways for businesses to minimize the effects of macro environmental factors. During the lockdown imposed to control pandemic, many small businesses had to close down, many are still trying to manage and maintain their position in the market. If these businesses are given proper assistance about what, when and where to invest their time and money, it would result in improvised situations.

IV. IMPLEMENTATION

The flow chart depicts the basic workflow of the system architecture. The first step is to attain the

relevant datasets for the research and analysis. The datasets aimed are of dynamic dataset of COVID-19, that will help us analyze the recovery and death stats as per the situation; the economic growth and downfall in business sectors (if observed) dataset in that period, etc. The data is achieved from various sources like social media, health departments, economical departments, hence the data is never clean or ready for instantaneous use, it needs to be cleaned using data cleaning techniques. The data cleaning techniques are specified.

Analyzing the past up till present data to get better insights for future development that will eventually help to get a brief overview of trends and patterns where macro environmental factors are affected, composing further decision making easier. The application would start with depiction of recovery and death rates during the pandemic till date by using a calculative formula as proposed by Dr. Bhapkar [6]. The aim is to use this method rather than going with traditional method of statistic calculations to procure a finer and more accurate rate. Meanwhile, per day recovery and death cases would also be represented in info-graphical format. Using automated machine learning algorithms to training and testing of datasets for statistical depiction of dynamic inflow of data regarding Gross Domestic Product (GDP) will also be done. Further, the prediction and analysis procedures to forecast the same will be followed. In addition to this system, another info-graphic representation of businesses in every state of the country, how the economy has been respectively. This will give the user an insight as to how the conditions are for business in their region. Once the models are defined and their feasibility gets established, we will be able to develop an apt user interface with the defined methods connected at their backend.

The goal is to let the user review the outcomes for next few days, how the rate of virus spread will be, whilst analyzing the situation for them as to how they

can keep their business steady. For ease of operation and understanding, the analysis will be described in visual forms. This will help the clientele to make a thought-through decision for their purposes.

The smooth inflow of data is ensured using Redux-saga as middleware which will help note the state change in the application. Testing is done using Junit test cases. The ultimate application will assist the client in the emergency situation or calamity that has extreme effects on the world using the past, present data and visualization in least complicated ways, for the measures and better decisions that can be taken to avoid future damage to their business. The macro environment effects on any trade can result in declination of its economic growth, providing assistance in these situations in a user-friendly way is the consequent outcome of this application.

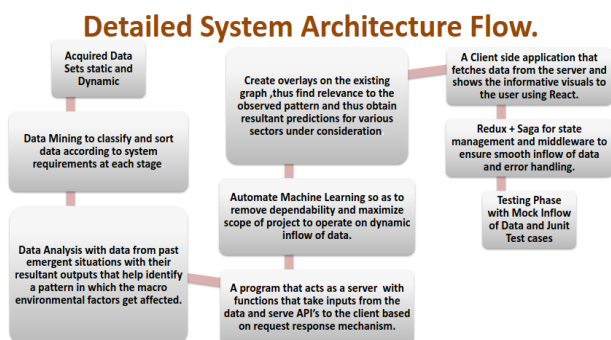


Fig. 2. Implementation Flow Chart

V. METHODOLOGY

Data Cleaning & Mining

Data cleaning is a technique or a process used for eliminating any errors and inconsistencies in the data to upgrade the data quality [1]. Considering the current situation, the data sets regarding any pandemic are huge and changes in the trends take place gradually. This bulk source of data contains patterns of uncertainties that may hinder the further process of prediction [5]. In order to avoid such situations, we need various methods of cleaning. Depending upon the need of the data sets, there are

few tools (ETL tools) for data cleaning among which the one that fits well must be used: COPYMANAGER (Information Builders), DATASTAGE (Informix / Ardent), EXTRICATE (ETI), POWERMART, DECISIONBASE (CA/Platinum).

These are more advanced methods of data processing; and instead, we can make use of some simpler means of cleaning of data, like clearing null values of certain columns by either discarding the data, which may lead to major loss. Hence another option of normalizing the data by means of statistics is also available. We can substitute the null values with the mean, median or mode of the column. This will not only give us a precise value of calculation, but will also not lead to data loss overall.

Data Analytics(Favourable algorithm prediction)

The application requires two major analytical approach to represent the per day recovery to death ratio as well as prediction of GDP of the country. Hence for these actions, certain research on types of algorithm and methods was required. In the current case-scenario, a very less amount of data is available. Due to this severely fast developing epidemic conditions, our model needs to be working among high uncertainties. Thus, in order to select the most appropriate and accurate algorithm, we will be making use of an extensive selection method, named as Group of Optimized and Multisource Selection, abbreviated as GROOMS [11]. This method is specifically designed in order to achieve highest rates of prediction with limited resource. This design is being used for group forecasting, by making use of algorithms, some of which are capable of taking multiple sources of input. After comparison among various optimized methods, the one that produces highest prediction accuracy will be implemented for defining the forecast on rate of growth/decline in spread of the virus.

a) Group of Optimized and Multi-source Selection
In this method, major three step process is implemented. Step one starts with the data set, no matter how small or large in scale, is being fed to the

collection of models defined by us. In the next step, each candidate model will present its best performance, either by optimizing the parameters, or by including other data sources. The characteristics of algorithms being used in this method are roughly categorized as: non-parametric (e.g., linear/logistic regression, simple statistics) that may not require fixture of parameters: parametric model (e.g. activation functions for neural networks, splitting criteria for decision tree and random forest classifiers); and dual models whose performances are sensitive to both types of input sources as well as selection of parameters. Step wise allocations for GROOMS are shown in Fig. 3.

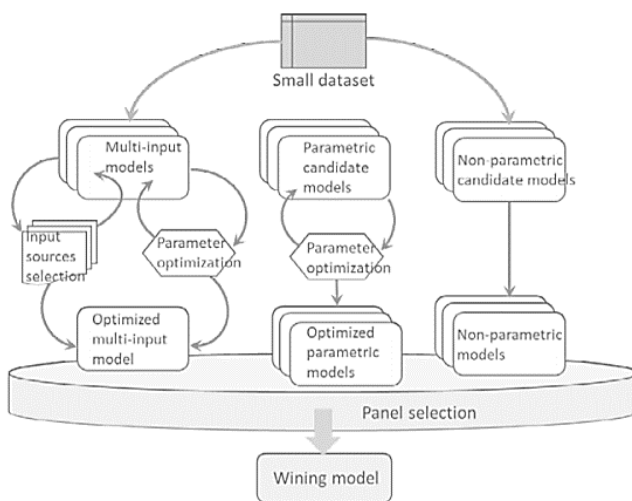


Fig. 3. GROOMS top to bottom approach.

All though GROOMS methodology analyses algorithms of various segregation, we chose this method to analyse various linear models of regression itself.

The linear regression models that were experimented were as follows: ElasticNet CV, Linear Regression, SGD Regressor, Bayesian Ridge, Lasso Lars, ARD Regression, Passive Aggressive Regressor and TheilSen Regressor.

Data Visualization(Infographic Visuals of analysis and prediction)

Visualizing information in the form of graphs, charts, diagrams that help getting a better overview of the large data is called Data Visualization. We need data visualization because the visual summary of any information makes it easier to understand the trends

and patterns of data.[31] Analysis of data is done to get insights from data, visualization adds to the betterment of data analysis. Data visualization of unstructured data does not give good insights hence it is necessary to have clean data, which helps in visualization.[33]

Types of data visualization techniques used in this system are as follows:

- a) Line Chart
- b) Tree Map
- c) Polar Chart

a) Line Chart:

Line chart is widely used for visualizing patterns, trends and fluctuations in data that changes continuously. We can easily compare two or more datasets that are different yet have similar trends [27]. Through this representation, we can easily display the per day recoveries and deaths happening. From python programming perspective, there are many ways of plotting the graphs, some statically, and some animated in nature. One such library of graph, known as plotly is being used in the system. The graph_objects module is being used for representing the recovery death analysis, as well as showing the prediction of GDP.

b) Tree Map:

Tree is used to display hierarchical data while Tree map is used to display hierarchy and compare values between categories and subcategories.[29] The concept of tree maps is simple: allocate one box for the parent node and children of the node are drawn as boxes within it.[34] Displaying the effect of lockdown in the metro cities of a country using tree maps. This type of representation would be effective for showing business economy state-wise in the applications case.

c) Polar Chart:

Polar charts are similar to pie charts but the angles measure the same and there is a change in radius depending on the value.[35] Polar charts here will show whether the certain sector should open for

business and will it earn profits amid lockdown or is it better to keep low and not add-up extra costs. This is based on data received in the past and government guidelines at that time.[30] The revenues earned by sectors in the past within a time span of 15 days are considered to depict the decision. The tree map below shows the after ease of lockdown some sectors were able to earn revenue. Fig. 4.

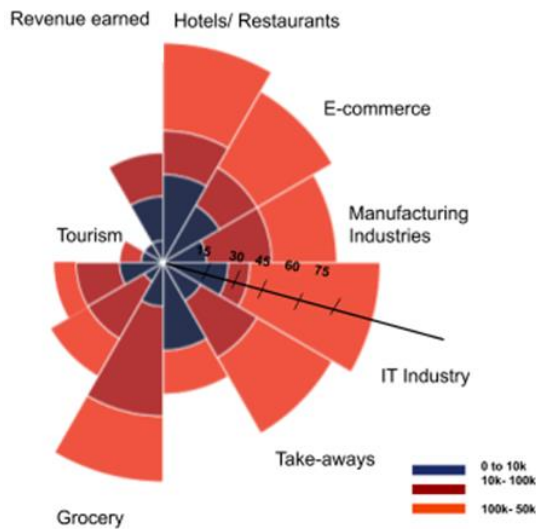


Fig. 4. Polar Chart showing the revenue earned by various sectors during Lockdown.

VI. EXPERIMENTAL ANALYSIS

The first step in the application is to have a calculative source, so as to display the recovery and death rate analysis during this pandemic. For that to happen, the main initial process is to find a relative dataset, and looking into the cleaning process of it. The dataset being used for recoveries vs deaths analysis is from World Bank Group [10]. The dataset for GDP analysis is from Macrotrends website that indulge in providing historical and present data for various aspects [9]. In both the cases, the cleaning and processing of data is being done by simpler methods of statistics. The null values in a column are substituted with mean values of the entire column, giving an even palette to work on. The next step is calculating the rate of recoveries and deaths occurred during the pandemic. After cleaning and normalization of data, we get the following representation of data as shown in table 1.1, with certain columns defined, that are useful for the inspection. For this process, we make use of a formula as defined by Dr. Bhapkar in his searches. According to the formula, since mortality rate changes over time during COVID-19 pandemic, we call it Progressive

S.No.	Date_YMD	Daily Confirmed	Total Confirmed	Daily Recovered	Total Recovered	Daily Deceased	Total Deceased	
0	30 Jan	2021-01-30	1	1	0	0	0	
1	31 Jan	2021-01-31	0	1	0	0	0	
2	01 Jan	2021-02-01	0	1	0	0	0	
3	02 Jan	2021-02-02	1	2	0	0	0	
4	03 Jan	2021-02-03	1	3	0	0	0	
.	
.	
.	
353	17 Jan	2021-01-17	13962	10572599	14513	10210569	145	151865
354	18 Jan	2021-01-18	9987	10582586	17127	10227696	137	152002
355	19 Jan	2021-01-19	13787	10596373	17229	10244925	161	152163
356	20 Jan	2021-01-20	15279	10611652	20071	10264996	152	152315
357	21 Jan	2021-01-21	14490	10626142	17726	10282722	161	152476

S. No.	Date_Y MD	Daily Confirmed	Total Confirmed	Daily Recovered	Total Recovered	Daily Deceased	Total Deceased	Recovery_r	Recovery_rate	Fatality_r	Fatality_rate
353	2021-01-17	13962	10572599	14513	10210569	145	151865	96.87	96.58	1.44	1.44
354	2021-01-18	9987	10582586	17127	10227696	137	152002	97.03	96.65	1.44	1.44
355	2021-01-19	13787	10596373	17229	10244925	161	152163	97.20	96.68	1.44	1.44
356	2021-01-20	15279	10611652	20071	10264996	152	152315	97.39	96.73	1.45	1.44
357	2021-01-21	14490	10626142	17726	10282722	161	152476	97.55	96.77	1.45	1.43

Table 1.1. Recovery and Death Records

S.No.	Algorithm	MSE	R2
0	Linear Regression	2261.878252	0.995431
1	ElasticNet CV	2310.806511	0.995332
2	SGD Regressor	3268.581138	0.993398
3	Bayesian Ridge	2320.479117	0.995313
4	Lasso Lars	2077.670415	0.995803
5	ARD Regression	1834.062541	0.996295
6	Passive Aggressive Regressor	6126.077982	0.987626
7	TheilSen Regressor	1822.249353	0.996319

Table 1.3. MSE and R² score comparison for best fit algorithm selection

Mortality Rate (PMR). In general, for any case/pandemic, C disease and N_{Pavg} refers to a total number of infections before average recovery time period P_{avg} [6].

Similarly, Recovery Rate (RR) used to follow classical mathematical equation, $RR = \frac{R}{N} \times 100$, where R refers to N total number of recovered cases. As before, when recovery time period is taken into account, Progressive Recovery Rate (PRR) is,

$$PRR = \frac{R_c}{N_{Pavg}} \times 100 ;$$

where R_c refers to total number of recovered cases till date [6].

According to the analysis, the general recovery time is 14 days. But in our case, we make the P_{avg} modification to 15 days for calculative simplicity. Based on the formula described above, an algorithm for the rate calculation is formulated.

In the initial stages, we make use of simple method of rate analysis just by division of deaths caused and total number of infections recorded. Alongside, we also implement Dr. Bhapkar's method of analysis. The results for both are projected below in table 1.2. The recovery and death rate by simple formula are represented by keywords 'recovery_rate' and 'deceased_rate'; whereas rate calculated by the progressive method are shown as 'recovery_r' and 'deceased_r'.

On establishing a visualization in the python programming itself, we implement a representation of recoveries and death per day, in comparison to the total confirmed cases of any given day.

The main purpose of the deceased and recovered rates are to understand the impact of COVID-19 on business sector, by looking into the mortality against the disease. The aim is that while the user can view the per day numbers of confirmed cases, along with recoveries and deaths, they would also be able to see the rate with which the recoveries are occurring in the UI. It will be precise as we are taking in consideration the 15 days recovery period.

The second functionality of the application is analysing the GDP and forecasting it. For this process, as mentioned in V section of the paper, we make use of a methodology that will help us decide the best fit model for the prediction, i.e., GROOMS (Group of Optimized and Multi-source Selection). In this, we make use of linear regression algorithms for smooth inflow of data during the visual representation of the forecasting. Not to mention, instead of using complex algorithms of ML and Deep Learning like SVM, K-Means or ANN that may lead to overfitting of data which is already not large in size, it's easier to use simpler methods for analysis.

For GROOMS method of selection, we chose algorithms such as ElasticNet CV, Linear Regression, SGD Regressor, Bayesian Ridge, Lasso Lars, ARD Regression, Passive Aggressive Regressor and TheilSen Regressor. We implement training and testing on the same data for all the algorithms. As mentioned earlier, the data being used is from Macrotrends [9]. After testing for all the methods, we look into the Mean Squared Error values and R^2 scores of each of them, which lets us finally choose the desired method of implementation. The lowest the MSE value and the highest

the R^2 values is, the better the algorithm. A tabulation of the algorithms implemented along with their MSE and R^2 scores are represented in table 1.3.

From the above observation, we can see that TheilSen Regressor produces minimum MSE and maximum R^2 score. Thus, the best fit model for the prediction is considered to be TheilSen Regressor.

VII. CONCLUSION

The paper thus, provides an informative insight into the real world application of data science and analytics. The demonstrations of problem solving using a combination of different methods and tools and how they will be beneficial for the business sector. Analyzing the effect of covid-19 by looking into the aspects of recovery and death rates, GDP of the country gets affected, not to mention forecasting the succession of GDP.

These insights will not only provide information to its business users, but will also be easy to understand due to its highly info-graphic visual representation.

VIII. SCOPE OF IMPROVEMENT

To make the system completely self-reliant and accurate, there may be additions to the set of tools and technologies used. The advanced versions, updated algorithms that are successful on boundary conditions and exceptions may be used for newer versions of the system with time. A major breakthrough could be applying the system for various other job sectors such as education, banking and finance as well as other tertiary sectors. These solutions can be implemented for any kind of calamities, as well as pandemic situations. Business Intelligence powered with AI will serve as a decision support system, that will help user make certain decisions for their business according to the type and give them fruitful results.

IX. ACKNOWLEDGMENT

We would express our sincere gratitude towards our project guide Dr. Meenakshi Thalor for her extensive support and expertise on the project research, development and execution.

X. REFERENCES

- [1]. Xuefang Xu, Yaguo Lei, Zeda Li(2020), "An Incorrect Data Detection Method for Big Data Cleaning of Machinery Condition Monitoring" published in IEEE transactions on industrial electronics.
- [2]. Mohammad Mahdavi, Felix Neutatz, Larysa Visengeriyeva, and Ziawasch Abedjan,(2019), "Towards Automated Data Cleaning Workflows" published in in IEEE transactions on industrial electronics.
- [3]. William E Winkler, "Data Cleaning Methods", U.S. Bureau of the Census Statistical Research, Washington, DC.
- [4]. Raju Dara,Dr.Ch. Satyanarayana, Dr.A.Govardhan(2013), "Front End Data Cleaning And Transformation In Standard Printed Form Using Neural Models" published in International Journal on Computational Sciences & Applications (IJCSA).
- [5]. Jaya Bajpai Pravin S. Metkewar(2016), "Data Quality Issues and Current Approaches to Data Cleaning Process in Data Warehousing" published in GRD Journals- Global Research and Development Journal for Engineering.
- [6]. H. R. Bhapkar, Parikshit N. Mahalle, Nilanjan Dey, and K. C. Santosh(2020), "Revisited COVID-19 Mortality and Recovery Rates: Are we missing Recovery Period" published in Journal of Medical Systems, Springer.
- [7]. <https://stackoverflow.com/questions/47230817/plotly-notebook-mode-with-google-colaboratory>
- [8]. <https://plotly.com/python/multiple-axes/>
- [9]. GDP datasets for prediction and forecasting: <https://www.macrotrends.net/countries/IND/india/gdp-gross-domestic-product>
- [10]. Recovery and Death Numbers Dataset: <https://api.covid19india.org/documentation/csv/>
- [11]. Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E. (2020). Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN Computer Science, 1(4). doi:10.1007/s42979-020-00209-9.
- [12]. Simon James Fong1, Gloria Li , Nilanjan Dey, Rubén González Crespo , Enrique Herrera-Viedma5.Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 6, No.1, DOI: 10.9781/ijimai.2020.02.002.
- [13]. Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, 15(1), 276. doi:10.1186/1471-2105-15-276.
- [14]. Neeraj Poonia, Sarita Azad. Short-term forecasts of COVID-19 spread across Indian states until 1 May 2020.
- [15]. Chimmula, V. K. R., & Zhang, L. (2020). Time Series Forecasting of COVID-19 transmission in Canada Using LSTM Networks. Chaos, Solitons & Fractals, 109864. doi:10.1016/j.chaos.2020.109864.
- [16]. Yoshiro Suzuki, Ayaka Suzuki, Shun Nakamura, Toshiko Ishikawa, Akira Kinoshita.Machine learning model estimating number of COVID-19 infection cases over coming 24 days in every province of South Korea (XGBoost and MultiOutputRegressor).
- [17]. Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model

- for COVID-19 pandemic in India. Stochastic Environmental Research and Risk Assessment. doi:10.1007/s00477-020-01827-8.
- [18]. Pai, C., Bhaskar, A., & Rawoot, V. (2020). Investigating the dynamics of COVID-19 pandemic in India under lockdown. *Chaos, Solitons & Fractals*, 138, 109988. doi:10.1016/j.chaos.2020.109988.
- [19]. Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., ... Mukherjee, B. (2020). Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the the COVID-19 Pandemic: Data Science Call to Arms. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.60e08ed5>
- [20]. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, Nanda C, Sharma S, Sharma YD, Rabaan AA, Rahmani J, Kumar P. Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health Surveill* 2020;6(2):e19115. doi: 10.2196/19115
- [21]. AutoRegressive Integrated Moving Average – <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp#:~:text=What%20is%20an%20Autoregressive%20Integrated,or%20to%20predict%20ofuture%20trends>.
- [22]. Introduction to ARIMA: nonseasonal models - <https://people.duke.edu/~rnau/411arim.htm>
- [23]. Understanding LSTM Networks - <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [24]. Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5), 216–233. doi:10.1002/widm.1157
- [25]. An End-to-End guide to Understand the Math behind XGBOOST – <https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/>
- [26]. Marcek, D. and M. Marcek, 2006. *Neural Networks and their Applications*. EDIS Publ., Slovakia.
- [27]. Yunhai Wang, Fubo Han, Lifeng Zhu, Oliver Deussen, and Baoquan Chen, “Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series” , *IEEE Transactions on Visualization and Computer Graphics* 2017.
- [28]. Benjamin Born, Alexander M. Dietrich, Gernot J. Müller, “The lockdown effect: A counterfactual for Sweden”, July 2020, CEPR Discussion Papers 14744, C.E.P.R. Discussion Papers.
- [29]. Mahipal Jadeja, Kesha Shah “TREE-MAP: A VISUALIZATION TOOL FOR LARGE DATA” Published in GSB@SIGIR 2015.
- [30]. Niyazi ARI, “Matplotlib In Python”, 2014 11th International Conference on Electronics, Computer and Computation (ICECCO), doi: 10.1109/ICECCO.2014.6997585
- [31]. “What is data visualization” <https://www.klipfolio.com/resources/articles/what-is-data-visualization#DataViz2>
- [32]. “The impact of COVID-19 — data visualization using Plotly and comparative analysis with SARS” <https://towardsdatascience.com/the-impact-of-covid-19-data-analysis-and-visualization-560e54262dc>
- [33]. “What is Data Visualization and Why Is It Important?” <https://www.import.io/post/what-is-data-visualization/>
- [34]. “Treemaps in python” <https://plotly.com/python/treemaps/>
- [35]. “Polar chart” https://doc-archives.microstrategy.com/producthelp/10.6/AdvancedReportingGuide/WebHelp/Lang_1033/Content/AdvancedReporting/Polar_chart.htm

- [36]. Website-
investopedia.com/terms/d/datamining.asp
- [37]. Gunther Schuh, Jan-Philipp Prote, Philipp Hunnekes(2019), "Data Mining Methods For Macro Level Process Planning", 13th CIRP Conference on Intelligent Computation in Manufacturing Engineering.

Cite this article as :

Tithi Choudhary, Muskan Shaikh, Shreya Khedekar, Richa Sirwani, Dr. Meenakshi A. Thalor, "Use of Data Science during Worldwide Pandemic : An Efficient and Reliable System to Support Decision Making in Different Sectors", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 2, pp. 17-27, March-April 2021. Available at doi : <https://doi.org/10.32628/IJSRST21825>
Journal URL : <https://ijsrst.com/IJSRST21825>