European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 7, Issue 3, 2020

PANDEMIC ANALYZER FOR EFFICIENT PREDICTION OF COVID-19 IN INDIA USING MACHINE LEARNING ALGORITHMS

OWK.MRUDULA[1], A. MARY SOWJANYA[2]

[1]Research Scholar, Department of CS&SE, Andhra University College of Engineering (A), Visakhapatnam, AP, India

Email id:mrudulaowk1990@gmail.com

[2]Assistant Professor, Department of CS&SE, Andhra University College of Engineering (A), Visakhapatnam, AP, India.

Email id:sowmaa@gmail.com

ABSTRACT

With the rapid growth of COVID-19 pandemic infectious disease caused by the Corona Virus. It was first identified in Wuhan in December 2019. It expanded its circle all over the world and finally spreading its route to India. The whole world is fighting against the spread of this deadly disease, cases in India also gradually increasing day by day since May after lockdown. This article proposes how to contribute to utilizing the machine learning and deep learning models with the aim for understanding its everyday exponential behaviour along with the prediction of future reachability of the COVID-2019 across the nations by utilizing the real-time information from the Johns Hopkins. This paper studies the COVID-19 dataset and explore the data by data visualization with different libraries that are available in Python. The paper also discusses the current situation in India while tackling the Covid-19 pandemic and the ongoing development in AI and ML has significantly improved treatment, medication, screening tests, prediction, forecasting, contact tracing, and drug/vaccine development process for the Covid19 pandemic and reduce the human intervention in medical practice. However, most of the models are not deployed enough to show their real-world operation, but they are still up to the mark. Within this paper, we present Exploratory Data Analysis, Data Preprocessing, Data Cleaning and Manipulations, Machine Learning Algorithms, Pandemic Analyzing Engine GUI, and Deep Learning. We have performed linear regression, Decision Tree, SVM, Random Forest and for forecasting, we performed FBPrompet, ARIMA model to predict the next 15 day's Pandemic situation.

KEYWORDS:ARIMA model, Coronavirus, Deep learning, Forecasting, FB Prophet ModelMachine learning, Prediction,Web Scraping

1. INTRODUCTION

Coronavirus disease is an infectious respiratory syndrome caused by a virus. In November 2019, it was first transmitted to human in Wuhan province China, this disease also named as COVID-19. Unfortunately, there is no vaccine for prevention or treatment of this disease. For

data analysis, Machine learning is an important aspect to understand the problems faced by an organization, and to explore data in meaningful ways. Machine learning is an automated method in various domains like medical engineering, financial sector, business sector, and other related sectors. Artificial Intelligence teaches machines from training datasets. We can identify patterns, analyzing data, and make correct decisions without human intervention. Machine learning algorithms are helpful to analyze thedata, Predict, interprets, classification and automatically identify data and retrieve the useful information accurately. Machine learning is divided into three parts1. Supervised learning, 2. Unsupervised learning, 3. Reinforcement learning.This paper focuses on the GUI application of data analytics on COVID-19-India. The contributions of this paper are summarized below:

1) The study has been analysed to find the aspects of COVID-19.
2) Data analytics has been applied on a currently available covid_19_india dataset to visualize the spread of COVID-19 as of for June 2020 for different states in India [10].
3) For data-driven diagnosis of COVID-19 dataset, the Data cleaning and algorithms like feature selection, regression and classification algorithms are suitable for applying on the dataset.
    i) Simple linear regression: It is a type of regression analysis where we have a single independent variable to predict the value.
    ii) Multiple linear regression: It is a type of regression which is also called as multiple regression, we have two or more independent variables to predict the dependent variableand
    iii) Polynomial regression:This is a form of regression analysis in which the relationship between the independent variable and the dependent variable is modelled. This model has been used to model the number of confirmed cases in India.
4) Creating Pandemic Analyzer Engine GUI using Tkinter[9].
5) Predictions of the current scenario are calculated.

2. BACKGROUND KNOWLEDGE

Tiwari, Upendra &Khan, Rizwan. [1], nowadays, a new challenge is in front of us all which has not only made our present life miserable. Keeping this into mind we have tried to use the machine learning to analyse the current situation created by COVID-19 and what may be its impact in future days.Punn et al [2], the artificial intelligence researchers are focusing their expertise knowledge to develop mathematical models for analyzing this epidemic situation. This article proposes to utilize the machine learning and deep learning models with the aim for understanding its everyday exponential behaviour along with the prediction of future reachability of the COVID-2019.Jiang F et al. [3] The purpose of this brief review is to summarize those published studies as of late February 2020 on the clinical features, symptoms, complications, and treatments of COVID-19 and help guide frontline medical staff in the clinical management of this outbreak.Panwar H et al [4] In this paper, we have proposed a deep learning neural network-based method nCOVnet, an alternative fast screening method that can be used for detecting the COVID-19 by analyzing the X-rays of patients which will look for visual indicators found in the chest radiography imaging of COVID-19 patients.Waleed Salehi A et al [5] The objective of this paper is to know the novel disease epidemiology, major prevention from spreading of Coronavirus Severe Acute Respiratory Syndrome, and to assess the machine and deep learning-based architectures performance that is proposed in the present year for classification of Coronavirus images such as, X-Ray and computed tomography. Specifically, advanced deep learning-based

algorithms known as the Convolutional neural network, which plays a great effect on extracting highly essential features, mostly in terms of medical images.Sowjanya Mary & Prasad, Ch. [6] In this paper comparing the forecasting models with the other models in the prediction of stock market trends using ARIMA model.Kolla Bhanu et al [7] Different prediction models are built using machine learning algorithms and their performances are computed and evaluated. Random Forest Regressor and Random Forest Classifier outperformed the other machine learning models. The correlation matrices are built to understand the relationship between the features of the datasets. The feature importance is computed for the classifiers built.Shinde GR et al [8] Forecasting of a pandemic can be done based on various parameters such as the impact of environmental factors, incubation period, the impact of quarantine, age, gender and many more. These techniques and parameters used for forecasting are extensively studied in this work. However, forecasting techniques come with their own set of challenges (technical and generic). This classifcation will surely help researchers to consolidate the forecasting methods. analysis of various forecasting models of COVID-19 will be more helpful for adapting better intervention.

3. PROBLEM STATEMENT

This paper mainly focuses on the analysis of COVID-19 dataset and predict the status for the next 15 days or for a coming months how the analysis of the disease is going to increase or decrease for that we need to createPandemic Analytics Engine Calculator GUI using Tkinker to calculate the IFR by using ifr function and calculate CMR (Crude Mortality Rate) by building a function for every state/district[9]. By calculating IFR and CMR and then we are comparing the one-state information with the nearby states and predict for the next week or month analysis by forecasting covid_india dataset using FBPrompet model.

4. PROPOSED SYSTEM

The proposed system workflow

1. Train the model with the dataset
2. Then perform data manipulation, data preprocessing and feature extraction
3. Input the preprocessed dataset to the trained model it will extract.
4. Calculating IFR and CMR, we are havinga formula
   IFR: Total no. of Deaths cases/ Total no. of Confirmedcases
   CMR: Total no. of Deaths cases/ Total no. of Population
5. Finding the rate of spread of the disease in the next 7 days with the help of regression analysis models.Given as input we will get the important category variables to perform the predictions
6. A machine learning-based regression analysis models were developed for exposed COVID-2019.
7. FB Prophetmodelis used for Forecasting of COVID-2019 in India with the next 7 days for better management for doctors and various government organizations.
8. The above steps will be repeated for various datasets.

A. Data acquirement

Interpreted datasets have been used for training purposes. The datasets used for training and testing must be labelled dataset. Here we used CSV file format as an input file.

(i) Data Preprocessing and Data Cleaning

(ii) Data Visualization using Seaborn, Matplotlib and Bokehanalysis for states
(iii) Feature Selection and Feature Extraction
(iv) Pandemic Analyzer GUI Tool
(v) Machine Learning Algorithms
(vi) Predictions and Forecasting

B. Preprocessing

For any data mining application or machine learning application, the preprocessing is a fundamental step. Adataset is regularly unstructured and noisier. For proper analyzing of the data,we want to prepare a structured data first. Data cleaning is required and plays a major role before modelling. Preprocessing is applied to the dataset and then we will apply machine learning techniques on it.First, we have to use the appropriate dataset, to make and improve Machine Learning models. As the dataset information is gathered or collected from different and various sources which are then merged to make an appropriate dataset format. Once the dataset is prepared you should save it in the.CSV or.XLSX file formats. This Preprocessing eliminates the noise and unwanted data from the dataset and organizes the dataset. Now we can perform IFR & CMRcalculations, machine learning algorithms.

C. Data Visualization

The key aim of the paper is to explore the data and report insights and then predict the analysis. The report insights are represented in the form of data visualization as a diagrammatic representation. In this paper, we have used matplotlib which is a multi-platform library built on NumPy and designed to work with the SciPy stack. We have also used the Seaborn library for statistical plotting. Bokeh is a python library which is used for interactive visualization which allows building complex statistical plots quickly through simple commands.

D. Pandemic Analyzer GUI:

A few Graphical User Interface (GUI) based services are available for users in Python to develop their interface. It is an open-source accessible GUI which is called as Thinker[9]. It is a powerful object-oriented interface to the Tk GUI toolkit. Creating a Pandemic Analyzer GUI application using Tkinter is a fast and easy task. All you need to perform the following steps: –

1. Import the *Tkinter* module.
2. Create the GUI application main window.
3. Add one or more of the above-mentioned widgets to the GUI application.
4. Enter the main event loop to take action against each event triggered by the user.

Tkinker provides various controls, such as buttons, labels and text boxes used in a GUI application. These controls are commonly called widgets like buttons, canvas, Checkbutton, Entry, Frame, Label, Listbox, Menubutton, Menu, Message, Radiobutton, Scrollbar, Text, PanedWindow, LabelFrame, tkMessageBox, etc.This Pandemic analyzer (GUI) shows the current data and calculates the IFR & CMR for that particular state in India. The data fetched will be dynamic and real-time.

E. Machine Learning Classification algorithms

After the Pandemic analyzer calculation of IFR & CMR, we offer the machine learning different classification algorithms to train binary classifiers including Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and AdaBoost. The selection of algorithms was focused on the effectiveness of the data. In this paper, we have done all the classification techniques on the COVID-19 dataset and find calculate their accuracy.

S.No	Models	Accuracy
1.	Linear Regression	87.99
2.	Random Forest	88.54
3.	Decision Tree	87.50
4.	AdaBoost	83.85
5.	SVM	86.46

Figure1: Machine Learning models and their accuracy

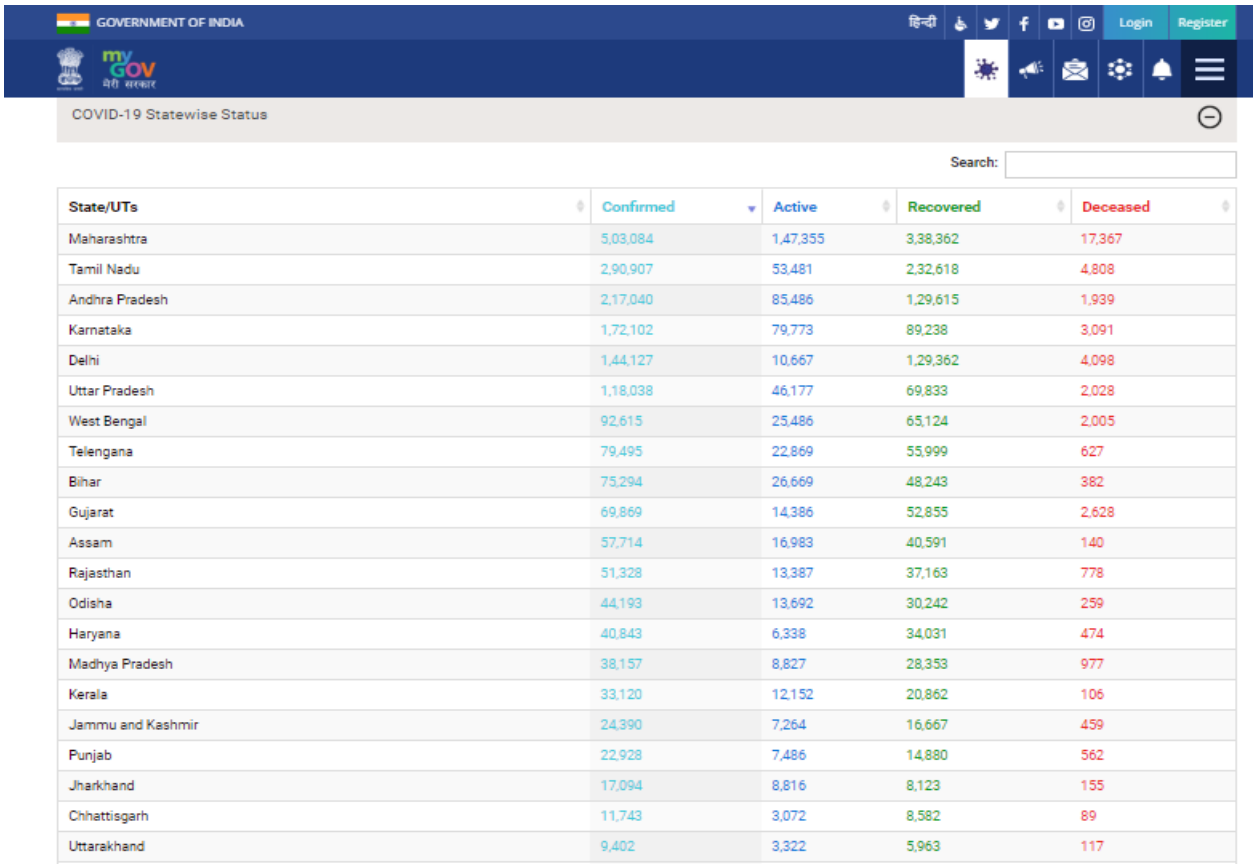
F. Predictions & Forecasting

Now we will apply FB Prophet model and can predict the next month analysis of the particular state. FB Prophet is an open-source time series model which is used for forecasting and works best with time series that have strong seasonal effects. First, we need to create a dataframe for future dates with the specified parameters that are how many days you want to predict and analyzed. But the actual target prediction is yhat. Predicting the increase of confirmed cases by looking into the model decomposition of the dataset When comparing with the ARIMA model and LSTM model monthly predictions and forecast is clearly shown by the FB Prophet. So we calculated the August month COVID-19 predictions of India and conclude that the rate of active cases and confirmed cases will increase.

S.No	Models	RMSE Errors	MSE Errors
1.	LSTM	9.89321	96.115824
2.	ARIMA	8.56067	68.718538
3.	FB Prophet	11.64295	132.66271

Figure 2: Comparing the forecasting models

5. RESULTS AND DISCUSSIONS



State/UTs	Confirmed	Active	Recovered	Deceased
Maharashtra	5,03,084	1,47,355	3,38,362	17,367
Tamil Nadu	2,90,907	53,481	2,32,618	4,808
Andhra Pradesh	2,17,040	85,486	1,29,615	1,939
Karnataka	1,72,102	79,773	89,238	3,091
Delhi	1,44,127	10,667	1,29,362	4,098
Uttar Pradesh	1,18,038	46,177	69,833	2,028
West Bengal	92,615	25,486	65,124	2,005
Telengana	79,495	22,869	55,999	627
Bihar	75,294	26,669	48,243	382
Gujarat	69,869	14,386	52,855	2,628
Assam	57,714	16,983	40,591	140
Rajasthan	51,328	13,387	37,163	778
Odisha	44,193	13,692	30,242	259
Haryana	40,843	6,338	34,031	474
Madhya Pradesh	38,157	8,827	28,353	977
Kerala	33,120	12,152	20,862	106
Jammu and Kashmir	24,390	7,264	16,667	459
Punjab	22,928	7,486	14,880	562
Jharkhand	17,094	8,816	8,123	155
Chhattisgarh	11,743	3,072	8,582	89
Uttarakhand	9,402	3,322	5,963	117

Figure 3:Statewise COVID-19 dataset of India

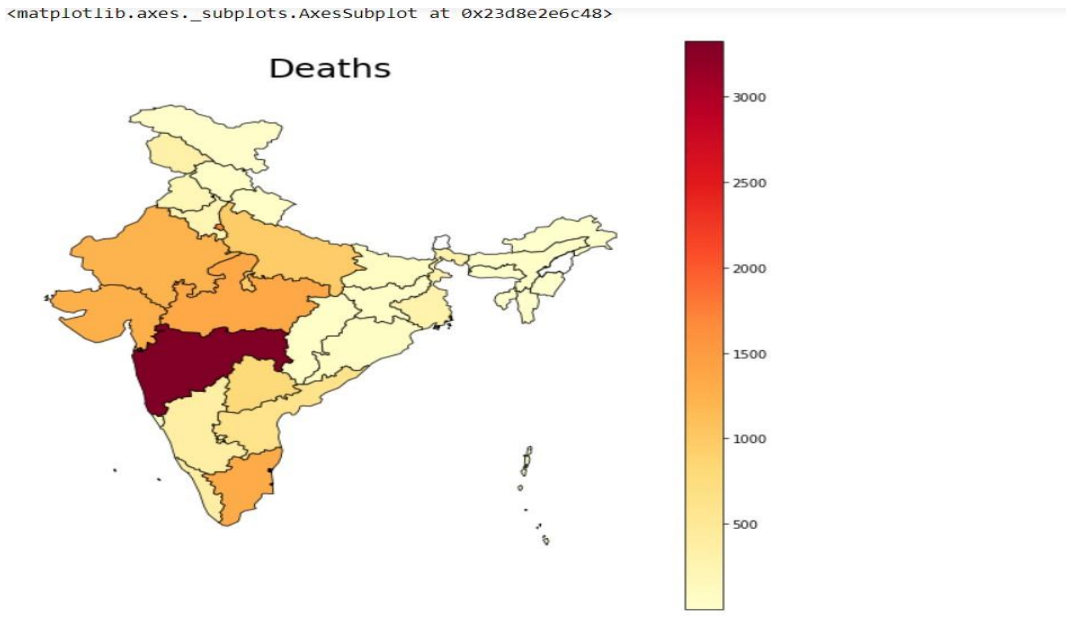


Figure 4: Plotting India map with most affected state

ISSN 2515-8260 Volume 7, Issue 3, 2020

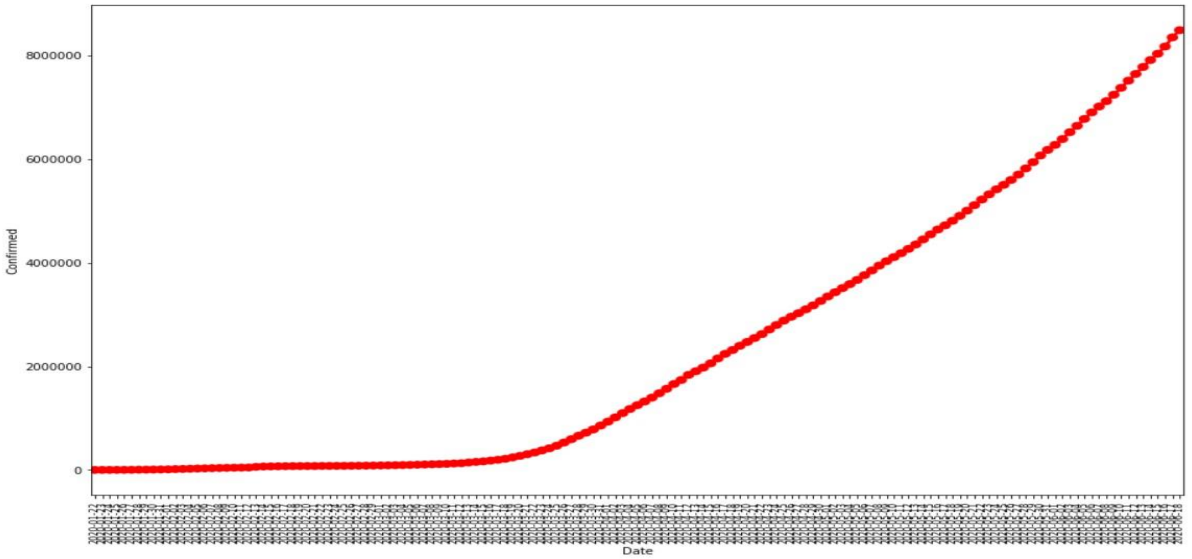


Figure 5: Plotting graph for India confirmed cases from December to till 12/07/2020

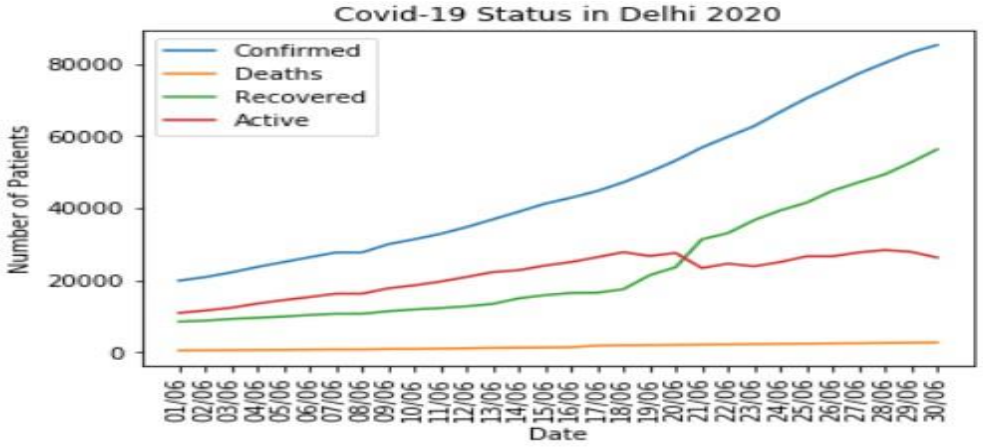


Figure6: Selecting the Delhi state with total no. of patients in June

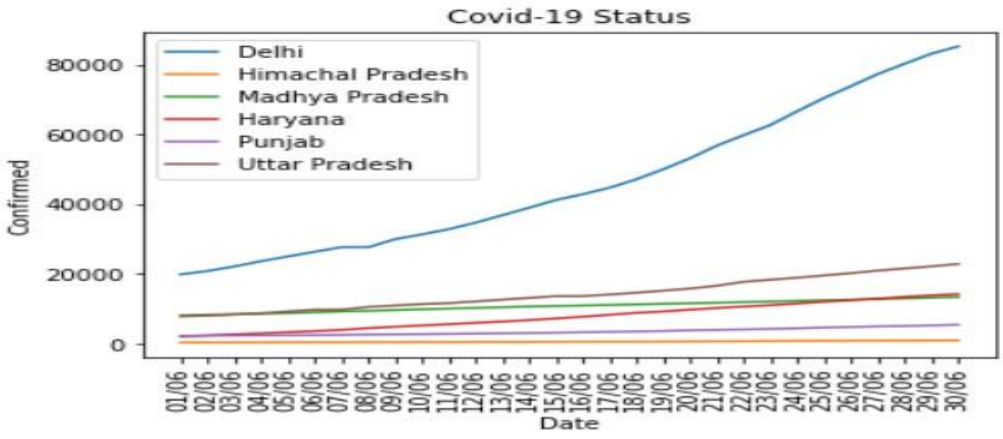


Figure7: Comparing Delhi state with the nearby states

ISSN 2515-8260 Volume 7, Issue 3, 2020

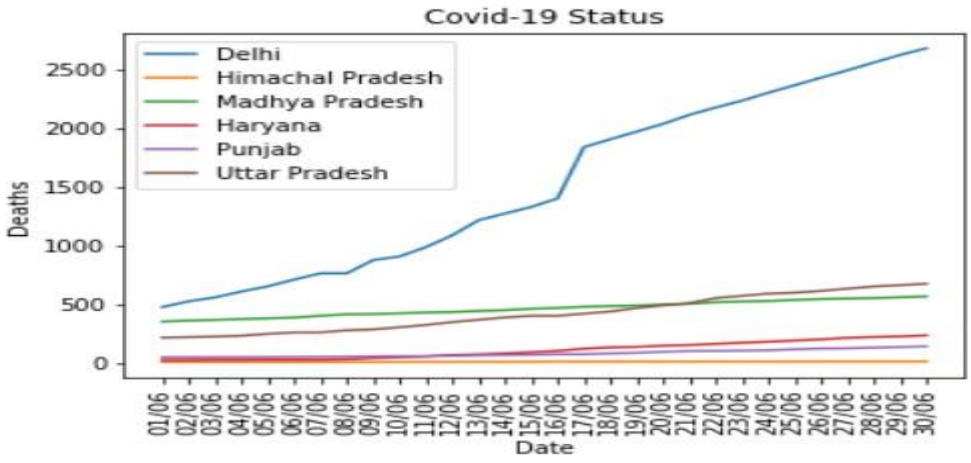


Figure8:Total number of deaths in June and comparing it with nearby states

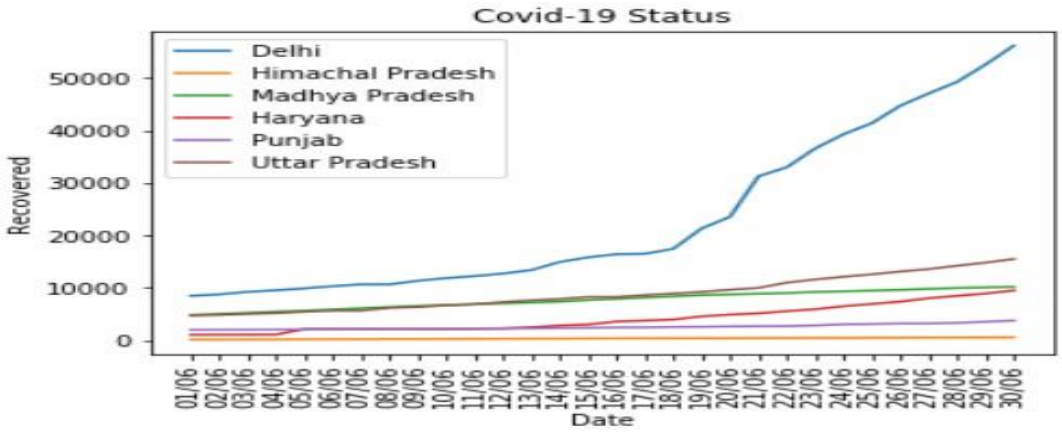


Figure9: Total number of Recovered in June and comparing it with nearby states

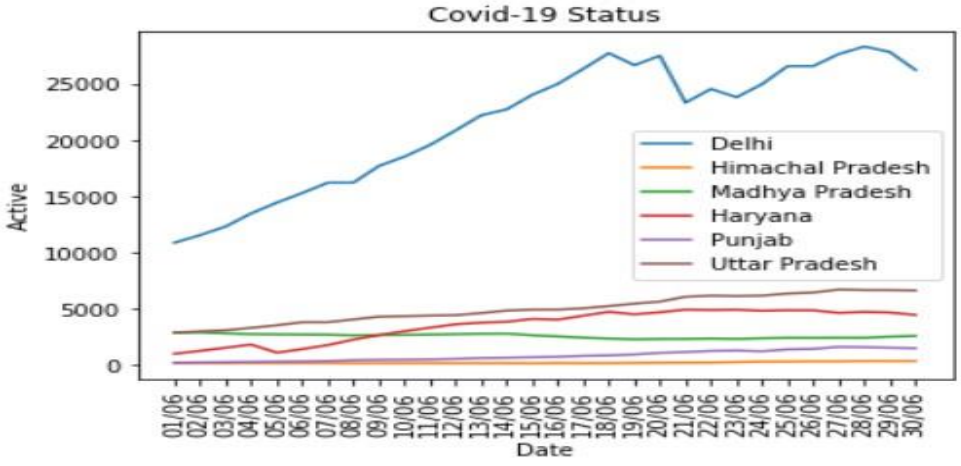


Figure10:Total number of Active cases in June and comparing it with nearby states

ISSN 2515-8260 Volume 7, Issue 3, 2020

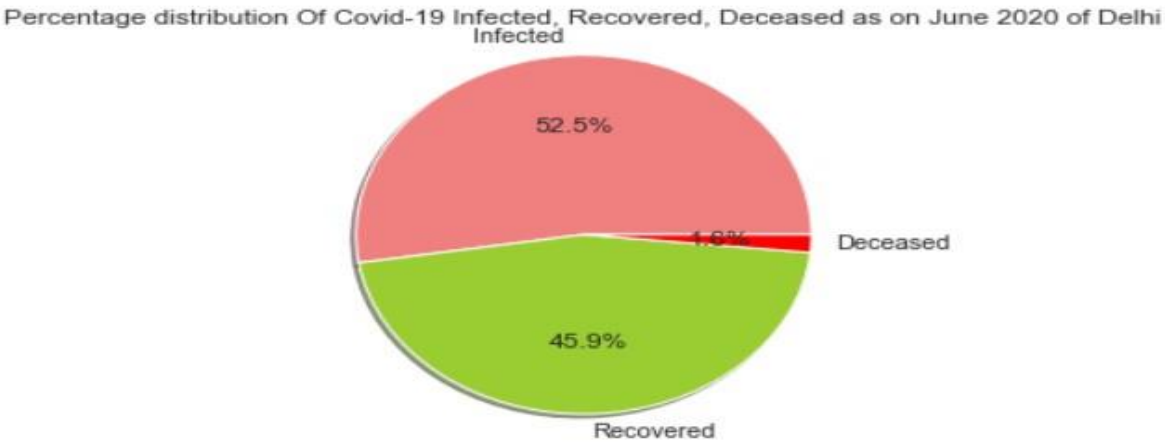


Figure11: Pie chart representation of cases for the Delhi state in June 2020

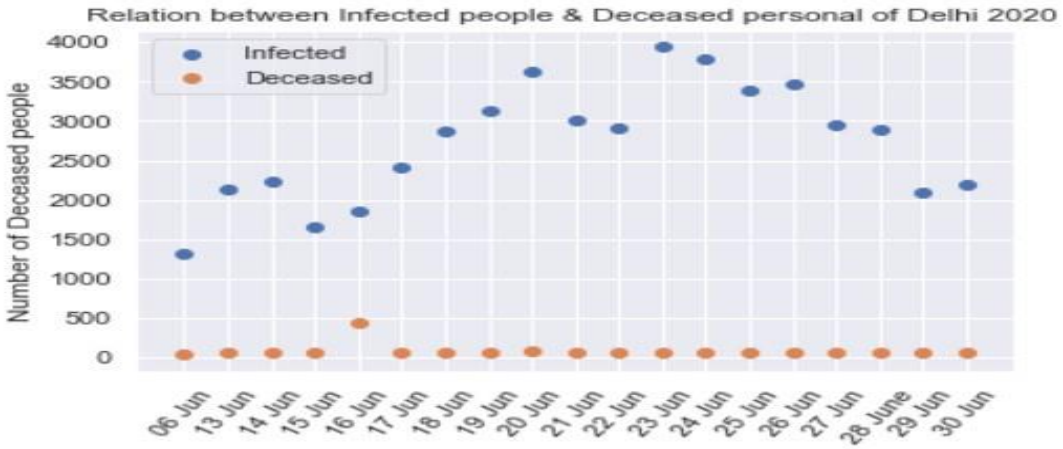


Figure12: Plotting the relation between Infected and Deaths of Delhi in June 2020

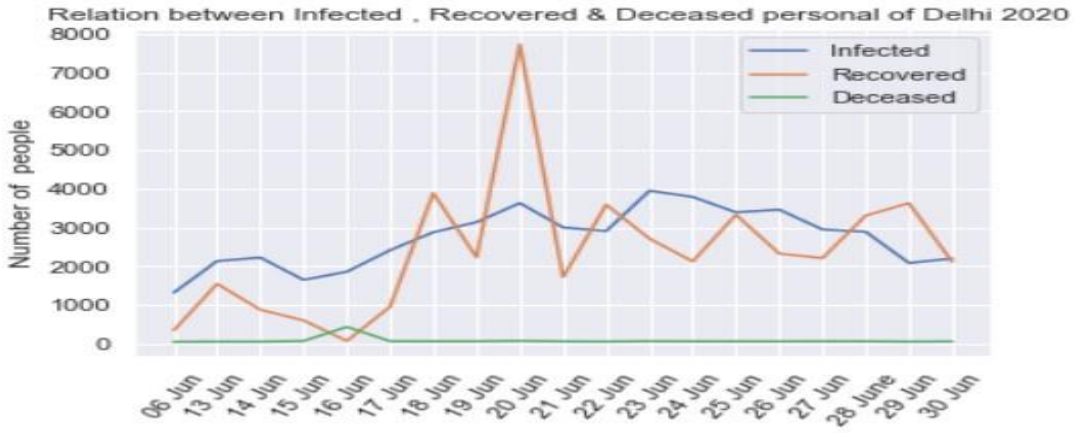


Figure13: Relation between infected cases,recovered cases and death cases of Delhi in June 2020

ISSN 2515-8260 Volume 7, Issue 3, 2020

Figure14: Plotting the relation between recovery rate and mortality rate of Delhi weekly

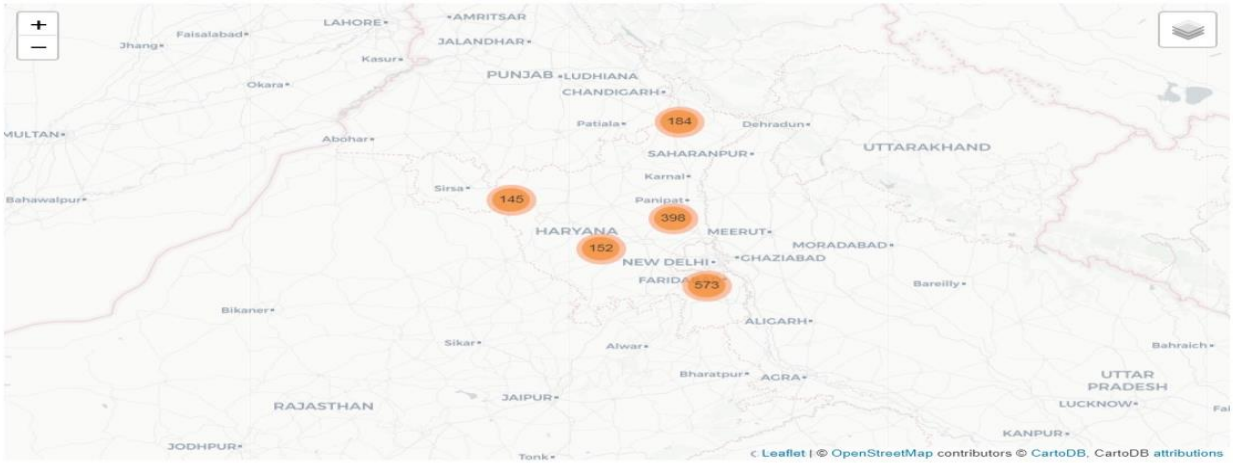


Figure15: Leaflet Map marking the Delhi with the nearby states

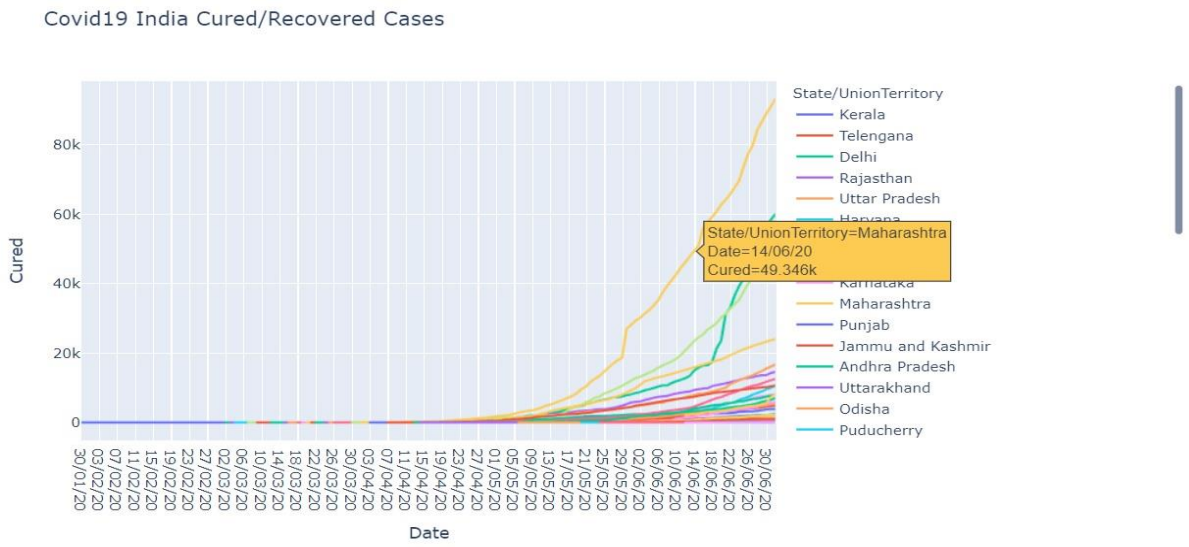


Figure16: Covid19 India cured cases with state name and date

ISSN 2515-8260 Volume 7, Issue 3, 2020

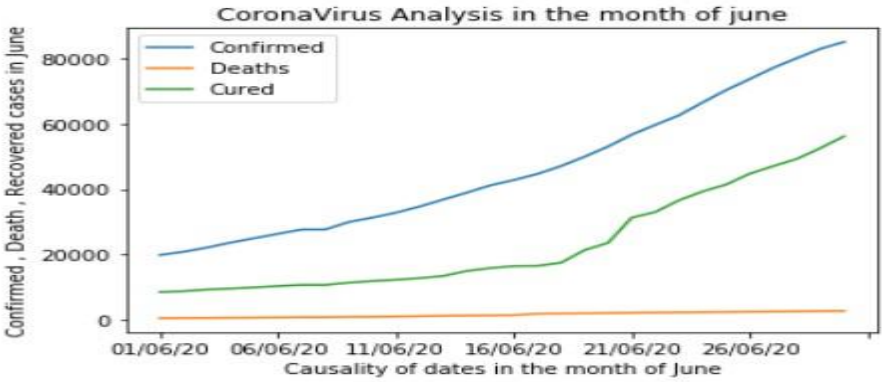


Figure17: Analysis report for June 2020

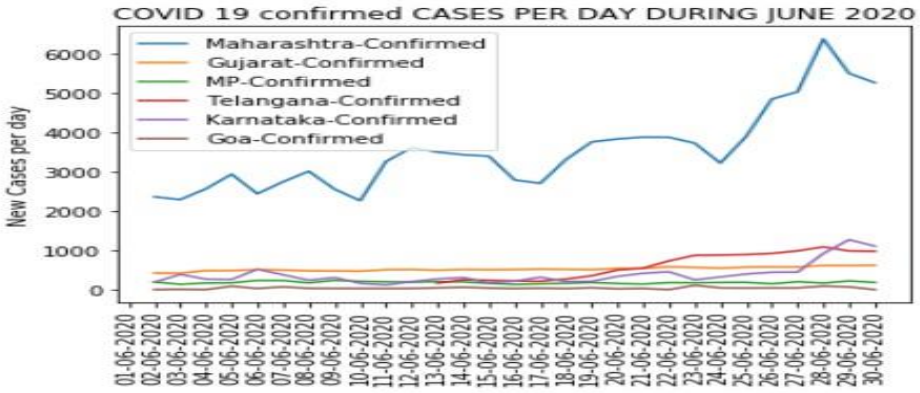


Figure18: comparing no. of cases per day with the nearby states

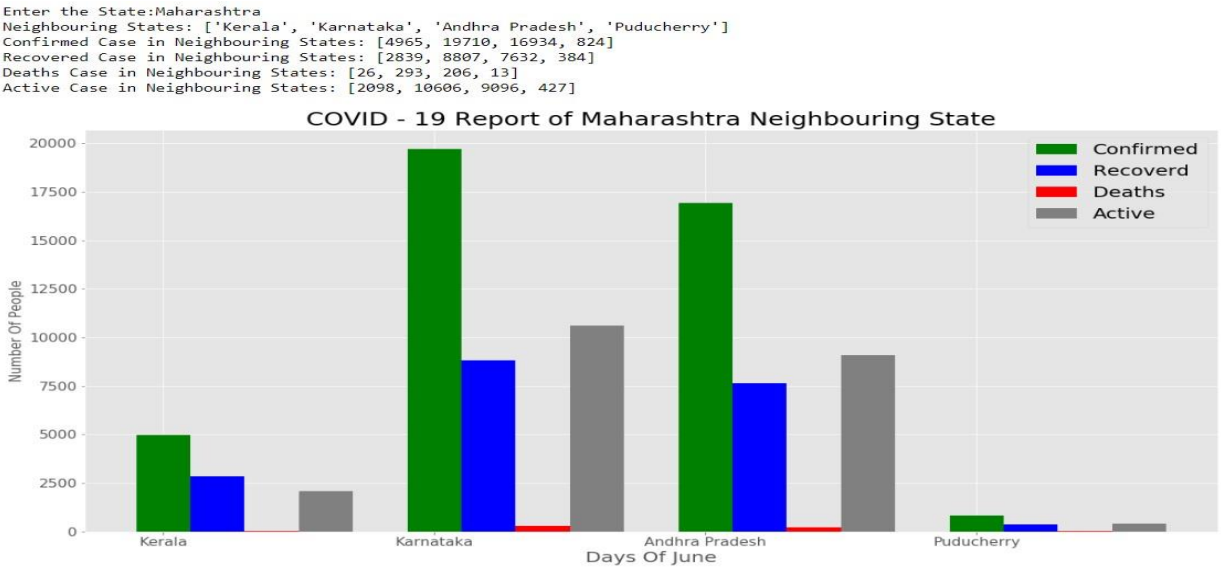


Figure19: Comparing the report of Maharashtra

ISSN 2515-8260 Volume 7, Issue 3, 2020

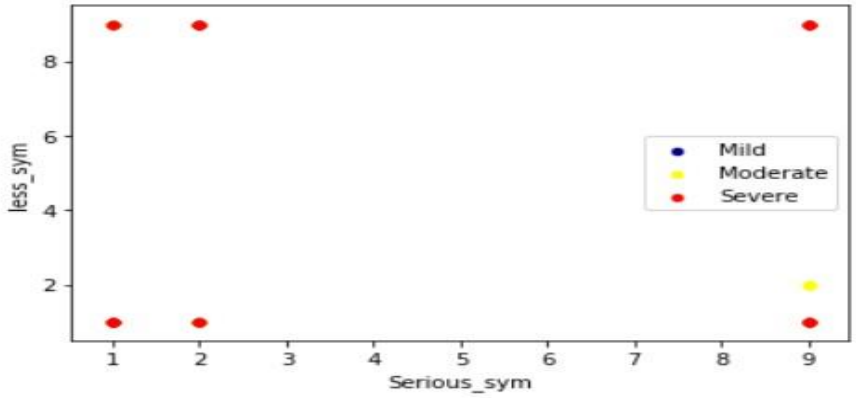


Figure20: Severity gauging of the covid19

	precision	recall	f1-score	support
0	0.28	0.10	0.15	204
1	0.36	0.71	0.48	218
2	0.34	0.21	0.26	205
accuracy			0.35	627
macro avg	0.33	0.34	0.30	627
weighted avg	0.33	0.35	0.30	627

```
Confusion matrix, without normalization
[[ 20 132  52]
 [ 29 154  35]
 [ 22 139  44]]
```

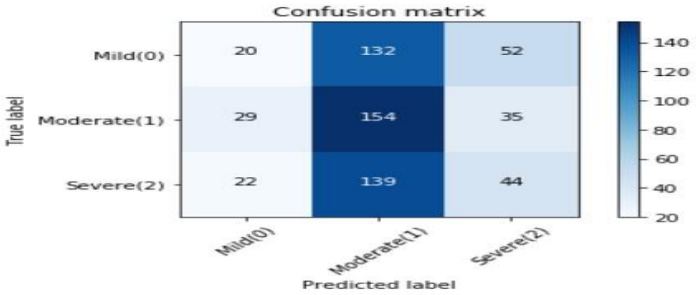


Figure21: Confusion Matrix for the severity gauging

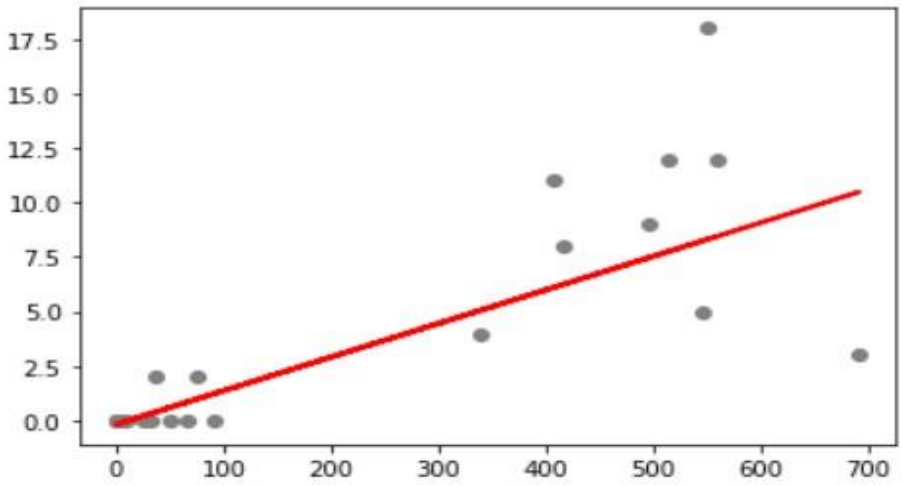


Figure22: Linear regression classifier

ISSN 2515-8260 Volume 7, Issue 3, 2020

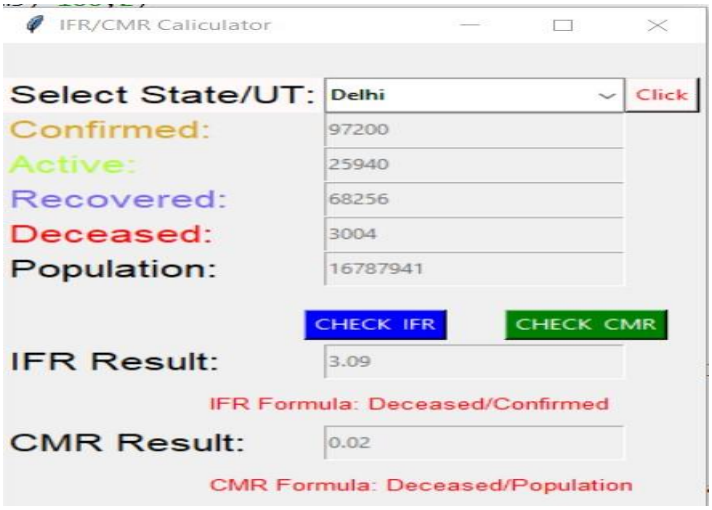


Figure23: Pandemic Analyzer Engine for Calculating IFR and CMR

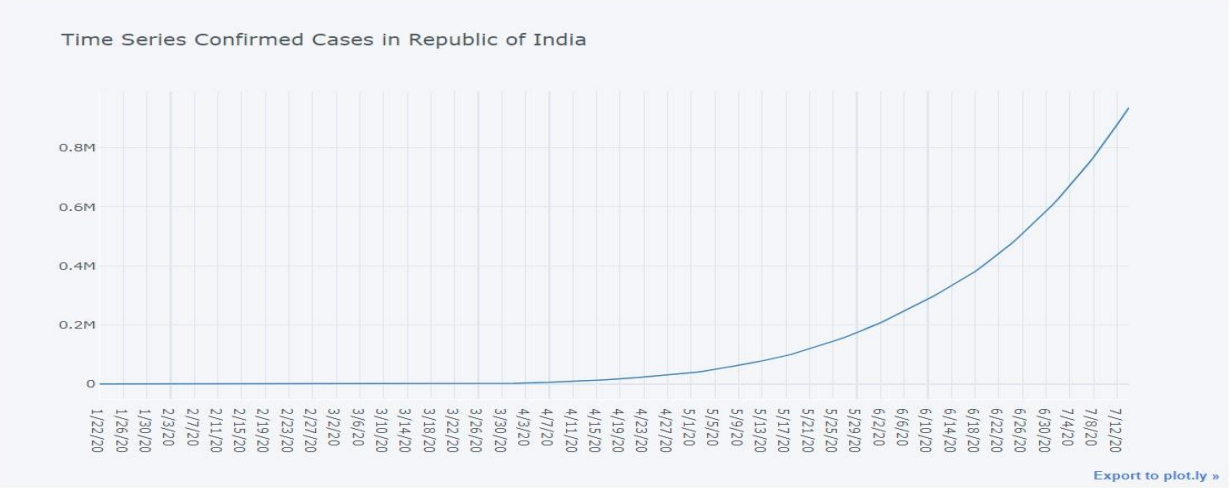


Figure24: Time series analysis for the confirmed cases in India from January 22nd - July 12th

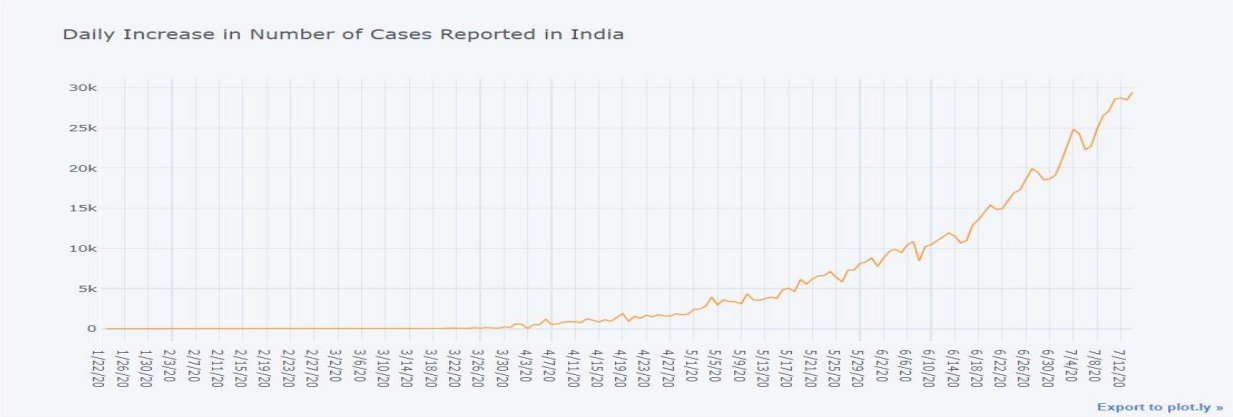


Figure25:Time series analysis for daily increasing cases in India from January 22nd - July 12th

ISSN 2515-8260 Volume 7, Issue 3, 2020

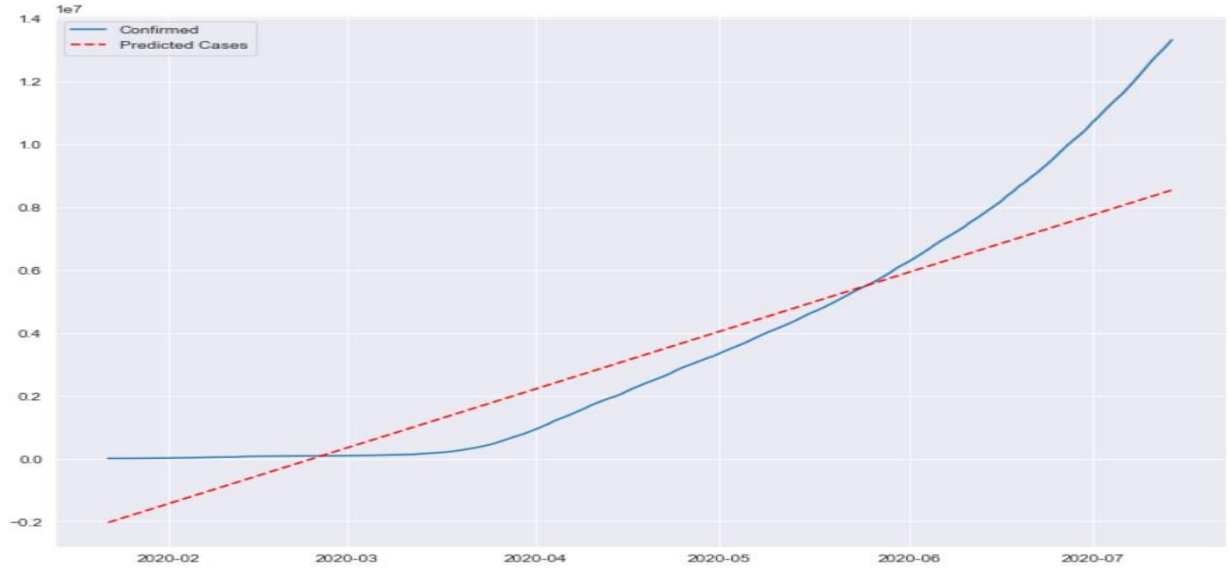


Figure26: Analysis for every month in India Confirmed vs. Predicted cases from February – July

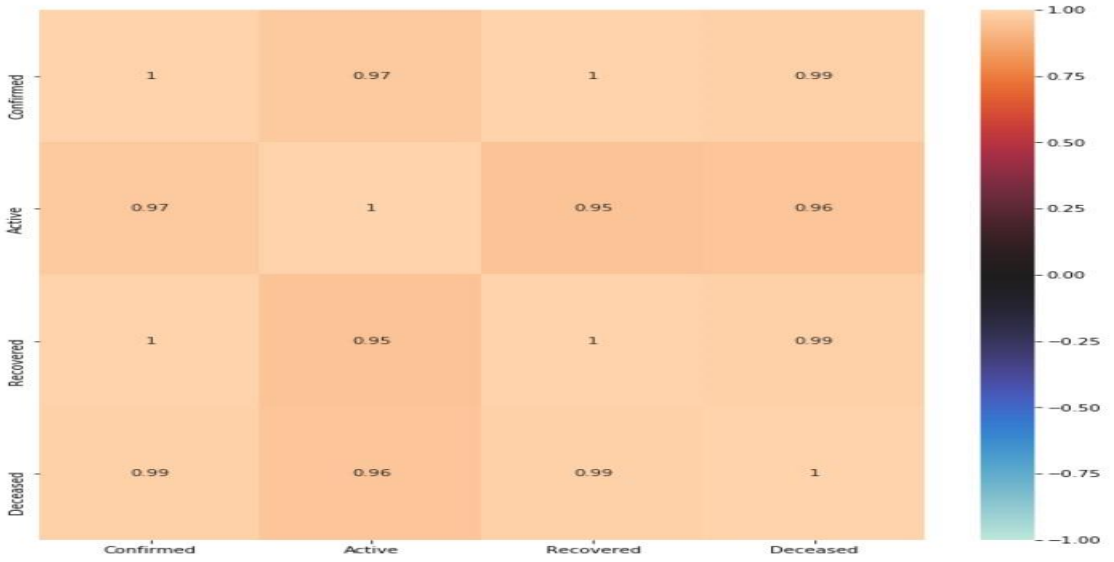


Figure27: Correlation Heat map for various parameters of covid19

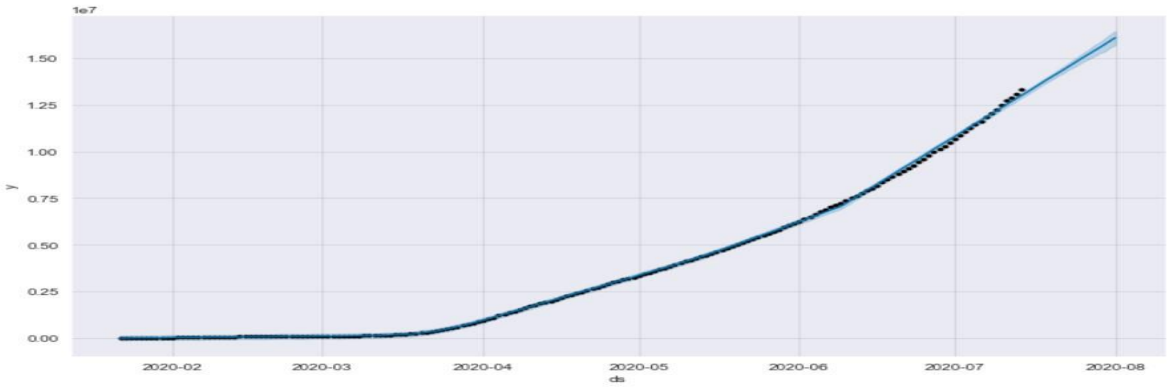


Figure28: Prediction using FBPrompet ARIMA for the August 2020 month

7. CONCLUSION AND FUTURE WORK

We can gain many primitive insights by calculating metrics, discover and handle incorrect data and missing values.To analyze and summarize covid19 dataset pandemic this framework IFR and CFR can be used. To predict the analysis and time series for the next coming month by using FBPrompet. The results show that Pandemic analyzer will give accurate information about IFR and CMR which gives better performance to implement machine learning techniques on the dataset. The future work will work on upgrading the analyzer by utilizing deep learning procedures to conquer and confinement of gathering information on datasets for preparing the machine learning models.

8. REFERENCES

[1] Tiwari, Upendra & Khan, Rizwan. (2020). Role of Machine Learning to Predict the Outbreak of Covid-19 in India. Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture & Technology. 12. 2663-2669.

[2] Punn, Narinder & Sonbhadra, Sanjay & Agarwal, Sonali. (2020). COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. 10.1101/2020.04.08.20057679.

[3] Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z (2020) Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). J Gen Int Med 35:1545–1549. https://doi.org/10.1007/s11606-020-05762-w

[4] Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals. 2020 Sep;138 109944. doi: 10.1016/j.chaos.2020.109944. PMID: 32536759; PMCID: PMC7254021.

[5] Waleed Salehi A, Baglat P, Gupta G. Review on a machine and deep learning models for the detection and prediction of Coronavirus. Mater Today Proc. PMCID: PMC7309744.

[6] Sowjanya, Mary & Prasad, Ch. (2018). Prediction of Stock Market Trends using Machine Learning Techniques. 5. 302-308.

[7] Kolla, Bhanu. (2020). Analysis, Prediction and Evaluation of COVID-19 Datasets using Machine Learning Algorithms. International Journal of Emerging Trends in Engineering Research. 8. 10.30534/ijeter/2020/117852020.

[8] Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. *SN Computer Science*. 2020;1(4):197. doi:10.1007/s42979-020-00209-9

[9]https://www.academia.edu/42794679/Pythonand_Tkinter_Programming_Graphical_user_interfaces_for_Python_programs

[10]https://www.mygov.in/covid-19