COVID-19 Prediction and Analysis of Spreading Rate Dashboard.

Sumit Rasal[1], Rahul Mali[1], Shubham Patil[1], Ganesh Kannor[1], Danish Majeed[1], Ms. Sunita Nandgave[2]

[1] *UG Students,* [2] *Assistant Professors, Computer Engineering Department,*
G H Raisoni College of Engineering and Management, Pune

ABSTRACT

covid-19 is started at the end of 2019 at the Wuhan city chain. After a couple of months, it spread all over the world. The Coronavirus Disease-2019 (COVID-19) badly impacts the health and well-being of the Indians and the global population. The number of positive cases of Covid-19 daily is increasing and it's created a lot of stress on governing bodies across India and global and they are finding it a hard time to tackle the situation. The main thing to control the covid-19 is social distancing, wearing the mask, and analysis of coivd-19 data so the government body can take the right measure to control the covid-19 spreading. So We build the Covid-19 Dashboard. The Dashboard is showcasing the total numbers of confirmed, active, recovered cases of each state and union territory of India. One advanced thing in the dashboard is prediction of coivd-19 positive patient for next 7 Days per state and union territory wise. For this machine learning and deep learning algorithm are used. Analyzed the covid-19 data by using the advance graph, that provides by seaborn and matplotlib python library. The dashboard will help the government and the people of India to take right decision and measurements to tackle the covid-19 situation. Project is deployed on the cloud that give the more flexibility and leverage. DevOps technology also used for doing the automation in the ML model for creating more accurate model, fast development and fast feedback.

Keywords— Keywords— AWS Cloud, DevOps, ML, LSTM, RNN, Covid-19, dash, Plotly.

I. INTRODUCTION

The covid-19 coronavirus disease started in Wuhan, China sometime during December 2019. Within two-three months, more than fifty thousand people were infected by the virus and hundreds died due to the virus [1]. In the first stage, the outbreak caused several deaths, as the medical systems of India, were not capable of handling many seriously ill patients. Till 31 Dec 2020, there were 1.8 million deaths [2] reported across the world due to this pandemic, and in India 2000 thousand reported. In a rapidly evolving pandemic, improper analysis and less data regarding the patient because of that Government body not able to take the decision. Government bodies badly need the analysis coivd-19 data and how many people will have infected for the next seven-eight days. We decided that we analyze the covid-19 data by using machine learning or deep learning technology. So we can predict how many people will have infected for the next seven-eight days. For the prediction, we developed and tested RNN predictive algorithms for India. It was noticed that the pattern of growth within the number of cases varied from state to state and union territory. The basic approach for the predictions was to coach the models based on the data-set provided, but these models were not sufficiently accurate, as they were trained on only a one-time series of data-set. As a result, the models were unable to accurately predict the number of new cases for the next seven-eight days and, consequently, the existing techniques failed to utilize the resources in an optimized way that why we are using the LSTM algorithm with proper data processing [5]. Insufficient training data is also one of the reasons for the models to have low accuracy but with time model will more mature. We tried LSTM with RNN standard machine learning (ML) algorithms for predicting the number of patients for the next 7 days. After getting a decent accuracy of 90 and above 90 Percent, we implemented these algorithms on data sets of a different state and union territory of India. We selected the state and union territory of India with the highest population and the highest density for our work. By using the data of these state and union territories, we trained standard prediction models using multiple ML and Deep Learning algorithms and obtained different accuracy for each of the models for different state and union territories. The RNN with LSTM algorithm gave high accuracy for state and union territory. However, there was variations inaccuracy because of the different parameters are consider for trends of change in COVID-19 patients for different state and union territory of India.

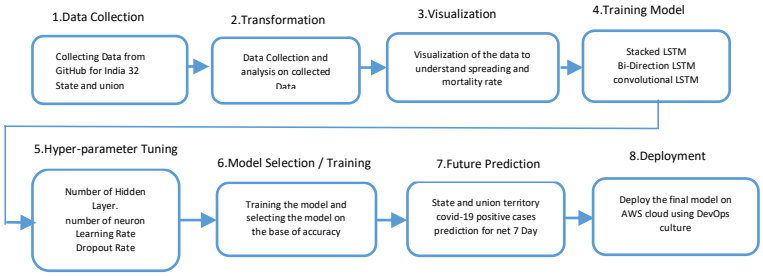
II. LITERATURE SURVEY

A lot of research is going for predicting the infection rate for the next 5 days or the next 2 days so the government body can take a decision on a particular situation. By studying the different research papers, we got some information about the algorithm. Tomar et al.

[6] have used data-driven estimation methods like long short-term memory (LSTM) and curve fitting for prediction for the monthly number of COVID-19 cases in India for the next five days and also give the idea about the effect of preventive measures like social-distancing, isolation, and lockdown on the area where the spread rate of COVID19 is high. Kumar et al. [7] have applied cluster analysis, to classify real groups of communicable disease of COVID19 on a knowledge set of various states and union territories of India, supported their high similarity and co-relation between one other. [8] Introduced an objective approach to predict a continuation of COVID- 19 by live forecasting. They produce ten-day-ahead point forecasts and prediction intervals. A susceptible-exposed-infectious-recovered (SEIR) metapopulational model was used to predict the spread across all major cities in China, with 95% credible intervals.[9] created a mathematical model for predicting the spread of COVID19 in countries using various types of parameters and tested their model on real data of countries.[10,11] Machine Learning (ML) and the Data Science community are striving hard to improve the forecasts of epidemiological models and analyze the information flowing over Twitter for the development of management strategies, and the assessment of the impact of policies to curb its spread. Various data sets in this regard have been openly released to the public. Yet, there is a need to capture, develop and analyze more data as the COVID-19 grows worldwide because it helps in the present and also in the feature for implementing such a model using ML and Deep Learning. We also get some idea about how to collect the data from a different location. [13] They are collected the data from the GitHub repository. achieved the highest accuracy of 95% for India using the ARIMA model forecasting. The ARIMA model for state and union territory for India achieved an accuracy of 90.55%, which was high as compared to an accuracy of 70% obtained by Gupta et al. [14] using the ARIMA model and Exponential smoothing.

III . MODEL FLOW

1.Data Collection
2. Transformation and scaling
3. Visualization
4. Training Models
5. Hyper-parameter Tuning
6. Model Selection
7. Future Prediction
8. Deployment



Data Collection

We Need Data for analyzing and predicting the spreading Rate of covid-19 for that we using the GitHub repository. https://api.covid19india.org/csv/latest/state_wise.csv
The data we are collecting from the repository are the time series data and in this repository we have Global confirm cases, death cases of covid-19. By using the pandas, we are directly fetching the data from the GitHub repository.

Transformation and scaling

In the transformation, we did Data Cleaning, data engineering, and exploratory data analysis on collected data. Transformation helps us for getting the right data in a proper format. In the transformation we did the one-hot encoding and replacing the missing data using the analysis of data by the hitmap. In the scaling, we normalize the range of independent variables of features of data that we collect. The alternative name is called data normalization.

Visualization

we doing the Visualization of the data to understand the spreading rate and mortality rate. Help us to understand a thing by visualization rather than using numbers. It helps us also for selecting the future parameter from the data set to train the model. some Data visualization is done on the data. For visualization, we used seaborne and matplotlib. Analyzed graph for global and Maharashtra we did at the initial stage.

Training Models

For training the model we used google colab. Google colab offer developer to use the GPU for developing the machine learning project. Google colab is free for use and google colab for flowing import thing that help us in the machine learning project.

- It provides the live interpreter so, we can write and execute the in the python on run time.
- Give the facility to integrate PyTorch, Tenserflow, Keras, OpenCV like more.
- Free Cloud services with free multiple GPU
- Import external Dataset e.g. REST Api, kaggle

In training the model, we flow some step like.

- Choose the algorithm, and defining it parameter or hyper parameter.
- Provide a dataset that is labelled, and has data compatible with the algorithm.
- After model is completed by suing the testing data set we check the accuracy of model.

So Here we choose the LSTM algorithm that useful for our data, because we have the time series. When data is in the time series mean it is belonging to the regression model. In the regression LSTM is best for our data. In the hyper parameter we use the relu as the activation function. Adam as optimizer, learning rate set 0.001 and did the 100 Epoch.

Hyper-parameter Tuning

In the Hyper-parameter we are tuning the parameter that can be change and showcasing the effect on the model accuracy, like Learning rate, Batch size, momentum, and weight decay. We can also do advance thing in the tuning like trying different convolution layer architecture, number of hidden layers and unit, Changing the optimization technique. Hyper-parameter helps us to get the more efficient model.

Model Selection

LSTM Networks: Long LSTM networks are modified version of recurrent neural network (RNN), which make it easier to recollect past data in memory. We face the vanishing gradient problem within the RNN is resolve within the LSTM. LSTM is well-suited to classify, process and predict statistic data for unknown duration. It trains the model by using back-propagation concept. In an LSTM network, three gates are present are following.

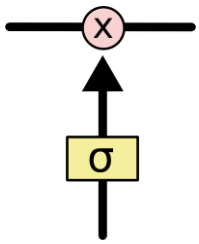


Figure 1
LSTM Gate

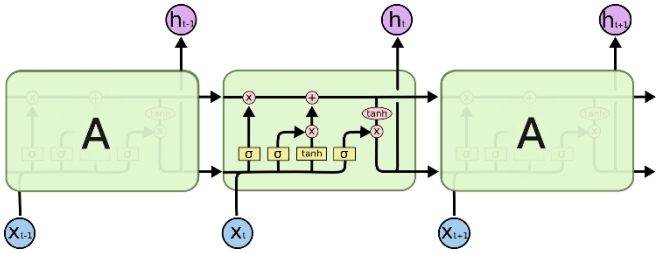
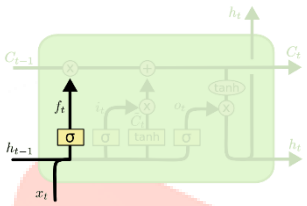


Figure 2
Architecture of LSTM

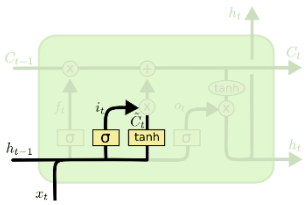
***Forget gate*:** forget gate discover what details to be discarded from the block. it's decided by the sigmoid function. it's at the previous state(ht-1) and thus the content input (Xt) and outputs variety between 0 (omit this) and 1 (keep this) for each number within the cell state ct1.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Figure 3
Forget Gate

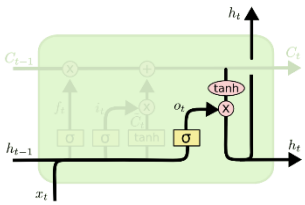
Input gate: input gate find out which value from input should be used to modify the memory. Input Sigmoid function decides which values to let through zero, one. And tanh active function gives weightage to the values and generate the output in the range of -1 to 1.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Figure 4
Input gate

Output Gate: The output gate determines the worth of subsequent hidden state. This state contains information on previous inputs. First, the values of the present state and former hidden state are passed into the third sigmoid function. Then the new cell state generated from the cell state is skilled the tanh function. Both these outputs are multiplied point-by-point. Based upon the ultimate value, the network decides which information the hidden state should carry. This hidden state is used for prediction.



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 5
Output Gate

Future Prediction

In the future prediction, we are predicting the number of peoples will found covid-19 infected for next 7 Day.

Deployment

Deployment is the Last step of our project. For the deployment we used the cloud technology

AWS Cloud: AWS cloud providing infrastructure for project. For the project what we need that is available on the AWS cloud. Like kubernets, Docker container. In the AWS cloud we used the ec2 instanced for launching the instance. S3 storage for storing the images and video.

Kubernetes: kubernets is the container management tool that help us to manage the container. Doing the automation in the scalability we used the Kubernetes that helps us in the scaling the infrastructure and reducing the infrastructure. That required because if the traffic increase on the covid-19 Dashboard, it will not crash.

Web UI: Bellow image is the web UI look. In the web UI, we showcasing the number of cases will infect by covid-19 for next 7 days for each state and union territory of India. We also showcasing some graph for confirm cases, recover cases, death cases. The web UI is implemented **Dash**(v1.19.0) and **Plotly** (v 4.14.3. with the help of Plotly we create interactive and dynamic visualization. Dash is Python framework for building web applications. It built on top of Flask, Plotly. js, React and React Js. It enables you to build dashboards using pure Python.

IV. RESULT & DISCUSSION

Below, some of the analysis graphs of Maharashtra and global.

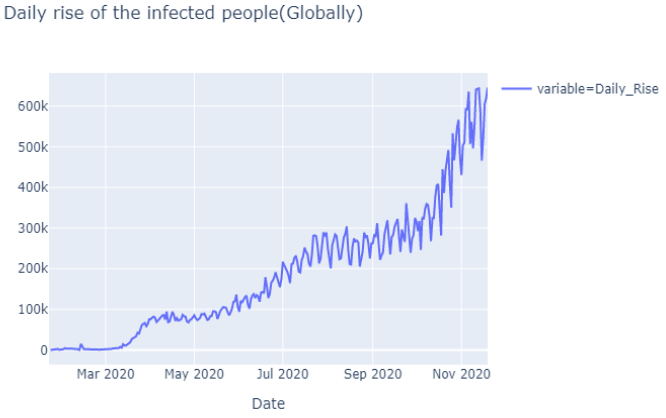


Figure 2

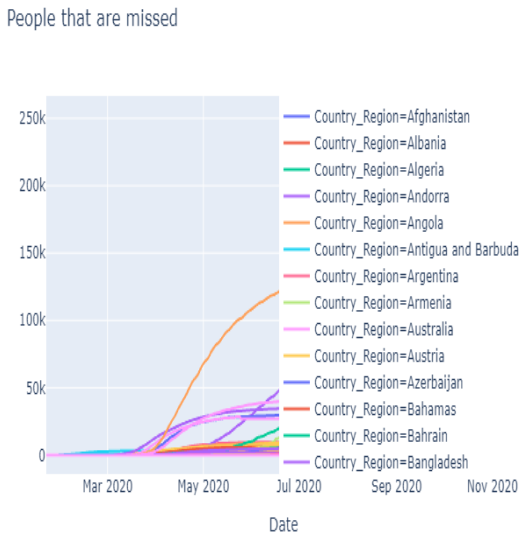


Figure 3

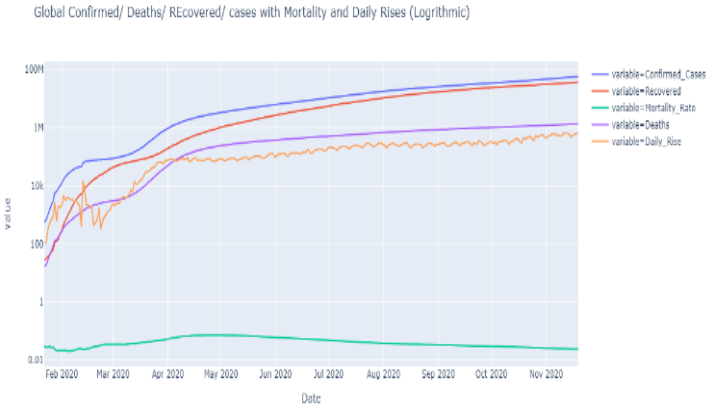


Figure 4

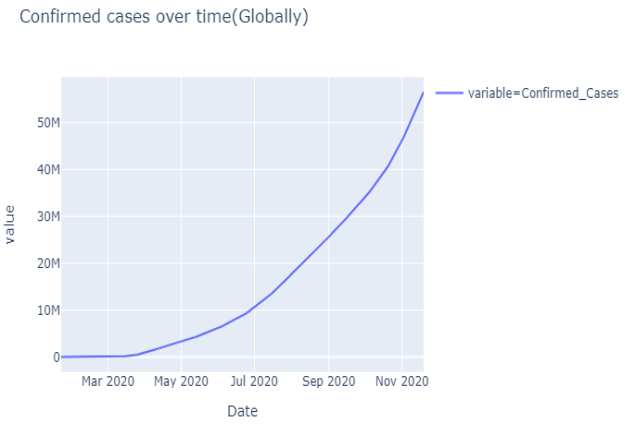


Figure 5

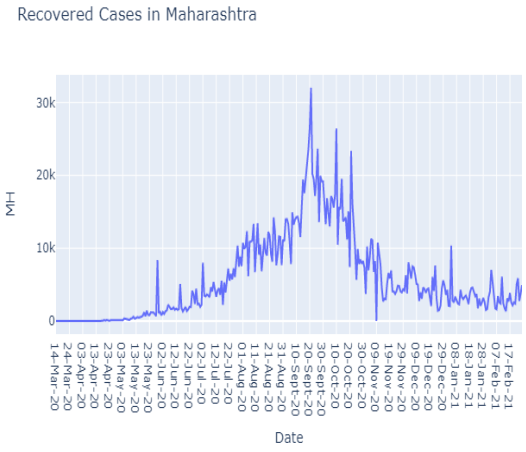


Figure 6

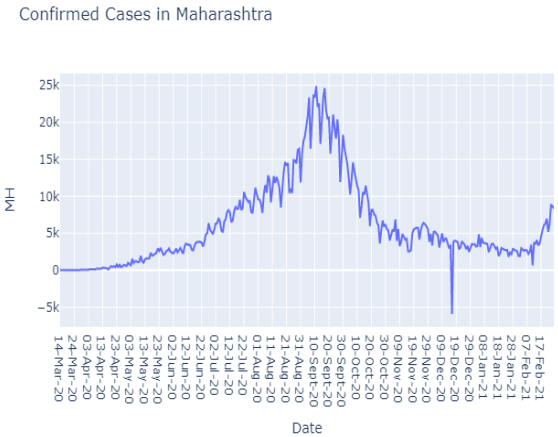


Figure 7

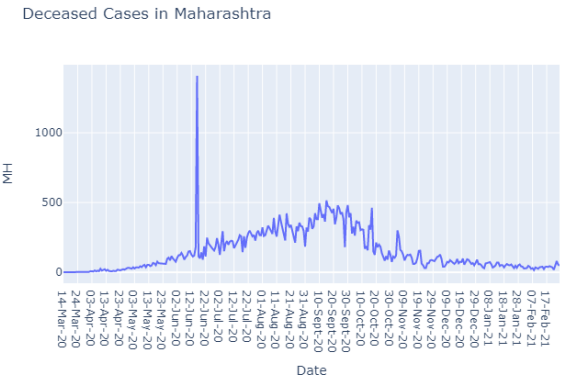


Figure 8

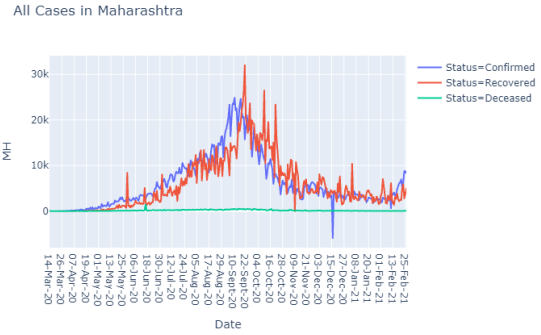
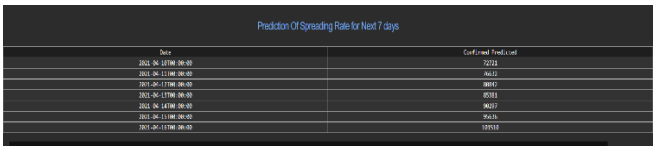
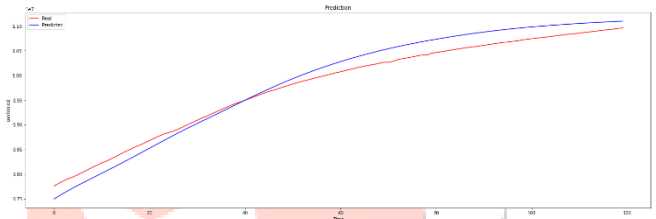
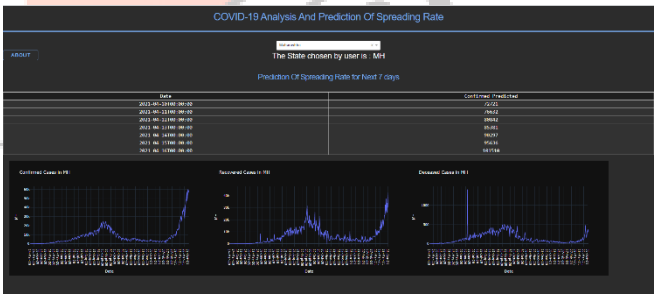


Figure 9







V. Reference

[1] Liu Y, Gu Z, Xia S, Shi B, Zhou XN, Shi Y, Liu J (2020) What are the underlying transmission patterns of COVID19 outbreak?—an age-specific social contact characterization. EClinical-Medicine 22:100354

[2] Temesgen A, Gurmesa A, Getchew Y (2018) Joint modeling of longitudinal cd4 count and time to-death of hiv/tb co-infected patients: a case of Jimma university specialized hospital. Ann Data Sci 5(4):659

[3] Olson DL, Shi Y, Shi Y (2007) Introduction to business data mining, vol 10. McGraw-Hill/Irwin, Englewood Cliffs

[4] Li J, Guo K, Viedma EH, Lee H, Liu J, Zhong N, Gomes LFAM, Filip FG, Fang SC, Ozdemir MS et al (2020) Culture ¨ vs policy: more global collaboration to effectively combat COVID-19. Innovation 1(2):100023

[5] Shi Y, Tian Y, Kou G, Peng Y, Li J (2011) Optimization based data mining: theory and applications. Springer, Berlin

[6] Annals of Data Science https://doi.org/10.1007/s40745-020-00314-9 Outbreak Prediction of COVID-19 for Dense and Populated Countries Using Machine Learning

[7] Kumar S (2020) Monitoring novel corona virus (COVID19) infections in India by cluster analysis. Ann Data Sci 7:1

[8] Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3):e0231236

[9] Bhatnagar MR (2020) Covid-19: mathematical modeling and predictions, submitted to ARXIV. http://web.iitd.ac.in/ manav/COVID.pdf

[10] E. Chen, K. Lerman, E. Ferrara, COVID-19: The First Public Coronavirus Twitter Dataset, arXiv preprint: 1603.07252 (2020).

[11] J.P. Cohen, P. Morrison, L. Dao, COVID-19 image data collection, arXiv preprint: 2003.11597 (2020).

[12] J.P. Cohen, P. Morrison, L. Dao, COVID-19 image data collection, arXiv preprint: 2003.11597 (2020).

[13] Poonia N, Azad S (2020) Short-term forecasts of COVID-19 spread across Indian states until 1 may2020. arXiv preprint arXiv:2004.13538