INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 7, Issue 1)
Available online at: https://www.ijariit.com

COVID-19 prediction and analysis of spreading rate

Sumit Rasal
sumit.rasal301@gmail.com
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

Rahul Mali
rahulmali65.rm@gmail.com
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

Ganesh Kannor
ganeshkannor88@gmail.com
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

Danish Majeed
danish.reshi.cs@ghrcem.raisoni.net
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

Shubham Patil
shubhamlpatil22@gmail.com
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

Sunita Nandgave
sunita.nandgave@raisoni.net
G.H. Raisoni College of Engineering and Management, Pune, Maharashtra

ABSTRACT

Covid-19 Disease started at the End of 2019 in china Wuhan city and Then spread in the whole world within three to four-month. 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 on a daily basis is increasing and it's created a lot of stress on governing bodies across the globe and they are finding it difficult to tackle the situation so they can handle the situation effectively. For solving this problem. We developed a COVID-19 Prediction And Analysis Of Spreading Rate system for every state and union territory of India. The purpose of the prediction model is to forecast the count of new cases likely to arise for successive 5 days using RNN and LSTM machine learning algorithms. A set of models for predicting the rise in new cases, having an average accuracy of 87.9% some time it will go above 90%. It was developed for every state and union territory of India. The highest accuracy of 93.93% was achieved for some states because of the availability of sufficient Data. For Eg. Maharashtra, Delhi. A data-driven approach with higher accuracy as here are often very useful for a proactive response from the govt and citizens. Through this project. we want to achieve our goal. Like 1.Analyzing the Spreading rate 2.Analyzing the recovery rate and mortality rate.3.Predicting future spreading rate for next 5 successive day.

Keywords: *COVID-19, RNN, and LSTM Algorithm, Machine Learning, Deep Learning*

1. INTRODUCTION

The SARS-CoV-2 coronavirus disease (COVID-19) orig-inated in Wuhan, China sometime during December 2019. Within a month, more than ten thousand people were infected by the virus and hundreds died due to the virus[1]. In the initial stage, the outbreak caused several deaths, as the medical systems were not capable of handling many seriously ill patients. Till 14 Dec 2020, there were 1.6 million deaths [2] reported across the world due to this pandemic, and in India 1.43 Lacks. In a rapidly evolving pandemic, improper analysis and predictions of the number of patients result in an inefficient distribution of medical resources and the wrong decision by the administration for controlling the virus spreading. Limited medical facilities and mismanagement of resource allocation can lead to additional severe cases and a decline in recovery rates and mortality rates. To cope with this situation, predicting the new cases which will arise in the future is very important and essential for a government body to make a decision. This can ensure optimal allocation of medical resources in the affected regions and the government body can take an effective decision like implementing the lockdown in affected regions. Data science in the predictive domain is an emerging field in any situation. In this study, we have incorporated the principles of data science and data analyst [3] for the prediction of COVID-19 progression. The outbreak of COVID-19 is a significant challenge for the Indian government or any government, concerning the capacity and management of public health systems to face the catastrophic emergency [4]. The prediction model can help hospitals, government, and healthcare management to properly allocate resources, thereby reducing the pressure and allowing the situation to be handled within the limited resources. 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 one time series of data-set. As a result, the models were unable to accurately predict the number of new cases for the next five 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 5 days. After getting a decent accuracy of 90 and above 90 Percent, we implemented these algorithms on data-sets of a different state. 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 territory. 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.

2. LITERATURE REVIEW

a lot of research is going for predicting the infection rate for the next 5 days or for 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 fit-ting 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 COVID-19 is high. Kumar et al. [7] have applied cluster analysis, to classify real groups of communicable disease of COVID-19 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-days-ahead point forecasts and prediction intervals. A susceptible-exposed-infectious-recovered (SEIR) metapopulation 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 COVID-19 in countries using various types of parameters and tested their model on real data of countries.[10,11] Machine Learn-ing (ML) and 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 the 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.

3. MODEL FLOW

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

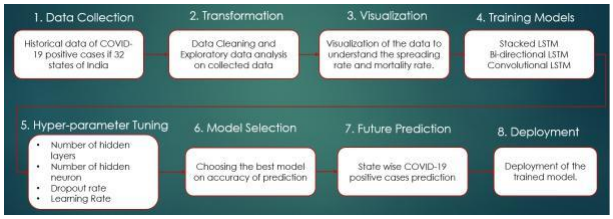


Fig. 1: Model Flow

A. Data Collection

We Need a Data for analyzing and predicting the spreading Rate of covid-19 for that we using the GitHub repository https://github.com/CSSEGISandData/COVID-19/tree/master/csse COVID 19 data/csse COVID 19 time series The Data we are collecting from the repository are the time series data and in this repository we have a Global confirm cases, death cases of covid-19. By using the pandas, we are directly fetching the data from the GitHub repository.

B. Transformation and scaling

In the transformation we did Data Cleaning and Exploratory data analysis on collected data. Transformation help us for collecting the relevant data and increase the accuracy. In our project we are collecting huge data of whole world. But we need only for India so we are using that much data only. If we find the missing data then we are removing the data or replacing the data with the mean and Categorical columns

C. Visualization

we doing the Visualization of the data to understand the spreading rate and mortality rate. Help us to understanding thing by visualization rather that number. It help us also for selecting the future in the data for taring the model. some Data visualization are done blow are some image of visualization

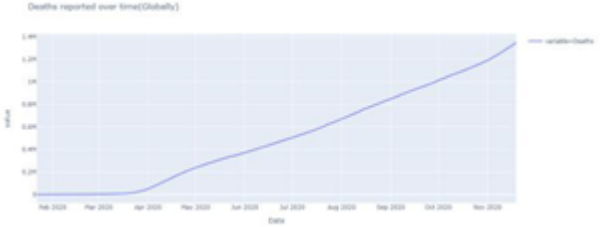


Fig. 2: Death Report over time

D. Training Models

For the training the model we are using the algorithm RNN

1) RNN-Recurrent Neural Networks: Recurrent Neural Network (RNN) are a type of Neural Network where the output from previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus, RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hid-den state, which remembers some information about a sequence. RNN has a "memory" which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

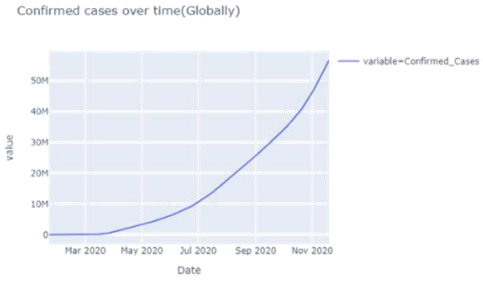


Fig. 3: Confirmed cases over time

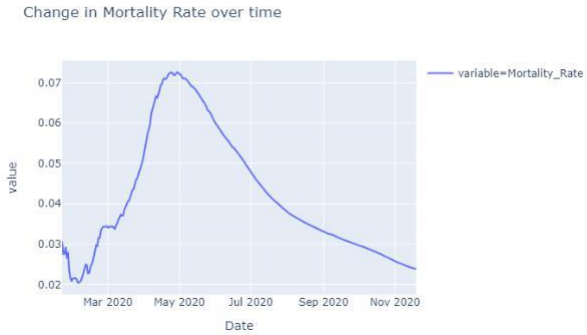


Fig. 4: Mortality Rate over time

2) Advantages and Disadvantages of RNN: -

Advantages: An RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short-Term Memory. Recurrent neural network are even used with convolutional layers to extend the effective pixel neighborhood.

Disadvantage: Gradient vanishing and exploding problems.

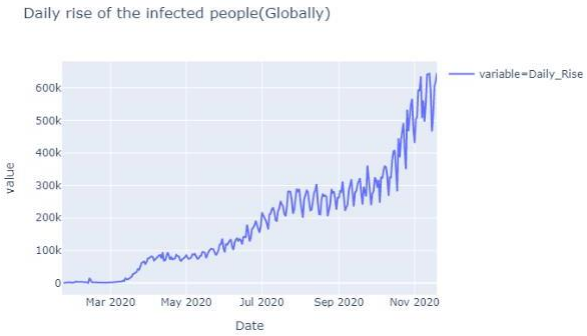


Fig. 5: Daily rise of the infection people

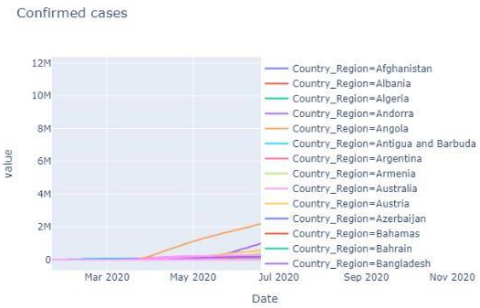


Fig. 6: Confirmed case

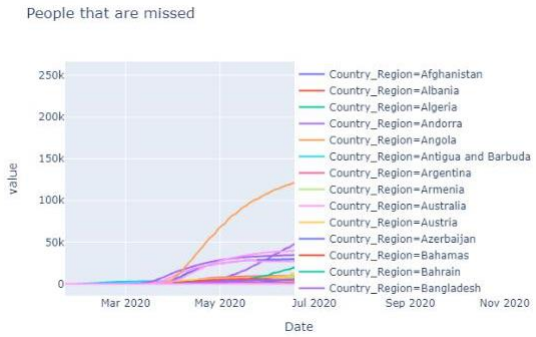


Fig. 7: People that are missed

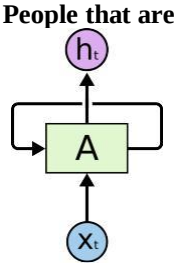


Fig. 8: Recurrent Neural Networks have loops.

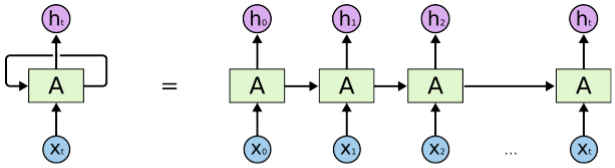
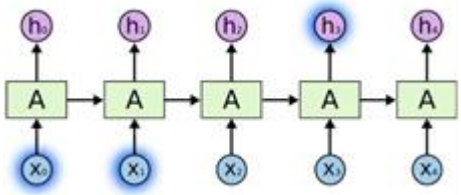


Fig. 9: Recurrent Neural Networks have loops

Training an RNN is a very difficult task.



E. Model Selection

1) LSTM Networks: Long Short-Term Memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory. The vanishing gradient problem of RNN is resolved here. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation. In an LSTM network, three gates are present.

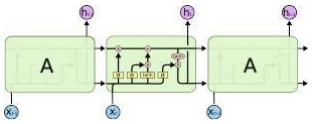
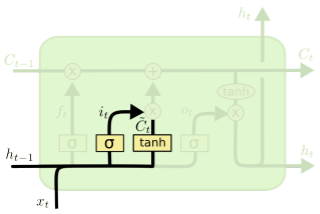


Fig. 10: The repeating module in an LSTM contains four interacting layers.

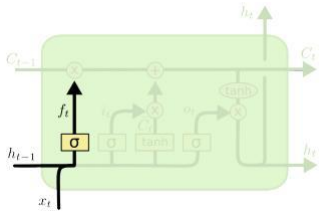
2) Forget gate: discover what details to be discarded from the block. It is decided by the sigmoid function. it looks at the previous state(ht-1) and the content input(Xt) and outputs a number between 0(omit this)and 1(keep this)for each number in the cell state Ct1.



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

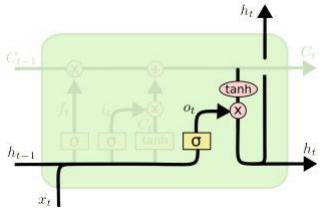
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

3) Input gate: discover which value from input should be used to modify the memory. Sigmoid function decides which values to let through 0,1. and tanh function gives weightage to the values which are passed deciding their level of importance ranging from-1 to 1.



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

4) Output Gate: the input and the memory of the block is used to decide the output. Sigmoid function decides which values to let through 0,1. and tanh function gives weight-age to the values which are passed deciding their level of importance ranging from-1 to 1 and multiplied with output of Sigmoid.



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

F. Future Prediction

In the future prediction we giving the prediction who many people will be infected by covid-19 ahead 5 day. so government body can take decision and save the people. our model give the prediction above the 90%.

G. Deployment

In the Deployment, we are using cloud technology. In cloud technology, we are using the AWS cloud for the deployment of the machine learning model. In the AWS cloud, we are using the ec2 instance for a compute unit, and then using the DevOps technology, we are training the model and deployed that inside the instance. After 24hr, we are training the model with new data and then again deployed the model using DevOps technology. DevOps Technology helps us to do the automation in the machine learning for Training the model and Deploying the model. In the DevOps, we are using Jenkins for doing the automation. For launching the Infrastructure we are using the terraform. because of automation, we get the right accuracy with the updated data.

4. CONCLUSION

By using the machine learning algorithm we analyzed the Covid-19 data and created the system for giving the spreading the rate for successive 5 days. This system helps the govern-ment body for taking the more efficient decision and control the covid situation.

5. REFERENCES

[1] Liu Y, Gu Z, Xia S, Shi B, Zhou XN, Shi Y, Liu J (2020) What are the underlying transmission patterns of COVID-19 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
[5] 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
[6] Shi Y, Tian Y, Kou G, Peng Y, Li J (2011) Optimization based data mining: theory and applica- tions. 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 (COVID-19) 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

