

# Machine Learning Algorithm to Forecast the Outbreak of Covid 19 using LSTM

1. Ketan Gaikwad, Computer Dept, MES College Of Engineering, Pune
2. Omkar Bhoite, Computer Dept, MES College Of Engineering, Pune
3. Sohail Ahmad, Computer Dept, MES College Of Engineering, Pune
4. Saurabh Wagh, Computer Dept, MES College Of Engineering, Pune
5. Dr. Shalaka Prasad Deore, Computer Dept, MES College Of Engineering, Pune

**Abstract:** Machine learning (ML) based on predicting and predicting algorithms and proving their value and importance in statistical prediction problems to improve the accuracy and results of a given problem. ML models have long been used in many application domains that required the identification and prediction of a number of information items. A few predictive methods are often used to manage prediction problems. This study will demonstrate the ability of machine learning models to predict the number of future patients with COVID-19, currently considered a potential threat to humanity. In particular, the implementation of the Long Short Term Networks algorithm is recommended to extend the efficiency of the currently available models. Three different types of speculation are made by each model, such as the number of newly infected cases, the death rate, and the recovery rate over the next 10 days.

**Keywords:** LSTM algorithm, Covid-19 Forecasting, ARIMA

## I. INTRODUCTION

This time series prediction problem has been considered and researched as a machine learning problem during this study, relying on other high-performance machine learning models such as regression to definition (LR), vector support (SVM), at least total reduction and selection operator (i - LASSO), and exponential smoothing (ES). training

models are trained using the COVID-19 patient statistical data set provided on the worldometer website. This database has been pre-processed and divided into two sub-sets: training set (85% records) and test set (15% records).

In the current crisis in humanity, our aim during this study is to develop the COVID-19 predictive system. Prediction completed three major disease variables in the next 10 days the number of newly confirmed Covid cases, the total number of deaths, the recovery rate using the Long Short-Term Memory (LSTM) algorithm.

Since the start of COVID-19 (corona virus disease 2019), government bodies all over the world have implemented several new policies and travel restrictions to curtail its spread.

The timely changes in policies and decisions made by the govt. helped in decelerating the outbreak of the COVID-19 virus to an outsized extent. Even with the help of these decisions, controlling the pandemic spread is a challenge. Future predictions about the outbreak may be helpful for future policy-making, i.e., to implement new plans and control the spread of COVID-19.

Further, it's observed throughout the globe that Covid positive people with no symptoms play a significant role with the spread of the disease. This has motivated us to incorporate such cases for accurate prediction of the trend of spread.

## II. LITERATURE SURVEY

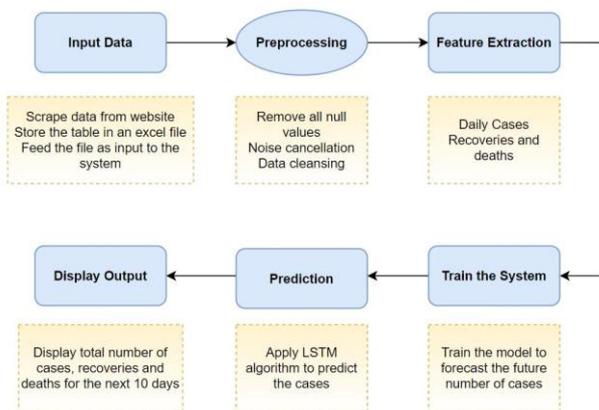
In this paper [1], during this paper the algorithm used is simple regression. the prediction and outbreak is taken into account to be a regression problem and it's implemented using two regression models namely linear and polynomial regression. The Covid-19 India data set is used for implementation purposes it predicts the number of confirmed new cases, deaths and recoveries based on data available from 12th March 2:30 1st October 2020. for the forecasting purpose and analysis statistic forecasting approach is employed. The output of this implementation suggests that polynomial regression shows better results than simple regression. The forecasting was done using Tableau and also the results are satisfactory, but the results may be more accurate upon use of a much bigger dataset considered for a extended duration. In Paper [2], the algorithm of the Supervised Machine Learning Model algorithm is used for prediction in this paper. Such models have long been used in many applications that require the identification and prediction of any kind of adverse factors or analysis of threats. A number of forecasting methods are used to deal with forecasting problems. This study demonstrates the potential of ML models to predict the number of future patients affected by COVID-19 using the Modeled Study Models. In particular, four common predictive models, such as vector support (SVM), linear rotation (LR), operator reduction and operator selection (LASSO), and exponential smoothing (ES) were used in this study to predict outbreaks. COVID-19. Three types of speculation are made by each model, such as the number of newly infected cases, the number of deaths, and the number of recovery over the next 10 days. The results produced by the study prove that it is a promising way to apply these methods in the current context of the COVID-19 epidemic. The results prove that ES performs well among all used models followed by LR and LASSO which performs well in predicting new confirmed cases, mortality rate and recovery rate, while SVM does not perform well in all

predictive data delivery conditions. This paper [3] analyzes the continuation of COVID-19 in India and the three most affected Indian regions (i.e., Maharashtra, Tamil Nadu and Andhra Pradesh) up to 29 August 2020 and develops a predictive and predictable approach to predicted COVID-19 behavior will be distributed in the coming months. They used time series data in India and used the Susceptible-Infective-Removed (SIR) model and the FbProphet model to predict the highest incidence of infection and the highest date of infection, i.e. the day with the highest surgery in India and the third highest. Covid 19 affected provinces. In this paper, they have performed a comparative analysis of predictive effects from SIR and FbProphet models. In this study, assuming that at least 5 percent of Indians are likely to be infected with covid-19, it is concluded that the global outbreak is expected to reach a high point by the end of Nov. 20. There is a good chance of an increase in the number of cases, especially if social cuts and other forms of control can be followed by people in the coming months.

## III. PROBLEM DEFINITION

The need for a prediction system to forecast the Covid situation is on the rise. So, we propose a machine learning algorithm to forecast the outbreak of Covid 19. From the literature survey we have done, it is evident that the available systems can be modified to a better system with higher accuracy and efficiency. This system can be implemented using LSTM (Long Short-Term Memory) based on real-time dataset.

#### IV. ARCHITECTURE DIAGRAM



**Fig 1. Architecture Diagram to Forecast the Outbreak using LSTM**

The proposed system developed a new forecasting system based on LSTM. In this system we are forecasting covid data for next 10 days.

**Algorithm: LSTM**

##### Module 1: Input Data

The data from the corona virus website is scrapped using API and database fetch queries and stored in a csv file. The scraped data will contain details including date, total number of cases, new cases, total number of deaths, new deaths, total recoveries, new recoveries, population, % of population affected and so on.

This csv file will be fed as input to the system.

##### Module 2: Pre - Processing

The csv file contains some null values and irrelevant information which needs to be cleansed. The `iloc()` and `shape()` functions can be used to remove the null values. Noise cancellation and data cleansing is done in this step.

##### Module 3: Feature Extraction

The only relevant data columns from the csv file are the new confirmed daily cases, new deaths and new recoveries. Only these features from the file will be extracted for further processing.

##### Module 4: Prediction

LSTM algorithm is applied to the extracted features to train the system and based on the time series forecasting, the prediction for the next 10 is generated.

##### Module 5: Output

The total number of new cases, estimated deaths and recoveries per day are displayed for a span of next 10 days.

#### V. CONCLUSION

In this study, a systematic study based on machine learning was proposed to predict the risk of COVID-19 outbreaks worldwide. The system analyzes a set of data containing actual past recorded daily data and makes predictions for future days using the LSTM machine learning algorithm. These includes algorithms like Linear Regression, LASSO, FbProphet and SIR. Neural networks is wildly used machine learning algorithm now a days but its implementation in this field is minimal. Thus use of LSTM which is based on recurrent neural networks and considers real-time changes in the database before prediction will be helpful and can give better results than the existing systems. Also, LSTM has better results for multivariate data and can predict the outbreak more precisely and efficiently.

#### VI. REFERENCE

- [1]. S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburg, "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 989-995, doi: 10.1109/Confluence51648.2021.9377137.
- [2]. F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in

- IEEE Access, vol. 8, pp. 101489-101499, 2020, doi: 10.1109/ACCESS.2020.2997311.
- [3]. N. Darapaneni, P. Jain, R. Khattar, M. Chawla, R. Vaish and A. R. Paduri, "Analysis and Prediction of COVID-19 Pandemic in India," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 291-296, doi:10.1109/ICACCCN51052.2020.9362817.
- [4]. Zhan C, Tse CK, Gao Y, Hao T. Comparative Study of COVID-19 Pandemic Progressions in 175 Regions in Australia, Canada, Italy, Japan, Spain, U.K. and USA Using a Novel Model That Considers Testing Capacity and Deficiency in Confirming Infected Cases. IEEE J Biomed Health Inform. 2021 Aug;25(8):2836-2847. doi:10.1109/JBHI.2021.3089577. Epub 2021 Aug 5. PMID: 34129512.
- [5]. S. Chordia and Y. Pawar, "Analyzing and Forecasting COVID-19 Outbreak in India," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 1059-1066, doi: 10.1109/Confluence51648.2021.9377115.
- [6] Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. COVID-19 Outbreak Prediction with Machine Learning. Algorithms. 2020; 13(10):249. <https://doi.org/10.3390/a13100>