JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction of Covid-19 Cases Using Machine Learning

[1]Colin Gonsalves, [2]Dr. Prashant Nitnaware

[1]Student, [2]Project Guide
[1]Department of Information Technology,
[1]PCE, Mumbai, India

Abstract : Covid 19 being the global pandemic, it has become an absolute necessity to make sure the epidemic is controlled and the state of the world with respect to health, finance and livelihood come to a better position. The motive of this research aims to contribute towards achieving this state. Every Local Body/Municipality, of a country, needs to be well equipped with the best technology available. Machine learning will play an important role to help Local Body/Municipality predict the trend of Covid19 cases based on the figures collected in their area. This tool will help the Local Body prepare for medical facilities, Administrative decisions and Citizen lifestyle. Making use of five ML algorithms, namely, Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), FB Prophet, Linear Regression (with FB Prophet) and Random Forest (with FB Prophet). Visualizing the trend on graph, testing the prediction against actual, ruling out the better suited algorithm and using the best of five to predict the coming days.

The country's count is derived from the sum of municipal and state statistics. These figures are displayed in front of the entire country to help people comprehend the current situation. There are several web portals that give data visualisation to show the current trend. Some of these sites, such as amso, make predictions based on the information they have. The example of a prediction site is shown above. The municipality-level forecast computation is done in reverse chronological sequence.As a result, when the country's production is conducted as "X," the country data is used to compute the same. The percentage rise or reduction is computed, and the results are applied to the data from the municipality. The same is done for the next 15 to 20 days' worth of data, which is then provided for analysis. Though this is a workaround, it has been in use for quite some time. Because the technology isn't available, the local government must rely on such computations.

This study estimates the total number of active cases in India over the next 20 days using machine learning. According to the conclusions of this study, the ARIMA and Random Forest models are most suited to this situation. The model's current-state projection will be beneficial in forecasting the future. Overall, this research can help authorities acquire caution, which could help control the COVID-19 outbreak.

I.INTRODUCTION

Belonging to a family of coronavirus, Covid 19 came into known in the year 2019. Thus, obtaining the abbreviation Covid 19 for CoronaVirisDeseases 2019. Similar to the rest of the viruses in the family, Covid 19 attacks the respiratory system of the human body. Originating from the Wuhan province in China, it gained momentum and spread across the world. Considering the behaviour and nature of the virus and the rate of spread, the World Health Organization (WHO) declared it as a global pandemic. Along with this pandemic announcement came a set of rules which restricted travel, imposed social distancing and wearing a mask. These steps were imposed to limit or control the spread of virus

India registered its first case on 30th January, 2020 in the state of Kerala. India underwent lockdown to harness the spread of the virus. Registering upto 90 Thousand cases highest in the month of September 2020 and 4 Lakh cases in May 2021 per day, marking the highest peak for the two waves. With availability of vaccines, India initiated the drive in January 2021 which included British Oxford–AstraZeneca vaccine (Covishield), The Indian BBV152 (Covaxin) and Russian Sputnik V vaccine (Emergency use).

Learning about Covid 19, a huge magnitude of data is generated. This data is broadly categorized into state and county data, growth rate variation, covid-19 projections, lab data, disparities and at-risk populations, economic data, mobility, lab data and mortality. Taking a part of the mentioned data available, we can use machine learning techniques to implement certain algorithms to perform predictions under test and actual environment.

Machine learning plays an important role in the medical sector in helping apply predictions based on certain algorithms. These algorithms used depend on how and what the data is. In this report we have tried to understand five machine learning algorithms and how they have been helpful in the past. These algorithms will perform predictions based on the data provided as reference. Algorithms such as Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), FB Prophet, Linear Regression and Random Forest Regressor have proved to be more popular and better with predictions. Random Forest has been more popular with the past research on other diseases such as pneumonia, diabetic, etc. Whereas FB prophet has gained popularity more specific to Covid-19.

As we have seen in the past, the first wave and second wave in India has tested the best effectiveness of our medical facility. Adding additional support to this facility will only provide benefits. Allowing this accessibility, a prediction tool made handy to the local bodies of the state, will help us prepare for medical requirements, administrative actions and citizen awareness.

There would be certain challenges that we expect the Municipality to face but the benefit outweighs the challenges. In the current scenario, the municipality has to depend on the figures available on the internet. The online prediction portal available is performing predictions on the country or state level data. calculation based on prorata is presented on the municipality level. Though this is a workaround, the same cannot be considered as appropriate. Also considering these tools are not easily available in the market and the ones available are expensive. Making such tools available will not only provide support to the municipality but also to the nation indirectly.

Municipality bodies in India do not have easy access to the prediction tools to predict the rise or fall of Covid 19 new cases on a day to day basis. These bodies have to be dependent on figures available on the internet. The drawback of these figures are that they are available at country/state/city level. Very few municipalities have their own tool to perform predictions. Considering these scenarios, having such tools made available at the local body level in itself will provide a greater impact. Close to correct prediction will help the medical facility, administrative department and citizens be prepared for the coming day. Medical facilities such as availability of hospital beds, oxygen cylinders, medical personnel, etc. Administrative department can focus on implementing lockdows accordingly. Individual citizens can be made aware to take necessary actions. Taking care at the local level will add to the hierarchy of City, State, Country and World. Hence making such a facility available is indirectly protecting the human kind.

Considering the algorithm that will be used to perform the prediction:

Random Forest Regressor has shown better results in the past, comparatively. Random forest is a learning method that is supervised. It creates a "forest" out of an ensemble of decision trees, which are generally trained using the "bagging" approach. The bagging method's basic premise is that combining several learning models improves the final outcome.

ARIMA (autoregressive integrated moving average) is a statistical analysis technique that use time series data to better understand a data collection or forecast future trends. If a statistical model predicts future values based on previous values, it is called autoregressive.

The control flow of an LSTM is comparable to that of a recurrent neural network. It processes data and passes information on as it moves along. The processes within the cells of the LSTM vary. The LSTM uses these processes to remember or forget information.

FBProphet aims to fit several linear and nonlinear time functions as components using time as a regressor. By default, FBProphet uses a linear model to fit the data, but it may be switched to a nonlinear model (logistics growth) using its parameters. There are a sufficient number of missing or outlier data points.

Linear regression is a supervised learning-based machine learning technique. Linear regression is a mathematical technique for predicting the value of a dependent variable (y) based on the value of an independent variable (x). As a result of this regression approach, a linear connection between x (input) and y (output) is discovered (output).

1.1 Introduction to Machine Learning

Machine learning (ML) is a subset of artificial intelligence (AI). Machine learning (ML) is the process through which a computer learns from its previous experiences. Unlike traditional programming, machine learning does not require explicit programming. Machine learning aims to develop programmes that can learn from the data they are given. It also has anything to do with personal development programmes. To begin the learning process, data or observations are needed. The information or instruction set provided is helpful in the learning process. The main goal is to enable computers to learn on their own, without the need for human interaction, and to adapt their behaviour accordingly.

ML is a branch of Artificial Intelligence that can anticipate a variety of facts based on input parameters. Machine learning is used in industries such as stock exchange markets to predict whether the price of a share will fall or rise based on a set of inputs. This dissertation will use machine learning techniques to predict the trend of covid 19 cases. Depending on the data loaded for a specific area, the system is expected to project the future trend of the cases that will be registered. This will also provide an opportunity to compare techniques with each other based on 80:20 ratio. The better suited technique can help to perform the actual prediction on the 100% data available. This prediction will be based on the input parameters of date, count, etc. The classification of machine learning algorithms is presented below:

1.1.1 Supervised Algorithms

ML algorithms that are supervised apply previously learnt experiences to fresh data. To produce predictions, supervised learning systems employ labelled data. The algorithms begin by learning from a set of previously known datasets. It even goes so far as to anticipate the data of unknown values using the set of rules that has been developed. After enough training, the system can provide outputs for fresh input. The predictions are compared to the actual output by the learning algorithms. Further adjustments are made to account for the inaccuracies in the output.

1.1.2 Semi-Supervised Algorithms

Semi-supervised machine learning techniques fall in the middle of the supervised and unsupervised learning spectrum. For training, they employ both labelled and unlabeled data. These approaches improve the performance of systems over time. Semi-supervised algorithms are frequently given additional unlabeled data. When a competent intervention is necessary, semi-supervised learning is utilised. The necessity for unlabeled data does not necessitate any additional expert actions.

1.1.3 Unsupervised Algorithms

Unsupervised machine learning algorithms are ones in which the data is not labelled. The data fed to the unsupervised algorithm is all unlabeled. There is no such thing as a right or wrong categorization. The programme looks at the data and divides it into clusters.

1.2 Motivation

Municipality bodies in India do not have easy access to the prediction tools to predict the rise or fall of Covid 19 new cases on a day to day basis. These bodies have to be dependent on figures available on the internet. The drawback of these figures are that they are available at country/state/city level. Very few municipalities have their own tool to perform predictions.

Considering these scenarios, having such tools made available at the local body level in itself will provide a greater impact. Close to correct prediction will help the medical facility, administrative department and citizens be prepared for the coming day. Medical facilities such as availability of hospital beds, oxygen cylinders, medical personnel, etc. Administrative department can focus on implementing lockdows accordingly. Individual citizens can be made aware to take necessary actions.

Taking care at the local level will add to the hierarchy of City, State, Country and World. Hence making such a facility available is indirectly protecting the human kind.

1.3 Beneficiaries

Covid 19 predictions are helpful to make sure necessary medical facilities are made available, necessary guidelines are set and citizens are made aware. Considering these aspects, local bodies/municipalities, State government, Central government, may use this tool. As we have gathered information, the local bodies do not have access to such tools as the state and central government do considering the cost and maintenance involved. Hence Municipal bodies will be the first level of users that will gain advantage from the information made available using this tool. They are also the first to report the data to the state and central government. Cumulative of all the local bodies is the figures that are presented for the state and country.

1.4 Problem Statement

Considering the current system, the method used cannot be termed as appropriate. Though it is a workaround to fill the gaps present. Hence the primary agenda is availability. To make this tool available to the municipality as they do not have easy access to perform prediction on the data available with respect to their own area.

Data visualization itself takes time, as the data needs to be furnished and calculated. This visualization itself has its own challenges when it comes to performing comparison. Currently municipalities do not have a scope to choose which algorithms best suit their area calculations. This can be resolved by providing options to perform operations based on different algorithms and compare the prediction of each. The best suited one can be used to perform the actual coming days prediction.

1.5 Scope of the Project

The agenda/scope of this project is to make this tool able to visualize 1) the current state of the Covid 19 cases. 2) Perform tests on the mentioned prediction algorithms. 3) Compare prediction test results against 80:20 ratio. 4) Perform prediction on 100% data on the better-resulted algorithm. This requires the tool to have a pre furnished .csv file ready wil will contain information of one selected area. This prediction will be performed on the data available on the file for that area only. Failure to preformat the file of incorrect data will result in ill-judged prediction.

II. LITERATURE SURVEY

For the purpose of this research, past research was referred to as providing a base and compounding effect to add better value. The search was categorized to look for research work done on Machine Learning, Prediction algorithms and Covid 19. Machine learning research in to health care include Pneumonia Detection by Chouhan V. (2020), Diseases Classification by Sharmila S. (2017), Chest diseases diagnosis by Er O. (2010), Role of machine learning to predict the outbreak of covid-19 by Tiwari U.K. (2020) etc. Diseases prediction research include Building predictive models for MERS-CoV infections by Turaiki I. (2016), Disease prediction using machine learning by Shirsath S.S. (2018), prediction of the epidemics trend of COVID-19 by Yang Z.F. (2020), Predicting active pulmonary tuberculosis by El-Solh A.A. (1999), Predicting the Growth and Trend of COVID-19 by Tuli S. (2020).

When we consider past works that have been performed on the algorithms, a summary for the same is provided below:

2.1 Evolution of Pandemic Forecasting

Time-series techniques, compartmental epidemiological models, agent-based models, metapopulation models, and metrology approaches are all types of forecasting methodologies for pandemic evolution. Machine learning (ML) and deep learning (DL) techniques are two new additions to this extensive list. ARIMA models were proposed by Soebiyanto, Adimi, and Kiang (2010) for one-step forward forecasting of influenza weekly cases. Andersson et al. (2008) advocated using regression methods to forecast the peak timing and volume (of cases) for a pandemic, citing promising empirical data from seven outbreaks to support their claim (in Sweden).

Yaffee et al. (2008) contrasted casual techniques with 16 time-series univariate methods in an exhaustive examination of several time series methods for forecasting the course of an epidemic using data from CDC8, and found that univariate methods were better at prediction than causal models. Petropoulos and Makridakis (2020) used ETS (Hyndman, Koehler, Snyder, & Grose, 2002) models to forecast the number of cases on a worldwide basis for COVID-19. In terms of true accuracy, they claimed excellent outcomes for both their point predictions and the prediction intervals they gave. This is an open-access paper in PLOS ONE that has already received a lot of attention despite the fact that it has only been published for two months, demonstrating the interest in and relevance of quantitative research in academia and practise. Finally, there have been a number of research concentrating on forecasting fatalities in the United States and Europe for the first wave of the COVID-19 pandemic in the coming months.

It's tough to think about planning for the next pandemic when one is in the middle of one, especially when the time is unknown. COVID-19 is often regarded as a "once in a lifetime" or "once in a century" pandemic; nevertheless, this is not the case. But when a future epidemic threatens, what is the real return time, and how can we assure that the world is ready the next time?

According to a team from Metabiota, the next pandemic may not be as far away as we assume. Human activities and their influence on the environment, according to historical models, are driving an increase in the frequency and severity of epidemics produced by animal zoonoses. They calculate that the chance of a future zoonotic spillover event resulting in a pandemic of COVID-19 scale or greater is between 2.5 and 3.3 percent each year. In other words, another outbreak of the scale of COVID-19 has a 22-28 percent probability of occurring within the next 10 years, and a 47-57 percent chance of occurring within the next 25 years.

Pandemic planning and preparation has taken place in governments and international organisations. Though the World Health Organization issues recommendations and guidelines, there is no ongoing system to assess nations' epidemic preparedness and quick response capabilities. National government activity is reliant on national governments. The governments of the United States, France, the United Kingdom, and others controlled strategic health equipment stocks in the years leading up to and during the 2009 swine flu pandemic, although they often cut supplies after the epidemic to save money.

2.2 Theoretical Implications, Research Problems, and Methodological Approach

The aforementioned research problems must be addressed during the epidemic, not later, emphasising the necessity of our current inquiry. This disclaimer is a contribution in and of itself, demonstrating the accessibility, reactivity, and timeliness of OR research.

We use an exploratory methodological approach to identify the best forecasting techniques because we don't have a set of formal hypotheses that tell us which methods/models will perform better. The surplus demand of items and services, which is driven by the rise of COVID-19 instances, is then forecasted through a series of simulations. Our research covers a significant portion of the current pandemic wave, from January 22 to April 15, 2020. We contribute to the stream of Phenomena-based research from a methodological viewpoint since we are involved in the very early stages of a scientific inquiry, observing, studying, and giving answers for the development of a unique phenomenon.

We also make contributions to Operations Research (OR) and Supply Chain Management (SCM). For the former, we conduct a comprehensive empirical study to determine the best reliable approach for projecting growth rates during a pandemic. We do so in the early stages of the phenomena, before the start-growth-maturity-decline process is completed. We contribute to the latter, the field of SCM, by giving an input that is required by decision-making algorithms including stock-control, replenishment, advance purchase, and even rationing, i.e. scenarios requiring a mean predicted demand over the lead-time. We also give simulations for the pandemic's surplus demand for items and services.

Finally, we add to predictive analytics theory by proposing novel data-driven prediction techniques. We accomplish so by utilising machine-learning clustering techniques and drawing on theory from non-parametric regression smoothing on Nearest Neighbors. To anticipate the pandemic's development, we draw on the experience of nations where the epidemic broke out earlier. We also assist policymakers by considering the influence of political choices, particularly the imposition of a lockdown/curfew11, on the pandemic's development as well as the resilience of the impacted supply networks.

2.3 Forecasting Methods

A collection of competing models may be found in Table 1. For further information on these popular approaches, interested readers should refer to either the article on the most recent forecasting competition (the M4 competition – Makridakis et al., 2020) or Hyndman and Athanassopoulos' free online forecasting textbook. In Appendix A, we offer a brief overview of the more advanced machine- and deep-learning approaches.

Sr No	Category	Methods
1	Time Series	Naïve, Moving Averages , ARIMA
2	Machine Learning	Multiple Linear Regression (MLR), Random forest (RF), Support Vector Machine (SVM).
3	Deep Learning	Long-Short Term Memory networks (LSTM)

Table 2.1 Forecasting Methods

III. Proposed System and Implementation

The proposed system is based on a machine learning framework. Implementing five machine learning tools to learn, identify and predict the pattern of covid 19 cases being registered. This system intends to implement ARIMA, LSTM, FB Prophet, Linear Regression and Random Forest Regressor.

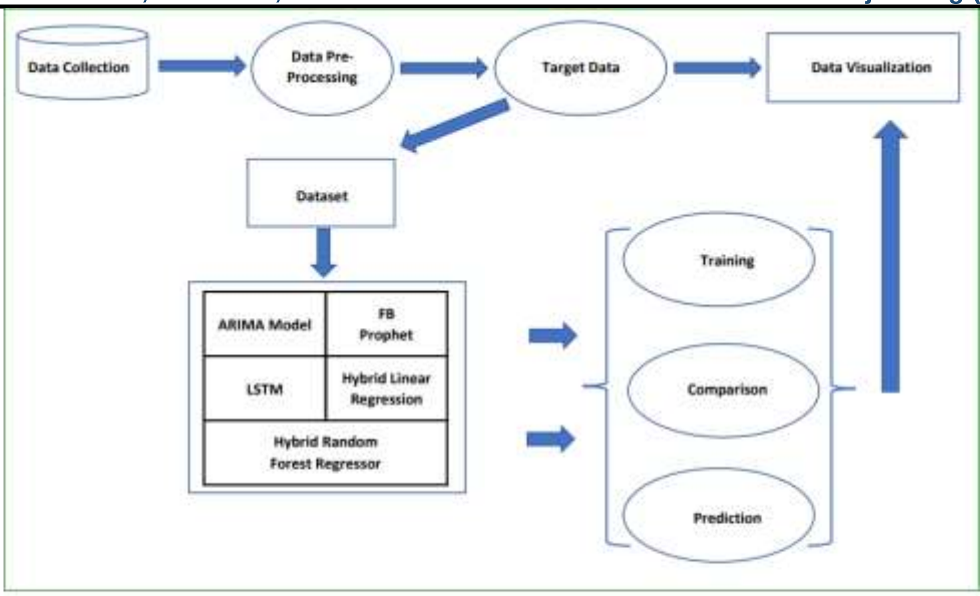

Fig 3.2 Proposed System

3.1. Data Collection

The process for collecting the data for the required operation from one or more sources is termed as data collection. The raw collected data in itself may be in one or multiple formats. All these formats of data are collected and placed ready. These sets of data formats are also known as datasets. For every machine learning system, data is the key ingredient. This will act as the input for the algorithm and relative to it the output will be derived.

We intend to use the data from the below mentioned source:

https://api.covid19india.org/

Covid19India.org makes the data available in a comma delimited file. Depending on the date range, the number of files are downloaded. The data is made available at City/Municipality level. Mumbai (BMC), Thane (TMC) and Palghar (VVMC) are the main raw data points that will be used to extract the

New Cases for the Day

New Deaths for the Day

New Recovered for the Day

In the next stage we will understand how the data will be furnished.

3.2. Data Pre-Processing

For processing the file, below is the accepted format with fields as:

ID,Day,Month,Year,Area,New,Recovered,Deceased

Pre-processing of files involves making the data ready and furnished for operational use. Defining this format is what can be called the main dataset for the prediction. The furnishing of the dataset will involve manual work or an automated script to perform the task. For the current scope, we will proceed with manual intervention to prebate the dataset. Explaining the fields:

ID : Unique Identity for the record

Day: Day of the Month

Month: Month of that Year

Year: Year of the Day and Month

(Bringing together to form dd/mm/yyyy)

Area: Name of the Municipality

New: Total New Patient Count for that date and Area

Recovered: Total Recovered Patient Count for that date and Area

Deceased: Total Deceased Patient Count for that date and Area

3.3. Target Data

Based on the data collected and furnished, it will be used to be imported. Data import necessarily requires the pre-set format. This allows the application to proceed with the functionality and the values that each of the fields hold. This imported data will be confirmed and taken for processing as per further requirements.

Supported format of file will be a comma separated (csv) file which will act as a read only database. The visualization depends on how the data is read by the instructions defined. These instructions we will understand further in detail in our next segment under Training Set and Prediction.

3.4. Data Visualization

3.4.1 Training Set

Data visualization is defined as presenting the data in an elaborative, readable and understandable way. The collected data is available in a comma separated which is readable but not quite explanatory. Hence for demonstration purposes, data visualization tools are used to make them understandable in public or closed forums.

Training Set is a part of data visualization. This can also be termed as a testing phase to verify the comparison and effectiveness between mentioned algorithms. We propose to use five algorithms in the training set helping us understand the effectiveness of each of the five algorithms.

Autoregressive Integrated Moving Average (ARIMA)

Long Short-Term Memory (LSTM)
FB Prophet
Linear Regression Hybrid (using FB Prophet)
Random Forest Regressor Hybrid (using FB Prophet)

The training set module is used by the data analyzer to make sure that the data is well tested and the prediction may work as per the expected. This provides the insights to tweak the algorithm pr classify based on the prediction vs actual. For this application we will work based on an 80:20 ratio. This will allow the application to use 80 percent of the sata as base reference and perform the next 20percent prediction. The 20 percent prediction will be compared with the actual values to verify now the algorithm is performing.

These selections may vary on multiple factors. but more importantly it depends how the actual vs prediction is portrayed. The data analyzer will then use the selection to actually predict the future trend. We shall understand the prediction module better in the next section.

3.4.2 Prediction

Prediction module and training dataset module have a dependency. The prediction module depends on the data analyzer to use the best effective algorithm out of the five in the training set. The selected is the one that will perform the prediction based on the instructions set. Even though the prediction module will only be used for one algorithm, we will provide the facility to verify the other algorithms to be used as reference.

The prediction module will make use of the same algorithms as listed below:
Autoregressive Integrated Moving Average (ARIMA)
Long Short-Term Memory (LSTM)
FB Prophet
Linear Regression Hybrid (using FB Prophet)
Random Forest Regressor Hybrid (using FB Prophet)

IV. Proposed System

The COVID-19 pandemic's precarity will result in a huge worldwide disaster. Experts and government organisations all around the world are concerned about a pandemic that may affect a large portion of the world's population. The potential of a COVID-19 epidemic was proposed as a worldwide ML-based prediction approach in this work. The system examines data sets comprising real-time data from previous days and uses machine learning techniques to anticipate future days.

Considering the use case as per chapter 4, we saw how the record file Mumbai4.csv can be used to serve as an input for the application and perform the set operation as per the application. In this chapter we will go step by step on how the file can be used and the results can be scaled to meet the user requirement from the user's perspective.

4.1 Home Page

Selecting the file Mumbai4.csv as input on the home page.

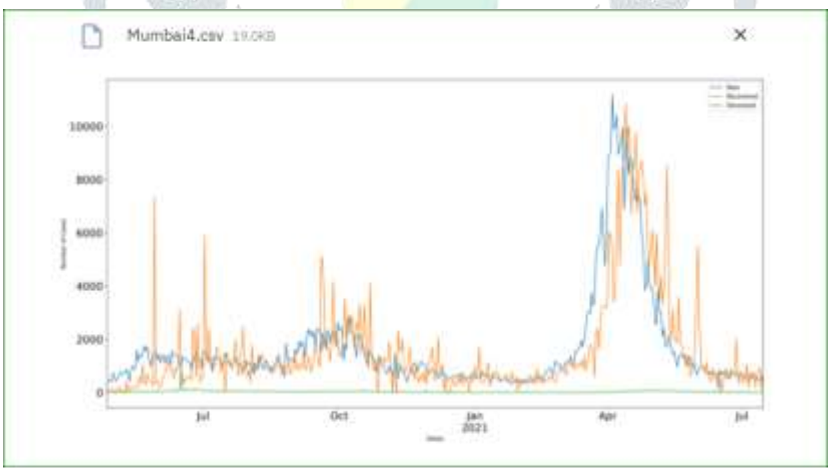

Fig 4.1 Graph View

4.2 Machine Learning Algorithm (Training Data)

The data you use to train an algorithm or machine learning model to anticipate the outcome you want it to predict is known as training data. Your data will be enhanced with data labelling or annotation if you use supervised learning or a hybrid technique that includes it. Test data is used to evaluate the algorithm you're using to train the machine's performance, such as accuracy or efficiency. You may use test data to assess how well your model can predict new responses based on its previous training. Machine learning models require both training and test data to improve and validate.The data you use to train a machine learning algorithm or model to properly anticipate a certain outcome, or response, is referred to as training data. In supervised learning, choosing and labelling the features in the data that will be used to train the machine requires a person in the loop. Unsupervised learning finds patterns in unlabeled data, such as conclusions or grouping of data points, using unlabeled data. Semi-supervised learning combines supervised and unsupervised learning methods.

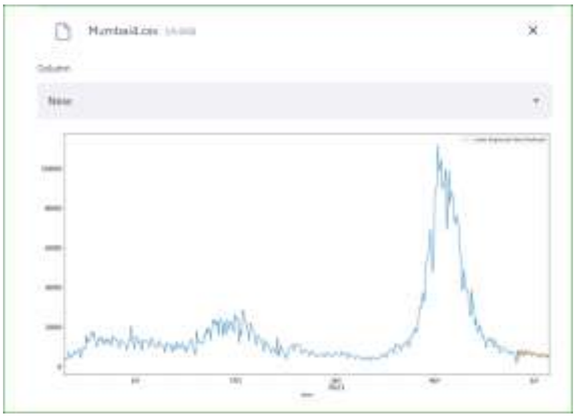

Fig 4.2 Linear Regression (Training Data)

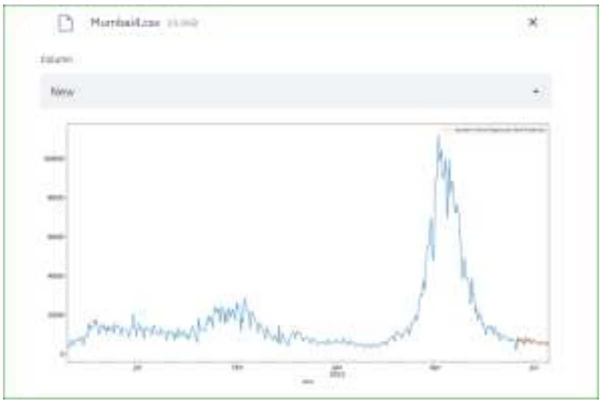

Fig 4.3 Random Forest (Training Data)

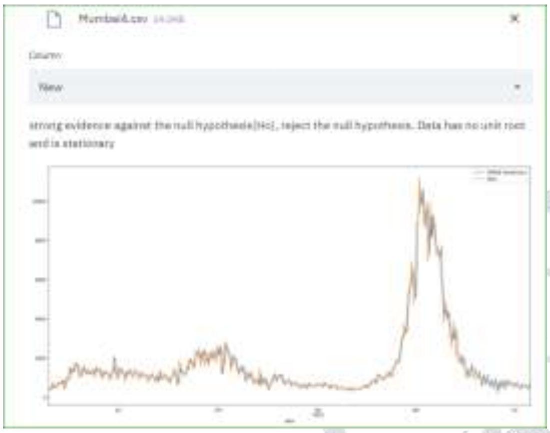

Fig 4.4 ARIMA (Training Data)

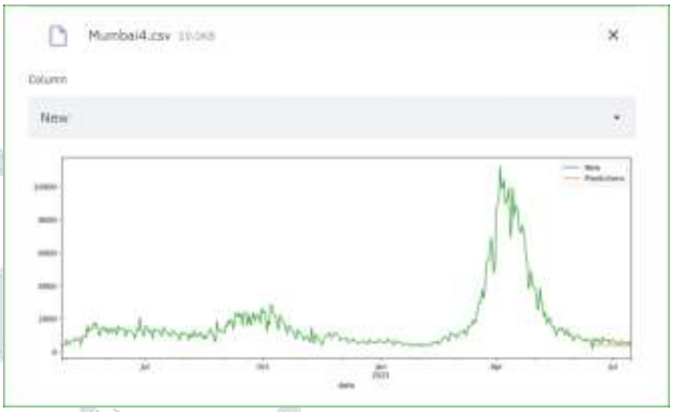

Fig 4.5 LSTM (Training Data)

Fig 4.6 Facebook Prophet (Training Data)

4.3 Graph Evaluation

In order to avoid overfitting to the training set, it's critical to evaluate our model with new data. However, it's occasionally useful to assess our model while we're developing it in order to identify the best parameters - but we can't do this with the test set since we'll wind up picking the parameters that perform best on the test data but not necessarily the values that generalise best. We generate a third subset of data known as the validation set to test the model while it is still being built and tuned. 60 percent of the data would be used for training, 20% for validation, and 20% for testing in a standard train/test/validation split.

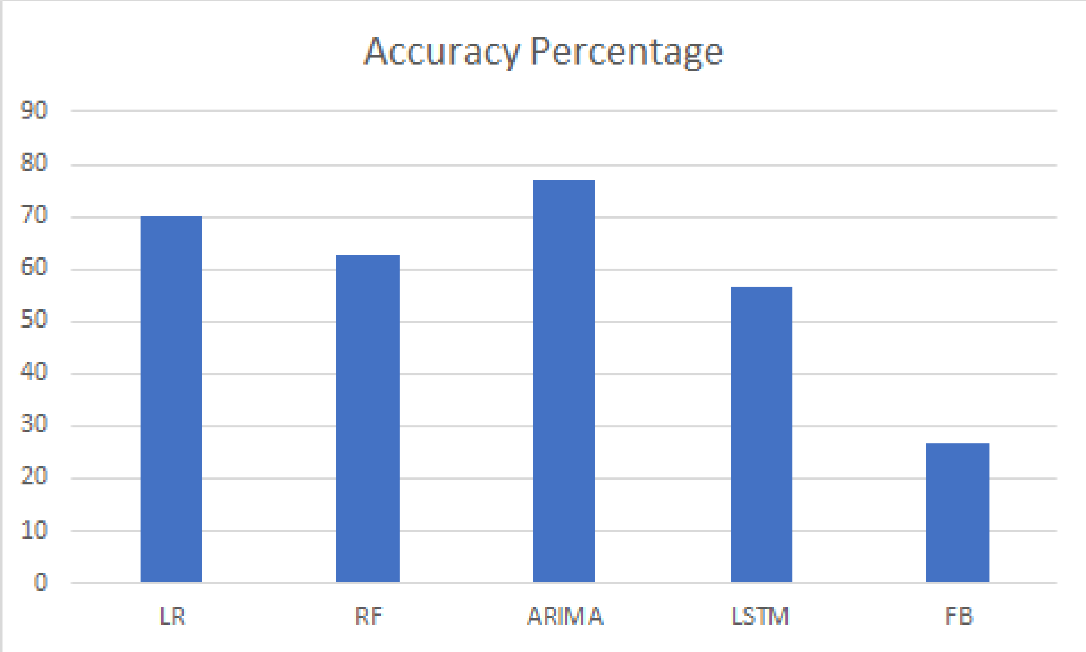

Fig 4.7: Accuracy Comparison

4.4 Final Prediction

As we saw in chapter 3, the final prediction is dependent on the analysis performed at the training data phase. Let's take a look at the predictions performed on the complete 100p percent data available.

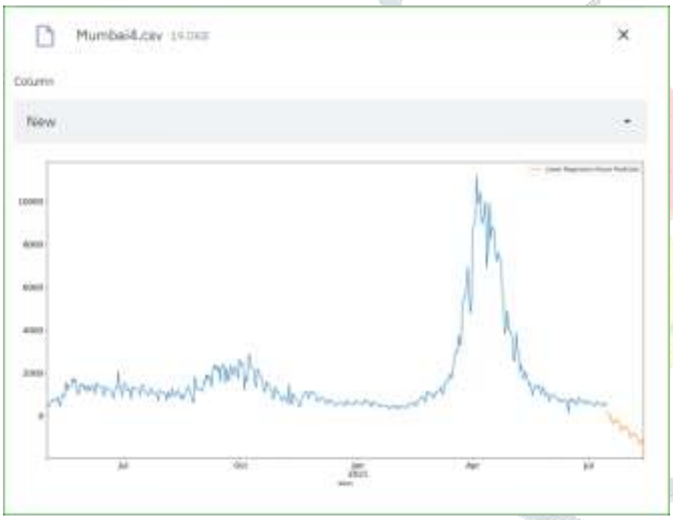

Fig 4.8 Linear Regression (Final Prediction)

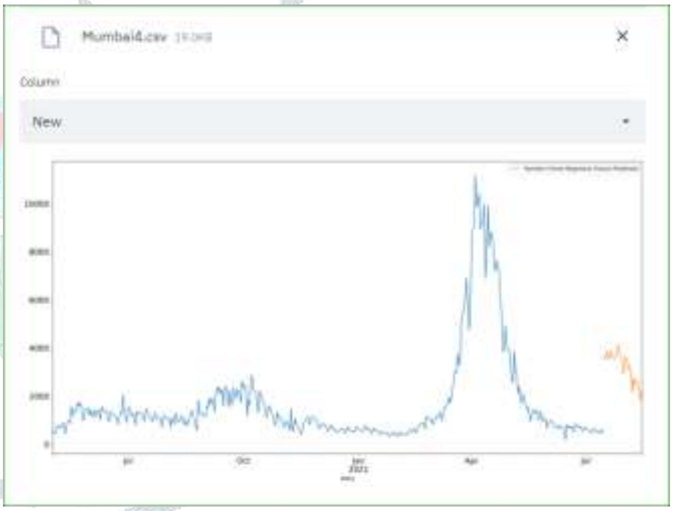

Fig 4.9 Random Forest (Final Prediction)

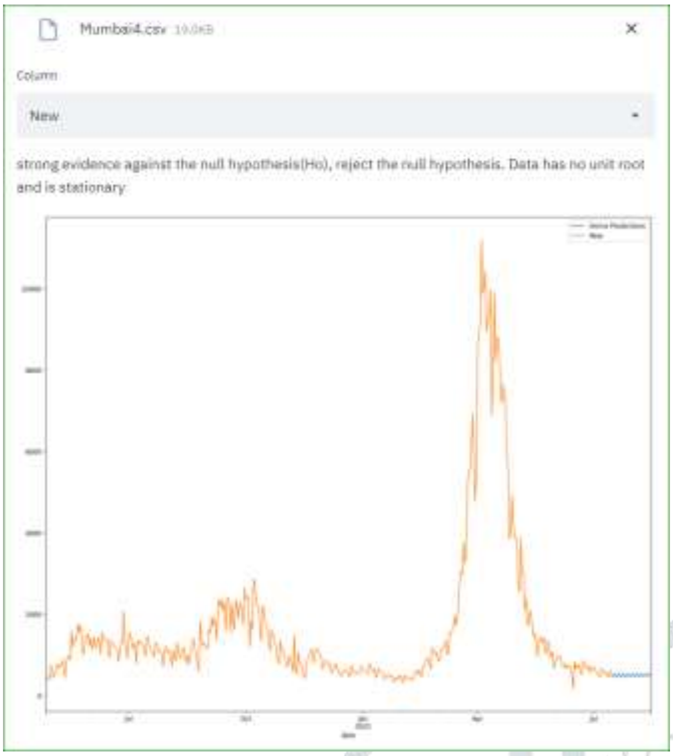

Fig 4.10 ARIMA (Final Prediction)

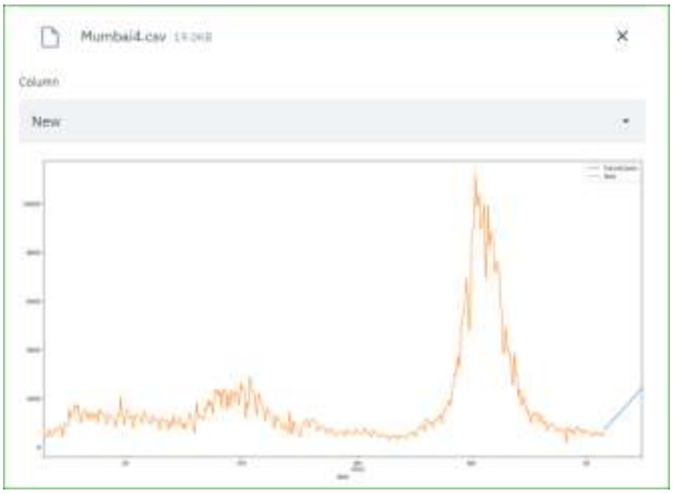

Fig 4.11 LSTM (Final Prediction)

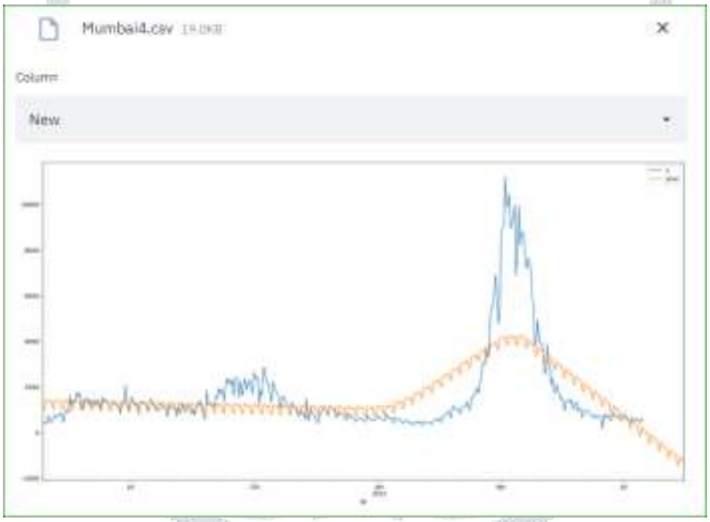

Fig 4.12 Facebook Prophet (Final Prediction)

V. APPLICATION

At the inception phase of the project, it was decided that this application will be used from the perspective of the municipality or local governing bodies. Let us se furthur how we can make use of the application and the use cases against the same.

5.1 Case Prediction

Looking at the past trend of COVID 19 cases, the effect of the new cases registered have greatly hampered the wellness and livability of the human kind. Having a prediction tool will provide readiness to face these challenges. The element of surprise has provided COVID 19 the upper hand. Leaving the governing bodies to take fast and sometimes rash decisions for the betterment of the people living in the area.

Previous loss, when we consider the two waves that hit Bombay Municipal Corporation and other Municipal Corporations, is that the readiness was not with respect to the COVID 19 pandemic. Now that such tools are available, it provides the country to be prepared for the coming waves. The same can be made available to the state level. We talk about the local governing bodies or the municipalities, not everyone that has the opportunity and access to these tools. Metropolitan cities may have access considering the education system and research teams available. Far remote bodies are not that facilitated. This tool can be made available to provide them reference for what is about to come. They can compare with the existing prediction system and see how this can benefit.

5.2 Medical Readiness

The primary objective of the prediction is to make sure the affected subjects are treated well with all the facilities available. Shortage of facilities will lead to unfavourable outcomes. As we saw during the second wave, shortage of remdeviver medicine and oxygen supply was creating many problems.

When the municipalities have access to these tools, the prediction can be analyzed and preparing the right medical facility and keeping ready in case of emergency can play an important role. Medical facilities pertaining to government hospitals, private hospitals or local clinics can be made aware of the coming situation and to gear up the medical equipment, medicines, support staff, doctors, etc.

5.3 Setting Rules

As the primary goal of predicting the COVID 19 trend is to make sure the medical facility is made available to all, the secondary goal is to make sure the spread of the diseases is controlled. This can be done by making sure necessary rules are placed on site. When we talk about municipalities, we understand that they are responsible for the governance and operation of the set area assigned. This means, the development of the municipality is the development of the state and country. In a similar way, preparing the municipality for the COVID 19 wave itself is preparing and providing ammunition to the country to fight the COVID 19 battle.

Setting up rules such as laying out social distancing rules, proper lockdown, non essential services to be closed or minimized, making the citizens aware of the upcoming situations, etc. All these steps will contribute and help strengthen the local body and take care of the area.

VI. Conclusion

6.1 Conclusion

Because the number of victims and deaths is continuously increasing, the COVID-19 continues to pose a potential threat throughout the world. The COVID-19 has clearly demonstrated its participation in the country's recent economic collapse on a massive scale. It has the capability of infecting anyone. There are serious worries that COVID-19's economic consequences will be similar to those of the Great Depression. Using machine learning, this study forecasts the total number of active cases in India over the following 20 days. The ARIMA and Random Forest model is best adapted to this circumstance, according to the findings of this study. The model's forecast of the current condition will be useful in predicting the future. Overall, this research can assist authorities in gaining caution, which might aid in the containment of the COVID-19 issue. This research will be improved over time as the course progresses.

6.2 Future Scope

A full scale system with live data feeds to the server can be developed in future. These feeds can be stored in the database for learning purposes. This will increase the usage of COVID 19 Prediction System.

References

[1] Quinlan R. Morgan Kaufmann Publishers; San Mateo, CA: 2014. C4.5: Programs for Machine Learning.

[2] Turaiki I., Alshahrani M., Almutairi T. Building predictive models for MERS-CoV infections using data mining techniques. J. Infect. Public Health. 2016;09:744–748.

[3] Chouhan V., Singh S.K., Khamparia A., Gupta D., Tiwari P., Moreira C., Damaševičius R., De Albuquerque V.H.C. A Novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl. Sci. 2020;10:559.

[4] Sreeja S., Bhavya L., Swamynath S., Dhanuja R. Chest x-ray pneumonia prediction using machine learning algorithms. Int. J. Res. Appl. Sci. Eng. Technol. 2019;07(04):3227–3230.

[5] Kose U., Guraksin G.E., Deperlioglu O. 2015. Diabetes Determination via Vortex Optimization Algorithm Based Support Vector Machines: Medical Technologies National Conference; pp. 1–4.

[6] Sharmila S., Dharuman C., Venkatesan P. Disease classification using machine learning algorithms – a comparative study. Int. J. Pure Appl. Math. 2017;114(06):1–10.

[7] Shirsath S.S. Disease prediction using machine learning over big data. Int. J. Innov. Res. Sci. 2018;07(06):6752–6757.

[8] Er O., Yumusak N., Temurtas F. Chest disease diagnosis using artificial neural networks. Expert Syst. Appl. 2010;37(12):7648–7655.

[9] Yang Z.F., Zeng Z.Q., Wang K. Modified SEIR and AI prediction of the epidemic trend of COVID-19 in China under public health interventions. J. Thorac. Dis. 2020;12(3):165–174.

[10] Müller M.A., Corman V.M., Jores J., Meyer B., Younan M., Liljander A., Bosch B.J., Lattwein E., Hilali M., Musa B.E., Bornstein S. MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983–1997. Emerg. Infect. Dis. 2014;20(12):2093.

[11] El-Solh A.A., Hsiao C.B., Goodnough S., Serghani J., Grant B.J. Predicting active pulmonary tuberculosis using an artificial neural network. Chest. 1999;116(04):968–973.

[12] Tiwari U.K. Role of machine learning to predict the outbreak of covid-19 in India. J. Xi'an Univ. Archit. Technol. 2020;12(4):2663–2669.

[13] Makridakis S., Wakefield A., Kirkham R. Predicting medical risks and appreciating uncertainty. Foresight Int. J. Appl. Forecast. 2019;52:28–35.

[14] Tuli S., Tuli S., Tuli R., Gill S.S. Internet of Things; 2020. Predicting the Growth and Trend of COVID-19 Pandemic Using Machine Learning and Cloud Computing.

[15] Punn N.S., Sonbhadra S.K., Agarwal S. MedRxiv; 2020. COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms.

[16] Jia L., Li K., Jiang Y., Guo X. 2020. Prediction and Analysis of Coronavirus Disease 2019. arXiv Preprint; p. 05447. arXiv:2003.

[17] Kalipe G., Gautham V., Behera R.K. IEEE; 2018. Predicting Malaria Outbreak Using Machine Learning and Deep Learning Approach: A Review and analysis; International Conference on Information Technology (ICIT) pp. 33–38.

[18] Sanjay Sharma discusses about How predictive models can aid in the battle against COVID-19. Available from: https://home.kpmg/in/en/home/insights/2020/04/how-predictive-models-can-aid-in-the-battle-against-covid-19.html. (Accessed 30 April 2020).

[19] Bohdan M.P. vol. 04. 2019. (Machine-Learning Models for Sales Time Series Forecast: MDPI). (15)

[20] Santosh K.C. AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. J. Med. Syst. 2020;44:93.

[21] Petropoulos F., Makridakis S. Forecasting the novel coronavirus COVID-19. PloS one. 2020;15(3) doi: 10.1371/journal.pone.0231236.

[22] Govindan K., Mina H., Alavi B. A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19) Transportation Research Part E: Logistics and Transportation Review. 2020

[23] Yang Z., Zeng Z., Wang K., Wong S.-.S., Liang W., Zanin M. Modified SEIR and AI prediction of the epidemic trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease. 2020;12(3):165–174.

[24] Nsoesie E.O., Marathe M., Brownstein J.S. Forecasting peaks of seasonal influenza epidemics. PLoS Currents. 2013;5(Outbreaks)

[25] Makridakis S., Winkler R.L. Averages of forecasts: Some empirical results. Management Science. 1983;29(9):987–996.

[26] Makridakis S., Spiliotis E., Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting. 2020;36(1):54–74.

[27] Nikolopoulos K. We need to talk about intermittent demand forecasting. European Journal of Operational Research. 2020 doi: 10.1016/j.ejor.2019.12.046.