WebApp for Covid-19 Prediction using Prophet

Venkata Sai Madhav Kaza
Computer Science Department
Central Michigan University
Mount Pleasant, Michigan
kaza2v@cmich.edu

Dr. Patrick Kinnicutt
Computer Science Department
Central Michigan University
Mount Pleasant, Michigan
kinni1p@cmich.edu

Dr. Lisa Gandy
Computer Science Department
Central Michigan University
Mount Pleasant, Michigan
gandy1l@cmich.edu

Abstract—**Coronavirus pandemic had been a huge drawback for 2020 year in United States. It is called as a pandemic because it affected almost every country on the earth. Everyone suffered a lot from the pandemic and vaccines are rolling out from March 2021. The main idea of the paper is not considering the vaccine option and if no one is able to get vaccinated then how the coronavirus cases are increased. Thus, the website application built by me undergoes the above said scrutiny while taking the dataset from covid-tracking project conducted by United States. I downloaded the dataset at start of the semester on 26th January, 2021. The dataset downloaded was for only one state i.e. Michigan State, where Central Michigan University is located. The research mainly focuses on building up a web-application for the overall covid-19 prediction system and uses the machine learning technique designed by facebook i.e. Prophet algorithm. Thus, results surprised everyone by using the prophet algorithm for building web-application using django tool of python. The accuracy of the prophet algorithm for the 80 percent test data out of 330 total datapoints is 95 percent. Thus, next phase of the research went upon using the django tool for the web-application. At last, total system is built for the covid-19 prediction system and submitted to the Department of Computer Science professors for review.**

Index Terms—**Coronavirus, Prophet, Machine Learning, Django, Pickle .**

I. INTRODUCTION

The information is cited in the papers [1], [5]–[10] about COVID-19, also known as coronavirus, is an ongoing pandemic of coronavirus disease started in 2019 caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV 2). The virus got spread first from Wuhan, China in December 2019. The WHO (World Health Organization) declared the disease as causing a Pandemic on 11 March 2020. As of 12 April 2021, more than 136 million cases have confirmed, with more than 2.94 million deaths attributed to COVID-19 in the world, making it one of the deadliest pandemics is history.

The COVID-19 pandemic has a huge effect on people in terms of economy and health aspects. The need for a forecasting model to predict the numbers of the COVID-19 pandemic has urged me to do a research project. The research has been done on the dataset of covid tracking project conducted US officials which is released in public. The dataset has been collected on January 26, 2021 at the start of the semester as part of ITC 691 (Information Technology Project) under professors at Central Michigan University.

The need of the COVID-19 prediction numbers made me see a lot of time-forecasting models. Out of all the forecasting models presented in various research papers, the FBProphet showed us the promising values to accept the approach for the forecasting prediction of the cases in this research. The research went on from collecting the dataset, pre-processing the data to building up the web application for the results display via interface designed by django tool.

Each of the sections in the paper is of the great interest of the reader. Abstract describes about the overall research from January 2021 - May 2021. Dataset describes the downloaded dataset from covid tracking project conducted by US officials. Methods indicates approaches taken in the research and how I started from collecting the data to finalizing the approach (FbProphet) and then to building an easy to use UI interface by django tool. Results depicts how better are the resutls by the research conducted and how it can be used by the hospitals in Michigan state to use the Predictor System at the offices for the benefit of the measures to take by them in future to mitigate the pandemic.

II. DATASET

The dataset was downloaded from URL mentioned in the cited Link [11] at the start of the SPRING 2021 semester. The dataset has various components describing the Coronavirus cases from Positive Cases to Total Tests Done in Michigan State. It mentions at the time of the data downloaded as 5 million active cases are present in the michigan state at the time of late january 2021. The dataset is then understood and the numbers are never changed from the public dataset. Thus, the overall dataset is used for the prediction of the covid-19 cases as part of the research. The below sections depicts the dataset features and what kind of data is present in the official dataset. And how the data is cleaned and processed before applying the prophet alogrithm for the research.

A. Columns Used

About the datset used, there are only two columns in the public dataset which is used. The Date column and PositiveIncrease column has been used in the research. These two columns are showed in the research to be the heart of the research from the dataset obtained. When we look at the dates, the dates are in the increasing order of the days from March 1, 2020 to January 26, 2021. And the PositiveIncrease column indicates the Positive Covid-19 cases in michigan

states counting all the counties in the state and counts positive cases day to day till the end.

B. Identifying Missing Values

There is a lot of missing values in the dataset and columns have been ignored for the accuracy of the research conducted. Atlast the one misisng value is on the day that started the covid-19 cases in the michigan state i.e. March 1, 2020. These missing values are considered as part of the research and showed in a good acccuracy at the final stage of research by using porphet algorithm.

C. Processing the Dataset

The dates provided are in the decreasing order of the dates from January 26, 2021 to March 1, 2020. So, processing of the dataset is a key aspect of research for further analysis using the prophet algorithm. Thus, the dataset is called in the jupyter notebook using python language and analysed further for accuracy results of the prophet algorithm.

III. METHODS

The methods section describes the two main approaches used in the research conducted as part the project under professors at Central Michigan University. The First approach is the machine learning approach to solve the problem of the COVID-19 Cases in the pandemic year from 2019 to present. And, the second approach is the Web Application Approach for designing the UI interface for possible Users who will use the appllication for easy access of the COVID-19 Future Cases.

A. Machine Learning Approach

The machine learning approach consists of the prophet algorithm chosen for two different variations on the same dataset. The two variations are with saturation and without saturation. According to the research conducted, based on accuracy tables mentioned, the prophet algorithm with saturation resulted in the promising values in terms of accuracy results.

1) Data Pre-Processing: Data needs to pre-processed before the algorithm can be applied to the dataset downloaded. This step consists of extraction of the required data from the dataset. Thus, preprocessing step has ordering the dates present in the Dates column of the dataset with link to indexes of the row. And, also the dataset shows the dates in decreasing order. Thus, it is in our hands to solve the above said problem by storing the reverse of the dates in an array with link to indexes of the row. Thus, retaining the value of PositiveIncrease column corresponding to the date given. Finally, the date and PositiveIncrease columns are stored in a pandas dataframe.

2) Prophet Algorithm without Saturation: Generally, the prophet alogrithm is used without saturation. And the prophet algorithm estimates the results of the dates predicted near to the timeline with perfect accruacy. But, as we go later in the future dates, the accuracy falls down below 0 percent for the later dates. The prophet algorithm api is referenced from the cited link [12] and the paper which is close in attaining the
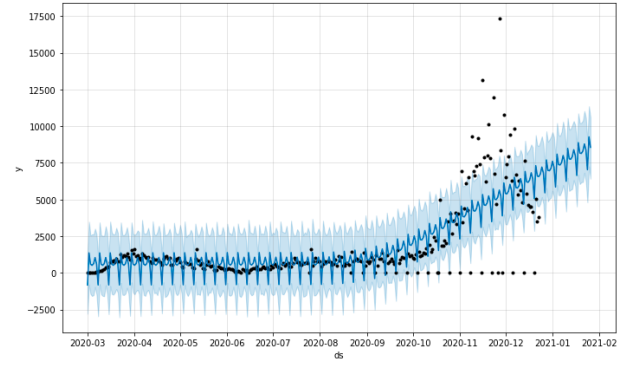

Fig. 1. The Figure shows the predicted values in blue and actual values in black for various dates in without saturation prophet algorithm approach.
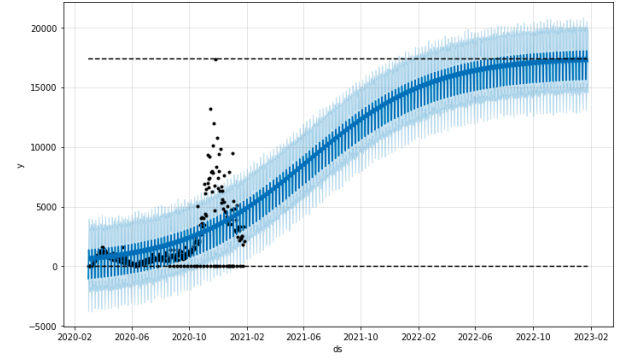

Fig. 2. The Figure shows the predicted values in blue and actual values in black for various dates in with saturation prophet algorithm approach.

least mean error rate is for without saturation approach as cited in [2]–[4], [13].

3) Prophet Algorithm with Saturation: The resarch conducted made me explore all the opportunities in the prophet algorithm applications as cited in [12]. There is no paper related to the use of saturation of the prophet forecasted model for the betterment of the results in the long-run. The saturation means capping the values both minimum and maximum values to the forecasted model for the betterment of the results. The maximum cap is found by the local maximum the dataset has seen in the previously deployed/used dataset. The minimum cap is intuitively known to the user by the no covid-19 cases can be in negative number. Thus, the maximum value obtained is around 17000 cases in a particular day in 332 data points and the minimum value is 0 as the covid-19 cases cannot go below negative number in a day.

4) Accuracy Table for Prophet without Saturation: The below table shows the accuracy of the prophet alogrithm in without saturation approach used as part of the research. The results shows a promising values that this approach fails in the usage of the prophet algorithm. The approach has a maximum accuracy for 35 percent split test data with accuracy as 93.52. And, the approach has a minimum accuracy as 0 percent for multiple values of the test split data percentage.

Serial Number	Test Split Data Percentage	Accuracy without Saturation
1	5	6.01
2	10	0.34
3	15	8.12
4	20	21.43
5	25	32.93
6	30	88.84
7	35	93.52
8	40	91
9	45	67.21
10	50	65.06
11	55	68
12	60	70.68
13	65	77.78
14	70	81
15	75	43.37
16	80	10.45
17	85	0
18	90	2.94
19	95	0

5) Accuracy Table for Prophet with Saturation: The below table shows the accuracy of the prophet alogrithm in with saturation approach used as part of the research. The results shows a promising values that this approach succeed in the usage of the prophet algorithm. The approach has a maximum accuracy for more than 84 percent split test data with accuracy as 100. And, the approach has a minimum accuracy as 39.87 percent for 5 test split data percentage.

Serial Number	Test Split Data Percentage	Accuracy with Saturation
1	5	39.87
2	10	70.9
3	15	69.25
4	20	67.29
5	25	65.86
6	30	63.52
7	35	60.65
8	40	57.5
9	45	53.55
10	50	48.79
11	55	43.33
12	60	36.09
13	65	27.35
14	70	28
15	75	74.69
16	80	98.5
17	85	100
18	90	100
19	95	100

B. Web Application Approach

1) Django Tool: As cited in the link [14], the django uses are many when building the framework for website on our local server. The django framework has many advantages when compared to other frameworks for building the website on local server. Django is an open-source framework for backend web applications based on Python. Its main goals are simplicity, flexibility, reliability, and scalability.Django has its own naming system for all functions and components (e.g., HTTP responses are called "views"). Django provides a dynamic CRUD (create, read, update and delete) interface, configured with admin models and generated via introspection. CRUD is used to describe the basic database commands, which means the interface facilitates viewing, changing, and searching for information.

2) Pickle and Simulation Files: The pickle files are the serialized files which store a snapshot of the model fit in the python notebook. Thus, we can store the pickle files as mentioned in the readme file attached with the predictor system. The pickle files has extension as ".sav" and can be used after fitting the model. And, simulation files are stored for the simulation results needed for the forecasting model to predict directly the future dates. The simulation files consists of the future dates in the ds (dates) column. And, y column which is the value for the PositiveIncrease on a particular date. Thus, these two files are must for the UI application built on django tool.

3) UI Interface Application: The User-Interface application built asks the user for his name and date which needs to be predicted. These two aspects are considered in building the UI application by django tool. By clicking on submit option, the date has two variation in the pages rendered by the django tool. The pages are just the HTML pages written with basic building block used in django. The 2-year window period from January 26, 2021 to October 23,2023 is the window taken into consideration for the date variations. If the date entered is between the 2-year window then, BeforeDate.html page is rendered by the django tool. And, if the date entered exceeds the 2-year window then, AfterDate.html page is rendered by the django tool. A brief description of the AfterDate.html and BeforeDate.html pages is shown in results section. And the UI build is also shown in the results section in detail-oriented way.

IV. RESULTS

The results surprised me and professors who reviewed the work I submitted in the april month. The accuracy results and web application part are so critical in getting the paper to be written and submit to the conference for a review. The accuracy results are shown below. And, the accuracy of the Prophet algorithm is so reliable that 95 percent obtained for the 80 percent split data. The accuracy results with days predicted on X-axis and accuracy of the model is determined in the accuracy section of the results. The web application results are shown below. And, the web application has two variations as the user can either enter a date between 2-year window
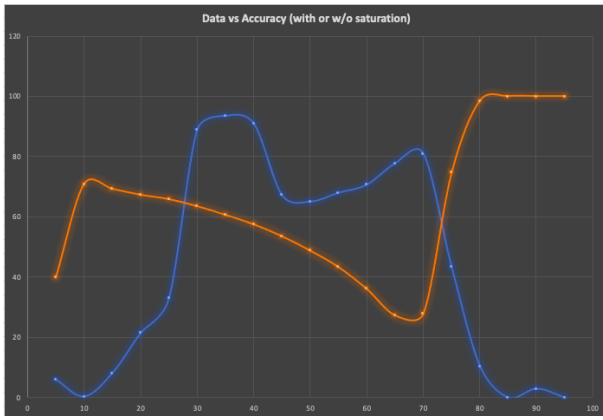
Fig. 3. The accuracy variation vs percentage of the train-test split in the 80-20 percentage is shown above.

Fig. 4. The accuracy variation vs days predicted is shown above.

Fig. 5. The Initial Page for the Day predicted in the 2-year window

or after the date of the 2-year window. The web application results are the key aspects of the django tool used in building up the predictor system which can be used by hospitals for the covid-19 cases predictions.

A. Accuracy as Y-Axis

1) Test Split Data Percentage as X-Axis: The test split data percentage increases from 5 percent to 95 percent with an increase of 5 percent each iteration. Thus, 5,10,15, ... 95 are the percentage values obtained for the test split data. The highest value of the accuracy for prophet with saturation results are the orange color line in the Fig.3 is about 100 percent. The highest value of the accuracy for prophet without saturation are the blue color line in the Fig.3 is about 94 percent. Thus, prophet algorithm with saturation is used as the accuracy results are a little higher when considering the model based on accuracy percentage.

2) Days Used for Prediction as X-Axis: The days used for prediction decreases from 316 to 17 with a decrease of 17 each iteration. Thus, 316, 299, 283, ... 17 are the values obtained for the days used for training the model. The highest value of the accuracy of the prophet algorithm with saturation results are only conisdered in the Fig.4 is about 100 percent. This highest value is same as the test split data percent for 95 percent used while checking the model for verification on the same datset.

B. Web Application

1) Date Entered within 2-Year Window: When User enters a date in between the 2-year window the following outputs through the UI of the application built. There are two html pages which are build as part of the application. First, initial page is rendered when we open url mentioned in the Readme file of the application. Next, when we click on get value after entering the date between March 1, 2020 and October 23, 2023, the final page is rendered and shown in the figures of the paper. So, as part of testing the present date is entered in the date field of the application which is April 12, 2021. And
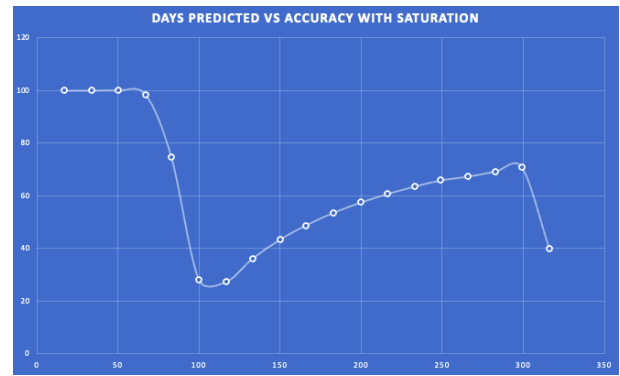
the predicted results are seen in the Fig.6. The daily predicted cases in michigan state are 7640 on April 12, 2021.

2) Date Entered After 2-Year Window: When User enters a date after the 2-year window the following outputs through the UI of the application built. There are two html pages which are build as part of the application. First, initial page is rendered when we open url mentioned in the Readme file of the application. Next, when we click on get value after entering the date after October 23, 2023, the final page is rendered and shown in the figures of the paper. So, as part of testing the present date is entered in the date field of the application which
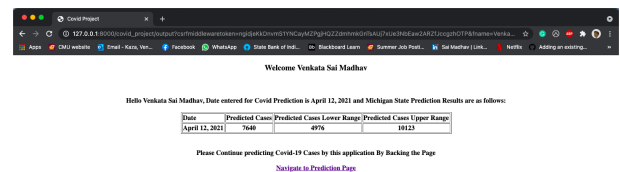

Fig. 6. The Final Page for the Day predicted in the 2-year window
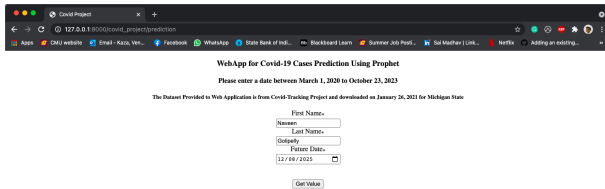
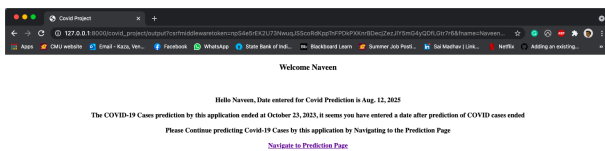Fig. 7. The Initial Page for the Day predicted after the 2-year window

Fig. 8. The Final Page for the Day predicted after the 2-year window

is August 12, 2025. And the predicted results are seen in the Fig.8. The daily predicted cases in michigan state cannot be determined by the application as UI detected entered date is not in the 2-year window.

V. Conclusion

The COVID-19 prediction has been done at the Central Michigan University in the time period from January 2021 - May 2021 as part of the course ITC 691 (Information Technology Project) under the professors Dr. Patrick Kinnicutt and Dr. Lisa Gandy. The results of the research done at CMU showed us a promising values that prophet algorithm can be used for time-related forecast for any sequence of numbers. The prophet algorithm showed us an accuracy of 95 percent plus for a training data of 80 percent of total datapoints (332 to be precise). Thus, prophet algorithm is used for future dates prediction in the web-application built. The website built by django tool is based on the results the prophet showed for future dates. The website takes parameters such as the name of the user and date he wants to predict for the covid-19 cases. There are two possible outputs for the prediction done. One is day predicted is in between the prediction period of the overall system and other is after the prediction period of the overall system i.e. any future date ofter 2 year window for the covid-19 prediction.

The future work for the research might to go on further for web-scraping of dataset from covid-tracking project officially launched by United States Government. We can frequently update the dataset and track the 2-year window future dates as we go on further with the days passed. We can eventually build a real-time analytics of the covid-19 prediction system by using prophet algorithm. Thus, the system build on real-time analytics platform can be submitted to hospitals in the michigan state for taking the measures in the hospitals who come for covid-19 case in near future.

VI. Acknowledgment

References

[1] Mahalle, P. N., Sable, N. P., Mahalle, N. P., and Shinde, G. R. (2020). Data analytics: Covid-19 prediction using multimodal data. In Intelligent Systems and Methods to Combat Covid-19 (pp. 1-10). Springer, Singapore.

[2] Indhuja, M., and Sindhuja, P. P. (2020). Prediction of covid-19 cases in India using prophet. International Journal of Statistics and Applied Mathematics, 5(4), 103-106.

[3] Mahmud, S. (2020). Bangladesh COVID-19 Daily Cases Time Series Analysis using Facebook Prophet Model. Available at SSRN 3660368.

[4] Gupta, A. K., Singh, V., Mathur, P., and Travieso-Gonzalez, C. M. (2020). Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. Journal of Interdisciplinary Mathematics, 1-20.

[5] Yadav, D., Maheshwari, H., and Chandra, U. (2020). Outbreak prediction of covid-19 in most susceptible countries. Global Journal of Environmental Science and Management, 6(Special Issue (Covid-19)), 11-20.

[6] Ndiaye, B. M., Tendeng, L., and Seck, D. (2020). Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting. arXiv preprint arXiv:2004.01574.

[7] Mahalle, P. N., Sable, N. P., Mahalle, N. P., and Shinde, G. R. (2020). Predictive analytics of covid-19 using information, communication and technologies.

[8] Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., ... and Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. Algorithms, 13(10), 249.

[9] Wang, P., Zheng, X., Li, J., and Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons and Fractals, 139, 110058.

[10] Santosh, K. C. (2020). COVID-19 prediction models and unexploited data. Journal of medical systems, 44(9), 1-4.

[11] Dataset URL: https://covidtracking.com/data/download

[12] Prophet URL: https://facebook.github.io/prophet

[13] N. Kumar and S. Susan, "COVID-19 Pandemic Prediction using Time Series Forecasting Models," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225319.

[14] Django URL: https://djangostars.com/blog/why-we-use-django-framework/