Final report_v2.docx

by Qiyuan Shen

Submission date: 17-Dec-2021 06:44PM (UTC+0800)
Submission ID: 1732709638
File name: Final_report_v2.docx (840.73K)
Word count: 4004
Character count: 21523

STATS 402 - Interdisciplinary Data Analysis
<Multi-variate Model for COVID-19 Prediction>

< zihao.wei@dukekunshan.edu.cn >
< qiyuan.shen@dukekunshan.edu.cn >

Abstract
Our work is devoted to predicting the volume of existing COVID-19 cases in the future to illustrate the situation of the pandemic to facilitate policymaking. It is noticed that policy interventions have played an important role in controlling the transmission of the flu, and hospitalization conditions also have an effect on the recovery rate of those infected people. Our proposed multi-variate method takes those factors into consideration based on an eXtreme Gradient Boosting (XGBoost) model with K-fold training. Unlike traditional compartmental models in epidemiology, the sequence of time is ignored during training with shattered data. The proposed model is tested on a Virginia state dataset with a comparison of SIR model. It turns out that the XGBoost model has an outstanding performance in prediction accuracy and much better generalization than the SIR, especially for a long-term prediction.

1. Introduction

The outbreak of the COVID-19 pandemic has affected our lifestyles a lot. Recently, we have encountered a new variant—omicron, which declares a new wave of such flu. It seems that there is no end to the pandemic, and we will live with the coronavirus for a long time. Therefore, we want to predict the volume of existing COVID-19 cases in the future for the sake of finding a turning point, where the trend in number begins to go down thus signing humans' success in the fight against COVID. By illustration the situation of the pandemic, we also want to facilitate policymaking. Take a view of the whole world, China has done a relatively good job in controlling the transmission of the COVID-19. The major difference between China and others is that China has adopted stricter public health policies than others. Policy interventions may be an important factor in influencing the situations of COVID-19. As a result, features concerning policy interventions will be introduced in the proposed eXtreme Gradient Boosting (XGBoost) model for prediction. At the same time, government investment in the medical system is also thought as another important aspect because the conditions of hospitalization are likely to limit the recovery rate for patients. Motivated by that, hospitalization condition is also taken into account when constructing the model. It is also noticed that with the appearance of new variants of coronavirus, the development of pandemic varies like a superposition of waves, thus does not have obvious relationships with time in the long term. Therefore, our proposed model will only focus on the shattered data and leave the influence of date-time aside. For the following part of this article, we will first go through works done by others in predicting the volume of COVID-19 cases. Secondly, our proposed XGBoost model, along with a comparison with the reimplementation of SIR (Susceptible-Infected-Resistant) model will be introduced. Those two models will be tested on the dataset from Virginia. Then, we will show the processing of dataset and the performance evaluation of the two models. Finally, a conclusion will be drawn showing that our proposed model has done satisfying work.

2. Related work/literature review

With the outbreak of the COVID-19 pandemic, the research in the transmission of COVID-19 comes to be a hot topic. There are a lot of previous works in the field of predicting the volume of COVID-19 cases which gives some insights and inspiration to be applied in our research. Most of them are SIR-related.

Talukder applied the traditional SIR model to predict the pandemic development in Bangladesh [1]. He used the Bangladesh data and sets the infectious rate, recovery rate and reproductive ratio at the maximum likelihood estimation. The traditional SIR model is a simple and fast model to adopt. It is easy to get the required data to train the model (suspectable, infective, recovered and death). However, the fixed infectious rate and recovery rate may neglect the changes in the development of pandemic, and it is hard for SIR to take some outside features like government intervention, medical system pressure into consideration. Another research uses a modified SIR model: SEIR model. The author applied the pandemic data of Wuhan as training data and then use the SEIR model to test the effectiveness of the quarantine policy [2]. Compared to the traditional SIR model, the SEIR model adds a new group of people: exposed which refers to the group of susceptible people who have contact with the infected group without any symptom. The SEIR model is more similar to the real situation of COVID-19 than the traditional SIR model which can provide a more precise prediction than the normal SIR model. However, the SEIR model also has its disadvantages. It is hard to detect the exposed cases and determine the transmission ratio since there is no difference between the susceptible and exposed. The third research applied a time-independent SIR model to get rid of the influence of time [3]. Though the robustness is increased, the prediction accuracy of the model is not outstanding. But the inspiration is that the influence of time should be removed for generalization. It is also noted that all the SIR models are mathematical deterministic which do not have the ability to tolerate random events that may affect the pandemic situation a lot. At the same time, it is hard to introduce other COVID-19

related features to help the prediction of SIR. As a result, XGBoost is proposed as a stochastic model to catch more randomness in the development of pandemics and allow more features to be counted in since we think that policy intervention and hospitalization conditions all play a role. To the best of our knowledge, we haven't found any XGBoost related models constructed to directly predict the volume of COVID-19 cases partly due to the lack of data. But XGBoost has been used to classify priorities for vaccination campaigns and achieved an excellent result under the COVID-19 context [4]. Given its performance on classification tasks, for the same topic, we think XGBoost can also do well on the regression task.

3. The proposed method

Our research goal is to predict the volume of existing cases of COVID-19 to illustrate the overall situation of the pandemic. As stated before, the final goal will be decomposed into three parts of forecasting the accumulated infections, total recovered cases and total death. Finally, the number of existing cases will be calculated by the sum of total recovered and the death subtracted by accumulated cases. Such a procedure of estimating the volume of existing cases is made under the assumption that unreported cases will not have a big influence on the general trend of the pandemic as well as the accuracy of our numbers concerning cases. Based on that, the XGBoost model is proposed to take policy interventions and hospitalization conditions into consideration during the prediction of the number of accumulated infections, total death and total recovered cases separately. As a comparison, a reimplementation of adapted SIR model is also adopted for prediction. A brief dataset analysis and the introduction of two models are as follow.

3.1. Dataset

In this project, the data from Virginia state is mainly used to build and test the proposed method.

The first dataset is mainly about the overall cases in Virginia which are collected by the Centers for Disease Control and Prevention [5]. The dataset covers the date from 2020/01/22 to 2021/11/14. The total/new cases and total/new death are the bases for our model. There are some missing values at the beginning of this dataset. Since it is the beginning of the pandemic in Virginia, zeros are used to replace the missing data. Here is the visualization of some important features.

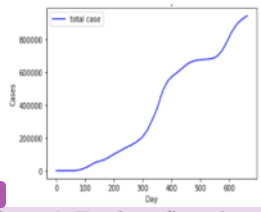

Figure 1. Total confirmed cases

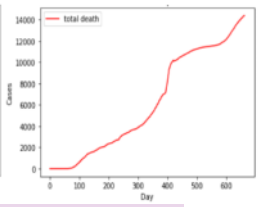

Figure 2. Total death

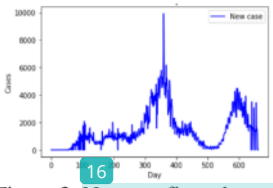

Figure 3. New confirmed cases

The total confirmed cases and total death have the same trend. But it seems that those three features do not have a strong relationship with date time and the trend of new cases fluctuates a lot especially.

The second used dataset is the policy intervention dataset: COVID policy tracker which is collected from the COVIDVIS team of UC Berkely [6]. It covers the data from 2020/01/01 to 2021/11/14. In this dataset, there are two kinds of features counted in the model. The first one is the index for certain policy which is one-hot encoded and uses 0-3 to represent the implementation of the policy. The higher the number is, the stricter the policy implements. The second kind is the overall policy index. It is the sum of a series of policies' indexes. Considering that there are some policies that haven't been changed since implemented in Virginia state, features with the same value and some data having less relationship with our model are removed. The correlation heatmap for some picked features are shown below:

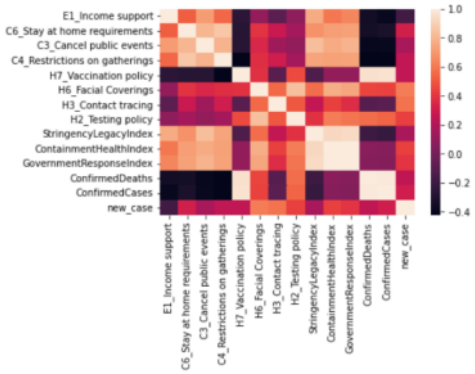

Figure 4. Correlation metrics for policy interventions

From the heatmap above we know that the total confirmed deaths and total confirmed cases relate to H7_Vaccination policy, H6_Facial Coverings, H2_Testing policy best. While H6_Facial Coverings, H2_Testing policy, Containment Health Index and Government Response Index have a straight influence on the increase in new cases.

The third dataset mainly focuses on hospitalization conditions which is collected by the Virginia Open Data Portal [7]. This dataset covers the date from 2020/04/14 to 2021/11/14. The features of existing hospitalized cases and ICU cases reflect the medical system pressure after the outbreak of the COVID-19. A simple visualization is shown below:

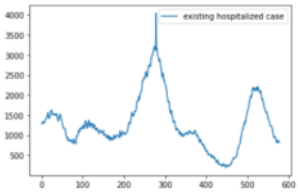

Figure 5. Existing hospitalized cases (left)

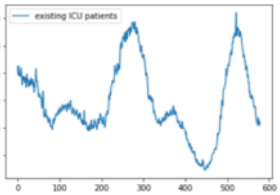

Figure 6. Existing ICU patients (right)

It can be seen that the trend of hospitalized and ICU cases are similar to that of newly confirmed cases. But till now, there is no information about recovered cases found in those datasets which are important for our model building. The final dataset is used to cover this.

The dataset is the recover data collected by Johns Hopkins University Center for Systems Science and Engineering [8]. It covers the data from 2020/04/14 to 2021/02/22. This dataset fulfills our lack of recover data, and the number of total recovered cases is mainly used. The data of new recovered cases is obtained by calculation: *New_recovered_cases(t)= Total_recovered_cases(t)- Total_recovered_cases(t-1)* where *t* refers to date time. Such a calculation causes some negative values in the new recovered cases which is unrealistic. Zero is used to replace those wrong values. After the preprocessing, the visualization of total recovered cases and new recovered cases is shown below:

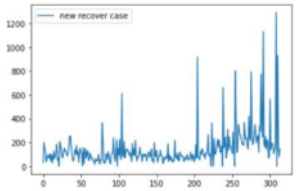

Figure 7. New recovered case

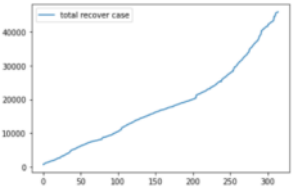

Figure 8. Total recovered cases

3.2. SIR

SIR model is a classic transmission model which is widely used in predicting the trend of the pandemic. We reimplement an adapted version in our research to make a comparison with the XGBoost. The main idea of the SIR model is to separate the population into 3 types of people: susceptible, infective and the removed group. Since the recovered cases and death cases don't have infectiousness which will have no influence on the prediction of the model. So, they are classified into the group of removed. In the model, we use S(t) to represent the susceptible cases at time t, I(t) refer to the confirmed cases at time t and the cases recover from the COVID at time t is R(t). Besides the main idea of the SIR, there are 3 main principles that are the bases for the model.

1. The first idea is that the amount of the population in the model is fixed. (1)
2. The susceptible cases will convert into infected cases when they have contact with each other, the infective rate is β . (2)
3. As time passes by the infected cases will be recovered by the recover rate γ . (3)

Here are the formulas of the principles correspondingly:

$$N(t) = S(t) + I(t) + R(t) \quad (1)$$

$$S(i) + I(j) \xrightarrow{\beta} I(i) + I(j) \quad (2)$$

$$I = (S_0 + I_0) - S + \frac{\gamma}{\beta} ln \frac{S}{S_0} \quad (3)$$

In the traditional model, the infective rate and the recovery rate are fixed which may cause some errors. In our model, adaptation is made by applying a daily updated infective rate and recover rate. By differential equation, we can get the formula below. Because the time is discrete, the differential equation can change into the difference between two near the value.

$$\frac{dS(t)}{dt} = \frac{-\beta(t)S(t)I(t)}{n} = S(t+1) - S(t) \quad (4)$$

$$\frac{dX(t)}{dt} = \frac{\beta(t)S(t)I(t)}{n} - \gamma(t)I(t) = I(t+1) - I(t) \quad (5)$$

$$\frac{dR(t)}{dt} = \gamma(t)I(t) = R(t+1) - R(t) \quad (6)$$

Then we can get the value of the β and γ

$$\gamma(t) = \frac{R(t+1) - R(t)}{I(t)} \quad (7)$$

$$\beta(t) = \frac{[R(t+1) - R(t)] + [I(t+1) - I(t)]}{I(t)} \quad (8)$$

After that we can make prediction for the existing cases and recovered cases by the infective rate and recover rate of previous data. The figure below is the procedure of adopting the SIR model.

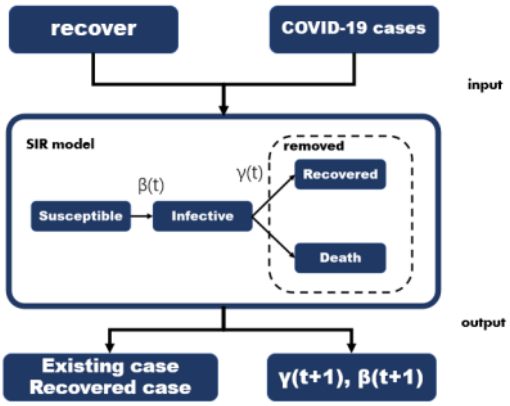

Figure 9: Pipeline for SIR model

3.3. XGBoost

XGBoost is a scalable model based on decision trees that provides regularization in a gradient boosting system [9]. This method makes use of individual decision trees and sequentially adds predictors to learn the previous residuals from models for correction under the gradient descent framework. XGBoost is applied as a prediction model for three targets: accumulated cases, accumulated death and accumulated recover because of its characteristics of high generalization performance, allowance for high dimensional data and high interpretability that outstand in dealing with tabular data. Compared to the SIR model, XGBoost

is a purely data-driven approach that requires no prior knowledge concerning epidemiology to set up a transmission rate. Instead, XGBoost only concentrates on the given training dataset. If the data is given with a shuffled sequence of time, the model is sensitive enough to detect the sparsity and answer the pattern discrimination question without giving a thought for time, which is the same as desired in the initiative. Besides, since XGBoost provides an easy way to introduce multi-dimensional features, the policy interventions and hospitalization situations can be fully taken into consideration. In this sense, XGBoost model is more likely to catch up with more patterns about policy and hospitalization.

For the tree ensemble model using N data points and T additive functions, namely regression trees, the objective function is as follow:

$$Obj = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{i=1}^{T} \Omega(f_i) \quad 9$$

Where l refers to the loss function between observed value y_i and the predicted value $\hat{y}_i$, Ω represents the regularization term for each small regression tree f_i.

XGB regressor is implemented with data from a combination of COVID-19 cases data, policy interventions, recover and hospitalization condition data to predict the volume of accumulated cases, accumulated recover, and accumulated death in the future 60 days separately with three models. The dataset will be shuffled first to disorganize the sequence of time. And after the prediction for three aspects, the number of existing cases will finally be estimated. For the model, the depth of a tree is limited to be less than 8 to avoid the potential overfitting problem and to reduce the model complexity considering that there are over 40 features introduced from the training. K-fold cross-validation is also adopted in the training process considering that one XGB model is sensitive to outliers. The training data is split into 5 different folds for validation to train 5 different sets of parameters. Those parameters are used for the 5 folds to return their predictions for the test data. The final forecast will be integrated by the answers from the 5 folds by taking the average. In this case, the bias for the final prediction would be reduced, thus a better generalization can be easily achieved. Finally, the prediction for accumulated cases, accumulated recover and accumulated death will be calculated as the equation below to estimate the volume of existing cases:

$$Existingcases = Ca - Re - De \quad 10$$

Where Ca, Re, De refers to accumulated cases, accumulated recover and accumulated death correspondingly.

And it should be noted that as the targets to be predicted, the number of COVID-19 total cases, total death and total recover is intentionally lagged behind for 60 days. By this lagging, the XGBoost model would be allowed to predict the situations 60 days later by simply looking at the condition 60 days before.

The above-proposed procedure for implementing XGBoost is illustrated in Figure 10.

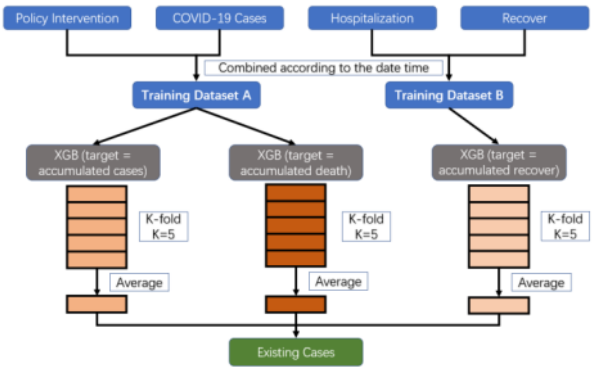

Figure 10. Procedure to implement XGBoost

Additionally, the volume of new cases is also calculated to help better illustrate the situation of the pandemic by:

$$Newcases(t) = Ca(t) - Ca(t-1) \quad 11$$

Where t represents the date time, and $t-1$ refers to the previous day.

Performance evaluation

It is worth noting here that, our research goal is to estimate the volume of existing COVID-19 cases in the future and the test dataset for final results only covers the date range from 2020/04/14 to 2021/02/22 which is limited by the available recover data. But in the separated parts for training, each model made use of the most range of data we had for the sake of better generalization. For the convenience of visualization, the date time is changed to the time index in figures starting from the first date being used.

4.1. Measures

The predictive performances for the SIR and XGBoost approaches were evaluated according to the following measures for regression: (a) Mean absolute percentage error (MAPE): an illustration in terms of relative error. (b) Coefficient of determination (R^2 score): an interpretation of fit which shows how much variation of targeted values can be explained. (c) Mean squared error (MSE): the average squared differences between the estimation and the real value. (d). Normalized mean squared error (NMSE): a normalized version of MSE by the division of real mean. (e) Normalized root mean square error (NRMSE): a measure of how concentrated the data is around the fitting.

4.2. Performance

As stated before, sliding window is adopted in the procedures of implementing SIR model. The training window size is set as 60 days and the predicting range is limited to 4 days. The performances of the SIR model on the confirmed existing cases and total recovered cases are shown below:

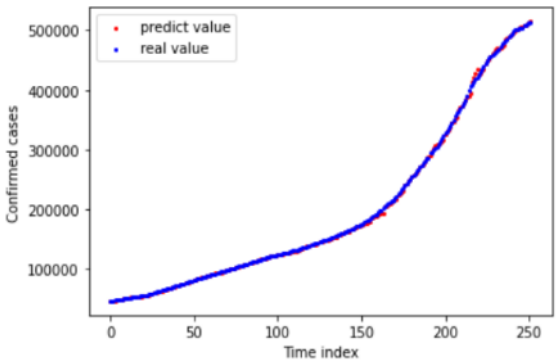

Figure 11. The performance of the SIR model in existing cases (window range: 60, predicting range 4)

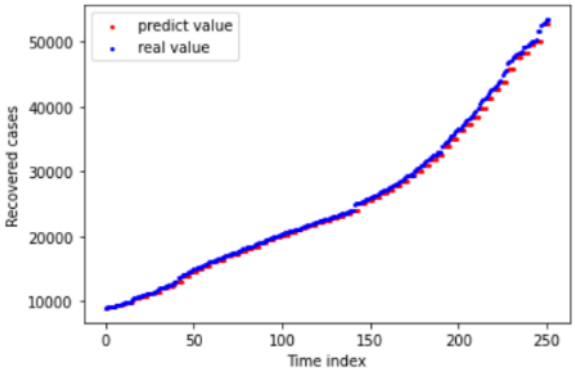

Figure 12. The performance of the SIR model in recovered cases (window range: 60, predicting range 4)

	Existing cases	Recovered cases
MAPE	0.006532	0.015698
R^2	0.999708	0.997740
MSE	4879773.0782786505	316234.1
NMSE	28.40200	12.75730
NRMSE	0.011937	0.022332

Table 1. The performance of the SIR model

In general, the predictions of the SIR model in existing cases and recovered cases are quite good. The model is able to catch the triggers of changes. But there still exist some divergence between the estimation and the actual value shown in the figure especially when it comes to recover data. Besides, it is found that the performance of the SIR model is not satisfying in a long-term prediction. When the window size is increased to 10, the predicting values won't fit the real values closely as is demonstrated in figure 13 and 14.

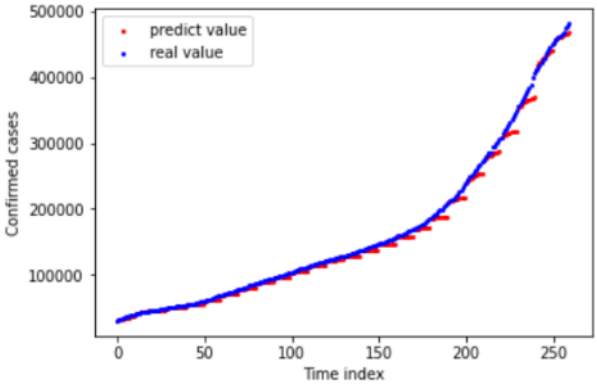

Figure 13. The performance of the SIR model in existing case (window range: 60, predicting range 10)

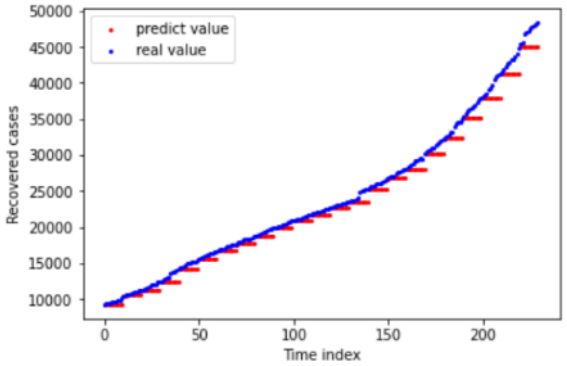

Figure 14. The performance of the SIR model in recovered cases (window range: 60, predicting range 10)

For XGBoost method, the evaluation for the separated models which predicts the volume of accumulated cases, total death and total recover correspondingly on dataset A and B is shown below:

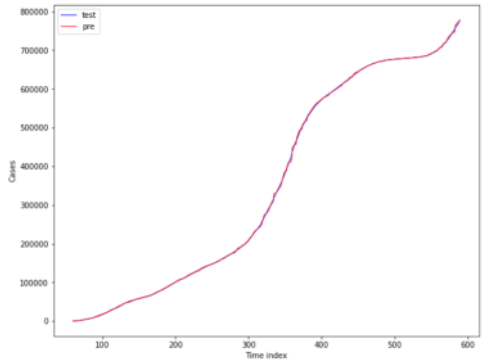

Figure 15. XGB prediction for total cases

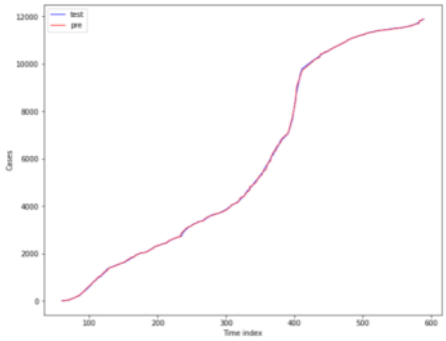

Figure 16. XGB prediction for total death

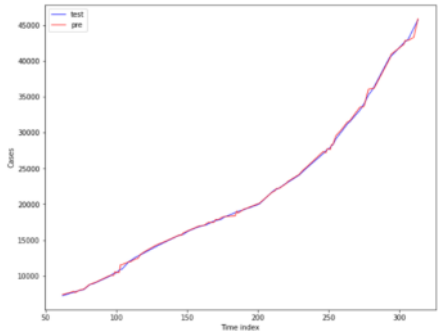

Figure 17. XGB prediction for total recover

The red line stands for the prediction and the blue line plots the real values. Overall, the XGB models have done a great job with loss measurement table below:

	Accumulated cases	Total death	Total recover
MAPE	0.017885	0.017445	0.010189
R^2	0.999874	0.999900	0.999280
MSE	8866957.9	1638.042	74202.99
NMSE	23.96579	0.269223	3.474250
NRMSE	0.008048	0.006651	0.012754

Table 2. The performance of individual XGB models

After the final calculation for existing cases as proposed before, the predictive results are shown below with a comparison to SIR.

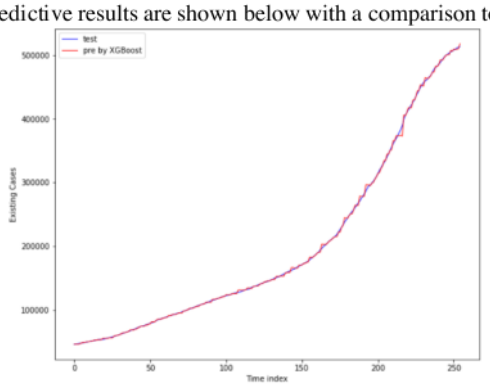

Figure 18. Existing cases predicted by XGBoost approach

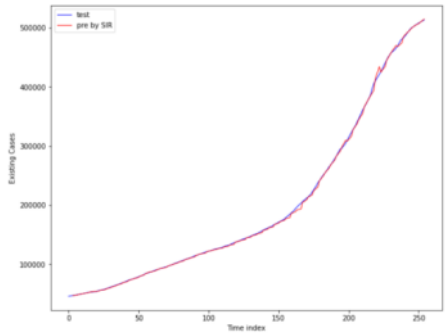

Figure 19. Existing cases predicted by SIR approach

	XGBoost	SIR
MAPE	0.006662	0.007136
R^2	0.999727	0.999687
MSE	5301608.4	6073990.5
NMSE	26.81811	30.44738
NRMSE	0.011647	0.012354

Table 3. The performance of XGB and SIR on exiting cases

It is hard to tell which method is better by simply looking at the two figures since both of the predicted lines are greatly close to the real one. There is almost no difference in the R^2 score. However, the MSE has something to say by illustrating a big gap between the two values. In this aspect, XGBoost outperforms. Though the MAPE and R^2 score indicate that our two models can approximately generalize the real model, considering the large number given for the targets, even a small relative error will result in a big disparity reflected on MSE.

Additionally, new cases predicted by XGBoost are shown below as a help to understand the trend of the pandemic.

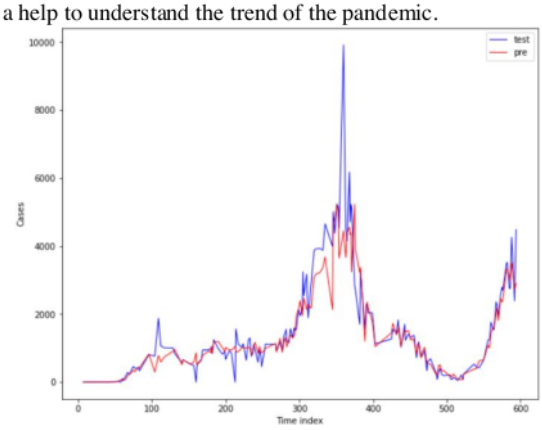

Figure 20. XGB prediction for daily new cases

	XGBoost
MAPE	5.004894
R^2	0.830882
MSE	382852.3
NMSE	252.1374
NRMSE	0.407494

Table 4. The performance of XGB on daily new cases

Though the evaluation metrics indicate that the divergence is relatively larger compared to the results for accumulated numbers. It can be seen from figure 20 that the model has the ability to learn and forecast fluctuations.

5. Conclusion and future work

In this paper, we have introduced four datasets and done the comparison between the SIR and XGBoost model on the prediction of COVID-19 cases volume. Both of the two approaches have achieved a satisfying result while XGBoost outperforms in a long-term prediction with a high generalization. We can easily predict the volume of existing cases 60 days later in the future by XGboost. Besides, XGBoost allows stochasticity and provides an easy way to introduce more related features. Our work also verified that the policy intervention and hospitalization conditions have a strong relationship with the development of the pandemic. Our work has the potential to play a fundamental role in facilitating policymaking by illustrating a general situation of COVID-19 which casts a light on the volume of existing cases. The proposed model provides indications on the progress of pandemic that although not definitive, would generate an insight simply given by the data of which levels of interventions are taken. One limitation of the proposed approach lies in the lack of a dataset. Policy interventions are hard to be quantized and there are only a little public data concerning policies being taken. We had tested our method on the Virginia state data and more tests need to be done on other states for validation. Another disadvantage of the XGBoost model is that it is too sensitive to outliers which may lead to a decrease in model generalization. But in our proposed approach, K-fold training has resolved such an issue to some extent. And since there exists only a little space for improvement on prediction accuracy, we expect some future works to focus on simplifying our proposed model. Rigorous features selections and feature combinations may be further adopted. Another direction of future work may be concentrated on adapting the SIR model to improve its model generalization.

References

[1] A. Talukder, (2020). Susceptible ‐ infectious ‐ recovered (SIR) model ‐ based forecasting of Covid ‐ 19 outbreak in Bangladesh. International Journal of Clinical Practice, 74(11). https://doi.org/10.1111/ijcp.13648

[2] C. Hou, et. al., (2020). The effectiveness of quarantine of Wuhan City against the Corona virus disease 2019 (Covid ‐ 19): A well ‐ mixed seir model analysis. Journal of Medical Virology, 92(7), 841–848. https://doi.org/10.1002/jmv.25827

[3] Y. -C. Chen, P. -E. Lu, C. -S. Chang and T. -H. Liu, "A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 4, pp. 3279-3294, 1 Oct.-Dec. 2020, doi: 10.1109/TNSE.2020.3024723.

[4] L. Romeo, E. Frontoni, A Unified Hierarchical XGBoost model for classifying priorities for COVID-19 vaccination campaign, Pattern Recognition, Volume 121, 2022, 108197, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2021.108197.

[5] CDC Case Task Force, "United States COVID-19 Cases and Deaths by State over Time", Centers for Disease control and prevention, November 14, https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36

[6] Toby Phillips, Saptarshi Majumdar, actions-user, tatlowhelen, aleserche and Sergio Leunissen, "covid-policy-tracker", Github, November 14, https://github.com/OxCGRT/covid-policy-tracker

[7] Virginia Department of Health, "VDH-COVID-19-PublicUseDataset-KeyMeasures-Hospitals," Virginia Open Data Portal, November 9, 2021, https://data.virginia.gov/Government/VDH-COVID-19-PublicUseDataset-KeyMeasures-Hospital/28wk-762y

[8] Johns Hopkins University Center for Systems Science and Engineering, "csse_covid_19_daily_reports_us", Github, December 1,2021, https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports_us

[9] T. Chen, C. Guestrin. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (2016), pp. 785-794.

Final report_v2.docx

ORIGINALITY REPORT

4%	3%	2%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Luca Romeo, Emanuele Frontoni. "A Unified Hierarchical XGBoost Model for Classifying Priorities for COVID-19 Vaccination Campaign", Pattern Recognition, 2021 Publication	1%
2	Anupam Prakash, Piyush Sharma, Indrajeet Kumar Sinha, Upendra Pratap Singh. "Spread & Peak Prediction of Covid-19 using ANN and Regression (Workshop Paper)", 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), 2020 Publication	<1%
3	doctorpenguin.com Internet Source	<1%
4	www.modernhealthcare.com Internet Source	<1%
5	docs.google.com Internet Source	<1%
6	Jagpreet Chhatwal, Yingying Xiao, Peter Mueller, Madeline Adee, Ozden O Dalgic,	<1%

Mary Ann Ladd, Turgay Ayer, Benjamin P Linas. "Changing Dynamics of COVID-19 in the U.S. with the Emergence of the Delta Variant: Projections of the COVID-19 Simulator", Cold Spring Harbor Laboratory, 2021
Publication

---

7 arxiv.org
Internet Source <1%

---

8 Choujun Zhan, Yufan Zheng, Haijun Zhang, Quansi Wen. "Random-Forest-Bagging Broad Learning System with Applications for COVID-19 Pandemic", IEEE Internet of Things Journal, 2021
Publication <1%

---

9 Www.Igi-Global.Com
Internet Source <1%

---

10 Yi-Cheng Chen, Ping-En Lu, Cheng-Shang Chang, Tzu-Hsuan Liu. "A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons", IEEE Transactions on Network Science and Engineering, 2020
Publication <1%

---

11 dora.dmu.ac.uk
Internet Source <1%

---

12 napier-surface.worktribe.com
Internet Source <1%

---

www.ijariit.com

13	Internet Source	<1%
14	www.ncbi.nlm.nih.gov Internet Source	<1%
15	www.researchgate.net Internet Source	<1%
16	Xiao-Hua Liang, Xian Tang, Ye-Tao Luo, Min Zhang, Ze-Pei Feng. "Effects of policies and containment measures on control of COVID-19 epidemic in Chongqing", World Journal of Clinical Cases, 2020 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		