American Journal of Theoretical and Applied Statistics
2021; 10(5): 208-215
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20211005.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

COVID-19 Prediction and Detection Using Machine Learning Algorithms: Catboost and Linear Regression

Justine Shinjae Kim

Emma Willard School, Troy, USA

Email address:
justinekim2424@gmail.com

To cite this article:
Justine Shinjae Kim. COVID-19 Prediction and Detection Using Machine Learning Algorithms: Catboost and Linear Regression. *American Journal of Theoretical and Applied Statistics*. Vol. 10, No. 5, 2021, pp. 208-215. doi: 10.11648/j.ajtas.20211005.11

Received: September 13, 2021; **Accepted**: October 4, 2021; **Published**: October 12, 2021

Abstract: A global pandemic COVID-19 has been rapidly spreading, and the predictions for infected rate shows how the cases will increase or decrease. Even though the number of people who get the corona vaccine is increasing, COVID-19 has been a serious worldwide problem. As machine learning and deep learning were implemented to predict COVID-19 in recent days, machine learning to predict the number of confirmed and death cases of COVID-19 was used. Prediction graphs of our proposed model play a crucial role for preventing more people getting infected. The project collected the number of daily infected cases in New York from March 21th 2020 to March 6th 2021. For precise results, the dataset in 6 different kinds of the machine learning methods was used. The methods were Decision Tree, Random Forest, Linear Regression, Gradient Boosting, XGboosting, and LGBM. RMSE and MAE values fluctuated from 9.95 to 68.85 and 5.99 to 58.76. The most accurate model was Linear Regression, RMSE and MAE with 9.96 and 5.99 for death cases and 597.61 and 346.04 for confirmed cases. Therefore, those prediction graph almost matched the same as the real number graph that the project drew with an actual dataset. The other dataset was about common COVID-19 symptoms, and the Catboost model listed from the most influential factor, breathing problem. Collecting data from other areas and specifying the patients' features could have improved the quality of the research, though overall the result was successful.

Keywords: COVID-19, Machine Learning, Linear Regression

1. Introduction

1.1. Background

Currently, with 83 million cases around the world, Coronavirus Disease (COVID-19) has caused a global pandemic with nearly 2.7 million deaths so far [1]. COVID-19 was first reported in the Wuhan province of China in December of 2019 [2]. WHO claimed that the virus in COVID-19 is mainly transmitted when a person coughs, sneezes, or speaks. The reason why COVID-19 is so fast in transmission is because it is spread through smear, and when the cough or sneeze of an infected person is inhaled through the respiratory system, the virus spreads [3]. Also, if you touch your eyes, nose, or mouth after touching contaminated objects, there is a high possibility that the virus will spread through contact. The COVID-19 incubation period ranges from 1 to 14 days (4 to 7 days on average). COVID-19 symptoms include fever, boredom, cough, shortness of breath, acute respiratory distress syndrome, and rare cases include small talk, sore throat, headache, bleeding and nausea and diarrhea. As corona-induced infections began in earnest, countries around the world implemented distancing and economic containment measures, which caused severe economic damage. Therefore, patients with this virus must be discovered before spreading the virus to more people [4]. COVID-19 has a significant impact on social, leisure, professional and family life as well as individual health problems. Many people have faced drastic changes in their lives and economic crises, and the COVID-19 crisis is creating many social changes. These changes could create problems with income polarization or educational inequality as the pandemic drags on. COVID-19 has had a huge impact on the entire national economy and industry from the daily lives of the entire nation, and many industries are directly and indirectly affected by entry restrictions and travel bans. Some predict that even if the COVID-19 situation calms down in the future, it will not be able to return to society before the

COVID-19 outbreak. The current impact of COVID-19 is unprecedented in the World Health Organization (WHO) to the extent that it declares COVID-19, and the tourism industry is almost shut-down. For diagnosing COVID-19, the real-time reverse transcription PCR (Real-time RT-PCR) used by corona tests in hospitals checks for DNA amplified by PCR reactions in real time, allowing quantitative analysis of the amount of DNA as well as the presence of target DNA in samples without electrophoresis [5]. Even though a vaccine such as Moderna, Pfizer, and AstraZeneca for this virus was invented, the antigenic drift recently occurred and some European countries started to lock down the country again [6]. 51.5% of people received at least one dose and 41.6% have been fully vaccinated as of June 5, 2021, however, in some countries, such as the Republic of Korea, only 13.7% of people got vaccinated at least once and 4.3% were fully vaccinated [7]. Figure 1 shows the cumulative sum graph of COVID-19 death cases in the US [8].

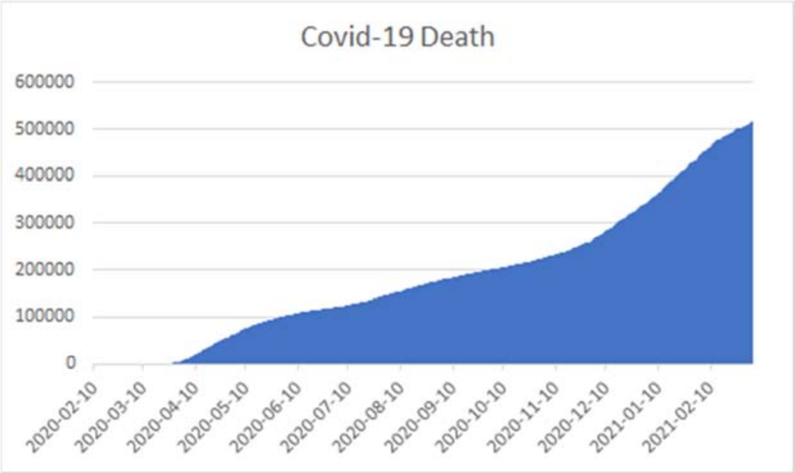


Figure 1. *COVID-19 death cases in US.*

1.2. Objective

A global pandemic COVID-19 has been infecting people all around the world, and the number of cases is rapidly increasing. The focus of the paper is to find out if machine learning programs are as accurate as they are to organize the status or predict the future situation. When the value of RMSE and MAE is close to 0, it is called a reliable predicted model. Tamhane and Mulge have shown large numbers of RMSE and MAE, which need to be improved by using a different machine learning model [9]. Furthermore, they did not show any significant symptoms for COVID 19. The goal of this paper is to show the important factors for COVID 19, choose a better model and improve the RMSE score of the prediction.

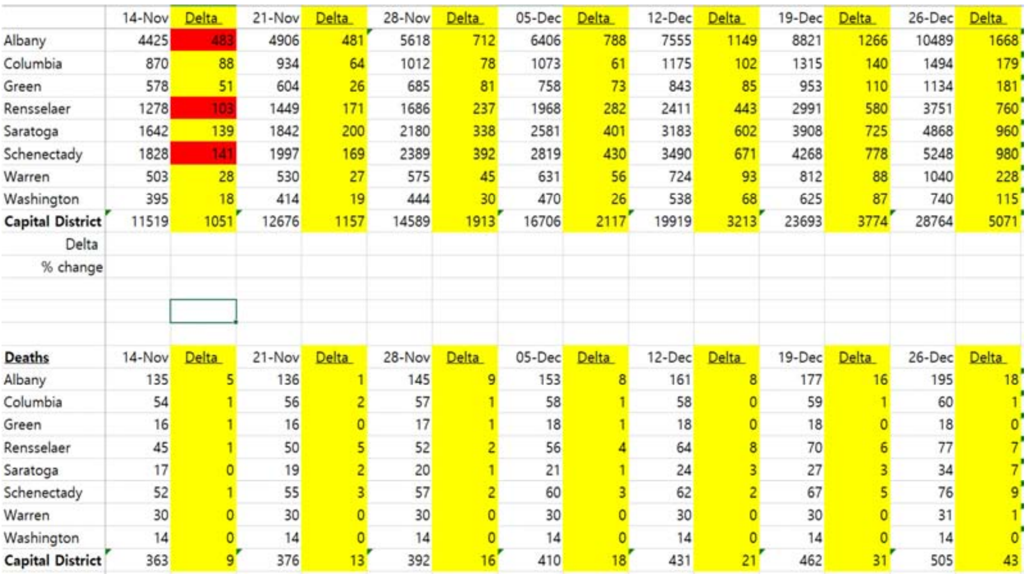
2. Related Works

Shrivastav and Jha collected data from the Ministry of Health and Family Welfare, resulting in MSE and RMSE by using the Gradient Boosting Model (GBM). Their objectives were to find the effect of the minimum temperature, maximum temperature, minimum humidity, and maximum humidity on the COVID-19. The greatest values of MSE and RMSE were 69855.46 and 272.49 in the active cases modeling, and 6614.20 and 81.33 in the recovered cases modeling [10]. Parbat and Chakraborty published the paper predicting the deaths, confirmed, and recovered cases was output by the Support Vector Regression model, about Novel Coronavirus 2019 dataset. Using the support vector regression model, total deaths and total recovered cases have shown a great accuracy with 0.00849 and 0.092142 of each MSE and RMSE for total deaths and 0.0030289 and 0.174036 for total confirmed [11]. Tamhane and Mulge utilized the data collected from Johns Hopkins University and one of the machine learning methods, Polynomial Regression and Support Vector Regression aimed at a successful prediction for COVID-19 outbreak. In order to find the most accurate prediction by leaving the least error rate, each MSE of Polynomial Regression and Support Vector Regression has drawn 185.7626758879287 and 2816.121760878107 [9]. Shahid and Muneeb analyzed data consisting of COVID-19 one from Brazil, Germany, Italy, Spain, UK, China, India Israel, Russia, and the USA. This research utilized LSTM, GRU, and BI-LSTM, ARIMA, SVR with polynomial and RBF kernels. The highest scores of MAE and RMSE are 0.007 and 0.0077 respectively through Bi-LSTM [12]. Gupta et al. analyzed the dataset through the Support Vector Machine, Prophet Forecasting Model, and Linear Regression Model, the paper makes reliable predictions of COVID-19 measuring. The super Vector Machine method is categorized into Active Rate, Cured Rate, and Death Rate, and the Active rate has shown 266.82 in MSE and 16.22463 in RMSE. Cured rate and Death rate have marked 139229.8 and 17.12 in MSE, and 273.1351 and 4.137632 in RSME as well [13].

3. Materials and Methods

3.1. Data Description

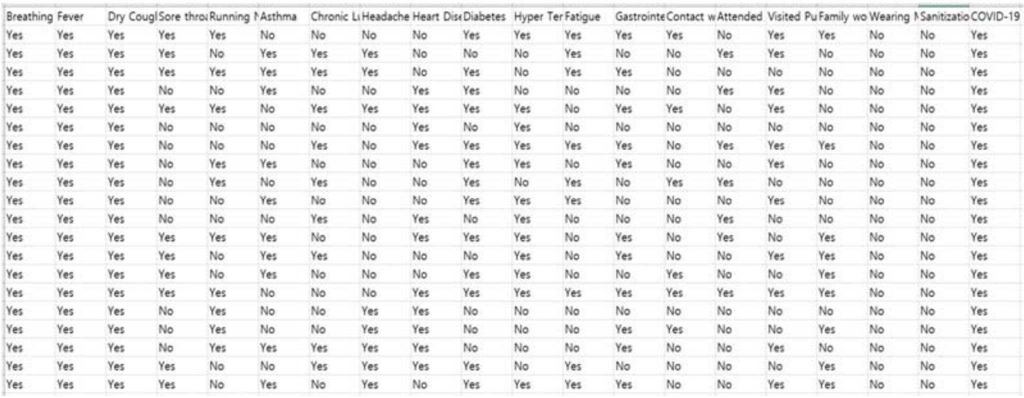
The number of confirmed cases of each day from March 24th, 2020 to March 6th, 2021 was collected to predict the prospects for the COVID-19 by finding the patterns. The columns were divided into 8 districts, and they are Albany, Columbia, Green, Rensselaer, Saratoga, Schenectady, Warren, and Washington. Figure 3 dataset consists of COVID symptoms and whether people suffer COVID-19 or not.



	14-Nov	Delta	21-Nov	Delta	28-Nov	Delta	05-Dec	Delta	12-Dec	Delta	19-Dec	Delta	26-Dec	Delta
Albany	4425	483	4906	481	5618	712	6406	788	7555	1149	8821	1266	10489	1668
Columbia	870	88	934	64	1012	78	1073	61	1175	102	1315	140	1494	179
Green	578	51	604	26	685	81	758	73	843	85	953	110	1134	181
Rensselaer	1278	103	1449	171	1686	237	1968	282	2411	443	2991	580	3751	760
Saratoga	1642	139	1842	200	2180	338	2581	401	3183	602	3908	725	4868	960
Schenectady	1828	141	1997	169	2389	392	2819	430	3490	671	4268	778	5248	980
Warren	503	28	530	27	575	45	631	56	724	93	812	88	1040	228
Washington	395	18	414	19	444	30	470	26	538	68	625	87	740	115
Capital District	11519	1051	12676	1157	14589	1913	16706	2117	19919	3213	23693	3774	28764	5071
Delta														
% change														

Deaths	14-Nov	Delta	21-Nov	Delta	28-Nov	Delta	05-Dec	Delta	12-Dec	Delta	19-Dec	Delta	26-Dec	Delta
Albany	135	5	136	1	145	9	153	8	161	8	177	16	195	18
Columbia	54	1	56	2	57	1	58	1	58	0	59	1	60	1
Green	16	1	16	0	17	1	18	1	18	0	18	0	18	0
Rensselaer	45	1	50	5	52	2	56	4	64	8	70	6	77	7
Saratoga	17	0	19	2	20	1	21	1	24	3	27	3	34	7
Schenectady	52	1	55	3	57	2	60	3	62	2	67	5	76	9
Warren	30	0	30	0	30	0	30	0	30	0	30	0	31	1
Washington	14	0	14	0	14	0	14	0	14	0	14	0	14	0
Capital District	363	9	376	13	392	16	410	18	431	21	462	31	505	43

Figure 2. *Preview of our dataset.*



Breathing	Fever	Dry Cougl	Sore throa	Running N	Asthma	Chronic Lu	Headache	Heart Dise	Diabetes	Hyper Ter	Fatigue	Gastrointe	Contact w	Attended	Visited Pu	Family wo	Wearing M	Sanitizatio	COVID-19
Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes
Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes	Yes	No	No	No	Yes
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No	No	No	No	No	Yes
Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No	Yes
Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	No	No	Yes
Yes	Yes	Yes	No	No	No	No	No	Yes	No	Yes	No	No	No	No	No	No	No	No	Yes
Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes
Yes	Yes	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes	No	No	Yes	No	No	No	Yes
Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes	Yes	No	No	No	No	Yes
Yes	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	No	Yes
Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	Yes	No	No	No	Yes	No	No	No	No	Yes
Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No	No	Yes
Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	Yes	No	Yes	No	No	Yes	Yes	No	No	Yes
Yes	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No	Yes
Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes
Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes	No	No	No	Yes
Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No	Yes	No	No	Yes
Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No	Yes

Figure 3. *Preview of dataset of COVID symptoms.*

3.2. Machine Learning

Machine learning mainly focuses on how computers can improve their performance through experiences [14]. Machine learning mainly consists of two different learnings, which are supervised learning and unsupervised learning. The dominant difference between them is whether they are labelled or not. For supervised learning, the input data should be labelled before the machine learning models trained them. On the other hand, unsupervised learning does not need any label for training [15]. Decision trees, random forest, logistic regression, support vector machines, neural networks, kernel machines, and Bayesian classifiers are the representative algorithms of supervised learning. Unsupervised learning mainly involves the clustering algorithms, and the K-means algorithm is symbolic one [14]. The main difference between machine learning and deep learning is that deep learning needs less human interventions. While the implement feature extraction manually before we put them into a model for classification or regression was needed, deep learning automatically extracts the feature from the given dataset.

3.3. Boosting Algorithm

Boosting algorithms are one of the representative models of ensemble algorithms. While a random forest, another model of ensemble algorithms, is based on bagging, which operates parallelly, while a boosting operates sequentially. Boosting is a decision tree based algorithm, which can not handle the categorical value by itself. In other words, the basic boosting models must go through the preprocessing stage before training. Even though the One-hot Encoding, Label Encoder function, or the dummy coding could be utilized, these methods could trigger a lot of memory usage and lower the speed of calculation. Furthermore, the basic boosting algorithm learns the residual error sequentially and predicts with the process. However, the boosting algorithm is vulnerable to overfitting [16].

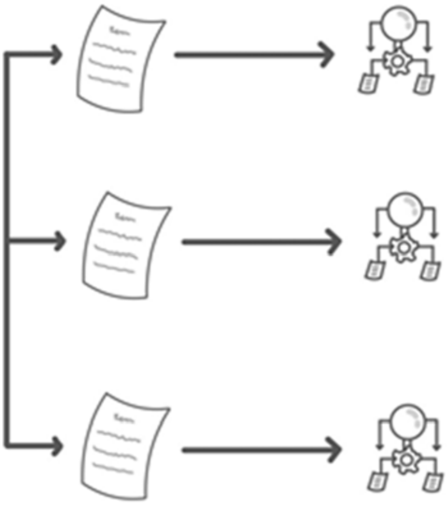


Figure 4. *Overall process of bagging algorithm.*

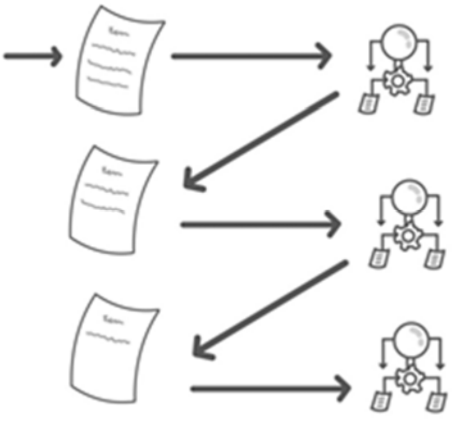


Figure 5. *Overall process of boosting algorithm.*

3.4. Catboost

Catboost algorithm is an ordered boosting algorithm, which focuses on solving the overfitting problem of the boosting algorithm. The ordered boosting is a method that builds the model for calculating the residual error from limited data and then calculates the whole data through the model. Furthermore, adding random permutation in the ordered boosting prevents overfitting effectively. For preprocessing the categorical variables, the catboost algorithm calculates a sample mean for the variables in the same category which goes through the random permutation. Catboost algorithm increases the training speed through feature combinations which gathers the variables with the same information gained into a group. Moreover, while the other ensemble algorithms such as random forest and gradient boosting should utilize the GridSearch or RandomizedSearch for finding the optimal hyperparameter, the catboost does not go through those stages as the default setting of the parameter is already satisfied [17].

3.5. Linear Regression

The regression with a single independent variable is known as univariate regression, while with multiple independent variables is called multivariate regression. Linear regression model implements a straight line to find the optimal fitted line. On the contrast, logistic regression and non linear regression utilizes a curved line for fitting into the dataset. By fitting a linear equation, it attempts to find the relationship between two variables. One is called an independent variable and the other is called a dependent variable [18]. The formula for linear regression can be found on formula (1).

$$y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$

while y denotes the dependent variable for the independent variable x, β_0 denotes the intercept, the predicted value of y when the x is 0, β_1 denotes the regression coefficient, ϵ denotes the error of the estimate. The purpose of the linear regression is to find the best fit line for the given dataset by seeking for the regression coefficient which can decrease the total error of the model.

4. Result

4.1. Analyzing Important Symptoms for COVID-19

According to the COVID-19 symptom dataset given by Kaggle, which is available on https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence, seven kinds of machine learning methods, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, XGBClassifier, LGBM Classifier, and Cat Boost Classifier were used for detecting and listing the factors (symptoms) in order of influence. Among the seven methods, Decision Tree Classifier, Random Forest Classifier, LGBM Classifier, and Cat Boost Classifier have shown the greatest accuracy of 98.41%. As the Cat-Boost algorithm is based on decision tree models, it could show the feature importances of each variable [17]. Therefore, by visualizing the feature importances through the function, The project could conclude that breathing problems and attending large gatherings are the most influential factors on COVID-19.

4.2. Results of RMSE and MAE Scores from Various Models

The models used to predict the numbers for the future were Decision Tree Regressor, Random Forest Regressor, Linear Regression Model, Gradient Boosting Regressor, XG Boosting Regressor, and Light Gradient Boosting Regressor. The Linear Regression model has shown the greatest accuracy of RMSE and MAE in both deaths and confirmed case predictions. As shown by Figure 8 and Figure 9, the RMSE and MAE for the predicted death number were each 9.96 and 5.99. As the values are close to 0, the predicted graph was drawn almost the same as the actual graph made after collecting the data. As demonstrated by Figure 10 and Figure 11, the RMSE and MAE for the predicted cases were each 597.61 and 346.04. Small values of RMSE and MAE resulted in a reliable prediction that had nearly the same shape as the real number graph.

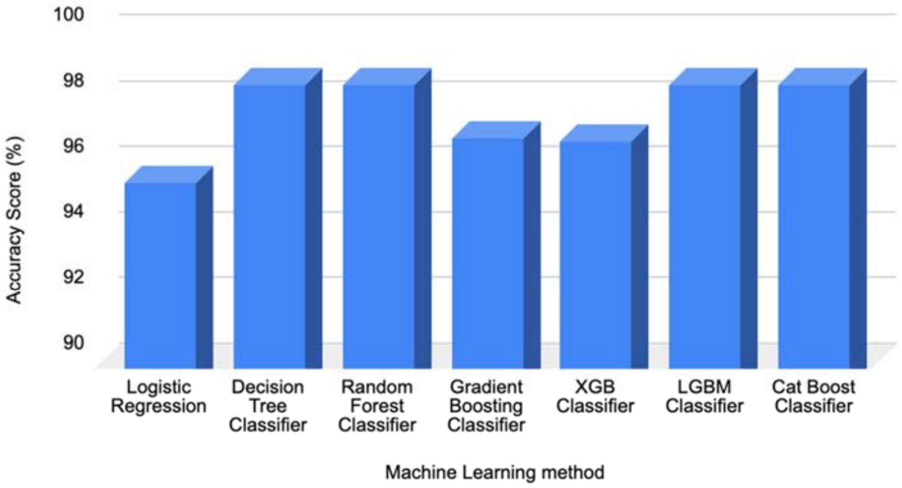


Figure 6. *Accuracy score from machine learning models.*

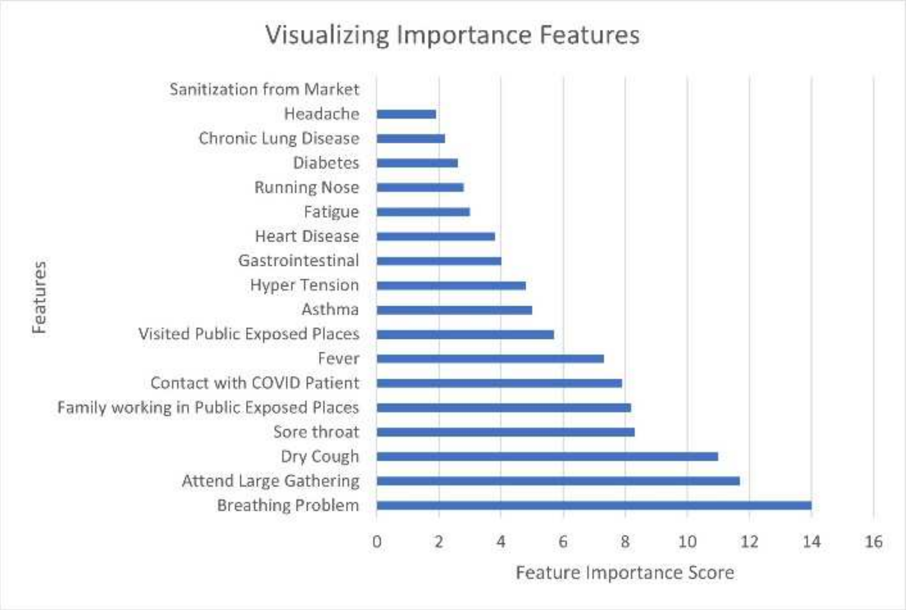


Figure 7. *Visualizing important symptoms of COVID-19.*

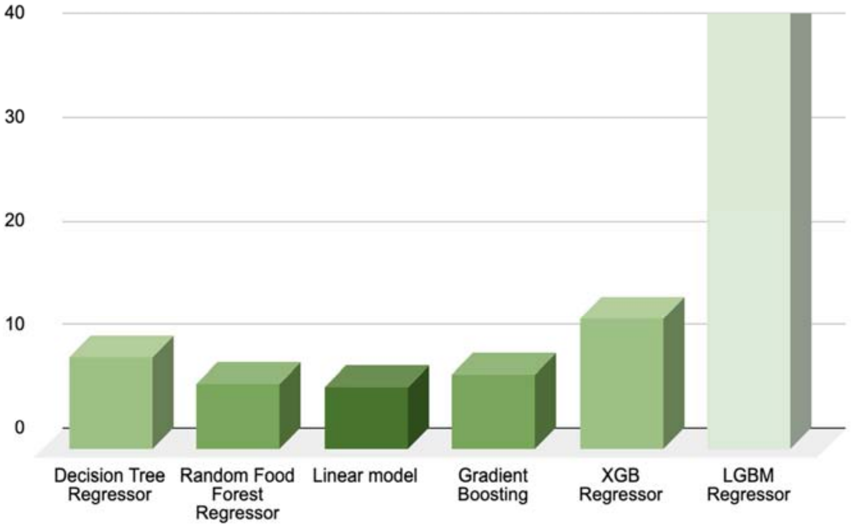


Figure 8. *Visualizing graph of RMSE scores from machine learning models for death cases.*

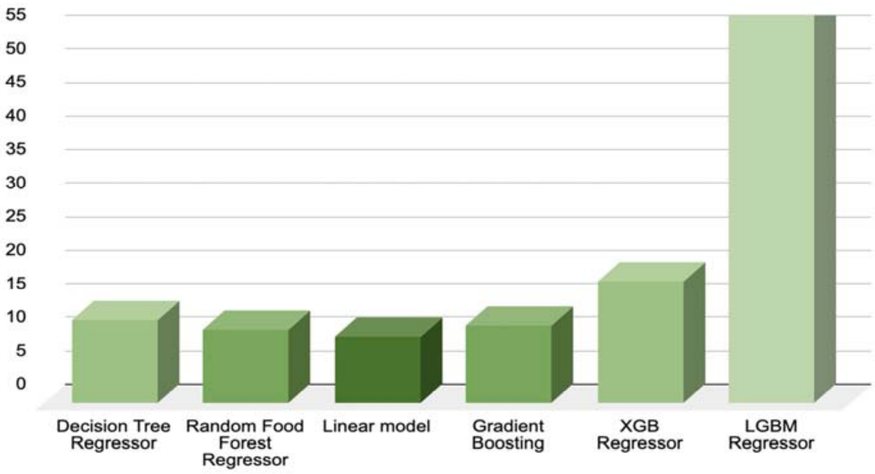


Figure 9. *Visualizing graph of MAE scores from machine learning models for death cases.*

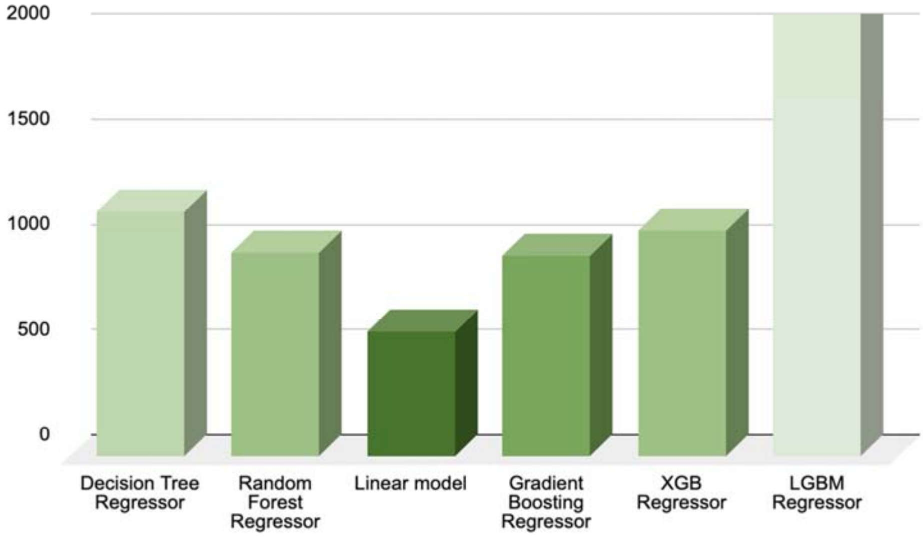


Figure 10. *Visualizing graph of RMSE scores from machine learning models for confirmed cases.*

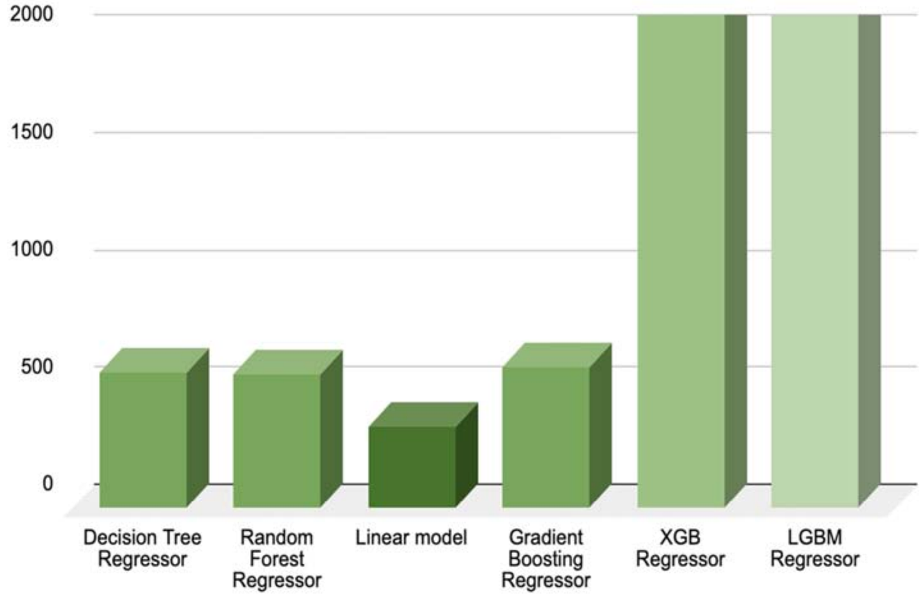


Figure 11. *Visualizing graph of MAE scores from machine learning models for confirmed cases.*

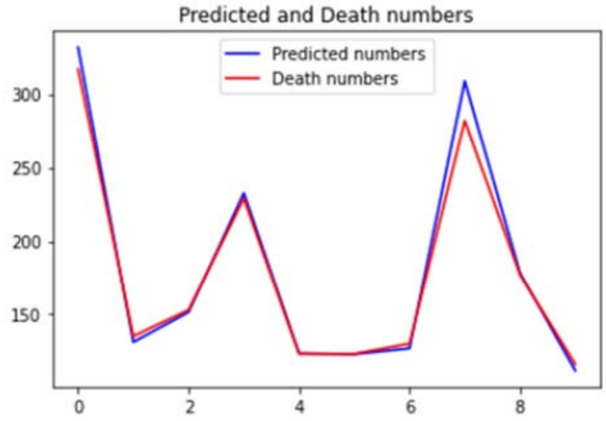


Figure 12. *Visualizing graph of predicted numbers and death numbers.*

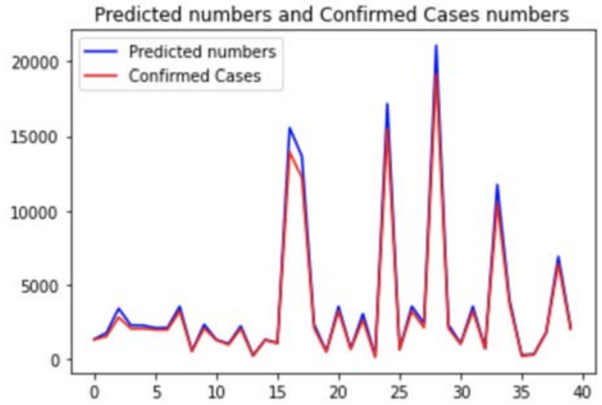


Figure 13. *Visualizing graph of predicted cases numbers and confirmed cases numbers.*

5. Discussion

5.1. Principle Finding

Through the symptom analysis, the project figured out which factor of COVID-19 is most influential among 18 common symptoms. Especially breathing problems, attending large gatherings, dry cough and sore throats are the most representative symptoms among them. The prediction graph and the real number graph matched almost the same with little difference. This explains how accurate the machine learning model was; the Linear Regression model had the best prediction between 6 other models. On the other hand, LGBM had the highest RMSE and MAE, which means that its prediction was the worst among other algorithms. Boosting Model or Random Forest have the greatest performance in accuracy and RMSE in general, but in this COVID-19 prediction case, Linear Regression model showed better accuracy.

5.2. Limitation

To output more accurate predictions, there has to be more data. If the dataset included more data, it could have used a Deep Learning model such as Deep Neural Network (DNN), Recurrent Neural Network (RNN), Long Short Term Memory LSTM) and Gated Recurrent Unit (GRU). As the project only focused on the confirmed cases in New York, there was a limitation for digging into advanced and precise predictions. The data recorded was just the number of new daily cases, which did not consider their age or gender as a feature. If the infected cases were more specified, the result would be even more credible.

6. Conclusion

The machine learning model helped find the most influential factor of COVID-19 symptoms and visualize the predictions for future cases in New York by constructing a graph. From the lists of the number of people who have the symptoms, the model found out which was the most influential factor for COVID-19 that confirmed people commonly had. Patients have shown that many of them had a breathing problem and attended large gatherings before being confirmed. The prediction graph for future confirmed cases in New York created by the Linear Regression model, which had the smallest value of RMSE and MAE, presented nearly the same shape as the real number graph. With more data and a more accurate machine learning model that digs deeper, it could have brought better predictions. Collecting more information about confirmed cases and symptoms from other areas can improve the quality and content of the research paper. Overall, machine learning allowed to make reliable predictions.

References

[1] Coronavirus in the WORLD: Latest case and death tolls in 24h per country. (n.d.). Retrieved March 24, 2021, from https://www.sortiraparis.com/news/in-paris/articles/212134-coronavirus-in-the-world-as-of-datadatestodayfrlatest-latest-case-and-death-toll/lang/en

[2] Novel Coronavirus – China. (2020, January 13). Retrieved January 09, 2021, from https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

[3] Coronavirus disease (COVID-19): How is it Transmitted? (n.d.). Retrieved March 24, 2021, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-COVID-19-how-is-it-transmitted

[4] Coronavirus. (n.d.). Retrieved March 24, 2021, from https://www.who.int/health-topics/coronavirus#tab=tab_3

[5] World Health Organization. (n.d.). Episode #14 - COVID-19 - Tests. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/science-in-5/episode-14---COVID-19---tests?gclid=Cj0KCQjw5PGFBhC2ARIsAIFIMNclWkv6n-pt-Zl06aTK2VBepUdH_u42soguf6QpPg28jJdtWnm7LmoaAuZVEALw_wcB.

[6] Several european countries under new COVID lockdown restrictions. (n.d.). Retrieved March 24, 2021, from https://www.voanews.com/COVID-19-pandemic/several-europeAn-countries-under-new-covid-lockdown-restrictions

[7] Bhatia, G., Dutta, P. K., & McClure, J. (2021, June 3). *COVID-19 vaccine rollout: charts, maps and eligibility by country*. Reuters. https://graphics.reuters.com/world-coronavirus-tracker-and-maps/vaccination-rollout-and-access/

[8] The COVID Tracking Project. (n.d.). https://covidtracking.com/

[9] Tamhane, R., & Mulge, S. (2020). Prediction of COVID-19 outbreak using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, *7* (5).

[10] Shrivastav, L. K., &amp; Jha, S. K. (2020). A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of COVID-19 in India. Applied Intelligence. https://doi.org/10.1007/s10489-020-01997-6

[11] Parbat, D., & Chakraborty, M. (2020). A Python Based Support Vector Regression Model for Prediction of COVID-19 Cases in India. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3591840

[12] Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, *140*, 110212. https://doi.org/10.1016/j.chaos.2020.110212

[13] Gupta, A. K., Singh, V., Mathur, P., & 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*, *24* (1), 89–108. https://doi.org/10.1080/09720502.2020.1833458

[14] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349* (6245), 255–260. https://doi.org/10.1126/science.aaa8415

[15] Cayir, A., Yenidogan, I., & Dag, H. (2018). Feature Extraction Based on Deep Learning for Some Traditional Machine Learning Methods. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. https://doi.org/10.1109/ubmk.2018.8566383

[16] Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, *14* (771-780), 1612.

[17] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin (2018), “CatBoost: unbiased boosting with categorical features“, *Advanced in Neural Information Processing Systems* 31, pp. 6639-6649.

[18] Khademi, F., Akbari, M., Jamal, S. M., & Nikoo, M. (2017). Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete. *Frontiers of Structural and Civil Engineering*, *11* (1), 90-99.