

Noble Machine Learning Approaches for Lock Downing Area during Coronavirus (COVID-19) Pandemic Waves


Motaleb Hossen Manik*Department of Computer Science
and Engineering, Khulna University of
Engineering & Technology,
Khulna-9203, Bangladesh

Abstract

Novel coronavirus has become a pandemic worldwide since December, 2019. All ongoing plans have been propounded due to COVID-19. Areas with large number of infected and death patients have been taken under lockdown. Continuous locking down areas have created a huge impact on economy especially for the under developed and developing countries. But most of the countries are locking down their areas without any assumption. In this situation, this paper proposes a model to determine which area of a country should be taken under lockdown immediately. The proposed work uses dataset containing the information of different region of the world with their infected, death, recovered patients, and area of the region, population and successfully presence of lockdown. Finally, Machine Learning algorithms have been applied to determine the area to lockdown with the highest accuracy of 95.185%.

Keywords: Novel coronavirus; COVID-19; Machine learning; Prediction

Corresponding author:

M.D. Motaleb Hossen Manik, Department of
Computer Science and Engineering, Khulna
University of Engineering & Technology
Khulna-9203, Bangladesh, E-mail: mkmanik557@gmail.com

Citation: Manik MMH (2021) Noble
Machine Learning Approaches for
Lock Downing Area during Coronavirus
(COVID-19) Pandemic Waves. Am J Compt
Sci Inform Technol Vol.9 No.8:107.

Received: August 03, 2021; **Accepted:** August 17, 2021; **Published:** August 24, 2021

Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus [1]. It was first identified in December 2019 in Wuhan, Hubei, China which has created the current pandemic. The initial confirmed case has been identified on 17 November 2019 in Hubei [2].

The severe acute respiratory syndrome (SARS) virus strain known as SARS-COV is an example of a coronavirus. Many health experts believe that the new strain of coronavirus likely originated in bats or pangolins [3]. The common symptoms of this disease are fever, fatigue, cough, shortness of breath. Recently discovered symptoms are loss of smell and taste. Almost 188 Million confirmed cases have been found till July, 2021 with 172 Million recovered patients and 4.05 Million deaths [4]. Different precautionary measures have been taken to stop spreading coronavirus. Cleaning hands, maintaining safe distance, wearing masks, staying at home, refraining from touching eyes, nose, and mouth are the main precautionary. In order to keep people at home and maintain social distance, different parts of the world are being kept under lockdown indefinitely. Undoubtedly, this lock-down is ruining the global economy. However, the matter of concern is that most countries are locking their territories without any statistical analysis which is creating a burden for the people of that region. People are losing their jobs and those who are not being fired off, cannot attend their workplace. A statistics shows that in the United States the proportion of people out of work has hit form 3.7% to 10.4%, according to the International

Monetary Fund (IMF) which is really a matter of concern. Among the locked down regions, some are marked which should not be taken under lockdown according to their infected, death and recovered patients and even Sweden has faced the coronavirus without a lockdown [5].

Recently more than 1,230,000 articles have been published on COVID-19 topic. Among them, Machine Learning based models are mostly on Forecasting, Diagnosis, Environmental Dependencies, Survey and Screening. Again almost all articles related to lockdown are mainly planning based on how to set up lockdown and how it can be succeed. So, Machine Learning models on how and where to set lockdown are few [6]. Cole et al. have proposed a two steps based model where first step uses Machine Learning and second step uses Augmented Synthetic Control Model. The full system reveals the impact of lockdown on air pollution and Health. They have showed that the reduction of concentration of NO₂ could have saved 10,822 deaths in China. But their model does not provide any suggestion on either to set lockdown or not. Alvarez et al. have suggested a SIR epidemiology model to control the fatalities by minimizing the cost of the lockdown. Authors could not provide any solution on which areas should be taken under lockdown. No other works have been found related to my research field [6,7].

In this situation, a model is required to determine which areas should be taken under lockdown and which should not. This paper proposes a Machine Learning based model to do so. Initially a dataset has been used which includes different region of world

with their infected, death, recovered patients, area of the region, population of the region and finally the successful presence of lockdown in those areas. The areas with successful presence of lockdown have been marked as 1 and other are marked as 0. Finally, various Machine Learning algorithms have been applied on them to determine the output either as to keep lockdown or not.

The rest of the paper is constructed as follows. Section 2 explains the materials and methods of this work. In section 3, result of the research has been composed. The discussion and some directions for future work have been explained in section 4. Finally section 5 concludes the paper.

Materials and Methods

Materials

Dataset collection: A dataset has been constructed from the available information of different territory that contained COVID-19 related data like patients information, success of their lockdown mechanism and density of that region. The created dataset contains 7 features including the name of different region of the world, infected, death and recovered patients of that region, population and area of that region and finally the target feature, which is the successfully presence of lockdown on that region. The dataset contains the mentioned information of 100 different regions of the world [8,9].

Dataset preparation: Since the numerical values used regarding the attributes are large, they require long time to be processed. That is why they have to be normalized and kept between 0 and 1. Again, there were some missing attribute values. They have been filled by average values of the corresponding attribute. Since the dataset is manually collected by the author, some derived attributes have been recalculated from the single valued attributes. The doting and ration attributes have been calculated from the ratio of death patients hence infected patients and recovered patients hence infected patients, respectively and the density has been calculated from the ration of population and area of a specific locality. Finally the target attribute has been converted to 0 and 1 based on succeed of lockdown.

Dataset separation: The used dataset contains 100 different region's data from which 70 have been used as training purpose and rest 30 have been used for testing purpose. They have been taken randomly for multiple times for training.

Methods

Support Vector Machine (SVM): Support Vector Machine is one of the renowned Machine Learning algorithms that can be applied on classification as well as regression. The actual concept is to draw hyper planes that can make a partition on the given dataset into desire classes. Actually two hyper planes are drawn to separate the dataset and the key objective is to maximize the distance between the hyper planes. A concept of kernel trick is used for higher dimensional data to convert them into acceptable format so the classification task can be easier. For the sake of the

used dataset, two classes have been predicted namely either lockdown should be imposed or not on the basis of six features.

Decision Tree (DT): Decision Tree is another mostly used Machine Learning tool that can be named a tree-structured classifier. Morally the internal nodes represent the used features of the dataset and the adjacent braches are the rules for applying the decision on which path the algorithm should follow to draw a conclusion on given input. The leaf nodes are the actual decisions that follow the path through root to leaf via decision branches. Multiple methods are available to build a decision tree where the basic idea is same on calculating the entropy and information gain or gene index. In a particular feature node, the path is selected either with higher information gain or lower gene index. In the used dataset, multiple paths from each node have been created and finally 22 leaf nodes are founded for the motive of decision placing.

Random Forest (RF): Like the previously recounted methods, Random Forest (RF) another remarkable methods on classification that is based on ensemble learning method. A RF can be marked as a combination of decision trees. A group of mentors' decision is better than a single decision that is the idea in the background of RF. The dataset is divided into sub datasets and assigned to each decision tree to set their decision on the input. Finally a voting mechanism is applied on the decisions of a particular input. The mostly used voting mechanism is average voting that averages the decision taken by different decision trees. In applied dataset, maximum result achieved on estimators value of 34.

Results and Discussion

Any Machine Learning algorithm's acceptability is measured on its accuracy on how the algorithm is classifying the data. In this research, actually three major Machine Learning algorithms have been applied namely Support Vector Machine, Decision Tree and Random Forest. Their accuracy have been displayed. From the measured data, the clear vision is that Random Forest algorithm is providing maximum accuracy with 95.185% while SVM and DT are providing 85.55% and 91.481% accuracy, respectively. A large dataset could have shown much more accuracy with the models (Figure 1).

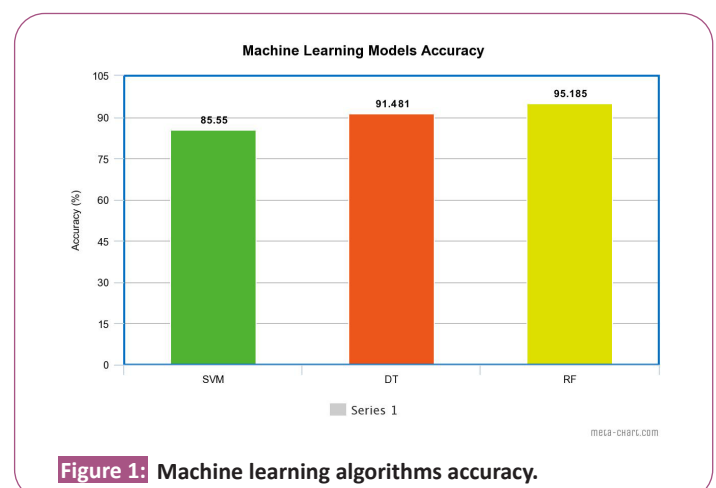


Figure 1: Machine learning algorithms accuracy.

COVID-19 has become the most talked topic recent days. More the days are passing, the more the efficient lock downing approaches are being needed. But a sudden lockdown a specific locality can lead the people towards a miserable life leading. Furthermore, people may not be provided with proper financial or food support from the authority though they are trying their best. Thus a lock downing approaches can be crucial remedy where locality selection can be one of them.

In this study, the area that should be kept on locked down during the latter wave of COVID-19 has been predicted based on the previous data on success rate of lockdown. The data that are fateful on these are the infected patients, death patients, recovered patients and the area of the locality. For the case study, three mostly damaged counties (USA, Brazil and India) during coronavirus pandemic have been considered. State wise lockdown success rate has been taken into account for the decision building. Three Machine Learning algorithms have been applied on them and most success rate have been found with Random Forest algorithm that is a package of decision trees.

Recommendation for future work

This research introduces the idea of lock downing specific region on the basis of previous coronavirus wave's data. Though the trained models expose remarkable accuracy on new data, a better result is anticipated on rich dataset with new data available. Furthermore, deep learning models can be promising one on larger dataset. In this research, the infected, recovered, death patients and the area of a particular region have been considered. But other minor factors like weather condition, habitual factors of folks can be brought on judgment for further approach.

Conclusion

Since during COVID-19 pandemic, randomly regions were being locked down, an approach to correctly determine the regions that should be brought under lockdown was the requirement. In this research, different Machine Learning models have been

applied on a manual dataset that has an annotated feature on the success of lockdown of a particular region. Among the models, Random Forest algorithm displayed higher accuracy than rest of the algorithms. Further work on this model can be incorporated on the basis of availability of new data.

References

- 1 World Health Organization (2020) Coronavirus.
- 2 Singhal T (2020) A review of coronavirus disease-2019 (COVID-19) *Indian J Pediatr* 87: 281-286.
- 3 Pandit MK, Banday SA, Naaz R, Chishti MA (2021) Automatic detection of COVID-19 from chest radiographs using deep learning. *Radiography* 27: 483-489.
- 4 Segrest SL, Hurley-Hanson AE, Giannantonio CM (2021) Refugees/displaced people in the workplace.
- 5 Marquioni VM, De Aguiar MA (2020) Quantifying the effects of quarantine using an IBM SEIR model. *Scalefree Networks* 1: 138-199.
- 6 Cole MA, Elliott RJ, Liu B (2020) The impact of the Wuhan COVID-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. *Environ Resour Econ* 76: 553-580.
- 7 Alvarez FE, Argente, Francesco L (2020) A simple planning problem for covid-19 lockdown. National Bureau of Economic Research (NBER).
- 8 Xia Y, Ning P, Wang Z, Chen X (2020) Calming the cytokine storm in pneumonia by biomimetic nanoparticles 3: 18-20.
- 9 Yanti B, Mulyadi E, Wahiduddin W, Novika RG, Arina YM, et.al, (2020) Community knowledge, attitudes, and behavior towards social distancing policy as prevention transmission of COVID-19 in indonesia. *J administrasi kebijakan Indones* 8: 4-14.