International Journal of Modern Developments in Engineering and Science
Volume 1, Issue 7, July 2022
https://www.ijmdes.com | ISSN (Online): 2583-3138

Vaccissure with Predictive Capacity of COVID-19

Chrissie Susan Alex[1*], Jinu Thomas[2]
[1]*M.Tech. Student, Department of Computer Science and Engineering, Saintgits College of Engineering and Technology, Kottayam, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Saintgits College of Engineering and Technology, Kottayam, India*

***Abstract*: COVID-19 infections can spread silently, due to the simultaneous presence of significant numbers of both critical and asymptomatic to mild cases. While, for the former reliable data are available (in the form of number of hospitalization and/or beds in intensive care units), this is not the case of the latter. Hence, analytical tools designed to generate reliable forecast and future scenarios, should be implemented to help decision-makers to plan ahead. With this application we will be able to check whether an area will be a hotspot or not by predicting the TPR rate and it also provide a way to check if a person is at risk or not at risk of having covid after getting fully vaccinated. We collected the individual details by sending out the Google form. The details are then categorized as fully vaccinated, dose 1 vaccinated, not vaccinated. This is again classified on the prediction-based category also. We use Random Forest algorithm for the predictions process. The technology we use is Machine learning.**

***Keywords*: Machine learning, Random Forest algorithm.**

1. Introduction

The COVID-19 outbreak, which began in China, has spread throughout the world. The World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020 [1]. The disease has disrupted global trade, employment, and travel, and many governments have had to take stringent measures to control the virus's spread and reduce the burden of morbidity and mortality so that health-care systems can function [2]. As a primary measure to prevent the spread of COVID-19, citizens in many countries around the world have been advised to stay at home and practise social distancing for as long as possible. The COVID-19 application of the Department of Health and Family Welfare, Government of Kerala summarizes the COVID-19 situation in Kerala. The application also contains information to various important information, guidelines, and other details for references. COVID infection prevention by following Covid appropriate behaviour got strengthened with introduction of COVID vaccination. Fast vaccination is very useful to safeguard people. Among all people the first priority is elderly people followed by people with comorbidities and others. COVID-19 vaccines protect people from getting infected and become severely ill, and significantly reduce the likelihood of hospitalization and death. The data as on date suggests the conversion rate is stable because the People are following instructions, taking vaccines. However, it can be further improved with one singular cooperation from all i.e., positive person having comorbidities should not delay the arrival in Hospital and take treatment as per the treatment protocols.

2. Literature Survey

Since the beginning of the COVID-19 Pandemic, machine learning has been used in a variety of ways to combat it. Predicting the severity (in the early stages) of COVID-19 patients is one area of research. Several studies have previously been conducted to predict the severity of Covid-19 patients using various forms of Machine Learning Models. In many studies, machine learning (ML) and deep learning (DL) models were used to diagnose COVID-19 patients. [3]-[6]

Y. Sun et al. [4] study indicates that utilizing the support vector machine (SVM) method, it is possible to accurately categorize 85% of COVID-19 cases.

The study shown in [5] used several classifiers, including XGBoost and multilayer perceptron (MLP), to diagnose corona virus patients with a classification accuracy of 91%.

Somani et al. [7] developed a model that predicts the risk of death in COVID-19 hospitalized patients. The XGBoost model was combined with baseline comparator models to create the prediction model. The model achieved AUC Scores for Mortality of 0.89 after three days, 0.85 after five and seven days, and 0.84 after ten days.

Yan et al. [8] developed a machine learning model to predict the mortality and criticality of COVID-19 patients. The ML method XGBoost was used, and it was discovered to be 93% accurate. In this model, the main predictors of mortality risk were LDH, lymphocytes, and high-sensitivity CRP.

3. Proposed Methodology

The pure predictive capacity of these COVID-19 predictors in terms of hospitalizations, in comparison to the TPR. While the TPR can be thought of as a measure of the number of infections that occur on a given day, taking into account unknown cases, indicators based on officially reported positive cases (incidence and growth rate) measure the variation of official cases.

We are implementing using the Random Forest Algorithm. Random Forest is a well-known machine learning algorithm from the supervised learning technique. It can be applied to both classification and regression problems in machine learning. It is based on the concept of ensemble learning, which is a process

*Corresponding author: chrissie.susan05@gmail.com

that involves combining multiple classifiers to solve a complex problem and improve the model's performance.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset, as the name implies. Instead of relying on a single decision tree, the random forest takes the predictions from each tree and predicts the final output based on the majority vote of predictions. The greater the number of trees in the forest, the higher the accuracy and the lower the risk of overfitting.

A. Why we use Random Forest?

Random forest is used because it requires less training time than other algorithms and predicts output with high accuracy, even for large datasets. It can also keep accuracy even when a large portion of the data is missing.

B. Implementation

Python is used to implement the Random Forest Algorithm. We will use the dataset from the Kerala government's Covid Portal for this. We obtained the information by data scraping. The process of importing information from a website into a spreadsheet or a local file saved on your computer is known as data scraping. It's one of the most efficient methods of retrieving data from the web and, in some cases, channeling that data to another website. We can compare the Random Forest classifier to other classification models using the same dataset, such as Logistic Regression, KNN, Random Forest Classifier, Gradient Boosting Classifier, and so on.

Implementation Steps are:

1. Data Scrapping
2. Data Pre-processing
3. Model fitting\training
4. Model prediction\validation
5. Result visualization

1) Data Scrapping

Data scraping is the process of importing information from a website into a spreadsheet or a local file saved on your computer. It's one of the most efficient ways to get data from the web and, in some cases, channel that data to another website. Here, we scrape data from the Kerala Government's Covid Portal.

2) Data Pre-processing

The transformations we apply to our data before feeding it to the algorithm are referred to as pre-processing. Data Pre-processing is a technique for converting raw data into a clean data set. In other words, whenever data is collected from various sources, it is collected in raw format, which makes analysis impossible. Data pre-processing is an essential step in Machine Learning because the quality of data and the useful information that can be extracted from it directly affects our model's ability to learn; thus, it is critical that we pre-process our data before feeding it into our model.

3) Model fitting/Training

A machine learning model is created by learning and generalizing from training data, then applying that knowledge to new data that it has never seen before in order to make predictions and fulfil its purpose. The training model is used to process the input data through the algorithm in order to correlate the processed output with the sample output. The correlation result is used to change the model. "Model fitting" refers to this iterative process. The precision of the model is dependent on the accuracy of the training or validation dataset.

4) Model Prediction/Validation

Model validation is the process by which a trained model is evaluated using a testing data set. The testing data set is a different piece of similar data from which the training set is derived. The testing data set is primarily used to test the speculation capacity of a prepared model. Model validation occurs after the model has been trained. Model validation, like model training, seeks to identify an ideal model with the best execution.

5) Data Visualization

The Graph shows the Accuracy of Random Forest with Logistic Regression, KNN and Gradient Boosting Classifier (Fig. 1).

Hotspot Prediction Accuracy Comparison:

When the accuracy of Random Forest is compared to that of other models, it is discovered that Random Forest has the highest accuracy (Table 1).

After Vaccination Chance of having Covid Prediction Accuracy Comparison:

When the accuracy of Random Forest is compared to that of other models, it is discovered that Random Forest has the highest accuracy (Table 2).

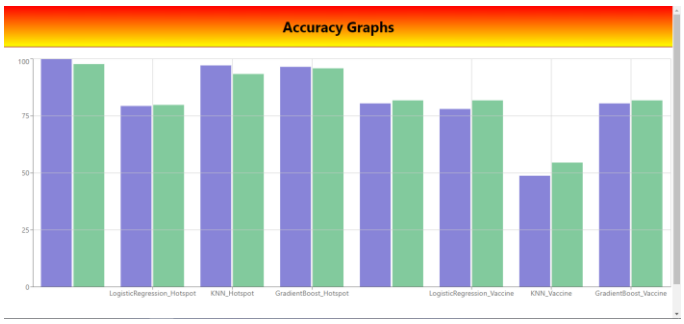

Fig. 1. Accuracy of Random Forest with Logistic Regression, KNN and Gradient Boosting Classifier

Table 1
Hotspot prediction accuracy comparison

Algorithm	Random Forest	Logistic Regression	KNN	Gradient Boosting Classifier
Accuracy Score on Train Set	1.0	0.79	0.97	0.96
Accuracy Score on test Set	0.97	0.79	0.93	0.95

Table 2
After vaccination chance of having covid prediction accuracy comparison

Algorithm	Random Forest	Logistic Regression	KNN	Gradient Boosting Classifier
Accuracy Score on Train Set	80.48	78.04	48.78	80.48
Accuracy Score on Test Set	81.81	81.81	54.54	80.81

4. Results

Given the District, confirmed cases, recovered cases, active cases, and deaths as input, we could predict whether a location is a hotspot or not. The TPR will be predicted; we have set a limit of 10; if the TPR is less than 10, there will be no hotspot; if the rate is greater than 10, there will be a hotspot. We also compared Random Forest to other Machine Learning models such as Logistic Regression, KNN, and Gradient Boosting Classifier and concluded that Random Forest provides the best prediction accuracy when compared to other models. The train set accuracy for each model is as follows: Random Forest has a score of 1.0, Logistic Regression has a score of 0.793, KNN has a score of 0.971, and Gradient Boosting Classifier has a score of 0.965. The validation accuracy for each model is as follows: Random Forest has a score of 0.977, Logistic Regression has a score of 0.798, KNN has a score of 0.933, and Gradient Boosting Classifier has a score of 0.958.

5. Conclusion

We present a forecasting method for predicting the impact of COVID-19 disease on the public health system in the short term. With various parameters, our model will predict whether or not an area should be declared a containment zone. Our model would assist health care workers in identifying hotspot areas and providing the necessary assistance. We tested the accuracy of various models and concluded that Random Forest will provide the best accuracy score. We attempted to conduct a study on people who developed Covid 19 following vaccination.

References

[1] Covid-19 Situation report (2020). Coronavirus disease 2019 (COVID-19): situation report – 78. World Health Organization. 2020. Apr 07, [2020-07-01]. https://www.who.int/docs/defaultsource/coronaviruse/situation-reports/20200407-sitrep-78-covid-19.pdf.

[2] Mahmood S, Hasan K, Colder Carras M, Labrique A (2020). Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill. 2020 Apr 16;6(2):e18980.

[3] A. F. de Moraes Batista, J. L. Miraglia, T. H. Rizzi Donato and A. D. Porto Chiavegatto Filho, "Covid-19 diagnosis prediction in emergency care patients: a machine learning approach", 2020.

[4] Y. Sun et al., "Epidemiological and Clinical Predictors of COVID-19", Clinical Infectious Diseases, vol. 71, no. 15, pp. 786-792, 2020.

[5] "Data analytics for novel coronavirus disease", Informatics in Medicine Unlocked, vol. 20, pp. 100374, 2020.

[6] G-M et al., "A machine learning algorithm to increase covid-19 inpatient diagnostic capacity", PLOS ONE, vol. 15, 2020.

[7] V et al., "Machine learning to predict mortality and critical events in a cohort of patients with covid-19 in new york city: Model development and validation", J Med Internet Res, vol. 22, no. 11, pp. e24018, Nov 2020.

[8] L. Yan, H. T. Zhang, Y. Xiao, M. Wang, Y. Guo, C. Sun, T et al., "Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan", 2020.