Journal of Pharmaceutical Research International

33(38A): 202-217, 2021; Article no.JPRI.70272
ISSN: 2456-9119
(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919, NLM ID: 101631759)

Pharmacy Impact on Covid-19 Vaccination Progress Using Machine Learning Approach

Shawni Dutta[1], Upasana Mukherjee[1] and Samir Kumar Bandyopadhyay[2*]

[1]Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India.
[2]The Bhawanipur Education Society College, Kolkata, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2021/v33i38A32076
Editor(s):
(1) Dr. Giuseppe Murdaca, University of Genoa, Italy.
(2) Dr. P. Veera Muthumari, V. V. Vanniaperumal College for Women, India.
(3) Dr. Sung-Kun Kim, Northeastern State University, USA.
Reviewers:
(1) S. Selvakanmani, Velammal Institute of Technology, India.
(2) Md Ismail Hossen, American International University-Bangladesh (AIUB), Bangladesh.
(3) Noor A. Ibraheem, University of Baghdad, Iraq.
(4) María Vanessa Villasana, Centro Hospitalar do Baixo Vouga, Portugal.
(5) Jin Zhang, Hunan Normal University, China.
Complete Peer review History: https://www.sdiarticle4.com/review-history/70272

Original Research Article

Received 19 May 2021
Accepted 18 July 2021
Published 24 July 2021

ABSTRACT

The novel coronavirus disease (COVID-19) has created immense threats to public health on various levels around the globe. The unpredictable outbreak of this disease and the pandemic situation are causing severe depression, anxiety and other mental as physical health related problems among the human beings. This deadly disease has put social, economic condition of the entire world into an enormous challenge. To combat against this disease, vaccination is essential as it will boost the immune system of human beings while being in the contact with the infected people. The vaccination process is thus necessary to confront the outbreak of COVID-19. The worldwide vaccination progress should be tracked to identify how fast the entire economic as well as social life will be stabilized. The monitor of the vaccination progress, a machine learning based Regressor model is approached in this study. This vaccination tracking process has been applied on the data starting from 14th December, 2020 to 24th April, 2021. A couple of ensemble based machine learning Regressor models such as Random Forest, Extra Trees, Gradient Boosting, AdaBoost and Extreme Gradient Boosting are implemented and their predictive performance are

**Corresponding author: E-mail: 1954samir@gmail.com;*

compared. The comparative study reveals that the Extra trees Regressor outperforms with minimized mean absolute error (MAE) of 6.465 and root mean squared error (RMSE) of 8.127. The uniqueness of this study relies on assessing as well as predicting vaccination intake progress by utilizing automated process offered by machine learning techniques. The innovative idea of the method is that the vaccination process and their priority are considered in the paper. Among several existing machine learning approaches, the ensemble based learning paradigms are employed in this study so that improved prediction efficiency can be delivered.

Keywords: COVID-19; vaccine; prediction; regression; ensemble learning; AdaBoost.

1. INTRODUCTION

A virus is not new to the world. For example, chicken pox affects many people but not as COVID-19. A virus is a tiny infectious particle. It can reproduce only by infecting a host cell. Viruses have some important features in common with cell-based life. They have nucleic acid genomes based on the same genetic code that is used in cells. Both make human sick but bacteria and viruses are very different at the biological level. Bacteria are small and single-celled. They are living organisms that do not depend on a host cell to reproduce. So, bacterial and viral infections are treated very differently. Application of antibiotics is useful against bacteria. However, to fight against viruses, vaccination should be applied. The most common type of viral disease is the common cold. It is caused by a viral infection of the upper respiratory tract. Chickenpox, Flu (influenza), etc. are also caused due to virus. Most of the viruses are regional but virus due to COVID-19 affects the entire world and it is a new type of virus. Scientists are trying to find out the nature of the virus but there are no in-depth suggestions available regarding the actual source of this virus. Scientists prepare vaccines but they are not in a position to assure people in the globe with drugs. Pharmacists are often amongst the most accessible and the most trusted healthcare professionals. Pharmacists can play an important role in disease prevention by advocating and administering immunizations in terms of vaccines.

The novel coronavirus is the deadliest virus that eventually caused a global pandemic claiming the lives of many. So far it has been proven that the epidemic started in the Wuhan province of China in December 2019 [1]. The virus that causes COVID-19 is known as “Severe Acute Respiratory Syndrome CoronaVirus-2” or SARS-CoV-2. Corona is a family of viruses that result in diseases like simple coughs and colds that are caused due to “Middle East Respiratory Syndrome CoronaVirus” or MERS-CoV and SARS-CoV [2]. The number of affected people by this virus is drastically increasing day by day. So just wearing masks, frequently cleaning hands and staying 1 meter apart from others is not enough to reduce our chance of being exposed to the virus or spreading it to others. The utility of vaccines is to boost up our immune system to fight against the disease if we are exposed. The vaccination process has begun in almost every affected county and territories. A safe, efficient and large scale vaccination process against Covid-19 is the only possible way out to eradicate the ongoing pandemic that has directed to severe social as well as economic turmoil globally. Despite of the present pandemic crisis, the world is inevitable to develop Covid-19 vaccines within an unprecedented short amount of time [3]. According to WHO, “The COVID-19 vaccines are safe for most people 18 years and older, including those with pre-existing conditions of any kind, including auto-immune disorders. These conditions include hypertension, diabetes, asthma, pulmonary, liver and kidney disease, as well as chronic infections that are stable and controlled” [4].

The current study has focused to predict the vaccination progress worldwide. This task is necessary because it will assist in having a forward-looking outlook on the various business industries. It is expected that taking vaccines will build the immunity among the people and they will get back to their normal lives. Analysing the vaccination data will benefit the industries, businesses, and market to contemplate more reasoning, precision, intelligence regarding their business strategies. Investigation of the vaccination progress will help the states or the countries to have a better understanding that how many vaccines have been administered. Based on this, the government can enlighten well-versed decision regarding the future vaccine supply to the community. For carrying out the vaccination uptake and coverage prediction, an

automated tool needs to be constructed. This automated tool will learn and get experience from the data without being explicitly programmed. This automated system prerequisites machine learning techniques for being implemented. Machine learning is an application of artificial intelligence (AI) that can be used in various fields such as image recognition, online fraud detection, medical diagnosis and many more [5]. When a machine learning method is trained on a labelled dataset it is known as supervised machine learning. This further can be classified into classification and regression methods. Regression techniques are used to map a data item to a real value prediction variable. The main goal of the regression techniques is to plot the best-fit line or a curve between the data.

In this paper, the vaccination coverage around the world is predicted which is essentially real value data. Hence, the Regressor model is implemented in this paper for carrying out the objective of this study. For implementing the Regressor model, ensemble machine learning techniques [6] are utilized. In ensemble based learning, multiple learners are assembled as a single entity for improvising the predictive power. This study implements a couple of ensemble based models namely Random Forest [7], Extra Trees [8], Gradient Boosting [9], AdaBoost [10] and XGB [11] Regressors. The performance of the regressor is determined based on Mean Absolute Error and Root Mean Squared Error.

The presented work is different from other Covid-19 related study as it has contributed in analysing vaccination uptake progress among world-wide people. The analysis has been carried out by applying machine learning based techniques that automatically gathers knowledge from prior experience. The primary objective of this study is to develop an automated tool that will predict the daily world-wide vaccination progress. To fulfil the objective of this presented study, a multi-step procedure is employed which can be summarized as follows-

1. To collect the data within the range between 14th December, 2020 and 24th April, 2021 from Kaggle repository [12].
2. Apply pre-processing methods on the collected data for retrieving a balanced data.
3. Transform the dataset in such a manner dedicated for tracking daily vaccination rate around the globe.
4. Employ ensemble based Regressor model to predict the daily vaccination progress Random Forest, Extra Trees, Gradient Boosting, AdaBoost and XGB Regressors.
5. Identify the best performing model that will provide close prediction with respect to the actual data.

2. RELATED WORKS

Various research works have been conduction while keeping an eye to Covid-19. This section gives brief discussion regarding those studies based on Covid-19. Zoabi Y. et al [13] have approached a machine learning method to predict test results of covid-19 with high accuracy. The trained set contains 51,831 tested persons and the test set contains 47,401 tested individuals. By using only eight binary features such as, sex, age ≥60 years, known contact with an infected individual and the appearance of five initial clinical symptoms from the dataset prediction is done. The model predicted with 0.09 auROC (area under the receiver operating characteristic curve) with 95% CI: 0.892–0.905 and by using predictions from the test set, the possible working points are: 87.30% sensitivity and71.98% specificity or 85.76% sensitivity and 79.18% specificity. The dataset is from the Israeli Ministry of Health [13]. Another study [14] has implemented machine learning models namely Polynomial Regression and Support Vector Machine Algorithm in the real-life dataset from the Ministry of Health and Family Welfare of India. A detailed analysis has been shown by this study regarding the trend of the transmission of Covid-19 in the world and also done a study on the spread of the virus outbreak situation in India. Experimental result has implied Polynomial Regression shows better accuracy (93% approximate) than SVM for predicting the rise of cases in the next 60 days. As a concluding statement, it was mentioned m that in the future deep learning models or hybrid models can be used to forecast the spread of the virus [14]. A machine learning as well as deep learning based epidemic analysis is observed in [15]. For forecasting Covid-19 transmission, Support Vector Regression, Deep Neural Network, Long Short Term Memory and Polynomial Regression models are used. Minimum Root Mean Square Error score is obtained by Polynomial Regression model as compared to other employed model by this study [15].

A comparative study was conducted for predicting the COVID-19 outbreak to an

alternative to susceptible–infected–recovered (SIR) and susceptible-exposed-infectious-removed (SEIR) models using machine learning and soft computing models. Implementations of these models were applied on the data collected from the Worldometer website. From the various models, two models namely multi-layered perceptron (MLP) and adaptive network-based fuzzy inference system (ANFIS) have exhibited the promising results. This study has indicated an initial benchmarking to demonstrate the potential of machine learning for future research and further suggested that a genuine novelty in outbreak prediction can be realized by integrating machine learning and SEIR models [16]. Tuli S. et al [17] have proposed a model using machine learning and cloud computing to predict the growth and trend of COVID-19. They have used the data set "Our World in Data" by Hannah Ritchie. In their paper, it is observed that an iterative weighting which is for the fitting of Generalized Inverse Weibull Distribution gives a better fit to the prediction model than the baseline Gaussian model. They have done this in a cloud computing platform which gives them better accuracy and real-time predictions [17].

Some studies have also presented survey based analysis for vaccination data. A survey based research has been presented to assess the willingness of Covid-19 vaccine intake among people. The authors [18] have found that 65% of respondents are willing and 27% are not deciding whether to take a COVID-19 vaccine or not if it becomes available from the dataset, Time 4 (COVID wave) of The Values Project. The reduced Rank Regression (RRR) technique is used in their model. It is concluded by them that further research is essential regarding the attitudes, beliefs, and potential mind-changing factors for those who are not able to decide whether to take the COVID-19 vaccine or not [18]. A survey has been conducted based on the data from 16th June to 20th June 2020, 13426 persons from 19 countries for determining the acceptance rate of the COVID-19 Vaccine [19]. For this, the authors have checked the factors which are influencing the acceptance of the vaccine. In the result, they saw that the highest rate is in China (almost 90%) and the lowest rate is in Russia (almost 55%).In the total dataset, 71.5% responded that they will take the vaccine if it is proven safe and 48% responded that they will take the Vaccine if their employer recommended it [19]. Another survey based study was conducted in the UK on April 2020, where 1653 residents were invited and 1194 were present. The authors have done a regression analysis of the individual and public health factor on behaviour towards COVID-19 Vaccination. Their result showed that 85% responded and 55% of vaccine sceptics were willing to get vaccinated. It was seen that those persons and their family have a higher risk for Covid-19 they are more willing to get vaccinated [20]. Charles F.Manski [21] has reduced the assumptions to acknowledge the various uncertainties for the purposes of evaluation to the policies against the Covid-19 Vaccination in his paper. The uncertainties are not only the effect of vaccination on the disease but also the health risks after vaccination. He mainly focuses on planning to build an algorithm that will determine the various policies of vaccines and the types of uncertainties [21]. A proposal to end the Covid-19 Pandemic is given by Ruchir Agarwal and Ms. Gita Gopinath in their paper [22]. Their proposal targets three points. The first target is to vaccinate a minimum of 40 per cent of the population by end of 2021 and by the half of 2022 at least 60 percent of them have to be vaccinated. The second target is to check the side effects and solve them and the last target is to maintain the public health measures and the stocks of therapeutics [22]. Ferranna. M et al [23] were focused on the Covid-19 Vaccination prioritization strategies and their goals in their paper. One of the employed strategies states that that if the essential workers are prioritized then the number of cases and life losses can be reduced. They also mentioned that by prioritizing the adults, a huge amount of deaths can be reduced [23]. According to Mare Lipstich and Natalie E. Dean [24] if the Covid-19 vaccine offers less protection in high-risk groups still can reduce the infection in younger adults. They also told the worst case that is the vaccine may not provide direct or indirect protection in case of high-risk groups [24].

From the aforementioned analysis, it can be easily implied that a wide range of researches have been conducted with the primary concern of Covid-19 disease. Some studies have presented on the vaccination intake possibilities observed among several people. This study has contributed in determining the global vaccination progress by a means of ensemble based machine learning approaches. Based on our knowledge, there is no such paper based on vaccination progress using Machine Learning Approach.

3. BACKGROUND

Regression is used to map a data item to a real-valued prediction variable. Regression assumes that the target data fit into some known type of function of this type that models the given dataset [21]. In this study, ensemble learning based regression methods are favoured. Ensemble learning methods basically combine the prediction of two or more methods. These models are robust in nature and can give better predictive skills than a single model [6]. There are three types of ensemble algorithms, which are Bagging ensemble, Boosting ensemble and Stacking ensemble. However, this research uses only bagging and boosting ensemble techniques.

3.1 Bagging Ensemble

The bagging ensemble method is a parallel ensemble method and also known as Bootstrap Aggregation. By this variance, of the prediction model decreases because this technique generates extra data in the training stage by random sampling with the replacement of the actual dataset. This training set data is used to train multiple models. So it develops an ensemble of different models and the average of every prediction from the different model is considered in the case of regression. These models are more robust than single models. Based on this technique, the Random Forest, Extra Trees regressors are employed in this work.

3.1.1 Random forest regressor

A Random Forest algorithm is a supervised machine learning algorithm. For both Regression and Classification problem: this algorithm can be used. It follows the concept of Ensemble Learning. Random Forest algorithm combines multiple decision trees and a technique called Boosting and Aggregation (also known as Bagging). The decision trees are the base learning models and random row sampling and feature sampling in the dataset is known as Bootstrap. The number of trees in the forest and the accuracy of the model is proportional to each other. Basically, instead of relying on a single decision tree, Random Forest takes the prediction from each tree and based on the majority votes of predictions and predicts the final output [7].

3.1.2 Extra trees regressor

The Extra Trees or Extremely Randomized Trees is an ensemble based approach which is highly scalable and efficient in nature. This approach proceeds by assembling the output of multiple de-correlated decision trees where each of these trees is learned on a training set. The collection of the base learners forms a “forest”. The acquired outcomes from these learners predictions are obtained by evaluating the mean prediction value as in case of regression problem. In the classification approach, predictions from all these learners are combined by employing majority voting techniques. While comparing this approach to bagging and random forest, minor difference can be perceived. Bagging and random forest construct each decision from a bootstrap instance of a training dataset whereas, the Extra Trees algorithm fits each decision tree on the whole training dataset [8]. Both Random forest and Extra Trees algorithm samples the feature points in a random order at each decision tree split point.

3.2 Boosting Ensemble

The Boosting technique is a sequential ensemble method that combines different types of prediction. By decreasing the bias error it generates a strong predictive model. Like the Bagging method, this method also generates multiple training datasets by doing random sampling. This technique gets multiple weak learners and by combining them, converts to a strong learner. During the training session, this method iteratively adjusts the weights to each resulting model. Learners with good prediction results assigned higher weights compared to those learners whose predictions are comparatively poor. Under this method, the Regressor models namely Gradient Boosting, AdaBoost are selected for implementation.

3.2.1 Gradient boosting regressor

Gradient Boosting Regression (GBR) is an extremely powerful technique for predictive models. It is also an ensemble machine learning algorithm and can be used for both classification and regression problems GBR involves three elements -a loss function (which needs to be optimized), a weak learner (used for making predictions) and an additive model (to add weak learners to minimize the loss function). This algorithm is used to construct the new base-learners to be maximally correlated with the

negative gradient of the loss function, associated with the whole ensemble. The loss function is for better intuition. If the error function is the classic squared-error loss then the learning procedure would result in consecutive error-fitting. The choice of the loss function is up to the researcher [9].

3.2.2 AdaBoost or adaptive boosting regressor

The first boosting ensemble algorithm is the Adaptive Boosting or AdaBoost and it was introduced in 1996 by Freund and Schapire [11]. This technique is also known as a meta-estimator. The AdaBoost method is built by combining weak learners like the decision trees and by assigning weights to the wrong values. It is used for both classification and regression problems but, mainly it was built for binary classification problems and for boosting the performance of the decision trees. In the case of less noisy datasets, this technique is robust to over-fitting. This algorithm also exhibits the feature of the self-averaging property. Here the parameter adjustment is not needed; this method automatically does its parameter adjustment on the data. Applying this regressor one can get a low generalization error.

3.2.3 XGB or extreme gradient boosting regressor

This algorithm provides a more efficient and effective implementation of the Gradient Boosting algorithm. XGBoost has the base learner that is constantly bad at the remainder so when every prediction is combined the bad ones are cancelled out. As a result, the rest of the good ones are summed up and form good predictions. The nature of XGBoost algorithm is robust enough not to be over fitting with increasing trees. A high value for a particular learning rate could degrade its ability in predicting new test data. Since it reduces the learning rate and increases the number of trees, the computation becomes expensive and take longer time in processing through computer [11].

3.3 Regressor Model Evaluation

The predictive performance of any Regressor model is evaluated based on the difference between the estimated prediction and the actual data. Two popular metrics such as Mean Absolute Error (MAE) [25] and Root Mean Squared Error (RMSE) [26] those are used to assess the efficiency of Regressor models. MAE is one of the loss functions which is used for the regression models. It is the sum of the absolute value of the difference between actual and predicted values and it is shown in equation (1). A perfect prediction is obtained when the value MAE is 0. RMSE is the square root of the Mean Squared Error (MSE) and it can be mathematically expressed as (2). MSE is the squared difference between actual and predicted values. Lower the value of RMSE, the higher the efficiency of the model.

$$\text{MAE} = \frac{1}{n}\sum_{i=0}^{n}|\text{Yi} - \text{Ypredi}| \quad (1)$$

$$\text{RMSE} = \frac{1}{n}\sum_{i=0}^{n}\sqrt{(\text{Yi} - \text{Ypredi})^2} \quad (2)$$

Here n is the number of training set instances, Y_i is the Actual Values and Y_{predi} is Predicted Values.

4. DATASET DESCRIPTION

This paper collects the vaccination progress data from kaggle repository [12]. The dataset contains total 13600 number of vaccination records which covers the data related to several countries such as United States, China, United Kingdom, England, India and many more. The dataset has vaccination information starting from 14th December, 2020 to 24th April, 2021. The daily vaccination rate around the entire globe is calculated within the aforementioned time span and is shown in Fig. 1. As shown in the Fig. 1, the daily vaccination rate has increased as the time elapses. Fig. 2 shows the vaccination progress of each country.

5. METHODOLOGY

This study herein will analyse the overall vaccination progress as the regression problem. In this paper, this regression problem is analysed using machine learning techniques. The use of machine learning techniques provides a favourable way to carryout prediction by a means of data understanding. In order to monitor the vaccination advancement, the machine learning based regression technique is favoured where the data labels are numeric. To accomplish the aforementioned task, a multi-step procedure is approached in the presented study. The collected dataset contains vaccination records among several countries. Instead of focusing on each country separately, overall vaccination progress throughout the globe has been

considered in this study. The collected dataset has been transformed by computing the daily vaccination rates around the world. The transformed data is partitioned into two parts. One part contains the daily vaccinated data ranging from 14th December, 2020 to 21st January, 2021 and the other part contains vaccination data from 22nd January,2021 to 24th April, 2021. The first part is used as training data for building the Regressor model whereas the second part is utilized as the testing data. The training set is utilized for establishing the regressor model and the testing set is for evaluating the model.

This research has approached to use both bagging and boosting based methods which are essentially ensemble based methods. The employed ensemble based machine learning models such as Random Forest Regressor, Extra Trees Regressor, Gradient Boosting Regressor, AdaBoost Regressor and XGB Regressor are implemented for retrieving the daily vaccination rate prediction. The Random Forest Regressor and Extra Trees Regressor are bagged ensemble based classifiers whereas Gradient Boosting Regressor, AdaBoost Regressor and XGB Regressor are based on boosting technique. These predictive models require a training set for acquiring the pattern identification and later that acquired knowledge is exploited using the testing dataset. All the mentioned models are implemented using the Scikit-Learn machine learning library in Python. During implementation, all these models have gone through several hyper parameters identification. The Random Forest Regressor has been constructed with 200 numbers of estimators (count of trees in the forest) whereas Adaboost Regressor, Gradient Boosting Regressor, Extra

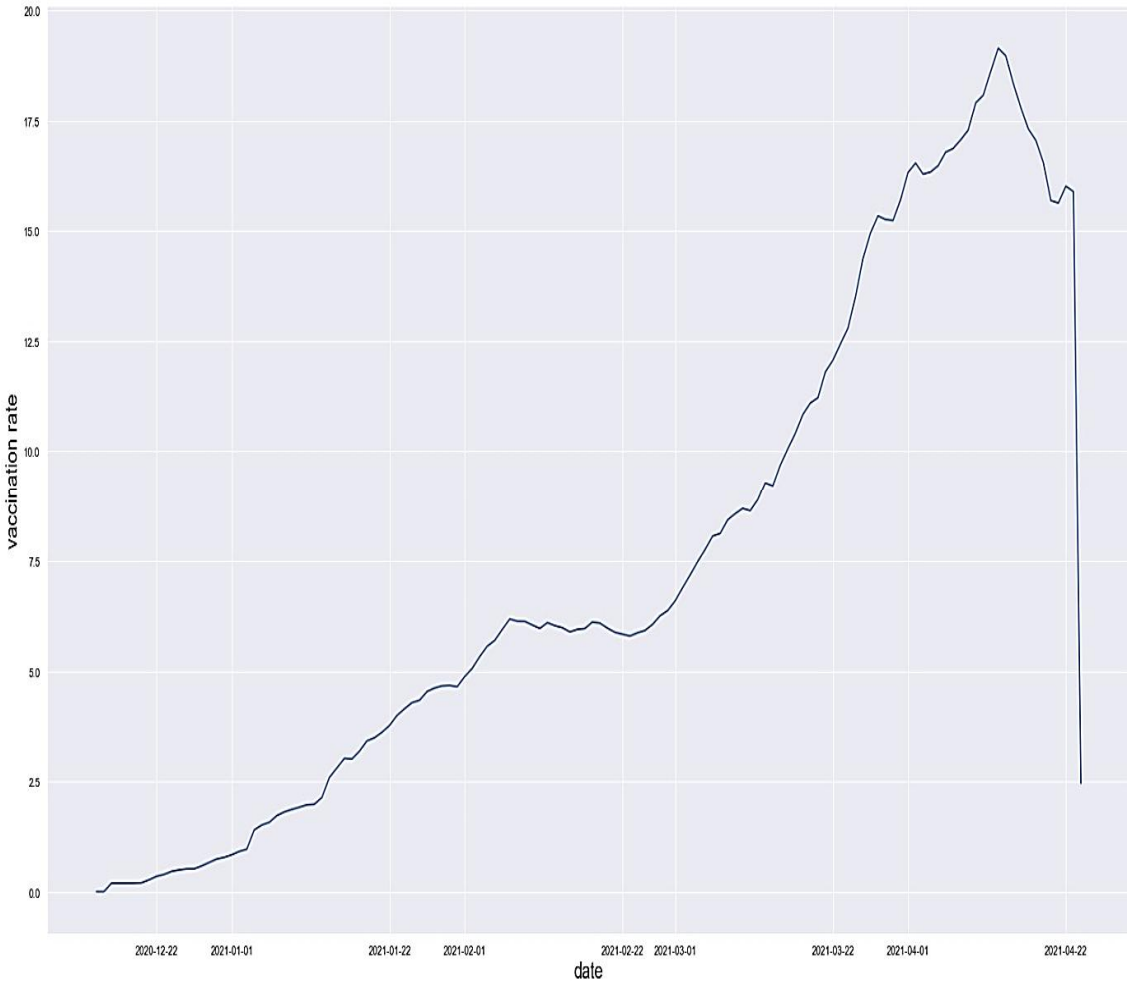

Fig. 1. Daily vaccination progress all over the world

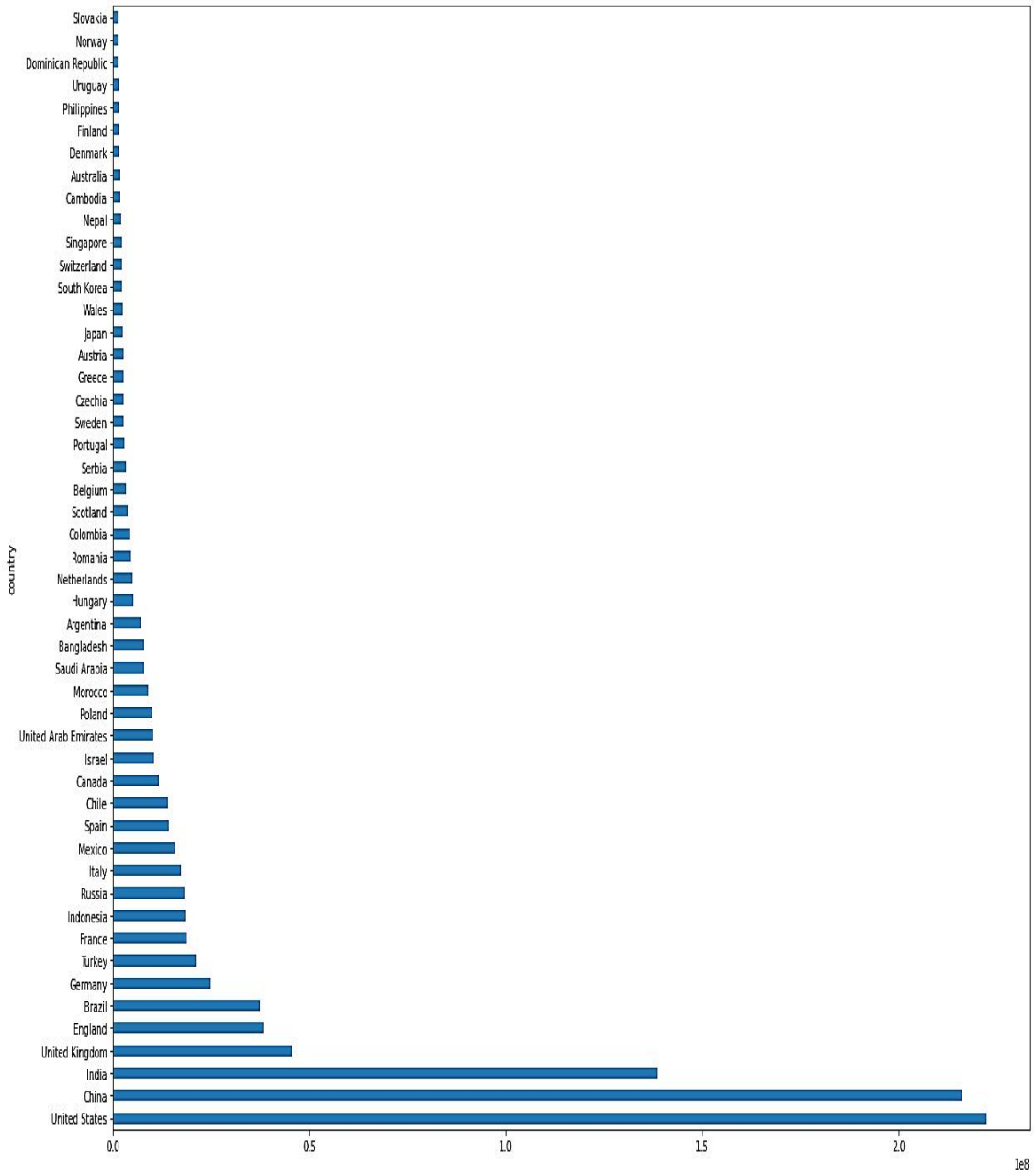

Fig. 2. Vaccination rate for each country

Trees Regressor, XGB Regressor receives a total of 50, 100, 100, 100 estimators respectively. These parameters have shown a promising outcome while considering the predictive result. The implemented models are evaluated using two metrics such as MAE and RMSE. The adjustment process of each Regressor model is specified in section 5.2. Regressor model that exhibits minimized MAE and RMSE is considered as the best predictive model. A complete multi-step procedure employed in this study is shown in the Algorithm 5.1. The block-diagram of the employed methodology is shown in Fig. 3.

5.1 Algorithm

Step 1: Dataset collection
Step 2: Transform the collected dataset by the daily vaccination rate around the world.

Step 3: Partition the transformed data into two parts. The first subset has the vaccination data from 14th December, 2020 to 21st January, 2021 and the other part contains vaccination data from 22nd January,2021 to 24th April, 2021.
Step 4: Use the first portion of the data for training purposes and the other portion for testing purposes.
Step 5: Then build Regressor models namely Random Forest Regressor, Extra Trees Regressor, Gradient Boosting Regressor, AdaBoost Regressor and XGB Regressor one by one using correct hyper-parameter adjustment.
Step 6: Use the training subset to acquire the l and test them with the testing data.
Step 7: Visualize the predicted result for each model.
Step 8: Evaluate MAE and RMSE for each model and compare them to find out the most efficient model.

5.2 Implementation Details

Application of ensemble based methods is often beneficial as it assembles the prediction outcomes from weak learners. However, the underlying parameters of any machine learning parameters need to be fine-tuned in order to obtain the enhanced predictive analytics. To accomplish this hyper-parameter adjustment, the number of estimators is varied for each ensemble models to find the best possible outcome. The concept of estimators is closely associated with the number of base learners. The prediction results acquired by each base learner will be assembled for attaining the most efficient prediction. Hence, it is necessary to check how these employed models are performing when the estimator count is varying. Following specification of estimator is considered while implementing each model.

- For random forest model, the estimator count has been varied within the range of 1 to 500 with an equal interval of 50. The performance for each underlying estimator is measured in terms of MAE. The relevance of different estimator counts for this ensemble model is depicted in Fig. 4. In this case, the estimator count of 50 shows the lowest error rate.
- The Gradient Boosting model receives the estimator count between 1 and 500 with a step size of 50 and for each case the performance is evaluated against MAE. The prediction performance for various estimator values is shown in Fig. 5. The estimator count of 100 provides the optimized MAE value for this underlying model.
- The AdaBoost and XGB Regressor model accepts the same estimator values between 1 and 500 and the performance is shown in Figs. 6 and 7 respectively. The estimator value of 150 and 100 shows the enhanced prediction outcome for AdaBoost and XGB Regressor model respectively.
- The extra tree classifier model is being trained for numerous times with the estimator count within the range of 1 to 500. But same prediction result is being observed in Fig. 8 for the specified estimator size. Hence, estimator value of 100 is being chosen at random.

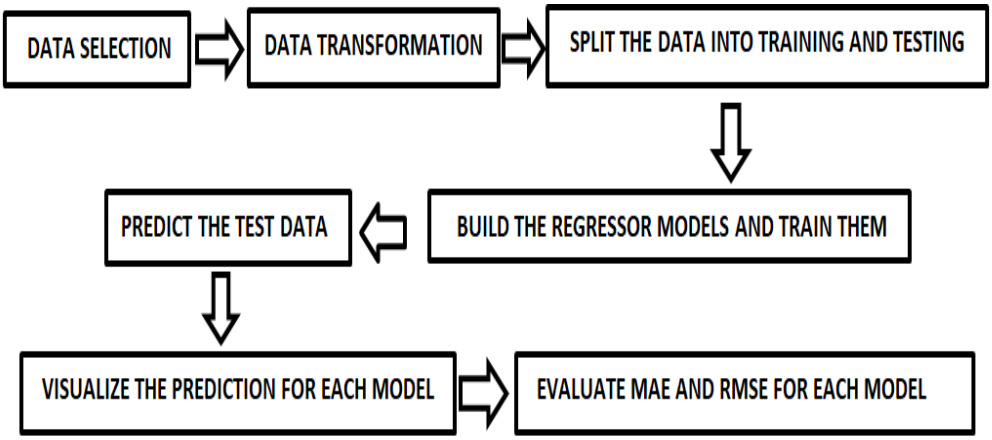

Fig. 3. Block diagram of proposed methodology

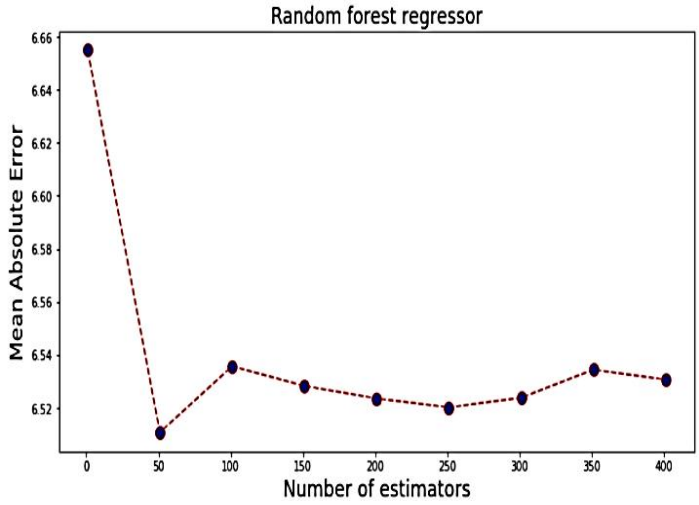

Fig. 4. Comparing prediction performance of random forest based on different estimator count

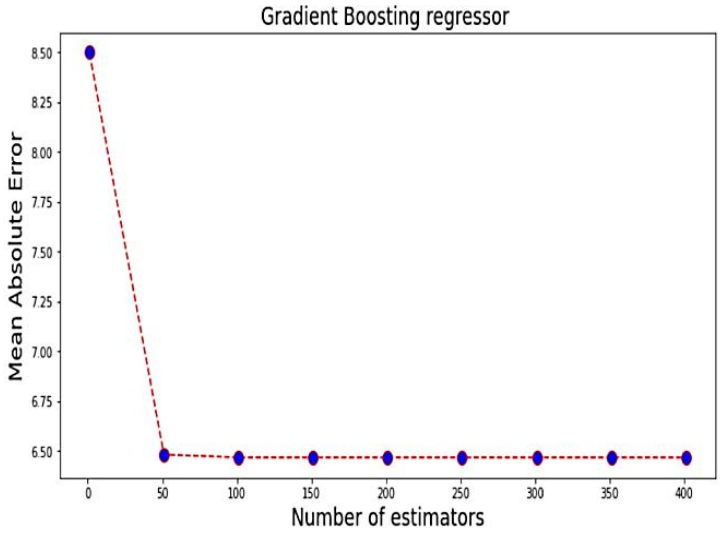

Fig. 5. Comparing prediction performance of gradient Boosting based on different estimator count

6. EXPERIMENTAL RESULTS AND DISCUSSION

The daily vaccination rate is being predicted in this study. For this purpose, all the employed Regressor models such as Random Forest Regressor, Extra Trees Regressor, Gradient Boosting Regressor, AdaBoost Regressor and XGB Regressor are implemented by adjusting the necessary parameters. Adjustment of these parameters will help the models to produce the best predictive results. All the employed models are evaluated by using the testing data. The testing data contains the vaccination records from 22nd January, 2021 to 24th April, 2021 for all over the world. Now, the actual vaccination data during the aforementioned period and the prediction data acquired from each ensemble based model are compared with respect to existing predefined metrics such as MAE and RMSE. These metrics will assist in understanding the closeness of actual and prediction values. The visualization of the actual and predicted vaccination rate obtained by each Regressor model namely Random Forest Regressor, Extra Trees Regressor, Gradient Boosting Regressor, AdaBoost Regressor and XGB Regressor is depicted in Figs. 9,10,11,12, and 13 respectively. The predictive efficiency of each of these models is evaluated against MAE

and RMSE. Table 1 shows the comparative study among these machine learning models. The comparative study implies the following observations-

- The Random Forest Regressor has obtained an MAE of 6.523 and an RMSE of 8.174.
- The Extra Trees Regressor shows the MAE of 6.465 and RMSE of 8.127.
- The Gradient Boost Regressor has reached the MAE of 6.466 and RMSE of 8.128.
- An MAE of 6.560 and an RMSE of 8.205 have been reached by AdaBoost.
- The XGB Regressor has given the MAE of 6.476 and RMSE of 8.137

Now, the model which will predict the vaccination progress with minimized MAE and RMSE value will be regarded as the best Regressor. From the above mentioned discussion, it can be observed that Extra Trees Regressor model exhibits the minimized MAE of 6.465 and RMSE of 8.127 for predicting the vaccination rate. Hence, this model may be regarded as the best predictive model.

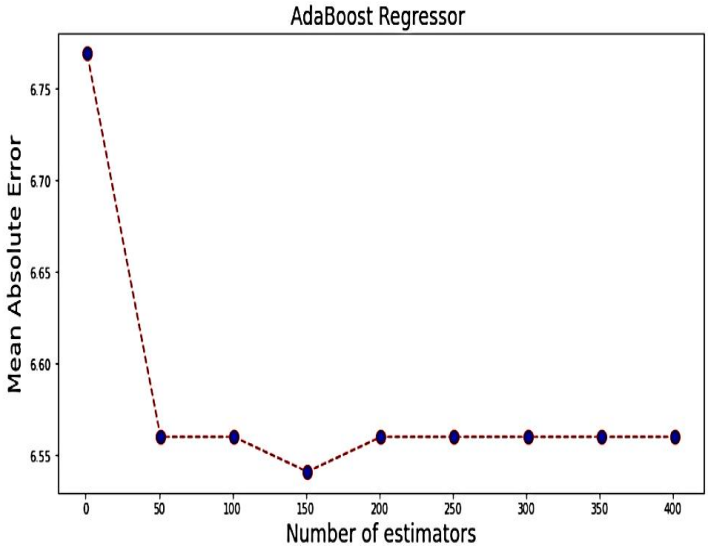

Fig. 6. Comparing prediction performance of AdaBoost based on different estimator count

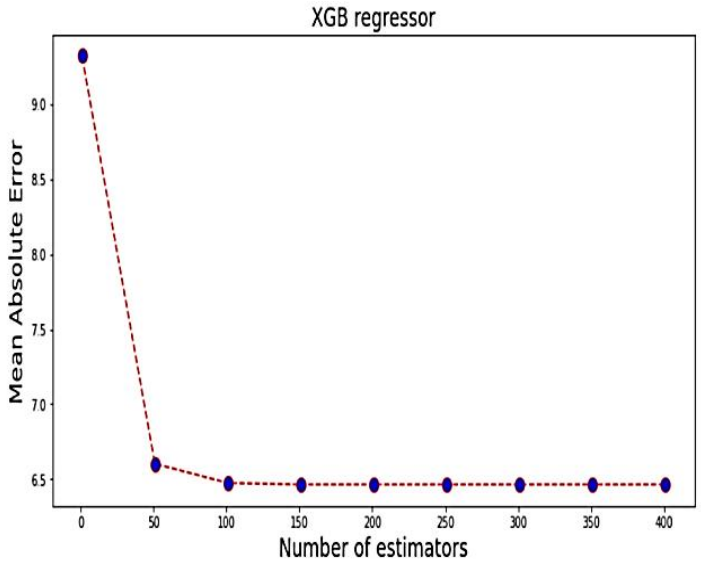

Fig. 7. Comparing prediction performance of XGB model based on different estimator count

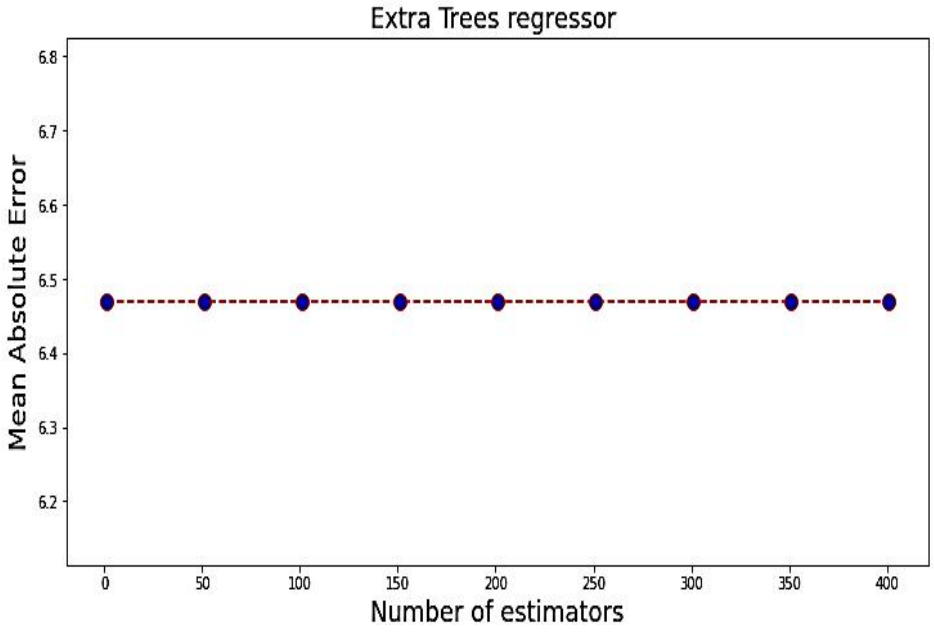

Fig. 8. Comparing prediction performance of Extra Trees model based on different estimator count

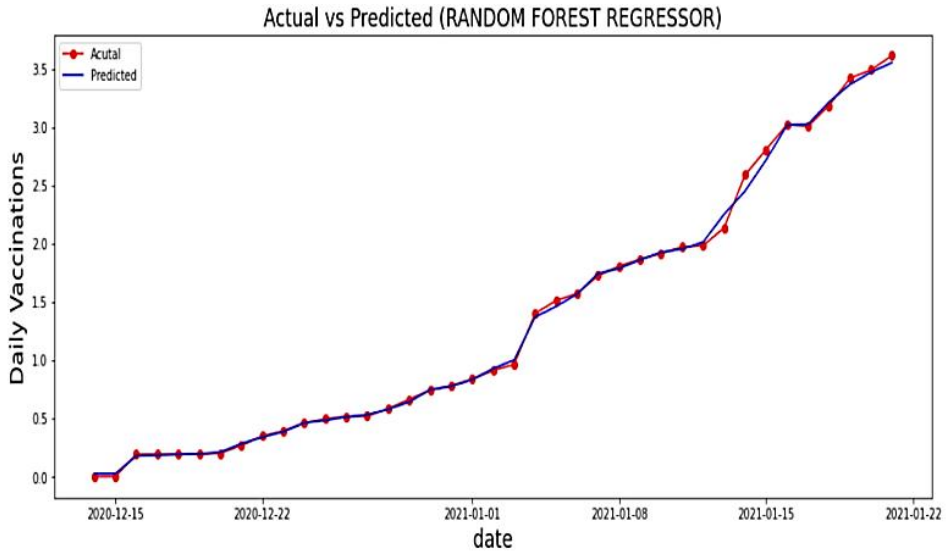

Fig. 9. Visualize actual and predicted daily vaccination rate using random forest

Table 1. Comparative analysis of the implemented regressor models in terms of MAE and RMSE

Regressor Name	MAE	RMSE
Random Forest Regressor	6.523	8.174
Extra Trees Regressor	6.465	8.127
Gradient Boosting Regressor	6.466	8.128
AdaBoost Regressor	6.560	8.205
XGB Regressor	6.476	8.137

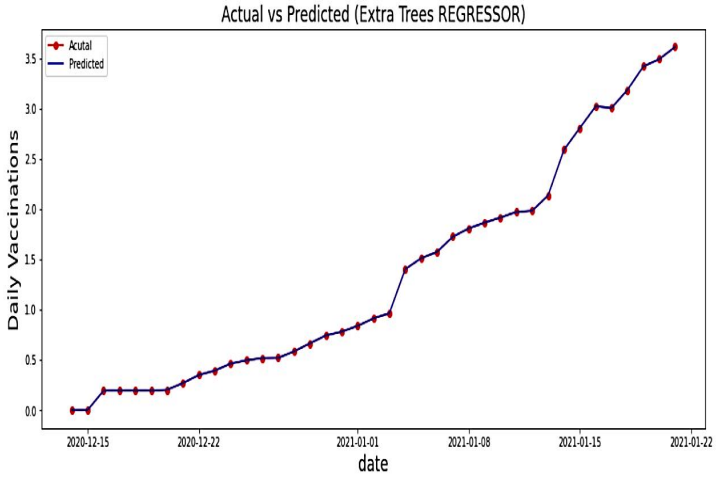

Fig. 10. Visualize actual and predicted daily vaccination rate using extra trees

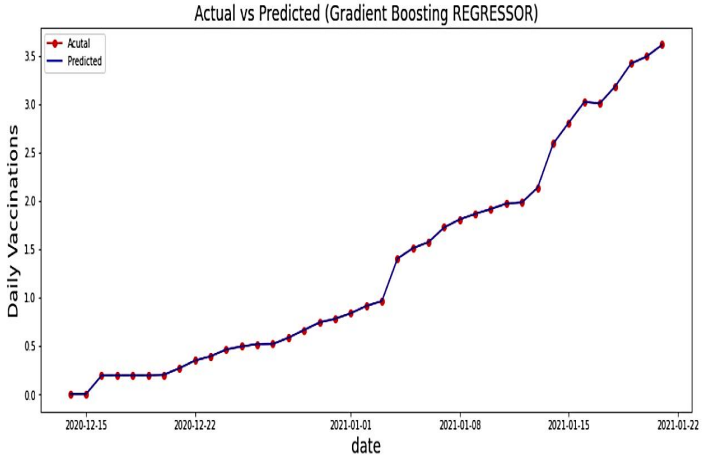

Fig. 11. Visualize actual and predicted daily vaccination rate using gradient boosting

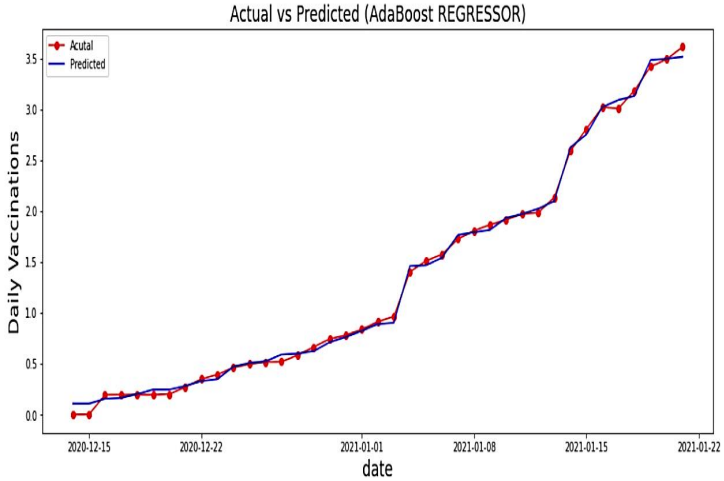

Fig. 12. Visualize actual and predicted daily vaccination rate using AdaBoost

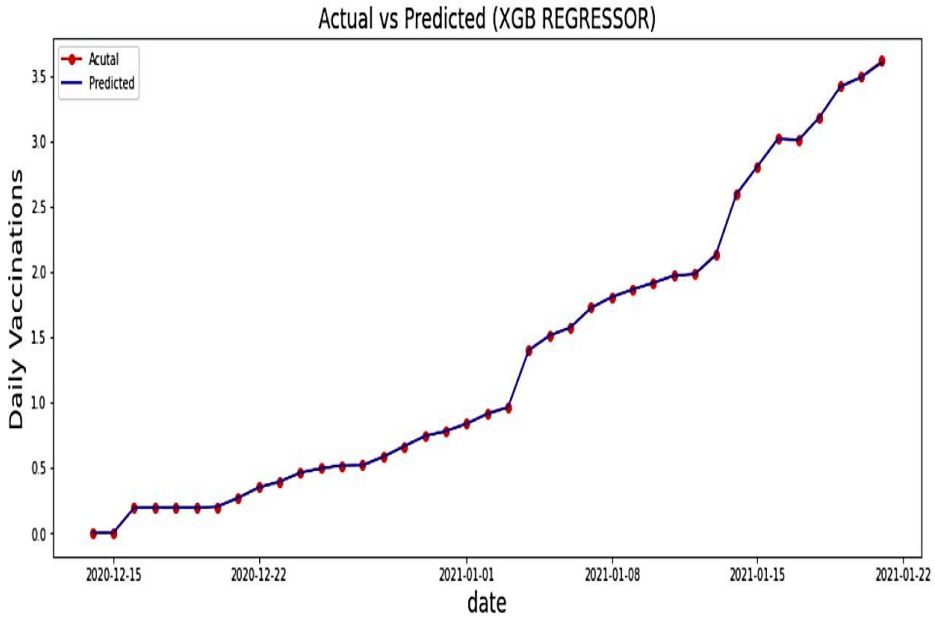

Fig. 13. Visualize actual and predicted daily vaccination rate using XGB

7. CONCLUSION

In order to prevent the highly infectious disease COVID-19, safe and effective vaccination process is compulsory. The vaccination process can eradicate the fear of being infected while being exposed with other people. The vaccination will also ensure the normalization of other businesses, industries and markets, thus by improving the social and economic status of the countries. Thus, it is necessary to monitor the vaccination rate so that government can take a supplementary decision for vaccination production and delivery. Early prediction of the vaccination process will also assist the government to find out the areas where less vaccinated people are there. This study has focused to construct an intelligent model that will follow a predictive analysis for daily vaccination rate assessment around the world. The data-driven model has been made using regression techniques. Use of ensemble based methods is approached in this paper so that improved efficiency can be obtained. The intelligent data-driven model can be implemented by using the Extra Trees algorithm as it exhibits the best possible results of minimized MAE of 6.465 and RMSE of 8.127.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Huang Chaolin, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 2020;395(10223):497-506.
DOI: https://doi.org/10.1016/S0140-6736(20)30183-5
2. Wu Peng, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Eurosurveillance. 2020; 25(3):2000044.
DOI: 10.2807/1560-7917.ES.2020.25.3.2000044
3. Vashishtha, Vipin M, Puneet Kumar. Emergency use authorisation of Covid-19 vaccines: An ethical conundrum. Indian Journal of Medical Ethics. 2021;1:1-3.
DOI: 10.20529/IJME.2020.122
4. Sarah Schaffer, et al. Planning for a COVID-19 Vaccination Program, JAMA. 2020;323(24):2458-2459.
DOI: 10.1001/jama.2020.8711

5. Raffaele Ciof, et al. Artificial intelligence and machine learning applications in smart production: Progress, Trends, and Directions, Sustainability. 2020;12:492.
DOI: 10.3390/su12020492
6. Dietterich, Thomas G. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems. Springer, Berlin, Heidelberg; 2000.
DOI: 10.1007/3-540-45014-9_1
7. Livingston, Frederick. Implementation of Breiman's random forest machine learning algorithm. ECE591Q Machine Learning Journal Paper. 2005;1-13.
DOI: 10.1109/ictai.2007.46Corpus ID: 29812549
8. Geurts, Pierre, Damien Ernst, Louis Wehenkel. Extremely randomized trees. Machine Learning. 2006;63(1):3-42.
DOI: https://doi.org/10.1007/s10994-006-6226-1
9. Natekin, Alexey, Alois Knoll. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics. 2013;(7):21.
DOI: https://doi.org/10.3389/fnbot.2013.00021
10. Schapire, Robert E. Explaining adaboost. Empirical inference. Springer, Berlin, Heidelberg. 2013;37-52. doi:
DOI: 10.1007/978-3-642-41136-6_5
11. Chen, Tianqi, Carlos Guestrin. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on Knowledge Discovery and Data Mining; 2016.
DOI: https://doi.org/10.1145/2939672.2939785
12. Gabriel Preda. COVID-19 World Vaccination Progress Daily and Total Vaccination for COVID-19 in the World from Our World in Data; 2021.
Retrieved on May 25, 2020.
Available:https://www.kaggle.com/gpreda/covid-world-vaccination-progress
13. Zoabi, Yazeed, Shira Deri-Rozov, Noam Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digital Medicine. 2021; 4(1):1-5.
DOI: https://doi.org/10.1038/s41746-020-00372-6
14. Gambhir, Ekta, et al. Regression Analysis of COVID-19 using Machine Learning Algorithms. 2020 International Conference on Smart Electronics and Communication (ICOSEC). IEEE; 2020.
DOI: 10.1109/ICOSEC49089.2020.9215356
15. Punn, Narinder Singh, Sanjay Kumar Sonbhadra, Sonali Agarwal. COVID-19 epidemic analysis using machine learning and deep learning algorithms. MedRxiv; 2020.
DOI: https://doi.org/10.1101/2020.04.08.20057679
16. Ardabili, Sina F, et al. Covid-19 outbreak prediction with machine learning. Algorithms. 2020;13:(10):249.
DOI: https://doi.org/10.3390/a13100249
17. Tuli, Shreshth, et al. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things. 2020;11:100222.
DOI: https://doi.org/10.1016/j.iot.2020.100222
18. Attwell, Katie, et al. Converting the maybes: Crucial for a successful COVID-19 vaccination strategy. PLoS One. 2021; 16(1):e0245907.
DOI: 10.1371/journal.pone.0245907
19. Lazarus Jeffrey V, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nature Medicine. 2021;27(2):225-228.
DOI: https://doi.org/10.1038/s41591-020-1124-9
20. Blanchard-Rohner, Geraldine, et al. Impact of COVID-19 and health system performance on vaccination hesitancy: Evidence from a two-leg representative survey in the UK.
Available at SSRN 3627335 (2020).
21. Manski Charles F. Vaccination planning under uncertainty, with application to covid-19. No. w28446. National Bureau of Economic Research; 2021.
22. Agarwal Ruchir, Gita Gopinath. A proposal to end the COVID-19 pandemic. Staff Discussion Notes. 2021;004.
23. Ferranna, Maddalena, Daniel Cadarette, David E Bloom. COVID-19 Vaccine Allocation: Modeling Health Outcomes and Equity Implications of Alternative Strategies. Engineering; 2021.
24. Lipsitch Marc, Natalie E Dean. Understanding COVID-19 vaccine efficacy.

Science (New York, N.Y.). 2020; 370(6518):763-765.
DOI: 10.1126/science.abe5938

25. Rong Shen, Zhang Bao-wen. The research of regression model in machine learning field. MATEC Web of Conferences. EDP Sciences. 2018;176.

26. Willmott Cort J, Kenji Matsuura. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005; 30(1):79-82.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/70272