International Journal of Early Childhood Special Education (INT-JECSE)

DOI:10.9756/INTJECSE/V14I4.181 ISSN: 1308-5581 Vol 14, Issue 04 2022

PREDICTING THE FUTURE COVID-19 USING SUPERVISED MACHINE LEARNING MODELS

Saïda Shake[1], Ch. Lavanya[2], S. Venkata Maha Lakshmi[3], B. Bhavani[4]
[1]Assistant Professor, Department of Information Technology,
Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India.
[2,3,4] Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh,
Indiasaida518@gmail.com

ABSTRACT:
Prediction mechanisms based on machine learning (ML) have proven their importance in several fields to improve the creation of choices on the long-term course of action. Technological advancements have a quick result in every area of life, be it the medical field or the other field. Many prediction methods are commonly used to handle prediction problems. This study demonstrates the ability of cubic centimeter models to predict the amount of approaching patients with COVID-19 that is currently considered a possible threat to clustering. Specifically, 3 common prediction models, like Extreme Gradient Boosting Machine (XGBM), LIGHT Gradient Boosting Machine (LIGHT GBM) and Random Forest mechanisms are used in this study to predict threatening factors of COVID-19. 3 styles of predictions are created by each of the models, such as the number of new infected cases, the number of deaths, as well as the range of recoveries in the next ten days. The results created by the study prove that it is a promising mechanism to use these means in this COVID-19 pandemic situation.

I. INTRODUCTION

COVID-19 is an associated contagious disease caused by the corona virus. In the healthcare sector, the COVID-19 pandemic is inflicting a global emergency and has had a high impact on human lives. This pandemic has caused the loss of millions and billions of human lives. The eruption of the latest coronavirus (COVID-19) infections has caused global concern as the disease has caused illness, as well as illness resulting in death and sustained human-to-human spread in several countries. Machine learning (ML) has proven to be an exceptional field of study over the past decade by solving several terribly advanced and complicated real-world problems. Various regression and neural network models have great relevance in predicting the conditions of patients in the future with a particular disease.

In particular, the targeted study is focused on the live prediction of confirmed COVID-19 cases and the study is further focused on the prediction of COVID-19 outbreak and early response. These prediction systems are often terribly useful in deciding to handle this situation to guide early interventions to manage these diseases terribly effectively. Therefore, the study relies on state-of-the-art supervised DC regression models such as Extreme Gradient Boosting Machine (XGBM), LIGHT Gradient Boosting Machine (LIGHT GBM). and random forest mechanisms. The performance analysis was drained in terms of the measurements needed as well as mean square error (MSE), mean absolute error (MAE), and mean square error (RMSE).

2.LITERATURE SURVEY

Number of works related to this study, provided few works and inspired too.[1]. The study included, machine learning approach called piecewise linear regression to predict the positive and negative cases of COVID-19. The study mainly focused on Indian states. The work is produced promising result compare to the linear regression. Later , they also planned to apply different ML and DL-based models to improve performance against

DOI:10.9756/INTJECSE/V14I4.181 ISSN: 1308-5581 Vol 14, Issue 04 2022

COVID-19 as well as other pandemics. Also, they provided main challenge about using of the picewise linear regression is identifying of the data points and considered only 7 days data to predict next day.[2] predicted chances of survival and performed data analysis over the real data admitted in the Tongji Hospital from January 10th to February 18th, 2020. Total 375 records . The study used the approach called XGBoost (XGB) and identified 3 features out of 300 features. Focused mainly on 29 patients and produced accuracy of 90%. .[3] study which can focus on classifier called the support vector machine (SVM) and is used for the prediction of COVID-19 patients. The result produced accuracy of 78% .

3. PROPOSED YSTEM

In this study, prediction of the covid-2019 cases done using various methods which include XGBOOST,LightBGM , and Random Forest . The study mainly focused to predict the positive cased for the next days.The complete idea is described in the Figure1.

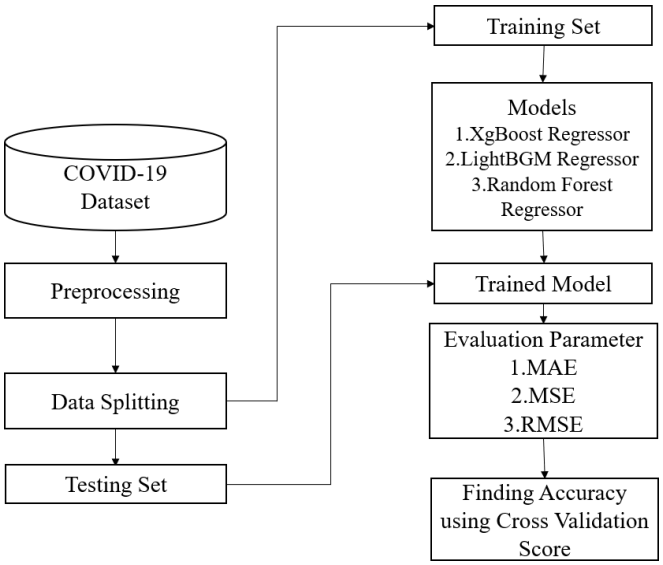

Fig1: System Workflow

From the workflow , it is shown raw data is preprocessing using normalizing ,missing data handling ,and then data is partitioned into 70 and 30 ratio and then it is migrated to the different classifiers. Finally modes are tested against the different benchmarks MAE,MSE and RMSE.

4. METHODOLOGY

The study is about a new coronavirus also called COVID-19 predictions. COVID-19 has tried a gift potential threat to human life. It causes tens of thousands of deaths and therefore the mortality rate is increasing day by day all over the world. To contribute to the current management of the pandemic scenario, this study tries to make future predictions on the mortality rate, to distribute the amount of the quantity of infected cases confirmed daily and therefore the number of cases of recovery within ten days to come. The prediction was made by a victimization of four cubic centimeters approaching this acceptable surface unit in the current context. The dataset used in the study contains daily statistical tables, as well as the number of confirmed cases,

4.1. XGBOOST REGRESSOR

Extreme Gradient Boosting, or XGBoost for short, is an efficient open source implementation of the gradient boosting algorithm. As such, XGBoost is an algorithm,

DOI:10.9756/INTJECSE/V14I4.181 ISSN: 1308-5581 Vol 14, Issue 04 2022

an open source project, and a Python library. It was originally developed byTianqi Chenand was described by Chen andCarlos Guestrinin their 2016 articletitled "System". It is designed to be both computerefficient (e.g. fast to run) and very efficient efficient, possibly more efficient than other open-source implementations. The two main reasons to use XGBoost are execution speed and model performance. XGBoost dominates structured or tabular datasets on classification and predictive regression modeling issues. The proof is that this is the go-to algorithm for contest winners on the competitive data science platform Kaggle.

RANDOM FOREST REGRESSOR

Random Forest a plusieurs arbres de décision comme modèles d'apprentissage de base. Nous effectuons de manière aléatoire un échantillonnage de lignes et un échantillonnage d'entités à partir de l'ensemble de données formant des exemples d'ensembles de données pour chaque modèle. Cette partie s'appelle Bootstrap. Nous devons aborder la technique de régression Random Forest comme toute autre technique d'apprentissage automatique. Concevez une question ou des données spécifiques et obtenez la source pour déterminer les données requises.

LIGHTGBM

LightGBM peut être un cadre d'amplification de gradient pris en charge par des arbres d'appels pour augmenter la puissance du modèle et réduire l'utilisation de la mémoire. Il utilise 2 nouvelles techniques :Gradient-based one-sided sampling and exclusive feature bundling (EFB) that meets the constraints of histogram-based rule which is mainly used in all GBDT (Gradient Boosting Call Tree) frameworks.

RESULTS

The chart related to the next 15 days as the prediction is:

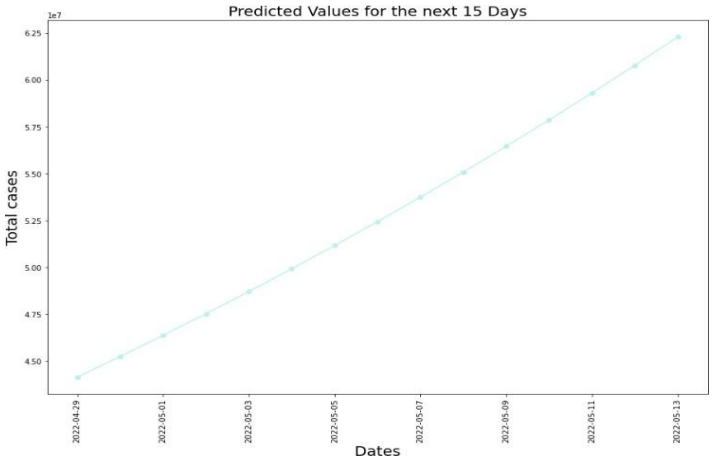

Figure 2: The result of various classifiers Light GBM, XGBM and Random Forest.

From the figure2, it is shown that prediction of 15 days from 29-04 to 13-05-2022 is predicted.

CONCLUSION

The dangerousness of the COVID-19 pandemic will trigger a huge international crisis. Some researchers and government agencies around the world fear that the pandemic will have an effect on a disproportionate proportion of the world's population. During this study, a prediction system based on the associated ML was planned to predict the

DOI:10.9756/INTJECSE/V14I4.181 ISSN: 1308-5581 Vol 14, Issue 04 2022

danger of outbreak of COVID-19 in the world. The system analyzes an information set containing the actual day-to-day past data and makes predictions for the victimization machine learning algorithms of days to come. Overall, we tend to conclude that the predictions of the model in phase with this state of affairs are correct, which can be useful for knowing the future state of affairs. The forecasts of the study can therefore even be of great help for the authorities to demand timely actions and build selections to contain the COVID-19 crisis. This study is going to be infinitely increased in the future course, then we tend to decide to explore the prediction methodology using the updated dataset and use the most correct and applicable mil strategies for the prediction. Live periodic predictions will be one of the first focuses of our future work. explore the prediction methodology using the updated dataset and use the most correct and applicable mil strategies for prediction. Live periodic predictions will be one of the first focuses of our future work. explore the prediction methodology using the updated dataset and use the most correct and applicable mil strategies for prediction. Live periodic predictions will be one of the first focuses of our future work.

REFERENCES

[1]. F. Petropoulos and S. Makridakis, "Prediction of the novel coronavirus covid-19", Plos one,flight. 15, no. 3, p. e0231236, 2020.

[2]. G. Grasselli, A. Pesenti and M. Cecconi, "Use of intensive care for the covid-19 epidemicin lombardy, italy: first experiences and forecasts during an emergency response", Jama, 2020.

[3]. S. Makridakis, E. Spiliotis and V. Assimakopoulos, "Statistical and machine learningforecasting methods: concerns and perspectives", PloS one, vol. 13, no. 3, 2018.

[4]. H. Asri, H. Mousannif, H. Al Moatassime and T. Noel, "Using Machine Learning Algorithmsfor the prediction and diagnosis of breast cancer risk", Procedia Computer Science, vol. 83, p. 1064–1069, 2016.

[5]. WHO. Name the coronavirus disease (covid-19) and the virus that causes it. [On line]. Available: https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-orientation/naming-of-the-disease-coronavirus-(covid-2019)-and-the-virus-that-causes-it

[6]. G. Bontempi, SB Taieb and Y.-A. Le Borgne, "Machine learning strategies for time series forecasting", at the European summer school on economic intelligence. Springer, 2012, p. 62–77.