Covid-19 Data Analysis using Machine Learning

Anika Bhardwaj
Galgotias College of Engineering and Technology,
Greater Noida, UP
anikabhardwaj31@gmail.com

N. Natarajan
National Informatics Centre
Delhi
natarajan@gov.in

2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) | 978-1-6654-3811-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAC3N53548.2021.9725367

***Abstract*-The COVID-19 epidemic began in Wuhan, China and has now expanded to the majority of a world's nations. The propagation of a pandemic is primarily determined according to each country's policies and social responsibilities. As per the WHO, the attack rate for 23 June 2020 is estimated to be between 1.4 and 2.5. In comparison to industrialized nations, India's position is rather manageable. It would be fascinating to learn about the facts and data surrounding corona cases throughout India. On world meters, many forms of data are provided. We aimed to assess similar information for India and created several predictions on the impacted rate, daily new cases, and daily total completed cases, among others. COVID-19 has cruelly stopped everything within civilization. An examination of COVID-19 records to determine which age groups are the most affected by the virus. Various Machine learning is used to develop predictive model. Algorithms as well as their related performance data are calculated and analysed. Regressor Random Forest and Random Forest The classification algorithm beat all other machine learning algorithm. such as Support Vector Machine, KNN+, Neighbourhood Component Analysis, decision tree classification, and Gaussian Classifier naive Bayesian, Multi linear Regression, various Logistic Classifiers based on the regression technique and the Extreme Gradient Boosting algorithm.**

Keywords- COVID, CoV-2, SVM, KNN+ NCA, Kaggle

I.INTRODUCTION

In December 2019, the Covid-19 pandemic struck Wuhan, China, and the origin was by a new severe sensitive Respiratory Syndrome Corona Virus 2 virus (SARS CoV2).
Corona virus Disease is caused by SARS CoV 2 in 2019 (COVID-19). The World Health Organization (WHO) proclaimed the epidemic a public health emergency and pandemic on January 30, 2020. COVID-19 causes respiratory disorders, exhaustion, a dry cough, and tiredness, among other symptoms, and 80 percent of people recover without treatment. COVID-19 is particularly dangerous for elderly men, youngsters, and men that already have cardiovascular disease, obesity, or diabetes.
COVID-19 causes respiratory disorders, exhaustion, a dry cough, and tiredness, among other symptoms, and 80 percent of patients recover without treatment. COVID-19 is particularly dangerous for elderly men, youngsters, and men who just possess cardiovascular disease, obesity, or diabetes.
According to World meter statistics, the count of COVID19 cases in India is 67,161 and the mortality toll is 2,212 by 11th MAY 2020. Worldwide, 4,180305 persons have been infected with a virus, resulting in a total of 283,865 fatalities from sickness.
In fact, Artificial Intelligence is the most effective instrument inside the battle against the COVID-19 catastrophe. AI is subdivided into subdomains such as Machine Learning and Deep Learning. It does have a variety of applications in Natural Language and Computer Vision. It aids in the diagnosis and prediction of COVID-19. Machine learning methods and deep learning methods are advantageous for monitoring Corona instances, forecasting, creating dashboards, diagnosing and treating patients, and generating alarms to preserve social distance and for other potential control mechanisms. The most effective method of preventing and slowing transmission is to maintain social distance. We must safeguard ourselves and others from illness by cleansing our hands or by using hand sanitizers and avoiding facial contact.

II.SARS-COV-2

SARS-CoV-2 is indeed an RNA virus with a single strand of ribonucleic acid (RNA). It is highly infectious among humans and has resulted in such a global epidemic. It endangers millions of people worldwide and also causes economic upheaval.
SARS-CoV-19 infects human cells as well as binds to a protein found inside the cells of numerous human being bodies nown as ACE2. Coronaviruses are a group of microorganisms that belong to the type Coronaviridae. Coronaviridae is indeed a family of single-stranded, enclosed RNA type viruses. Yang et al. describe the aetiology of SARSCoV-2.

A. Symptoms

COVID-19 is most frequently associated with flu-like symptoms. Due of the mild and atypical symptoms, it is becoming more hard to diagnose and confine.

Tiredness, fever, and a dry cough are by far the most common symptoms. Headache, diarrhoea, body aches, painful throat, loss of taste or inability to smell, reactions on skin, and chills are moderate symptoms. Severe symptoms include breathing difficulties or conciseness of breathing, discomfort in upper body or stress, and hammering of words.

B. Determination of Diagnosis

The virus that causes COVID-19 is an infection with SARS-CoV-2 which is detected by viral testing. If the result of test is positive this means the individual is infected. The test for diagnosing is generally depends on the location. Rapid Diagnostic Test (RDT) helps in detecting the presence of viral proteins, termed antigens, produced by the SARS CoV2 virus in a band of people respiratory.

Following around 30-40 minutes, if sufficient SARS-CoV-2 antigen is present in sample collected, it may combine too many antibodies linked to somewhat like a strip of paper in a storage box. It creates a readily detectable indication.

The RDT tests are performed to diagnose sensitive or premature infections with SARS-CoV-2, since the virus's

produced antigens are out only after successful replication. These tests are believed to be very accurate in diagnosing COVID-19. Another kind of RDT was offered for COVID19; it is a experiment that identifies the survival of auto antibodies of individuals supposed of being infected with COVID-19.

Antibodies develop between days or even weeks after viral infection. The suggested approach for evaluating and analysing COVID-19 cases and specimens is a chemical analysis of respiratory system test.

C. Treatment

Patients infected with COVID-19 do not have a specified treatment.

The drug is prescribed depending on the severity of the symptoms. This may include analgesics, cough syrup, rest, and increased liquid ingestion. Patients with minor symptoms may remain at ease and get therapy in remoteness. Otherwise, hospitalisation is obvious.

III. RESULTS

Some Machine Learning methods are used to comprehend COVID-19 affects on individuals, and to validate and forecast recovery. The Fig 1 illustrates the different groups of age and percentages of instances from the dataset of kaggle. These categories of age between 20 and 50 are at high risk for COVID-19 contraction.

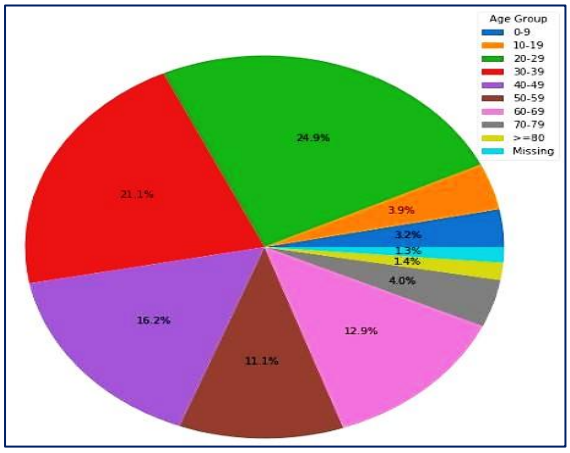

Fig. 1 As per age group, Percentage of COVID-19 cases

The databases Covid-19 in India and Covid-19 Data both are analysed and machine learning models for performance evaluation are built. The correlation matrices for the datasets are shown in Fig 2 and Fig 3.

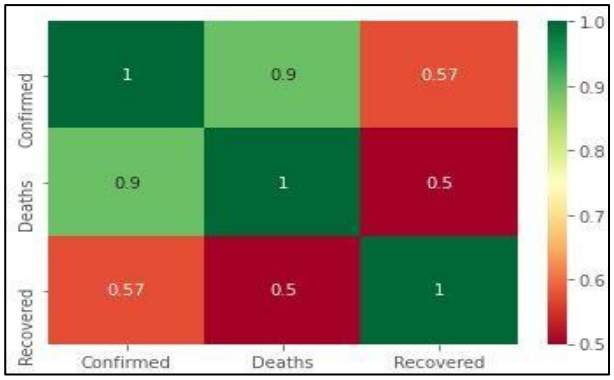

Fig 2(a) Correlation Matrix for Covid-19 in India Dataset

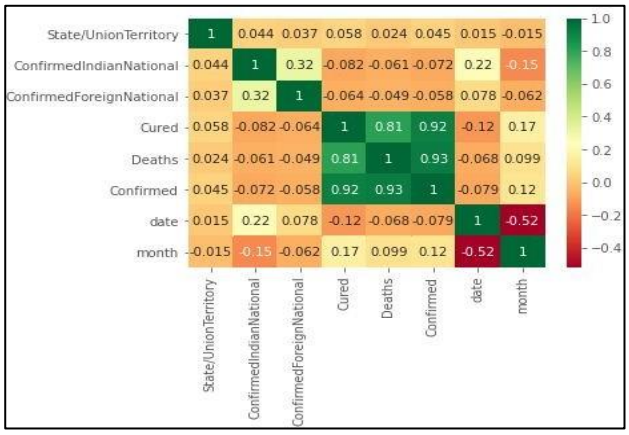

Fig 2(b) Matrix of Correlation for Covid-19 Data

The two datasets are used to build machine learning models using Support Vector Machine Algorithm, KNN+, NCA, Decision Tree Classifier, Gaussian Nave Bayes Classifier, Multi linear Regression algorithm, Logistic Regression algorithm, Random Forest Classifier and Extreme Gradient Boost Classifier. The RSquared (coefficient of determination) regression scores and accuracy are calculated using a 70:30 ratio of train to test datasets. The picture Fig 3(a) – Fig 3 illustrates the significance of the features in the Covid-19 India Dataset.

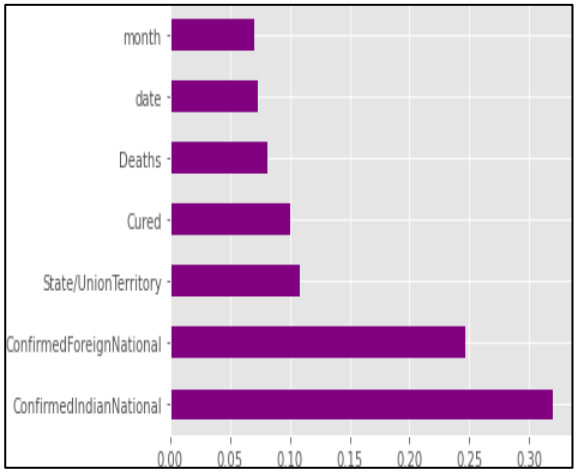

Fig 3(a) Importance of feature using Decision Tree Classifier

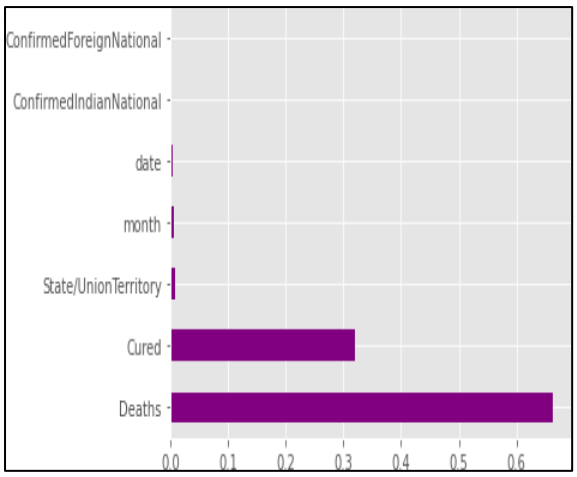

Fig 3(b) Importance of feature using Radom Forest Classifier

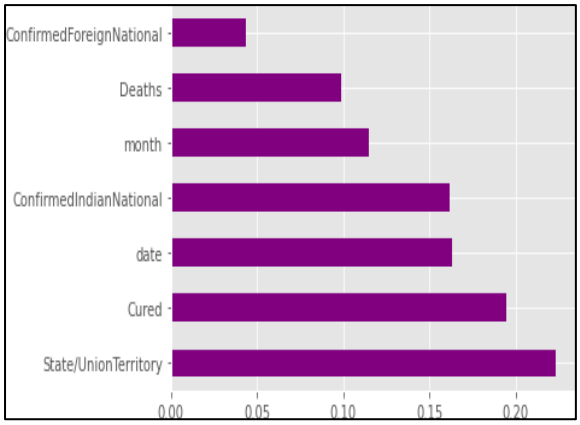

Fig 3(c) Feature extraction using Random Forest Regressor algorithm

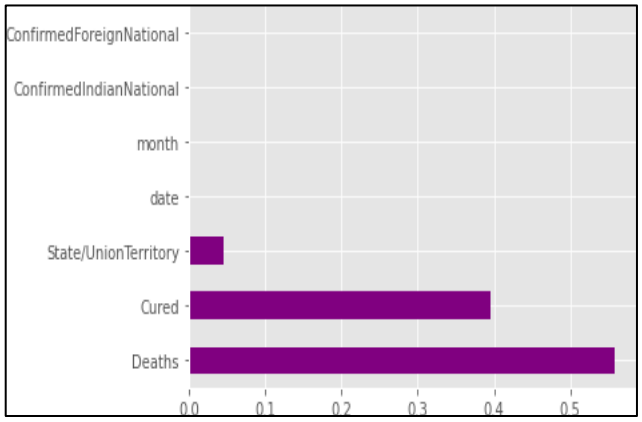

Fig 3(d) Feature extraction using Extreme Gradient Boosting Classifier

Below Figure 4(a) demonstrate the Coefficient of Determination popularly known as COD. It is also known as R-Square, and Figure 4(b) determines accuracy for models constructed using the COVID-19 in India Dataset. The Coefficient of Determination and Accuracy for something like the replicas generated on the COVID-19 Data Dataset are shown in Figs below 5(a) and 5(b).

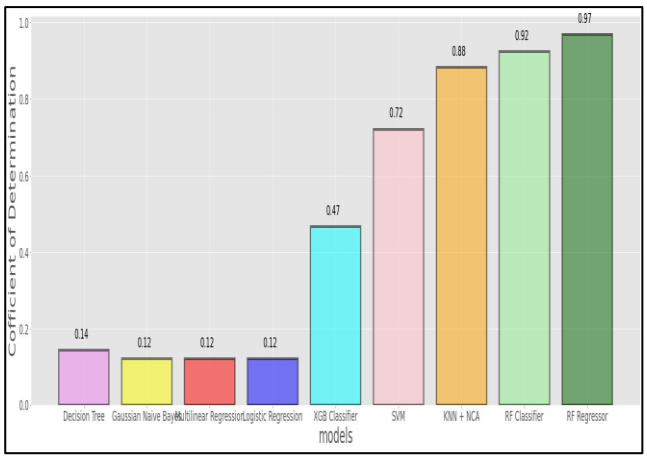

Fig 4(a) Coefficient for Determination for COVID-19-in India

The findings indicate that the Regressor and the algorithm known as Random Forest Classifier surpass the all machine learning models.

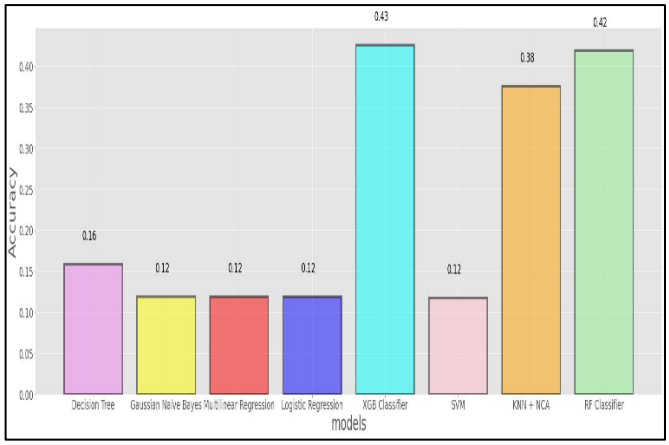

Fig 4(b) Accuracy rate for COVID-19 in India

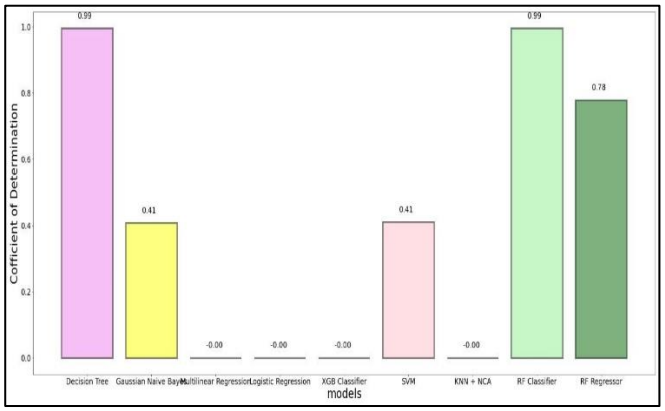

Fig 5(a) Coefficient for Determination for COVID-19 Data in India

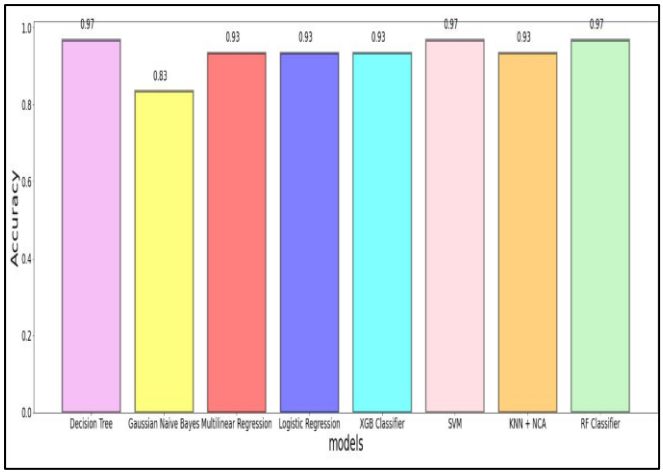

Fig 5(b) Precision rate of COVID-19 Data in India

IV. CONCLUSION

The investigations demonstrate that individuals in the age categories are infected with COVID-19. Matrices of correlation are constructed in order to comprehend the link among the datasets' attributes. The relevance of features is calculated for the classifiers that are constructed. Along with classifiers and suppressors, prediction is also constructed. The findings indicate that the algorithm known as Random Forest Regressor and the algorithm known as Random Forest Classifier outperform other various models in conditions of coefficient of determination and precision. In the upcoming years, further machine learning regression models and classifiers will be assessed on the expanding COVID-19 datasets.

REFERENCES

[1] https://doi.org/10.1038/s41591-020-0916-2.

[2] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a- detail/q-a-coronaviruses

[3] https://www.cdc.gov/ coronavirus/2019-ncov/symptoms-testing/ symptoms.html

[4] https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=AgeGroupDetails.csv

[5] https://doi.org/10.14419/ijet.v7i2.8.10332

[6] Kolla B.P., Raman A.R. "Data Engineered Content Extraction Studies for Indian Web Pages." Advances in Intelligent Systems and Computing, 2019: 505-512.

[7] Prakash, K.B. 2017, "Content extraction studies using total distance algorithm", Proceedings of 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016, pp. 673.

[8] Ismail, M., Prakash, K.B. & Rao, M.N. 2018, "Collaborative recommendation filtering based of online social voting", International Journal of Engineering and Technology (UAE), vol. 7, no. 3, pp. 1504-1507.

[9] Prakash, K.B., Rajaraman, A. & Lakshmi, M. 2017, "Complexities in developing multilingual on-line courses in the Indian context", Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017, pp. 339.