International Journal of Research in Engineering, Science and Management
Volume 4, Issue 6, June 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Analysis On the Progression of COVID with the Impact of Climatic Conditions

G. Swetha[1*], V. Parthiban[2], D. Shanti Chelliah[3], D. Padma Subramanian[4]
[1,2,3,4]*Department of Electronics and Communication Engineering, Sri Muthukumaran Institute of Technology, Chennai, India*

***Abstract*: Globally since December 2019 Coronavirus disease (COVID- 19) has been identified as a potentially severe contagious disease and WHO has declared the spread of COVID-19 as pandemic. Hence, Study of COVID-19 is most essential for prevention and proper treatment of the disease. COVID is a contagious disease in mammals and birds. In humans, coronavirus causes respiratory track infections that can be mild, such as some cases of the common cold, and the others can be lethal, such as SARS, MERS and COVID-19. Machine learning is a powerful technique which is used to train computer programs involving big data to take automated decisions. It is proposed to develop an efficient machine learning algorithm for analysis of progression of COVID-19 with the impact of climatic conditions such as temperature rainfall and humidity. Feature selection has several objectives such as enhancing model performance by avoiding overfitting in the case of supervised classification. Thus EDA is performed for a large set of data that is fed for different algorithms in order to arrive at efficient and lesser RMSE values.**

***Keywords*: COVID, Machine Learning algorithms.**

1. Introduction

The climate of India is interesting and complex because India's topography is so diverse. India sees a wide range of weather conditions throughout the enormous continent, which ranges from desert plains, to mountains, to forests, to beaches. That being said, the climate ranges from equatorial to alpine. The temperatures can differ drastically from one region of the country to another. Climate in South India is generally hotter and extremely humid than that of North India. South India is more humid due to nearby coasts. The southern half of the nation doesn't experience temperatures below 10 °C (50 °F) in winter, and the temperature usually tends to exceed 40 °C (104 °F) during summer.

The COVID-19 pandemic in India is part of the worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case of COVID-19 in India, which originated from China, was reported on 30 January 2020. India currently has the largest number of confirmed cases in Asia, [6] and, as of March 2021, has the third-highest number of confirmed cases in the world after the United States and Brazil with more than 11.4 million reported cases of COVID-19 infection and more than 158,000 deaths as of March 15, 2021.The per day cases peaked mid-September in India with over 90,000 cases reported per day and have since come down to below 15,000 as of January 2021.

SARS-CoV-2, the virusthat causes COVID-19, spreads through the environment. In this module we will deduct the effect of temperature and rainfall on COVID spread

2. Objective

The main aim of our project is to draw the impact of COVID with climatic conditions like temperature and rainfall. So, this project provides an opportunity to get precision on the spread of COVID with environmental conditions. The project deals with collection of COVID data for a particular interval of time and applying different algorithms for the calculation of the impact factor. It is achieved by Feature selection and hyper tuning the ML classifiers to get good accuracy. The different classifiers used are SVM – Support vector machine, DT – Decision tree, RF – Random forest.

3. Literature Survey

[1]. "Autism Spectrum Disorder Detection with Machine Learning Methods" by Uğur Erkan*, Dang N.H. Thanh In Current Psychiatry Research and Reviews Volume 15, Issue 4, 2019. In this work, Autistic Spectrum Disorder (ASD) is a disorder associated with genetic and neurological components leading to difficulties in social interaction and communication. According to statistics of WHO, the number of patients diagnosed with ASD is gradually increasing. Most of the current studies focus on clinical diagnosis, data collection and brain images analysis, but do not focus on the diagnosis of ASD based on machine learning. To classify ASD data to provide a quick, accessible and easy way to support early diagnosis of ASD. Three ASD datasets are used for children, adolescents and adults. To classify the ASD data, we used the k-Nearest Neighbours method (kNN), the Support Vector Machine method (SVM) and the Random Forests method (RF. The data was randomly split into training and test sets.

[2]. "Detection of Autism Spectrum Disorder (ASD) using Machine Learning Techniques: A Review "by Mamata V. Lohar and Suvarna S. Chorag. In International Journal of Future Generation Communication and Networking Vol. 13, No. 1, (2020), pp. 426-438. ASD depends on the factor of gender. The

*Corresponding author: reachmeatswetha@gmail.com

similarities and weaknesses in autistic children and adults are distinct. The irreversible loss is observed if ASD is not detected at an earlier stage. Hence there is a need for automated techniques for early and accurate detection. There are many developments in current research in the field of biomarkers for risk assessment, diagnosis and tracking of disease progression. Machine learning used in healthcare has made enhancements in diagnosis by processing and analysis of the huge amount of data. Present research work focuses on automated methods of identification to diagnose ASD accurately. Fusion method is used to combine any number of instruments, allowing data from various reliable sources to be fused, all within an objective framework that can be converted to the desired metric. Pre-processing techniques can be streamlined to incorporate techniques for data fusion to minimize ambiguity in feature evaluation. Classification between ASD and TC subjects for sMRI and fMRI using the K-NN classifier for different feature sets.

[3]. "Machine learning approach for early detection of autism by combining questionnaire and home video screening" by Halim Abbas, Ford Garberson, Eric Glover, Dennis P. Wall. In National Library of Medicine- J Am Med Inform Assoc. 2018 Aug 1;25(8):1000-1007 In this work, Existing screening tools for early detection of autism are expensive, cumbersome, time intensive, and sometimes fall short in predictive value. In this work, we sought to apply Machine Learning (ML) to gold standard clinical data obtained across thousands of children at-risk for autism spectrum disorder to create a low-cost, quick, and easy to apply autism screening tool. They used 2 algorithms to identify autism, one based on short, structured parent-reported questionnaires and the other on tagging key behaviours from short, semi-structured home videos of children. A combination algorithm is then used to combine the results into a single assessment of higher accuracy. Determination where appropriate in order to boost screening accuracy when conclusive. The performance is then validated in a controlled clinical study.

4. Methodology

The important steps in the analysis is collection of data. Once the data is collected it is necessary to perform EDA-Exploratory data analysis. This is done in order to remove the out layers that are present in the data which Is "Filling the missing data". Following the pre-processing is testing and training of data. During the training process various models are selected and trained. After the training the models are tested. If the results do not meet the expectation again the process is repeated. The parameter used for comparison is RMSE.

The various models that are iterated here are as follows:

A. SVM

A support vector machine (SVM) is machine learning algorithm that analyses data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible.

B. RF

The random forest algorithm is a supervised classification algorithm. As the name suggests, this algorithm creates the forest with a number of trees. In general, the more trees in the forest the more robust the forest looks like.

C. DT

Decision Tree Algorithm is a supervised Machine Learning Algorithm where data is continuously divided at each row based on certain rules until the final outcome is generated. Let's take an example, suppose you open a shopping mall and of course, you would want it to grow in business with time.

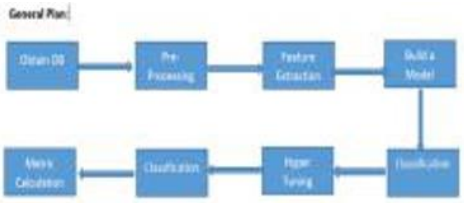

Fig. 1. Block diagram of COVID analysis

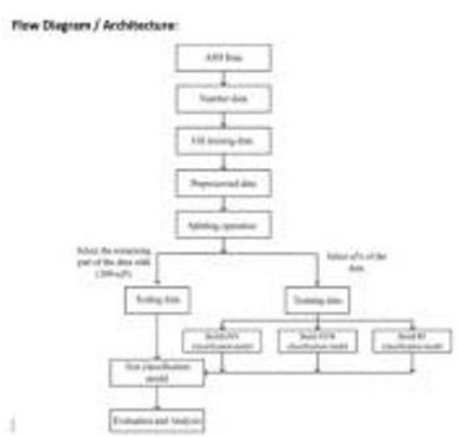

Fig. 2. Flow diagram

5. Software Description

A. Anaconda

Anaconda Enterprise facilitates machine learning and deep learning by enabling you to develop models, train them, and deploy them. You can also use AE to query and score models that have been deployed as a REST API. To help get you started, Anaconda Enterprise includes several sample notebooks for common repetitive task.

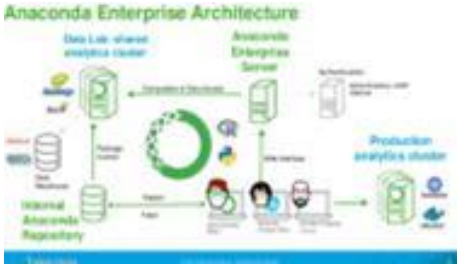

Fig. 3. Anaconda Enterprise Architecture

6. Verification and Results

A	B	C	D
Temperatu	Rainfall	State	Covid
27	16	TamilNadu	9829
29	2.3	Delhi	12000
23	4.54	Assam	3523
30	12	TamilNadu	9829
26	1.1	Assam	3523
9	4.3	Uttarpradesh	5000
34	1.5	Uttarpradesh	5000

Fig. 4. Sample data for COVID analysis

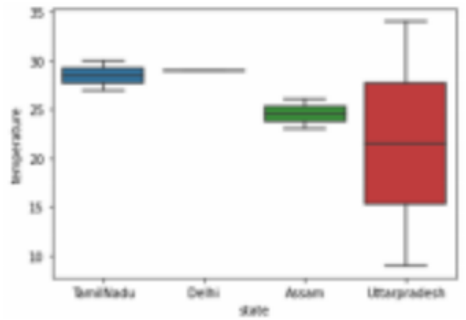

Fig. 5. Box plot after EDA

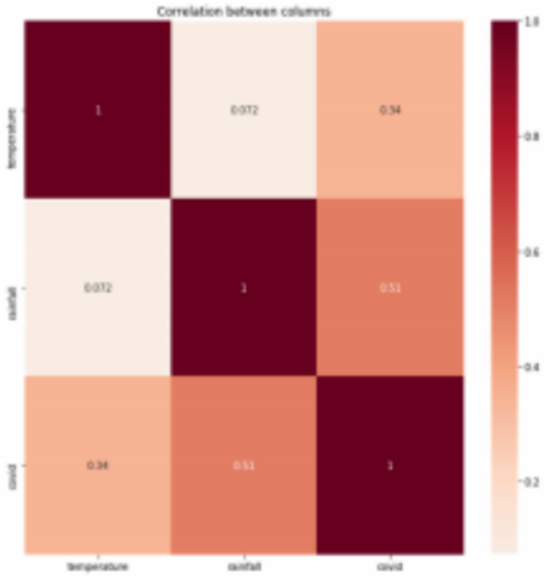

Fig. 6. Correlation after EDA

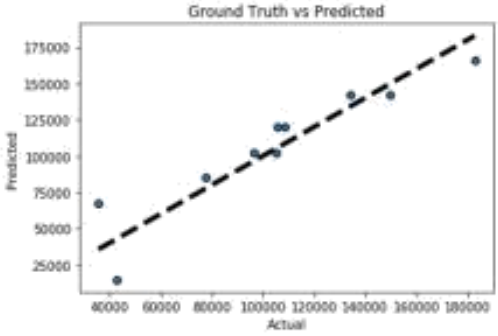

Fig. 7. Output of Decision Tree algorithm

Random Forest

Fig. 8. Output of Random Forest algorithm

SVM

Fig. 9. Output of SVM algorithm

7. Applications

Using the above analysis, the impact of climatic conditions on COVID is drawn It is extremely appropriate for huge collection of data for drawing appropriate RMSE value.

8. Conclusion and Future Enhancement

Using Machine Learning the various models have been iterated and compared. In the decision tree algorithm, the line of best fit does not pass through most of the points were as using SVM and Random Forest algorithms the RMSE values are quite high. Therefore, for fine tuning larger dataset is taken on which EDA is performed to obtain the best algorithm with efficient RMSE values.

References

[1] K. C. Howladar, M. S. Satu, A. Barua, and M. A. Moni, “Mining significant features of diabetes mellitus applying decision trees: A case study in Bangladesh,” bioRxiv, 2018.

[2] M. Duda, R. Ma, N. Haber, and D. Wall, “Use of machine learning for behavioral distinction of autism and adhd,” Translational psychiatry, vol. 6, no. 2, p. e732, 2016.

[3] A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, and I. Castiglioni, “Use of machine learning to identify children with autism and their motor abnormalities,” Journal of autism and developmental disorders, vol. 45, no. 7, pp. 2146–2156, 2015.

[4] F. Thabtah, F. Kamalov, and K. Rajab, “A new computational intelligence approach to detect autistic features for autism screening,” International journal of medical informatics, vol. 117, pp. 112–124, 2018.

[5] C. Cong and C. Tsokos, “Theory and applications of decision tree with statistical software,” age, vol. 58, p. 250, 2009.

[6] F. Thabtah, “An accessible and efficient autism screening method for behavioural data and predictive analyses,” Health informatics journal, 2018.

[7] H. Talabani and A. Engin, “Performance comparison of svm kernel types on child autism disease database,” in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). IEEE, 2018, pp. 1–5.

[8] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn and R. Di Napoli, "Features evaluation and treatment coronavirus (COVID-19), 2020. https://www.ncbi.nlm.nih.gov/books/NBK5 54776/