International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

Climate, Air Pollution and Forest fires : AI based models for Indian case study.

Dr.Mrs Nupur Giri, Mr. Aashish Ramesh Raheja, Mr. Kartikey Verma, Mr. Ashwin Pansare

Abstract—Forest fires are a regular phenomenon in India that are often observed during summers (March - May). Severe fires occur in different types of forests, particularly dry deciduous forest, while evergreen, semi-evergreen and montane temperate forests are comparatively less prone. Every year large areas of forests are affected by fires of varying intensity and extent. Important forest resources like the carbon locked in the biomass etc are lost due to forest fires every year, which adversely impact the biodiversity of the planet earth. This project includes an approach on predicting the chances of forest fire in 13 Indian states. Climatic factors such as humidity of the region ,rainfall level ,maximum and minimum temperature to determine the chances of occurrence of forest fire were considered for building the model. The model aims to study the gradual increase of wildfires in Indian districts and its direct impact on the natural vegetation and wildlife residing in the forests. The implementation of dataset collected and analytics obtained gives us a clear picture of the gravity of the situation and the detrimental effect of wildfire on the Ecology.

Index Terms—Artificial Intelligence,Forest Fire, Climate Change, Air Pollution, Supervised Machine Learning, Data Analytics, Web development .

1 Introduction

A study done by the Forest Survey of India [12] showed an increase in forest fires in India since 2004, the number of forest fires have kept increasing every year. Every year the fires affect large areas of forest with varying intensity and extent. Based on the forest inventory records, 54.40% of forests in India are prone to occasional fires, 7.49% are exposed to moderately frequent fires and 2.40% to high incidence levels while 35.71% of India's forests have not yet been exposed to fires of any real significance. As a forest burns, large amounts of smoke are released into the atmosphere. Air pollution from fires has the potential to travel great distances and oftentimes may pose a threat to human health.Our project aims to cover maximum districts from India covering 13 most vulnerable states which are- Madhya Pradesh, Mizoram, Maharashtra, Odisha, Assam, Manipur, Chhattisgarh, Meghalaya, Andhra Pradesh, Telangana, Tripura, Nagaland, Uttarakhand. The aim of this project is to use Machine Learning algorithms to predict the forest fire and analyze its effect on the level of pollutants(NO2,SO2 and RSPM). Our dataset analyzes patterns for a period of 9 years (2010-2018) and gives an efficient solution. With early prediction of the forest fires the damage to the environment and humans can be mitigated and millions of natural resources and wildlife can be saved.

In September 2015, the UN General Assembly adopted the Sustainable Development Goals (SDGs). These objectives form a program of sustainable, universal and ambitious development, a program of the people, by the people and for the people, conceived with the active participation of UNESCO. The main motivation for this project was to adheres to with the Sustainable Development goal 13 and Goal 15 i.e. Climate Action and Life on Land.

2 Related Work

[1] elaborates on how forest fires occurrences are affected by different parameters. It explains how environmental parameters like temperature, oxygen level in the atmosphere and humidity levels affect the spread of forest fires through a region. This research also elaborates on the various other methods used in the field of forest fire detection and prevention ,thus explaining the drawbacks of using alternative systems. The lacuna in the given research is that the dataset used for training the machine learning algorithms is not based on actual data values collected during forest fires occurrences in a particular region and does not clearly define the scope of the regions to be evaluated and considered for model training.

The research [10] has applied parallel SVM for modeling the prediction on forest fires which aims to improve the efficiency of the prediction. The parallel SVM model divides large data in smaller, independent subsets. The SVM is trained through

International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

partial SVMs. Each subSVM is used as a filter that takes us towards the global optimum. The computations are parallelized which reduces the computing time. This model can be useful to improve efficiency and maintain high accuracy when working with large datasets. In the research, the basic parameters considered for making the model which greatly affect the occurrences of the forest fire were relative humidity, temperature, rain and wind speed. On comparing the linear SVM with parallel SVM it was found that parallel SVM was more reliable than linear SVM as it had better accuracy and efficiency as compared to the linear SVM and these results were stable for different datasets. Linear SVM cannot perform well when it comes to large datasets, but for smaller datasets the linear SVM is faster as it has to perform less computation to compute the linear relationship among the features. The parallel SVM when compared with existing models such as Convolutional Neural Network (CNN), Radial Basis Function (RBF), Polynomial Neural Network (PNN), and Multilayer Perceptron Neural Network (MPNN) showed a lower error value in forest fire prediction.

The research [11] studies the effect of different pollutants in the air on the air quality using machine learning. It uses the data available from Central Pollution Control Board, India which has observations from January 2015 to July 2020. Through statistical analysis it was found that pollutants PM10, PM2.5, CO, NO2, SO2, NOx and NO had higher correlation with AQI and are responsible for higher AQIs.
For predicting the AQI levels five models, KNN, Gaussian Naive Bayes, SVM, RF and XGBoost were used. SMOTE resampling was also used to handle the data imbalance. The models were evaluated using the standard error metrics like MAE, RMSE, RMSLE. It was found that all models performed better when SMOTE resampling was applied and the XGBoost model performed better in terms of accuracy as well as error statistics.

3 Data Gathering, Mapping And Pre-processing

3.1 Data gathering

1) Climatic Data:
Climatic data was collected from Skymet Weather Services [16], 2010-2018 daily data. Climatic data for every state was arranged for parameters like Rainfall, Relative Humidity, Minimum Temperature and Maximum Temperature.

2) Pollution Data:
The pollution data over a time period of 9 years(2010-2018) is collected from the Open Government Data Platform India for the districts of the states selected. Pollution levels of pollutants like SO2, NO2, RSPM were obtained,

3) Forest Fire Data:
Forest fire occurrence data for states across India was collected from the Forest Survey of India portal. Fire points for 13 states across 9 years(2010-2018) were collected.
4) Data Preprocessing:
The missing values for Maximum temperature, Minimum temperature, Relative Humidity and Rainfall in climatic dataset are filled in by the average values of the parameters respectively for every state individually.
The climatic dataset comprises the values for the climatic parameters namely Rainfall, Relative Humidity, Minimum temperature, Maximum temperature and also includes the Fire occurrence data from 2010 to 2018 for every state arranged district wise. The dataset is found to be highly imbalanced and is inclined towards the value when fire occurrence is false. In order to solve this, the following steps were performed.:-
    1) The dataset is divided into two parts. The division is made based on the condition of fire occurrence is true or false on the given date.
    2) Now, the average values of climatic parameters are calculated when fire occurrence is true and when fire occurrence is false.
    3) The instances of fire occurred on a given date is changed from false to true if any of the points below are true:-
        1. The maximum temperature and minimum temperatures on the given date when fire occurred is false is greater than the average of maximum temperature and minimum temperature when fire occurred is true respectively.
        2. The rainfall and relative humidity values when fire occurred is false is less than the average of rainfall and average of relative humidity values when fire occurred is true respectively.

Fig. 1 shows the distribution of fire occurrence before and after the processing of data points based on climatic parameters.

International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

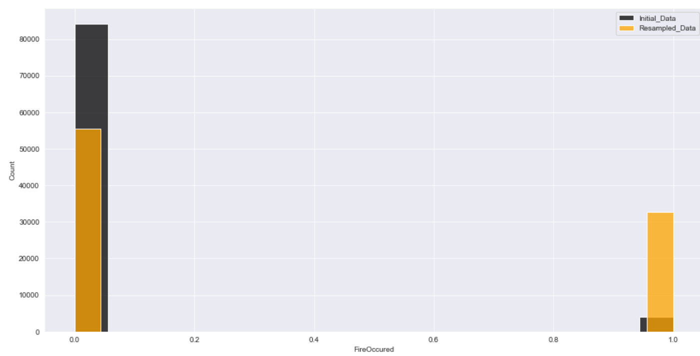

Fig. 1 : Comparison of fire occurrence distribution before and after resampling of data for the state of Telengana.

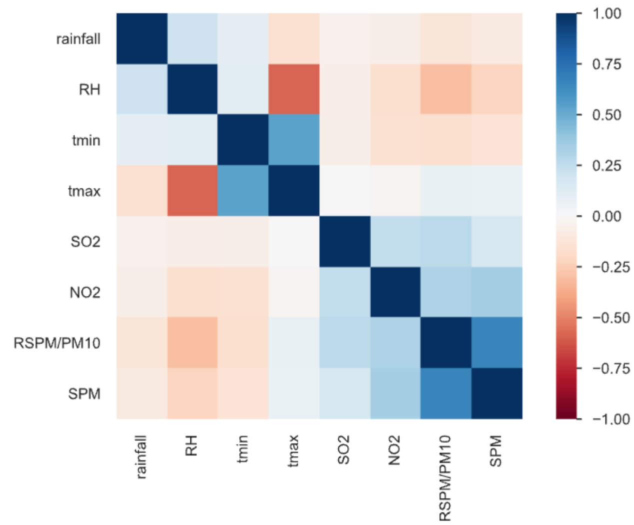

Fig. 3 : Correlation Model

4 METHODOLOGY

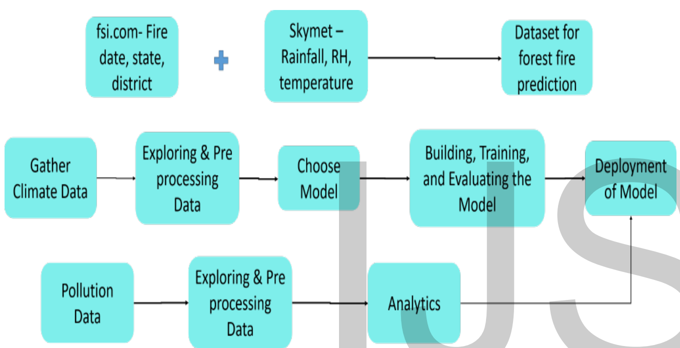

Fig 2. Flowchart for methodology

4.1 Understanding the data and the correlation between climatic variables and pollutants:

The Spearman rank correlation as shown in fig 3, was used to understand and establish the relation between the climatic data parameters of Temperature, Relative humidity and Rainfall and pollutants.

In order to elaborate the point of correlation between climatic data and fire occurrence, as shown in figure 4 we plotted a graph for relative humidity and fire occurred from 2010 to 2018, and indirect proportionality between fire occurred and relative humidity could be seen.

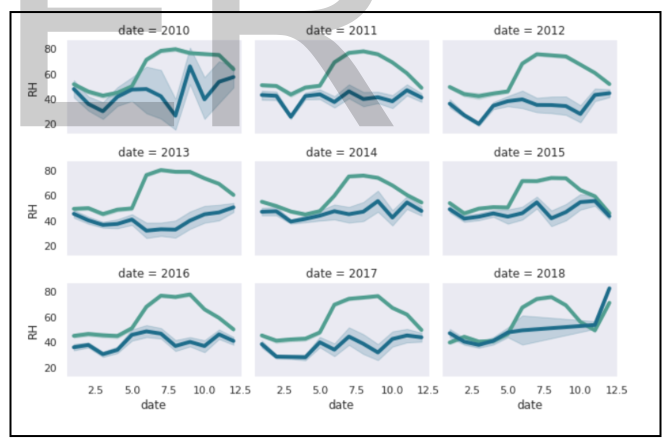

Fig. 4 : Variation of Relative Humidity from 2010 - 2018

Fig. 5 depicts how fire occurrence is affected by considering all the combinations that can be derived from the selection of 2 parameters from Maximum temperature, Minimum temperature, Relative humidity, Rainfall was plotted for further analysis of data.

International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

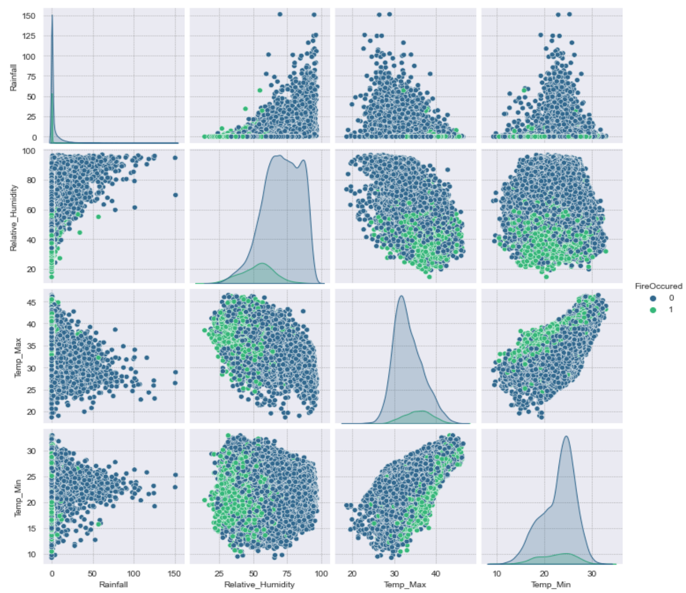

Fig. 5 : Fire Occurrence variation plot

4.2 Prediction model for Fire Occurrence:

In order to detect fire occurrence based on the dataset created, machine learning classification techniques like Random Forest Classifier, Decision Tree Classifier, Support Vector Classifier, Gaussian Naive Bayes Classifier, were used for building the classification model for fire occurrence in every state.

Input Features to Classification Models:

1) Maximum Temperature.
2) Minimum Temperature.
3) Relative Humidity.
4) Rainfall.

Output of the Classification Model:

1) Fire Occurrence (0 or 1).

Train Test split is 70% and 30% testing set. Fig. 1 shows all the models implemented along with parameters and evaluation metrics for each model is specified. A graph created in a Tableau dashboard, which forecasts forest fire occurrences for the next 5 years(2019-2024) district wise according to every state.

Model	Metric
Gaussian Naive Bayes	Accuracy:86.64 Precision:76.51 Recall:93.70 F1 score:84.24
Support Vector Classifier	Accuracy:95.81 Precision:97.43 Recall:91.40 F1 score:94.32
Decision Tree Classifier	Accuracy:94.51 Precision:92.55 Recall:93.08 F1 score:92.81
Random Forest Classifier	Accuracy:96.95 Precision:99.60 Recall:92.37 F1 score:95.85

4.3. Deployment and Prediction of Machine Learning Models:

After the training of machine learning models on the climatic dataset for all the states, in order to perform predictions, a web based application system is created.

The user can select the state from the dropdown menu and enter the values of 4 climatic factors namely Rainfall, Relative Humidity, Minimum Temperature and Maximum Temperature,

Upon submission of the values, a fire occurrence classification result of 0 or 1 is displayed . 0 indicates a low chance of forest fire occurrence and 1 indicates a higher chance of forest fire occurrence.

The web based application consists of 3 modules:

1) Front-end Application : The front-end part of the web application is developed using HTML, CSS and Javascript.
2) Back-end Application : The back-end application of the website is implemented using the Flask framework in Python. The POST method is used to collect values from the form displayed on the website. Upon submission of the values, the received variables are fed as an input to the machine learning model package.
3) Machine learning model package : Inorder to enable predictions of forest fire occurrence in various states, the ML model which is trained and tested on the dataset of respective states is converted into a packaged file which can be used to perform predictions on custom data using Pickle. Pickle is a module in Python which is used for serializing and de-serializing Python object structures. The model which is pickled can be downloaded as a .pkl file. Upon loading the .pkl file and feeding the model with respective values which are received using the Flask application, various functions like prediction,score are performed. After the prediction function is performed, the result generated is stored and sent back to the webpage using the Flask server for displaying the results as shown in Fig 6.

International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

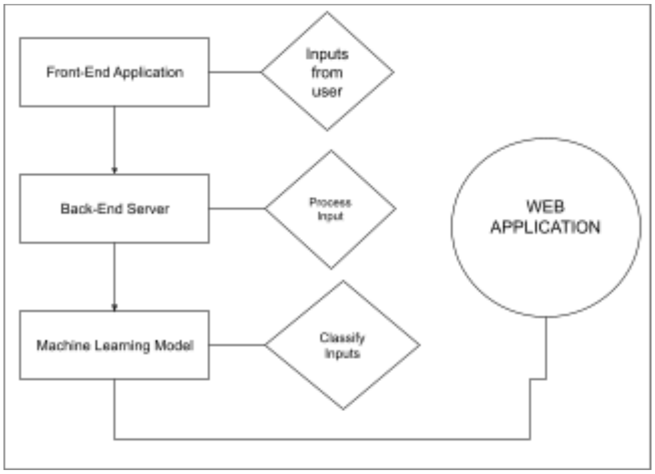

Fig. 6 : Flowchart for web application

5 RESULTS:

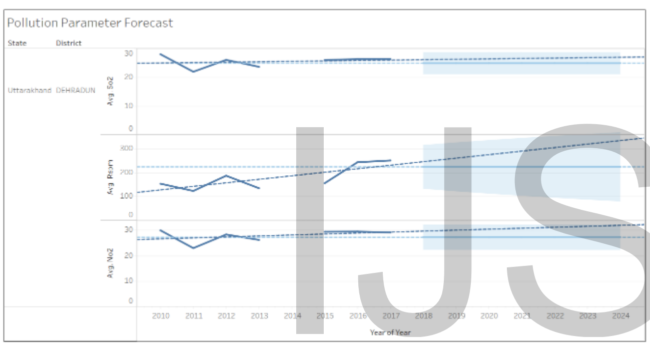

Fig. 7 : Pollution forecast.

Fig. 7 shows the forecast of parameters such as SO_2,NO_2,Rspm for the next 5 years.From the graph we can predict that the amount of pollutants such as NO_2,SO_2,and RSPM will increase in the next 5 years.

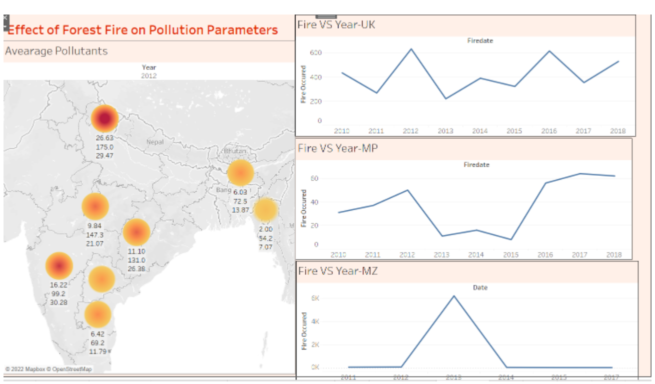

Fig. 8 : Effect of Forest Fires on SO_2.

Fig. 8 depicts the effect of forest fire on pollution parameters.The orange color depicts the intensity of the average SO_2 in different states.The above depicts that as the number of fires have increased the amount of pollutants have also increased.

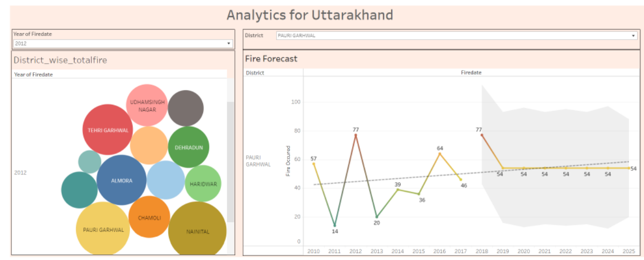

Fig. 9 : Fire forecast for 5 years.

Fig. 9 depicts the forecast of fire for next 5 years,total fire occurred in that particular district.

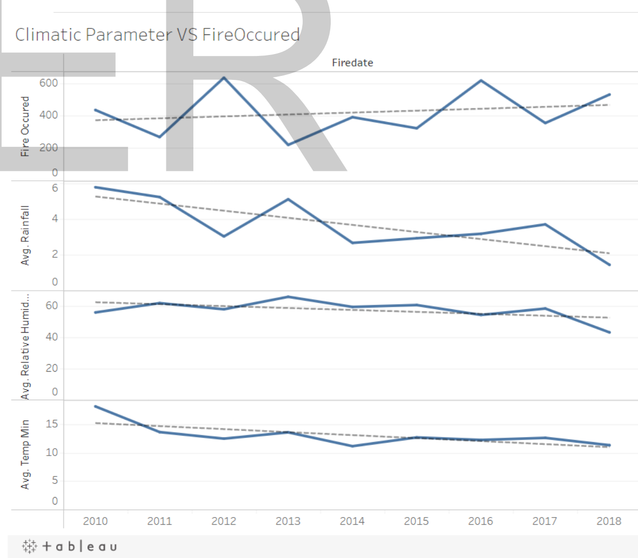

Fig. 10 : Climatic Parameters variation with fire occurrence.

Fig. 10 depicts the fire that occurred and the variation of climatic parameters with it.It is clear from the above graph that the decrease in the amount of rainfall and relative humidity have led to an increase in the number of forest fires.

The forest fire portal web application is shown in Fig 11. As mentioned above, the user can enter the values of parameters and the result will be displayed. Additionally, the analytics

International Journal of Scientific & Engineering Research, Volume 13, Issue 10, October 2022
ISSN 2229-5518

dashboard is attached to the main portal. As a result the user can also study the effects of forest fires across India.

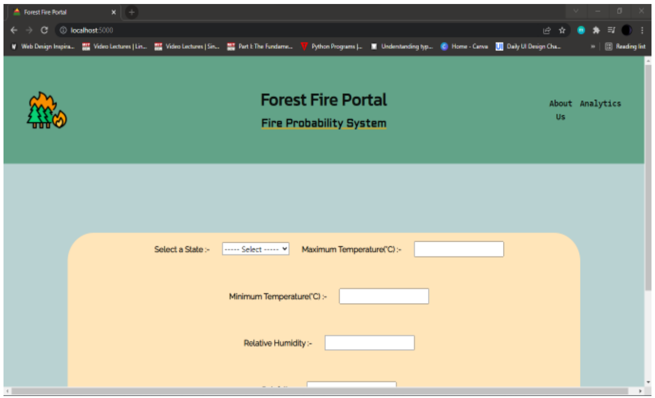

Fig. 11 : Web application

REFERENCES:

[1] Nachiket Hebbar, "*Prevention of Forest Fire with AI*", International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) Volume IX, Issue V, May 2020

[2] Mubarak A. I. Mahmoud, Honge Ren, *"Forest Fire Detection Using a Rule-Based Image Processing Algorithm and Temporal Variation"*, Mathematical Problems in Engineering, vol. 2018, Article ID 7612487, 8 pages, 2018.

[3] Salman ÖZÜPEKÇE, *"USE OF LAND SURFACE TEMPERATURE (LST) DATA IN THE DETERMINATION OF HIGH AREAS OF FOREST FIRE RISK IN ERDEMLI DISTRICT*", The Journal of International Social Research Volume: 11 Issue: 57 June 2018 Issn: 1307-9581

[4] Zhu, Dixian, Changjie Cai, Tianbao Yang, and Xun Zhou. 2018. *"A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization"* Big Data and Cognitive Computing 2, no. 1: 5.

[5] S.D. Wang, L.L. Miao, G.X. Peng, "*An Improved Algorithm for Forest Fire Detection Using HJ Data*", Procedia Environmental Sciences, Volume 13, 2012, Pages 140-150, ISSN 1878-0296

[6] Mauro Castelli, Fabiana Martins Clemente, Aleš Popovič, Sara Silva, Leonardo Vanneschi, "*A Machine Learning Approach to Predict Air Quality in California*", Complexity, vol. 2020, Article ID 8049504, 23 pages, 2020.

[7] Forest Survey of India, "*Rapid Assessment of Fire affected Forest Areas in the Country based on MODIS-detections following a Sampling Approach*", FSI Technical information series, Volume 2 No. 2 2020.

[8] O, S., Hou, X. & Orth, R. *"Observational evidence of wildfire-promoting soil moisture anomalies."* Scientific Reports 10, 11008 (2020).

[9] D. Chaparro, M. Vall-llossera, M. Piles, A. Camps, C. Rüdiger and R. Riera-Tatché, *"Predicting the Extent of Wildfires Using Remotely Sensed Soil Moisture and Temperature Trends"*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 6, pp. 2818-2829, June 2016.

[10] Kajol R Singh, K.P. Neethu, K Madhurekaa, A Harita, Pushpa Mohan, *"Parallel SVM model for forest fire prediction"*, Soft Computing Letters, Volume 3, 2021, 100014, ISSN 2666-2221

[11] Kumar, K., Pande, B.P. *"Air pollution prediction with machine learning: a case study of Indian cities."* International Journal of Environmental Science and Technology (2022).

[12] Forest Survey of India. (2022, July). *"Forest fire activities"* https://fsi.nic.in/forest-fire-activities Accessed 29 Aug. 2022.