Forest Fire Analysis and Prediction using ML Regressors

Soumyashri Singha, Dibyanshi Tripathy, Dev Parmar
School of ECE, MITWPU Pune, Maharashtra.
1032190167@mitwpu.edu.in, 1032190271@mitwpu.edu.in, 1032190154@mitwpu.edu.in

Abstract - Forest fire is a major disaster or major environmental problem which has effects such as loss of biodiversity, extinction of plants and animals, loss of wildlife habitat and depletion of wildlife, loss of natural regeneration and reduction in forest cover, global warming and many more. Prediction of forest fires is expected to reduce the impact of forest fires in the future. Many fire detection algorithms are available with different approaches towards the detection of fire. The data set used in this paper is presented within the UCI machine learning repository that consists of climate and physical factors of the Montesinos park in Portugal and it has a total of 517 entries (rows) and 13 features (columns) for each row. This paper displays visualization of data and different machine learning algorithms of regression to predict the area of land burned or prone to catch fire w.r.t the features provided. Here the comparison between all the methods are shown and thereby get a better idea of which one is more suitable.

Keywords- Machine Learning, Linear Regression, Decision Tree Regression, Random Forest Regression, XGBoost Regression, REC curves, Forest Fires.

I. INTRODUCTION

Forest fires are a major environmental issue, they pose a threat not only to the forest wealth but also to the entire regime to fauna and flora disturbing the biodiversity and the ecology of a region.The spread of forest fires depends on many factors such as human careless activities, Climatic changes, Deforestation, Power line failure. Earlier authorities used to collect the satellite images of forest and if there was any such kind of emergency they found caused by the fires then they used to give signals to neutralize its effects. But by that time the fires would have already caused a lot of damage to the specific area resulting in more loss.
For this, fast detection is a key element for controlling such a phenomenon. To achieve this, one alternative is Forest Fire Prediction and Analysis using Machine Learning and Regressors..This prediction represents an essential tool to predict fire risks, back up the forest fire monitoring and extinction phase, and to assist in the fire control planning and resource allocation.

The application of this predictive model helps in analyzing the burned area according to the given data and also predicting the burned area if parameters are changed. In this approach the data is modeled as a regression task where various algorithms are used and its efficiency is calculated by splitting the data into train and test.

II. LITERATURE SURVEY

A. *Forest Fire Detection using Machine Learning Techniques* -
This research proposes three machine learning approaches, linear regression, ridge regression, and lasso regression algorithm. This paper uses two versions, all features are included in the first, and 70% of the features were included in the second. The paper uses a training set which is 70% of the data set, and the test set is 30% of the data set. The accuracy of the linear regression algorithm gives more accuracy than ridge regression and lasso regression algorithms

B. *A Data Mining Approach to Predict Forest Fires using Meteorological Data* -
Explore a Data Mining (DM) approach to predict the burned area of forest fires. Five different DM techniques, e.g. Support Vector Machines (SVM) and Random Forests, and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on recent real-world data collected from the northeast region of Portugal.

C. *Forest Fire Forecasting Using Fuzzy Logic Models* -
. Fuzzy Inductive Reasoning (FIR) and the Adaptive Neuro-Fuzzy Inference System (ANFIS) are two powerful fuzzy techniques for modeling burned areas of forests in Portugal. The results obtained from them were compared with those of other artificial intelligence techniques applied to the same datasets found in the literature.

D. *Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo* -
This research aims to evaluate random forest models in predicting forest fires based on the climate variables and satellite data of burned areas. Prediction of forest fires is expected to reduce the impact of forest fires in the future. Based on analysis of spatial and annual variability, the random forest model with all selected climate variables can represent the forest fires event over Borneo.

E. *Forest Fire Prediction Using Machine Learning Techniques* -
In the existing work processes, the fire-affected region is predicted based on the satellite images. To predict the occurrences of a forest fire the proposed system processes using the meteorological parameters such as temperature, rain, wind and humidity were used. Random forest regression and Hyperparameter tuning using RandomizedSearchCV algorithm we used various sub-samples of dataset on which it fits several decision trees and uses averaging to improve the predictive accuracy and control over-fitting. Based on the analysis of the models, all the selected meteorological parameters can represent forest fire events.

F. *Forest fire prediction using ML and AI* -

A new forest fire risk prediction method is described, which is based on Support Vector Machines, Logistic Regression, KNN, decision trees, and Random Forest. The findings show that forest fire danger can be predicted with reasonable accuracy.

G. *Predicting Forest Fires using Meteorological Data: an ANN Approach* -

In effect, meteorological conditions (e.g. temperature, wind) are known to influence forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In this work, we explore a Just Neural Network (JNN) approach to predict the burned area of forest fires that were tested on recent real-world data collected from the northeast region of Portugal.

H. *Forest Fire Prediction System using Machine Learning*

This system presently analyzes forest fire prediction methods based on machine learning . A novel forest fire risk prediction algorithm, based on support vector machines, is presented. The algorithm depends on previous weather conditions and data in order to predict the fire hazard level of a forest. The implementation of the algorithm using the present data and accurately predicts the hazard of fire occurrence.

I. *An intelligent system for forest fire risk prediction and fire fighting management in Galicia* -

This system was developed for the region of Galicia in NW Spain, which is one of the most affected regions of Europe by fires. This system fulfills three main aims of predicting forest fire risks and acts as a crucial preventive tool, it backs up the forest fire monitoring and extinction phase, and it assists in planning the recuperation of the burned areas. The model is based on a neural network whose output is classified into four symbolic risk categories and obtaining an accuracy of 0.789. The CommonKADS methodology was used to develop the system.

J. *Artificial Intelligence for Forest Fire Prediction* -

This paper presents a description and analysis of forest fire prediction methods based on artificial intelligence. A new forest fire risk prediction algorithm, based on support vector machines, is presented. The algorithm depends on previous weather conditions in order to predict the fire hazard level of a day. Using data from Lebanon, the ability of the algorithm to predict the fire occurrence is demonstrated.

K. *PREDICTING BURNED AREA OF FOREST FIRES* -

This research proposes various Machine learning approaches such as Naïve Bias, Decision Trees, SVR, Random forest, Stochastic Gradient Descent and Bagging for predicting the amount of land burnt in the forest. A Clusterized Forest Fires Data Set was used for this analysis. Here the predictive model is built using the outbreaks of fire caused in the northeast region of Portugal.

III. DATA ANALYSIS

The objective of the data analysis step is to increase the understanding of the problem from the data. There are two approaches to describe a given dataset. Summarizing and Visualizing data. This dataset is publicly available for research. The details are described in [Cortez and Morais, 2007].The data can be used to test regression (difficult task), feature selection or outlier detection methods. There are 517 instances and 13 attributes in my dataset.[11]

Dataset information :

The attributes in the dataset include:

X - x-axis spatial coordinate within the Montesinho park map: 1 to 9

Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9

Month - month of the year: "jan" to "dec"

Day - day of the week: "mon" to "sun"

FFMC - FFMC index from the FWI system: 18.7 to 96.20

DMC - DMC index from the FWI system: 1.1 to 291.3

DC - DC index from the FWI system: 7.9 to 860.6

ISI - ISI index from the FWI system: 0.0 to 56.10

Temp - temperature in Celsius degrees: 2.2 to 33.30

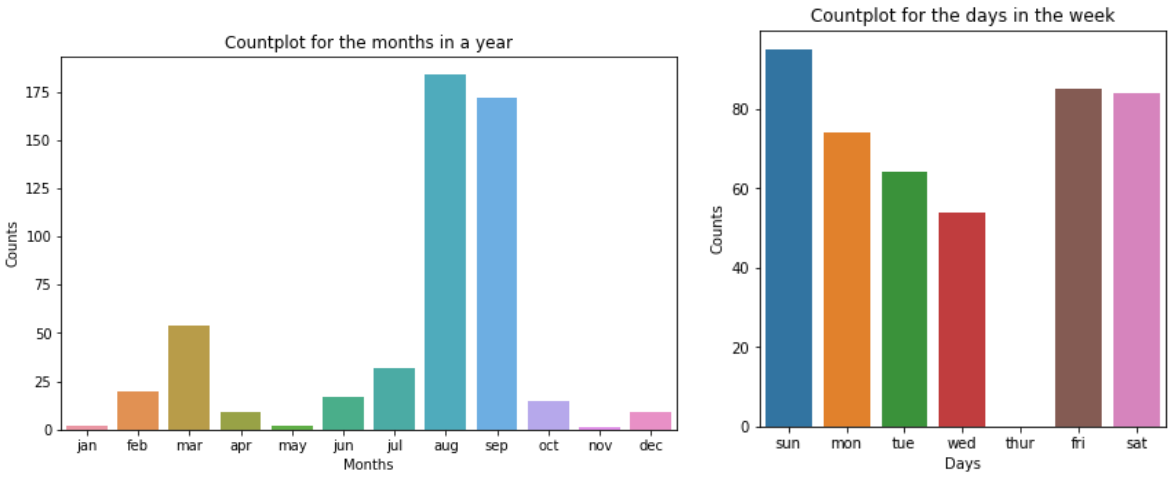
RH - relative humidity in %: 15.0 to 100

Wind - wind speed in km/h: 0.40 to 9.40

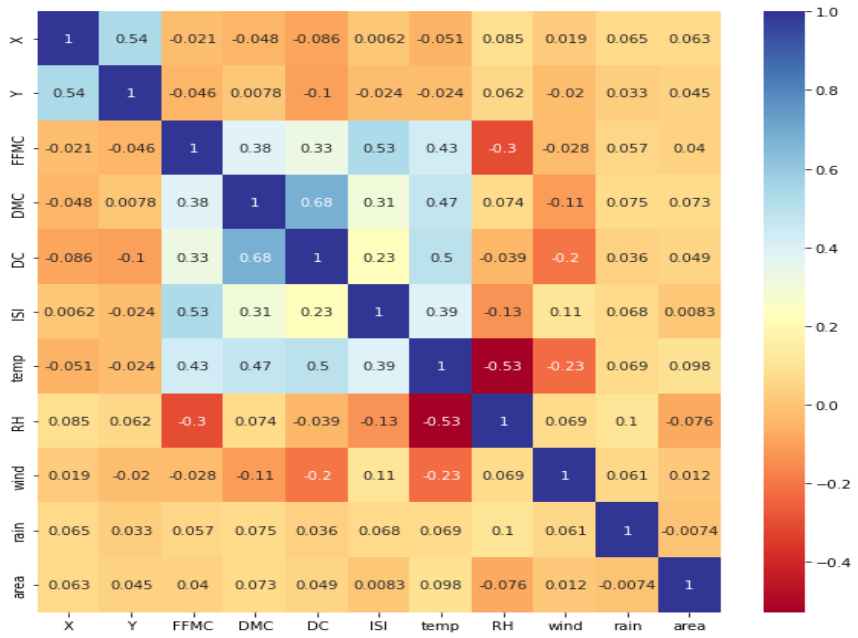
Rain - outside rain in mm/m2 : 0.0 to 6.4

Area - the burned area of the forest (in ha): 0.00 to 1090.84 [11]

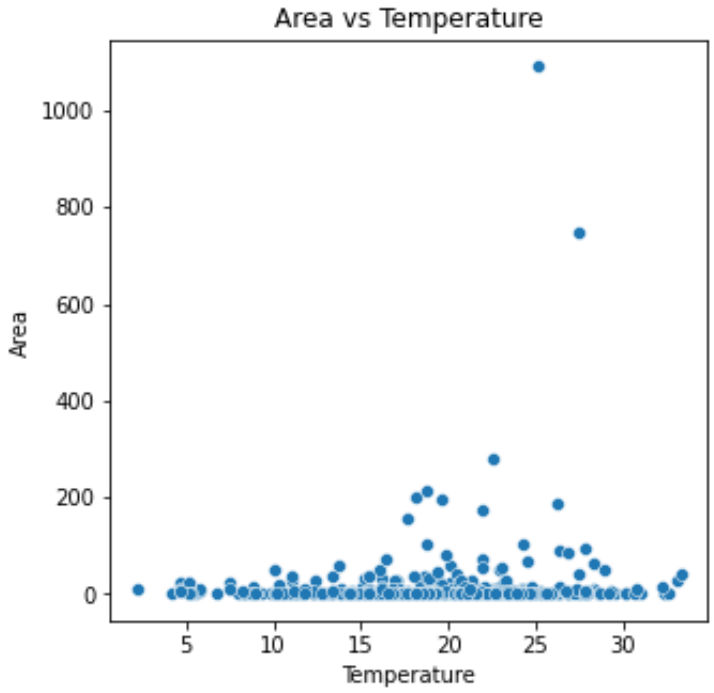
Countplot of months and days:



Heatmap:



Area vs. Temperature:



IV. METHODOLOGY (ALGORITHMS)

In this project we tried to make a prediction for the burned area using the Forest Fires Data Set from Kaggle. Stepwise regression methods were applied to choose one best predictor. It is interesting to see which one of them has the biggest impact on the burned area in each cluster.

Method 1: Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. The problem here is modeled into a Regression task since the over motive is to predict the area of the land burnt. Normally a regression equation is Y(Dependent variable) = a (intercept) + b (slope of the line) * X (Independent or explanatory variable) [5]

Method 2: Random Forest Regressor

The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees. It is among the most popular machine learning algorithms due to its high flexibility and ease of implementation.A forest is created using decision trees, each decision tree is a strong classifier on its own. These decision trees are used to create a forest of strong classifiers. This forest of strong classifiers gives a better prediction than decision trees or other machine learning algorithms. [5]

Method 3: Decision Tree Regression

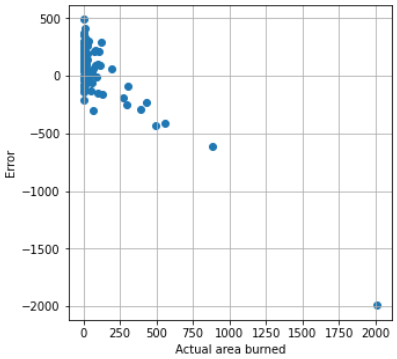
Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs, and utility. It observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values. It falls under supervised learning algorithms. It works for both continuous as well as categorical output variables. It is a bagging-based algorithm with a key difference wherein only a subset of features is selected at random. [5]

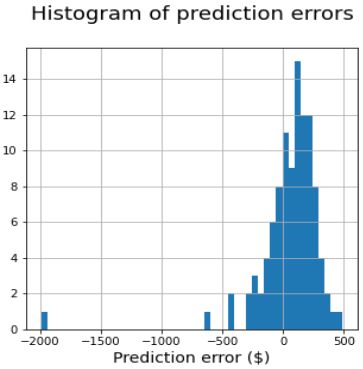
Method 4: XGBoost

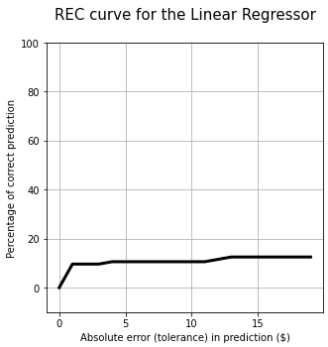
XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now. It is a perfect combination of software and hardware optimization techniques to yield superior results using less computing resources in the shortest amount of time. [5]

V. PREDICTION

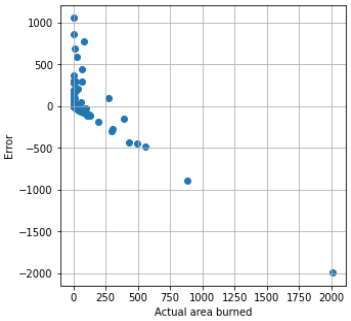
1. Using Linear Regression:
   Area of Burned forest - (-0.78)

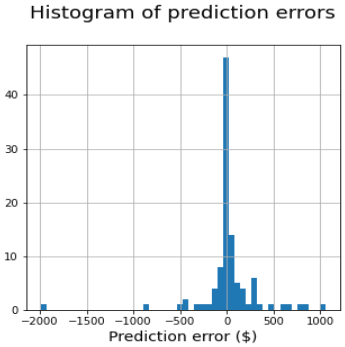


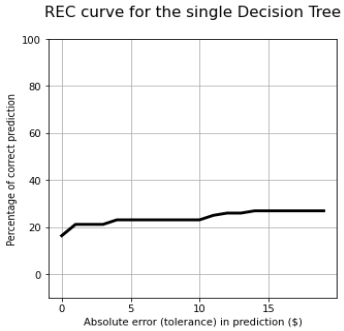




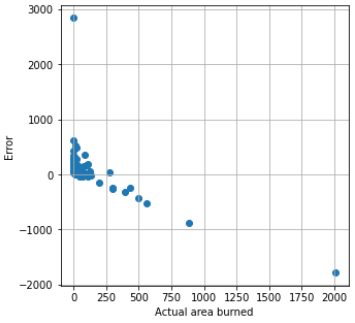
2. Decision Tree Regression
   Area of Burned Forest - (26.00)

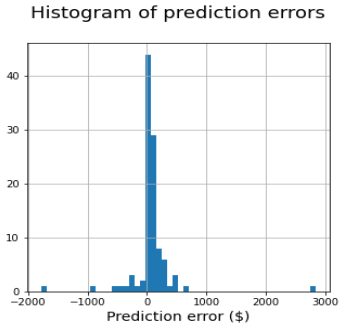


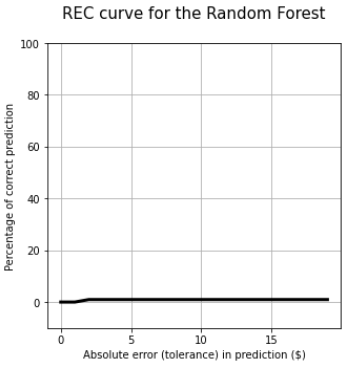




3. Random Forest Regression:
   Area of Burned Forest - (33.00)

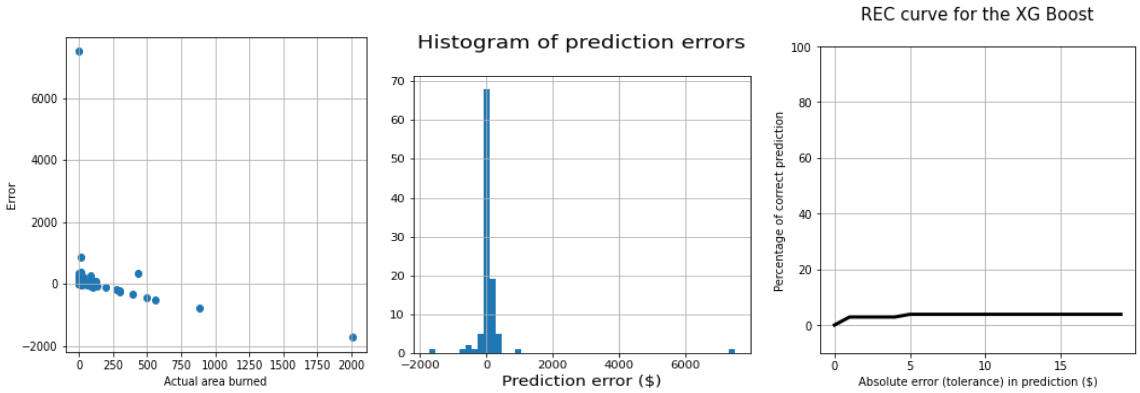






4. XGBoost

   Area of Burned Forest - (77.00)



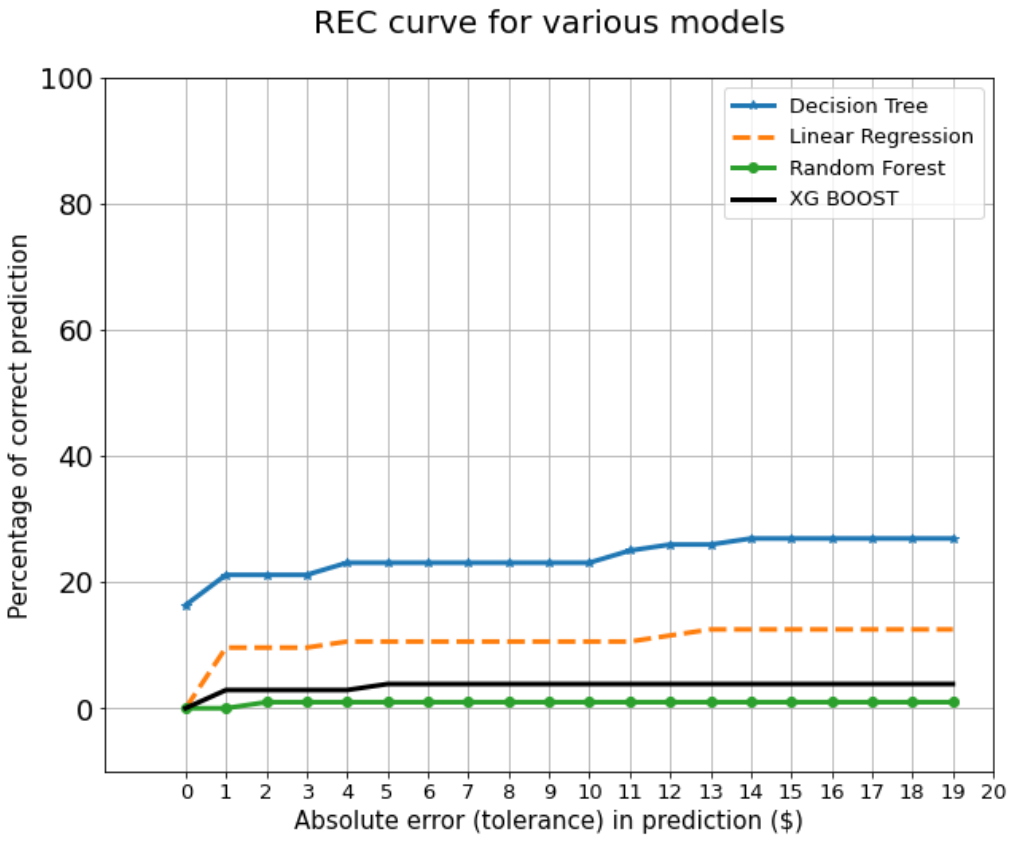
REC Curves: Receiver Operating Characteristic (ROC) curves provide a powerful tool for visualizing and comparing classification results. Regression Error Characteristic (REC) curves generalize ROC curves to regression. REC curves plot the error tolerance on the x-axis versus the percentage of points predicted within the tolerance on the y-axis. The resulting curve estimates the cumulative distribution function of the error. The REC curve visually presents commonly-useds statistics. The area-under-the-curve (AUC) is a biased estimate of the expected error. The R2 value can be estimated using the ratio of the AUC for a given model to the AUC for the null-model. Users can quickly assess the relative merits of many regression functions by examining the relative position of their REC curves. The shape of the curve reveals additional information that can be used to guide modeling.

VI. METRICS

ALGORITHMS	MSE	MAE	R2 score	RMSE
Linear regression	771.9643	17.896	-3.198	_
Random Forest Regressor	1469.77	17.5338	17.5338	38.3376
Decision Tree Regression	969.985	14.453	1.5309	31.1445
XG Boost	6055.0258	21.0036	0.1280	0.1280

VII. RESULTS

Comparison of all the Regressors:



VIII. CONCLUSION

In this model, since our dataset is very small,we don't have to make too many decisions to predict the data. Hence, in this case we can see that Decision tree Regression is better than

Random forest Regression to predict values. But if we don't give the same input sample as that in the csv file, Random Forest gives better predictions since it gets better trained as it has to focus on many decisions. In such cases, the Decision tree gives constant output and fails to give accurate results.

We have to consider these cases as it is really important to understand the behavior of the model changing according to the environment set.

IX. REFERENCES

1- Ahmed M. Elshewey, Amira. A. Elsonbaty, "Forest Fire Detection using Machine Learning Techniques", Journal of Xi'an University of Architecture & Technology, Volume XII, Issue IX,, pp. 510-517, 2020.

2- Paulo Cortez and An´ıbal Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data" ,Article · January 2007, Published on 11 November 2014.

3- Nebot, À.; Mugica, F. "Forest Fire Forecasting Using Fuzzy Logic Models." Forests 2021, 12, 1005, Published: 29 July 2021.

4- Arnida L. Latifah, Ayu Shabrina, Intan N. Wahyuni and Rifki Sadikin, "Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo", 2019 International Conference on Computer, Control, Informatics and its Applications, pp. 4-7,2019.

5- Preeti T, Dr.Suvarna Kanakaraddi, Aishwarya Beelagi, Sumalata Malagi, Aishwarya Sudi, "Forest Fire Prediction Using Machine Learning Techniques", 2021 International Conference on Intelligent Technologies (CONIT) Karnataka, India. June 25-27, 2021.

6- Adithi M. Shrouthy, Syed Matheen Pasha, Yamini S. R. E., Navya Shree S., Lisha U. "Forest fire prediction using ML and AI",International Journal of Advance Research, Ideas and Innovations in Technology, (Volume 7, Issue 3 - V7I3-2136), 2021, pp. 2197-2199, 2021,

7- Aladdin Khaled Al-Zebda, Mutasim Mahmoud Al-Kahlout, Ahmed Mahmoud Abu Ghaly, Donia Zaher Mudawah, "Predicting Forest Fires using Meteorological Data: an ANN Approach",International Journal of Academic Information Systems Research (IJAISR) , Vol. 5 Issue 1, January - 2021, pp: 51-57.

8- Pratima Chaubey , Nidhi J. Yadav , Abhishek Chaurasiya , Prof. Satish Ranbhise, "Forest Fire Prediction System using Machine Learning", (IJRASET), Volume 8 Issue XII Dec 2020,pp. 539-546, 2020.

9- A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdi˜nas, E. Hern´andez-Pereira, M. Inmaculada Paz Andrade, E. Jim´enez, J. Luis Legido Soto, and T. Carballas, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia," Expert systems with applications, vol. 25, no. 4, pp. 545–554, 2003.

10- George E. Sakr, Imad H. Elhajj, George Mitri and Uchechukwu C. Wejinya, "Artificial Intelligence for Forest Fire Prediction", 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics Montréal, Canada, July 6-9, 2010, pp. 1311-1315.

11- T.Niranjan Babu, D.Swetha, V.Charitha, A.J.Stephen, “PREDICTING BURNED AREA OF FOREST FIRES”, International Research Journal of Computer Science (IRJCS), Issue 04, Volume 6 (April 2019), pp. 132-136, 2020.