INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Forest Fire Prediction System using Machine Learning

Pratima Chaubey[1], Nidhi J. Yadav[2], Abhishek Chaurasiya[3], Prof. Satish Ranbhise[4]
[1, 2, 3, 4]*Shree L.R Tiwari College of Engineering, Mumbai University, Maharashtra, Mumbai, India*

Abstract: Forest fires, usually occurring in the forest area or wild land and are uncontrolled fires and cause significant damage to natural and human resource, which are one of the most dangerous disasters to the ecological environment .The Recommended system use various technology like Machine learning techniques and Artificial Intelligence and Wireless network utilized for collecting 24-hour weather data continuously, which provides a high chance to reflect accurately the status of forest environment. Depending on the system, we can decide which days have the high possibility of forest fires and danger and paid special attention to prevent forest fire for forest guards. Forest fire prediction constitutes a significant component of forest fire management. It contains a major role in resource allocation, mitigation and recovery efforts. This system presently analyzed of the 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 predict the hazard of fire occurrence.
Index Terms: wildfires; susceptibility mapping; machine learning; random forest; spatial-cross validation ; correlation and regression

I. INTRODUCTION

Forest fires have become one of the major disasters occurring in recent years. The effects of forest fires have a lasting impact on the environment as it led to deforestation and global warming, which is also one of its major cause of occurrence. Forest fires can be dealt by collecting the satellite images of forest and if there is any emergency caused by the fires then the authorities are notified to neutralize its effects. By the time the authorities get to know about condition, the fires would have already caused a lot of damage to the specific sector. By adopting Data mining and machine learning techniques it can provide an efficient prevention approach where data associated with forests can be used for predicting the places with high possibility of forest fires. Numerous algorithms like Logistic regression, Support Vector Machine, Random forest, K-Nearest neighbors in addition to Bagging and Boosting predictors are used, both with and without Principal Component Analysis (PCA). Among the models in which PCA was applied, Logistic Regression gave the highest F-1 score of 68.26 and among the models where PCA was absent, Gradient boosting gave the highest score of 68.36. Geostationary satellite remote sensing systems are a useful tool for forest fire detection and monitoring because of their high temporal resolution over large areas. These computerized system is capable of capturing, storing, analyzing, and displaying geographically referenced information that is, data identified according to location.

II. PROBLEM STATEMENT

To develop a system that can provide predictions of wildfires using random forest approach and detect this fire in the corresponding map space image using HSV color model. Therefore, the proposed solution is designed to:

A. Train a prediction model using Random forest algorithm.
B. Predict the possibility of wildfire of the given set of attributes.
C. Detect fire in the same map space confirming the presence of wildfire in the predicted area.

Forest Fire are a major environmental issue, creating economic damage and ecological imbalance while threatening human lives. Fast detection is a key element for controlling such phenomenon. To achieve this, one alternative is to use automatic tools based on local sensors, such as provided by meteorological stations.

III. LITERATURE REVIEW

A. Related Work

1) Several techniques for making an intelligent decision were introduced to predict forest fire automatically. Fuzzy reasoning system is one of that is capable of making real time decision and Fuzzy triangular number can better express fuzzy linguistic terms, it has been integrated with multi-attribute decision-making which have been applied in fields such as risk evaluation and performance evaluation. A fuzzy logic algorithm uses five membership functions as temperature, smoke, light, humidity and distance was introduce to forecast the probability of fire. A decision-making tool was developed using fuzzy logic to designate a fuel model for forest-fires, which is subject to surface fire spread techniques to develop a real time fire prediction system.

2) Before applying the machine learning techniques from it is necessary to investigate the case including the provided raw data-set and the existing method provided by forest fire records. Based on conditions, we introduce four simple traditional machine learning methods with their algorithmic descriptions. In addition, necessary data pre-processing is provided. The fourth algorithm is the k-mean clustering algorithm. It is an unsupervised learning mechanism, so the number of clusters and initial centroid point(s) need to be pre-defined. Algorithm 4 shows the mechanism in detail.

3) Data mining is one such efficient approach in which the forest fires can be predicted based on their past occurrences. Data mining requires an authentic and a clean set of data for prediction. If the dataset is not clean or if there are many unknown values then those values must be taken care of before we use them for modeling. The dataset present in the UCI Machine learning packages about the forest fires is used for prediction. proposed a related work to predict the area burned by 2 the forest fires using the dataset. Initially, the feature 'area' was transformed using ln (1+ x) function. Data mining models were applied and fitted. Post-processing was done on the outputs with the inverse of transform. The experiment $\ln(x+1)$ was conducted using 10-fold (cross-validation) x 30 runs. The metrics used there for regression were MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error). Support Vector Machines with Gaussian kernel using 4 features, namely temperature, relative humidity, wind speed and rain, and Naïve mean predictor obtained the best MAD and RMSE values. The results also suggest that SVM predicted small fires with better perfections. proposed a forest fire prediction method based on meteorological data [6]. The results suggest that SVM gave a higher accuracy for a two-class prediction and for a four-class prediction. [4]Convolutional neural network (CNN) is one of the most notable DL approaches and has exhibited robust performance in feature learning for image classification and recognition. It is a feed forward neural network whose parameters are trained by using the classic stochastic gradient descent based on the back-propagation algorithm. Widely, the CNN consists of several building blocks—convolutional, pooling, and fully connected layers. The different types of computing layers play different roles. The convolutional layers, which perform linear convolution operations between the input tensor and a set of filters, output the feature maps. Typically, each feature map is then followed by a nonlinear activation function. The rectified linear unit (ReLU), which performs the nonlinear transformation of the feature map generated by the convolution layer and introduces nonlinearity into the system, is the most used activation function. The function of the convolution operation is to extract different input layer features and achieve weight sharing.

B. Existing System

1) The current system consists of Data Mining and sensor which are capable of sensing the smoke and fire. In effect, meteorological conditions (e.g. temperature, wind) are mostly the cause of forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In this work, we explore a Data Mining (DM) approach to predict the burned area of forest fires.

2) 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 real-world data collected from the northeast region of Portugal. The best configuration uses a SVM and four meteorological inputs (i.e. temperature, relative humidity, rain and wind) and it is capable of predicting the burned area caused by small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management (e.g. prioritizing targets for air tankers and ground crews).

3) Our system consists of high temporal and spatial image to prevent these destructions. Using of Geostationary satellite remote sensing systems which are a useful tool for forest fire detection and monitoring because of their high temporal resolution over large areas. In this, we propose a combined 3-step forest fire detection algorithm (i.e., thresholding, machine learning-based modeling, and post processing) with the help of geostationary satellite.

4) This threshold-based algorithm filtered the forest fire candidate pixels using adaptive threshold values considering the diurnal cycle and seasonality of forest fires while allowing a high rate of false alarms. The random forest (RF) machine learning model then effectively removed the false alarms from the results of the threshold-based algorithm (overall accuracy ~99.16%, probability of detection (POD) ~93.08%, probability of false detection (POFD) ~0.07%, and 96% preventing the false alarmed pixels for validation), and the remaining false alarms were removed through post processing using the forest map. This threshold-based algorithm filtered the forest fire pixels using adaptive threshold values considering the diurnal cycle and seasonality of forest fires while allowing a high rate of false alarms.

5) The random forest (RF) machine learning model then effectively removed the false alarms from the results of the threshold-based algorithm (overall accuracy

6) ~99.16%, probability of detection (POD) ~93.08%, probability of false detection (POFD) ~0.07%, and 96% reduction of the false alarmed pixels for validation), and the remaining false alarms were removed through post processing using the forest map.

IV. METHODOLOGY

In this project we tried to make a prediction for the burned area within the Montesinho park. Forest Fires Data Set was used for this analysis. The data was clusterized. 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.

A. Method 1: Linear Regression

The problem here is modelled into a Regression task since over motive is to predict the area of the land burnt.The variable that we need to predict is in numerical value.

The regression allows us to model mathematically the relationship between two or more variable. This linear regression model is used to find whether there is a positive or negative relationship between the variables.

Normally a regression equation is Y(Dependent variable) = a (intercept) + b (slope of the line) * X (Independent or explanatory variable)

B. Method 2: Gradient boosting

Gradient boosting is a technique for producing regression models consisting of collections of regressor. It is an ensemble algorithm where the repressor predictions are combined usually by some sort of weighted average or vote in order to provide an over all prediction.

Boosting is a method in which learners are learned sequentially with early learners fitting simple models to the data and analyzing the data for errors .

C. Method 3: Bagging

Bagging is a classic ensemble method known as bootstrap aggregation. Bagging algorithm consist of many classifiers each uses only some portions of data in each iterations and then combining them through a model averaging techniques. The idea behind this is reduce the over fitting in the class of models. The bootstrap method in bagging creates a random subset of data from a given dataset by sampling

D. Method 4: Random forest

Random forest is a powerful algorithm which can be used for both regression and classification. The algorithm first creates bootstrap samples from the original data. A regression tree is developed from the each bootstrap samples. Then it randomly sample the number of predictors and the best split is chosen from the variables. Now the aggregation method is used for predicting the new data

E. Method 5: SVM regression

SVM looks at the extremes of the dataset and draws a decision boundary known as hyper plane near the extreme points in the dataset. It is a method which uses epsilon loss function and performs linear regression in high dimensional space. The SVM always follows a kernel trick where we can use different kernels like RBF, linear, polynomial, Sigmoid

F. Method 6: Logistic regression

This method uses sigmoid function where we get the probability of the given event. It just assign probability to every single event of area burned. This produces logistic coefficients from where we can find the probability of present or absent. The data should be distributed and there should be some relationship between the attributes in the data
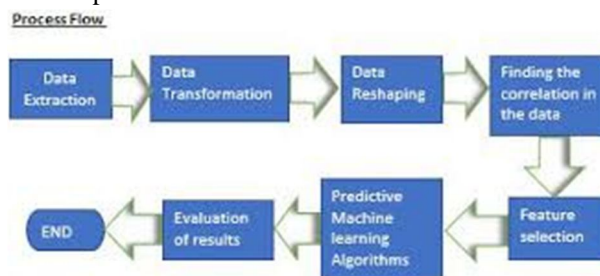

Figure 1: Process flow

V. IMPLEMTATION

The implementation of the linear regression, ridge regression, and lasso regression algorithms are done using the Jupyter Notebook .Jupyter notebook helps to write and execute Python in the browser, where it is open-source and widely used for the implementation of machine learning algorithms such as regression, classification, and clustering.

A. Data Extraction

The data is extracted from the UCI machine learning repository. The data consist of meteorological. .FWI system data and amount of area burned during fires over a period of 2000-2003 in Montesinho park in Portugal. The factors that mainly affect the forest fire are the climatic conditions of the forest. The data set has clear description of the climatic conditions such as Relative humidity, temperature of the forest, Wind speed and rainfall in the forest. These data is collected from the local sensors with are available in the Portugal. The Portugal has around 162 weather stations so getting this data is not a big deal. The FWI system is the which is widely used as a fire danger rating system. The data also contains the day , month and X and Y axis values where the fire occurred. The getting the day and month we can separate the fires into week day and weekend. The next FWI data is like moisture code, Fire index, Drought code and spread index which are mainly depend on the weather conditions. These values calculated by the FWI system is a direct indicator of the fire intensity. The Relative humidity value is a changing one because it will be high in the morning and keep reducing to the minimum value as hours past. The wind speed is a major factor since it can make the fire to spread rapidly. From looking at the data we can say when the wind speed is around 15/hr the chances of fire is high. One of the most important feature in the dataset is temperature of the forest which can cause fire.

B. Clustering Data

First, the coordinates were clusterized. The cluster amount was chosen using the elbow method. The K-Elbow Visualizer implements the "elbow" method of selecting the optimal number of clusters for K-means clustering. K-means is a simple unsupervised machine learning algorithm that groups data into a specified number (k) of clusters. Because the user must specify in advance what k to choose, the algorithm is somewhat naive – it assigns all members to k clusters even if that is not the right k for the dataset. The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters.
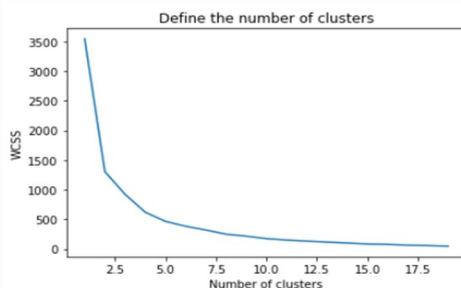

Figure 2: Define the number of cluster data

There is last bend somewhere near fifth point, and then the curve is more smoothed. So, as we can see, the optimal number of clusters is 5 So, kmeans algorithm with the same configurations was applied to find the clusters.
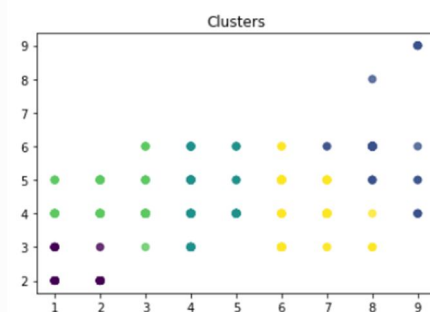

Figure 3: Clusters

And for each cluster, the burned area prediction was found using regression method in machine learning. There is really small correlation values between data and dependent variable, so the stepwise regression methods was applied to chose the best predictors.

C. Finding the correlation

The correlation matrix is used to find which attribute has a significant correlation on the output target variable. The correlation ranges from Negative and positive. If the correlation is zero then there is no relationship between two attributes. We can see that the temperature has more positive correlation with the area burnt. The wind has more negative correlation with the output variable. The correlation of each cluster are given below.
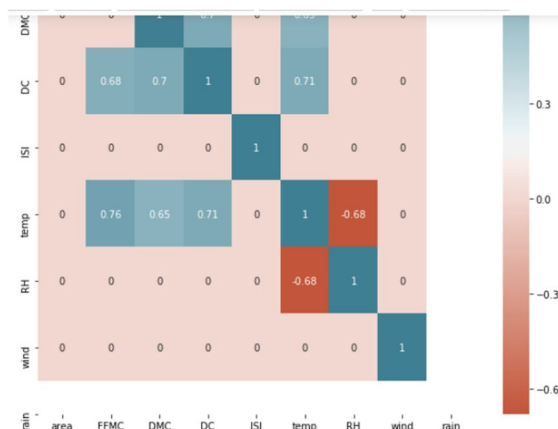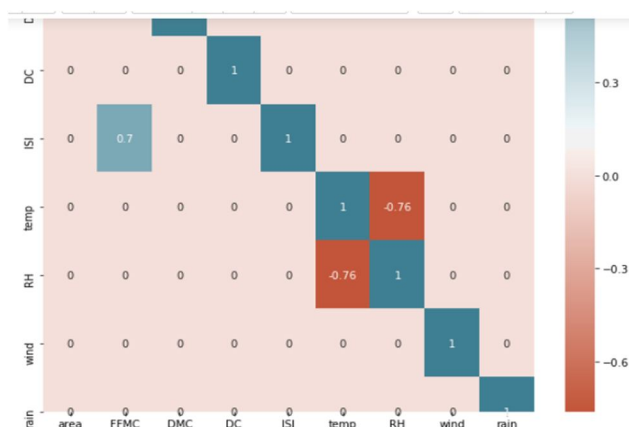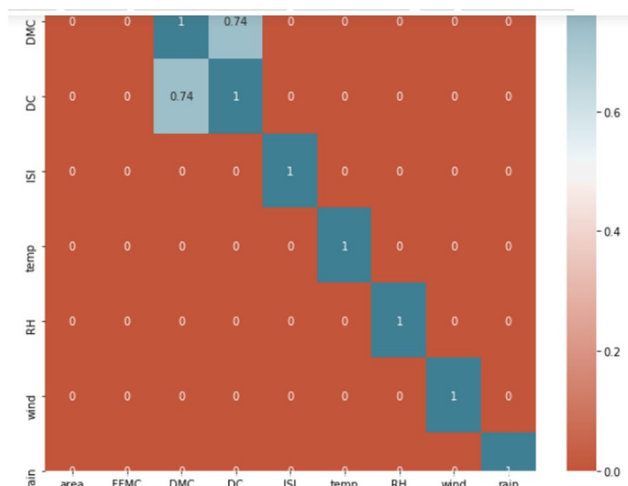

Figure 4: cluster 0

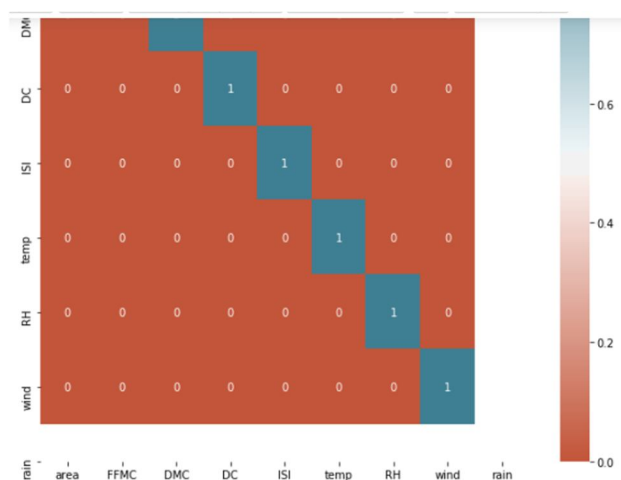Figure 5 :cluster 1

Figure 6: cluster 2
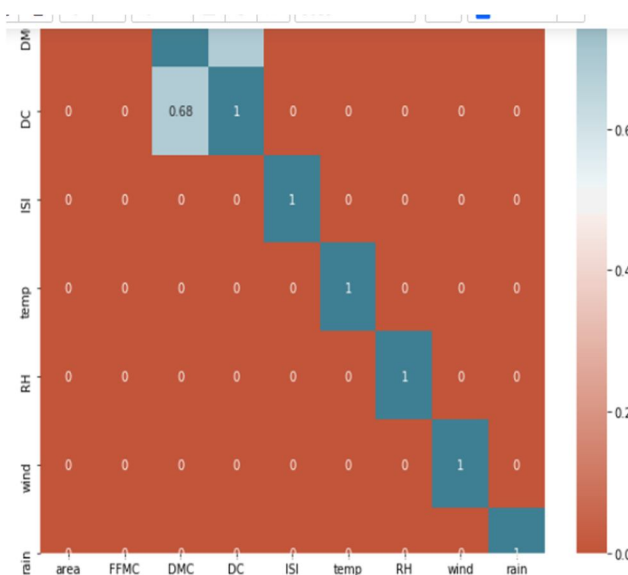

Figure 7: cluster 3

Figure 8: cluster 4

D. Training the model

There is no predictor was chose by forward selection. The reason of that are un-significant correlation values between the data, because the forward selection train the model using each predictor separately, so it is hard to choose really significant results. So the backward elimination algorithm was applied, to train the model using all predictors, and then choose the best one.
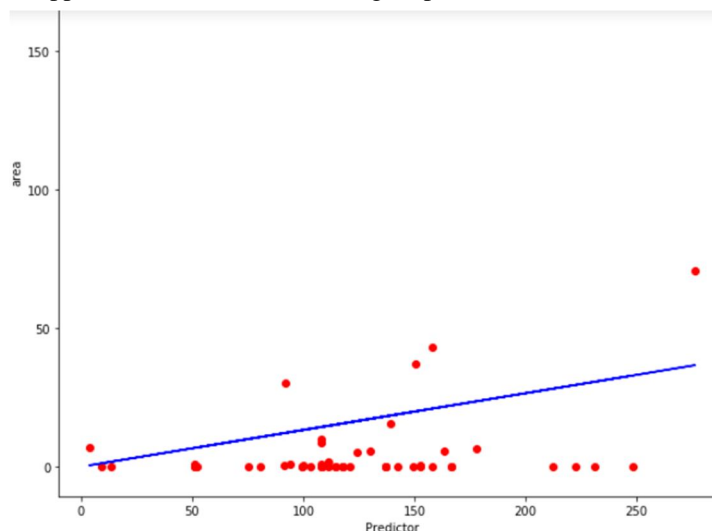

Figure 9: Predictor

Figure 10: Regression Result

The DMC predictor was chosen with r squared 14 %.So, the best model for this cluster is area = 0.1324 * DMC

E. Testing the Prediction Of each Cluster

We have created total 5 cluster named as cluster 0, cluster 1, cluster 2, cluster 3, cluster 4 each cluster data are trained using regression model and the features selected are used to find which attributes we need to use in our prediction model. We find the attributes Temperature and RH are the are the best features for doing a prediction model. By including the attributes like temperature,RH,DMC and DC in our prediction model and by splitting into training and testing set help us to get better prediction accuracy By doing this the overfitting of the model is reduced. The training time of the model is reduced because we are eliminating the attributes that are less contributing to the output variable(Area burnt).Here we are using random forest classifiers for the feature ranking of the attributes. Before building a machine learning algorithm we need to split the data into Train and Test.This split is used to validate the model. The training part is used to create a model and testing part is used to verify the model created.Here the data split is done by separating 70 % to the test data and 30 % to the train data. Then a standard scalar function is introduced on the training set. Standardization of the dataset is the common usage for many machine learning algorithms. The standard transform is applied to both test and training set. Now the these data is loaded into a dataframe.Now the prediction models are implemented using this dataFrame.A confusion matrix is build from which we can calculate True positive.

F. Evaluation

In the evaluation section we will be selecting the best model in terms of Accuracy Various results of the predictive models . The best model for this cluster is backward elimination algorithm result: area = 0.9878 * temp

VI. CONCLUSION AND FUTURE WORK

Forest fires are an environmental issue, creating economic damage and ecological distortion while threatening human lives. Fast detection is a key element for controlling such phenomenon. Using satellite data and machine learning methods the monitors are able to detect the natural disasters in real time. The world is moving towards automation and this Big data era urges us to build more solutions for the complex problems. In this project with all the use of Big data and machine learning techniques we build a model for the prediction of area burned during the forest fires. This model should be incorporated in all the areas which have more probability of capturing fire. By further tuning the parameters and by adding some other attributes like vegetation of the forest, Forest cover, type of trees in the forest and Buildup Index we can improve the Accuracy of random forest and Boosting algorithm. This project mostly aims in developing a predictive model with the usage of climatic conditions. By detecting the area burned we can separate the fires into small and large. This classification of fires helps the FMS team to send adequate crews and air tankers to the following danger zone. The future work in this project can be done by creating a probabilistic models that can identify the origin of fire by using some conditions. Those probabilistic models should be integrated with the model provided in this study to handle more risky conditions in the case of large or big fires. The use of GIS data and satellite view can also be included with this model which provides better accuracy

REFERENCES

[1] Pradeep Kumar Singh, Amit Sharma, "An insight forest fire detection techniques using wireless sensor networks", Signal Processing Computing and Control (ISPCC)2017 4th International Conference on, pp 647653, 2018.

[2] Diwakar Pant, Sandeep Verma, Piyush Dhulia, "A study on disaster detection and management using WSN in Himalayan region of Uttarakhand", Advances in Computing Communication & Automation (ICACCA)(Fall)2018 3rd International Conference on, pp. 1-6 2018.

[3] Evizal Abdul Kadir, Sri Listia Rosa, An Yulianti, "Application of WSNs for Detection Land and Forest Fire in Riau Province Indonesia", Electrical Engineering and Computer Science (ICECOS)2018 International Conference on pp. 25-28, 2019.

[4] George E. Sakr, Imad H. Elhajj, George Mitri and Uche Chukwu C. Wejinya "Artificial Intelligence for Forest Fire Prediction", Advanced Intelligent Mechatronics ,Montréal, Canada, July 6-9, 2019

[5] Hanchao Li, Xiang Fei, "Study on Most Important Factor and Most Vulnerable Location for A Forest Fire Case Using Various Machine Learning Techniques", Electrical Engineering and Computer Science (ICECOS)2019 International Conference on pp. 25-28, 2019.

[6] R. Rishikesh, A. Shahina, A. Nayeemulla Khan "Predicting Forest Fires using Supervised and Ensemble Machine Learning Algorithms", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019

[7] Taylor, S. W. and Alexander, M. E. (2006). Science, technology, and human factors in fire danger rating: the Canadian experience., International Journal of Wild land Fire 15(1): 121–135.

[8] Stojanova, D., Panov, P., Kobler, A., Dˇzeroski, S. and Taˇskova, K. (2006). Learning to predict forest fires with different data mining techniques, Conference on Data Mining and Data Warehouses (SiKDD 2006), Ljubljana, Slovenia, pp. 255–258.

[9] Stocks, B. J., Lynham, T., Lawson, B., Alexander, M., Wagner, C. V., McAlpine, R. and Dube, D. (1989). Canadian forest fire danger rating system: an overview, The Forestry Chronicle 65(4): 258–265.

[10] Ozbayoˇglu, A. M. and Bozer, R. (2012). Estimation of the burned area in forest fires ¨ using computational intelligence techniques, Procedia Computer Science 12: 282–287

[11] Boubeta, M., Lombard´ıa, M. J., Gonz´alez-Manteiga, W. and Marey-P´erez, M. F. (2016). Burned area prediction with semiparametric models, International Journal of Wildland Fire 25(6): 669–678.

[12] Ammann, H., Blaisdell, R., Lipsett, M., Stone, S. L., Therriault, S., Jenkins, J. and Lynch, K. (2001). Wildfire smoke: a guide for public health officials, California Air Resources Board. http://www. arb. ca. gov/smp/progdev/pubeduc/wfgv8. pdf (accessed 06/02/08) .

.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)