Predicting Forest Fires

Prashi Doval
prashi012mca18@igdtuw.ac.in

Shruti Mishra
shruti005mca18@igdtuw.ac.in

Abhipriya Sharma
abhipriya004mca18@igdtuw.ac.in

May 25, 2020

Abstract

Fires have been a major contributing factor in the loss of our forests and ecosystem worldwide and also affect human animal lives and their habitat but when fires burn too hot and uncontrollable or when they're in the the places where woodlands and homes or other developed areas meet, they can damage which could be life threatening. We are predicting the burned area of forest fires, in the northeast region of Portugal on the basis of spatial, temporal, and weather variables where the fire is spotted. Thus, early detection of forest fires is of paramount importance. But methods that give us accurate predictive results for the same are not available.

Contents

1 Introduction

Forest fires have been some of the oldest threats to our nature. A forest fire destroys not only the flora and fauna, but also disrupts the ecological cycle of the area.Wildfires can be awfully threatening,as they cause of loss of not only the trees and the green cover of the area but also result in loss of wildlife, charred, damaged soil, loss of houses and other structures and smoke and respiratory diseases. Forest fires are being a common occurrence these days, like the (2019) Amazon Rainforest wildfires, Brazil; (2020) Australia Bushfires, (resulted in loss of 46.03 million acres) etc have been ravaging the lands of our planet causing huge losses.

1.1 Problem Statement & Objectives

Forest Fires have been a major contributing factor in the loss of our forests ecosystem.

- Thus, early prediction of forest fires is of paramount importance to prevent huge losses to our eco-system , human and wild life and habitat.
- Also, the available methods which give us predictive results for the same are not available or outdated.

1.2 Motivation

The current methods concentrate more on the cure rather than the prevention as they only try and control the fire from spreading further, after it has occurred, hence causing immense destruction. Hence, methods to predict or detect forest fires beforehand are urgently required to prevent more such incidents from happening.

These methods will directly benefit the green cover, wildlife (the whole eco-system) and the people associated or dependent on the forests for their livelihood. Indirectly, benefits the whole government as it prevents loss of habitat, housing structures and respiratory problems, in short prevents a global disaster from occuring.

1.3 Scope and Limitation

Fire occurrence estimation by modeling the relations between fire threat and the influence factors. Our scope will be limited to the prediction and estimation of possible occurrences of forest fires depending on the influential geographical and partial factors of the area.

1.4 Organization of Report

Section 2 involves the details of the data used for the research i.e. description of the fields in the dataset used and methods used for cleaning the data i.e. the outliers. In Section 3, we explain the experimental setup and the results. Section 4 gives the conclusions and the future work.

2 Methodology

The methodology used is explained in the sub-sections.

2.1 Dataset Description

The Data is taken from the online source.The dataset consists of the following features.

Attribute	Details
X	coordinate for the area within the park where the analysis has been done.
Y	coordinate for the area within the park where the analysis has been done.
FFMC	Fine Fuel Moisture Code,the moisture content of litter and cured fine fuels.
DMC	Duff Moisture Code, moisture content of the forest floor within a medium depth.
DC	Drought code which is the moisture content of deep compact organic layers.
ISI	Initial Spread Index i.e. the head fire indicator and the rate of fire spread.
Temp	Temperature of the area.
RH	Relative Humidity of the area.
Wind	Speed of wind in that area.
Rain	Amount of rainfall in that area.
Area	The size of the area.
FIRE	newly added attribute derived from the 'area' feature

Table 1: Details of the dataset.

FFMC is the numeric rating of the moisture content of litter and cured fine fuels. DMC is the moisture content of the forest floor within a medium depth. FIRE is a newly added attribute derived from the 'area' feature where the values of 0 in the column 'area' corresponds to 0 in column 'fire' and values above 0 corresponds to 1 in the column 'fire', denoting the presence of forest fires.

2.2 Data Pre-processing

Various data cleaning and exploration methods are used.In some cases data is not of high quality. The data is checked for possible outliers and null values, categorical columns analysis and many more. Methods like skew and kurtosis is used in order to check symmetry and distribution of the dataset.Below is given the snapshot of the data.

	count	mean	std	min	25%	50%	75%	max
X	517.0	4.669246	2.313778	1.0	3.0	4.00	7.00	9.00
Y	517.0	4.299807	1.229900	2.0	4.0	4.00	5.00	9.00
FFMC	517.0	90.644681	5.520111	18.7	90.2	91.60	92.90	96.20
DMC	517.0	110.872340	64.046482	1.1	68.6	108.30	142.40	291.30
DC	517.0	547.940039	248.066192	7.9	437.7	664.20	713.90	860.60
ISI	517.0	9.021663	4.559477	0.0	6.5	8.40	10.80	56.10
temp	517.0	18.889168	5.806625	2.2	15.5	19.30	22.80	33.30
RH	517.0	44.288201	16.317469	15.0	33.0	42.00	53.00	100.00
wind	517.0	4.017602	1.791653	0.4	2.7	4.00	4.90	9.40
rain	517.0	0.021663	0.295959	0.0	0.0	0.00	0.00	6.40
area	517.0	12.847292	63.655818	0.0	0.0	0.52	6.57	1090.84

Figure 1: Dataset

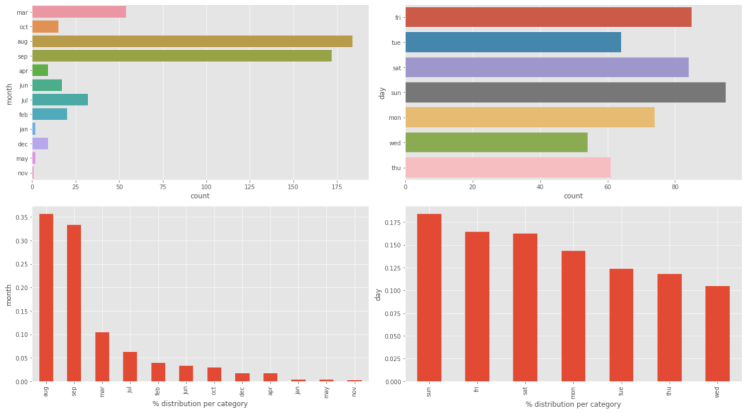


Figure 2: Categorical data w.r.t month and day

The figure below shows outliers in respective attributes of the data using skew and kurtosis method. It is found that many attributes like FFMC and DMC consist of good amount of outliers.

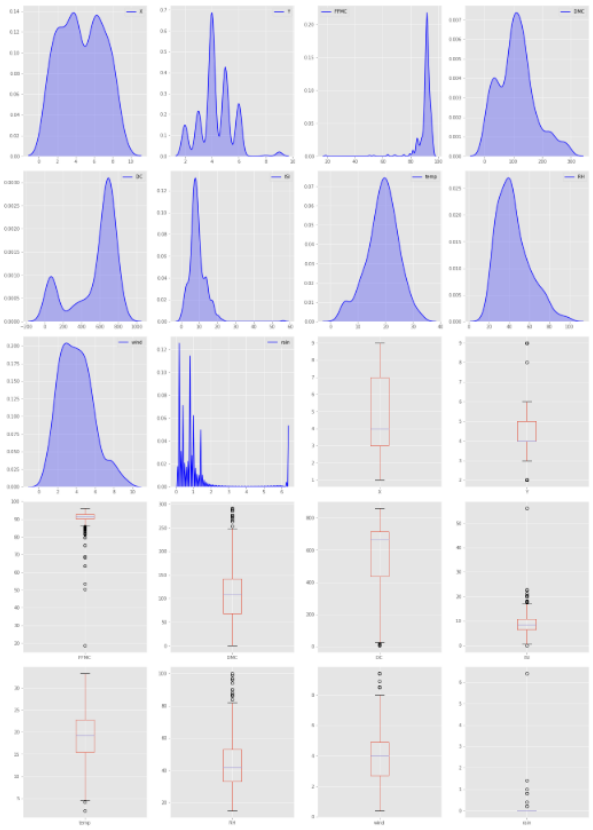
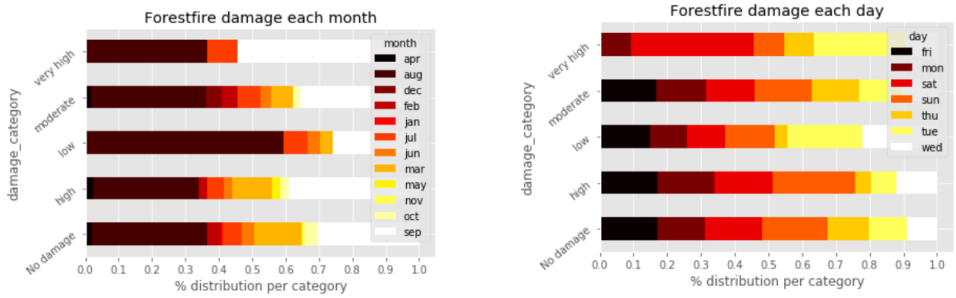


Figure 3: Outlier Detection

Next, new attribute named 'damage-category' is added to divide the data into damage categories namely No damage, low, moderate, high and very high. It is found after bivariate analysis of categorical columns that there is very high damage in the month of august, july and september.



(a) damage category w.r.t month (b) damage category w.r.t day

Figure 4: Damage category for each month and day

After performing bivariate analysis,features including X,Y, FFMC, DMC, DC, ISI, temp, RH, wind, rain, area are selected for multivariate analysis of data.

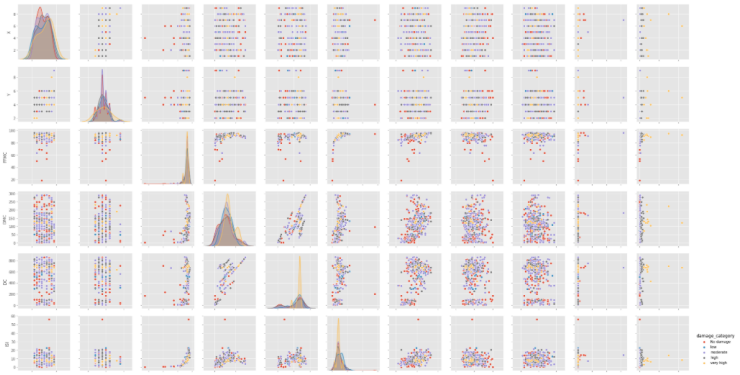


Figure 5: A snapshot of Multivariate Data Analysis

More Outliers are detected in the above step performed and data transformation is being done. To remove the outliers one such method used is zscore method. After pre-processing of data,various Machine learning algorithms applied on data which are shown in the next section.

2.3 Proposed Methodology

Algorithms and techniques:

Regression: It is the set of processes for estimating the relationships between different variables used in the analysis. It focuses on the relationships between one dependent variable and one or more independent variables. The following regression algorithms have been used in the research:

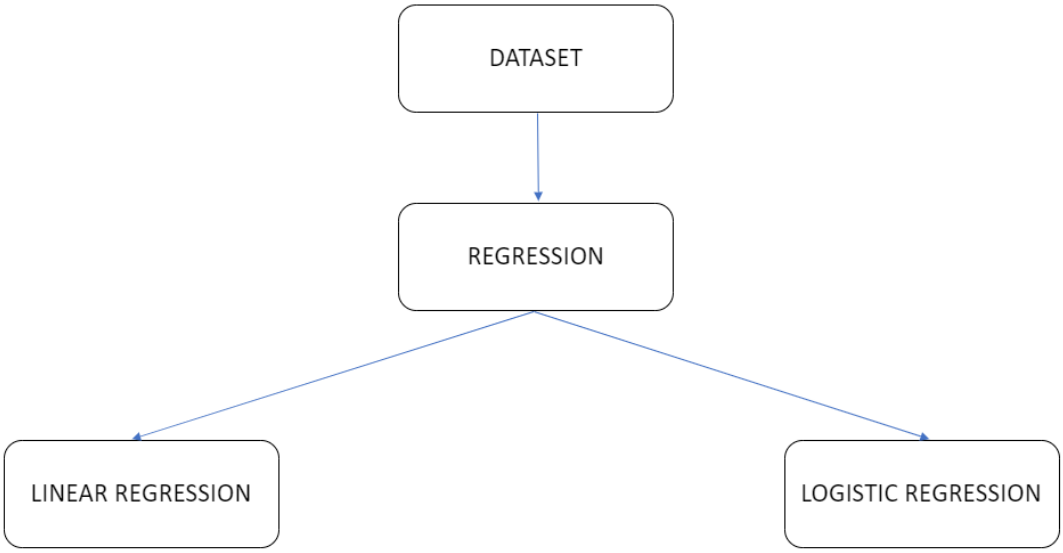


Figure 6: Regression Flowchart

- **Linear Regression(Supervised Learning/Regression)**: The simplest form of regression, linear regression is used to understand the relationship between two continuous variables.It involves finding the line, that most closely fits the data according to a specific mathematical criterion.

- **Logistic Regression (Supervised Learning/Regression)** : Logistic Regression is a machine learning method used for modeling a binary dependent variable. It is a form of binomial regression.The dependent variable takes a binary form – 1 or 0, yes or no. The relationship between the dependent variable and the independent variable helps it to predict the target variable. It uses sigmoid function to determine their probability and map them to some discrete values. The sigmoid function is as follows:-

$$\frac{1}{1+e^{-}z} = f(x) \tag{1}$$

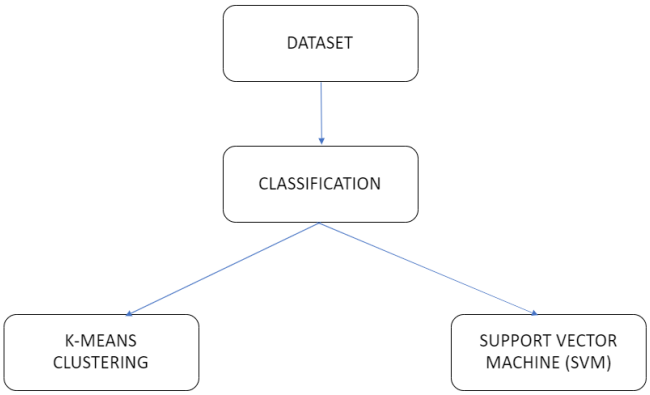


Figure 7: Classification Flowchart

Classification : Classification is a task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. It specifies the class to which data elements belong to and is best used when the output has finite and discrete values. It predicts a class for an input variable as well.

- **Support Vector Machine Algorithm (SVM) (Supervised Learning)**: The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.It can be used for both regression and classification purposes.
- **K-Means(Unsupervised Learning/Clustering)**: that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. K-means algorithm only converge to local minima of the minimum-sum-of-squares clustering problem defined as:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 . \tag{2}$$

Figure 8: k-means Clusters

3 Experiment Setup & Results

Algorithm	Accuracy
SVM linear Classifier	1.00
SVM Gaussian Classifier	0.56
SVM Sigmoid Classifier	0.47
Linear Regression	0.0769
K-Means Clustering	0.6426
Logistic Regression	0.5823

Table 2: Accuracy of Algorithms

4 Conclusion and Future Work

Our results showed that SVM (Support Vector Machine Algorithm) linear classification is the best model to predict the burn area of the forest. Further, we would like to implement deep learning algorithms and compare various classifiers and explore the most accurate forest fire prediction model.

References

1. Daniela Stojanova, Pance Panov, Andrej Kobler, Saso Dzeroski, Katerina Taskova : "Learning to predict forest fires using different data mining techniques",2006

2. T. Niranjan Babu, D. Swetha, V. Charitha , A.J. Stephen: "Predicting burnt area of forest fires", 2019 (IRJCS), Vol VI, 132-136

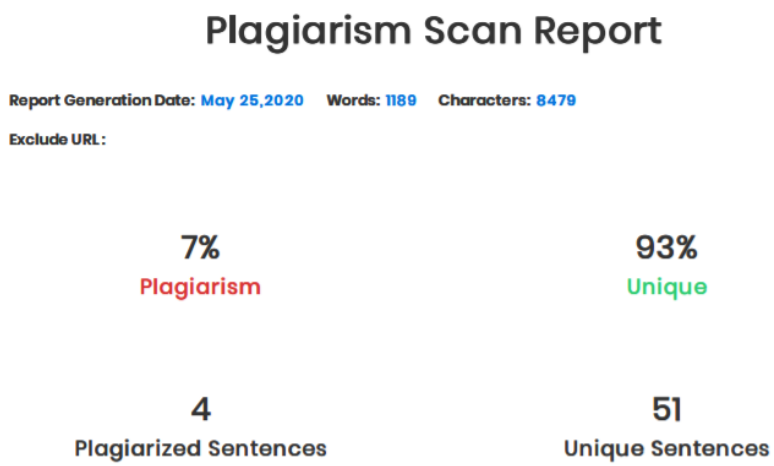
3. Anupam Mittal, Geetika Sharma, Ruchi Aggarwal: "Forest Fire Detection using various Machine Learning Techniques using Mobile agent in WSN" ,2016, IRJET, Vol III, Issue VI

4. Guoli Zhang, Ming Wang, Kai Liu: " Forest Fire Susceptibility Modeling using a Convolutional Neural Network for Yunnan Province of China", 2019, IJDRS

5. Paolo Cortez and Anibal Morais: "A Data Mining Approach to Predict Forest Fires using Meteorological Data

6. R. Rishickesh, A. Shahina, A. Nayeemulla Khan: "Predicting Forest Fires using Supervised and Ensemble Machine Learning Algorithms", 2019, IRJTE, Vol 8

5 Appendix: Similarity Report



Plagiarism Scan Report

Report Generation Date: May 25,2020 **Words:** 1189 **Characters:** 8479

Exclude URL:

7%	93%
Plagiarism	Unique
4	51
Plagiarized Sentences	Unique Sentences

Content Checked for Plagiarism

Fires have been a major contributing factor in the loss of our forests and ecosystemworldwide and also affect human animal lives and their habitat but when fires burntoo hot and uncontrollable or when they're in the the places where woodlands andhomes or other developed areas meet, they can damage which could be life threaten-ing. We are predicting the burned area of forest fires, in the northeast region of Portugalon the basis of spatial, temporal, and weather variables where the fire is spotted. Thus,early detection of forest fires is of paramount importance. But methods that give usaccurate predictive results for the same are not available.d.1.2 MotivationThe current methods concentrate more on the cure rather than the prevention as they onlytry and control the fire from spreading further, after it has occurred, hence causing im-mense destruction. Hence, methods to predict or detect forest fires beforehand are urgentlyrequired to prevent more such incidents from happening.These methods will directly benefit the green cover, wildlife (the whole eco-system)and the people associated or dependent on the forests for their livelihood. Indirectly, bene-fits the whole government as it prevents loss of habitat, housing structures and respiratoryproblems, in short prevents a global disaster from occuring.1.3 Scope and LimitationFire occurrence estimation by modeling the relations between fire threat and the influencefactors. Our scope will be limited to the