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 prediction and estimation of possible occurrencesof forest fires depending on the influential geographical and partial factors of the area.1.4 Organization of ReportSection 2 involves the details of the data used for the research i.e. description of the fieldsin 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 thefuture work.1 2 MethodologyThe methodology used is explained in the sub-sections.2.1 Dataset DescriptionThe Data is taken from the online source.The dataset consists of the following features.AttributeDetailsXcoordinate for the area within the park where the analysis has been done.Ycoordinate for the area within the park where the analysis has been done.FFMCFine Fuel Moisture Code,the moisture content of litter and cured fine fuels.DMCDuff Moisture Code, moisture content of the forest floor within a medium depth.DCDrought code which is the moisture content of deep compact organic layers.ISIInitial Spread Index i.e. the head fire indicator and the rate of fire spread.TempTemperature of the area.RHRelative Humidity of the area.WindSpeed of wind in that area.RainAmount of rainfall in that area.AreaThe size of the area.FIREnewly added attribute derived from the 'area' featureTable 1: Details of the dataset.FFMC is the numeric rating of the moisture content of litter and cured fine fuels. DMCis the moisture content of the forest floor within a medium depth. FIRE is a newly addedattribute derived from the 'area' feature where the values of 0 in the column 'area' cor-responds 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-processingVarious data cleaning and exploration methods are used.In some cases data is not of highquality. The data is checked for possible outliers and null values, categorical columns anal-ysis and many more. Methods like skew and kurtosis is used in order to check symmetryand distribution of the dataset.Below is given the snapshot of the data.Figure 1: Dataset2 Figure 2: Categorical data w.r.t month and dayThe figure below shows outliers in respective attributes of the data using skew andkurtosis method. It is found that many attributes like FFMC and DMC consist of goodamount of outliers.Figure 3: Outlier DetectionNext, new attribute named ’damage-category’ is added to divide the data into damagecategories namely No damage, low, moderate, high and very high. It is found after bivari-ate analysis of categorical columns that there is very high damage in the month of august,july and september.3 (a) damage category w.r.t month(b) damage category w.r.t dayFigure 4: Damage category for each month and dayAfter 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 AnalysisMore Outliers are detected in the above step performed and data transformation isbeing 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 shownin the next section.2.3 Proposed MethodologyAlgorithms and techniques:Regression:It is the set of processes for estimating the relationships between differentvariables used in the analysis. It focuses on the relationships between one dependent vari-able and one or more independent variables. The following regression algorithms havebeen used in the research:4 Figure 6: Regression Flowchart•Linear Regression(Supervised Learning/Regression): The simplest form of regres-sion, linear regression is used to understand the relationship between two continu-ous variables.It involves finding the line, that most closely fits the data according toa specific mathematical criterion.•Logistic Regression (Supervised Learning/Regression): Logistic Regression is amachine learning method used for modeling a binary dependent variable. It is aform of binomial regression.The dependent variable takes a binary form – 1 or 0,yes or no. The relationship between the dependent variable and the independentvariable helps it to predict the target variable. It uses sigmoid function to determinetheir probability and map them to some discrete values. The sigmoid function is asfollows:-11 +e−z=f(x)(1)5 Figure 7: Classification FlowchartClassification :Classification is a task that requires the use of machine learning algo-rithms that learn how to assign a class label to examples from the problem domain. Itspecifies the class to which data elements belong to and is best used when the output hasfinite and discrete values. It predicts a class for an input variable as well.•Support Vector Machine Algorithm (SVM) (Supervised Learning): The objectiveof the support vector machine algorithm is to find a hyperplane in an N-dimensionalspace(N — the number of features) that distinctly classifies the data points.It can beused for both regression and classification purposes.•K-Means(Unsupervised Learning/Clustering): that aims to partition n observationsinto 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-meansalgorithm only converge to local minima of the minimum-sum-of-squares clusteringproblem defined as:arg minSk∑i=1∑x□Sill x−μill2.(2)Figure 8: k-means Clusters6 3 Experiment Setup & ResultsAlgorithmAccuracySVM linear Classifier1.00SVM Gaussian Classifier0.56SVM Sigmoid Classifier0.47Linear Regression0.0769K-Means Clustering0.6426Logistic Regression0.5823Table 2: Accuracy of Algorithms4 Conclusion and Future WorkOur results showed that SVM (Support Vector Machine Algorithm) linear classification isthe best model to predict the burn area of the forest. Further, we would like to implementdeep learning algorithms and compare various classifiers and explore the most accurateforest fire prediction model.7 References1. Daniela Stojanova, Pance Panov, Andrej Kobler, Saso Dzeroski, Katerina Taskova :”Learning to predict forest fires using different data mining techniques”,20062. T. Niranjan Babu, D. Swetha, V. Charitha , A.J. Stephen: ”Predicting burnt area offorest fires”, 2019 (IRJCS), Vol VI, 132-1363. Anupam Mittal, Geetika Sharma, Ruchi Aggarwal: ”Forest Fire Detection using var-ious Machine Learning Techniques using Mobile agent in WSN” ,2016, IRJET, Vol III,Issue VI4. Guoli Zhang, Ming Wang, Kai Liu: ” Forest Fire Susceptibility Modeling using aConvolutional Neural Network for Yunnan Province of China”, 2019, IJDRS5. Paolo Cortez and Anibal Morais: ”A Data Mining Approach to Predict Forest Firesusing Meteorological Data6. R. Rishickesh, A. Shahina, A. Nayeemulla Khan: ”Predicting Forest Fires using Su-pervised and Ensemble Machine Learning Algorithms”, 2019, IRJTE, Vol 8

Sources

Jan 3, 2015 - In Section 3, we explain the experimental setup and the results of Study 1 (trust game), which is the main focus of the paper. In Section 4, we ... **7%**

http://www.pwl.uni-koeln.de/sites/personalwirtschaftslehre/team/publications/Social_Preferences_Finance.pdf

Predicting Forest Fires using Supervised and ... - IJRTE

Jul 15, 2019 - to predict the target variable. Logistic regression uses sigmoid function to determine their probability and map them to some discrete values. **3%**

https://www.ijrte.org/wp-content/uploads/papers/v8i2/B2878078219.pdf

Classification - Machine Learning | Simplilearn

Classification is a type of supervised learning. It specifies the class to which data elements belong to and is best used when the output has finite and discrete values. **3%**

https://www.simplilearn.com/classification-machine-learning-tutorial

k-means clustering - Wikipedia

k-means clustering is a method of vector quantization, originally from signal processing, 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. **3%**

https://en.wikipedia.org/wiki/K-means_clustering