Predicting Future Forest Fires Using Meteorological Data

Dillon Alterio Benjamin Byers

Southern Connecticut State University Southern Connecticut State University

501 Crescent Street, New Haven, CT 06515 501 Crescent Street, New Haven, CT 06515

*alteriod1@southernct.edu byersb1@southernct.edu*

Isaiah Brooks

Southern Connecticut State University

501 Crescent Street, New Haven, CT 06515

*brooksi2@southernct.edu*

***Introduction:***

**Problem Statement:** How to prevent and determine when and where the next forest fire will take place by using meteorological data for the northeast region of Portugal.

**Objective:** A key component of preventing forest fires is early intervention and detection. This can be achieved by conducting surveillance with early detection methods. These methods include using the meteorological data that we have available to us to help predict and prevent any potential future wildfires.

**Motivation:** Forest fires are a very costly and dangerous event. They can cause massive amounts of destruction to not just the forest, but to the wildlife and to people’s property. Having a method in which we are able to predict future events is vital to the longevity of forest in all regions of the world. One of the biggest culprits in how wildfires are formed is meteorological events. Some of these events include things like relative humidity, atmospheric conditions, strong winds, and lighting storms. Using meteorological data, we will have the capability to predict where and when the next possible forest fire might occur in the northeast region of Portugal.

**Related Work:** A student at the University of Minho created a data mining approach to predict forest fires by using given meteorological data. He conducted a similar analysis in regards to meteorological data forest fire predictions

**Data**

The dataset that was used in our experiment is from UCI Machine learning repository <https://archive.ics.uci.edu/ml/datasets/Forest+Fires>. Our dataset is called Forest fires and the data show us the data of forest fires that happen in Montesinho natural park. They were trying to predict forest fires using the artificial intelligence and Meteorological data, The forest Fire Weather Index (FWI) is the Canadian system for rating fire danger and it includes six components. Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI) and FWI. The first three are related to fuel codes: the FFMC denotes the moisture content surface litter and influences ignition and fire spread, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect fire intensity. The ISI is a score that correlates with fire velocity spread, while BUI represents the amount of available fuel. The FWI index is an indicator of fire intensity, and it combines the two previous components. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions: 16 hours for FFMC, 12 days for DMC and 52 days for DC. The dataset has 13 attributes and no missing values. We also have the excel spreadsheet with the data on it.Graphical user interface, application, table, Excel

Description automatically generated

**Data & Processing**

The program will be written in Python using Scikit-learn using the Anaconda interpreter in a PyCharm IDE. We going to use that access the data.

We going to use the data and use histogram graphs to figure out what temperature cause the most forest fires.

Since the dataset is regression dataset were going to use some data models from the experiment . The overall performance is computed by a global metric, namely the Mean Absolute Deviation (MAD) and Root Mean Squared (RMSE), which can be computed as [27]: MAD = 1/N × PN i=1 |yi − ybi | RMSE = qPN i=1 (yi − ybi) 2/N (1) . This supports the SVM algorithm which is used for regression datasets

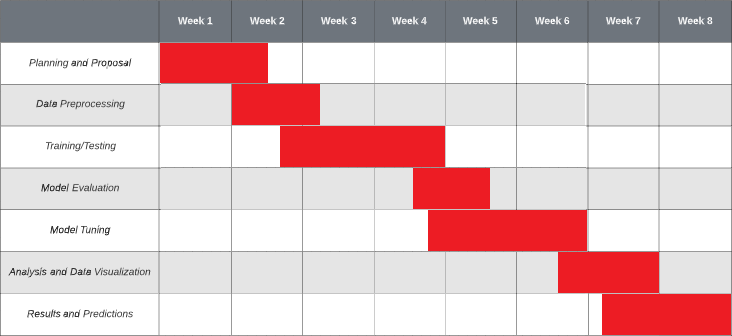
.

**Procedures and Features:** The features of the data set include coordinates, date, FFMC, DMC, DC, ISI, temperature, humidity, wind speeds, and amount of rainfall. The output variable is the area of the forest burned. FFMC, DMC, DC, and ISI are components of the Fire Weather Index System (FWI). FFMC, DMC, and DC indicate moisture levels of fuels based on how long they take to adjust to the current humidity. ISI is a composite of wind speeds and fuel moisture that estimates the potential for forest fires to spread. The temperature is measured in degrees celsius. The rain is measured in mm/m2 and the wind speeds are measured in km/h. The burned area is measured in hectares. Affected areas less than 100 m2 are rounded to 0 ha and therefore the data set is very skewed towards 0. Because of this, the area will be transformed using A’ = ln(A + 1) as recommended by the paper “A Data Mining Approach to Predict Forest Fires” by Cortez and Morais.

From these features, Principal Component Analysis (PCA) will be used in order to extract the features that are most impactful on the output variable. It will reduce the dimensionality of the data set as well as ensure that the variables are independent of each other. As this data has multiple dimensions, and will be used to solve a regression problem, the Support Vector Machine (SVM) algorithm will be used. We will use the Gaussian Radial Basis Function as the kernel for SVM unless another kernel function is determined to be a better fit.

**Evaluation Metric:** The evaluation metric will be the Root-Mean Square Error of the predicted data against the target data.

**Anticipated Results:** The day, and the x and y coordinates should have the least correlation with the final results. The x and y coordinates of a given forest fire will be based on the conditions of the resulting area, but should effectively be randomly distributed. The day of the week should also have very little impact. The month will not have a linear correlation with the area of forest burned. The FWI characteristics (FFMC, DMC, and DC) and humidity should have the greatest negative correlation with the output variable. The drier it is, the easier it is for the fire to spread. The temperature should have a slight positive correlation, but it is expected that the other features will have greater impact. The rain will have a negative correlation, but if the larger fuels (indicated by the DC characteristic) are dry, it is still possible to have a large fire. Based on the predicted strong correlations between the features and output variable, the resulting model should be fairly accurate in predicting the size of forest fires.

**Timeline:**   


**References:**

*1.4. Support Vector Machines*. scikit. (n.d.). Retrieved September 29, 2021, from

https://scikit-learn.org/stable/modules/svm.html#.

*Cortez, Paulo & Morais, A.. (2007). A Data Mining Approach to Predict Forest Fires*

*using Meteorological Data.*

*Fire weather index (FWI) system*. NWCG. (n.d.). Retrieved September 29, 2021,

from https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system.

*SKLEARN.DECOMPOSITION.PCA*. scikit. (n.d.). Retrieved September 29, 2021,

from https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html.