The Impact of Forest Fire Weather Index and Geographical on Total Burned Area

**Executive Summary**

**Introduction**

Every year, forest fires cause a lot of damages to the environment and threaten both animals and human’s lives. There are many online research papers have been done and guidelines to help firefighters save the environment and to warn visitors avoid fire hazard locations. By having these guidelines, visitors would be more likely to avoid accessing fire hazard area during the fire seasons. Moreover, firefighters also have an idea of the fire patterns to help them prioritize their resources. If the fire area is larger around a specific station, they could employ more firefighters to that station or build more fire hydrants around that zone. If the fire in that area tend to be smaller, they could easily send less resources to the battle and have little to no impact. To make sure that the guidelines are safe for everyone, this paper will answer the following questions: What factors contribute to the area of the burned fire? What are the usages of the model? What are the future directions to help improve the model? This analysis will answer the questions by exploring a dataset that contain forest fire and related information from the Montesinho Natural Park in Portugal. To conclude the findings, this analysis will also develop a model using logistic regression to help firefighter determine the impact of fire burn area. All analysis will be done using R programming language.

**Exploratory Data Analysis**

The data originated from the Monteshinho Natural Park with forest fires information that contains a total of 517 observations with 4 space-time variables, 8 fire indices, and 1 continuous target variable – total burned area (hectares). In terms of predictors, horizontal coordinate, vertical coordinate, month of the year, and day of the week are considered as space-time variables. Eight fire indices used to determine the intensity and velocity of fire spread are Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), outside temperature in degrees Celsius, percentage of outside relative humidity (RH), outside wind speed in km/hr, and outside rain volume in mm/m2. There are two observations with relatively large burned area of 746 and 1,091 hectares and no missing values in the dataset. Since the proportion of zeros in total burned area is 47.8%, this variable was converted to a dichotomous variable (0 = area smaller than 0.01ha or 100m2, 1 = area bigger than 0.01ha). With an approximately of 48/52 of the proportion of small burned area versus large burned area, respectively, a binomial distribution will be well presented for this analysis.

Before diving into the variables exploratory, it is important to reshape shape the dataset to change the unit of analysis to be distinct to each horizontal coordinate, vertical coordinate, and month of the year. Although there are 517 rows of data, there is only a total of 136 unique values from a combination of horizontal coordinate, vertical coordinate, and month of the year. This indicates that there might be one or multiple fires happened at the same location and month with the highest historical fire count of 34. The median values of all eight fire indices were extracted from this aggregation to accommodate the outliers from both end points of the variables. Furthermore, the mean of the total area burned was created to represent the average total area burned in each location and at a given month. After aggregating all variables in the dataset, a new feature was engineer to account for the historical fire count happened in the area.

Before diving into model development, it is important to access the relationship between the response variable, total area burned, with each of the independent variables.

**Statistical Analyses**

**Conclusions**

**References**

**Appendix**