**Initial Draft for CMPUT 466 Mini-Project**

Jinzhu Li

1461696

**Introduction:**

One major environment concern nowadays is the occurrence of the forest fires (wildfires), which is a threat to the forest preservation, economical and ecological environment, and human activities. Such phenomenon is caused by multiple reasons such as lightings and human negligence, and each year millions of forest hectares are destroyed worldwide.

Therefore, this study is going to predict the forest fire using meteorological data, using three different machine learning algorithms and answering the question that which one works the best, that is has the smallest error.

**Description of the Data Set:**

The data is the real-world data, collected from the Montesinho natural park, from the northeast region of Portugal, with the aim of predicting the burned area (or size) of the forest fires. The data used in the experiment was collected from January 2000 to December 2003.

There are total of 517 instance and 12 attributes: spatial location X, spatial location Y, Month of the year, Day of the week, Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), temperature, relative humidity, wind speed, and rain. The target value is the total burned area. All these variables are numerical, and Mouth and Day have been converted to numbers accordingly.

**Learning Methods:**

There learning methods are chosen for this project: Neural Network (NN), Support Vector Machines (SVM), and Random Forest (RF).

* Neural Network (NN): This study will consider multilayer perceptron with one hidden layer of 32 hidden nodes and the rectified linear unit function as the activation function and one output node with linear function.
* Support Vector Machines (SVM):
* Random Forest (RF):

**Design of experiment:**

All the experiment reported in this study were conducted using the scikit-learn, an open source library for Python that facilitates data mining and data analysis.

Before fitting the model, some preprocessing was required by NN, SVM and RF models. First, month and day were transformed into numerical attributes. Also, for NN and SVM methods, all attributes were standardized to a zero mean and one standard deviation. Next, the regression model was fitted.

**Final Results:**

# tables and graphs remain to be added

**Conclusion:**

# To be filled