A Novel Forest Fire Prediction Tool Utilizing Historical Weather Data and Machine Learning Methods

Leo Deng [A,C], Clemen Deng [A], Marek Perkowski [A], John Saltenberger [B]

[A] Department of Electrical and Computer Engineering, Portland State University, Portland, OR, USA

[B] Predictive Services Fire Weather Program, Northwest Interagency Coordination Center, Portland, OR, USA

[C] Corresponding Author. Email: leodeng5@gmail.com

Table of Contents

Non-Specialist Summary. In this paper a tool is developed to predict the intensity or magnitude of wildfires in a given area. It utilizes 10 fire weather attributes and six machine learning methods – data mining models – to classify a fire by the amount of acres burnt into one of seven intensity levels.

Abstract. A novel Forest Fire Prediction Tool (FFPT) was developed, constructed by testing six different Machine Learning (ML) methods. 3 methods were based on multiple-valued logic: Disjunctive Normal Form (DNF) rule based method, Decision Trees, and Naïve Bayes. One method was the Support Vector Machine (SVM) based on continuous representation along with the radial basis and polynomial kernel functions. Historical meteorological data from the past 32 years were collected from local sensors including National Fire Danger Rating System (NFDRS) indices and fire weather data. To screen out weather attributes, different combinations of fire weather attributes were tested with the 6 methods. Results showed that using all 10 attributes was most optimal. The six different methods were then trained on all attributes to contend for the most accurate predictor. When testing learners multiple times on random data, the DNF method yields the highest accuracy of 98%. This tool takes forecasted fire weather data of 10 different attributes, which can be obtained locally with relatively low costs ahead of time, and predicts the intensity of wildfires in specific areas. This paper demonstrates that the FFPT achieved accuracy high enough such that it can be used realistically for efficient wildfire prediction.

Additional Keywords: Machine Learning (ML); Forest Fire Prediction; Meteorological Data; Proactive Wildfire Support

Introduction

Forest fires are a major environmental issue causing vast ecological and economic damage while endangering human lives. They are caused to start by multiple events (e.g. humans or lightning). Every country around the globe is endangered by wildfires except Antarctica, and the need for better control of fires is increasing exponentially due to climate change and its side effects. This is especially true for large and very large fires as their numbers have been increasing dramatically; compared to an average year in the

1970's, there are 7 times more fires that are greater than 10 thousand acres each year. Overall, the amount of acres burnt due to forest fires in the world is expected to double by the late 21st century. According to the Northwest Interagency Coordination Center (NWCC) and National Interagency Fire Center (NIFC), year 2014's wildfire season has been one of the busiest they've ever experienced.

Wildfire response organizations currently use different methods of fast fire detection including satellite images and infrared scanners. However, reacting to such wildfire detectors can lose a significant amount of time before help can arrive on the scene, allowing the fire to intensify. The spread of fires is vastly dependent on the meteorological conditions in which it exists, including both the live and dead fuel moisture in the area.

There are many regression tools and similar methods to predict forest fires, many with unfortunately insufficient accuracy rates. In order to increase the accuracy, there needs to be relevant meteorological information as well as a competent method to predict the wildfire.

However, very little results can be found in the forest fire prediction area using meteorological data. Paulo Cortez and Anıbal Morais [1] tested 6 different data mining tools on weather attributes and the Fire Weather Index (FWI) for wildfires in Portugal, and the Support Vector Machine surpassed the accuracy of all others marginally. This model, however, only gave a high accuracy in predicting small fires.

Data Mining Techniques Applied to the Fire Detection Domain

Cristina Vega-Garcia et al. adopted Neural Networks [2] to predict human-caused wildfire occurrences, with an 81% overall correctly predicted no-fire and yes-fire occurrences.

HSU used FASTCiD [3] (a spatial clustering algorithm) to detect forest fire spots in satellite images, and satellite images from North America forest fires were fed into a Support Vector Machine (SVM), obtaining 75% accuracy at finding smoke at the 1.1-km level.

Also, Stojenova et al. used decision tree algorithms, random forest methods, and logistic regression on data from fires in Indonesia to create models that predict active fire locations, and through much pre-processing of the data, reached the best model of 80% with a decision tree algorithm [4].

Many other techniques have been used to help increase fire awareness on Earth [5]. The best model for detecting fire occurrences had 80% accuracy. None of the Machine Learning (ML) methods work 100% of the time – different techniques may work better for certain kinds of data more than others.

In contrast to these previous works, this paper introduces a novel Machine Learning (ML) forest fire prediction method, where the emphasis is on predicting future forest fires with the use of real-time and easily-obtained meteorological data.

The six different machine learning methods used in the development of this prediction tool have been established and known to be successful in other areas of prediction such as cancer; their accuracies vary based on what kind and the amount of data given to learn with.

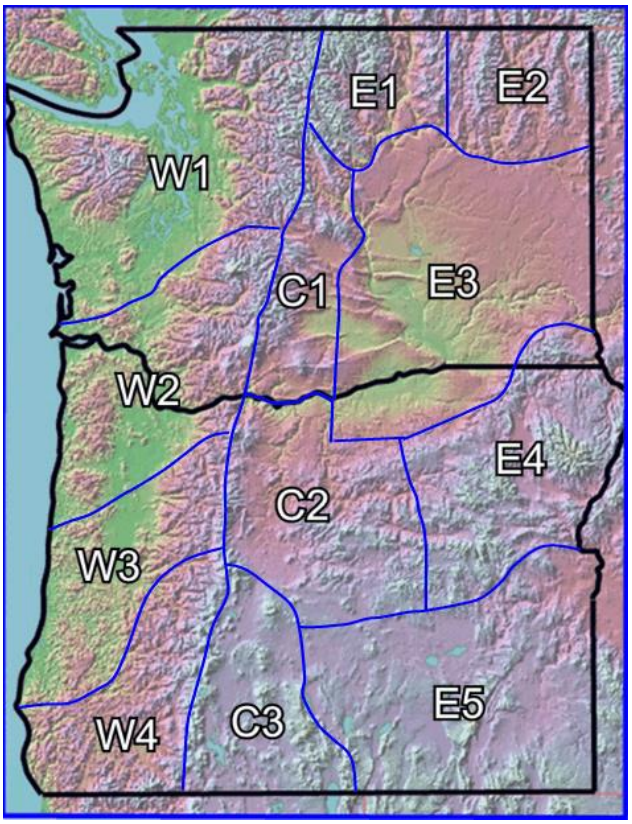
In this paper, we develop a Forest Fire Prediction Tool (FFPT) using machine learning methods with strategically selected fire weather attributes. The method is unique in that it takes advantage of easily obtainable fire and weather information from existing local sensors, and minimizing the cost for both the detectors and the recovery process. Also, rather than allowing response teams to react quickly to a fire that has already started, this new method allows them to be proactive in their fire management movements, providing for the areas that will have potentially huge fires quickly before it can spread to its full potential.

The rest of the paper is organized as follows: Materials and Methods section describes the fire weather database used for testing the new forest fire prediction tool in this paper and details the methodology for maximizing the accuracy of the tool. Results section shows the experimental results for the individual fire weather attributes' accuracies, and comparative results for the six final fire prediction models' average accuracies. The conclusion summarizes the paper.

Materials and Methods

Fire Weather Database

Fig. 1. Map of the predictive service areas in the Northwest United States from which the fire weather data is collected from



The data collected from local sensors by the Northwest Interagency Coordination Center was used for testing and developing the tools in this paper. Fig. 1 shows the areas of predictive service from which data was collected. The data includes the monthly averages of 10 different fire weather attributes including National Fire Danger Rating System (NFDRS) indices.

The attributes are displayed below:

Attribute Domain

1. 100 hour dead fuel moisture (F100)

2. 1000 hour dead fuel moisture (F1000)

3. Live Fuel Index 1-100 (LFI)

4. Sum of Rain Duration (hours) (RainDur)

5. Sum of Rain Amount (inches) (RainAmt)

6. Average Temperature (°F) (Temp)

7. Maximum Temperature (°F) (Max Temp)

8. Minimum Relative Humidity % (MinRH)

9. Wind Speed (mph) (Wind)

10. Duff Moisture Code (DuffMC)

Number of instances (months): 1443

Each of these attributes of the fires in the database are multi-valued, and this data was integrated from the 12 predictive service areas in Fig. 1 in Oregon and Washington States over a time period of 32 years. Each of the 1443 instances are classified as one of 7 intensity levels based on the number of acres burned.

Machine Learning Methods

Six different machine learning methods were tested [6] in this paper on the fire weather data provided. 3 are based on multiple-valued logic: a Disjunctive Normal Form (DNF) rule based method, Decision Trees, and Naïve Bayes. Three based on continuous representation: the Support Vector Machine (SVM) along with the radial basis and polynomial kernel functions.

As a result of the varying concepts these methods are based on, one cannot be absolutely named better than another; their ability to optimize with precision is dependent on the type of data that is being tested.

DNF rule based method (CN2 learner) [7, 8] is a logic-based method adopted from binary function minimization to data with multiple input attribute variables. Decision Trees learning method is a practical inductive inference method and is based on creating a decision tree to classify the data [7]. Naïve Bayes is a probability based Machine Learning method and assumes each attribute of the data is unrelated to any other attribute [7, 9]. SVM is an optimization-based, non-probabilistic binary linear classifier that operates on continuous representation, it selects the optimal hyper plane used as the threshold for classifying the data [7]. Two kernel tricks are also used, adding additional dimensions (classifiers could be round or polynomial shaped rather than linear) to the input data to make it more easily separable. We intentionally selected different types of methods and different representations, believing that this should improve the results.

Strategically Testing and Selecting the Attributes

The 10 fire weather attributes utilized in this paper were tested in different combinations to confirm their relevance to the intensity of a wildfire. The Support Vector Machine and rbf kernel were tested with each attribute individually to identify their individual potential for predicting the intensity of a fire. For each of the 7 intensity levels, the machine classifies the fire into one of two categories – less than some number and greater than some number – multiple times as shown in Fig. 2 resulting in the final intensity of the fire.

The 10 attributes are ranked from highest to lowest average accuracy. The Support Vector Machine then takes three different cases of input data – all attributes, top 7 attributes, and top 4 attributes – resulting in 3 different final average accuracies. Each of the 3 combinations is tested 5 times, each trial using a randomly selected 97% of the data, and testing on the remaining 3%. The combination with the highest average accuracy is deemed to contain the optimal attributes which from then on will be used to test each of the six different machine learning methods.

Fig. 3 illustrates how the best set of input data is determined using the two support vector machine methods. The “4 best attributes” and “7 best attributes” are determined from the previous step; individual accuracies are used to rank them by importance. The “Medians of attributes” option takes the median value of each of the 10 attributes and uses those single values to train on the data.

These four training set options are used to train a linear SVM as well as the radial basis function (RBF) kernel SVM using a 97%/3% training/testing set on random data. This is repeated 5 times for each method to finally see which preprocessing option gives the highest accuracy and ultimately the best training set to use going forward.

Fig. 2. Final FFPT illustration

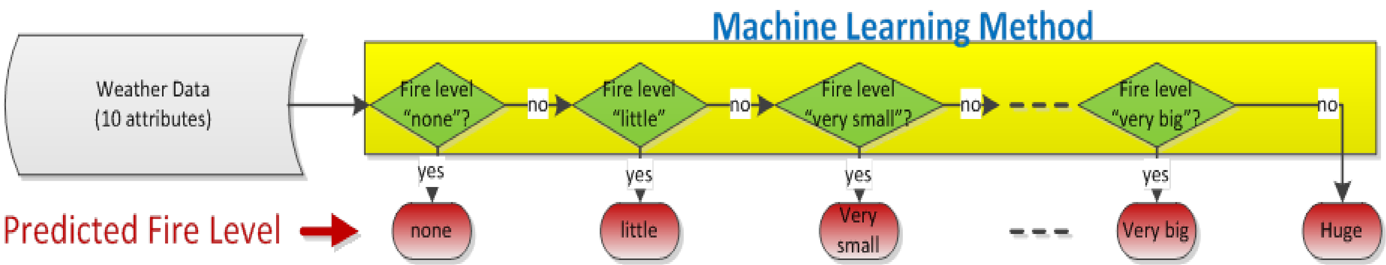
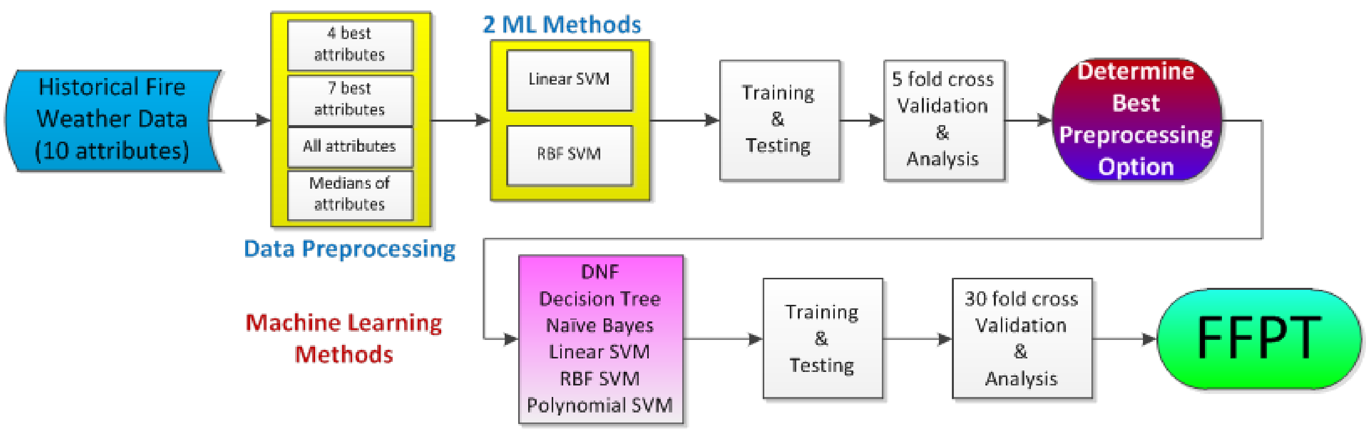


Fig. 3. Determining best set of input data from 4 different preprocessing options and applying to 6 ML Methods to develop FFPT



Testing Six Methods on Selected Attributes to Find the Best FFPT

Once the most optimal training set is determined, this set of data is used to train on all 6 different machine learning methods using the same 97%/3% training/testing method repeating each intensity/method combination 30 times using randomly selected data each time.

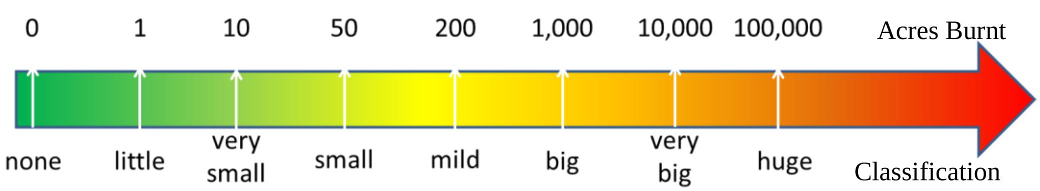
The tool is made to classify fires into one of 7 specific intensity levels as shown in Fig. 4. The “none” and “little” levels are combined as one in my project. Therefore, the machine must perform 6 different separations between each intensity level so that it will give the final fire intensity in the end as shown in Fig. 2. This results in 6 different methods and 6 different intensity separations resulting in a total of 36 accuracies. Each of these 36 values are defined by the average or mean of the 30 random trials from the previous step. Finally, for each of the six machine learning methods, the overall average accuracy is derived from the 6 intensity separations it makes.

In Fig. 4, the six numbers on the top row represent each intensity classification line. For example, the “0” is “little or none vs. yesfire” where the machine predicts whether or not there will be any damage at all. “6” is “not huge vs. huge” where the machine predicts whether or not the amount of burned land will surpass 100,000 acres. The values in the “method avg.” column will be used to compare the six methods to each other and determine which of the methods is most optimal for the FFPT.

Fig. 4. Sample Table for testing 6 machine learning methods’ accuracies

	0	1	2	3	4	5	6	method avg
DNF								
tree								
bayes								
svm rbf								
svm polynomial								
svm linear								

Fig. 5. Classification levels FFPT uses: values determined based on the median and range of values for acres burnt.



Results and Discussion

Fig. 6-9 show the accuracy results of each of the 10 attributes tested individually with the support vector machine, the 6 ML methods accuracy for each of the 7 intensity levels, and the final master average accuracies of each version of the FFPT respectively.

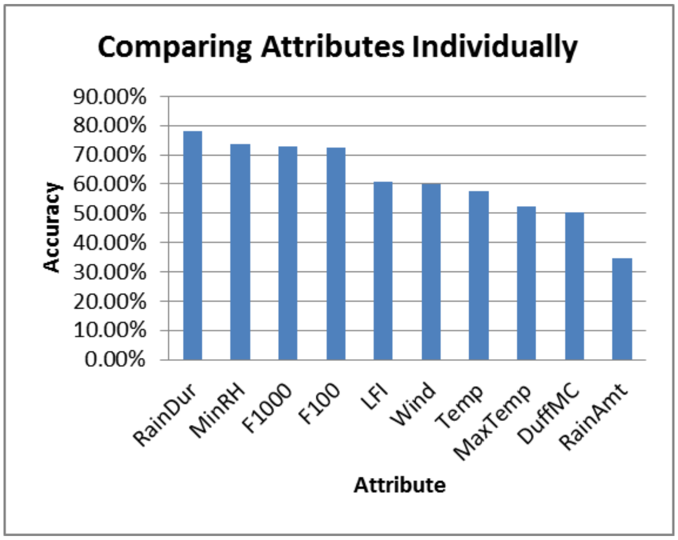


Fig. 6. Accuracy of each individual attribute using SVM

The results in Fig. 6 show each of the 10 attributes' individual accuracies when tested with a 97%/3% training/testing set. The top 4 attributes – Rain Duration, Minimum Relative Humidity, F1000, and F100 –, top 7, all attributes, and all attributes' median values will be tested with 2 machine learning methods in the next step to determine which set of data is most optimal.

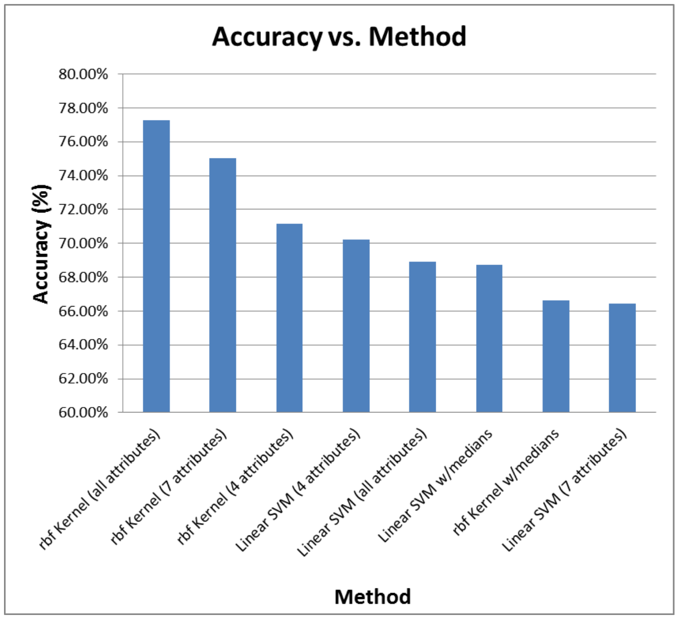


Fig. 7. Different data sets results

The results in Fig. 7 show:

1. The highest accuracy occurs when using all attributes to train.
2. Using 7 attributes with the rbf Kernel SVM gave a much higher accuracy than when using 4 attributes.
3. The least optimal training set was one that was using median values.
4. Rbf Kernel SVM is almost always better than the Linear SVM.

In the sequel of this paper, all attributes will be used to train the 6 machine learning methods, as they all have a significant role to play in the intensity of a fire.

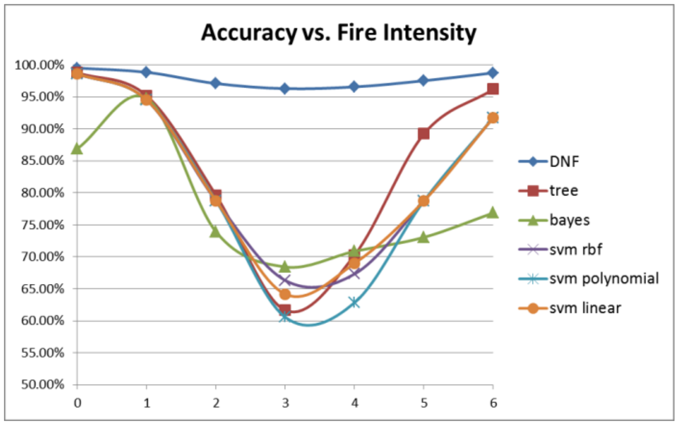


Fig. 8. Accuracy of each machine learning method for predicting each of the 7 intensity levels of wildfires

The trend seen in Fig. 8 for all 6 methods shows that this tool is most accurate at predicting very small fires or very huge fires, and the accuracy drops by some margin as the intensity level nears the center of the pool of data. This can be explained, because for classifying mild fires, the machine has to deal with an equal amount of data on both sides of the separation. When looking at whether a fire is huge or not, on the other hand, is much easier with almost all the data being less intense than "huge" and only a small portion of months with wildfires having burnt more than 100000 acres of land.

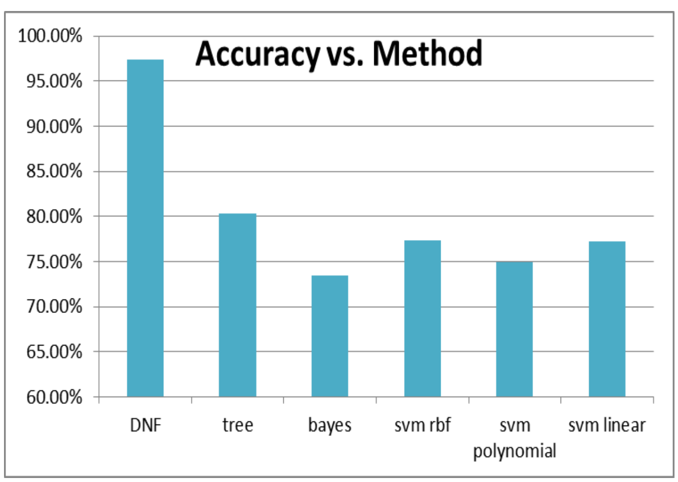


Fig. 9. Average accuracy for each method

Fig. 9 gives the final average accuracies of each machine learning method when tested on all attributes with a 97%/3% training/testing set. The DNF rule based method gives the highest accuracy by far at an impressive 97.8% overall accuracy, and a maximum of 98.8% when predicting huge fires.

With so much data from the past 32 years – 1443 months of fires – the DNF method shows the best ability to optimize the large data set with precision. Its lowest accuracy when classifying fires into very small vs. mild fires stayed very high at 96.3% unlike the other 5 methods whose accuracies all dropped significantly by at least 25%.

Conclusion

A new forest fire prediction tool (FFPT) utilizing a disjunctive normal form (DNF) rule based method was developed and was used for wildfire prediction for the Northwest United States area (Oregon and Washington states). This method was the best chosen out of six different machine learning methods all tested with random data from the database. The fire weather database compiled by the Northwest Interagency Coordination Center of the American Meteorological Society was used in testing the tool. Each of the fire weather attributes were compared to one another using statistical analysis. Individual attributes were then compared separately using a constant machine learning method to authorize the different correlations of each with the fire intensity. The optimal relationship between all of these variables to the intensity of a fire was then established. A data mining tool was developed that implements the 10 fire weather attributes for predicting the intensity of wildfires.

In contrast to previous methods of fast fire detection, this FFPT makes it possible to use proactive resource management for firefighting response teams, promoting the conservation of water as well as many other valuable resources. This will inevitably result in a significantly better control of large fires and lowered costs for land restoration.

Results showed that of the six machine learning methods used, given a 97%/3% training/testing combination, the tool utilizing DNF rule based method and all 10 fire weather attributes gave the highest average accuracy of 97.8%, significantly better than the rest. All 6 machine learning methods show the prediction accuracy is higher for small and large scale fires. Fire levels in the middle range are hardest to predict. However, the DNF method showed a significantly smaller drop in accuracy when predicting mild fires.

This tool will result in a much greater awareness for wildfire management, allowing proactive movement for response teams. Ultimately, the FFPT will decrease the use of water resources, human resources, and promote the greater protection of our environment.

The FFPT developed in this paper is demonstrated to yield the highest accuracy out of all previous tools for forest fire intensity prediction. The Northwest Interagency Coordination Center has approved of this tool, and will start a pilot for beta testing.

References

[1] Cortez, Paulo, and Anibal Morais. <i>A Data Mining Approach to Predict Forest Fires Using Meteorological Data</i>. Guimaraes, Portugal: Department of Information Systems/R&D Algoritmi Centre, U of Minho.

[2] Vega-Garcia, Cristina, Bryan Lee, Paul Woodard, and Stephen Titus. Trans. Array applying neural network technology to human-caused wildfire occurrence prediction. . Vol.10 No.3, 1996. Alberta: Department of Forest Science, 1996. Web. 18 Feb. 2014.

[3] W. Hsu, M. Lee, and J. Zhang. Image Mining: Trends and Developments. Journal of Intelligent Information Systems, 19(1):7–23, 2002.

[4] D. Stojanova, P. Panov, A. Kobler, S. Dzeroski, and K. Taskova. Learning to Predict Forest Fires with Different Data Mining Techniques. In D. Mladenic and M. Grobelnik, editors, 9th International multiconference Information Society (IS 2006), Ljubljana, Slovenia, 2006.

[5] I.S. Sitanggang, R. Yaakob, N. Mustapha and A.N. Ainuddin, 2013. Predictive Models for Hotspots Occurrence using Decision Tree Algorithms and Logistic Regression. Journal of Applied Sciences, 13: 252-261.

[6] Zupan, Blaž, et al. "Orange and Decisions-at-Hand: Bridging Predictive Data Mining and Decision Support." n.pag. Web. 18 Feb 2014.

[7] Orange Machine Learning software suite http://orange.biolab.si/.

[8] Clark, EE., & Niblett, T. (1989). "The CN2 induction algorithm". Machine Learning, 3, pp. 261-284

[9] "Bayesian Reasoning and Machine Learning", David Barber. http://www.cs.ucl.ac.uk/staff/D.Barber/brml, pp. 203-213.