

Predicting Algerian Forest Fire with Accessible Classification Algorithms

Grasiele Romanzini Bezerra

Overview

Wildfires are one of the contemporary world's biggest problems. It is responsible for ecological, economical and ethnic losses, consequently a rapid response is crucial to reduce its impacts. In such situations, where fast decisions are mandatory, the use of machine learning has becoming extremely popular because of its ability to make precise predictions based on previous data.

It is well known that meteorological conditions, such as temperature, relative humidity of the air, wind speed and rain volume can impact forest fires. Therefore, several numerical index that use these meteorological measures were created in order to predict wildfires, as for example, the Canadian Fire Weather Index (FWI) that was created in the 1970's [1, 2] and shall be used in this report.

The FWI is composed of six numerical components easily calculated with meteorological data using pre defined tables. Firstly there is the Fine Fuel Moisture Code (FFMC), that is calculated using the rain volume, the relative humidity of the air, the temperature and the wind velocity. Subsequently the FFMC is used to calculate the Initial Spread Index (ISI), a score that represents the capability of the fire initial spread. Other FWI components are the Duff Moisture Code (DMC), that is calculated using the rain volume, the relative humidity of the air and temperature, and the Drought Code (DC), calculated with the rain volume and the temperature. Furthermore they are used to calculate the Buildup Index (BUI), a measurement of the amount of fuel available. Finally the BUI and the ISI are used to calculate the Fire Weather Index (FWI), the index that measure the fire risk [1].

Finally, the data used in this report was collected at the summer in Algeria, a country located in the north of the Sahara desert that suffer with wild fire mainly during this period. Annually, on average 35024ha of forest are burned, furthermore, between 2008 and 2012 it has burned in total 320409 ha of forest [3, 4]. In summary, the data that I will use in this report are from June 2012 to September 2012, from two Algerian Regions, Bejaia and Sidi Bel-abbes.

The main objective is to use a decision tree algorithm, with the fire weather index as input, to predict whether there will or there will not be a forest fire in those places. This algorithm was chosen because of it's simplicity and low computational cost that make it accessible to use even in distant and isolated locations. However, I will also use random forest as an example of more complex algorithm to test how much information is lost when using an over simplified model such as classification trees.

This report is organized as follow. In the upcoming section I will describe the data set with more details and then describe the models used. Afterwards, I will discuss the results of the predictions and in the final section I will present the conclusions.

Methods

This section will be divided in two. First I will describe the data set utilized, with details about each variable, also I shall point out some of the data cleaning process. Following that, I will delineate the training process,

from the splitting of the data set between training and test set to some elucidation about the two model that will be utilized.

The Data Set

Firstly, this data can be found in the UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++#>). The Raw data set consisted in 15 variables and 244 observations, and the first step was to check for missing values. Since there was only one missing value of FWI, I decided to exclude this observation from my analyses, as a result, the data set ended up with a total of 243 lines.

The first three columns are the day, month and year, respectively, when the data was collected, the values on it include the date from June 1st to September 31st of 2012. The following four columns are the meteorological data. It starts with the temperature (Temperature), in the fourth column, these values are measured in Celsius degrees and it goes from 22°C to 42°C. The fifth column contains the values of relative humidity of the air (RH), these values are given in percentage and it has a range between 21% and 90%. The wind speed (Ws), in km/h, can be found in the sixth column, its values are between 6 km/h and 29 km/h. The last meteorological value is in the column 7, it is the daily rain volume (Rain), that is given in mm and it has a minimum value of 0 and a maximum value of 16.8.

The next six variables, from column 8 to column 13, are the values of the Canadian Fire Weather Index System. Firstly there are three fuel moisture codes, in the column number 8 is the Fine Fuel Moisture Code (FFMC), its observations go from 28.6 to 92.5. This value is an evaluation of the moisture content of the surface litter and can give a measurement of the ignition and initial spread of the fire. In the columns 9 and 10 you can find the Duff Moisture Code (DMC), which vary from 1.1 to 65.9, and the Drought Code (DC), with values between 7 and 220.4. These two last codes represent the moisture in the shallow and deep organic layers respectively [1, 2].

After that, there are the two “Fire Behavior Indexes”. To begin with, in the column 11 there is the Initial Spread Index (ISI), its values go from 0 to 18.5, this variable represents the spread velocity of the fire. In column 12 there is the Buildup Index (BUI), with values between 1.1 and 68, this code denotes the amount of fuel available. Finally the Fire Weather Index (FWI) can be found in the 13th column, its values go from 0 to 31.1, this index is calculated through the five codes previously described and it gives us a measurement of the fire risk, consequently it will be the one used to make the predictions.

The final two columns are the outcomes (Classes) and the place where the data was collected (Region) respectively. The data found in the 15th are the names of the two Algerian regions where this data came from, Bejaia and Sidi Bel-abbes. Originally there were 122 lines per region, nevertheless one line from Sidi-Bel Abbes will be kept out due to a missing value in the FWI column. And ultimately, in the column 14 there are the Classes that have two values, “fire” and “not fire”, moreover these values shall be defined as a factor, with the “fire” as the positive class.

The Models

Before applying the machine learning algorithms it is important to define a validation set, that will be used to test the predictions. For this, I used the “createDataPartition” function from the Caret package and divided the 243 observations into two sets, one test set with 20% of these values and one train set with the remaining 80% of the observations. It is also important to state that the test set will not be used at any moment in the training process, it is exclusively to test and validation purposes, by this way I can avoid overtraining the model.

The first model that shall be used will be classification trees, also known as decision trees. Decision tree is a supervised classification algorithm where a parameter is tested in the so-called nodes and divided into branches that leads to child nodes, where another parameter is tested, or in leaves, where the final classification is displayed [3, 5, 6]. This method was chosen because it is a simple algorithm with a extremely low computational cost, consequently it can be accessibly performed even in remote and isolated places. In addition to that, the

classification tree model can be easily interpreted, once it is trained, or in other words once the tree is grown, it can be used straightforwardly to make fast decision.

Following that, a more complex algorithm was trained in order to compare the results. The model chosen was the random forest, a model proposed in the early 2000s that use an ensemble of classification trees randomly grown [7, 8]. It has been proven that this method has improved considerably the classification algorithms by showing highly accurate results, and that is why it was used as a comparative model.

Results

After some data exploration we can have some insights of how the Fire Weather Index (FWI) can be used to predict forest fires. In the figure 1 there is a boxplot of the FWI values for “fire”, the pink box, and “not fire”, the blue box. Firstly it is interesting to notice that there is a clear separation between the FWI values for each class, for most observable classified as fire the FWI value is above 5, with its median above 10. On the other hand, the great majority of FWI values classified as “not fire” are close to zero.

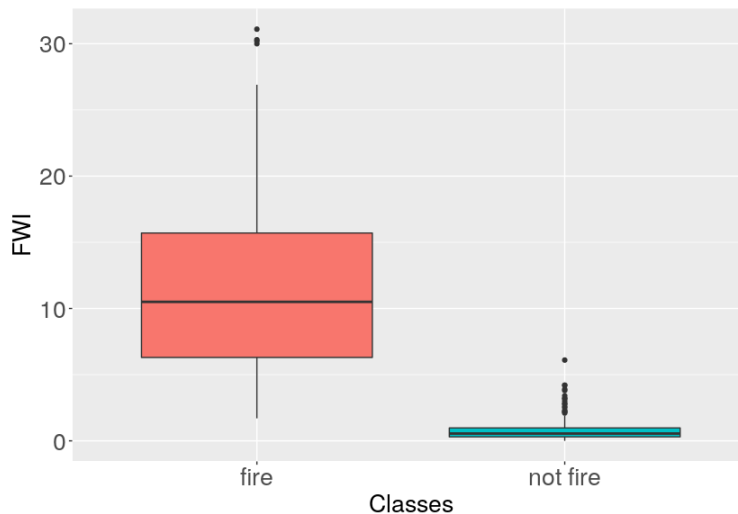


Figure 1: FWI Distribution

This plot confirm that the Canadian system of fire prediction is a substantial measurement and the Fire Weather Index can be precisely used to predict wildfires. In the next subsection I shall present the results of the decision tree and random forest trained and tested in this work. In order to perform the algorithms, it was used the train function from the Caret package, with the features “rpart” for classification trees and “rf” for random forest.

The classification tree

After training the classification tree model, using the feature “rpart” from the Caret package, it was possible to plot the resulting tree of the trained model. The figure 2 shows how this decision tree operate, its node is testing if the FWI is higher or equal to 2.95, if this statement is true, the branch leads to the “fire” classification, otherwise it will lead to “not fire”. If we compare this figure with the figure 1 we can get an insight in why the limit value for FWI was set in 2.95, although 75% of the “fire” values are above 5, these values have a high dispersion, and its minimum value is well below this mark. Oppositely, the “not fire” values have a smaller dispersion and are more concentrated close to zero.

In order to determine if this model is reliable it is necessary to test it. Furthermore, after test the model using the test set, I obtained the accuracy of 0.94, which is a significantly promising value, however it does not



Figure 2: Final Tree

Table 1: Confusion Matrix of the Classification Tree Algorithm

Prediction	Reference	
	fire	not fire
fire	26	1
not fire	2	21

say much about the binary classification model. In order to have a better insight in how the model performed it is better to take a look into the confusion matrix, so it is displayed in the table 1.

In summary, the confusion matrix in the table 1 shows that the model predicted 26 fire and 21 not fire correctly, it had 2 false negatives and 1 false positive. It give us a true positive rate of 93% which is a satisfactory result for such a simple model. It is also important to observe that the figure 1 also shows an outlier for the not fire values, what explain the one false positive value.

The Random forest

In the former subsection I presented the results for the implementation of a classification tree algorithm. In this section I intend to use random forest algorithm as an example of a more complex classifier in order to measure how the classification can be improved by changing the method.

After training the random forest classifier using the training set and the “rf” feature from the Caret package, it was used the test set to test how efficient this model was in predicting the wildfires. The accuracy improved from 0.94 to 0.96, however, as mentioned before, although the accuracy is a important measure for an algorithm, it is not the best evaluation of a binary classification, so the confusion matrix is shown in the table 2.

The main change in the confusion matrix of the table 2 is that the true positive values increased from 26 to 27, consequently the false negatives decrease to one. Therefore the true positive rate rose from 93% to 96%, this is a slight improvement in the classification model. Nevertheless, that was no change in the false positive value, the reason for that was presented in the section above, there is an outlier among the not fire values, as can be seen in the figure 1.

Table 2: Confusion Matrix of the Random Forest Algorithm

Prediction	Reference	
	fire	not fire
fire	27	1
not fire	1	21

Conclusion

Wildfires are a disaster that causes economical, ecological and humanitarian harm. In consequence, a fast and precise prediction can decrease the damages and, in some cases, even avoid the fires. Because of that, it has become increasingly common the use of machine learning and artificial intelligence to make such predictions. Today is massively customary for prediction models to make use of the Canadian Fire Weather Index (FWI), that is a numerical index calculated using meteorological data.

The main objective of this work was to show that it is possible to train accessible models for forest fire prediction using elemental algorithm such as decision trees with the FWI as input. Besides training and testing the classification tree algorithm, I also trained its more elaborated version, the random forest, in order to compare the results and test how much information is lost by the simplified model.

In conclusion, the accuracy for decision tree was 0.94, and it also resulted in 26 true positives out of 28 values in the positive class, it gives a true positive rate of 93%. In addition to that, the random forest results indicated a slight increase in the accuracy to 0.96 and in the true positive rate to 96%. Although, an improvement in the model performance was expected, this growth is not enough to discard the use of decision trees. This is an exceptionally satisfactory result, and it shows that classification tree is a substantial model for wildfire prediction and it can be used, in conjunction with other methods, for prevention and response to the forest fires.

Bibliography

1. Cortez, P., & Morais, A. de J. R. (2007). A data mining approach to predict forest fires using meteorological data.
2. Taylor, S. W., & Alexander, M. E. (2006). Science, technology, and human factors in fire danger rating: The canadian experience. *International Journal of Wildland Fire*.
3. Abid, F., & Izeboudjen, N. (2019). Predicting forest fire in algeria using data mining techniques: Case study of the decision tree algorithm. In *International conference on advanced intelligent systems for sustainable development*. Springer.
4. Meddour-Sahar, O., Gonzalez-Caban, A., Meddour, R., & Derridj, A. (2013). Wildfire management policies in algeria: Present and future needs. In *Proceedings of the fourth international symposium on fire economics, planning, and policy: Climate change and wildfires*.
5. Rutkowski, L., Jaworski, M., Pietruczuk, L., & Duda, P. (2014). The cart decision tree for mining data streams. *Information Sciences*.
6. Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*.
7. Breiman, L. (2001). Random forests. *Machine learning*.
8. Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*.