# Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm

**Chapter** · February 2020

DOI: 10.1007/978-3-030-36674-2_37



**2 authors:**

Faroudja Abid
Centre Algérien de Développement des Technologies Avancées


Nouma Izeboudjen
Centre Algérien de Développement des Technologies Avancées

# Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm

Faroudja Abid(✉) and Nouma Izeboudjen

DMN, Centre de développement des Technologies Avancées, Cité 20 août 1956, Baba Hassen, Algiers, Algeria
fabid@cdta.dz

**Abstract.** Forest fire is a disaster that causes economic and ecological damage and human life threat. Thus predicting such critical environmental issue is essential to mitigate this threat. In this paper we propose a decision tree based system for forest fire prediction. The aim being the integration of the decision tree classifier as a part of the smart sensor node architecture that allows fire prediction in automated and intelligent way without requiring human intervention. The fire prediction is based on the meteorological data corresponding to the critical weather elements that influence the forest fire occurrence, namely temperature, relative humidity and wind speed. We have obtained accuracy about 82.92% regarding the software implementation of the proposed DT based forest fire prediction system.

**Keywords:** Data mining · Decision tree · Fire prediction · Smart sensor node

## 1 Introduction

Forest fire is a critical concern over the world, each year millions of hectares are destroyed all around the word. Algeria is one of the countries affected by this phenomenon, mainly in summer. According to [1], over 320409 hectares (ha) of forest are burned between 2008 and 2017 with over 31513 fires. Over 5110 fires and more than 99061 hectares of burnt area had been recorded in Algeria in 2012. As reported in [2], Algeria is the fourth most affected country among the countries covered by the European Forest Fire Information System (EFFIS).

Forest fires detection and prediction become very important concerns for avoiding the damage caused by this disaster. The exploration of new methods for fire detection and prediction as alternatives to the old ones becomes an emergency. Several technologies and systems have been proposed to detect fire, e.g., systems employing charge-coupled device cameras and infrared detectors, satellite systems and wireless sensor networks (WSNs).

The introduction of new communication technologies such as WSN [3–8] in monitoring systems is a promising direction that can be exploited for fires prediction. The trend is toward the integration of artificial intelligence to the WSN to automate the prediction and detection of fire occurrence. This work is a part of an ongoing research


M. Ezziyyani (Ed.): AI2SD 2019, AISC 1105, pp. 363–370, 2020.
https://doi.org/10.1007/978-3-030-36674-2_37

on a Field-programmable gate array (FPGA)-based smart sensor node. We propose a fire prediction based on integrating one of the data mining techniques [9], namely decision tree (DT) in the smart sensor node architecture. The integration of the DT classifier as a part of the sensor node architecture allows fire prediction in automated and intelligent way.

The idea is the hardware implementation of the DT classifier as an intellectual property (IP) core in the sensor node architecture, thus the sensor node will locally and automatically predict the fire using the IP_DT classifier. In addition to the weather data sensing by mean of a combination of several sensors for meteorological data; the sensor node will be able to predict fire occurrence locally using the DT classifier without requiring human intervention.

In this paper we expose a decision tree based fire prediction, which constitutes the intelligent part of the sensor node. Three meteorological attributes that influence the fire occurrences are used, namely temperature, relative humidity (RH) and wind speed. Since the classification is two classes, the binary tree is considered. It is noteworthy to precise that, we focus in this paper on the software implementation of DT classifier as a predictive model, which will be implemented on hardware. The hardware implementation and the sensor node architecture are not in the scope of this paper.

This paper is organized as follow: In Sect. 2, we review related work for both data mining and WSN based forest fire prediction and detection. The methodology for the predictive model generation is presented in Sect. 3. In Sect. 4, we expose the performances evaluation of the DT classifier and it ensembles and we discuss the obtained results. Finally, Sect. 5 concludes this paper.

## 2 Related Work

Substantial studies have investigated the application of the data mining techniques for fire detection and prediction purpose [3–6, 8, 10–12]. For example a comparison study of various data mining techniques in WSN based fire detection is presented in [3]. According to this study, the prediction is better using the neural network classifier in the case of larger dataset, whereas in the case of small dataset, classifiers such as the OneR (one-level decision tree) or FURIA (Fuzzy Unordered Rule Induction Algorithm) give reasonable results. They stated that the selection of a correct data mining algorithm depends on the application and the compatibility of the observed dataset. In [4] Giuntini *et al.* proposed a self-organizing and fault-tolerant (at the application level) WSN model for wildfire detection. The forest fires dataset of the UCI machine learning repository was used for three decision trees components included in their model. The work in [5] discussed a dataset analysis to extract the most contributing features for wildfire and residential fire detection using data mining approaches in WSN. The reported accuracies are about over 81% for residential fire detection and over 92% for the wildfire detection. In [6] authors applied the data mining technique to reduce the data size in the WSN, which includes different sensors, namely temperature, humidity, smoke and light sensors. The naive Bayes classifier is adopted for forest fire detection in this study, the reported precision is about 94%.

In another work [7], Heffeda *et al.* proposed a WSN for forest fire detection based on the fire weather index (FWI) system components, namely the fine fuel moisture code (FFMC) and the FWI, which are computed using the weather conditions. The detection system is modeled as a k-coverage model. The computed FFMC and FWI values are sent to a processing center for action. Authors in [8] exposed a WSN system for forest fire detection; they adopted a multi-criteria detection implemented by an artificial neural network. Their prototype is based on the TelosB sensor node.

The decision tree learning algorithm is widely applied for this purpose due to its simplicity and ease readability compared to others machine learning algorithms. For example, a system that predicts the forest fire based on data mining approach is exposed in [10]. Stojanova *et al.* evaluated the forest fire occurrence in Slovenia using the logistic regression and the J48 decision tree (J48 is an implementation of the C4.5 on Weka tool [13]) and its ensembles. They used both satellite-based and meteorological data. According to this study the bagging DT gives the best results in term of accuracy, precision and Kappa statics in case of the continental Slovenia dataset. The reported accuracies are about 0.812, 0.849 and 0.844 for the J48, the bagging DT and the boosting DT, respectively. Fire occurrence prediction is also studied in [11]; authors applied the decision tree and the back propagation forward artificial neural network.

A solution based on the support vector machine (SVM) classifier to predict burned area is discussed in [12]. Cortez *et al.* used the meteorological data from the northeast region of Portugal; the dataset includes direct weather inputs, namely temperature, rain, relative humidity and wind speed. They concluded that the SVM based solution is suitable for small fires detection and presents a limitation for large fires.

In contrast to the systems, which detect the fire occurrence, we propose a system that predicts the fire based on weather conditions. Compared to the works that predict fires, we propose a hardware implementation for speeds up the process by predicting and giving the decision locally at sensor node level. Hence, there is no need for communicating the collected data to the base station (the sink) to give decision.

## 3 Methodology for Fire Prediction Model Development

In this section, we expose our approach for forest fire prediction based on the machine learning technique. We first give a brief overview of the machine learning algorithm adopted as a predictive model, namely the decision tree algorithm. After that, we present the software implementation that allows the DT construction.

### 3.1 Brief Review of Decision Tree

A decision tree [14] is a supervised learning algorithm, known as an Iterative Dichotomiser (ID3) developed by Quinlan, he later presented the C4.5. The third learning algorithm of the decision tree is the Classification and Regression Trees (CART) tree, which was proposed by Breiman *et al.* The CART algorithm combines classification and regression.

The Decision tree algorithm allows predicting, explaining and classifying outcomes. The result is a tree structure, where the internals nodes represent the tests on an attributes, branches represent outcomes of the test and leaves represent the class labels (decision). The root node is the most significant attribute or split.

An example of decision tree for the concept buys_computer is shown in Fig. 1. The internal nodes represent the tests on attributes, and leafs indicate classes ('yes' for buys computer class and 'no' for cannot buys computer class). The possibility to translate the obtained tree into a set of IF-ELSE rule makes the DT understandable and easy to use thus easy to translate to hardware implementation.

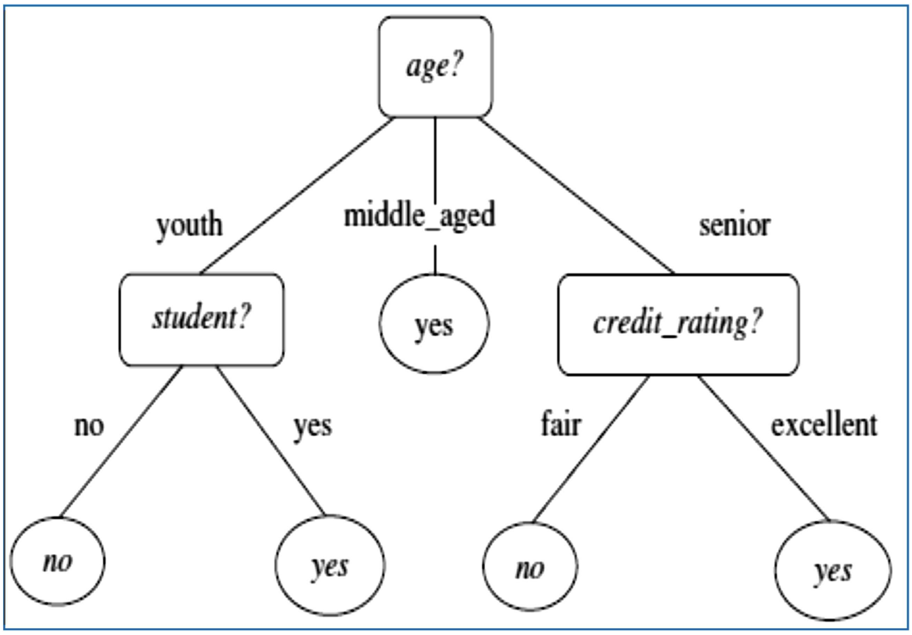


**Fig. 1.** Decision tree example for the concept buys_computer [9]

### 3.2 Software Implementation and DT Generation

The binary DT classifier is chosen for its simplicity, yet with good performance. The input of the learning algorithm is a set of labeled data and the output is a binary tree (two classes, 'fire' and 'not fire'), the generated tree is used for the classification of a new input. The predictive model is developed and evaluated under Weka tool [13].

The DT training is performed using the fire dataset based on the meteorological observations of two regions in Algeria. The generated tree is used for the classification with the aim to hardware implementation (IP_DT). The DT will be implemented as a succession of a comparison using if/else statements; the inputs of the IP_DT are the three attributes representing the meteorological observations, namely temperature, relative humidity and wind speed. The outputs are classes, namely 'fire' and 'not fire'; these two classes can be replaced by an integer number (0 for 'not fire' and 1 for 'fire').

## 4 Results and Discussion

### 4.1 Dataset Description

We have used a dataset that regroups a data of two regions of Algeria: the Sidi Bel-abbas region located in the northwest of Algeria and the Bejaia region located in the northeast of Algeria. The used dataset includes 244 instances (a combination of the data collected in Sidi Bel-abbas and Bejaia, with 122 instances each one).

We have used the meteorological observations for the period of summer of 2012, from June to September since the fire occurrence is high on this period and the 2012 is the year where the recorded fire occurrence is the highest from 2007 to 2018. The used dataset includes the critical weather elements influencing the wildfire occurrence, namely temperature, RH and wind speed. The rain attribute is not used since it is an irrelevant attribute, i.e., it has no impact on the decision tree. The numerical attributes values are used to predict two possibilities, namely 'fire' and 'not fire'.

We first classified the forest fires dataset instances as "fire" and "not fire" classes using the two components of the FWI system, namely the FWI and the FFMC. The FFMC indicates the moisture in the fine fuels and it is an indicator of the ignition probability. The FFMC is calculated through the meteorological elements (temperature, relative humidity, wind speed and rain) using the Fire Weather Index Calculator [15]. The 244 instances include 138 'fire' classes and 106 'not fire' classes. The dataset description is presented in Table 1.

**Table 1.** Dataset description

| Attributes name | Description | Values interval Sidi-Bel-Abbes region | Values interval Bejaia region | Values intervals The two regions |
|---|---|---|---|---|
| Temperature | Temperature in Celsius degrees | 24 to 42 | 22 to 37 | 22 to 42 |
| Relative humidity | Relative humidity in % | 21 to 90 | 45 to 89 | 21 to 90 |
| Wind speed | Wind speed in km/h | 6 to 29 | 11 to 26 | 6 to 29 |
| Rain | Outside rain in mm/m$^2$ | 0 to 16.8 | 0 to 8.7 | 0 to 16.8 |

### 4.2 Performances Evaluation

We have used the Weka tool for the performances evaluation and the decision tree (J48) classifier development. Table 2 tabulates the performances of the decision tree (J48); its confusion matrix is presented in Table 3.

**Table 2.** Performances by classes

| Classes | TP rate | FP rate | Precision | Recall | F-Measure | AUC |
|---|---|---|---|---|---|---|
| Fire | 0.92 | 0,27 | 0,78 | 0,92 | 0,85 | 0,82 |
| Not fire | 0,72 | 0,07 | 0,89 | 0,72 | 0.80 | 0,82 |
| Average | 0,82 | 0,18 | 0,83 | 0,82 | 0,82 | 0,82 |

TP rate: true positives rate; FP rate: false positives rate; AUC: area under the receiver operating characteristic (RoC) curve.

**Table 3.** Confusion matrix

| a | b | Classified as |
|---|---|---|
| 37 | 3 | a = fire |
| 10 | 26 | b = not fire |

### 4.3 Performances Comparison of DT and Its Ensembles

In this section we expose the performances comparison between the decision tree (J48) and its ensembles, namely the boosting DT (AdaboostM1), the bagging DT and the random forest. Table 4 tabulates the performances comparison in terms of accuracy, recall and precision, which are calculated for the 'fire' class.

**Table 4.** Performances comparison of DT and DT ensembles algorithms

| Algorithms | Recall | Accuracy (%) | Precision |
|---|---|---|---|
| Simple decision tree (J48) | 0.92 | 82.89 | 0.787 |
| Random Forest (RF) | 0.750 | 72.36 | 0.732 |
| Boosting (adaboostM1) | 0.95 | 84.21 | 0.792 |
| Bagging | 0.825 | 78.94 | 0.786 |

The tabulated results show that the obtained results with the simple DT are significant, mainly in terms of recall and accuracy. The recall value is about 0.92 and the accuracy is about 82.92%, these results are better than the ones obtained with the RF and the bagging DT algorithms; and comparable to the result obtained with the boosting DT.

The results show that the boosting of decision tree (AdaBoostM1) gives the best results in term of recall, accuracy and precision. However the boosting DT needs significant resources and effort to be translated to hardware implementation. This is why we have chosen the simple DT for the hardware implementation of the fire predictive model. It is noteworthy to precise that the simple DT classifier gives the tradeoff between good performances and resources utilization on hardware. In addition the results obtained with the simple DT are close to that obtained with the boosting DT. Hence the simple DT classifier responds to our purpose.

### 4.4 Comparison with Existing Works

The comparison between the existing works and this work is performed for the DT based fire prediction models, since we have selected the simple DT (J48) as a model for fire occurrence prediction. The comparison is summarized in Table 5.

**Table 5.** Comparison with existing works (those that used DT (J48))

| References | Accuracy (%) | Recall | F-measure |
|---|---|---|---|
| This work | 82.89 | 0.92 | 0.85 |
| [10] Stojanova *et al.* | 81.2 | 0.81 | 0.81 |
| [3] Maksimović *et al.* | Not given | 0.96 | 0.96 |
| [4] Giuntini *et al.* | Not given | 0.83 | 0.86 |

The classification in our case is for the fire prediction purpose, the important issue is to know how much the classifier predicts correctly the positive classes. The recall value indicates the sensitivity and gives this estimation of correctly predicted positive cases (fire occurrence), thus we compared our work to the existing works in terms of recall. The second compared performance is the F-measure, which combines the recall and the precision values. It indicates the robustness of the classifier, i.e. its ability to correctly classify instances without missing a significant number of instances.

The results tabulated in this table show that the obtained recall is better than the ones reported in [3] and [5] and comparable to the one reported in [4]. The obtained F-measure is comparable to the others reported results and better than the one reported in [4]. Note that this comparison is indicative because of the difference in the datasets used in each study; nevertheless the obtained results are promising.

## 5 Conclusion

In this paper, we presented a development of a predictive model based on the decision tree for forest fires prediction in Algeria. The main objective is the generation of a predictive model with the aim being its implementation in hardware. Two aspects should be considered: First the model should be ease to convert to hardware since the generated tree will be implemented as an IP core, which will be integrated to the smart sensor node architecture. Secondly it should give a good and acceptable performances and accurate prediction.

Promising results are obtained with the simple DT in terms of recall (recognition rate) and accuracy, with about 0.92 and 82.92% for the recall and the accuracy, respectively. Compared to existing works, the obtained results are quite comparable and somewhat better.

Results show that the decision tree is suitable for our purpose, since its gives significant performances and it can be translated to rule based, hence its hardware implementation will be relatively simple and will requires less resources when implemented as an IP core.

Our future work will be to perform the hardware implementation of the developed predictive model. The IP_DT development will be done using the Vivado_HLS tool for its integration in the smart sensor node architecture. This later will be used for predicting forest fires, which is a critical environmental issue in Algeria.

**Acknowledgments.** This work has been done under the socio-economic project No. 14/CDTA/DGRSDT/2017.

## References

1. Benkheira, A.: Ministère de l'agriculture et de développement rurale et de la pèche, direction des forets Les feux de forêts en Algérie analyse et perspectives. Technical report (2018)
2. San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., de Rigo, D., Ferrari, D., Maianti, P., Vivancos, T.A., Costa, H., Lana, F.: Advance EFFIS report on Forest Fires in Europe, Middle East and North Africa (2017)
3. Maksimović, M., Vujović, V.: Comparative analysis of data mining techniques applied to wireless sensor network data for fire detection. J. Inf. Technol. Appl. (JITA) **3**(2), 65–773 (2013)
4. Giuntini, F.T., Beder, D.M., Ueyama, J.: Exploiting self-organization and fault tolerance in wireless sensor networks: a case study on wildfire detection application. Int. J. Distrib. Sens. Netw. **13**(4), 1–16 (2017)
5. Bahrepour, M., Van der Zwaag, B.J., Meratnia, N., Havinga, P.: Fire data analysis and feature reduction using computational intelligence methods. In: Phillips-Wren, G., Jain, L. C., Nakamatsu, K., Howlett, R.J. (eds.) Advances in Intelligent Decision Technologies. Smart Innovation, Systems and Technologies, vol. 4, pp. 289–298. Springer, Heidelberg (2010)
6. Saoudi, M., Bounceur, A., Euler, R., Kechadi, T.: Data mining techniques applied to wireless sensor networks for early forest fire detection. In: International Conference on Internet of Things and Cloud Computing (ICC), Cambridge, United Kingdom (2016)
7. Hefeeda, M., Bagheri, M.: Wireless sensor networks for early detection of forest. In: The IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Pisa, Italy, pp. 1–6 (2007)
8. Liu, Y., Gu, Y., Chen, G., Ji, Y., Li, J.: A novel accurate forest fire detection system using wireless sensor networks. In: IEEE Seventh International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), New York, pp. 52–59 (2011)
9. Han, J., Kamber, M., Pei, J.: Data Mining Concepts and Techniques, 3rd edn. Elsevier, Amsterdam (2012)
10. Stojanova, D., Panov, P., Kobler, A., Džeroski, S., Taškova, K.: Learning to predict forest fires with different data mining techniques. In: The 9th International Multi Conference Information Society, Jubljana, Slovinia (2006)
11. Karouni, A., Daya, B., Chauvet, P.: Applying decision tree algorithm and neural networks to predict forest fires in Lebanon. J. Theor. Appl. Inf. Technol. **63**(2), 282–291 (2014)
12. Cortez, P., Morais, A.: Data mining approach to predict forest fires using meteorological data. In: Neves, J., Santos, M.F., Machado, J. (eds.) New Trends in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence (EPIA), Guimaraes, Portugal, pp. 512–523 (2007)
13. Weka Tool. https://www.cs.waikato.ac.nz/ml/weka/
14. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
15. https://fire.synopticlabs.org/tools/fwi/