Criminal profiling and classification in provoked wildfires: an approach using Bayesian networks

R. Delgado[1,*]

Departament de Matemàtiques
Universitat Autònoma de Barcelona
Edifici C- Campus de la UAB. Av. de l'Eix Central s/n.
08193 Cerdanyola del Vallès (Spain)

J. L. González[*]

Gabinete de Coordinación y Estudios
Secretaría de Estado de Seguridad
c/. D. Quijote, esquina con c/. de los Artistas
28020 Madrid (Spain)

A. Sotoca[*]

Sección de Análisis del Comportamiento Delictivo
Unidad Técnica de Policía Judicial
Guardia Civil - c/. Príncipe de Vergara, 246
28016 Madrid (Spain)

X.-A. Tibau[*]

Institut d'Economia de Barcelona
Universitat de Barcelona
Facultat d'Economia i Empresa - Av. Diagonal, 690
08034 Barcelona (Spain)

Abstract

Bayesian Networks (BN) are probabilistic models very effective in criminal

*Corresponding author

Email addresses: delgado@mat.uab.cat (R. Delgado), jlga@interior.es (J. L. González), asotoca@guardiacivil.es (A. Sotoca), xavitibau@gmail.com (X.-A. Tibau)

[1]The author is supported by Ministerio de Economía y Competitividad, Gobierno de España, project ref. MTM2015 67802-P

profiling and classification. In this paper we complete the study of a BN built from a database set in a previous paper by the same authors, as a profiler in arson-caused wildfires: from the features of a particular provoked wildfire (crime variables), the model infers the author characteristics (author variables). One of the strengths of this type of model is that it allows to study the most significant and meaningful relationships among the variables. After analyzing the different author variables from a point of view of centrality, it is determined that the most valuable is *motivation*, which acts as a "bridge" connecting author and crime variables. On the other hand, a sensitivity analysis is carried out to analyze to what extent *motivation* is affected by the crime variables, purpose for which two new measures are introduced. For *motivation* we construct different classifiers from the crime variables, including a Naive Bayes and an Augmented Naive Bayes, and compare their behaviour with each other and with the constructed BN itself. Different performance measures are computed for this: accuracy, precision, recall, Cohen's kappa and Matthews correlation coefficient. We confirm the surprisingly good behaviour of Naive Bayes as classifier, although it is outperformed by the Augmented Naive Bayes.

Keywords: Bayesian network, Bayesian profiler, Bayes classifier, provoked wildfire, sensitivity analysis, centrality metric, performance measure

1. Introduction

Wildfire can be regarded as an environmental disaster which is one of the most relevant threats to nature ecosystems and human societies, according to FAO survey [14], and is triggered by either natural forces or anthropogenic activities. In this paper we are interested in the latter, the arson-caused wildfire, understood as the large and destructive fire caused by humans that spreads quickly out control over woodland or brush, calcining forest fuels located in the mount and affecting vegetation in principle was not destined to burn, which is one of major environmental problems in Spain and other areas with mediterranean climate.

Although arson is one potential cause of many fires, yet the rate of clarification of arson-caused wildfires is extremely low compared to other criminal activities. According to the interim report of the Ministry of Agriculture, Food and Environment of Spain [29], 11,928 wildfires happened in 2015 in Spain, of which 429 authors have been identified, representing a clarification rate of $6-6.5\%$ since the estimated percentage of wildfires in Spain that were deemed arson in 2015 ranges from 55 to 60%. This fact highlights the importance of developing methodologies that can assist investigators to hunt down the culprits by means of the implementation of the criminal profiling, consisting of inferring features (behavioural, criminological, socio-demographic and of personality) of the offender from the analysis of the evidences obtained from the place and time where fire started. Section 2 is devoted to this issue. In addition, classification of provoked wildfires by author motivation is of great importance from a criminological point of view. For this reason we dedicate Section 5 to construct different Bayes classifiers from our database set, and to compare their performance. Two intermediate sections are relevant. In Section 3 we deal with the question of justify from a quantitative point of view the pertinence of the election of motivation, among all the author variables, as classification criterion for arson-caused wildfires. Section 4 contains a sensitivity analysis in which we highlight the more influent crime variables for classification, by following two criteria, stating that the results obtained with both are equivalent to practical effects.

Bayesian Networks (also denoted BN from now on) are an increasingly popular methodology for modeling uncertainty in complex domains, and in the opinion of many Artificial Intelligence researchers, the most significant contribution in this area in the last years (see Korb and Nicholson [25]). BN were introduced in the 1920s as a tool that describes probabilistic understanding of cause and effect and other relations between variables, and are the soundest framework for reasoning under uncertainty. From then, this methodology has proved to be useful in assisting in many decision-making procedures in a variety of fields, and its use in risk analysis is gaining popularity and includes applications in areas as diverse as economy (Adusei-Poku [1]), public health (Spiegelhalter [36] and Cruz-Ramírez et al. [8]), environmental risk (Borsuk et al. [5] and Pollino et al. [31]), emerging diseases (Walshe and Burgman [39]), ecological disasters (Ticehurst et al. [37]) or nuclear waste accidents (Lee and Lee [27]). Within the criminology scope, BN have been introduced as a novel methodology in assessing the risk of recidivism of sex offenders in Delgado and Tibau [9].

Regarding wildfires, there are some recent works as Papakosta and Straub [30] in which the authors apply a BN to spatial datasets from the Mediterranean island of Cyprus. Dlamini develops a BN model in [11] to determine factors that influence wildfire activity in Swaziland, and in [12] provides estimations of fire risk in Swaziland using GIS and remote sensing data by means of a BN, generating fire risk maps. Focusing on the case of arson-caused wildfires in Spain, the only statistical approach to arsonist profiling stems from the work of González, Sotoca and collaborators [33]. Their approach to this problem is through the application of different techniques of multivariate data analysis. In this paper, however, we use the methodology of BN, that has only recently been used for criminal profiling (see, for instance, Baumgartner et al. [3] and [4]). Up to our knowledge, there are no previous studies on the use of BN for profiling of the perpetrator of a wildfire, nor for classification attending to wildfire motivation, which justifies the interest of the present work.

As in our preliminary work [10], from which the present paper is a considerable improvement and enlargement, in Section 2 we learn a BN model from available data and expert knowledge considering as variables both the features of the provoked wildfire (crime variables) and the characteristics of the author (author variables). Using this model we can predict the characteristics (profile) of the offender from the particular features of an arson-caused wildfire. The inferred arsonist characteristics include confidence levels that represent their expected probability, which could help investigators to know who sparked the blaze. Roughly speaking, we construct the most probable BN given the observed cases (learning procedure), and this model provides the optimal prediction for future cases (inference procedure). Considering that its effectiveness for prediction purposes essentially depends on the sufficiency of the training database, we introduce different performance metrics to address this issue, and conclude that our database is actually large enough for our purposes. Having done that, we learn (*train*) both BN structure and parameters, to subsequently validate them both by split-validation and by $k-$fold cross-validation, providing performance information.

Moreover, in Section 5 we compare the performance of the (full) BN as a classifier attending to the wildfire motivation, with other classifiers including a Naive Bayes and an Augmented Naive Bayes, which represents a novelty with respect to [10]. Motivation turns out to be the most important author variable. Indeed, motivation has a key role in the model since acts as a bridge connecting author and crime variables, fact that is evidenced using different centrality metrics borrowed from graph theory and the social networks analysis in Section 3.

In regard to the reasons for choosing this methodology, BNs have several properties that make them useful in many real-life data analysis in scenarios with uncertainty. In particular, they allow to combine data with domain knowledge and facilitate learning about causal relationships between variables. The probabilistic modeling of the interactions is one of the key points of BNs, and it allows for the estimation of risks and uncertainties better than models that only account for expected values. The usefulness of BNs lies in the fact that by using Bayes's theorem (named after Reverend Thomas Bayes, 1702-1762), one can calculate the distributions of the parents given the values of their children from the probability distributions of children given the values of their parents. That is, one can proceed not only from causes to consequences, but also deduce the probabilities of different causes given the consequences. Because BNs are solved analytically, once the model is compiled they can provide fast responses to different queries. The compiled form of a BN contains a conditional probability distribution for every variable with respect to its parents, and thus, unlike simulation models, allows to get instantly any distribution related to the variables.

In contrast with what happens with other machine learning architectures as neural networks, BN is not black-box, allowing to infer relationships between the variables from the model itself, generating new knowledge in the corresponding field. In our setting, these relationships will show, from a quantitative and objective point of view, the central paper of variable *motivation* in provoked wildfire studies, which gives rise to construct and compare different classifiers attending to motivation.

To summarize, the organization of the paper is as follows. In section 2 we introduce the database and brief background in BNs, with special emphasis on the foundations of learning structure and parameters. Internal consistency, robustness and validation of the constructed model as Bayesian profiler are also considered. Section 3 is devoted to analyze the central role played by the author variable A_{15} =motivation, by means of different centrality measures. In Section 4 we carry on a sensitivity analysis to know to what extent variable A_{15} is affected by the different crime variables in the model, by introducing two new different measures: the Maximum Sensitivity Range (MSR) and the Corrected Maximum Sensitivity Range (CMSR). From this analysis we identify four crime variables as those that most affect motivation. Finally, in Section 5 we analyze the constructed BN as a Bayes classifier for motivation, and compare it with other classifiers, including the Naive Bayes and the Augmented Naive Bayes. For comparison we use different score functions and compute some performance measures for classification including, as usual, accuracy, precision and recall (sensitivity), but also the something less commonly used F-score, Cohen's Kappa statistic and Matthews Correlation

Coefficient. A discussion about the surprisingly good behaviour of the Naive Bayes has been included. We also highlight the problems associated to the extension of some performance measures that are found usually in the literature, from binary to multi-class classification. Final Section 6 comprehends some concluding remarks and prospective about future work.

2. Materials and Methods

2.1. The Database

The 1,423 cases used to construct our probabilistic model come from a database of policing clarified arson-caused wildfires throughout the entire Spanish territory, under the leadership of the Prosecution Office of Environment and Urbanism of the Spanish state. According of the experts, $n = 25$ variables were considered because of their usefulness and predictive relevance. These variables are relatives to crime $(C_1, \ldots, C_{10})$ and to the author $(A_1, \ldots, A_{15})$, and are described in Table 1.

2.2. Basics on Bayesian Networks

BNs are graphical structures for representing the probabilistic relationships among a large number of variables, and for doing probabilistic inference with those variables.

A BN of a set of random variables $V = \{X_1, \ldots, X_n\}$ is a model that represents the joint probability distribution P over those variables. In our case, all the variables in V are categorical. This model is based on expert knowledge and/or sample observations that are assimilated through training (also known as learning). The graphical representation of the BN consists of a *directed acyclic graph (DAG)*, whose n nodes represent the random variables and whose directed arcs among the nodes represent conditional dependencies.

It is said that node X is a parent of node Y if there is an arc in the DAG from X to Y. We denote by $PA(Y)$ the set of parents of Y. If there is a *path* from node Z to node T (that is, a set of directed arcs connecting them), then we say that T is a descendant of Z. What characterizes the BN is *Markov condition*, which can be expressed as follows: *each variable in V is conditionally independent of any of its non-descendants known the state of all its parents*. Moreover, P is equal to the product of the conditional distributions of all nodes given the values of their parents, whenever these

Table 1: Variables in the model with outcomes.

Variables	Outcomes
C_1 = season	spring/winter/summer/autum
C_2 = risk level	high/medium/low
C_3 = wildfire start time	morning/afternoon/evening
C_4 = starting point	pathway/road/houses/crops/interior/forest track/others
C_5 = main use of surface	agricultural/forestry/ livestock /interface/recreational
C_6 = number of seats	one/more
C_7 = related offense	yes/no
C_8 = pattern	yes/no
C_9 = traces	yes/no
C_{10} = who denounces	guard/particular/vigilance
A_1 = age	up to 34 / 35 to 45 / 46 to 60 / more than 60
A_2 = way of living	with parents/in couple/single/others
A_3 = kind of job	handwork/qualified
A_4 = employment status	employee/unemployed/sporadic/retired
A_5 = educational level	illiterate/elementary/middle/upper
A_6 = income level	high/medium/low/without incomes
A_7 = sociability	yes/no
A_8 = prior criminal record	yes/no
A_9 = substance abuse	yes/no
A_{10} = psychological problems	yes/no
A_{11} = stays in the scene	no/remains there/remains and gives aid
A_{12} = distance home-scene	short/medium/long/very long
A_{13} = displacement means	on foot/ by car/ all terrain / others
A_{14} = residence type	village/isolated house/city/town
A_{15} = wildfire motivation	profit/gross negligence/slight negligence/pulsional/revenge

conditional distributions exist (*chain rule*):

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \,/\, PA(X_i))$$

for all the possible values (*instantiations*) x_i of X_i, $i = 1, \ldots, n$. The probability values of these conditional distributions, when unknown, are the parameters to be estimated of the BN.

Using a BN to compute a posteriori probability is called (Bayesian) inference, propagation or belief updating: from an evidence of the form $E = \{X_{i_1} = x_{i_1}, \ldots, X_{i_t} = x_{i_t}\}$, where $\{X_{i_1}, \ldots, X_{i_t}\} \subset V$ are the *evidence* variables, an inference consists in the computation of probabilities of the form

$$P(X_{j_1} = x_{j_1}, \ldots, X_{j_s} = x_{j_s} \,/\, E)$$

with $\{X_{j_1}, \ldots, X_{j_s}\} \subset V \setminus \{X_{i_1}, \ldots, X_{i_t}\}$ the *query* variables. Variables of the BN that do not appear either as query or evidence are treated as unobserved. The prediction of a query variable X given the evidence $E = \{X_{i_1} = x_{i_1}, \ldots, X_{i_t} = x_{i_t}\}$ is the instantiation of X with the largest posterior probability. That is, if $x_1, \ldots, x_r$ are the possible instantiations of X, then $x^* = arg\,max_{k=1,\ldots,r} P(X = x_k \,/\, E)$ is the prediction for X, and $P(X = x^* \,/\, E)$ is said to be the *confidence level (CL)* of the prediction.

2.3. Constructing the BN

The structure of the BN is learned from the training data set. We adopt the *score-based* structure learning method (*Search-and-score*), which is an algorithm that attempts to find the structure that maximizes the score function, in our case the Bayesian Information Criterion (BIC). This score function is intuitively appealing because it contains a term that shows how well the model predicts the observed data when the parameter set is equal to its Maximum Likelihood (ML) estimation, which is the log-likelihood function, and a term that punishes for model complexity.

Minimum Description Length (MDL) scoring metric is equivalent to the BIC score for Bayesian networks, while the Akaike's Information Criterion (AIC) differs from BIC but only on the penalty term, which is less than that of BIC, implying that AIC tends to favor more connected (complex) networks. Some works suggest that BIC consistently outperforms AIC (see [28]), reason by which we have chosen the first as score function, besides that we prefer a network not very connected that allows to interpret the relations expressed by the oriented arcs. The *greedy search-and-score algorithm* with the BIC score function has *local scoring updating*, that it, this algorithm only needs locally recompute a few scores in each step to determine the score of the model in the next step, and performs a search through the space of possible network structures by adding, removing or reversing an arc, given a current candidate, subject to the constraint that the resultant graph does not contain a cycle, and greedily chooses the one that maximizes the score function, stopping when no operation increases this score.

A problem with this algorithm is that it could halt at a candidate solution that locally maximizes the score function rather than globally maximizes it. One way for dealing with this problem is a variant of the algorithm named *iterated hill-climbing*, in which local search is done until a local maximum is obtained. Then, the current structure is randomly perturbed, and the process is repeated. Finally, the maximum over local maxima is used.

Taking into account that in this section we are interested in constructing a Bayesian profiler, we decided to construct a BN of type *diagnosis*, that is, inhibiting arcs from any crime variable to any author variable. We performed a comparison of the two models, one with this restriction on the process of learning structure, and the other without any restriction, obtaining very close BIC values (respectively, -30111.66 and -30200.33), ensuring a similar degree of adjustment of the two models to the data set, fact that justifies the election of the first one, whose DAG is drawn in Figure 1).

With regard to inference, we use an exact procedure that relies on transforming the BN into a *junction tree*. Once the junction tree has been built and its conditional probability distributions have been computed, we can input the evidence into it. The local distributions of the nodes to which the evidence refers (evidence variables) are then updated, and the changes are propagated through the junction tree.

Sufficiency of the training database to learn a BN model has been addressed in [10]. For that, the existence of a saturation point which is reached before attaining the total number of available cases (approximately at 900 cases), from which the BIC score does not improve even if we increase the number of training cases, shows that the training database is sufficient for learning.

2.4. Internal consistency

Before proceeding to the validation of the learned BN, the internal consistency of the training data set has been studied in [10]. To this end, some metrics of concordance/discordance between the learned DAG from training subsamples of increasing size (partial DAG) with respect to that learned from the whole training database (final DAG), are considered. The measures of concordance are the *sensitivity*, which is the proportion of arcs of the final DAG that are already in the partial DAG, and the *accuracy*, which is he proportion of arcs in partial DAG that are also in the final DAG. Both measures increase to 1 as the size of the subsamples increase, showing the internal consistency of our model.

Crime variables	Author variables
C_1 = season	A_1 = age
C_2 = risk level	A_2 = way of living
C_3 = start time	A_3 = kind of job
C_4 = starting point	A_4 = employment status
C_5 = main use of surface	A_5 = educational level
C_6 = number of seats	A_6 = income level
C_7 = related offense	A_7 = sociability
C_8 = pattern	A_8 = prior criminal record
C_9 = traces	A_9 = history of substance abuse
C_{10} = who denounces	A_{10} = history of psychological problems
	A_{11} = stays in the scene
	A_{12} = distance from home to the scene
	A_{13} = displacement mean
	A_{14} = residence type
	A_{15} = wildfire motive

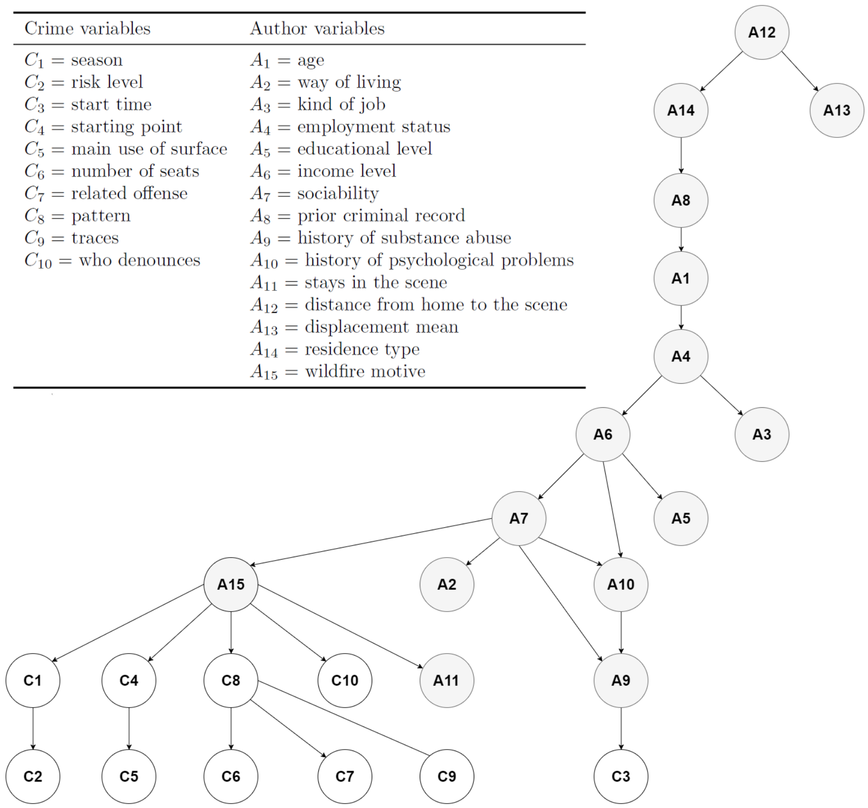

Figure 1: Learned structure (DAG) of the BN.

The first measure of discordance we considered is the *Structural Hamming Distance* (SHD) [23] that describes the number of changes that have to be made to the partial DAG for it to turn into the final DAG, and indeed, SHD tends to zero as the size of the subsamples increase. Other measures are the *Jaccard dissimilarity* [22], which is a measure of discrepancy between binary matrices, and the dissimilarity index of Sokal-Michener [20]. Note that both Jaccard and Sokal-Michener dissimilarity measures take values in [0, 1], the first one taking into account the concordances only in the sense of presence of arcs in both networks, and the second one also considering matches in absences, whilst SHD index takes values in $\mathbb{Z}^+$ and does not take concordances into account, only discrepancies. The evolution of these two dissimilarity measures also shows consistency, as for the SHD index, since they converge to zero as the size of the subsamples increase.

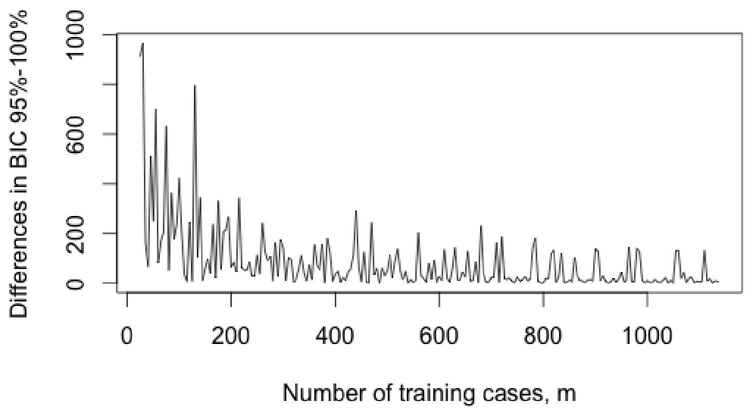

Figure 2: Discrepancies on the BIC score. Increment $\Delta m = 5$.

2.5. Robustness

A BN defined over a fixed universe of variables is said to be *structurally robust* if its arcs structure is insensitive to small changes in the training set. To evaluate if our BN model is robust in this sense, we randomly choose 5 % of the cases and remove them, from each of the subsamples of the training database, and learn the BN again. In this way we obtain a perturbed network from the 95 % of the cases for each subsample. Then, we compute the dissimilarity measures introduced in the previous subsection but now to compare the structure of the networks obtained from the 95 % and the 100 % of cases. We observe that discrepancies decrease to zero as subsample size increase. This it because the spurious dependencies that are introduced by the structure algorithm when the data is insufficient, disappear when the subsample size increases, and the structure of the network becomes more reliable and robust.

One last look at the robustness of the BN: we compute the discrepancy between the BIC score function for the network learned from each subsample with the 95 % and with the 100 % of cases. In Figure 2 we observe that, indeed, discrepancies converge to zero when the subsample size increases as expected.

2.6. Validation and assessment of predictive performance

Both split validation and $k-$fold cross validation were carried out in [10] to study the performance of our BN model as predictor for author variables from evidences on crime variables (profiling). With split validation, once learned the BN model from the training data set, we counted the matches between predicted and observed values of the arsonist variables for the cases of the validation data set, using as evidences the corresponding values of the crime variables. The success rate on predicting each arsonist variable from the evidence given by the crime variables, for the cases of the validating set, was called IPA (*Individual Predictive Accuracy*), and is obtained by dividing the number of matches by the total number of predictions (cases in the validation data set). See Table 2, where the average percentage of matches for all the arsonist variables, named OPA (*Overall Predictive Accuracy*), is also given. $k-$fold cross validation with $k = 10$ was also used in [10]. This procedure consists in reusing the whole data set to generate k splits into non-overlapping training and validate sets with approximately proportions $(k-1)/k$ and $1/k$, respectively. For each one of the k splits we obtained the IPA and OPA values just as we had done with split validation. From these $k = 10$ values we computed the mean and the coefficient of variation, also known as relative standard deviation RSD, which is a standardized measure of dispersion obtained by dividing the standard deviation by the absolute value of the mean (see corresponding columns in Table 2).

From Table 2, we see that arsonist characteristics A_3, A_7, A_8, A_9, A_{10} and A_{15} have IPA$\geq 60\,\%$, what makes them useful for narrowing the list of suspects of a provoked wildfire. Especially important are A_8, A_9 and A_{10}, since they are operative variables that greatly help the researcher to identify the offender, which is precisely the aim of profiling. Also of major importance is A_{15}, whose central role in our model will be evidenced in the next section. It should also be borne in mind that for any variable we choose as prediction the outcome that maximizes the probability, causing failures in prediction when the second most likely outcome has a close probability.

Finally, we computed the DIPA (*Disincorporate Individual Predictive Accuracy*), which is the percentage of matches between predictions and observed values, for each arsonist variable according to the prediction that was made for it from the evidence given by the crime variables. Particularly of interest is the DIPA for variable A_{15}: if prediction for A_{15} were gross negligence, the accuracy rate would be $79\,\%$, as can be seen in Table 3, while if instead it were revenge, for example, this rate plummets to $16.67\,\%$.

Table 2: IPA and OPA (split validation), and the corresponding mean and relative standard deviation (RSD) when using k-fold cross validation.

Arsonist variable IPA (%)	split validation	k-fold Mean	k-fold RSD
A_1 = age	27.57	30.30	0.10
A_2 = way of living	53.21	53.46	0.06
A_3 = kind of job	76.44	74.62	0.04
A_4 = employment status	34.68	37.74	0.09
A_5 = educational level	46.63	52.24	0.10
A_6 = income level	44.27	44.05	0.11
A_7 = sociability	69.19	66.70	0.07
A_8 = prior criminal record	80.07	79.41	0.06
A_9 = substance abuse	78.50	82.07	0.03
A_{10} = psychological problems	79.70	82.71	0.06
A_{11} = stays in the scene	49.10	49.35	0.09
A_{12} = distance home-scene	41.91	42.46	0.07
A_{13} = displacement means	44.53	42.79	0.11
A_{14} = residence type	40.52	45.12	0.12
A_{15} = wildfire motivation	61.05	62.94	0.07
Total OPA (%)	54.04	55.19	9.06×10^{-4}

Table 3: Disincorporate Individual Predictive Accuracy (DIPA) for A_{15}.

If prediction were	DIPA (%)
profit	46.67
gross negligence	79.00
slight negligence	51.65
pulsional	54.79
revenge	16.67

Since the predictive performance improves as the confidence level of the prediction does, better performance must be observed when only those predictions with a high confidence level are considered. Indeed, we computed the OPA but only considering those predictions for which the confidence level is greater than a fixed threshold. If our predictive procedure is reasonable in the sense that if we only consider "good" predictions (with high confidence level), performance is good, then OPA should grow towards 100 % as the threshold increases to 100 %, as well as the number of matches and the number of predictions should tend to coincide. Actually, this is what happens, as shown in Figure 3.

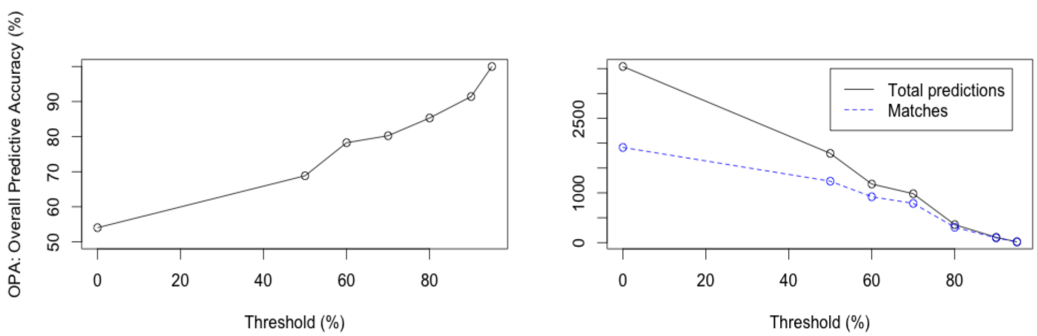

Figure 3: Evolution of the OPA, and evolution of the number of total predictions and the number of matches, with respect to the threshold for the CL.

3. Why classify provoked wildfires by motivation?

Variable A_{15}: wildfire motivation, which answers the question "Why did the author perpetrate the wildfire?", reveals itself as the natural choice when looking for a criterion by which to classify provoked wildfires, given that:

1. As argued in [33], from a psychological point of view it seems reasonable to classify provoked wildfires by motivation although other variables, such as the main use of surface or the starting point, could also be considered. It is remarkable the fact that motivation, as a classification criterion, is directly related to the author and not to the blaze itself.
2. All crime variables except C_3 are descendants of A_{15} in the DAG of the BN (see Figure 1). Informally speaking, we can say that A_{15} acts as a "bridge" connecting author and crime variables.
3. If we rank the nodes corresponding to author variables by means of a *centrality metric*, the key role of A_{15} is evidenced.

Indeed, in graph theory and network analysis, indicators of centrality identify the most important nodes within a graph. In this work two measures of centrality are considered:

a) *Geisberger's betweenness* ([6]) considers the number of times a node acts as a bridge in a graph along the shortest path between two other nodes. It is a variant of a centrality measure introduced for quantifying the control of a human on the communication between other humans in a social network by Freeman [15]. In his conception, nodes that have a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes have a high betweenness. With this measure, the farther away from the source, and hence the closer to the target, the more influential a node is. Given a node X_v, this measure is defined by:

$$\sum_{i,j\in\{1,\ldots,n\},i\neq j\neq v} \frac{d_{iv}}{d_{ij}} \frac{g_{ivj}}{g_{ij}} \quad \text{(with the convention } \frac{0}{0} = 0\text{)}$$

where g_{ij} denotes the number of paths from X_i to X_j, g_{ivj} denotes the number of them that pass through X_v, d_{iv} is the length of the shortest path between X_i and X_v and similarly for d_{ij}. Conceptually, high-betweenness nodes lie on a large number of non-redundant shortest paths between other nodes; they can thus be thought of as "bridges".

b) *Freeman's degree of centrality* ([16]) for each node is defined as the number of directed arcs starting or finishing at it.

The results are shown in Table 4. The numerical values have been obtained by scaling both Geisberger's betweenness and Freeman's degree of centrality, such that the sum of the set of the 25 variables in the model (including both author and crime variables) is 100. We see that, indeed, the most central node is A_{15}, whether one centrality measure is considered or the other, followed by A_7, A_6 and C_8.

At criminological level, it is a known fact that for drawing up profiles it is essential to address criminal motivation, which is evidenced in the crime scene by the signature elements ([13], [38]). However, a critical view from the scientific community is that many studies are not exhaustive in describing indicators corresponding to each motivation, or in building motivational typologies ([34]), so since late last century empirical studies from the psychological point of view have been carried out, which include both objective

Table 4: Centrality measures (in %) of the variables with the three higher values, which play a more prominent role. Ranks are indicated in brackets.

Geisberger's betweenness	Freeman's degree of centrality
$\mathbf{A_{15}}$ **: 19.76 (1st)**	$\mathbf{A_{15}}$ **: 11.54 (1st)**
A_7: 18.76 (2nd)	A_7: 09.62 (2nd)
A_6: 17.77 (3rd)	A_6, C_8: 07.69 (3rd)

variables of crimes and also the characteristics of offenders in order to establish profiles using multivariate statistical analysis. On arsonists, in particular, see [2], [7], [18], [21] and [32]. In this paper we follow the same vein, in the sense of including only the objective variables of the crime, but with the particular feature that the statistical relationships between these variables and motivation (A_{15}) will be considered, since the criminal behaviour is different depending on the motivation. The key role of motivation in the criminal act is indeed clearly caught by our Bayesian network, supporting its suitability as model for the analysis of wildfire arsonists. Thus, from information about "what" and "why" of an arson-caused wildfire, the network is able to provide information about "who", and therefore this model can be used to bear out previous studies on archetypes for wildfire arsonists.

4. Sensitivity Analysis

In the previous section we showed that variable A_{15} plays a central role in our model. Criminal motivation is essential for drawing up profiles and our BN captures this fact showing the centrality of A_{15} by means of two different centrality metrics.

The graphical display of the DAG (Figure 1) shows the relationships among variables learned from training data. We rank in Table 5 the relationships between A_{15} and the criminal variables which are its children, measured through the arc strength of the arcs from A_{15} to them.

Arc strength is a measure of force for each arc, while keeping fixed the rest of the network structure. The strength is measured by the BIC score gain/loss which would be caused by the arc's removal. In other words, it is the difference between the score of the network including the arc and without it. Negative values correspond to a decrease in the network score,

Table 5: Ranking by arc strength of crime variables who are children of A_{15}.

	Crime variable	Arc strength
(stronger) $\uparrow$	C_{10} = who denounces	-193.710772
	C_4 = starting point	-107.755970
	C_8 = pattern	-92.681936
(weaker) $\downarrow$	C_1 = season	-5.313557

while positive values correspond to an increase (the stronger the relationship, the more negative the difference).

In this section we carry on a sensitivity analysis for more deeply analyze to what extent variable A_{15} is affected by the different crime variables in the model, not only by its children. Formalization of sensitivity analysis is as follows. For each crime variable C_i, we would like to compute a measure of the effect of changes in it on the probability distribution of variable A_{15}. For that, we obtain from the BN the Conditional Probability Table (CPT) of A_{15} conditioned separately to each one of the crime variables (see tables 12-21 in the Appendix, where the CL of the prediction for A_{15} for each of the values of the crime variable is highlighted in boldface), and then consider two different (but related) metrics to measure discrepancies in the probability distribution of A_{15} depending on the instantiation of the crime variable (represented by columns in each CPT).

4.1. The Maximum Sensitivity Range (MSR)

We introduce the Maximum Sensitivity Range (MSR) associated to crime variable C_i as a measure (in %) of the extent to which this variable affects variable A_{15}, defined in the following way:

$$MSR(C_i) = \max_{y \in \mathcal{A}_{15}} \Big(\max_{x \in \mathcal{C}_i} P(A_{15} = y \,/\, C_i = x) - \min_{x \in \mathcal{C}_i} P(A_{15} = y \,/\, C_i = x) \Big)$$

where $\mathcal{A}_{15}$ denotes the set of outcomes of A_{15} and $\mathcal{C}_i$ that of C_i, $i = 1, \ldots, 10$. That is, for each fixed $y \in \mathcal{A}_{15}$ we compute the range ($= \max - \min$) of the set of probabilities $\{P(A_{15} = y \,/\, C_i = x)\}_{x \in \mathcal{C}_i}$, and then take the maximum when varying y in $\mathcal{A}_{15}$.

In Table 6 the standardized values of MSR to sum 100 % are collected. Comparing with Table 5, we see that the most influencing crime variables match: C_4 = starting point , C_{10} = who denounces , and C_8 = pattern . Variable C_1 = season has slightly lower influence. These crime variables are the children of A_{15} in the DAG (Figure 1). Unlike what happens with analysis carried on in Table 5, sensitivity analysis allows to quantify the influence of any of the crime variables on A_{15}, not only children.

The rest of crime variables have a weaker influence in motivation, especially C_3, the only one which is not descendant of A_{15}. It is quite natural: the longer the chain of uncertainty reasoning through paths in the DAG, the more tenuous influences, since each step adds additional uncertainty, and the resulting degree of belief will not be very sensitive to changes in the link probabilities.

4.2. The Corrected Maximum Sensitivity Range (CMSR)

Because MSR (introduced in Section 4.1) does not consider if different instantiations of a crime variable produce different predictions for variable A_{15}, it seems appropriate to introduce a correction in it that does take account of this fact.

Let $a = \#\mathcal{A}_{15}(= 5)$ and $c_i = \#\mathcal{C}_i$, where # denotes the cardinal of a finite set, and let d_i denote the number of different predictions obtained from the BN for A_{15} given the evidence $E = \{C_i = x\}$, x varying in $\mathcal{C}_i$. Then, define

$$\gamma(C_i) = \frac{d_i}{\min(a,\, c_i)} \in (0,\, 1]\,,$$

which is the proportion of different predictions actually obtained from the BN for A_{15} among the possible we could obtain from an evidence on C_i. Therefore, $\gamma(C_i)$ is a measure of influence of variable C_i on A_{15}, and we can use it to make the correction of MSR by introducing the Corrected Maximum Sensitivity Range (CMSR), which is a percentage, as:

$$\boxed{CMSR(C_i) = MSR(C_i) \times \gamma(C_i)}$$

In Table 6 we also collect the values of γ, and CMSR, ordering crime variables in descendent order with respect to CMSR. Note that CMSR has been standardized to sum 100 %.

Glancing at Table 6, we realize that crime variables are ordered in a slightly different way, according to the used criterion, either MSR or CMSR, although stand out C_4, C_8, C_{10} and C_1 as the most influential crime variables in motivation for both.

Table 6: Crime variables ordered by decreasing Corrected Maximum Sensitivity Range (CMSR)

Crime variable	MSR	γ	CMSR$\downarrow$
$\mathbf{C_4}$	**27.39**	3/5	**28.85**
$\mathbf{C_8}$	**14.61**	2/2	**25.65**
$\mathbf{C_{10}}$	**21.76**	2/3	**25.46**
$\mathbf{C_1}$	**12.47**	1/4	**5.47**
C_7	4.32	1/2	3.70
C_2	6.18	1/3	3.62
C_5	7.78	1/5	2.73
C_9	2.19	1/2	1.92
C_6	2.05	1/2	1.80
C_3	1.34	1/3	0.78
	100 %		100 %

5. Classifying provoked wildfires by motivation

Classification is a task of great importance in machine learning and data mining. In classification, the goal is to construct a model (classifier) that given evidence regarding features (crime variables in our case), assigns it a class (motivation) label. Naive Bayes (NB) is the simplest form of classifier. A NB is a Bayesian Network characterized by the fact that all the features are conditionally independent given the class known. The DAG of the NB, which is named BC1, is fixed (see Figure 4a) and the parameters are estimated from the training data set. Its simplicity makes it easy to apply, requiring less data to get a good result and being a very efficient model. Indeed, it is surprisingly good in classification, considering that the hypothesis of conditional independencies is hardly fulfilled in practice. This is so to the extent that sometimes it is difficult in practice to find more complex classifiers that improve its performance considerably. Papers [40, 41] and [17] on this subject can be consulted by the interested reader.

5.1. Constructing different classifiers

In Section 5.2 we will compare the behaviour of BC1 with three other classifiers: one is that obtained from the full BN in previous sections by cutting the bridge linking A_{15} with its ancestors; we name this model BC2 and its DAG has been drawn in Figure 4b. Regarding classification, model BC2 behaves exactly as the full BN whose DAG is in Figure 1. The second one is named BC3 and is like BC2 but obtained by optimizing the score function AIC instead of BIC; its DAG is in Figure 4c. The final classifier is an Augmented Naive Bayes (ANB) in which in addition to the directed arcs from class node A_{15} to the features (crime variables) in BC1, there also exist directed arcs among them. We name this classifier BC4 and its DAG is shown in Figure 4d. This classifier has been learned from the data by imposing all the directed arcs from A_{15} to the crime variables, and using the *greedy search-and-score* algorithm with the AIC score function. We use this score function here instead of BIC since it penalizes less for complexity, allowing to obtain an optimal classifier different from BC1. We will see that although the NB performs quite well, the ANB outperforms it. So, we solved the question of finding a classifier that improves the NB.

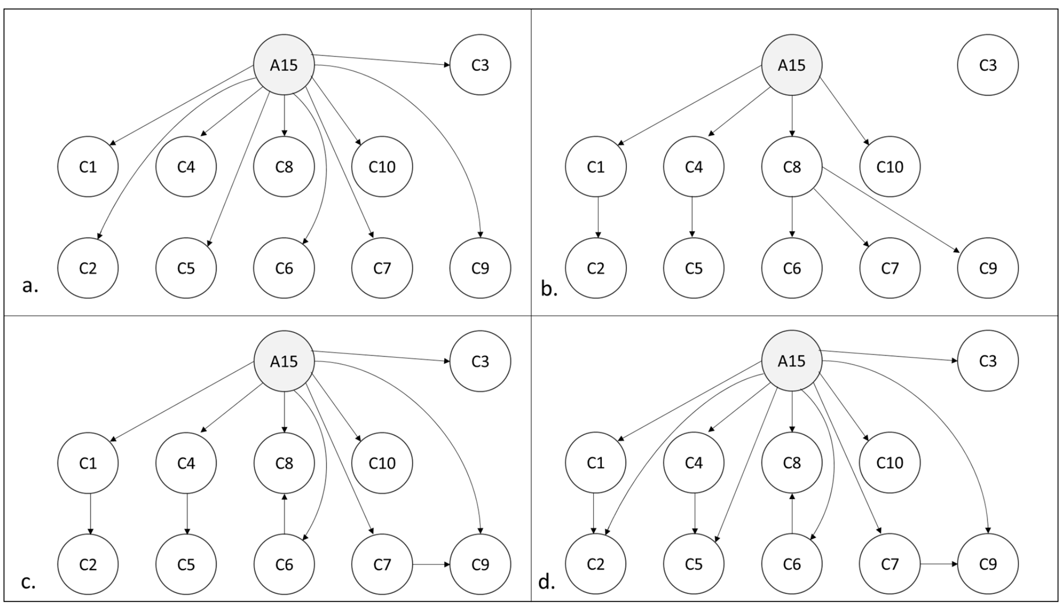

Figure 4: DAGs of the different classifiers: a) Naive Bayes BC1, b) BN learned with BIC score BC2, c) BN learned with AIC score BC3, d) ANB classifier learned with AIC score BC4.

Table 7 shows the values of the log-likelihood, BIC and AIC score functions for the four models. We can observe that when no penalization for complexity is considered, the best model (the most likely) is BC4, while in the other cases, the best is BC2 when considering BIC score, and BC3 when AIC has been used instead, which seems reasonable. In no case the best is BC1. Nevertheless, from the point of view of performance as classifiers, things are different. This disagreement could seem a contradiction but it is not, as shown in [17], where this question has been well studied. Indeed, it is possible to learn a network form data with a relatively good score that performs poorly as classifier, especially when there are many attributes. This happens in our case, in effect, with BC2, as we will see in the sequel.

Table 7: Comparing the four models by score functions.

Score function →	Log-likelihood	BIC	AIC
BC1	-12,670.60	-13,363.75	-12,867.60
BC2	-12,474.34	**-13,065.45**	-12,642.34
BC3	-12,286.36	-13,289.14	**-12,571.36**
BC4	**-12,105.02**	-14,022.61	-12,650.02

We use the four models to predict the value of variable A_{15} (that is, to *classify* the provoked wildfire by motivation), given the evidence of some of the crime variables $C_1, \ldots, C_{10}$. This is done via Bayes inference. Actually, variables C_4, C_8, C_{10} and C_1, which are the ones that most affect A_{15}, especially the first three (see Table 6), give all information needed from a provoked wildfire to classify its motivation. In this sense, Table 8 shows some particular values of one (or a combination) of these variables from which we can get the full range of possible classifications. Far from being an exhaustive list of evidences based on variables C_4, C_8, C_{10} and C_1 that allow to obtain each one of the possible classifications, the purpose of Table 8 is merely to show that, in fact, only taking into account these crime variables, without intervention of the rest, the full range of classes could be obtained.

In the same vein, we could ignore other crime variables except C_1, C_4, C_8 and C_{10} (*feature subset selection)* and construct a *selective* naive Bayes classifier (see [26] and [17]). We have really done it, resulting in a classifier with the same performance as the full model BC2. This finding has a double reading. On the one hand, it tells us that although they only introduce noise to the classifier, removing irrelevant attributes does not improves classification. On the other hand, neither does it make it worse, that is, we can reach the same performance with a simpler model, which results in a saving in calculations and time. In addition, this validates sensitivity analysis methods used

Table 8: Examples of classifications for wildfire motivation from some specific evidences on crime variables C_1, C_4, C_8 and C_{10}.

Evidence	Classification by motivation A_{15}
C_4 =road or forest track	Pulsional
C_4 =crops or C_{10} =guard	Gross Negligence
C_4 =houses or C_{10} =particular	Slight Negligence
C_1 =winter and C_8 =yes	Profit
C_1 =summer, C_4 =road, C_8 =no and C_{10} =particular	Revenge

to discriminate which are the most significant crime variables from the point of view of the classification according to motivation.

5.2. Comparing classifiers by performance

In addition to IPA and DIPA measures associated to A_{15} introduced in Section 2.6, we can compute some other performance measures used in machine learning classification for comparing classifiers. As mentioned in [24], this is one of the most critical questions in machine learning and can be carried out by means of different performance measures. Our setting is that of multi-class classification and then we must be careful since extensions of performance measures in binary classification to this more general setting is not always straightforward. We distinguish between performance classification measures that are *global*, in the sense they summarize in a single number he confusion matrix obtained by split-validation (see Table 9 for model BC2), where for the cases of the validation database set, observed values of A_{15} are given in columns while predicted values (classification) are given in rows, and the *averaged* performance classification measures that are obtained as the average over all the classes. In the sequel, we denote the latter group by the corresponding name with an asterisk. Actually, IPA belongs to the first group of measures and will be renamed as *accuracy*, as usual in this context, while DIPA is computed for each class and its average is, indeed, one of the measures of the second group (*precision**).

Table 9: Confusion matrix by split-validation using the BC2 model.

Pred.↓ Obs.→	Profit	G. Negligence	S. Negligence	Pulsional	Revenge
Profit	7	2	0	6	0
G. Negligence	10	79	1	9	1
S. Negligence	4	22	47	10	8
Pulsional	10	7	6	40	10
Revenge	0	0	2	3	1

From Table 9 we can compute the following global performance measures denoting the elements of the confusion matrix by $a_{i,j}$ with $i, j = 1, \ldots, r$, $r = 5$, and $\sum_{i=1}^{r} \sum_{j=1}^{r} a_{ij} = N$:

i) *Accuracy*$= \dfrac{\sum_{i=1}^{r} a_{ii}}{N}$ is the fraction of correctly classified cases, and it is naturally extended to multi-class from binary classification.

ii) *Cohen's Kappa statistic* (or simply *kappa*)$= \dfrac{P_0 - P_e}{1 - P_e}$ is intended to measure agreement between observed and predicted values in a contingency table as the confusion matrix. $P_0 = \frac{\sum_{i=1}^{r} a_{ii}}{N}$ is the relative observed agreement among observed and predicted classes (that is, accuracy), and P_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of randomly choose each category, that is, $P_e = \sum_{i=1}^{r} \frac{a_{i\cdot} \times a_{\cdot i}}{N^2}$, where $a_{i\cdot} = \sum_{j=1}^{r} a_{ij}$ is the sum of row i, and $a_{\cdot i} = \sum_{\ell=1}^{r} a_{\ell i}$ is the sum of column i.

iii) *Matthews Correlation Coefficient* MCC ([19], [24]), which is the extension to multi-class setting of the binary ϕ-coefficient (i.e., the square root of the averaged χ^2 statistic on the number of observed cases, for the 2×2 contingency table of the binary classification). Definition is as follows:

$$MCC = \frac{\sum\limits_{k,\ell,m=1}^{r} (a_{kk}\, a_{\ell m} - a_{mk}\, a_{k\ell})}{\sqrt{\sum\limits_{k=1}^{r} \Big(\big(\sum\limits_{\ell=1}^{r} a_{k\ell}\big) \big(\sum\limits_{u,v=1,\, u\neq k}^{r} a_{uv}\big) \Big)} \sqrt{\sum\limits_{k=1}^{r} \Big(\big(\sum\limits_{\ell=1}^{r} a_{\ell k}\big) \big(\sum\limits_{u,v=1,\, u\neq k}^{r} a_{vu}\big) \Big)}}$$

Moreover, from Table 9 we also can obtain the binary collapsed confusion matrix for each one of the five classes separately. We introduce the following commonly used notation for binary classification, say *class*$= +$, *no-class*$= -$.

t_p =true positive (number of correctly recognized cases as +),
f_p =false positive (number of cases with an incorrect assignation of +),
f_n =false negative (number of cases that are not recognized as +), and
t_n =true negative (number of correctly recognized cases as −).

As an example we show in Table 10 the binary collapsed confusion matrix corresponding to class **Profit** for the BC2 classifier.

Table 10: Binary collapsed confusion matrix by split-validation for class **Profit** using the BC2 model.

Pred.↓ Obs.→	Profit (+)	No Profit (−)
Profit (+)	$t_p = 7$	$f_p = 8$
No Profit (−)	$f_n = 24$	$t_n = 246$

We compute the usual averaged performance classification measures *, first obtaining them for each class from the corresponding binary collapsed confusion matrix by using the following definitions, and then averaging them over the five classes (see, for example, [35]):

iv) *Precision* $= \dfrac{t_p}{t_p + f_p}$ measures the proportion of class agreement among all the cases with a given predicted classification.

v) *Recall* (also known as *Sensitivity*) $= \dfrac{t_p}{t_p + f_n}$ measures the proportion of class agreement among all the cases that actually belong to a given class.

vi) *F-Score* is a measure that combines precision and recall into a single number, being their harmonic mean. Harmonic mean is used rather than the common arithmetic mean since both precision and recall are expressed as proportions between zero and one, and then can be interpreted as rates. This measure has the advantage of describing two aspects of the model performance into a single number, and provides a convenient way to compare several models side by side. Its definition is:

$$\text{F-Score} = \frac{2}{\frac{1}{\text{Precision}^*} + \frac{1}{\text{Recall}^*}}$$

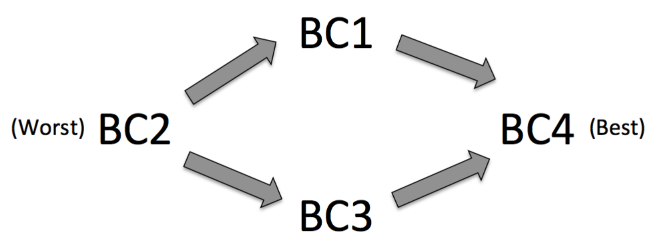

Figure 5: Ranking of the four classifiers.

Results for the four models are shown in Table 11, where the best (highest) values correspond to the Augmented Naive Bayes BC4, according to all the considered performance measures. Then, BC4 is the best classifier of the four we compare, but BC1 has also a quite good performance, being simpler, while BC2 has the worst performance (see Figure 5).

Table 11: Comparing classifiers: performance measures.

	Accuracy	kappa	MCC	Precision*	Recall*	F-Score
BC1	0.65263	0.53204	0.53677	0.55165	0.55080	0.55123
BC2	0.61053	0.46903	0.47636	0.49755	0.48430	0.49084
BC3	0.64912	0.52743	0.53520	0.57281	0.55552	0.56403
BC4	**0.65965**	**0.53789**	**0.54151**	**0.57803**	**0.56210**	**0.56995**

Unlike what is done in other works, such as for example [35], we do not consider nor *specificity* nor accuracy as an averaged performance measure. The corresponding definitions for each class would be $(t_p+t_n)/(t_p+t_n+f_p+f_n)$ and $t_n/(f_p+t_n)$, respectively, and then the average would be calculated for the five classes. Observe that in both formulas the number of true negatives t_n appears as a quantity that computes positively at the performance classification level. Although this is obviously true in the binary case, in which non-classification in (or non belonging to) one class necessarily implies classification in (or belonging to) the other, it is not in the multi-class setting, where t_n includes both well and misclassified cases. For this reason, we decided to introduce accuracy as a global measure, not as an average, and do not consider specificity as performance measure in the multi-class context.

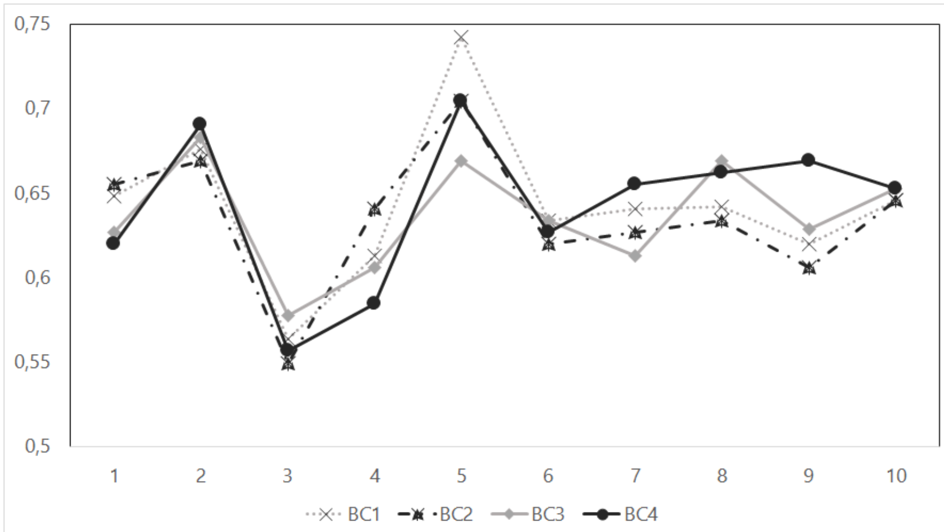

Figure 6: Accuracy curves comparing the four classifiers. The horizontal axis lists the subsets.

We could extend the previous comparative study based on different performance measures and split-validation, to k-fold cross-validation (with $k = 10$, for example). That is, we could carry out the comparison of the models from ten different realizations of the six performance measures already considered. To not extend too much, we will only show some of the results in figures 6 and 7.

6. Conclusion

All fires are destructive, especially wildfires. Once the blaze begins, there's no telling how far it can spread. Fire's consequences can be deadly, and are almost always devastating. This is why we must tenaciously investigate the cause of every single wildfire. It is therefore crucial the development of tools that help the investigators in the task of gather evidence, analyze data and determine whether it is an arson-caused wildfire and, if it is the case, who set the blaze and why.

Bayesian Networks can be a useful addition to the toolkit of arson-caused wildfire investigators, especially if their work is related to profiling or classification. Explicit accounting for uncertainty can add substantial insight to the

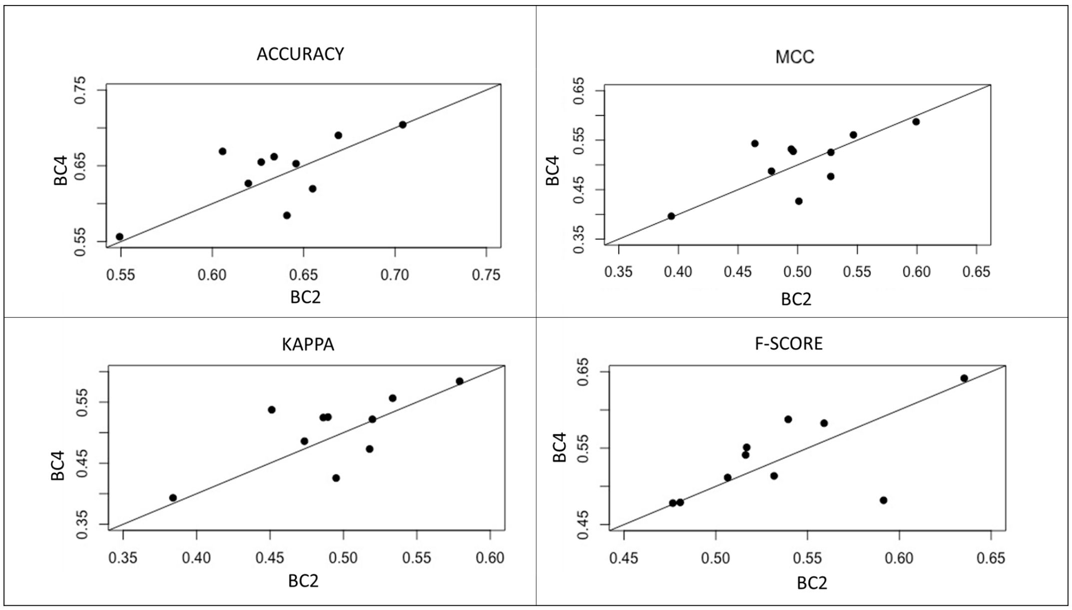

Figure 7: Scatter plots for paired comparisons between the best (BC4) and the worst (BC2) of the models. Each one corresponds to a different performance measure. Each point represents a data set, where the x coordinate of a point is the performance measure for model BC2 and the y coordinate is the same performance measure for model BC4. Thus, points above the diagonal line correspond to data set on which BC4 performs better with respect to the performance measure than BC2, while points below the diagonal correspond to the opposite case. For the four performance measures we can see that BC4 outperforms BC2.

question of identifying the perpetrator of a criminal act, and to classify it, and the graphical representation of model structures and probability distributions is very useful to understand the relationship between variables, allowing to contribute to the development of knowledge about the phenomenon under study, and to communicate this knowledge in a visual friendly way.

BNs are not a household name in criminal investigation yet. However, they are gaining popularity in the field and are likely to establish their position as one of the standard methods of analysis in fields with uncertainty. The learned BN represents a good approach to reality, from which data has been produced, and can be analyzed numerically to extract hidden knowledge.

We present BN as a novel methodology in provoked wildfires investigations, both for the author's profile and for classification by motivation, after showing the key role of this feature of the author by means of different centrality metrics. We learn the BN from the database and validate it, by estimating the accuracy of predictions. This work also shows internal consistency, robustness and validity of our model. Therefore, on one side we can use learned BN to predict the profile of the offender from the information about a particular arson-caused wildfire, and obtain confidence levels for the predictions of the arsonist variables. On the other side, we compare the BN as a Bayesian Classifier to classify wildfires by motivation with other classifiers we build, including a Naive Bayes and an Augmented Naive Bayes, the latter showing the best performance although the Naive Bayes also performs quite well. This is in concordance with previous literature, in which the surprisingly good performance of Naive Bayes classifier with respect to more complex models is highlighted and analyzed.

We think this approach is really innovative and helpful. The introduction of MSR and CMSR measures in the sensitivity analysis of motivation to reveal which crime variables are most important, is novel and displays useful for our purpose. Prospects for future work are to extend this analysis to different real situations including, but not limited to, other types of crimes. Alternative algorithms of structure learning should be tried and compared with the ones we use in this paper.

Appendix

The Appendix contains the Conditional Probability Tables of A_{15} conditioned separately to any of the crime variables. Confidence level (CL) of the prediction for A_{15} conditioned to each of the possible outcomes of each crime variable is highlighted in boldface. In each CPT, MSR indicates the value of the *Maximum Sensitivity Range* of A_{15} with respect to the crime variable, which is the maximum of the Range column.

Table 12: CPT of A_{15} conditioned to C_1 (in %).

$C_1 \rightarrow$	Spring	Summer	Autum	Winter	Range
Pulsional	15.20	31.97	6.61	9.35	25.36
Gross Negligence	**48.02**	**34.24**	**47.11**	**45.93**	13.78
Slight Negligence	27.66	17.01	26.46	28.05	11.04
Profit	3.95	9.30	14.05	13.82	10.10
Revenge	5.17	7.48	5.79	2.85	4.63
				MSR:	**25.36**

Table 13: CPT of A_{15} conditioned to C_2 (in %).

$C_2 \rightarrow$	High	Medium	Low	Range
Pulsional	25.53	17.18	12.96	12.57
Gross Negligence	**38.56**	**43.86**	**45.57**	7.01
Slight Negligence	20.47	24.86	26.49	6.02
Profit	8.83	8.81	10.42	1.61
Revenge	6.58	5.29	4.56	2.02
			MSR:	**12.57**

Table 14: CPT of A_{15} conditioned to C_3 (in %).

$C_3 \rightarrow$	Morning	Afternoon	Evening	Range
Pulsional	18.93	19.48	21.63	2.70
Gross Negligence	**42.66**	**42.23**	**39.94**	2.72
Slight Negligence	23.68	23.51	22.76	0.92
Profit	9.20	9.20	9.55	0.35
Revenge	5.55	5.59	6.12	0.57
			MSR:	**2.72**

Table 15: CPT of A_{15} conditioned to C_4 (in %).

$C_4 \rightarrow$	Pathway	Road	Houses	Crops	Interior	Forest Track	Others	Range
Pulsional	25.44	**35.43**	12.16	0.39	11.76	**56.08**	7.54	55.69
Gross N.	**36.69**	15.75	40.54	**63.78**	**47.71**	14.86	**53.77**	48.92
Slight N.	16.57	12.60	**41.89**	29.92	28.76	6.08	30.65	35.81
Profit	10.65	18.11	1.35	5.12	11.11	14.19	6.03	16.76
Revenge	10.65	18.11	4.05	0.79	0.65	8.78	2.01	17.46
							MSR:	**55.69**

Table 16: CPT of A_{15} conditioned to C_5 (in %).

$C_5 \rightarrow$	Agricultural	Forestry	Livestock	Interface	Recreational	Range
Pulsional	11.24	26.59	22.12	19.44	19.53	15.35
Gross Negligence	**51.08**	**35.27**	**39.62**	**39.33**	**42.13**	15.81
Slight Negligence	26.45	20.02	22.00	25.82	23.48	6.43
Profit	7.34	11.32	10.04	8.24	9.23	3.98
Revenge	3.90	6.80	6.20	7.17	5.63	3.27
					MSR:	**15.81**

Table 17: CPT of A_{15} conditioned to C_6 (in %).

$C_6 \rightarrow$	One	More	Range
Pulsional	18.87	23.04	4.17
Gross Negligence	**42.65**	**39.19**	3.46
Slight Negligence	23.94	21.03	2.91
Profit	8.91	11.01	2.10
Revenge	5.64	5.74	0.10
		MSR:	**4.17**

Acknowledgments

The authors would express their acknowledgment to the Prosecution Office of Environment and Urbanism of Spain for providing data and promote research.

[1] K. Adusei-Poku, Operational Risk management - Implementing a BN for Foreign

Table 18: CPT of A_{15} conditioned to C_7 (in %).

$C_7 \rightarrow$	Yes	No	Range
Pulsional	27.47	18.89	8.58
Gross Negligence	**35.07**	**42.67**	7.60
Slight Negligence	17.88	23.93	6.05
Profit	13.31	8.91	4.40
Revenge	6.27	5.60	0.67
		MSR:	**8.58**

Table 19: CPT of A_{15} conditioned to C_8 (in %).

$C_8 \rightarrow$	Yes	No	Range
Pulsional	**39.44**	9.73	29.71
Gross Negligence	22.44	**50.36**	27.92
Slight Negligence	19.44	4.29	15.15
Profit	9.44	30.33	20.89
Revenge	7.22	5.29	1.93
		MSR:	**29.71**

Table 20: CPT of A_{15} conditioned to C_9 (in %).

$C_9 \rightarrow$	Yes	No	Range
Pulsional	16.41	20.86	3.99
Gross Negligence	**44.90**	**40.91**	3.99
Slight Negligence	25.68	22.54	3.14
Profit	7.64	9.92	2.28
Revenge	5.37	5.77	0.40
		MSR:	**4.45**

Exchange and Money Market Settlement. Unpublished PhD thesis. Göttinger University (2005).

[2] L. Almond, L. Duggan, J. Shine, D. Canter, Test of the arson action system model in an incarcerated population, *Psychology, Crime & Law* **11 (1)** (2005) 1-15.

[3] K. Baumgartner, S. Ferrari, G. Palermo, Constructing Bayesian networks for criminal

Table 21: CPT of A_{15} conditioned to C_{10} (in %).

$C_{10} \rightarrow$	Guard	Vigilance	Particular	Range
Pulsional	18.36	19.93	18.02	1.91
Gross Negligence	**61.84**	**66.67**	22.44	44.23
Slight Negligence	0.00	0.00	**44.17**	44.17
Profit	16.91	9.80	7.07	9.84
Revenge	2.90	3.59	8.30	5.40
			MSR:	**44.23**

profiling from limited data, *Knowledge-Based Systems* **21** (2008) 563-572.

[4] K. Baumgartner, S. Ferrari, C. Salfati, Bayesian Network Modeling of Offender behaviour for Criminal Profiling, *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, 2702-2709 (2005).

[5] M.E. Borsuk, C.A. Stow, K.H. Reckhow, A BN of eutrophication models for synthesis, prediction, and uncertainty analysis, *Ecological Modeling* **173** (2004) 219-239.

[6] U. Brandes, On Variants of Shortest-Path Betweenness Centrality and their Generic Computation, *Social Networks* **30** (2008) 136-145.

[7] D. Canter, K. Fritzon, Differentiating arsonists: A model of firesetting actions and characteristics, *Legal and Criminological Psychology* **3** (1998) 73-96.

[8] N. Cruz-Ramírez, H.G. Acosta-Mesa, H. Carrillo-Calvet, L. Alonso Nava-Fernández, R.E. Barrientos-Martínez, Diagnosis of breast cancer using BN: A case study, *Computers in Biology and Medicine* **37** (2007) 1553-1564.

[9] R. Delgado, X.A. Tibau, Las Redes Bayesianas como herramienta para la evaluación del riesgo de reincidencia: Un estudio sobre agresores sexuales, *Revista Española de Investigación Criminológica* **Artículo 1, Número 13** (2015). (In Spanish).

[10] R. Delgado, J.L. González, A. Sotoca, X.A. Tibau, A Bayesian network profiler for wildfire arsonists. In P.M. Pardalos et al. (Eds.), *Proceedings of the Second International Workshop in Machine Learning, Optimization, and Big Data MOD 2016, Lecture Notes in Computer Science (LNCS)* **10122**, pp. 379-390, Springer International Publishing AG 2016. DOI: 10.1007/978-3-319-51469-7 31.

[11] W.M. Dlamini, A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland, *Environmental Modelling & Software* **25** (2010), 199-208.

[12] W.M. Dlamini, Application of Bayesian networks for fire risk mapping using GIS and remote sensing data, *GeoJournal* **76** (2011) 283-296.

[13] J.E. Douglas, C.M. Munn, Modus operandi and the signature aspects of violent crime. In J.E. Douglas, A.W. Burgess, A.G. Burgess, R.K. Ressler (Eds.), *Crime classification manual* (pp. 259-268). New York: Lexington Books, 1992.

[14] FAO *Fire Management - Global Assessment 2006. A Thematic Study Prepared in the Framework of the Global Forest Resources Assessment 2005* FAO Forestry Paper **151** Rome (2007).

[15] L.C. Freeman, A set of measures of centrality based upon betweenness, *Sociometry* **40** (1977) 35-41.

[16] L.C. Freeman, Centrality in Social Networks I: Conceptual Clarification, *Social Networks* **1** (1979) 215-239.

[17] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian Network Classifiers, *Machine Learning* **29** (1997) 131-163.

[18] K. Fritzon, D. Canter, Z. Wilton, The application of an action system model to destructive behaviour: the examples of arson and terrorism, *Behavioural Sciences and the Law* **19** (2001) 657-690.

[19] J. Gorodkin, Comparing two K-category assignments by a K-category correlation coefficient, *Computational Biology and Chemistry* **28** (2004) 367-374.

[20] J.C. Gower, P. Legendre, Metric and Euclidean properties of dissimilarity coefficients, *Journal of Classification* **3** (1986) 5-48.

[21] H. Häkkänen, P. Puolakka, P. Santtila, Crime scene actions and offender characteristics in arsons, *Legal and Criminological Psychology* **9** (2004) 197-214.

[22] P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura, *Bulletin de la Société Vaudoise des Sciences Naturelles* **37** (1901) 547-579. (In French).

[23] M. Jongh, M.J. Druzdzel, A comparison of Structural Distance Measures for Causal Bayesian Network Models, *Recent Advances in Intelligent Information Systems*, Publisher: EXIT, Editor: M.A. Klopotek, A. Przepiórkowski, S.T. Wierzchon, K. Trojanowski, ISBN: 9788360434598 (2009).

[24] G. Jurman, S. Riccadonna, C. Furlanello, A comparison of MCC and CEN Error Measures in Multi-class Prediction, *PLoS ONE* **7(8)** (2012) e41882. doi:10.1371/journal.pone.0041882

[25] K.B. Korb, A.E. Nicholson, *Bayesian Artificial Intelligence*, Second Edition. CRC Press (Taylor & Francis Group), 2011.

[26] P. Langley, S. Sage, Introduction to selective Bayesian classifiers. In R. López de Mántaras & D. Poole (Eds.), *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence* (1994) 399-406. San Francisco, CA: Morgan Kaufmann.

[27] C. Lee, K.J. Lee, Application of BN to the probabilistic risk assessment of nuclear waste disposal, *Reliability Engineering and System Safety* **91(5)** (2006) 515-532.

[28] Liu, Z., Malone, B., Yuan, Ch., Empirical evalutation of scoring functions for Bayesian network model selection, BMC Bioinformatics 2012, 12(Suppl 15):S14 (2012). http://www.biomedcentral.com/1471-2105/13/S15/S14

[29] MAGRAMA - Ministerio de Agricultura, Alimentación y Medio Ambiente (2016). Los Incendios Forestales en España: Avance informativo. 1 de enero al 31 de diciembre de 2015. (In Spanish). http://www.magrama.gob.es/es/desarrollo-rural/estadisticas/iiff_2015_def_tcm7-416547.pdf

[30] P. Papakosta, D. Straub, A Bayesian network approach to assessing wildfire consequences, *Proceedings ICOSSAR, 2013*, New York.

[31] C.A. Pollino, O. Woodberry, A. Nicholson, K. Korb, B.T. Hart, Parameterisation and evaluation of a BN for use in an ecological risk assessment, *Environmental Modelling and Software* **22** (2007) 1140-1152.

[32] P. Santtila, H. Häkkänen, L. Alison, C. Whyte, Juvenile firesetters: Crime scene actions and offender characteristics, *Legal and Criminological Psychology* **8** (2003) 1-20.

[33] A. Sotoca, J. L. González, S. Fernández, D. Kessel, O. Montesinos, M. A. Ruíz, Perfil del incendiario forestal español: aplicación del perfilamiento criminal inductivo, *Anuario de Psicología Jurídica 2013* **23** (2013) 31-38. (In Spanish).

[34] B. Snook, R.M. Cullen, C. Bennell, P.J. Taylor, P. Gendreau, The criminal profiling illusion: What's behind the smoke and mirrors?, *Criminal Justice and behaviour* **35** (2008) 1257-1276.

[35] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Information Processing and Management* **45** (2009) 427-437.

[36] D.J. Spiegelhalter, Incorporating Bayesian ideas into healthcare evaluation, *Statistical Science* **19** (2004) 156-174.

[37] J.L. Ticehurst, L.T.H. Newham, D. Rissik, R.A. Letcher, A.J. Jakeman, A BN approach for assessing the sustainability of coastal lakes in New South Wales, Australia, *Environmental Modelling and Software* **22 (8)** (2007) 1129-1139.

[38] B. Turvey, *Criminal profiling. An introduction to behavioural evidence analysis.* California: Elsevier, 2008.

[39] T. Walshe, M. Burgman, A Framework for Assessing and Managing Risks Posed by Emerging Diseases, *Risk Analysis* **30(2)** (2010) 236-249.

[40] H. Zhang, The Optimality of Naive Bayes, *In Proceedings of FLAIRS2004 Conference*, American Association for Artificial Intelligence, 2004.

[41] H. Zhang, L. Jiang, J. Su, Hidden Naive Bayes, *In Proceedings of the Seventeeth Florida Artificial Intelligence Research Society Conference 2004*, 562-567, American Association for Artificial Intelligence, 2005.