Characterising the Personality of the Public Safety Offender and Non-offender using Decision Trees: The Case of Colombia

VÍCTOR HUGO MASÍAS¹,*, MAURICIO A. VALLE², JOSÉ J. AMAR³, MARCO CERVANTES³, GUSTAVO BRUNAL⁴ and FERNANDO A. CRESPO⁵

¹Department of Management Control and Information Systems, Universidad de Chile, Santiago, Chile
²Facultad de Economía y Negocios, Universidad Finis Terrae, Santiago, Chile
³Facultad de Humanidades y Ciencias Sociales, Universidad del Norte, Barranquilla, Colombia
⁴Fundación Universitaria Luis Amigó, Centro Regional Montería, Montería, Colombia
⁵CEDYTEC, Dirección de Investigación, Facultad de Ingeniería y Administración, Universidad Bernardo OHiggins, Santiago, Chile

Abstract
The aim of this paper was to create a decision tree (DT) to identify personality profiles of offenders against public safety. A technique meeting this requirement was proposed that uses the C4.5 algorithm to derive decision rules for personality profiling of public safety offenders. The Mini-Mult test was used to measure the personality profiles of 238 individuals. With the test results as our database, a C4.5 DT was applied to construct rules that classify each profile into one of two groups, those without and those with records of offences against public safety. The model correctly classified 80% of the personality profiles and delivered a set of decision rules for distinguishing the profiles by group, and the principal personality profiles were interpreted. We conclude that DTs are a promising technique for analysing personality profiles by their offender or non-offender status. Finally, we believe that the development of a classifying model using DT may have practical applications in the Colombian prison system. Copyright © 2016 John Wiley & Sons, Ltd.

Key words: public safety offender; decision trees; personality; psychological profile; decision rules; Colombia

INTRODUCTION

A gap has long existed between criminal profiling research and actual forensic practice (Alison, Goodwill, Almond, Heuvel, & Winter, 2010). A problem currently facing researchers in this field is transmitting knowledge about personality profiles to forensic and clinical psychologists. As Turvey has stated, ‘[g]iven the advanced level of knowledge required, it is unclear how a criminal profiler could perform these examinations
competently without receiving a baseline of formal education and ongoing training in these areas from non-law enforcement forensic and behavioural scientists’ (Turvey, 2011, p.149). One way of addressing this problem would be to use alternative methods of analysis for supporting the transmission of knowledge regarding criminal profiles.

In Colombia (South America), the focus of the present article, multiple studies have attempted to explain the issue of violence from the perspectives of law, economics, and sociology, but very few forensic studies of crime have been conducted from the standpoint of personality theory (Amar, Cervantes, Brunal, & Crespo, 2011; Uzlov & Araslanov, 2013). The latter approach posits that personality traits can be used to distinguish between criminals and law-abiding persons (Akers, 1997; Brathwaite, 2009; Gearing, 1979; Gray et al., 2004; Hampson & Kline, 1977; Khan, 2014). Based on this theory, the present article uses decision trees (DTs) to identify patterns that can differentiate between persons with and without a criminal record.

Existing research shows that uncovering criminal personality profiles is a multi-dimensional analytical problem (Alison, West, & Goodwill, 2004; D. Canter, 2000; Egger, 1999). There are a range of personality theories that contribute to explaining the offender’s personality (Jackson & Bekerian, 1997; Pinizzotto & Finkel, 1990) and many instruments for exploring the criminal’s personality profile (Archer, Buffington-Vollum, Stredny, & Handel, 2006; K. M. Davis & Archer, 2010; Kline, 2013; Mullen & Edens, 2008; Nikolova, Hendry, Douglas, Edens, & Lilienfeld, 2012; Rabin, Borgos, & Saykin, 2008; Wood, Nezworski, Lilienfeld, & Garb, 2009). A variety of statistical techniques are also available for analysing these personality tests (Bennell, Goodwill, & Chinneck, 2014; Goodwill, Allen, & Kolarevic, 2014; Goodwill et al., 2013; Homant & Kennedy, 1998; Neuman & Wiegand, 2000). However, criminal personality profiling has occurred largely in the absence of a well-defined profiling framework and an empirical knowledge base (Becker, 2004; Snook, Cullen, Bennell, Taylor, & Gendreau, 2008; Snook, Eastwood, Gendreau, Goggin, & Cullen, 2007).

Criminal behaviour analysts frequently use logistic regression (LR) among other statistical techniques (Kleck, Tark, & Bellows, 2006). However, studies have shown that even experienced researchers do not always have the training to properly interpret the results of LR analysis (King, Tomz, & Wittenberg, 2000; Mood, 2010) or the necessary skills to communicate them (Wouda & van de Wiel, 2012). This practical problem is described by Tonkin, Woodhams, Bull, Bond, and Santtila (2012), who explain that LR ‘requires the analyst to perform several analytical steps (…) Although this process could be automated, the analyst would still need to understand how the logistic regression function has arrived at a particular decision so that they can explain their decision-making processes to investigating officers and/or the courts (which they often have to do). This is inherently difficult for a decision that is based on a mathematical equation, which can be difficult to break down into its constituent parts’ (p. 237–238). Clearly, then, there is a need for techniques that can create models of criminal profiles, which are easy to interpret without losing their analytical potential.

In this context, we propose an experimental study using DTs to create risk indices of public safety offenders. According to the penal code of Colombia, crimes against public safety include conspiracy to commit a public safety offence, terrorism, and threats. In 2013, a total of 43,026 public safety offences were registered nationally (Cárdenas-Guzmán, 2013). These crimes have a high impact on the safety of the civilian population because they can lead not only to the death of civilians but can also cause insecurity and terror, as well as the displacement of entire families. A clear example of this is that in 2013, death threats caused forced intra-urban displacements in Medellín, the capital of Colombia, and there
were attempts against the life, integrity, freedom, and personal safety of 6,004 people (Personería de Medellín, 2014). Crimes against public safety, understood as acts of violence, are among several factors that contribute to Colombia being the country with the second highest number of internally displaced people in the world (Ruiz, 2011; Salaya & Rodríguez, 2014; Vargas, 2012). Therefore, to obtain data on the personality of public safety offenders, which are generally members of armed criminal groups within their own community, it is important to develop strategies to ensure public and citizen safety.

The present article proposes the use of DTs to analyse personality profiles that characterise individuals according to whether or not they have criminal records for public safety offences as defined in Colombian law. A DT has been defined as ‘a way to represent rules underlying data with hierarchical, sequential structures that recursively partition the data’ (Murthy, 1998, p. 345). DT is a technique that learns to recognise patterns in data and has performed well in various areas of application (Quinlan, 1993; Rokach, 2007; Wu & Kumar, 2009), including criminal and forensic investigations (Abbasi & Chen, 2005; Bennell, Mugford, Ellingwood, & Woodhams, 2014; Berk & Bleich, 2013; Gutierrez & Leroy, 2008; Ngai, Hu, Wong, Chen, & Sun, 2011; Yang & Olafsson, 2011). Furthermore, DTs are used as pedagogical support tools to produce easy-to-interpretable models generally (Anaya, Luque, & García-Saiz, 2013; Breiman, 2001b; Jormanainen & Sutinen, 2012). Among DT’s various positive characteristics, those that have prompted us to apply them to the study of criminal personality profiling include the following (Zhao & Zhang, 2008, p. 1956):

- They are easy to understand.
- They are easily converted to a set of decision rules.
- They can classify both categorical and numerical data (but the output attribute must be categorical).
- There are no a priori assumptions about the nature of the data.

To explore the performance of DTs, we will apply the C4.5 algorithm to the problem of classifying individuals by their criminal record status for public safety offences using the Mini-Mult personality test (Kincannon, 1968; Table 1). The analytical problem thus becomes one of identifying what personality subscale values contribute to the correct profiling of public safety offenders. This is particularly challenging because identifying public safety offenders requires long forensic training and expert knowledge.

The rest of the paper is organised as follows. First, some basic background on DTs is provided. Second, the experimental method is described. Third, the results of the experiment are presented, and finally, our conclusions and some practical implications are discussed.

**DECISION TREES AND THE C4.5 ALGORITHM**

In this section, we introduce DTs and the C4.5 algorithm. DTs are one of the most popular type of classification model because they are simple and easy to understand. They are constructed by recursively partitioning a space containing $N$ instances, progressively dividing the set into groups according to some division rule that attempts to maximise the homogeneity or purity of the variable class (Giudici & Figini, 2009). In each stage of the DT procedure, the rule is applied to a predictor variable that is split, thus establishing how the instances are partitioned. The end result is the final partition of the instances.

A DT is made up of internal (or decision) nodes, edges, and leaf nodes. The internal nodes represent the predictive variable used to partition the instance space, while the edges
are the specific values of this variable that determine the partition. The leaf nodes are the tree’s terminal nodes with class labels. Thus, the internal nodes and the edges configure the decision rules that assign a new instance to a leaf node depending on its attributes, and thereby determine the class it belongs to.

There are two types of DTs: regression trees and classification trees. In the first type, the variable to be predicted is continuous, while in the second type, it is discrete. With classification trees, the decision rules for determining the class of an instance are found in the leaf nodes. Thus, each tree path represents a classification rule that is simple to interpret graphically. A DT is, therefore, a flow-chart-like tree structure in which each internal node denotes a test on an attribute, each branch represents an outcome of the test, and the leaf nodes, as already mentioned, represent the classes. The topmost node in a tree is the root node and is selected using an information gain measure.

The division rule is a distinctive element in a DT that represents the mechanism by which the instances in a given group form the tree’s nodes. The rule is used to choose the best partition for a predictor variable. The choice criterion is the maximisation of a goodness-of-fit measure. In the case of the C4.5 algorithm (Quinlan, 1993; Wu & Kumar, 2009), the splitting criterion is the information gain ratio. This avoids problems of bias towards choosing multi-valued attributes (Dai & Ji, 2014). The measure takes into account the probability that an instance belongs to one of the classes. If, following Quinlan (1993), we let $D$ be the set of instances, $C$ be the number of classes, and $p(D, j)$ be the proportion of cases in $D$ that belong to the $j^{th}$ class, then the uncertainty regarding the class $D$ belongs to is expressed as

$$\text{Info}(D) = \sum_{j=1}^{C} p(D, j) \times \log_2 p(D, j)$$

Then, the information gain ratio with $k$ possible classes is then defined as

$$\text{Gain}(D, T) = \text{Info}(D) - \sum_{i=1}^{k} \frac{|D_i|}{|D|} \times \text{Info}(D_i)$$

The information gain ratio expresses the ability of a predictor variable to discriminate between instances in any of the $k$ possible classes in terms of information gained. It is influenced by the number of possible classes that an instance may belong to and especially by the partition of subset $D_i$. Thus, all possible partitions of subset $D_i$ are executed, and the one that maximises information gain is chosen.

In general terms, the DT construction procedure is as follows. Given a set of data $D$ with $n$ instances, a measurement function is applied to all of the predictor variables in order to find the candidate that produces the best partition (as indicated, for example, by the information gain ratio). Then, for each of the partitions obtained, the process of searching for the best predictor variable is repeated using the measurement criterion until the partitions are assigned to a class or a stopping criterion is satisfied. For a classification tree, this criterion may be a certain minimum information gain ratio value.

Finally, training a DT does not require an assumption on the distribution of data compared with parametric methods, such as LR. In the parametric method, it is assumed that the data should have a normal distribution and a similar variance. DT can be used in cases where data assumptions are not satisfied. For further discussion of this issue, refer to Genser, Cooper, Yazdanbakhsh, Barreto, and Rodrigues (2007), Khemphila and Boonjing (2010), and Stark and Pfeiffer (1999).
In the next section, we describe an experimental setup for profiling public safety offenders using DT.

**MATERIAL AND METHODS**

To build a model capable of classifying individuals by their criminal record status for public safety offences, we designed the experimental setup depicted in Figure 1, which combines psychometric and machine learning analysis techniques. As can be seen, the method consists of six stages: personality diagnostics using a psychometric instrument (Mini-Mult), data coding, running experiments using a DT, calculating performance measures of the DT model obtained, comparison of the DT model performance with LR as a benchmark using statistical hypothesis testing, and extracting and interpreting personality profiles from the DT model. These stages are described in turn later.

**Measuring personality using a psychometric instrument**

The analysis unit is the ‘personality profile’ measured using the Mini-Mult personality test (Kincannon, 1968). The test employs an abbreviated Minnesota Multiphasic Personality Inventory (MMPI; Pope, Butcher, & Seelen, 2006) scale with subscales for measuring possible psychopathological behaviours that are latent in the subject (Table 1). Mini-Mult is regularly applied in Colombia by the Instituto Nacional Penitenciario y Carcelario (INPEC), the corrections service of the country’s Ministry of Justice, to assess personality characteristics of individuals who enter the national prison system. A translated version of the test was validated for the Colombian population (Diazgranados & Amar, 2011; Instituto Nacional Penitenciario y Carcelario & Universidad Pontificia Bolivariana, 2009).

We used this test because it is one of the most frequently used by psychologists in Colombia’s penitentiary context (Instituto Nacional Penitenciario y Carcelario & Universidad Pontificia Bolivariana, 2009). The Mini-Mult test began to be applied by INPEC in 2004. This test is used to evaluate eight clinical personality scales, which are required to assess both the accused and convicted inmate population that participate in the penitentiary attention and treatment programmes, as stipulated in Law 65 of 1993 and Law 1709 of 2014. These laws indicate that a quantitative personality study must be conducted permanently on convicted inmates so that under measurable criteria combined with clinical assessment, it can be decided which penitentiary treatment programmes the inmate can have access to.

**Data and coding**

The coded sample contained 238 individual personality profiles. All participants were informed of the scope and objectives of our work and signed consent forms permitting the use of their personality test data for research purposes provided that anonymity and confidentiality were maintained. Of the 238 personality profiles registered, 112 were from individuals who had criminal records for public safety offences (Criminal Group), and 126 were from individuals with no such records (non-criminal group). The codification of the independent variables (personality subscales) and the dependent variable (criminal record status) is described later.
Figure 1. Methodological setup.
Codification of independent variables

The personality profiles were codified in eight numerical variables representing clinical subscales defined by the Mini-Mult test (Kincannon, 1968). The personality scales measured by the test (Kincannon, 1968) are summarised in Table 1.

<table>
<thead>
<tr>
<th>Personality Subscales</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypochondriasis (Hs)</td>
<td>Obsession with bodily symptoms</td>
</tr>
<tr>
<td>Depression (D)</td>
<td>Hopelessness, slowed thinking</td>
</tr>
<tr>
<td>Hysteria (Hy)</td>
<td>Use of physical or mental symptoms to avoid problems</td>
</tr>
<tr>
<td>Psychopathic deviate (Pd)</td>
<td>Disregard for social customs, emotional shallowness</td>
</tr>
<tr>
<td>Paranoid (Pa)</td>
<td>Delusions, suspiciousness</td>
</tr>
<tr>
<td>Psychasthemia (Pr)</td>
<td>Worry, guilt, anxiety</td>
</tr>
<tr>
<td>Schizophrenia (Sc)</td>
<td>Bizarre thoughts and perceptions</td>
</tr>
<tr>
<td>Hypomania (Ma)</td>
<td>Overactivity, excitement, impulsiveness</td>
</tr>
<tr>
<td>Validity Subscales</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Subscale of Lie</td>
</tr>
<tr>
<td>F</td>
<td>Subscale of integrity</td>
</tr>
<tr>
<td>K</td>
<td>Subscale of correction</td>
</tr>
</tbody>
</table>

Codification of dependent variables

The dependent variable is the criminal record status of the individual, that is, whether or not he or she has a record of public safety offences. Individuals with such records (Criminal Group, n = 112) were selected from the Monteria Municipal Prison in Córdoba, Colombia, where a considerable number of persons convicted of public safety offences are held. For our purposes, this type of offence was defined to include the following categories listed in Colombia’s Penal Code:

- **Conspiracy to commit a public safety offence**: Agreement between persons to carry out genocide, forced disappearance, torture, forced displacement, murder, kidnapping, or extortion.
- **Terrorism**: Creating or maintaining a state of anxiety or terror in the general public through acts that endanger the life, physical integrity, or liberty of any person.
- **Threats**: Frightening or threatening a person, family, community, or institution by any means with intent to cause alarm, anxiety, or terror in the general public.

The volunteers with no criminal records (non-criminal group, n = 126) were selected from residents of the municipalities within the city of greater Córdoba (Tierralta Cereté, Lorica, and Sahagún). Persons with a history of mental retardation and psychiatric or organic brain disorders were excluded.

Running experiments

As already explained, two different techniques were used to classify the psychological profiles. Our first choice was the C4.5 algorithm because the trees it creates are easy to interpret and perform well, but for purposes of comparison, we also used the LR. To estimate the performance of the two techniques, we applied the 10-fold
cross-validation approach (10-fold-CV) (Japkowicz & Shah, 2011; Kohavi, 1995), which trains on 90% and tests on 10%; the latter repeated 10 times on a different 10% each time.

Performance measures

The results of the DT and LR models were classified by a binary confusion matrix (Rokach, 2007; Figure 2). Based on this matrix, four performance measures were applied: overall accuracy (Alberg, Park, Hager, Brock, & Diener-West, 2004), precision (J. Davis & Goadrich, 2006), recall (J. Davis & Goadrich, 2006), and area under the receiver operating characteristic curve (ROC area) (Hanley & McNeil, 1982). Their definitions are given by the formulas in Table 2. The relative performance of the two models was thus measured by comparing the values obtained for these indicators. The ROC curve can be constructed by plotting the false positive values on the x-axis and the true positive values on the y-axis with different classification thresholds. For further information of this measure, refer to Fawcett (2004).

![Confusion matrix for classifying the predictions of each model. TP: is the number of correct predictions that an instance is positive (true positive), FN: is the number of incorrect predictions that an instance is negative (false negative), FP: is the number of incorrect predictions that an instance is positive (false positive), and TN: is the number of correct predictions that an instance is negative (true negative).](image)

Table 2. Performance measures for classification tasks

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>(\frac{TP + TN}{TP + FP + FN + TN})</td>
</tr>
<tr>
<td>Precision</td>
<td>(\frac{TP}{TP + FP})</td>
</tr>
<tr>
<td>Recall</td>
<td>(\frac{TP}{TP + FN})</td>
</tr>
<tr>
<td>Receiver operating characteristic (AUC)</td>
<td>(\frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right))</td>
</tr>
</tbody>
</table>

*Note: Formulas are based on the confusion matrix classifications of Figure 1. Each performance measure varies between 0 and +1. In every case, values close to +1 indicate a good performance.*
Statistical evaluation of decision tree and logistic regression performance

A comparative evaluation of DT and LR classifications was determined using McNemar’s test \( \chi^2_{Mc} \) (Bostanci & Bostanci, 2013; Dietterich, 1998). In a binary classification problem, a contingency matrix of 2*2 can be constructed (Figure 3) to compare the performance of C4.5 and LR according to Formula 3. If the statistic is \( Z=0 \), then both algorithms have a similar performance. If \( Z \) diverges from zero, the classifiers have a different performance. The null hypothesis \( (H_0) \) was that the two models performed similarly, while the alternative hypothesis \( (H_1) \) was that they did not. In particular, we compare the respective performances of DT and LR to assess whether DT’s performance is significantly different than LR and in which direction.

\[
Z = \frac{(|N_{sf} - N_{fs}| - 1)}{\sqrt{N_{sf} + N_{fs}}}
\]  

Figure 3. Possible results of two classification algorithms. \( N_{ss} \): is the number of times when both algorithms succeed, \( N_{ff} \): is the number of times when both algorithms failed, and \( N_{sf} \) and \( N_{fs} \): are cases where one of the algorithms succeeded and the other failed (Bostanci & Bostanci, 2013).

Extracting decision rules from the decision tree model

In this stage, we show how to extract and interpret personality profiles using the DT technique.

Softwares

For the statistical analyses, the following R packages were used: C4.5 implementation (Hornik, Buchta, & Zeileis, 2009), LR (Kuhn, 2008), McNemar’s test (Bilder & Loughin, 2014), and ROC (Sing, Sander, Beerenwinkel, & Lengauer, 2005).

RESULTS

This section sets out the performance measure results for the two classification techniques DT and LR and displays the tree generated by the best-performing algorithm. Finally, an interpretation of the DT is proposed.
Comparison of decision tree and logistic regression performance

The DT and LR performance measure results are summarised in Table 3 in the form of the average performance scores and their standard deviations. As can be seen, the LR algorithm achieved the highest values on the accuracy, precision, recall, and ROC area indicators.

The McNemar significance test ($\chi^2_{Mc}$) did not reject the null hypothesis that states that both models perform similarly ($\chi^2_{Mc} = 0.0238$, df = 1, $p$-value = 0.8774). Thus, the performance of C4.5 was found to be statistically similar to LR.

Table 3. Performance measures of decision tree and logistic regression in classifying personality profiles of individuals by their criminal record status for public safety offences

<table>
<thead>
<tr>
<th>Performance measures</th>
<th>C4.5</th>
<th>Logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (AVG)</td>
<td>0.77</td>
<td>0.802</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.087)</td>
<td>(0.0943)</td>
</tr>
<tr>
<td>Precision (AVG)</td>
<td>0.767</td>
<td>0.804</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.124)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>Recall (AVG)</td>
<td>0.742</td>
<td>0.791</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.157)</td>
<td>(0.0657)</td>
</tr>
<tr>
<td>ROC (AVG)</td>
<td>0.77</td>
<td>0.803</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.087)</td>
<td>(0.0906)</td>
</tr>
</tbody>
</table>

Note: AVG, average performance scores; SD, standard deviation.

Extracting and interpreting decision rules from decision tree model

The DT graph generated by C4.5 is displayed in Figure 4. The nodes represent the personality subscales (the independent variables) that characterise the psychological profiles of the individuals with and without criminal records (the dependent variables). The root node is Sc, the most important subscale for classifying personality profiles, that is, the attribute with the highest information gain or greatest entropy reduction, and therefore the best splitting point. Each path from a root node to a leaf node represents a decision rule that characterises a personality profile. The conjunction of variables and values in a decision rule is called a rule antecedent (the ‘IF’ part), and each class to be predicted (leaf nodes) is called a rule consequent (the ‘THEN’ part). These set of rules are equivalent to risk indicators.

The complete set of rules correctly classified 80% of cases, a good level of classification performance. As can be seen, the DT has 27 leaves, and therefore, 27 of these paths passing through internal nodes (Appendix B). Each of the paths consists of a conjunction of Mini-Mult scales (variables) whose numerical values describe the two groups of individuals. For example, an individual with a personality profile given by $Sc = 97$, $Pd = 69$, and $Pt = 75$ would be classified in the criminal group given that the decision rule number 1 states that IF $Sc > 70 \land Pd > 58 \land Pt > 56$, THEN ‘Criminal Group’. However, an individual with a personality profile given by $Sc = 46$, $Pd = 50$, and $Pa = 45$ would be classified in the non-criminal group because the decision rule number 13 states that IF $Sc \leq 70 \land Pd \leq 70 \land Sc \leq 53 \land Pa > 41$, THEN ‘Non-Criminal Group’.

For teaching purposes, it is not only a DT’s performance that matters but also its size and the number of rules it requires. This is so because smaller size and fewer leaves mean that the graph a student will have to learn to interpret will have fewer rules and objects. The smaller size and fewer leaves of the C4.5 DT will thus aid in simplifying the interpretability of the model obtained.
Figure 4. The tree characterises personality profiles of criminal and non-criminal individuals. The rectangular green nodes represent Mini-Mult subscales (independent variables), the hexagonal nodes represent the variable values, and the oval nodes (leafnodes) indicate which of the two groups (individuals with and without a criminal record for public safety offences). Each path from the root node (i.e. Sc) to the leaves is a decision rule that classifies different personality profiles.
Interpreting the decision tree model

A key aspect is the interpretation of the main findings generated by the DT. First, the number of rules makes it possible to interpret how homogeneous or heterogeneous the classified groups are. This is because the combination of attributes and their values describe groups that do not overlap. The tree obtained uses 15 rules to classify individuals with records for crimes against public safety, whereas 17 rules classify individuals with no criminal records. Therefore, the first interpretation of the DT is that the personality profiles are heterogeneous for both groups.

A second interpretation of the tree comes from the rules that permitted classification of a substantial percentage of the total sample. In effect, when exploring the ability of the rules to cover the cases, we found that with the sole use of four decision rules, 68.2% of all the cases were covered (162 of 238 cases). Of these 162 cases, the model correctly classified 97.5%, that is, only four cases (2.5%) were classified erroneously by these four rules. The decision rules extracted from the DT are presented in Table 4.

A third source of information comes from the psychological interpretation of these four decision rules. We think that these are fundamental to interpreting the predominant personality in both study groups, particularly because personality is relatively stable in adults. The personality profiles associated with the rules are as follows:

- **Rule #3**: This profile suggests a personality disorder and describes a distinctly alienated individual, with interpersonal difficulties and occasionally losing touch with reality. In addition, this individual can have moderate or marked antisocial and impulsive behaviour in addition to presenting some obsessive ideas and/or rigid compulsive behaviours.
- **Rule #10**: This profile suggests a personality disorder and describes a distinctly antisocial and amoral individual. This individual has a deficit of empathy and an inclination to violent behaviour and contradicting authority. However, staying in touch with reality and the ability to develop relationships with others are maintained. Finally, this is a rational personality, whose emotional reactions (i.e. anger) stay between limits controlled by repression mechanisms.
- **Rule #13**: This profile suggests a sociable personality that does not lose touch with reality or the ability to develop empathy with others. In addition, it is a personality inclined to trust in social relationships. It is a personality open to experience.
- **Rule #17**: This profile suggests a sociable personality that does not lose touch with reality or the ability to develop empathy with others. This individual may be suspicious of the motives of others. Finally, this profile presents low levels of hypochondria, and depressive emotions are kept relatively controlled.

Beyond the individual analysis, it should be mentioned that the DT technique allows interpretation of the results on a graph (Figure 4), in a table (Table B1), and the accuracy of the

Table 4. Decision rules that can classify a higher percentage of offenders against the public safety and non-offenders

<table>
<thead>
<tr>
<th>Rule ID</th>
<th>IF (Rule antecedent)</th>
<th>THEN (Rule consequent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>Sc &gt; 70 ∧ Pd &gt; 58 ∧ Pr &gt; 56</td>
<td>Criminal group</td>
</tr>
<tr>
<td>#10</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy &gt; 59 ∧ Hy ≤ 70</td>
<td>Criminal group</td>
</tr>
<tr>
<td>#13</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc ≤ 53 ∧ Pa &gt; 41</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td>#17</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs ≤ 52 ∧ D ≤ 68</td>
<td>Non-criminal group</td>
</tr>
</tbody>
</table>
model can be ascertained to classify the personality of the study groups (Table 3). Therefore, we believe that the DT technique can support the task of identifying criminal profiles, in particular if we consider that one of the meanings of the word *profile* is defined as ‘a graph, table, or list of scores representing the extent to which a person, field, or object exhibits various tested characteristics or tendencies’ (Profile, 2015). These are fundamental aspects that must be complemented by other data and analyses, interviews, and clinical observation.

**DISCUSSION AND CONCLUSION**

The purpose of this study was to use the DT technique for criminal personality profiling. The results presented earlier show that the C4.5 DT model was able to correctly classify 80% of the psychological profiles of individuals according to their criminal record status for public safety offences.

**Comparison with previous research on personality profiles**

The results of this study showed that the dimensions *Sc* and *Pd* are fundamental to classifying public safety offenders correctly. Indeed, it was observed that *Sc* increased dramatically in rule #3, and *Pd* markedly so in rule #10. We think that these two types of psychological profiles discovered are the results of a personality differentiation process.

There is evidence to show that some dimensions of the personality are able to classify homicidal population, and that these display different personality patterns (code types). On one hand, the work by Craig (2008) reported that 11 out of 30 studies using the MMPI test showed the inmate population of male murderers exhibits elevated scores in the variables *Pd* and *Sc*. Also, Dahlstrom, Panton, Bain, and Dahlstrom (1986) reported that the individual murderers on death row present elevated scores in the dimensions *Pd* and *Sc*. However, McKee, Shea, Mogy, and Holden (2001) compared three groups of female murderers (filicide, mariticide, and simple homicide groups) and found that the dimension *Pd* was elevated in two groups and that in the three groups, *Sc* and *Pa* had high scores. Finally, Friedman, Bolinskey, Levak, and Nichols (2014) indicate that individuals with a profile where the combination of codes *Pd–Sc* or *Sc–Pd* are significant:

‘(…) become social isolates or nomads, whereas others become involved in antisocial or criminal activity. When these individuals commit crimes, they often seem senseless, brutal, and poorly planned and may include sexual or homicidal attacks. Individuals with this codetype often become involved with others who are equally marginally adjusted. They were often born to parents who were rejecting of them and they, in turn, often have children early and develop ambivalent attachments to them’ (Friedman *et al*., 2014, p. 339).

However, as has been shown, the exploration made with the DT indicates that there is no rule of thumb. On the contrary, it shows us a detailed phenomenology in which the criminal personality can be formed. This means that the DT, by means of a non-overlapping set of rules, reveals a hierarchy of partitions that characterises the different personalities of public safety offenders. Indeed, prior research proposed different prototypical profiles: undifferentiated psychopath (Hill, Haertzen, & Davis, 1962), with high scores for the dimension *Pd*; classic psychopath profile (Eisenman, 1980; Hill *et al*., 1962), profile with marked elevations on the *Pd* and *Ma* scales; neurotic psychopath, profile with marked elevations on
the D and Pa scales; neurotic triad (Cox, Chapman, & Black, 1978), profile with marked elevations on the Hs, D, and Hy; psychotic triad, profiles with marked elevations on the Pa, Pt, and Sc; psychotic tetrad (Polimeni, Moore, & Gruenert, 2010), profile with marked elevations on the Pa, Pt, Sc, and Ma; depressed neurotic psychopath (Hill et al., 1962), with elevations on the D, Hy, and Pt scales. The profiles discovered by the DT are also different from the clusters described by Megargee (1984). Therefore, exploring the constructed DT makes it possible to go beyond describing the personality profiles according to ideal prototypes or dichotomous hierarchies.1

One of the more interesting findings of our DT model is that Sc (i.e. failure to recognise what is real) is the root node, and therefore the most important classifying variable. This variable indicates that detachment from reality is the attribute that helps separate individuals with or without records against public safety. Thus, rules #13 and #17 show that in the people with no criminal record, both Sc and Pd are maintained at levels equal to or lower than the levels usually considered in clinical terms as indicators of a personality pathology. In fact, in a study conducted by Boscán et al. (2002), university students were compared with the Mexican inmate population, and Sc was found to be high. However, it should be remembered that it is the combination of attributes and values that contributes to distinguishing between individuals with and without criminal records. This shows how a DT reveals the complexity of the combination of personality attributes and values in criminal profiling.

Finally, it has been documented that there are moderator variables that contribute to certain personality attributes appearing high. For example, Megargee and Bohn (1979) found that men who have been incarcerated for more than 2 years showed elevations in the dimensions Pd, Sc, and Ma of the MMPI. Additionally, intelligence quotient, social status, and occupational status correlations influence the values reached in inmate population personality profiles (Holcomb & Adams, 1982; Holcomb, Adams, & Ponder, 1984; McDonald & Paitich, 1981). Culture and nationality are also sources of variability in the criminal personality (Boscán et al., 2000; Catalá-Miñana, Walker, Bowen, & Lila, 2014). More research is required to explain the relation among these variables with the groups studied here.

Comparison of decision tree and logistic regression with previous research on offender risk assessment

There are several offender risk assessment applied studies that show that the classifying performance between DT and LR is similar in different classification problems: Steadman et al. (2000) compared the Chi-Squared Automatic Interaction Detector (CHAID) and LR to develop violence risk assessment models; S. Thomas et al. (2005) compared the performance of the Classification And Regression Tree (CART) decision algorithm and LR to develop models to predict violence in psychotic illness; Rosenfeld and Lewis (2005) compared the CART decision algorithm and LR regarding the problem of assessing violence risk in stalking cases; R. Berk, Sherman, Barnes, Kurtz, and Ahlman (2009) compared random forest and LR to construct a predictive model of murder within a population of probationers and parolees; Liu, Yang, Ramsay, Li, and Coid (2011) compared the CART, neural network and LR in predicting violent re-offending; Neuilly, Zgoba, Tita, and Lee (2011) compared classification tree analysis, random forest, and LR to predict recidivism in homicide offenders; another important work was conducted by Hamilton, Neuilly, Lee, and Barnoski (2014), who suggested that machine-learning techniques,

1For a discussion about this topic, see D. Canter (2000) and D. V. Canter, Alison, Alison, and Wentink (2004).
including DT algorithms, have a classifying performance similar to LR in offender risk assessment. Finally, the results of the present study should be seen as complementary to those of an earlier work (Tonkin et al., 2012), in which the relative performances of DT and LR were compared for the case linkage problem.

Although DT’s performance was found to be very similar to that of LR, the former’s advantage over the latter is that it provides a simple graphic representation of the logical rules that characterise the personality profiles of the groups under study. By contrast, LR can only aspire to determining which variable has the greatest impact on the classification of a set of individuals into a given group. The authors found that DT can perform similarly to LR, but nevertheless suggested that more research was needed before DT could be considered a practical alternative. We concur with this judgement. In our view, the goal should be to find the optimal balance between these properties of classification capability and interpretability, thus achieving a decision support system that is useful as well as reliable.

Contribution of decision tree technique to forensic science

An important theoretical advantage to using the DT is that it captures the nonlinear relations between the predictor variables and the class. However, the LR supposes there is a linear relation between the predictor variables and the class. Yet this theoretical assumption of each technique has its own consequences in the empirical investigation. As Landwehr, Hall, and Frank (2003) indicate, LR ‘fits a simple (linear) model to the data, and the process of model fitting is quite stable, resulting in low variance but potentially high bias. The latter [a DT], however, exhibits low bias but often high variance: it searches a less restricted space of models, allowing it to capture nonlinear patterns in the data, but making it less stable and prone to overfitting’ (Landwehr et al., 2003, p. 241). From this point of view, the DT shows us it can find the nonlinear relations if they exist in the data in the process of learning by induction. In addition, DTs contribute to an exploration of the entire instance space in the search for non-overlapping decision rules. Indeed, other techniques used to classify an inmate population such as the discriminant analysis (Klecka, 1980) or the analysis of variance (Girden, 1992) can result in models with potentially superposed classes that do not necessarily try to cover all the instance space. However, the set of decision rules extracted by the DT are non-overlapping rules that include the entire instance space. Finally, whereas in LR analysis, there is little consensus on evaluating the importance of a predictor, and various criteria have been proposed for measuring it (D. R. Thomas, Zhu, Zumbo, & Dutta, 2008); the DT technique quickly shows which is the attribute with the highest information gain or greatest entropy reduction in classifying individuals into different groups.

Practice implications

Currently, the use of the Mini-Mult scale is a tool being used by INPEC for a variety of purposes. It is used to orient the inmate population and to inform the sentencing judges about how the sentence is progressing. This means that from the point at which an individual is sentenced by a judge of the Republic, he must pass through different phases of penitentiary treatment, requiring assessment and classification (i.e. phases of observation, high security, medium security, and the confidence phase when freedom is obtained). Additionally, the Mini-Mult by itself is used to evaluate the inmate when he asks for an administrative benefit or permission for an unsupervised absence for 72 hours. INPEC is currently seeking new ways to use this instrument. In this context, we believe that the development of a classifying
model with greater performance than that reported in this experimental study may have two main practical applications in the Colombian prison system. First, the uncovered model can be easily transformed into risk indicators that make it possible to identify the personality of an individual associated with crimes against public safety in the prison population. Indeed, INPEC, the state entity responsible for the national prison population, routinely applies the Mini-Mult to obtain a personality profile of their prison population. A practical use for the model obtained is to train forensic investigators to identify the personality of individuals who commit these types of crimes.

Second, recognising the personality of these individuals is fundamental to the psychiatrists and psychologists who work with and evaluate degrees of freedom of individuals convicted of crimes against public safety (Instituto Nacional Penitenciario y Carcelario & Universidad Pontificia Bolivariana, 2009). Accordingly, the use of the DT can contribute to making a better clinical diagnosis and to the development of new sensitive treatments for the personality of this type of individual. It must be remembered that the DT is not transformed into an equation as in the LR but into a diagram with the potential to be used for didactic purposes. Public safety offenders are armed groups that act in hidden cells, which is why the direct study of the personality of these individuals is very difficult compared with the clinical investigation of the personality of non-criminal individuals.

A third practical implication is that the DT may help explain how certain configurations of personality are related to certain criminal activities. The DT shows several personality profiles that can be correlated with crimes against public safety. Therefore, the practical value is that personality profiles may help to explain why an individual engages in activities related with threats, conspiracy to commit a public safety offence, and terrorism. Thus, a deep exploration about how a specific personality has been constructed, its social context, life history, and significant episodes may contribute to understand how an individual has become a public safety offender.

Limitations and future research

This paper has three principal limitations. First, a machine learning approach was used to classify individuals by personality profile rather than a matched case-control approach. Because machine learning uses available cases (i.e. is a naturalistic experiment), there is no control for the effects of known potential confounding variables. A methodological design that combines the techniques of the two methods would be an interesting line of research for a future study.

Second, previous research has shown that DT can be unstable (Kitsantas, Hollander, & Li, 2007; Last, Maimon, & Minkov, 2002) in the sense that the addition of new examples to the training data may generate changes in the tree obtained. A future study should also include a stability analysis of the tree such as the one developed in Dwyer and Holte (2007) to determine how stable the tree is.

Third, the highest gain ratio used here by the C4.5 algorithm is only one of a variety of measures available for analysing variable importance (Rokach & Maimon, 2005). Other options include the information gain rate or the Gini index (Raileanu & Stoffel, 2004). Alternatively, different trees such as random forest (Breiman, 2001a), which incorporates a more robust measure of variable importance, could be employed.

Conclusion

As Weiner and Hess (2006) have pointed out, collaboration between psychology and law has grown prodigiously. This is reflected in the increase in the number of journals, textbooks, and
workshops on forensic psychology. With the growing public interest in crime issues, clinical psychologists have expanded their studies to include questions in this field. Based on results using machine learning, we posit that DT techniques can be useful tools for psychological criminal profiling. Further study is needed, however, and greater accuracy may be attainable through further research into the performance of other psychometric tests.

Based on the exploratory and comparative results of our study, we conclude that DT techniques have great potential for criminal personality profiling. DT techniques can help coordinate the work and language of teaching activities involving forensic psychologist instructors and practitioners. Our aim has been to link the development of better classification models to better forensic practices.

REFERENCES


Characterising the personnality of the public safety offender


Copyright © 2016 John Wiley & Sons, Ltd.  
DOI: 10.1002/jip


### APPENDIX A

#### Logistic Model

| Estimate | Standard error | z value | Pr (>|z|) |
|----------|----------------|---------|-----------|
| (Intercept) | −11.8174 | 2.6057 | −4.54 | 0.0000 |
| Hs | 0.0611 | 0.0324 | 1.89 | 0.0590 |
| D | −0.0012 | 0.0278 | −0.04 | 0.9656 |
| Hy | −0.0308 | 0.0456 | −0.68 | 0.4995 |
| Pd | 0.0572 | 0.0233 | 2.45 | 0.0141 |
| Pa | 0.0374 | 0.0276 | 1.36 | 0.1747 |
| Pt | −0.0103 | 0.0238 | −0.43 | 0.6658 |
| Sc | 0.0959 | 0.0309 | 3.11 | 0.0019 |
| Ma | −0.0206 | 0.0297 | −0.69 | 0.4887 |
### Appendix B

**Decision Tree Rules**

Table B1. Decision rules that discriminate between offenders against the public safety and non-offenders

<table>
<thead>
<tr>
<th>Rule ID</th>
<th>IF (rule antecedent)</th>
<th>THEN (rule consequent)</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1</td>
<td>Sc &gt; 70 ∧ Pd ≤ 58 ∧ Hs ≤ 70</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 2</td>
<td>Sc &gt; 70 ∧ Pd ≤ 58 ∧ Hs ≤ 70</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 3</td>
<td>Sc &gt; 70 ∧ Pd &gt; 58 ∧ Pt &gt; 56</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 4</td>
<td>Sc &gt; 70 ∧ Pd &gt; 58 ∧ Pt ≤ 56 ∧ Pa &gt; 67</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 5</td>
<td>Sc &gt; 70 ∧ Pd &gt; 58 ∧ Pt ≤ 56 ∧ Pa ≤ 67 ∧ Hs ≤ 54</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 6</td>
<td>Sc &gt; 70 ∧ Pd &gt; 58 ∧ Pt ≤ 56 ∧ Pa ≤ 67 ∧ Hs ≤ 54</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 7</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy ≤ 59 ∧ Pt ≥ 54</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 8</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy ≤ 59 ∧ Pt &gt; 54 ∧ Ma ≤ 53</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 9</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy ≤ 59 ∧ Pt &gt; 54 ∧ Ma &gt; 53</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 10</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy ≤ 59 ∧ Hy ≤ 70</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 11</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy &gt; 70 ∧ Hs ≤ 75</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 12</td>
<td>Sc ≤ 70 ∧ Pd &gt; 70 ∧ Hy &gt; 70 ∧ Hs &gt; 75</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 13</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc ≤ 53 ∧ Pa ≥ 41</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 14</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc ≤ 53 ∧ Pa ≥ 41 ∧ Ma ≤ 48</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 15</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc ≤ 53 ∧ Pa ≥ 41 ∧ Ma ≥ 48</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 16</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa &gt; 70</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 17</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa &gt; 70 ∧ Hs ≤ 52 ∧ D ≤ 68</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 18</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa &gt; 70 ∧ Hs ≤ 52 ∧ D &gt; 68</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 19</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa &gt; 70 ∧ Hs &gt; 52 ∧ Hy ≤ 55</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 20</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa &gt; 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 21</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 22</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy &gt; 65 ∧ Hy &gt; 70 ∧ Hy ≤ 73</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 23</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy &gt; 65 ∧ Hy &gt; 70 ∧ Hy &gt; 73</td>
<td>Criminal group</td>
</tr>
<tr>
<td># 24</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy ≤ 65 ∧ Hy &gt; 60 ∧ Pt ≤ 52</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 25</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy ≤ 65 ∧ Hy &gt; 60 ∧ Pt &gt; 52</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 26</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy ≤ 65 ∧ Hs ≤ 60 ∧ Pd ≤ 60 ∧ Sc ≤ 69</td>
<td>Non-criminal group</td>
</tr>
<tr>
<td># 27</td>
<td>Sc ≤ 70 ∧ Pd ≤ 70 ∧ Sc &gt; 53 ∧ Pa ≤ 70 ∧ Hs &gt; 52 ∧ Hy &gt; 55 ∧ Hy ≤ 65 ∧ Hs ≤ 60 ∧ Pd ≤ 60 ∧ Sc &gt; 69</td>
<td>Non-criminal group</td>
</tr>
</tbody>
</table>