Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 22 Number 10, November 2017 ISSN 1531-7714

Advocating the Broad Use of the Decision Tree Method in Education

Cristiano Mauro Assis Gomes, *Universidade Federal de Minas Gerais*
Leandro S. Almeida, *Universidade do Minho*

Predictive studies have been widely undertaken in the field of education to provide strategic information about the extensive set of processes related to teaching and learning, as well as about what variables predict certain educational outcomes, such as academic achievement or dropout. As in any other area, there is a set of standard techniques that is usually used in predictive studies in the field education. Even though the Decision Tree Method is a well-known and standard approach in Data Mining and Machine Learning, and is broadly used in data science since the 1980's, this method is not part of the mainstream techniques used in predictive studies in the field of education. In this paper, we support a broad use of the Decision Tree Method in education. Instead of presenting formal algorithms or mathematical axioms to present the Decision Tree Method, we strictly present the method in practical terms, focusing on the rationale of the method, on how to interpret its results, and also, on the reasons why it should be broadly applied. We first show the modus operandi of the Decision Tree Method through a didactic example; afterwards, we apply the method in a classification task, in order to analyze specific educational data.

Predictive studies have been widely used in the field education to provide strategic information about the extensive set of processes related to teaching and learning, as well as about what variables should predict certain educational outcomes (Osborne, 2000). There are many situations in which predictive studies are applied in the field of education. Just to mention a few, some studies aim to understand the role of school climate, teaching styles, curriculum, school management, study habits, students' personal characteristics, as well demographic and socioeconomic variables that impact learning or academic dropout (Knowles, 2015; Miller, Soh, Leen-Kiat, & Samal, 2015). Because of its broad scope, the results of predictive studies usually cast relevant evidence on the decision-making process in educational politics, while providing a clear perspective of the variables that are associated to specific phenomena (Osborne, 2000).

As in any other area, there is a set of standard techniques used in predictive studies in the field of education (Hsu, 2005). Even though the Decision Tree Method is a well-known and standard approach in Data Mining and Machine Learning, and is broadly used in data science since the 1980's, this method is not part of the mainstream techniques used in predictive studies in the field of education. In this paper, we support a broad use of the Decision Tree Method in education. Instead of presenting formal algorithms or mathematical axioms to present the Decision Tree Method, we strictly present the method in practical terms, focusing on the rationale of the method, on how to interpret its results, and also, on the reasons why it should be broadly used.

But why apply the Decision Tree Method in education? There are some advantages in applying the Decision Tree Method. Standard techniques, such as

linear regression and logistic regression, make certain important assumptions about the structure of the data, or about the model that is used to analyze the data and predict certain target variables. Logistic regression makes the assumption that the data should follow a logistic distribution. On the other hand, linear regression makes the assumption that data are normal, also requiring the homoscedasticity and normality of the model's residuals. Unlike these standard techniques, the Decision Tree Method does not demand any assumption about the data nor requires an a priori model to predict target variables. In the words of Nisbet, Elder, and Miner (2009), being a robust approach from Data Mining and Machine Learning, the Decision Tree Method "... doesn't start with a model; it builds a model with the data" (p. XXV).

Since the Decision Tree Method does not assume or require any structure for the data or an a priori model, this approach is suitable to deal with non-linear relationships between variables. We will describe the effectiveness of the Decision Tree Method to analyze non-linear relationships, illustrating this with an example of the method's rationale. Beside these advantages, the Decision Tree Method produces results that are very intuitive and easy to interpret. The resulting trees are clear and do not require any relevant statistical knowledge to be read, becoming appropriate to communicate evidence to a broad and diverse audience of people, such as educational managers, teachers, parents, students, and so on.

The Decision Tree Method Rationale: Explaining the Basic Concepts through an Example

Instead of introducing formal algorithms or mathematical axioms to present the Decision Method Tree, we will strictly present this method in practical terms, focusing on the method's rationale, on how to interpret its results, and also, on why it should be broadly used. Considering this goal, we present an example of a classification task.

Let us imagine that we have a set of educational data that contain the variable "enrollment", which informs about the number of enrolled and non-enrolled students in a specific university. Supposing that we are interested in understanding what explains students' non-enrollment, we aim to predict which students enroll and which students do not enroll in university, and for this reason we have a classification task. If our prediction is correct, we will produce an accurate classification of students according to the categories of the target variable.

Still according to this example, suppose that we have 10,000 students in this data set, with 5,000 students enrolled in university, and the other 5,000 students not. We have 50% of non-enrolled students and 50% of enrolled students. Our base line value for the prediction is 50%, since that, if we select students randomly from the 10,000 group, we have a predictive performance of 50%, just like when tossing a coin. Of course, when performing a prediction study, we expect that the model is capable of predicting more than the base line value.

What does the Decision Tree Method do exactly? This method recurrently and consecutively produces cut-offs in the data, aiming to achieve the best classification of the categories of the target (or dependent) variable. Imagine that you wish to cut a watermelon. You have a knife and you want to cut this fruit in the best place, that is, in the best location to separate the spoiled seeds from the healthy ones. Of course, when you cut the watermelon, you expect that one piece contains, if possible, only spoiled seeds, while the other piece contains only healthy seeds. This is the same "desire", the same essential rationale of the Decision Tree Method: to recurrently generate cuts in the data, in order to produce many pieces in the data, which improves the differentiation of the categories related to the target variable.

Now let us imagine that, instead of watermelons and seeds, we have a target variable that possesses two categories. For this target variable, the Decision Tree Method will search for the best independent variable in the data that is capable of identifying the best "cut in the data". In doing that, it will produce two pieces in the data. However, the method does not stop at the first cut. After that, the method will verify if it is possible to cut-off these two pieces in order to produce further new pieces, and so on, until it is not possible to produce more pure pieces, that is, pieces that best discriminate enrolled from non-enrolled students.

Let us suppose that our data contain five independent variables, and that the variable "I like to read" was the best one to produce the first "cut in the data", since this variable provides us with the best separation of the enrolled from the non-enrolled students. Figure 1 shows the product of that cut-off.

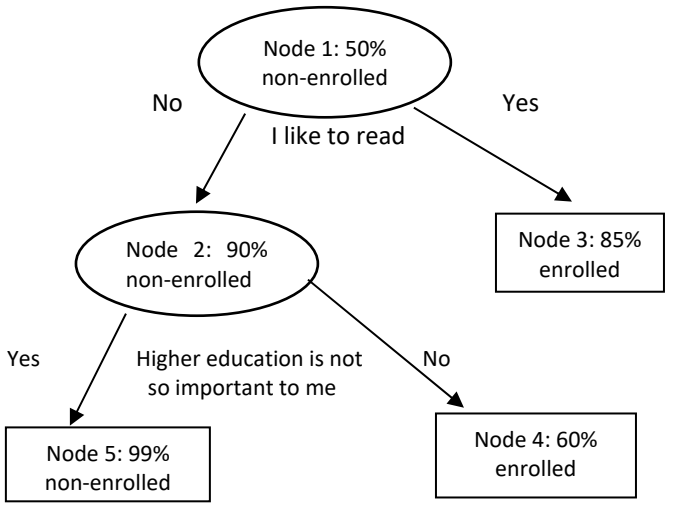

Figure 1. Example of a Tree from the Decision Tree Method.

From Figure 1, you should note that there is an oval object on the top. It is named the root node and it represents the 10,000 students of our example. Inside the root node there is the information that there is 50% of non-enrolled students. You should observe too in Figure 1 that, when the method selected the variable "I like to study" to generate the best cut-off in the data, it also produced one rule of separation, which is the following; rule 1: if the students like to read, they must be placed in one group, or else they must be placed in another group. This cut-off produced two new pieces of data, named node 2 and node 3. The term "node" is not used by chance; on the contrary, it is a standard name used in the Decision Tree Method given its allusion to trees, which possess branches, nodes, and leaves.

In Figure 1, the two new pieces of data (node 2 and node 3) have achieved a better classification than the original base line value of 50% to predict non-enrolled students. After the first cut-off, we now have two pieces of data that do a better job in separating the two categories of our target variable ("enrollment"). Node 2 possesses 90% of non-enrolled students, and node 3 contains 85% of enrolled students, which is a much better separation than the original condition when the data was divided.

As stated, the method will search for new cut-offs if it is capable of generating new pieces of data that better separate the categories of the target variable. Note that, in Figure 1, the method encountered a second cut-off, selecting the variable "Higher education is not so important to me", hence generating node 4 and node 5. The nodes that do not possess any descendant nodes are terminal nodes, and are named "leaves". Usually the leaves of the trees are represented by rectangles, while the other nodes are represented by ovals (cf. Figure 1).

The results of the tree that was produced must be interpreted by reading the information of the leaves. One of the advantages of the Decision Tree Method is that its results are easy to understand, while providing a clear scenario of what variables are related to the categories of the target variable. Supposing that in our example the final tree is represented by Figure 1, we must read the three rectangles that represent the leaves of the tree. Reading these leaves helps us to interpret the substantive results from our imaginary study.

Node 3, which is one of the three leaves, informs us that from the group of students that like to read, 85% of them are enrolled students. This information tells us that the variable "I like to read" is an interesting variable to predict if the student will enroll in university, since there is a probability of 85% that a student will enroll in university if he or she likes to read. Node 5, another leaf of the tree, informs us that if students do not like to read, and if they feel that higher education is not so important to them, there is a likelihood of 99% that they will not enroll in university. Finally, the third leaf tells us that if students do not like to read, yet do not think that higher education is not so important to them, there is a chance of 60% that they will enroll in university.

The information from the tree is very interesting; from it, we are able to interpret that, if students do not like to read and do not conceive higher education as being very important to them, there is a strong chance that they will not enroll in university. However, if they do not think that higher education is not so important, even if they do not like to read that much, the likelihood of non-enrollment drops from 99% to 40%, which is a powerful decrease of 54%. As you can see, the information contained in the leaves is straight, easy to understand, and enables a rich interpretation of the data.

Table 1. Sociodemographic Variables of Minho University Students: Frequencies and Missings

Variables	1	2	3	4	5	missings
Brothers in higher education	905 (Yes)	1537 (No)				35
Expectation of Course Conclusion	26	29	145	576	1615	86
Expectation of University Conclusion	27	31	149	480	1666	124
Age	264 (< 20 years-old)	2207 (>= 20 years-old)				6
Sex	1096 (Male)	1381 (Female)				0
GPA	105.0 (minimum)	152.0 (mean)	200.0 (maximum)			46
Vocational Orientation	1126 (Yes)	1325 (No)				26
Retention	2050 (No)	65 (Basic Education)	347 (Secondary)			15
Hours of study	0.0 (minimum)	6.0 (mean)	72.0 (maximum)			61
Course as the first option	1439 (Yes)	1012 (No)				26
University as the first option	1747 (Yes)	674 (No)				56
Student employed	2269 (No)	131 (Part Time)	70 (Full Time)			7
Changing of residence to Study in University	990 (Yes)	1470 (No)				17
Father education	1232 (Basic)	715 (Secondary)	333 (College)	171 (Beyond)		26
Mother education	1054 (Basic)	746 (Secondary)	448 (College)	215 (Beyond)		14
Socioeconomic status	596 (Low)	877 (Low-Middle)	814 (High-Middle)	189 (High)		1

Describing the Educational Data to be Analyzed

We applied the Decision Tree Method to an educational data set from the University of Minho, in Portugal. Our target variable is the enrollment of students in this university in 2015, and it is composed of two categories: enrolled students and non-enrolled students. The non-enrolled students are those that, even though having registered in the first phase of college national access and having been accepted at University of Minho, did not actually enroll, but rather they opted for a different vacancy in another degree course/higher education institution that better served their academic interests. From a total 2,477 students, only 131 are non-enrolled students (5.29% of the students), while 2,346 are enrolled students, which represents the large majority.

Apart from our target variable, the educational data set that we analyzed possesses a set of other interesting variables. We will use a group of variables from this bigger set, and these variables will be our independent variables. These variables will be grouped in two categories. The first one is formed by students' demographic information. Table 1 presents these variables, as well as their frequencies and missing values. The second category concerns students' expected difficulties in relation to university or the academic context. Table 2 presents these variables, as well as their frequencies and missing values.

Table 2. Expected Difficulties of the Minho University Students about the University or the Academic Context: Frequencies and Missings

Expected Difficulties on	No	Little	Middle	High	Very High	missings
Understanding academic contents	70	715	1463	191	31	7
Management of activities/time	109	659	1148	425	130	6
Support daily expenses	241	852	997	297	81	9
Interacting with colleagues	363	1114	777	174	43	6
Interacting with teachers	281	1074	874	192	47	9
Leaving home/family	772	637	620	286	154	8
Active participation in classes	243	867	982	303	76	6
Obtaining good achievement	83	737	1346	241	64	6
Obtaining family support	804	1176	412	57	21	7

The Technical Aspects of the Decision Tree Method Implementation

As stated, the Decision Tree Method essentially involves to cut the data as many times as possible, to achieve recurrently a better classification of the categories of the target variable. In our previous example, we used an analogy of a person who wishes to separate the healthy seeds of a watermelon from the spoiled ones, and for that reason she or he aims to cut the watermelon in the best place to separate the seeds. According to the literature, there are many different manners to perform the cut in the data using distinct algorithms (Rokach & Maimon, 2015). For our data, we choose the CART (Classification and Regression Trees) algorithm, which was originally developed by Breiman, Friedman, Olshen, and Stone (1984), who provided details about the mathematical features of the algorithm. To perform the CART algorithm, we used the rpart R package (Therneau & Atkinson, 2015). The default strategy in the CART algorithm to cut data was employed, which is the GINI index.

The data set to which we applied the Decision Tree Method presents a small frequency of missings, in comparison with the total data (see Table 1 and Table 2), and there are no missings in the dependent variable (non-enrolled students). We included all missing data of the independent variables in the analysis, and we treated these missings through the default strategy in the rpart R package, which is to use a surrogate split resembling the original split, in order to estimate the tree nodes. Further details of this technique are provided in Therneau and Atkinson (2015).

From the total 2,477 students in our data, only 131 belong to the group of non-enrolled students, while 2,346 are in the group of enrolled students, which shows a strong unbalanced sample of enrolled versus non-enrolled students. The Machine Learning and Data Mining classification literature argues that unbalanced samples tend to achieve bad accuracy. For this reason, the same literature strongly recommends that researchers treat the data before the classification task (Rokach & Maimon, 2015). Nisbet et al. (2009) claim that a ratio greater than about 10 to 1 generates troubles for many algorithms. Since our data shows a ratio of enrolled students to non-enrolled students around 25 to 1 (2,346/131), we employed the technique of weighting the cases of the dependent variable, taking the ratio between non-enrolled versus enrolled students as a reference. We employed the process of weighting the cases just in the train sample, as recommended in the literature (Flach, 2012; He & Ma, 2013). We explain ahead what is the train sample.

The Decision Tree Method does not assume any structure for data nor does it employ any a priori models to analyze data. This is an advantage, but it produces a problem that is very common for many methods in the fields of Machine Learning and Data Mining: the problem of overfitting. . It is defined as all the situations where the results of certain analyses fit better for the analyzed sample and fit worse for other samples. Usually, the literature in the fields of Machine Learning and Data Mining determines that the researchers divide the data in one or more samples, leaving at least one sample where the algorithm will be trained, and leaving at least one sample to test the generality of the model created in the trained sample.

Nisbet et al. (2009) recommend the use of "resample tools, such bootstrap, cross-validation, jackknife, or leave-one-out" (p. 736). We randomly

divide the data in a train sample (75% of the sample) and a test sample (25% of the sample), taking the proportion of cases in the dependent variable as a reference. For a better estimative, we employed a 10 N-fold cross-validation in the train sample, which is a resampling technique that divides data into mutually exclusive subsets. The advantage of this strategy is that the errors from each portion are averaged, and the algorithm is trained on n -1 folds and tested by the only fold that does not participate in that training. Afterwards, this fold that served as test comes back and participates in the n – 1-fold in training. Thus, a new fold from the training folds is chosen to serve a test, and the processes continue until all folds have been chosen as test. The 10-fold cross-validation enables many trainings and testings, reducing considerably the risk of some relevant overfit in the train sample.

There are other strategies to avoid overfitting. After running cross-validation, we employed the cost complexity pruning. A pruning decision tree informs us about the number of splits that should be pruned. In the pruning process, the nodes that only produce overfitting and disturb the generalization of the model are eliminated from the tree. We used the rpart R package to perform all these strategies to avoid overfitting (Therneau & Atkinson, 2015), except the strategy concerning data splitting in a train sample and in a test sample. For the later, we used the caret R package (Kuhn, 2017).

We built the Decision Tree through the train sample. However, as mentioned earlier, the train sample usually produces overfit, since the algorithm tends to "learn" excessively about the data, producing cuts that are proper only for the specific analyzed data, yet not proper for other samples. The Machine Learning and Data Mining literature recommends that researchers evaluate the quality of the generated model, verifying the prediction not in the train sample, but in the test sample. So, we evaluated the quality of the Decision Tree generated in the train sample, examining how this Decision Tree was capable of predicting the target variable in the test sample. As recommended in the literature, we performed this analysis employing the caret R Package (Kuhn, 2017), through a confusion matrix and the following indexes: (1) accuracy; (2) sensitivity or recall; (3) specificity.

Presenting the Results of the Implementation

Figure 2 shows the Decision Tree that was produced. We employed the cost complexity pruning, so this tree has been pruned. After examining the cost complexity of the tree that was originally generated, we used the value of 0.023190 to perform the pruning, thus generating the final tree. This value has been chosen because this cut-off value indicates a tree where the nodes do not increase the error of prediction in the samples generated by the 10-fold cross-validation. As stated, the goal of pruning the tree is to avoid overfitting.

Before interpreting the leaves of our Decision Tree, we need to inspect the quality of the prediction. Table 3 shows the confusion matrix and the indexes that inform us about the ability of the created model in the train sample to predict the target variable in the test sample. Since we split the data in two parts, a train sample and a test sample, when inspecting the quality of the prediction just in the test sample, it should be noted that the confusion matrix in Table 3 informs about the data of the test sample.

Table 3. The Confusion Matrix and the Indexes about the Model Quality in Classifying the Enrollment Variable

	True Enrolled	True Non-enrolled
Predicted as Enrolled	462	17
Predicted as Non-enrolled	124	15
Accuracy	0.772	
Recall	0.788	
Specificity	0.469	

By observing the "true non-enrolled" column in Table 3, we should identify 32 non-enrolled students in the test sample. From these 32 students, the model correctly predicts 15 students. You should see this information by examining the cell that crosses the true non-enrolled column with the line of students predicted as non-enrolled. Additionally, we can see that the model incorrectly predicted 17 non-enrolled students, since the model predicted them as enrolled students. This weak performance of predicting the true non-enrolled students is represented by the index of specificity, which is the number of true non-enrolled students that was correctly predicted divided by the

total number of the true non-enrolled students. The specificity of 0.469 indicates that only 46.9% of the non-enrolled students was correctly predicted by the model.

The model performs better in predicting the true enrolled students than the non-enrolled students. This can be observed when examining, in Table 3, that 462 true enrolled students were correctly predicted by the model, while 124 true enrolled students were incorrectly predicted as non-enrolled students. The index that represents this performance is recall (or sensitivity). The value of 0.788 in recall indicates that the model correctly predicted 78.8% of the true enrolled students.

The index of accuracy informs us about the model's ability to predict all cases, while specificity only focuses on the true non-enrolled students, and recall focuses on the true enrolled students. The value of 0.772 indicates that the model was capable of correctly predicting 77.2% of all students in the test sample. In sum, the model presents a better performance to predict all the students, as well as the enrolled students, while presenting a worse performance to predict the non-enrolled students.

While Table 3 was used to interpret the quality of the model, Figure 2 will be used to examine the substantial results from the Decision Tree. Seeing that the Decision Tree is created in the train sample, the values of this tree are related to this sample. So, if in Table 3 we observed data related to the test sample, in Figure 2 we shall analyze data of the train sample, which is a standard approach in the Decision Tree Method.

First, we will read the creation of the nodes, analyzing the different cuts before interpreting the leaves. Each node possesses relevant information (see Figure 2). The first information reports the name of the node (Enrolled or Non-enrolled). The name of each node is determined by the majority of students in the node. If the node possesses more non-enrolled students, it is named "Non-enrolled"; if it possesses more enrolled students, the node is named "Enrolled". Since the first node has the same percentage for both categories, the name of this node has been chosen randomly. Below the name of the node there is the percentage of the enrolled and non-enrolled students in the node. The reader's left side concerns the percentage of enrolled students, while the reader's right side relates to the percentage of non-enrolled students. In the first node, you can see .50 to your left side, and .50 to your right side, corresponding to the percentage of enrolled and non-enrolled students, respectively. Below this, there is information about the percentage that represents the relative frequency of the sample students in the respective node. For example, the first node possesses all students, so inside of this node you can read "100%".

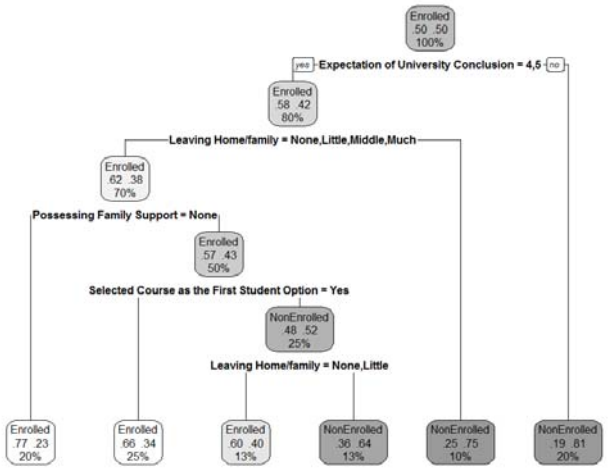

Figure 2. The Generated Tree from the Decision Tree Model: Classification of the Enrollment Variable

In Figure 2, there is the root node at the top of this Figure. Because we have weighted the train sample, since our data is unbalanced, we should see that this node possesses 50% of enrolled students and 50% of non-enrolled students. The first variable used to cut the data was the "Expectation of University Conclusion". Students who answered 1, 2 or 3 in the "Expectation of University Conclusion" scale were placed in a new node (Non-enrolled .19; .81; 20%), which corresponds to 20% of students in the train sample. As this node did not generate any other nodes, it should be seen as one leaf of the tree. On the other side, students who answered 4 or 5 in the "Expectation of University Conclusion" scale were placed in the other new node (Enrolled.58; .42; 80%). These students correspond to 80% of the train sample. This node was split in two other nodes through the cut by the variable "Leaving Home/Family".

Students who answered a scale about how hard it is "Leaving Home/Family", choosing the option "very high difficulty", were placed in a new node (Non-enrolled .25; .75; 10%), corresponding to 10% of the

train sample. This node is a leaf node, since it did not generate any other nodes. On the other side, students who answered the same scale stating that they found "no difficulty, little difficulty, middle difficulty or high difficulty" "Leaving Home/Family" were placed in the other new node (Enrolled .62; .38; 70%), corresponding to 70% of the train sample. This node was split in two new nodes through the cut by the variable "Possessing Family Support".

Students who responded to a scale saying that they did not have ("none" option of the scale) any difficulty about "Possessing Family Support" were placed in a new node (Enrolled .77; .23; 20%), corresponding to 20% of the train sample. This node is a leaf node, since it did not produce any new nodes. On the other side, students who chose the other options of the scale ("little difficulty, middle difficulty, high difficulty and very high difficulty") were placed in the other new node (Enrolled .57; .43; 50%). This node was split in two new nodes, by the cut of the variable "Selected Course as the Student's First Option". If the students answered "yes", they were placed on a new node (Enrolled .66; .34; 25%) that is a leaf node, while the students that answered "no" were placed in the other new node (Non-enrolled .48; .52; 25%). This node was split in two new nodes. Again, the variable "Leaving Home/Family" was employed to generate the cut-off. Students who answered the scale supposing to have "no" or "little difficulty" related to the variable "Leaving Home/Family" were placed in one new node (Enrolled .60; .40; 13%), while the students who answered the other options of the scale ("middle difficulty, high difficulty, and very high difficulty") were placed in the other new node (Non-enrolled .36; .64; 10%). Both nodes are leaves.

The Decision Tree possesses six leaves, which are disposed on the bottom of Figure 2. The first three leaves (on your left side) represent nodes where the majority is composed of enrolled students, while the other three leaves represent nodes where the majority is formed of non-enrolled students. We must read the leaves observing what rules have created this node. Observing the rules, we are capable of understanding what is node and what it tells us.

Reading the first leave in Figure 2 (on your left side), we may say that this leaf possesses 20% of the train sample students and that it has the majority of enrolled students (77%). This leaf informs us of the following: (1) if students answer options 4 or 5 (high expectations) regarding their "Expectation of University Conclusion", and (2) if they perceive "No, Little, Middle, or High" difficulty related to "Leaving Home/Family", and also, (3) if they see "No" difficulty related to "Possessing Family Support", there is a 77% likelihood that these students are enrolled. This is the message contained in this leaf.

Following, the second leaf in the Figure (on your left side) possesses 25% of the train sample students, and its majority is composed of enrolled students (66%). This leaf informs that: (1) if students answer options 4 or 5 (high expectations) concerning their "Expectation of University Conclusion", and (2) if they answer "No, Little, Middle, or High" difficulty related to "Leaving Home/Family", and (3) if they see "Little, Middle, High or Very High" difficulty related to "Possessing Family Support", and also, (4) if they answer "Yes" for the "Selected Course as the Student's First Option", there is a 66% chance that these students are enrolled.

The third leaf is composed of 13% of the train sample, and the majority is composed of enrolled students (60%). This leaf informs that: (1) if students answer options 4 or 5 (high expectations) concerning "Expectation of University Conclusion", and (2) if they perceive "No, Little, Middle, or High" difficulty related to "Leaving Home/Family", and (3) if they see "Little, Middle, High or Very High" difficulty related to "Possessing Family Support", and (4) if they answer "No" for the "Selected Course as the Student's First Option", and finally, (5) if they answer "None or Little" difficulty for "Leaving Home/Family", there is a 60% likelihood that these students are enrolled.

In sum, the results show that students who have high expectations (options 4 or 5 of the scale) in relation to the "Expectation of University Conclusion", and who do not anticipate difficulties related to "Leaving Home/Family", will more likely enroll in university. At the same time, there is a higher probability that students will enroll in university if they anticipate no difficulties related to "Possessing Family Support", as well if they have been admitted in their first choice-course and university.

The fourth leaf is very similar to the third one. It possesses the same rules and splits as the third leaf, except for the last split. Instead of answering "No or Little" in the last split, students choose one of the options "Middle, High, or Very High" for "Leaving

Home/Family". As an effect of this specific change, the probability of students enrolling in university dropped from 60% to 36%. This is evidence that the variable "Leaving Home/Family" is decisive, playing a relevant role as a predictive variable of enrollment, and as a protective factor against non-enrollment. This leaf contains 13% of the train sample students.

The fifth leaf is formed by two splits only. This leaf tells us that if students declare high expectations for the variable "Expectation of University Conclusion" (options 4 or 5), and at the same time expect very high difficulty in "Leaving Home/Family" (option "Very High"), there is a chance of only 25% that they will enroll in university. Again, it evidence in favor of the protective factor of the variable "Leaving Home/Family". This leaf contains 13% of the train sample students.

The sixth leaf contains 20% of the train sample students, and informs that if students choose the options 1, 2, or 3 (low or middle expectations) for "Expectation of University Conclusion", there is a likelihood of only 19% that they will enroll in university. Concerning the fourth, fifth and sixth leaves, we may state that negative expectations related to the variables "Expectation of University Conclusion" and "Leaving Home/Family" are decisive to improve the probability of non-enrollment considerably.

Conclusions

In this paper we have proposed that the Decision Tree Method should be broadly used in the field of education. Instead of showing the mathematical aspects or the specific algorithms of this approach, we focused exclusively on presenting the method's rationale, and on how to read and extract meaningful information from the results of the Decision Tree through an argumentative example, as well by applying the method in an educational data set.

Throughout the paper, we have focused on the Decision Tree's substantial aspects, namely: (1) the recurrent character of data splitting; (2) the generation of the tree through the splits and the rules created by the selected variables to perform the splits; (3) the strategies used to avoid overfitting, such as the pruning process; (4) how easy and straightforward it is to interpret and to extract meaning from the Decision Tree; (5) the potential of the results from the Decision Tree to produce information to ground educational interventions. For example, we have seen in our results that the expectation of difficulty in "Leaving Home/Family" is decisive on university enrollment. For this reason, it is possible that interventions aimed at reducing this kind of expectation should decrease the numbers of non-enrolled students.

As any other method, the Decision Tree Method is not perfect, and possesses limitations. Because it performs splits, generating distinct nodes, the method works much better in bigger samples. Other methods, such as random forest, bagging, and so on, generally tend to produce a higher accuracy in prediction, since they perform classes of trees instead of just one. Despite of that, if the Decision Tree Method tends to be worse in terms of accuracy in relation to other methods, this approach is superior in terms of interpretability and, because of that, in the production of meaningful results. Random forest, bagging, and other potent methods for prediction are all "black box" methods, so they fail to explain what and how the predictive variables are related to the target variable. This last aspect is maximized by the Decision Tree Method. We hope that this paper motivates researchers to use this method in their studies in the field of education.

References

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. New York: Chapman & Hall/CRC.

Flach, P. (2012). Machine learning: The art and science of algorithms that make sense of data. New York: Cambridge University Press.

He, H. & Ma, Y. [eds.] (2013). Imbalanced learning: Foundations, algorithms, and applications. Hoboken, New Jersey: John Wiley & Sons.

Hsu, T. (2005).Research methods and data analysis procedures used by educational researchers. *International Journal of Research & Method in Education*, 28(2), 109-133. DOI: 10.1080/01406720500256194.

Knowles, J. E. (2015). Of needles and haystacks: building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining*, 7, (3), 18-67.

Kuhn, M. (2017). caret: Classification and regression training. Retrieved from https://CRAN.R-project.org/package=caret

Miller, L. D., Soh, Leen-Kiat, & Samal, A. (2015). A comparison of educational statistics and data mining

approaches to identify characteristics that impact online learning. *Journal of Educational Data Mining*, 7 (3), 117-150.

Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis & data mining applications. London: Elsevier.

Osborne, J. W. (2000). Prediction in multiple regression. *Practical Assessment, Research& Evaluation*, 7(2). ISBN 1531-7714.

Rokach, L., &Maimon, O. (2015). Data mining with decision trees: theory and applications. Singapore: World Scientific Publishing.

Therneau, T. M., & Atkinson, E. J. (2015). An introduction to recursive partitioning using the rpart routines. Retrieved from https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Citation:

Gomes, Cristiano M. A., Almeida, Leandro S. (2017). Advocating the Broad Use of the Decision Tree Method in Education. *Practical Assessment, Research & Evaluation*, 22(10). Available online: http://pareonline.net/getvn.asp?v=22&n=10

Corresponding Author

Leandro S. Almeida
University of Minho
Portugal

email: leandro [at] ie.uminho.pt