

SELECTION OF ACADEMIC TUTORS IN HIGHER EDUCATION USING DECISION TREES

SELECCIÓN DE TUTORES ACADÉMICOS EN LA EDUCACIÓN SUPERIOR USANDO ÁRBOLES DE DECISIÓN

Argelia B. Urbina Nájera¹

Universidad Popular Autónoma del Estado de Puebla, México

Jorge de la Calleja

Universidad Politécnica de Puebla, México

ABSTRACT

In this paper, we present a method for the tutoring process in order to improve academic tutoring in higher education. The method includes identifying the main skills of tutors in an automated manner using decision trees, one of the most used algorithms in the machine learning community for solving several real-world problems with high accuracy. In our study, the decision tree algorithm was able to identify those skills and personal affinities between students and tutors. Experiments were carried out using a data set of 277 students and 19 tutors, which were selected by random sampling and voluntary participation, respectively. Preliminary results show that the most important attributes for tutors are communication, self-direction and digital skills. At the same time, we introduce a tutoring process where the tutor assignment is based on these attributes, assuming that it can help to strengthen the student's skills demanded by today's society. In the same way, the decision tree obtained can be used to create cluster of tutors and clusters of students based on their personal abilities and affinities using other machine learning algorithms. The application of the suggested tutoring process could set the tone to see the tutoring process individually without linking it to processes of academic performance or school dropout.

¹ *Correspondencia:* Argelia B. Urbina-Nájera. 17 sur No. 901 Barrio de Santiago. 72410. Puebla, México. Correo-e: abunajera@gmail.com

Key words: Academic Tutoring, Educational Data Mining, Decision trees, machine learning, tutors.

RESUMEN

En este documento se presenta un método para mejorar el proceso de tutoría académica en la educación superior. El método incluye a identificación de las habilidades principales de los tutores de forma automática utilizando el algoritmo árboles de decisión, uno de los algoritmos más utilizados en la comunidad de aprendizaje automático para resolver problemas del mundo real con gran precisión. En el estudio, el algoritmo arboles de decisión fue capaz de identificar las habilidades y afinidades entre estudiantes y tutores. Los experimentos se llevaron a cabo utilizando un conjunto de datos de 277 estudiantes y 19 tutores, mismos que fueron seleccionados por muestreo aleatorio simple y participación voluntaria en el caso de los tutores. Los resultados preliminares muestran que los atributos más importantes para los tutores son la comunicación, la autodirección y las habilidades digitales. Al mismo tiempo, se presenta un proceso de tutoría en el que la asignación del tutor se basa en estos atributos, asumiendo que puede ayudar a fortalecer las habilidades de los estudiantes que demanda la sociedad actual. De la misma forma, el árbol de decisión obtenido se puede utilizar para agrupar a tutores y estudiantes basados en sus habilidades y afinidades personales utilizando otros algoritmos de aprendizaje automático. La aplicación del proceso de tutoría sugerido podría dar la pauta para ver el proceso de tutoría de manera individual sin vincularla a procesos de desempeño académico o deserción escolar.

Palabras clave Tutoría académica, minería de datos educativa, árboles de decisión, aprendizaje automático, tutores.

Introduction

Higher education institutions have many responsibilities to educate and form students. These responsibilities not only include creating strategies that contribute to the development of professional skills, but also to form competent and integral professionals (Garcia M. and Pérez L. 2008). In particular, the tutoring process has gained significant importance as a complimentary strategy in order to support this purpose by tracking and monitoring students on their academic life (de la Cruz F., Chehaybar K. and Abreu 2011).

Tutoring actions are based on the establishment of a relationship between tutors and students. This relationship goes beyond scholar context; it is a link that opens a new space in which the student is known in other dimensions; it is a process of support and guidance on other aspects of their personal life. In addition, the relationship requires trust, communication, understanding, self-knowledge, empathy and proactivity, among others (Coriat B. and Sanz O. 2010; de la Cruz Flores, Chehaybar K. and Abreu 2011; Sinclair G. 2013).

Fernández J. (2004) notes that tutoring is effective if this is well suited to the needs of students and the characteristics of the educational environment; in addition it offers a personalized way, and also it is a part of a comprehensive educational psychology and guidance service. On the other hand, Gómez-Collado (2012), Castaño P., Blanco F. and Asensio C. (2012) state that it is necessary that all individuals must participate actively to achieve the same aim: to guide and to advise to the students in selecting classes each semester, to help them in mobility, job, social service, professional practices, among others. In order to achieve success in academic tutoring, Troyano and Garcia (2009), Cardozo-Ortiz (2011), Duran-Aponte and Duran-Garcia (2012)

mention that the tutor should be constantly trained to acquire skills to enable it to give timely follow-up to each of their students to encourage the completion of his university studies.

As mentioned, tutors must have skills to adequately perform the task of academic tutoring. These skills include those identified from several universities, particularly from the five main Mexican universities which are ranked in *The Economist Newspaper* in 2015. The National Autonomous University of Mexico (Universidad Nacional Autónoma de México [UNAM], 2015) considers that the skills are: research, teaching, communication, accuracy and personal. The Technological of Monterrey (Instituto Tecnológico y de Estudios Superiores de Monterrey [ITESM], 2014) considers as skills: cognitive, emotional and social. The National Polytechnic Institute (Instituto Politécnico Nacional [IPN], 2013) considers the skills of referential knowledge, know-how, and knowing how to be and stay. The Autonomous Metropolitan University of Mexico (Universidad Autónoma Metropolitana [UAM], 2016) proposed as a tutor profile: high institutional commitment, knowing all the academic regulations and possess skills such as communication, attention, affection, respect and ethics. Finally, the Autonomous University of Nuevo León (Universidad Autónoma de Nuevo León [UANL], 2011) suggests as skills: knowledge of institutional regulations, empathy, compassion, closeness and willingness to know the student and be guide in the process of school career.

Nowadays, one of the main challenges facing the tutoring process is to identify students' skills and affinities from large educational data repositories in order to create tutoring processes to improve academic tutoring in higher education. Recently, Educational Data Mining (EDM) has emerged as an alternative approach based on data mining, for exploring and extracting useful information from an educational context in order to improve the learning processes (Romero and Ventura 2007). Several studies have been carried out using different data mining approaches to perform tasks such as clustering, classification, prediction, and text categorization, among others; with the main goal of discovering the factors that affect the students' performance.

In this paper we introduce a method for the university academic tutoring process. We used the decision tree algorithm for identifying the main attributes of tutors that permit to classify teachers as tutors, and also to match them with similar students. Our preliminary results show several interesting attributes to be considered in order to take into account for academic tutoring, thus communication, self-direction and digital skill are the most important characteristics.

The remainder of the paper is organized as follows: in Section 2 we give some concepts of educational data mining and machine learning. In Section 3 is described related work. The methodology is introduced in Section 4, while experimental results are described in Section 5. Finally we present conclusions and directions for future work.

University Tutoring Process

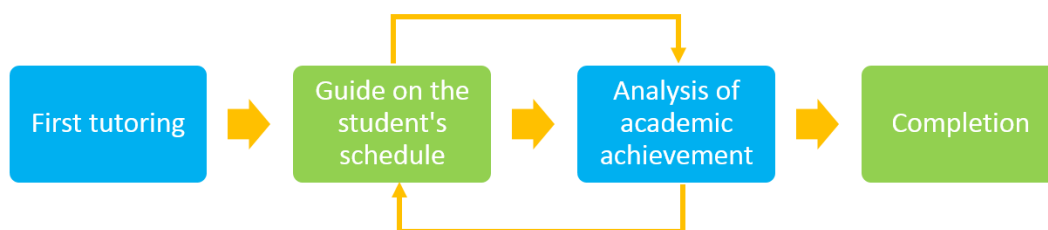
In the beginning, mentoring was considered as a tool to help in order to reduce problems such as academic failure, school dropout and the academic lag, through monitoring and support students throughout their academic trajectory. Recently, it is intended that through academic tutoring it can be met the challenges of modern society that requires students to incorporate working life skills to innovate in knowledge, make collaborative work and not be limited to benefit their school success (de la Cruz F., Chehaybar-Kury, and Abreu, 2011).

The tutorial activity is carried out by means of an action tutorial program (ATP) which serves

as a tool to design the content and implementation of university tutoring. In the ATP is included the action of tutors with students, in his action individually or in groups (Romo L., 2011). Moreover, Lepeña P., Sauleda P. and Martínez R. (2011) state that a consolidated program can be highly effective in developing specific or transversal skills, i.e, which can promote the development of cognitive abilities, methodological, technological, linguistic and oral, self-reflection, teamwork, social commitment, decision-making capacity, among others.

Figure 1 shows the sequence of activities that are part of the ATP, which include: 1) First tutoring: involves the mutual presentation (teacher-student), the tutor provides information on the structure, regulatory framework and services offered in the institution. He gives an overview of the academic career that tells the student alternatives or possible events throughout their education, also the student gives personal information as: health condition, personal data, and family data, among others. 2) Guide on the student's schedule. The tutor reviews your academic history to add subjects to study and offer alternatives for improvement to prevent academic failure. 3) Analysis of academic achievement. The tutor performs an analysis for applying strategies in order to improve or to refer students to other areas such as psychological support, academic counseling, scholarships, induction courses, specific courses, etc. 4) Completion. The tutor provides guidance on postgraduate programs, job search techniques, and search for research grants, among others (Lepeña P., Sauleda P. and Martínez R. 2011 and Romo L. 2011).

FIGURE 1. University tutoring process



Source: Self made

In Figure 1, the cycle is repeated while the student has studied materials. The cycle ends when the student graduates from college. As noted, there is no tutorial before the first phase where the form is known about how a tutor is assigned to a particular group of students. That is why, in this work we present a method in order to establish a process that includes this phase.

Educational data mining and machine learning

Learning may be the most distinctive characteristic of human intelligence that includes knowledge acquisition, skill development, knowledge organization and discovery of facts, among others (Mitchell 1997). Data mining and machine learning are two computer sciences areas that try to mimic these processes by studying and modeling them computationally (Bishop 2007).

Data mining can be defined as the process of extracting knowledge hidden from huge volumes of raw data. Technically, data mining is the process of finding correlations or patterns among thousands of fields in large databases. These patterns must be new, not obvious, and one must be able to use it. In classical databases manager systems, database records are returned according to a query; while in data mining, what is retrieved is not explicit in the database, i.e. implicit patterns. Data mining finds these patterns and relationships using data analysis tools and techniques to build models, hence machine learning (Witten, Frank and Hall 2011).

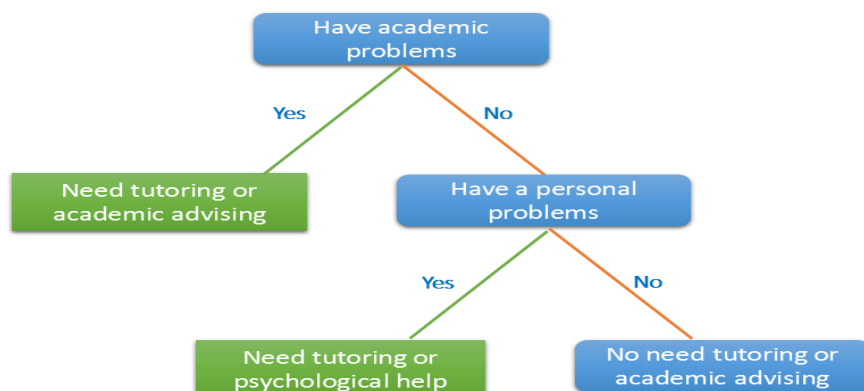
Educational Data Mining is an emerging discipline, concerned with creating methods for exploring the unique types of data that come from educational settings, and using those methods to solve and improve learning processes in an automated manner. Educational data mining methods are drawn from a variety of areas, including data mining and machine learning, psychometrics and other areas of statistics, information visualization, and computational modeling (Romero and Ventura 2007).

Nowadays, there are many algorithms that have been applied in several real-world problems with high accuracy, these algorithms include artificial neural networks, Bayesian learning, instance-based methods, kernel methods, and decision trees, among others. In this paper, we only give a brief description of decision trees and the metrics used for their evaluation.

Decision trees

Decision trees are one of the most widely used machine learning methods, and they have been successfully applied to a broad variety of practical problems. A decision tree describes graphically the decisions to be made, the attributes selected, and the outcomes associated with combinations of decisions and attributes. Basically, decision trees are formed by nodes and branches. Nodes are the points where a choice must be made, while branches represent one of the possible alternatives or courses of decision. A root node is selected where begins the decision process (Witten, Frank and Hall 2011). In Figure 2 we show a simple example of a decision tree to identify students for tutoring.

FIGURE 2. Decision tree to identify the need for tutoring, academic advising or psychological help for students



Source: Self made

Figure 2 describes a decision that the tutor should take if the student has academic or personal problems. This is identified by a student interview, or watching the student's academic performance. In both cases, the tutor identifies the type of problem of the student, if this one is an academic problem then the tutor schedules an academic counseling. Otherwise, this is a personal problem, so the tutor should send him to psychological care. And if the tutor identifies a problem outside of these two criteria, then it may be of an administrative nature so the student should assist to other agencies and the tutoring process ends.

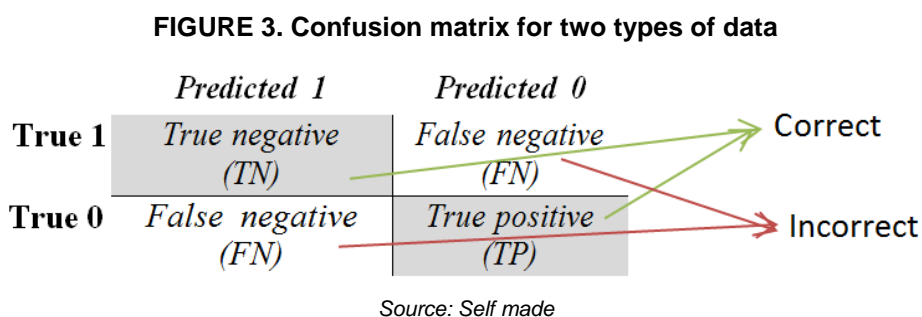
Building decision trees

The algorithm for building decision trees is based on the simple divide-and-conquer algorithm. Witten, Frank and Hall, (2011) describe this algorithm, which consist of selecting the best attribute to place it as the root node and make one branch for each possible value. This splits up the example set into subsets, one for every value of the attribute. Now the process can be repeated recursively for each branch, using only those instances that actually reach the branch.

Ideally, the process terminates when all leaf nodes are pure, that is, when they contain instances that all have the same classification. However, it might not be possible to reach this happy situation because there is nothing to stop the training set containing two identical sets of attributes but different classes. Consequently, it stops when the data cannot be split any further. Alternatively, one could stop if the information gain (which can be interpreted as the informational value of creating a branch on the attribute) is zero. This is slightly more conservative because it is possible to encounter cases where the data can be split into subsets exhibiting identical class distributions, which would make the information gain zero.

Performance of decision trees

In order to know the performance of machine learning algorithms in general, several metrics have been defined such as accuracy, precision, recall and f-measure. A confusion matrix (see Figure 3) summarizes the classification performance of an algorithm with respect to some test data and permits to calculate the metrics mentioned previously.



From Figure 3 the following metrics can be defined:

- The *accuracy* is the proportion of the total number of predictions that were correct, i.e. True negative + True positive / Total of instances.
- The *precision* is the proportion of the predicted positive cases that were correct, i.e. True positives / False positive + True positive.
- The *recall* is the proportion of positive cases that were correctly identified, i.e. True positive / False negative + True positive.
- Finally, *f-measure* combines precision and recall to give a single score, i.e. $2 \times (\text{precision} \times \text{recall} / (\text{precision} + \text{recall}))$.

Given a classifier and an instance, there are four possible outcomes. If the instance is positive and it is classified as positive, it is counted as a true positive; if it is classified as negative, it is counted as a false negative. If the instance is negative and it is classified as negative, it is counted as a true negative; if it is classified as positive, it is counted as a false positive. Given a classifier and a set of instances (the test set), a two-by-two confusion matrix can be constructed representing the dispositions of the set of instances. This matrix forms the basis for many common metrics.

Related work

Relevant literature about educational data mining can be found in Romero and Ventura (2010), Shu-Hsien, Pei-Hui and Pei-Yuan (2012) and Peña-Ayala (2014). However, from these surveys we can highlight that educational data mining has been applied for analysis and visualization of data, providing feedback for supporting instruction, recommendations for students, predicting students' performance, student modeling, detecting undesirable student behaviors, grouping students, social network analysis, developing concept maps, constructing courseware, planning and scheduling.

Data mining and machine learning methods that have been widely used include Bayesian learning, decision trees, instance-based learning, and hidden Markov models (Peña-Ayala 2014). Thus, we only present some of these works, for example, Vialardi, Bravo, Shafti and Ortigosa (2009) proposed to use a subject recommender system based on data mining to help students to make better decisions when they have to take courses. They analyzed data of the last seven years of students in engineering at the University of Lima, Peru. Their experiments were performed using the algorithm called C4.5 that is a decision tree-based method. Experimental results show a global accuracy of 77.3%, modifying the conditions for training and testing. The authors also comment that their approach permits to obtain relevant information of students that would not be available when using descriptive statistical techniques.

In 2011, Anupama K. and Vijayalakshmin used decision trees, particularly ID3 and C4.5, to predict the performance of students in their final exam. This prediction aided to their tutors to identify weaks in order to improve student's performance. Experiments were performed using a data set of 116 and 117 examples, obtaining an average accuracy of 88% and 91%, respectively. They concluded that data mining brings several advantages in higher learning institutions, and these techniques can be applied in other education areas to optimize resources, to predict the retention of faculties, to predict the feedback of tutors, among others.

Kumar B. and Pal (2011) used decision trees to extract a set of academic features to assess the performance of students. The data set consisted of 50 examples from students of the computer applications department at the VBS Purvanchal University, India. Some features considered were grades of previous semesters, seminar performance, general proficiency performance, and attendance, among others. The knowledge extracted and represented by the decision tree permitted to obtain if-then rules to classify the students. With this work, authors predicted the performance of students in the end of semester, as well as they identified students who need special attention to reduce the failure rate.

Yukselturk, Ozekes and Türel (2014) experimented with four algorithms: k-nearest neighbors, decision trees, the naive Bayes classifier and artificial neural networks to classify students who dropped out of school. The data set was collected by applying an online test for 189 students enrolled in 2007 to 2009. The machine learning algorithms were trained and tested using the 10-fold cross-validation technique. The best results were obtained using 3-nearest neighbors and

decision trees, with an accuracy of 87% and 79.7%, respectively. These results were useful since they permitted to predict dropout of students on the online program dataset. Finally, the authors concluded that data mining methods might help to predict different reasons when students decide to drop out before to finish their study programs.

Another problem that has been tackled using machine learning algorithms is the introduced by Kakavand, Mokfi and Tarokh (2014), with the purpose of predicting loyalty of students using decision trees. The authors investigated the external factors that may cause loyalty, in order to identify those students who have decided to continue studying, thus University can invest on them, increasing its educational quality. The experiments were performed using a data set of 135 instances for training, 33 for testing and 35 for validation, with 14 attributes per instance (gender, age, amount of income, among others). The best result was obtained using the CART decision tree algorithm with 94% accuracy.

By doing this search, it was not found that the decision tree algorithm has not been used to identify the characteristics of tutors or students; mainly focused on skills and personal affinities. Therefore, we identified the opportunity to experiment with this algorithm to find the principal skill of the tutors.

Methodology

In this section, the methodology for the proposed method is described. The first step was to create a data set using an instrument to collect information. This instrument considers six skills: communication, self-direction, interpersonal, intrapersonal, autonomy, digital, and personal affinities. Next, we use the Weka software (Hall et al., 2009) to build a decision tree with the data set, in order to identify the most important skills for tutors and students. In next subsections, we describe these stages.

Data collection

The dataset was created from engineering students and teachers using an H-A instrument (Urbina-Nájera, de la Calleja, Vega L., Lopez M. and Pico G. 2014). However, we perform a selection process in order to have a representative sample. For selecting students, random sampling was performed considering 1199 individuals. Thus, with the purpose of obtaining a sample with a confidence level of 90% and an acceptable error of 5%, we only select 277 students: 143 men, and 134 women; all of them between 18 and 28 years old. Regarding the tutors, they were selected under the criterion of voluntary participation, thus, from a population of 35 teachers, only 19 participate in this study: 6 women and 13 men; all of them between 31 and 45 years old.

The H-A instrument for collecting data

The H-A instrument includes evaluating skills demanded by modern education as digital, autonomy, self-direction and personal affinities, according to (Urbina-Nájera, de la Calleja, Vega L., López M., and Pico G., 2014). Also, the instrument consists of 38 items divided into three groups: 18 items to identify skills, 16 items to identify affinities and 4 items to identify

demographic characteristics. The instrument offers a wide range of skills and personal affinities that should have both, students and tutors, like a significant factor for living-group. It is important to mention that this instrument was validated using the *Cronbach alpha* metric, obtaining a value of 0.8924, that indicates its reliability. Finally, the scale used for the H-A instrument was as follows: (5) strongly agree to (1) strongly disagree.

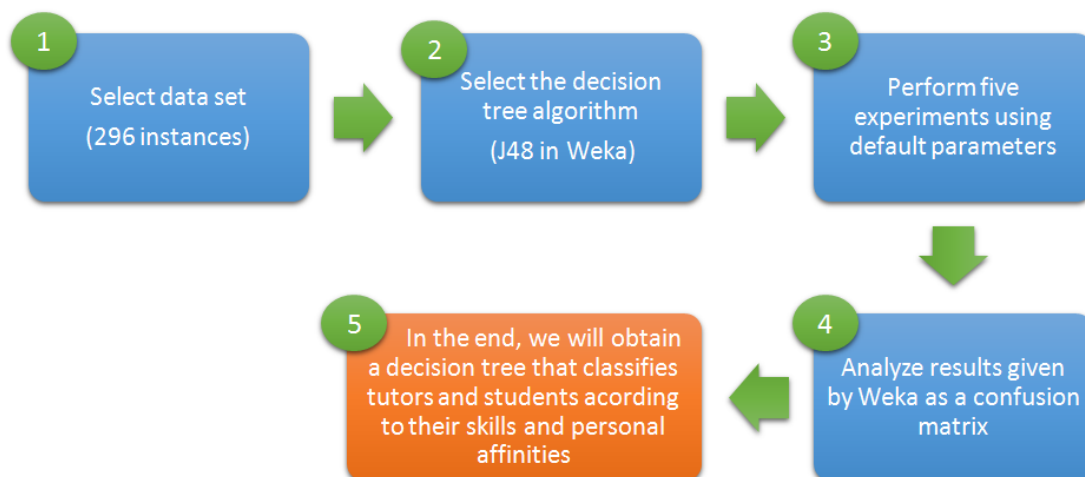
Creating decision trees using Weka

Although there are several software tools for data mining experimentation such as Knime, Apache Mahout, CLUTO, DMelt, Rattle, among others; Witten, Frank and Hall (2011) proposed the used of Weka (Hall, y otros, 2009) due to it was designed to rapidly experiment with the most common data mining methods on datasets. It provides extensive support for the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically and visualizing the input data and the result of learning. This diverse and comprehensive toolkit is accessed through a common interface so that its users can compare different methods and identify those that are most appropriate for the problem at hand.

Decision trees enable to deploy a problem visually, facilitates the interpretation of decisions, provides a high degree of understanding of the knowledge used in the decision-making, as well as, has a low computational cost (Coelho B., Porto B., de Carvalho and Freitas 2012; Chen 2011; Mitchell 2000).

Figure 4 describes the experimentation process performed in Weka: 1) Selecting the decision tree algorithm in Weka, named J48, 2) Applying the algorithm to the dataset of tutors and students (296 instances), 3) Performing five runs in order to obtain the average of the results, 4) Analyzing the confusion matrix of each experiment to determine the classification done by the algorithm, and finally 5) Obtaining the tree with the prevailing characteristics for the tutor and tutoring.

FIGURE 4 Experimentation process using Weka



Source: Self made

In table 1 we described the default parameter used for experimentation with decision trees (J48).

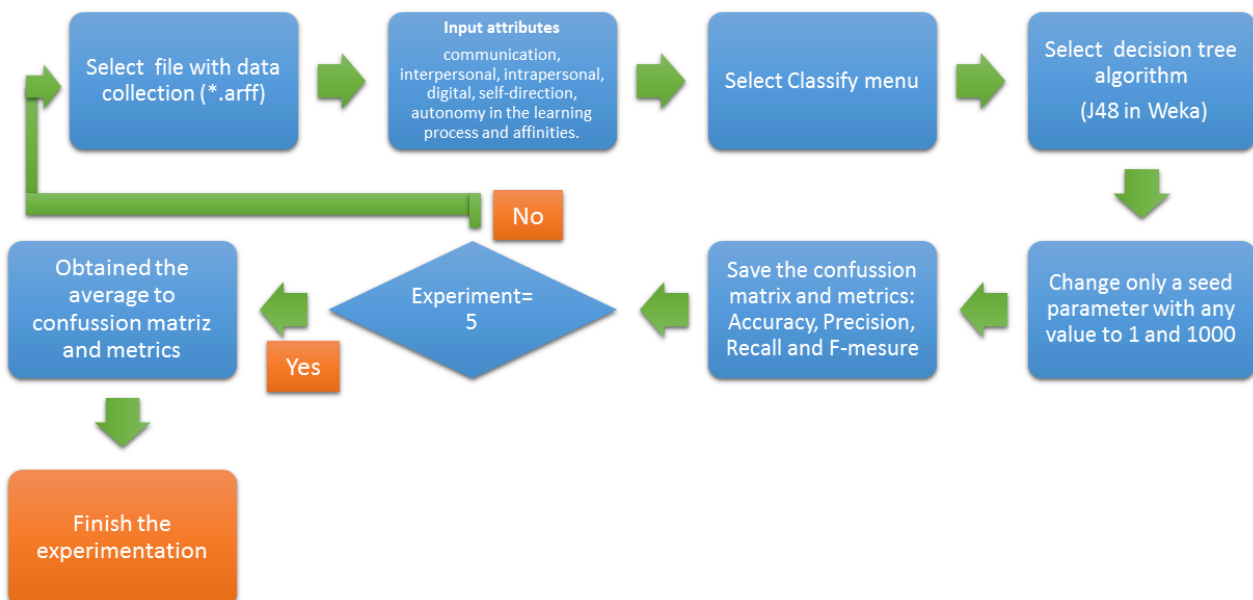
TABLE 1. Default parameters using the J48 algorithm in Weka

Name	Description	Value
binarySplits	Whether to use binary splits on nominal attributes when building the trees.	False
confidenceFactor	The confidence factor used for pruning (smaller values incur more pruning).	0.25
Debug	If set to true, classifier may output additional info to the console.	False
minNumObj	he minimum number of instances per leaf.	2
numFolds	Determines the amount of data used for reduced-error pruning. One fold is used for pruning, the rest for growing the tree.	3
reducedErrorPruning	Whether reduced-error pruning is used instead of C.4.5 pruning	False
saveInstanceData	Whether to save the training data for visualization.	False
Seed	The seed used for randomizing the data when reduced-error pruning is used.	1
subtreeRaising	Whether to consider the subtree raising operation when pruning.	True
Unpruned	Whether pruning is performed	False
useLaplace	Whether counts at leaves are smoothed based on Laplace.	False

Source: Self made

In Figure 5, we showed the process to perform the experiments in Weka, thus we can see that first, it is necessary to select the data set in .arff format file, then select the J48 algorithm and to examine the confusion matrix obtained.

FIGURE 5. Process to experiment with decision tree on Weka



Source. Self made

Experimental results

In this section, we present the experimental results in two sections. First section describes the process to identify skills using the decision tree algorithm. In the second section we introduce the academic tutoring process from the attributes identified by the decision tree.

Identification of skills

We use the decision tree algorithm that is implemented in Weka, using default parameters described in Table 1. The input attributes for the algorithm correspond to the seven characteristics that were evaluated with the H-A instrument: skills (communication, interpersonal, intrapersonal, digital, self-direction, autonomy in the learning process) and affinities. It is important to note that the demographic variables were omitted because they did not provide relevant and sufficient information for classification.

Weka reports the classification results through a confusion matrix and different measure metrics; thus, in Table 2 we show the results for classifying tutors and students. These results were obtained by averaging the results of five runs of the decision tree algorithm using 10-fold cross-validation. This validation technique is commonly used in the machine learning community for performing experiments, that is, the original data set is randomly divided into ten equally sized subsets and 10 experiments are performed, using in each experiment one of the subsets for testing and the other nine for training. By doing this, the entire data set is used for testing and training. As we can observe from Table 2, the decision tree algorithm was able to obtain good results above 92% accuracy, .90 for precision, .92 for recall and .91 for f-measure. These results indicate that the algorithm correctly classified 274 samples, while 22 were incorrectly classified. Therefore, if we want to classify new instances, we expect that the decision tree will perform with an accuracy of 92%.

TABLE 2. Results for the decision tree algorithm for classifying tutors and students

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Mesure</i>
<i>Average</i>	92.0270	0.9072	0.9204	0.9134

Source: Self made

Also, we show in Table 3, the confusion matrix for the best decision tree obtained, which is shown in Figure 6. In order to explain the values of the confusion matrix we describe the rows and columns as follows:

- Top left, 6 tutors were correctly classified.
- Bottom left, 9 tutors were incorrectly classified
- Top right, 13 students were incorrectly classified
- Bottom right, 268 students were correctly classified
- The values of the main diagonal correspond to the correctly classified instances (tutors and students), i.e. 274 instances.

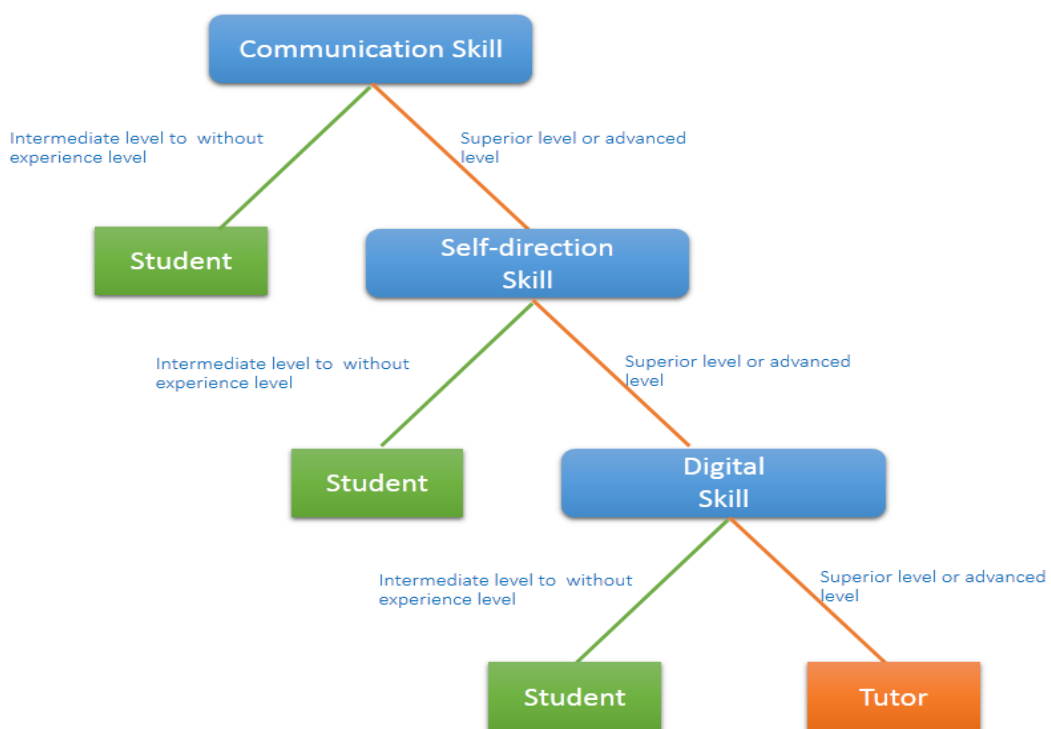
TABLE 3. Best confusion matrix for classifying tutors and students

	TUTOR	STUDENT	Total
TUTOR	6	13	19
STUDENT	9	268	277
Total	15	281	296

Source: Self made

Figure 6 shows the decision tree obtained. We can observe that there is only one way to classify a teacher as a tutor, that is, the algorithm classifies a tutor as an individual who has developed skills such as communication, self-direction and digital skills. Also, we can see that the most important attribute was the communication skill, since the algorithm placed it as the root of the tree. Otherwise, the decision tree classifies the individuals as students.

FIGURE 6. Decision tree for classifying tutors and students

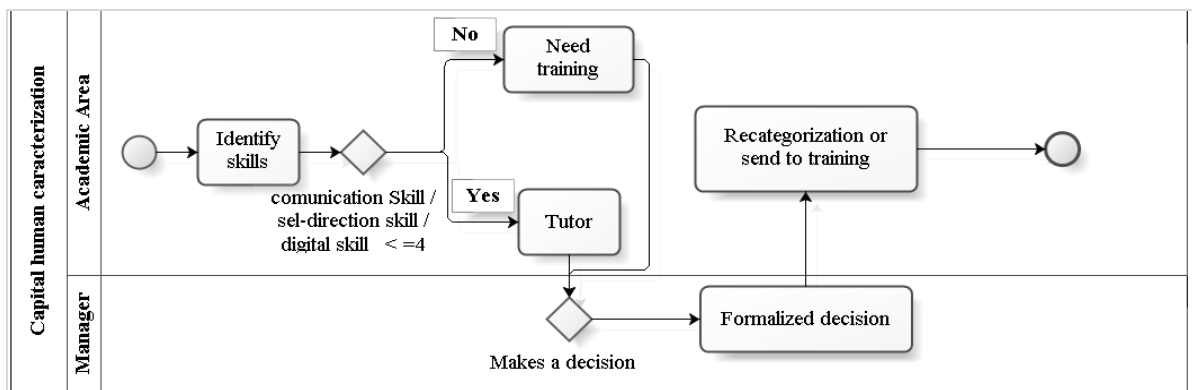


Source: Self made

The proposed academic tutoring process

From the attributes identified by the decision tree (Figure 6), we now propose a process (see Figure 7) for identifying tutors. This process consists of the following stages: 1) Identification of skills through the H-A instrument; 2) Verification if communication, self-direction and digital skills have a value greater than 4; if this is true, the teacher can be a tutor; 3) If such skills have a value less than 4, i.e. regular or bad, then teachers need training in order to develop them; 4) Back to the characterization, the academic area sends the information obtained to managers; 5) The directors or heads of each area are responsible for making decisions. 6) Formalizes the decision and sent to the response to the academic area. 7) The latter is responsible for notifying the reclassification or sent to training. 8) Finish the process.

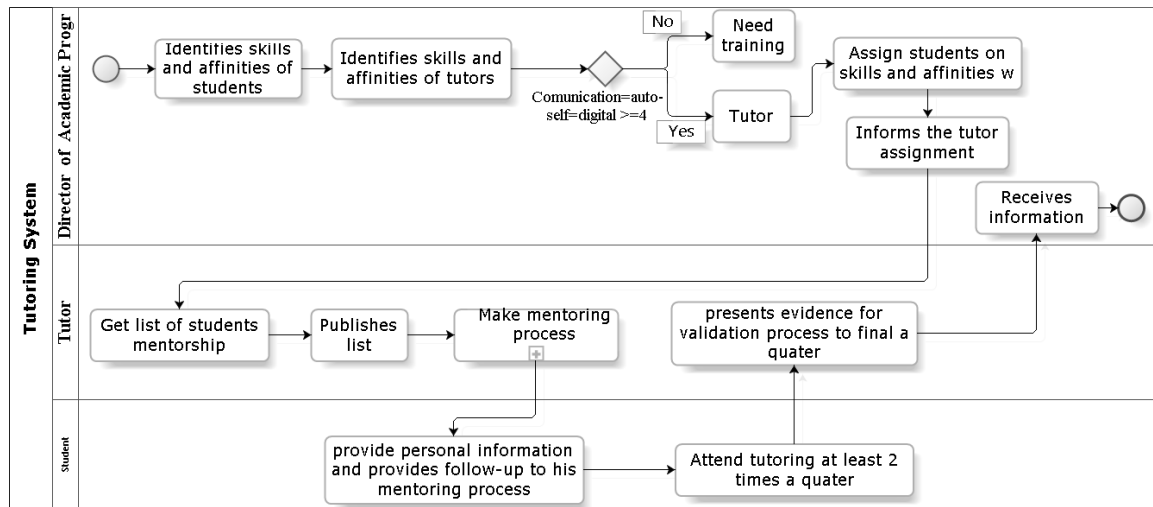
FIGURE 7. Process for identifying tutors



Source: Self made

Once the tutors have been identified, we propose a process for academic tutoring, which is presented in Figure 8. In this process we propose to match the best tutor, according skills, for each student group to promote the development of skills in students, in order to contribute for their professional skills and to form competent and integral professionals.

FIGURE 8. Proposed process for academic tutoring



Source: Self made

As shown in Figure 8, the process proposed includes selecting the tutor depending on skills and personal affinities to the group. The process is repeated every semester until students graduate or dropout of the institution.

Conclusions

Data mining and machine learning has recently been applied in the academic context, allowing educational institutions allocate better human and material resources, managing performance of students and improving the effectiveness of performance throughout their education.

Particularly, academic tutoring has become more important in the formation of students. Tutoring promotes academic supervision and permits to reduce dropout significantly. Therefore, in this paper we have introduced a method to identify skills for tutors and students using decision trees, and also, we have presented an academic process in order to attend the tutoring process in higher education.

Our experimental results show that the decision tree algorithm was able to identify those attributes that permit to identify relevant skills for tutors. Then, a tutor is characterized by having a superior level in communication, self-direction and digital skills. However, in the experimentation carried out with 19 tutors, only 44% of them match with these characteristics. This tells us that these abilities were not developed in their training or work experience; or that higher education institutions do not regard as part of their training. Therefore, there is a need to make actions to strengthen the figure of the tutor, for example universities may promote tutoring as part of teacher training, courses that allow to develop different skills, or any type of coaching; in addition, incorporating tools and information technologies.

The proposed academic tutoring process, does not guarantee dropouts, but ensures adequate and timely academic tracing to prevent it. To test the effectiveness of the process in all areas of

tutoring, we would need to test it, improve it and implement it for a period of time to determine whether this process contributes to remedy university academic problems.

The application of the suggested tutoring process could set the tone to see the tutoring process individually without linking it to processes of academic performance or school dropout.

Future work may include creating automated methods for identifying skills of students, improving making decision, assignment of tutors using some criteria, recommendation of lessons and materials, improving the admission process, predicting student performance, among others. Also, many other machine learning approaches can be applied such as dealing with imbalanced data sets, selecting the best attributes, using ensembles of algorithms and applying data reduction techniques.

References

- Anupama K., S, Vijayalakshmi, M.N. (2011). Efficiency of decision trees in predicting student's academic performance. *Computer Science and Information Technology*, 2, 335-343.
- Bishop, C. M. (2007). *Pattern recognition and Machine Learning*. Singapore: Springer.
- Cardozo-Ortiz, C. E. (2011). Tutoría entre pares como una estrategia pedagógica universitaria. *Educación y Educadores*, 14(2), 309-325.
- Castaño P., E., Blanco F., A., and Asensio C., E. (2012). Competencias para la tutoría: experiencia de formación con profesores universitarios. *Revista de Docencia Universitaria*, 10(2), 193-210.
- Chen, M.-Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38(9), 11261-11272.
- Coelho Barros, R., Porto Basgalupp, M., C.P.L.F. De Carvalho, A. & Freitas, A.A (2012). A Survey of Evolutionary Algorithms for Decision-Tree Induction. *IEEE Transactions on Systems, Man and Cybernetics*, 42(3), 291-312.
- Coriat B., M., y Sanz O., R. (2010). *Orientación y tutoría universitaria*. Recuperado de Universidad de Salamanca: http://campus.usal.es/~ofeees/nuevas_metodologias/tutorias/orienta001.pdf
- Curto D., J., and Conera I Carat, J. (2010). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.
- Durán-Aponte, E. y Durán-García, M. (2012). Competencias sociales y las prácticas profesionales. Vivencias y demandas para la formación universitaria actual. *Cultura y Educación*, 24(1), 61-76.
- Fernández J. F. (2004). La tutoría académica personalizada y su importancia en la eficiencia escolar. *REMO*, 1(2), 31-35.
- Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. The Netherlands: Springer-Verlag.
- Gómez-Collado, M. E. (2012). La percepción de los estudiantes sobre el Programa de Tutoría Académica. *Convergencia: Revista de Ciencias Sociales*, (58), 209-233.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1), 10-18.
- Hitt, M. A., Black, J. S., and Porter, L. W. (2014). *Management*. Edinburg Gate, Harlow: Pearson Education Limited.

- Kakavand, S., Mokfi, T., and Tarokh, M. J. (2014). Prediction the Loyal Student Using Decision Tree Algorithms. *International Journal of Information and Communication Technology Research*, 4(1), 32-37.
- Kumar B., B., and Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications, IJACSA*, 2 (6), 63-60.
- Lepeña P., C., Sauleda P., N., and Martínez R., Á. (2011). Los programas institucionales de acción tutorial: una experiencia desarrollada en la Universidad de Alicante. *Revista de Investigación Educativa*, 29(2), 341-361
- Mitchell, T. M. (1997). *Machine Learning*. Singapore: McGraw-Hill.
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert System with Applications*. 41, 1432-1462.
- Romero, C., Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert system with Applications*, 33,135-146.
- Romero, C., Ventura, S. (2010).Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 40(6) 601-618.
- Romo L., A. (2011). *La tutoría: una estrategia innovadora en el marco de los programas de atención a estudiantes*. México, D.F.: ANUIES.
- Shu-Hsien, L., Pei-Hui, C., Pei-Yuan, H. (2012). Data mining techniques and applications- A decade from 2000 to 2011. *Expert System with Applications*. 39, 11303-11311.
- Sinclair G. (2013). *Mentoring and Tutoring by Students*. London, UK.: Sinclair Goodlad.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. EUA.: Addison-Wesley.
- Urbina-Nájera, A. B., De La Calleja, J., Vega L., C. A., López M., N., and Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores en la modalidad virtual. *V Congreso Internacional sobre Calidad y Accesibilidad de la Formación Virtual* (pp. 227-234). Antigua Guatemala: ESVI-AL.
- Vialardi, C., Bravo, J., Shafti, L., and Ortigosa, Á. (2009). Recommendation in Higher Education Using Data Mining Techniques. *Second International Conference on Educational Data Mining* (págs. 190-199). Córdoba, España: International Working Group on Educational Data Mining.
- Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. USA: MK:Morgan Kaufmann.
- Yukselturk, E., Ozekes, S., and Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and eLearning*, 17 (1), 118-133.

Electronic References

- De La Cruz Flores, G., Chehaybary Kury, E., y Abreu, L.F. (2011). Tutoría en educación superior: una revisión analítica de la literatura. [Electronic version]. *Revista de la Educación Superior*, Vol XL(1), 189-209. Recovered on 12 de noviembre de 2015, from <http://www.scielo.org.mx/pdf/resu/v40n157/v40n157a9.pdf>
- El Economista. (2015). *Las mejores Universidades de México. Ranking 2015*. Retrieved from Las

10

mejores

universidades:

<http://eleconomista.com.mx/especiales/americaeconomia/2015/07/20/las-mejores-universidades-mexico-ranking-2015>

García M. V., y Pérez L. M.C. (2008). Análisis de las competencias lingüísticas y digitales en el marco de los estudios universitarios de turismo en España. [Electronic version]. *Revista electrónica "Actualidades Investigativas en Educación"*. Vol 9(1), 1-27. Recovered on 12 de noviembre de 2015, from <http://revista.inie.ucr.ac.cr/index.php/aie/article/view/339/338>

IPN. (2013). *Instituto Politécnico Nacional*. Retrieved from Programa Institucional de Tutorías: <http://www.tutorias.ipn.mx/pi-tutorias/Documents/recursos/PROGRAMA%20INSTITUCIONAL%20DE%20TUTOR%20C3%8DAS.pdf>

ITESM. (2014). *Instituto Tecnológico y de Estudios Superiores de Monterrey*. Retrieved from Tutorea alumno: http://sitios.itesm.mx/prepanet/apoya_unete_tutoreo.htm

Mitchell, T. M. (2000). *Decision Tree Learning*. Retrieved from Washington State University: <http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CD0QFjAA&url=http%3A%2F%2Fwww.cs.princeton.edu%2Fcourses%2Farchive%2Fspr07%2Fcos424%2Fpapers%2Fmitchell-dectrees.pdf&ei=QAY7U-ngA4Xo2QXw7oCwAQ&usq=AFQjCNGMko1QbJWpXS8K8lzSZ>

Troyano, Y. y García, A. J. (2009). Expectativas del alumnado sobre el profesorado tutor en el contexto del Espacio Europeo de Educación Superior. *Revista de Docencia Universitaria*, 7(2), En línea: <http://red-u.net/redu/index.php/REDU/article/view/97/pdf>

UAM. (2016). *Universidad Autónoma Metropolitana*. Retrieved from ¿Qué necesito para ser un tutor?: <http://www.docenciadcsh.azc.uam.mx/tutorias/index.php/2014-03-13-02-37-51/que-necesito-para-ser-tutor>

UANL. (2011). *Universidad Autónoma de Nuevo León*. Retrieved from Manual del tutor: <http://www.uanl.mx/sites/default/files/dependencias/del/manualdetutor.pdf>

UNAM. (2005). *Universidad Nacional Autónoma de México*. Retrieved from Perfil de competencias del tutor de posgrado de la UNAM: https://www.aui.org/images/stories/DATOS/PublicacionesOnLine/seminario_taller_santo_domingo/PerfilCompetencias.pdf

Fecha de entrada: 20 Enero 2017

Fecha de revisión: 16 Marzo 2018

Fecha de aceptación: 20 Marzo 2018