

Vol. 19, No. 4, 2017

Brief Review of Educational Applications Using Data Mining and Machine Learning

Breve revisión de aplicaciones educativas utilizando Minería de Datos y Aprendizaje Automático

Argelia Berenice Urbina Nájera (1) argeliaberenice.urbina@upaep.mx
Jorge de la Calleja Mora (2) jorge.delacalleja@uppuebla.edu.mx

(1) Universidad Popular Autónoma del Estado de Puebla

(2) Universidad Politécnica de Puebla

(Received: March 9, 2016; accepted for publishing: June 24, 2016)

How to cite: Urbina, A. B., & De la Calleja, J. (2017). Brief review of educational applications using data mining and machine learning. *Revista Electrónica de Investigación Educativa*, 19(4), 84-96. <https://doi.org/10.24320/redie.2017.19.4.1305>

Abstract

The large amounts of data used nowadays have motivated research and development in different disciplines in order to extract useful information with a view to analyzing it to solve difficult problems. Data mining and machine learning are two computing disciplines that enable analysis of huge data sets in an automated manner. In this paper, we give an overview of several applications using these disciplines in education, particularly those that use some of the most successful methods in the machine learning community, such as artificial neural networks, decision trees, Bayesian learning and instance-based methods. Although these two areas of artificial intelligence have been applied in many real-world problems in different fields, such as astronomy, medicine, and robotics, their application in education is relatively new. The search was performed mainly on databases such as EBSCO, Elsevier, Google Scholar, IEEEExplore and ACM. We hope to provide a useful resource for the education community by presenting this review of approaches.

Keywords: Education, Data mining, Machine learning.

Resumen

La gran cantidad de datos utilizados en la actualidad han motivado la investigación y el desarrollo en diferentes disciplinas buscando extraer información útil con el fin de analizarla para resolver problemas difíciles. La Minería de datos y el Aprendizaje automático son dos disciplinas informáticas que permiten analizar enormes conjuntos de datos de forma automática. En este documento proporcionamos un panorama de varias aplicaciones que utilizan estas disciplinas en la Educación, particularmente aquellas que utilizan algunos de los métodos más exitosos en la comunidad de aprendizaje automático, como redes neuronales artificiales, árboles de decisión, aprendizaje bayesiano y métodos basados en instancias. Aunque estas dos áreas de la inteligencia artificial se han aplicado en muchos problemas del mundo real en diferentes campos, como la Astronomía, la Medicina y la Robótica, su aplicación en la Educación es relativamente nueva. La búsqueda se realizó principalmente en bases de datos como EBSCO, Elsevier,

Google Scholar, IEEEExplore y ACM. Esperamos proporcionar un recurso útil para la comunidad educativa con esta revisión de enfoques.

Palabras clave: Educación, Minería de datos, Aprendizaje automático.

I. Introduction

The huge amount of digital content available in different areas has motivated research in, and the development of, different disciplines that make it easier to search, organize and analyze this content. Data mining and machine learning are disciplines that have emerged to analyze this information in an automated manner, by finding patterns and relationships in raw data, and they have also contributed to solving complex problems.

Data mining and machine learning technologies have already achieved significant success in many areas including medicine (Cruz & Wishant, 2007; Er, Yumusak & Temurtas, 2010; Heckerling, Canaris, Flach, Tape, Wigton & Gerber, 2007; Kadhim A-Shayea, 2011; Pérez, De La Calleja, Medina & Benitez, 2012), astronomy (De la Calleja, Benitez, Medina & Fuentes, 2011; De La Calleja, Huerta & Fuentes, 2010), business (Alhah, Abu Hammad, Samhoury & Al-Ghandoor, 201; Salles, 2011; Tirenni, Kaiser & Herrmann, 2007;), robotics (Conforth & Meng, 2008; Engedy, 2009), and computer vision (Isik, Leibo & Poggio, 2012; Yokono & Poggio, 2009), to name just a few.

In an educational context, there are many interesting and difficult problems that may arise from four aspects: administrative problems, and problems associated with school, academic staff, and students. All these problems entail a large amount of data to be analyzed. The first aspect encompasses problems such as poor infrastructure, inadequate human resources, insufficient support and training for the incorporation of ICTs (Information and Communication Technologies) in the classroom, legislative instability, and excessive administrative bureaucracy, among others. The second is related to problems such as inadequate management training, indifference to current educational needs, incorrect grouping of students (many groups, diversity in the classroom, different skills and abilities), poor coordination with academic staff, bureaucratic overload for the academic staff, etc. The third includes the dearth of management and executive involvement, lack of engagement (motivation, commitment, self-esteem), quality of teaching (prioritizing content, monotony in the teaching process, dependence on textbooks, poor digital resources used), an absence of collaborative attitude and lack of training. Finally, the last aspect presents problems such as insufficient motivation (gap between society and the world of education), absence of discipline, high dropout rates (family problems, family destruction, economics, parental authority), and non-use of ICTs for learning (Arnaut & Giourguli, 2010; De Ibarrola, 2012; OCDE, 2010; Zorrilla & Barba, 2008).

In recent years, data mining and machine learning have been applied in education, providing some solutions to the above-mentioned problems. In this paper, we give an overview of several applications of these two computer disciplines in the context of education. Our search was based on: 1) using keywords such as “education” and “data mining”, and the application of data mining or machine learning algorithms in solving problems related to education; 2) performing a search in databases such as EBSCO, Elsevier, Google Scholar, IEEEExplore and ACM; and finally, 3) disregarding the year of publication of certain papers, as the study presented is a review of the state of the art.

The rest of the paper is organized as follows. In section 1 we give a general overview of the main processes involved in data mining, and we also briefly describe the most commonly used machine learning algorithms such as artificial neural networks, decision trees, instance-based methods and Bayesian learning. Then, in section 2, we introduce several successful applications of machine learning approaches to solve or automate different tasks in education. After that, in section 3 we offer a discussion, and finally, in the last section, we give conclusions and directions for future work.

II. Knowledge Discovery in Databases

Data mining can be defined as the process of extracting hidden knowledge 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 and usable, and must not be obvious. In classic database management 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 & Hall, 2011).

The data will take the form of a set of examples, while the output takes the form of predictions about new examples. These examples are the “things” that will be classified, associated or clustered, and are commonly named instances. Each instance is an individual, independent example of the concept to be learned. In addition, each instance is characterized by the values of a set of predetermined attributes, which will be the input parameters to machine learning algorithms. The value of an attribute is a measurement of the quantity to which the attribute refers (Han, Kamber & Pei, 2011). For example, in the context of education, a record of a student is an instance, the attributes of which may be the student’s name, age, grades, hobbies, etc; thus, the student may be classified or clustered in a particular problem according to these characteristics.

In order to provide a better understanding of data mining, we describe the data mining process that focuses on Knowledge Discovery in Databases (KDD), which mainly involves the following tasks: data collection, data processing, attribute selection (feature selection) and application of machine learning algorithms (figure 1).

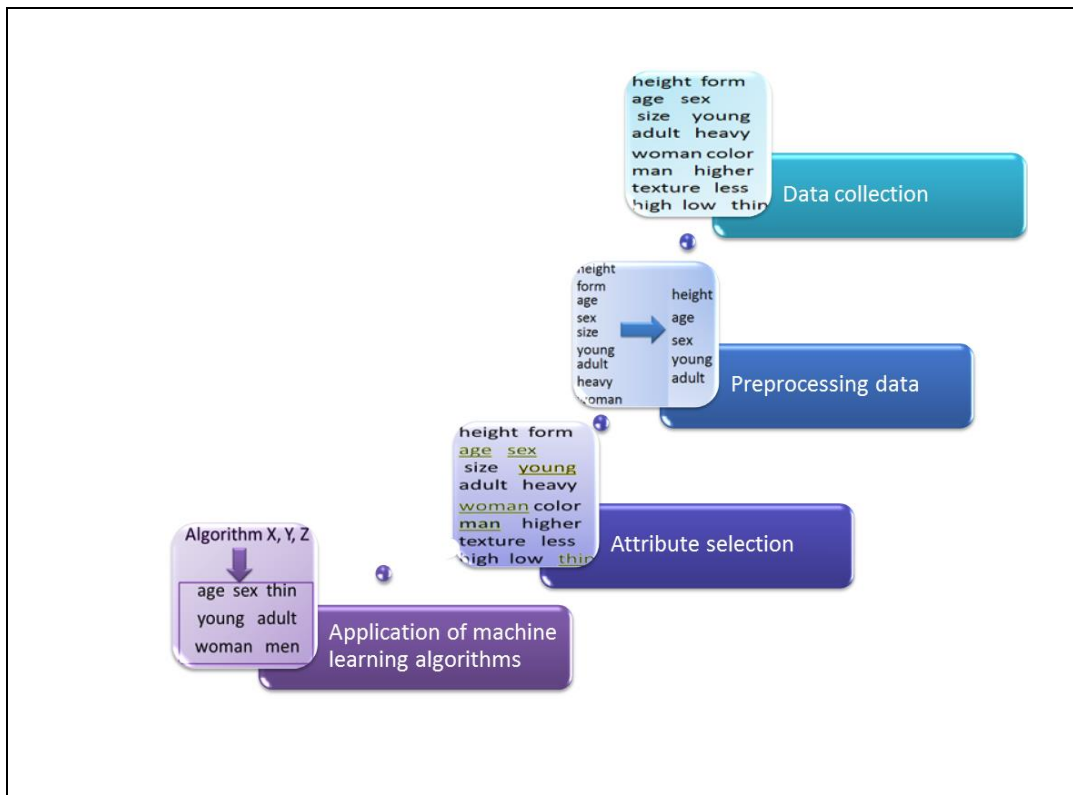


Figure 1. General process for Knowledge Discovery in Databases (KDD)

Data collection. Data are any facts, numbers or text that can be processed by a computer. Today, there exists a tremendous amount of data in different formats and databases, such as sales, cost, inventory,

forecast data, medical reports, images, and student information.

The process of data collection includes all the steps required to obtain the desired data in a digital format. Methods of data collection include acquiring and archiving new observations, querying existing databases according to the science problem at hand, and performing as necessary any cross-matching or data combining (Han et al., 2011).

Data processing. Data processing can be divided into steps that enable a more meaningful reading of the data, and steps that transform the data in some way to make it appropriate for a given machine learning algorithm (Han et al., 2011).

Attribute selection. In order to perform their task (classification, prediction, recognition, clustering, decision-making, etc.), machine learning algorithms require the attributes, i.e., the numerical or nominal values in the data fields that describe the properties of each instance. Numerical attributes, sometimes called continuous attributes, measure real numbers or integers, for example age, width, height, weight, etc. Nominal attributes take on values in a prespecified, finite set of possibilities, and they are sometimes called categorical attributes, and include, for example, name, size, country, color, etc. It is possible to transform numerical data to categorical data and vice versa (Witten & Frank, 2005).

Frequently, a large number of attributes will be available for each instance in a data set; however, not all will be required to solve a problem. Indeed, if we use all the attributes, this may, in many cases, cause a poor performance of the machine learning algorithms (Witten & Frank, 2005). In formal terms, the large number of attributes represents a high-dimensional space; therefore, sometimes it is necessary to carry out a reduction of the dimensionality by choosing only a few attributes. This small set of attributes must retain as much of the relevant information as possible (Bishop, 2007). For example, you might have a file with data on students, with several attributes, which could then be reduced in order to identify the most relevant characteristics for a given purpose, either clustering or classification.

Machine learning. Learning may be the most distinctive characteristic of human intelligence, and includes knowledge acquisition, skill development, knowledge organization and discovery of facts, among other aspects (Mitchell, 1997). Machine learning tries to mimic these processes by studying and modeling them computationally, and is generally divided into two approaches: supervised learning (classification) and unsupervised learning (clustering), which are also known as predictive and descriptive, respectively (Bishop, 2007).

Supervised learning relies on a training set of examples for which the target property, for example a classification, is known with confidence. Then, an algorithm is trained on this set of examples and the resulting mapping is applied to further examples (a testing set) for which the target property is not available. The training set must be representative, meaning that the parameter space covered by the input attributes must span that for which the algorithm is to be used (Mitchell, 1997). For example, we may wish to classify the performance of students as good or bad, in which case we would need to provide the machine learning algorithm with a training data set with examples of students that fit these two types of classification. Then, once the machine learning algorithm has been trained, it can classify data for unknown students.

In contrast to supervised learning, the unsupervised approach does not use information about the target property for the instances and attempts to identify patterns in the data by creating natural groups. Unsupervised algorithms usually require some kind of initial input. For example, for a given data set of student records, we can let the machine learning algorithm make clusters of these records, according to the value attributes they have in common.

Artificial neural networks. Artificial neural networks have been one of the most commonly used methods for machine learning tasks. This method is based on the observation of biological neural systems, which are formed by sets of units called neurons, which are interconnected. A neuron is simply a switch with input and output information. The switch is activated if there are enough stimuli from other neurons yielding the input information, then a pulse is sent to other neurons. Incoming signals from other

neurons are transferred to a neuron via special connections, the synapses (Kriesel, 2013).

Technically, an artificial neural network consists of simple processing units (neurons) and directed weighted connections (synapses) between those neurons. Each unit has an activation function to produce a stimulus in order to activate other neurons. Generally, an artificial neural network works as follows: each unit (except for the input) receives the output (information) from all units in the previous layer and produces its own output, which then feeds the nodes in the next layer. For a given network architecture, the first step is training the artificial neural network, which is when the weights of the interconnections are determined (Mitchell, 1997). Once the artificial neural network has been trained, it can be used for several tasks, for example predicting students' grades. In figure 2 we show the typical architecture of an artificial neural network.

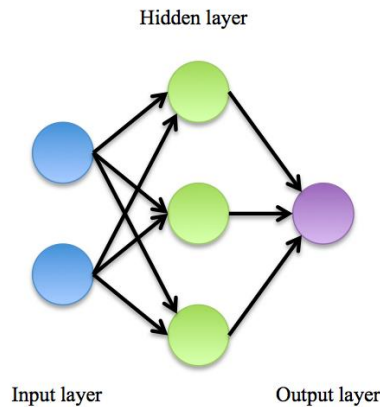


Figure 2. A simple artificial neural network

Note: Neurons are represented by circles and synapses by arrows. Blue circles are the input neurons (which receive attribute information), green circles are the hidden neurons, while the output is given by the purple circle.

Decision trees. Decision tree learning is another of the most widely used methods for inductive inference, and decision trees have been successfully applied to a broad variety of practical problems. A decision tree provides a graphical description of the decisions to be made, the attributes selected, and the outcomes associated with combinations of decisions and attributes. Basically, decision trees are made up of nodes and branches. Nodes are the points where a choice must be made, while branches represent one of the possible alternatives or decision routes. A root node is selected where the decision process begins (Mitchell, 1997). In figure 3 we show an example of a decision tree; we can observe that there are two pathways for accreditation, and five pathways for non-accreditation.

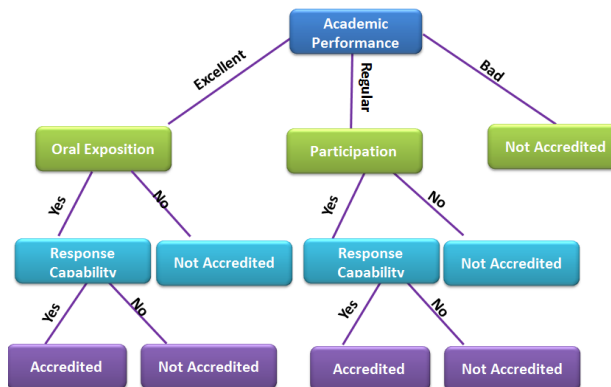


Figure 3. Decision tree for classifying students as accredited or not accredited

Instance-based methods. These methods are simple and easy to use, and they obtain results as good as those obtained by artificial neural networks or decision trees. We can use the well-known duck test to exemplify how these kinds of methods perform. The test says “If it looks like a duck, swims like a duck and quacks like a duck, then it probably is a duck”; i. e. in instance-based learning, data will be classified according to the most similar examples previously stored.

The method most often used in this kind of learning is k-nearest neighbors, where the k closest examples are selected to identify the target property of new data. Then, a distance function has to be defined in order to find the closest examples; generally the Euclidean distance is used. In figure 4 we show an example of the classification of a new instance considering the k-nearest examples.

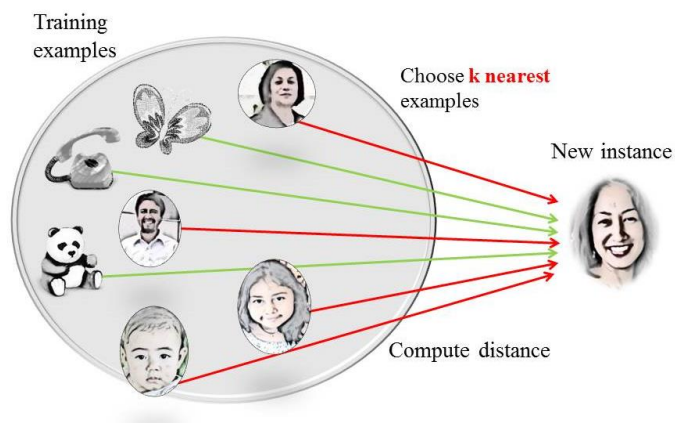


Figure 4. The k-nearest neighbor algorithm selects the k closest examples in order to classify new instances

Bayesian learning. Bayesian learning is, generally speaking, a method used for selecting the best hypothesis in terms of how well it can explain the observed training data. The naive Bayes classifier is one of the most practical and widely used inductive learning algorithms. This method is based on the assumption that the attribute values for the examples are conditionally independent given the target values (Mitchell, 1997). The naive Bayes algorithm applies to data mining tasks where each example can be described as a tuple of attribute values and the target function can take on any value from a finite set. This algorithm, in order to predict the target properties of new instances, assigns the most probable target value to these instances according to probabilities for the data set, by using their value attributes.

Applications of Data Mining and Machine Learning. Several approaches have been proposed so far to solve different problems in education. Recently, data mining and machine learning have been used to offer solutions to these problems in an automated manner. We present studies that have used supervised learning methods, and in particular the most successful ones used in the machine learning community, such as artificial neural networks, decision trees, Bayesian learning and instance-based methods. We also performed a search in several of the most important databases such as EBSCO, Elsevier, Google Scholar, IEEEXplore and ACM, using the keywords “data mining” and “machine learning in education”. Because Educational Data Mining (EDM) emerged recently, around 2006, we decided to present work from 2003 onwards.

In 2003, Minaei-Bidgoli, Kashy, Kortemeyer and Punch presented an approach to predict the final grade of students, based on several features extracted from logged data in a web-based education system. The data set consisted of 227 examples, which were grouped in different cases according to their final grades. They used the Bayes classifier, k-nearest neighbors, artificial neural networks and decision trees. They reported that k-nearest neighbors obtained the best results, with 82.3% accuracy; however, decision trees also performed well.

Thomas and Galambos (2004) used regression analysis and decision trees with the CHAID (CHi-squared Automatic Interaction Detection) algorithm to identify the academic satisfaction of students in three

major aspects: academic experiences, social integration and campus services-facilities. The results obtained using the decision tree analysis revealed that social integration is a determining factor for students' satisfaction, particularly for students who are less academically engaged. On the other hand, campus services and facilities had less effect on their satisfaction.

Kotsiantis, Pierrakeas and Pintelas (2004) performed two experiments to predict new students' performance in computer engineering in the Hellenic Open University, Greece. They experimented using six algorithms: decision trees, neural networks, naive Bayes, k-nearest neighbors, logistic regression and support vector machines. In the first experiment, they used 354 cases, while for the second experiment, only 28 cases were considered, as it was difficult to obtain more than 30 cases per academic year. After performing the second experiment, they found that these 28 cases are probably insufficient if greater accuracy is desired; they suggested using at least 70 cases for better accuracy. When the six algorithms were compared, the authors concluded that the naive Bayes algorithm is the most appropriate to build a software support tool to predict the performance of new students. The overall accuracy for the two experiments was best with the naive Bayes algorithm, with about 72% accuracy, and the overall sensitivity was about 78%. Moreover, the authors advise that the naive Bayes algorithm is easier to implement.

One of the objectives of a web-based education system is that students are able to learn despite their different learning styles. In order to achieve this goal, first it is necessary to identify how students learn. Therefore, we introduce three approaches that address this problem using Bayesian learning. In the first study, presented by García, Amandi, Schiaffino and Campo (2005), learning styles were identified using Felder's model. In the second paper, García, Amandi and Campo (2007) identified the students' learning style based on their behavior, which was recorded while they interacted with the educational system. Then a Bayesian network was used to infer the learning style from the modeled behavior. Their proposed approach was evaluated on an artificial intelligence web course. Finally, Schiaffino, García and Amandi (2008) introduced an approach that uses an intelligent agent (e-teacher) to detect learning styles based on students' actions on an e-learning system. They tested their approach using the students' profile information in order to suggest customized courses for them.

In 2007, Acevedo, Caicedo and Loaiza (2007) used intelligent systems to decrease subjectivity in the selection processes of the navy of the Republic of Colombia. Also, they used artificial neural networks to classify psychological patterns with the goal of identifying factors in cadet desertion. They used two types of neural networks: multilayer perceptron and radial basis functions. To perform their experiments they used a data set of 200 psychological profiles for cadets admitted from 1995 to 1998. The results show that the system was able to predict the admission performance with a success rate of about 70%. The authors state that the prediction accuracy increases when the neural network is trained with more data concerning psychosocial problems of high complexity. They also highlight the efficiency of radial basis neural networks to solve pattern classification problems.

Oladokun, Adebajo and Charles-Owaba (2008) proposed a neural network model to predict the performance of possible candidates for admission to university. Their model was based on a multilayer perceptron topology, with family history, age, and score of the entrance exam taken as input variables, among others. They reported about 74% accuracy in their results, but recommended conducting an oral interview to obtain more information to supply to their model.

In 2008, Baylari and Montazer introduced a multi-agent e-learning system to recommend customized materials for students, based on personalized Item Response Theory (IRT) using an artificial neural network. They took students' goals, as well as their experiences, knowledge, abilities and interactivity, as input parameters. They found that one of the agents was able to diagnose learning problems, thus students could receive adaptive testing in order to obtain appropriate learning materials. Their experimental results show that the approach can recommend customized materials in accordance with courses, with an accuracy of 83.3%.

Karamouzis and Vrettos (2008) used artificial neural networks to predict grades for college graduate students. The neural network model had three-layer perceptron architecture, and was trained using the

backpropagation algorithm. To perform the experiments, they used a data set of 1,407 student profiles from Wabauunsee University (USA); 1,100 profiles were used for training, and the remainder for testing. They reported a mean squared error of 0.22, with an accuracy of 70.27%.

Data mining was used by Ranjan and Khalil (2008) with two main objectives: 1) planning a course for education management through new data mining applications and to explore the effects on probable changes in the recruitment and admission processes and guiding courses; and 2) ensuring quality evaluations, student performance, courses and tasks. Also, they were interested in finding patterns in how students interact with others, how the admission process is carried out, and the mechanisms in counseling to choose courses. The framework was tested to find what types of courses are interesting to certain types of students. In addition, the proposed framework had three major processes that usually occur in all management institutions, namely admissions (planning, evaluation and registration), counseling, and allocation of specialization subjects. The results were presented as a conceptual framework to adopt data mining in management in institutions. The authors concluded that data mining is useful for predicting the success of educational programs, and also understanding learning styles in order to promote proactivity in students.

Márquez, Ortega, Gonzalez-Abril and Velasco (2008) used a Bayesian network to find a route to acquire professional competences. This method was able to predict the probability of success, considering student profiles and historical data. Their results show that the route varies when students' preferences and needs are changed, taking into account pedagogical unit learning and social behavior. In addition, they introduced an approach to design a useful path for dynamic learning using the ant colony optimization method.

In 2009, Vialardi, Bravo, Shafti and Ortigosa suggested using a subject recommender system based on data mining to help students to make better decisions when they have to take courses. They analyzed data relating to engineering students at the University of Lima, Peru, from the last seven years. Their experiments were performed using the C4.5 algorithm, which is a decision tree-based method. Experimental results show a global accuracy of 77.3%, with conditions modified for training and testing. The authors also comment that their approach provides relevant information on students that would not be available when using descriptive statistical techniques.

In 2011, Anupama and Vijayalakshmin used decision trees, particularly ID3 and C4.5, to predict the performance of students in their final exam. This prediction aided their tutors in identifying weak points in order to improve students' 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 offers several advantages in higher learning institutions, and these techniques can be applied in other areas of education to optimize resources and predict the retention of members of faculty and tutor feedback, among other purposes.

Kumar and Pal (2011) used decision trees to extract a set of academic characteristics to assess students' performance. The data set consisted of 50 examples from students of the computer applications department at the vbs Purvanchal University, India. The characteristics considered included grades in previous semesters, seminar performance, general proficiency performance, and attendance. The knowledge extracted and represented by the decision tree enabled the authors to obtain if-then rules to classify the students. With this work, the authors predicted students' end-of-semester performance and identified students who needed special attention to reduce the failure rate.

In 2013, Ayinde, Adetunji, Bello and Odeniyi described an approach to predict and classify the academic performance of students using the naive Bayes classifier and decision trees. The experiments were performed using a data set of 473 students from the Department of Computer Science and Engineering, at the Ladoke Akintola University of Technology, Nigeria. The results were obtained by applying 10-fold cross-validation and percentage split. The authors reported that the best result was obtained using the Bayesian classifier, with 70% accuracy.

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 administering 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 prediction of student dropout in the online program data set. Finally, the authors concluded that data mining methods might help to predict different reasons why students decide to drop out before finishing their study programs.

Another problem that has been tackled using machine learning algorithms was introduced by Kakavand, Mokfi and Tarokh (2014), with the purpose of predicting student loyalty using decision trees. The authors researched the external factors that may generate loyalty, in order to identify students who have decided to continue studying, and thus the university may invest in them and increase 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, and income, among others). The best result was obtained using the CART decision tree algorithm, with 94% accuracy.

III. Software to experiment with Data Mining and Machine Learning

As mentioned throughout this paper, data mining can be defined as the process of extracting hidden knowledge from huge volumes of raw data. We describe software tools used to perform data mining tasks and include some important characteristics.

WEKA is a popular piece of software that implements machine learning algorithms for data mining tasks. This software contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization. Some advantages of Weka are that it is free under GNU Public License, it contains a wide range of modeling and data processing techniques, and it has a very simple user interface that makes it accessible even for inexperienced users. WEKA supports several standard data mining tasks, including data preprocessing, clustering, classification, regression, visualization and feature selection (Hall et al., 2009).

Orange is a data mining component and machine learning software tool that offers a visual, fast and versatile program for exploratory data analysis. It has a friendly and intuitive interface. This software allows preprocessing, information filters, data modeling, and model evaluation and exploration techniques (Demsar et al., 2013).

RapidMiner is an open-source data science platform. This tool offers advanced analytics through template-based frameworks. It offers functions such as data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. Additionally, it provides learning schemes, models and algorithms from WEKA and R scripts (RapidMiner, 2016)

Apache Mahout. The goal of this tool is to build an environment for quickly creating scalable performance machine learning applications. It provides three major features: 1) a simple and extensible programming environment and framework for building scalable algorithms; 2) a wide variety of premade algorithms for Scala + Apache Spark, H2O, and Apache Flink; and finally, 3) Samsara, a vector math experimentation environment with R-like syntax that works at scale. It implements algorithms such as canopy clustering, k-means, fuzzy k-means, streaming k-means, and naive Bayes (The Apache Software Foundation, 2014).

Rattle is a free and open-source data mining toolkit, written in the statistical language R, which uses the Gnome graphical interface. It runs on GNU/Linux, Macintosh OS X, and MS/Windows. Rattle is used in business, government, research and for teaching data mining in Australia and internationally. It presents statistical and visual summaries of data, transforms data that can be readily modeled, builds both unsupervised and supervised models from the data, presents the performance of models graphically, and scores new datasets (Toware, 2014).

KNIME is integration-friendly, intuitive and easy-to-use software that allows processing, data analysis and data exploration. Graphy allows you to create data flows, selectively execute an analysis, study the results, model and generate interactive views to facilitate decision-making at the management level. At the same time, it allows text edition, and image and time series processing, among other functions. It has modules for connectors for all major file formats and databases, it supports a wealth of data types: XML, JSON, images, documents, and many more; it offers native and in-database data blending and transformation, and advanced predictive and machine learning algorithms, among other features (Hofer y KNIME, 2016).

DMelt. DataMelt is a successor of the popular jHepWork (2005-2013) and SCaVis (2013-2015) programs, which have been under intensive development since 2005. It is fully backward compatible with jHeoWork 3.9 and SCaVis 2.3. DMelt is a portable application. No installation is needed: simply download and unzip the package, and you are ready to run it. DataMelt exists as an open-source portable application, and as JAVA libraries under a commercial-friendly license. This software is used for numeric computation, statistics, analysis of large data volumes ("big data") and scientific visualization. The program can be used in many areas, such as the natural sciences, engineering, and the modeling and analysis of financial markets. It can be used with different programming languages on different operating systems (jWork.ORG & Chekanov, 2016).

CLUTO is a software package for clustering low- and high-dimensional data sets and analyzing the characteristics of the various clusters. Its main features are multiple classes of clustering algorithms (partitional, agglomerative and graph-partitioning-based), multiple similarity/distance functions (Euclidean distance, cosine, correlation coefficient, extended Jaccard, user-defined), numerous novel clustering criterion functions and agglomerative merging schemes, traditional agglomerative merging schemes, extensive cluster visualization capabilities and output options (gif, xgif, postscript, etc.), and it can scale very large datasets containing hundreds of thousands of objects and tens of thousands of dimensions (Karypis, 2015).

As we can see, there are several software tools for data mining and machine learning tasks. However, Witten, Frank and Hall (2011) suggested using **WEKA** as it was designed to rapidly try out existing methods on new data sets in flexible ways. It provides extensive support for the whole process of experimental data mining, including preparing 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.

IV. Discussion

Data mining has been used in recent decades to extract patterns from raw data in order to obtain valuable information. In addition, machine learning has been applied to perform several tasks including recognition, classification, identification, prediction and detection, in an automated manner. Also, these computer disciplines have been very useful to discover relationships not previously suspected, make automated decisions and define competitive strategies to examine large amounts of data.

Choosing a machine learning algorithm to perform a task is difficult, as several issues have to be considered, such as the amount of information, the quantity and types of attributes, the number of classes, the number of instances/examples and, primarily, the application domain.

Nevertheless, we can describe some advantages and disadvantages of the most common algorithms. The most remarkable advantages of neural networks include their good predictive power and their hardiness to deal with irrelevant or redundant attributes, and their main disadvantage is that, as black box models, it is difficult to interpret their operation, and they are also very time-consuming. The strengths of decision trees are that they are interpretable models, robust to outliers and noisy or redundant attributes, and also their input and output may use numerical or categorical variables. The main disadvantage of decision trees is that they may generate a very large tree that can lead to poor predictions. The advantages of

Bayesian learning include a strong mechanism for processing uncertain information, flexible applicability and the ability to handle missing data, while the main disadvantage is that it needs a large data set to make reliable estimations of the probability for each class. Finally, instance-based methods use all available information, they do not require training and have good predictive power, yet time-consuming computational processing and performance may be affected by noise or irrelevant attributes.

V. Conclusions

Data mining and machine learning have recently been applied in the academic context, allowing educational institutions to better allocate human and material resources, manage student performance and improve the effectiveness of performance throughout students' education. The approaches described in this paper have shown different ways to solve many interesting academic problems by applying some of the most widely used data mining and machine learning techniques such as artificial neural networks, k-nearest neighbors, Bayesian learning and decision trees.

Future work may include creating automated methods for identifying students' skills, improving decision-making, tutor assignment using certain criteria, lesson and material recommendation, improving the admission process, and predicting student performance, among other purposes. Many other machine learning approaches can also be applied, such as dealing with imbalanced data sets, selecting the best attributes, using ensembles of algorithms and applying data reduction techniques.

References

- Acevedo, G. L., Caicedo, E. F., & Loaiza, H. (2007). Selección de personal mediante redes neuronales artificiales [Personnel selection through artificial neural networks]. *Revista de Matemática: Teoría y aplicaciones*, 14(1), 7-20.
- Anupama, S., & Vijayalakshmi, M.N. (2011). Efficiency of decision trees in predicting student's academic performance. *Computer Science and Information Technology*, 2, 335-343.
- Arnaut, A., & Giourguli, S. (2010). *Los grandes problemas de México: VII educación* [The main issues in Mexico: VII Education]. Mexico City: El Colegio de México.
- Ayinde, A. Q., Adetunji, A. B., Bello, M., & Odeniyi, O. A. (2013). Performance evaluation of naive bayes and decision algorithms in mining students' Educational Data. *International Journal of Computer Science Issues*, 10(4), 147-151.
- Baylari, A., & Montazer, G. A. (2008). Design a personalized e-learning system based on item response theory and artificial neural network approach. *Expert Systems with Applications*, 39, 8013-8021.
- Bishop, C. M. (2007). *Pattern recognition and Machine Learning*. Singapore: Springer.
- Cheng, S.-C., Huang, Y.-M., Chen, J.-N., & Lin, Y.-T. (2005). *Automatic Leveling System for E-Learning Examination Pool Using Entropy-Based Decision Tree*. Advances in Web-Based Learning-ICWL 2005: 4th International Conference, Hong Kong, China, July 31st - August 3rd.
- De Ibarrola, M. (2012). Los grandes problemas del sistema educativo mexicano [The main issues in the Mexican education system]. *Perfiles educativos*, XXXIV (especial), 16-28.
- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., et al. (2013). Orange: data mining toolbox in python. *Journal of Machine Learning Research* 14, 2349-2353.

- García, P., Amandi, A., Schiaffino, S., & Campo, M. (2005). *Using bayesian networks to detect students' learning styles in a web-based education system*. 7o. Simposio Argentino de Inteligencia Artificial-ASAI2005, 115-126. Rosario, Argentina: Universidad del Centro de la Provincia de Buenos Aires.
- García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, 49, 794-808.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1).
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Amsterdam: Morgan Kaufmann.
- Hofer, H., & KNIME. (2016). *Oper for innovation*. Retrieved from <http://www.knime.org>
- jWork.ORG, & Chekanov, S. (2016). *Data melt: computation and visualization enviroment*. Retrieved from <http://jwork.org/dmelt>
- Kakavand, S., Mokfi, T., & Jafar, M. (2014). Prediction the loyal student using decision tree algorithms. *International Journal of Information and Communication Technology Research*, 4(1), 32-37.
- Karamouzis, S. T., & Vrettos, A. (2008). *An artificial neural network for predicting student graduation outcomes*. Proceedings of the World Congress on Engineering and Computer Science. WCECS 2008, 1-4. San Francisco.: International Association of Engineers.
- Karypis, G. (2015). Karypis Lab. Retrieved from <http://qlaros.dtc.umn.edu/gkhome/cluto/cluto/overview>
- Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting Students' performance in distance learning using machine learnings techniques. *Applied Artificial Intelligence*, 18, 411-426.
- Kriesel, D. (2013). *A brief introduction to neural networks*. Retrieved from http://www.dkriesel.com/en/blog/2009/1011_a_brief_introduction_to_neural_networks_published_in_epsilon_version
- Kumar, B., & Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6), 63-60.
- Márquez, J. M., Ortega, J. A., Gonzalez-Abril, L., & Velasco, F. (2008). *Creating adaptive learning paths using ant colony optimization and bayesian networks*. Neural Networks, 2008. IEEE World Congress on Computational Intelligence, 3834-3839. Hong Kong, China.
- Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). *Predicting student performance: an application of data mining methods with the educacional we-based system LON-CAPA*. 33rd ASEE/IEEE Frontiers in Education Conference, 1-6. Boulder, CO.
- Mitchell, T. M. (1997). *Machine learning*. Singapore: McGrawHill.
- OCDE. (2010). *Mejorar las escuelas: estrategias para la acción en México. Resumen ejecutivo* [Improving schools: Strategies for action in Mexico. Executive summary]. Autor.
- Oladokun, V. O., Adebajo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: a case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72-79.

Ranjan, J., & Khalil, S. (2008). Conceptual framework of data mining process in management education in India: an institutional perspective. *Information Technology Journal*, 7(1), 16-23.

RapidMiner. (2016). *RapidMiner*. Retrieved from <http://rapidminer.com>

Russell, P., & Norving, S. (2010). *Artificial Intelligence: a modern approach* (3rd. ed.). Prentice Hall.

Schiaffino, S., García, P., & Amandi, A. (2008). E-teacher: providing personalized assistance to e-learning students. *Computers & Education*, 51, 1744-1754.

The Apache Software Foundation. (2014). *Mahout*. Retrieved from <http://mahout.apache.org>

Thomas, E. H., & Galambos, N. (2004). What satisfies students? Mining student-opinion data with regression and decision tree analysis. *Research in Higher Education*, 45(3), 251-269.

Toware. (2014). *Data mining resources*. Retrieved from <http://datamining.togaware.com>

Vialardi, C., Bravo, J., Shafti, L., & Ortigosa, Á. (2009). *Recommendation in Higher Education using data mining techniques*. Second International Conference on Educational Data Mining (pp. 190-199). Cordoba, Spain: International Working Group on Educational Data Mining.

Witten, I., Frank, E., & Hall M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd. ed.). New Zealand: Morgan Kaufmann.

Yukselturk, E., Ozekes, S., & 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.

Zorrilla, M., & Barba, B. (2008). Reforma Educativa en México: Descentralización y nuevos actores. *Sinéctica*, 30, 1-32. Retrieved from <https://sinectica.iteso.mx/index.php/SINECTICA/article/view/189>