Predicting academic performance in traditional environments at higher-education institutions using data mining: A review

César A. Del Río / Julio A. Pineda Insuasti
Professors at the Universidad Técnica del Norte, Ibarra, Ecuador
cadelrio@utn.edu.ec

ABSTRACT

The purpose of this review paper is to survey the academic literature of the past five years in the area of data mining applied to the educational domain in a traditional classroom-based environment at Institutions of Higher Education (IHEs). We describe Educational Data Mining (EDM) and its main methods and techniques as they are applied to predicting academic performance in a traditional setting, and we proceed then to review 56 primary-research articles on the subject. We also examine 5 other review papers that have preceded this one in the last five years. To our knowledge, this is the first review article to focus exclusively on applying EDM to predict academic performance in a traditional setting. We determine that classification is by far the most popular method used by the primary-research studies, followed by clustering and association rule mining. We conclude that the success experienced by researchers abroad in predicting academic performance can be replicated in Ecuador, provided that we avoid the pitfalls of over-reliance on software and that we do not underestimate the complexities and need for human intervention that are involved in an EDM project. Moreover, the shortage of Big Data expertise in the country will need to start getting addressed.

Keywords: EDUCATIONAL DATA MINING, EDM, DATA MINING, PREDICT ACADEMIC PERFORMANCE, TRADITIONAL EDUCATIONAL ENVIRONMENT

RESUMEN

Predicción del rendimiento académico en instituciones de educación superior con ambientes tradicionales mediante el uso de la minería de datos: Una revisión

El propósito de este artículo de revisión es examinar la literatura académica de los últimos cinco años en el área de la minería de datos aplicada a la educación en un ambiente tradicional de aulas de clases en Instituciones de Educación Superior. Se describe a la Minería de Datos para la Educación (MDE) y sus métodos y técnicas principales tal como son aplicadas para predecir el rendimiento académico en un ambiente tradicional, y después se revisa 56 artículos de investigación primaria sobre el tema. Se examina también otros 5 estudios de revisión que han precedido a éste en los últimos 5 años. Se conoce, éste es el primer artículo de revisión que está enfocado exclusivamente en la predicción del rendimiento académico en un ambiente tradicional. La clasificación es el método que más se usó en los estudios de investigación primaria, seguida por el agrupamiento y la minería de reglas de asociación. Se concluye que el éxito experimentado por investigadores en el extranjero al predecir el rendimiento académico puede replicarse en el Ecuador, siempre que se evite el exceso de confianza en los paquetes de software de minería de datos, y que no se subestimen las complejidades y la necesidad de la intervención humana involucrados en un proyecto de MDE. Asimismo, la escasez de experticia en el área de "Big Data" en el país tiene que empezar a ser resuelta.

Palabras Claves: MINERÍA DE DATOS EN LA EDUCACIÓN, PREDECIR EL RENDIMIENTO ACADÉMICO, AMBIENTE EDUCATIVO TRADICIONAL

Introduction

Institutions of higher education (IHEs) are very focused on their students' academic performance (CES, 2016; IESALC, 2006; WCHE, 1998). Not only is this an accreditation and evaluation criterion, but it is also at the heart of the institution's contribution to successfully preparing the next generation for their future ahead.

Students' academic performance is measured by, among other things, course grades (Durairaj & Vijitha, 2014; Sweeney, Lester, & Rangwala, 2015), semester Grade Point Average (GPA) (Jacob, Jha, Kotak, & Puthran, 2015; Kolo, Adepoju, & Alhassan, 2015), cumulative GPA (CGPA) and final GPA (Arsad, Buniyamin, & Manan, 2014; Tekin, 2014). Alternatively, an IHE is interested in other academic success measurements such as the dropout rate (Iam-On & Boongoen, 2015), whether its graduates go on to pursue a higher-level degree (Borkar & Rajeswari, 2013), and how long it takes its graduates to obtain employment (Ramanathan, Geetha, Khalid, & Swarnalatha, 2016).

Given the importance of academic performance and success, an IHE would benefit from being able to predict its students' future academic performance. This would allow the IHE to take corrective measures to assist its at-risk students and to enhance their chances of graduating, and to do so with better grades.

One way to predict academic performance that is being explored by an increasing number of IHEs is the use of data mining techniques (Peña-Ayala, 2014). Data mining is the process of discovering useful patterns and trends in large data sets. Data mining in the field of education, known as Educational Data Mining (EDM), is a growing area of research (Shahiri, Husain, & Rashid, 2015). Educational data mining (EDM) is concerned with developing, researching, and applying computerized methods to detect patterns in large collections of educational data. EDM can also be described as the application of data mining techniques to data sets that come from educational environments to address important educational questions (Romero & Ventura, 2013).

There are three main research avenues in EDM (Romero, Ventura, Pechenizkiy, & Baker, 2011a): developing computational tools and techniques, finding educational stakeholders that could benefit from EDM, and determining what questions to ask of the data. Within the realm of determining which questions to ask, EDM has been applied, among other things, to the goal of predicting students' academic performance (Romero, Ventura, Pechenizkiy,

& Baker, 2011b).

The data sets that are available at IHEs providing education via a traditional classroom environment contain a range of student data, such as grades, attendance, socio-economic information, and high-school transcripts (Aziz, Ismail, Ahmad, & Hassan, 2015). Additionally, IHEs providing internet-based distance education and other types of electronically-delivered education via Learning Management Systems and other means, collect vast amounts of data captured in server logs, where the student's interaction with the learning environment can be traced in minute detail (Sheard, 2011). Due to the difference in volume and in the type and range of data between data sets from IHEs offering traditional educational environments, and those from IHEs offering electronically-delivered education, the EDM techniques that are used in each environment also vary.

This paper reviews the academic literature since 2011 for the application of EDM to predicting students' academic performance in a traditional educational environment. This is still the most widely used education delivery medium, based mainly on face-to-face contact between educators and students through lectures, class discussion, small groups, and so forth (Romero & Ventura, 2013). To our knowledge, this is the first EDM survey article that examines the literature on the application of EDM to the prediction of academic performance only as it applies to traditional classroom environments.

Methodology. Research questions

The research questions for this article were determined using the PICO (Population, Intervention, Comparison, Outcomes) criteria adapted by Kitchenham & Charters (2007) from the medical field for use in the framing of research questions in software engineering. For examples of other studies using the PICO criteria, see Shahiri et al. (2015) and Bolaji (2015). The application of the PICO criteria for this review is shown in Table 1.

Table 1. *PICO criteria for research questions*

Criterion	Question element
Population	The student population of an IHE offering traditional classroom-based education
Intervention	EDM techniques, such as classification, clustering and association
Comparison	Findings in the 2011 survey by Romero, Ventura et al, and in the 2014 survey by Peña-Ayala
Outcomes	Predicting academic performance in a traditional environment

Putting together the question elements arising from each one of the four criteria, then, the research question pursued by this survey article is: how is EDM used to predict students' academic performance in a traditional educational environment at IHEs?

Literature search strategy

The core list of peer-reviewed articles and conference proceedings to be examined by this study was obtained by searching the Scopus database for all English-language documents dated 2011 or later using the following search strings:

- "educational data mining" (with quotes), and
- predict performance

The search was performed on August 31, 2016. The list was subsequently pruned by discarding all documents that did not deal with a traditional education environment. The list was then supplemented with other articles on the subject from Google Scholar, Sci-Hub, and Google Search. The full text of all the documents on the resulting list was obtained. The final list contained 56 primary research journal articles, book chapters and conference proceedings that dealt with applying EDM to predict academic performance in a traditional environment.

Additionally, five EDM survey articles and edited books were reviewed, and they are listed in Table 2.

Table 2. Survey literature reviewed

#	Author (year)	Title
1	Thakar, Mehta, & Manisha (2015)	Performance analysis and prediction in educational data mining: A research travelogue
2	Shahiri et al. (2015)	A review on predicting student's performance using data mining techniques
3	Al-Razgan, Al-Khalifa, & Al-Khalifa (2014)	Educational data mining: A systematic review of the published literature 2006-2013
4	Peña-Ayala (2014)	Educational data mining: A survey and a data mining-based analysis of recent works
5	Romero et al. (2011a)	Handbook of Educational Data Mining

Educational data mining in a nutshell

This section represents a backgrounder on EDM, with an emphasis on those approaches, methods, and techniques used for the prediction of academic performance in a traditional classroom environment. The following subsections are included:

- Knowledge discovery process with EDM: This subsection describes the way EDM ferrets out knowledge and insights from data from educational sources
- Classification overview: This subsection provides a description of the most important method to predict academic performance in EDM, classification
- Classification techniques: Here the most important techniques to classify a data set containing educational data are presented
- Other methods that classify: Finally, here other techniques are presented that sometimes are used as classifiers.

Knowledge-discovery process with EDM

Combining the approaches discussed in Romero and Ventura (2013), and Larose and Larose (2015), the general knowledge-discovery process for an EDM project comprises the following steps:

1. Raw-data pre-processing
2. Exploratory data analysis
3. Data mining
4. Interpretation of results

Raw-data pre-processing. The data contained in databases, whether of an educational nature or not, is often incomplete and inaccurate; there may be obsolete or redundant fields, values may be missing, outliers may be present, the data may not be in a data-mining-ready form, or the values may defy common sense. These raw data need to be put through the steps of data cleaning, handling missing data, and data transformation. Depending on the data set, the data pre-processing step alone may account for some 10-60% of the time and effort for the entire EDM project (Larose & Larose, 2015; Romero & Ventura, 2013). Here are several ways to pre-process the data so it can be ready for the data mining step proper:

- Data cleaning: each field in the data set needs to be verified to make sure the values are correct. Incorrect values need to be fixed to the greatest degree possible
- Addressing missing data: For fields with gaps in their values, the missing values need to be supplied. This can be done by replacing the missing value(s) with some constant. For numeric fields, the missing values could be replaced by the field mean; for categorical values, the mode may be used. Another approach is to replace the missing values with a value generated at random from the observed distribution of the variable. Finally, missing values may be replaced with imputed values based on the other characteristics of the record
- Data transformation: This may involve normalizing numeric variables in order to standardize the scale of effect each variable has on the results. For some EDM algorithms, the data may need to be transformed to achieve normality. In other situations, categorical variables may need to be replaced with indicator variables (a variable than can take only two values, 0 or 1). Sometimes, numerical values may need to be partitioned into bins or bands for certain algorithms. Two other transformation techniques are feature extraction and feature selection. In feature extraction, new attributes are produced by

transforming and combining the original ones. In feature selection, an optimal set of attributes is selected

Exploratory data analysis. When approaching an EDM problem, a data mining analyst may already have an a-priori hypothesis regarding the relationships between the variables. For example, a hypothesis may be that class attendance is a good predictor of a course's final grade (target variable). In this case, the analyst would use standard hypothesis testing procedures. However, the analyst may not have a hypothesis about the data. Particularly when confronted with unknown databases, an analyst may prefer to use Exploratory Data Analysis (EDA).

EDA allows the analyst to delve into the data set, examine the interrelationships among the attributes, identify interesting subsets of the observations, and develop an initial idea of possible associations among the predictors and the target variable. Graphs, plots, and tables often uncover important relationships that could indicate important areas for further investigation (Larose & Larose, 2015).

Data mining. Once the data has been properly pre-processed and the EDA has produced a hypothesis to be tested, then the analyst is ready to undertake the data mining proper. There are a number of methods that are part of the EDM toolkit, but the most important one to predict academic performance is classification, which will be explored below. Classification is the placing of an object into a class or category based on its other characteristics (Larose & Larose, 2015; Romero & Ventura, 2013).

Another method that is sometimes used when predicting academic performance is estimation. Estimation is similar to classification, except that, for estimation, we approximate the value of a numeric target variable (rather than a categorical target variable, as is the case for classification) using a set of numeric and/or categorical predictor variables. Because of the similarity between classification and estimation, estimation approaches will be discussed in conjunction with classification ones below.

Finally, two other methods that will be covered as well are association and clustering. Association is the process of finding correlations between variables in a data set. Clustering is aimed at discovering the natural grouping structure of data.

Interpretation of results. Once the data have been put through the appropriate EDM algorithms, it is important to do the following:

- We need to determine whether the project objectives and requirements have been met. For example, with the data available, if our hypothesis was that socio-economic status is a statistically significant predictor of whether a student will graduate, were we able to prove this
- If we were not able to verify our hypothesis, we need to determine whether this is a problem. Do we go back to the drawing board and formulate a different hypothesis? Perhaps the technical strategy or modeling techniques that we followed to prove our hypothesis were not sound? Perhaps it is a data quality problem? Or is there a significant facet of the research problem that we did not account for?
- We need to ask ourselves, whether we proved or disproved the hypothesis, if the results are nevertheless useful. Do they allow us to draw inferences or conclusions that otherwise we would not have been able to draw?
- We need to evaluate what other avenues of research open up for us now that we have obtained the results that we have.

Classification overview

In classification, we place a target variable into categories based on numeric or categorical predictor (or explanatory) variables, based on a classification model called a classifier. For example, if the target variable is whether a student passes a course or not, the classifier will classify students taking the course in question into the two categories, pass or fail, based on the appropriate explanatory variables such as grades from previous courses, midterm exam grade, and so forth.

The basic idea is a follows: first we need to choose a classification method, such as decision trees, Bayesian networks, or neural networks. Second, we need a sample of data where the values of the target variable are known for all subjects. The data is then divided into two parts, a training set and a test set. The training set is given to a learning algorithm, which will derive a classifier. The classifier is then tested with the test set, to see if the values of the target variable predicted by the classifier match the actual values of the target values for the subjects in the test set. If the classifier made too many mistakes, then we can change the settings of the learning algorithm, or use a different classification method (Hämäläinen & Vinni, 2011).

Typically, the learning task is an iterative process, where one has to try different data manipulations, classification approaches and algorithm settings before a good classifier is found. Here are a few considerations with regards to classification.

Classification accuracy. Classification accuracy is measured by the classification rate (cr), which defines the proportion of correctly classified rows in the data set. The classification error (err) is the proportion of misclassified rows in the data set. Thus, err = 1 – cr.

When the data set is the training set, the error is called the training error. If the data set has the same distribution as the entire population, the training error is a good estimate for the generalization error as well. This is seldom the case, however, in the educational domain. The training data sets are generally too small to capture the real distribution, and the classifier will therefore be biased. A common solution is to reserve a part of the data set as a test set (Hämäläinen & Vinni, 2011).

Overfitting. Usually, the accuracy of the provisional model or classifier is not as high on the test set as it is on the training set, often because the provisional model is overfitting on the training set. Overfitting results when the provisional model tries to account for every possible trend or structure in the training data set (Larose & Larose, 2015). For example, a data set that was collected for predicting student success in a programming course contained one female student, who had good IT skills and self-efficacy, and knew the idea of programming beforehand, but still dropped out of the course. The model would be overfitting on the data set if it concluded that all female students with IT skills and self-efficacy would drop out of the course (Hämäläinen & Vinni, 2011).

There is an eternal tension in model building between model complexity (resulting in high accuracy on the training set) and generalizability to the test and validation sets. Increasing the complexity of the model in order to increase the accuracy on the training set eventually and inevitably leads to a degradation in the generalizability of the provisional model to the test set (Larose & Larose, 2015).

Overfitting is a critical problem in the educational domain because there are many attributes available to construct a complex model, but only a little data to learn it accurately. There are two things that we can do: (1) use simple classification methods requiring fewer model parameters, and (2) reduce the number of attributes and their domains by feature selection and feature extraction.

The most commonly used feature extraction technique is discretization. Here, the range of numeric values is divided into intervals, which will be used as new attribute values. In extreme cases, all attributes can be binarized. Even if some information is lost, the resulting model can produce a more accurate classification. Generally, discretization smooths out the effect of noise and enables simpler models, which are less prone to overfitting.

Feature selection can be done by analyzing the dependencies between the target variable or attribute and the explanatory attributes, with techniques such as correlation analysis or association rules. For example, one can first learn a decision tree, and then use its attributes for a K-nearest neighbour classifier (Hämäläinen & Vinni, 2011).

Classification methods

This section introduces the most commonly used methods and algorithms for classification: decision trees, Bayesian classifiers, neural networks, nearest neighbor classifiers, and support vector machines. Linear regression, though not a classification but an estimation method because the target variable is numeric, will also be covered here.

Decision trees. Perhaps the most popular classification method involves the construction of a decision tree, a collection of decision nodes, connected by branches, extending downward from the root node until terminating in leaf nodes. Beginning at the root node, which by convention is placed at the top of the decision tree diagram, attributes are tested at the decision nodes, with each possible outcome resulting in a branch. Each branch then leads either to another decision node or to a terminating leaf node.

As an example, a recent study (Al-Barrak & Al-Razgan, 2016) used a decision tree to predict the final GPA from grades for all mandatory courses in their curriculum: Computer Architecture, Software Engineering 2, and Information Security. After running the algorithm on these courses, the Information Security Course became the root node; the next node down was Computer Architecture. Software Engineering 2 did not appear at all on the resulting decision tree. The classification result is shown in Fig. 1. This result tells us, for example, that if a student gets a C in Information Security and a B in Computer Architecture, his final GPA is predicted to be "Very good".

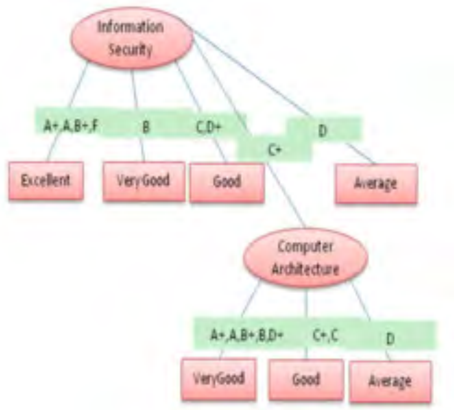

Fig. 1. *Classification result yielded by J48 decision tree to predict students' final GPA from grades in three mandatory courses (Al-Barrak & Al-Razgan, 2016)*

Decision trees have many advantages: they are easy to understand, they can handle both numeric and categorical variables, they can classify new examples quickly, and they are flexible. The main restriction of decision trees is the assumption that all records can be deterministically classified into exactly one class. As a result, all inconsistencies are interpreted as errors. Another problem is that decision trees are very sensitive to overfitting, especially in small data sets. Often overfitting can be avoided if we learn a collection of decision trees and then average their predictions. This approach is called model averaging or ensemble learning (Hämäläinen & Vinni, 2011).

Bayesian classifiers. In the field of statistics, there are two main approaches to probability. The usual way probability is taught in most typical introductory statistics courses, represents the frequentist or classical approach. In the frequentist approach to probability, the population parameters are fixed constants whose values are unknown. These probabilities are defined to be the relative frequencies of the various categories, where the experiment is repeated an indefinitely large number of times. For example, if we toss a fair coin 10 times, it may not be very unusual to observe 80% heads; but if we toss the fair coin 10 trillion times, we can be fairly certain that the proportion of heads will be near 50%. It is this long-run behavior that defines probability for the frequentist approach.

However, there are situations for which the classical definition of probability is unclear. For example, what is the probability that a student will flunk a given course? In the frequentist approach to probability, the parameters are fixed, and the randomness lies in the data, which are viewed as a random sample from a given distribution with unknown but fixed parameters.

The Bayesian approach to probability turns these assumptions around. In Bayesian statistics, the parameters are considered to be random variables, and the data are considered to be known. The parameters are regarded as coming from a distribution of possible values, and Bayesians look to the observed data to provide information on likely parameter values (Larose & Larose, 2015).

Bayesian classifiers are a classification technique based on Bayes' Theorem with an assumption of independence among predictors. The variant of this technique that is by far the most frequently used in EDM is the Naïve Bayes model. In simple terms, a Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple, and that is why it is known as 'Naïve' (Ray, 2015).

For an example of the use of the Naïve Bayes model as a classification technique, see Aziz et al. (2015).

Neural networks. Artificial neural networks are very popular in pattern recognition. Still, they can be problematic when applied to educational data, unless we have a lot of numeric data, and know exactly how to train the model.

Feed-forward neural networks (FFNNs) are the most widely used type of neural networks. The FFNN architecture consists of layers of nodes: one for input nodes, one for output nodes, and at least one for hidden nodes. The most general model contains just one hidden layer.

The learning algorithm is an essential part of the neural network model. Even if neural networks can represent any kind of classifiers, we are seldom able to learn the optimal model. The learning is computationally difficult, and the results depend on several open parameters such as the number of hidden layers, number of hidden nodes in each layer, initial weight, and the termination criterion. The selection of the network topology and the termination criterion are especially critical, because neural networks are very sensitive to overfitting. Unfortunately, there are no foolproof instructions, and the parameters have to be determined by trial-and-error.

FFNNs have several attractive features: they can easily learn nonlinear boundaries, and in principle they represent any kind of classifiers. Additionally, FFNNs are robust to noise, and can be updated with new data.

Their main disadvantage is that FFNNs need a lot of data, much more than what educational data sets typically contain. They are very sensitive to overfitting, and the problem is even more critical with small training sets. The data should be numeric, and categorical data must be somehow quantized before it can be used. This increases the model complexity, and the results are sensitive to the quantization method used.

The neural network model is a black box, and it is difficult for people to understand the explanations for the outcomes. Additionally, neural networks are unstable, and the expertise of the analyst is critical for success with this method (Hämäläinen & Vinni, 2011).

For an example of the use of neural networks, see Musso, Kyndt, Cascallar, and Dochy (2013).

K-nearest neighbor classifiers. K-nearest neighbor classifiers represent a different approach to classification. They do not build any explicit global model, but approximate it only locally and implicitly. The main idea is to classify a new object by examining the target variable's values (or categories) of the K most similar data points. The selected value can be either the most common value among the neighbors, or a value distribution in the neighborhood. The only learning task in K-nearest neighbor classifiers is to select two important parameters: the number of neighbors K, and the distance metric d.

The nearest neighborhood classifiers have several advantages: there are only two parameters to learn or select, the classification accuracy can be very good in some problems, and the method is quite tolerant of noise and missing values.

The main disadvantage is the difficulty to select the distance metric. The lack of an explicit model can be either an advantage or a disadvantage. On the plus side, there is no need to update the classifier when new data are added. On the other hand, if the data set is large, we need some index to find the nearest neighbors efficiently. Furthermore, an explicit model is useful for human evaluators and designers (Hämäläinen & Vinni, 2011).

For an example of the use of K-nearest neighbor classifiers, see Gray, McGuinness, & Owende (2014).

Support vector machines. Support Vector Machines (SVMs) are an ideal classification method when the class boundaries are nonlinear, but there is too little data to learn complex nonlinear models. The underlying idea is that when the data is mapped to a higher dimension, the classes become linearly separable.

The main advantage of SVMs is that they always find the global optimum. Another benefit is that the accuracy does not depend on the dimensionality of data. Moreover, the system is very robust to overfitting. This is an important advantage when the boundaries between the target variable's values are nonlinear. Most other classification methods produce, for nonlinear boundaries, models that are too complex.

SVMs, however, have the same restriction as neural networks: the data should be continuous numerical, or quantized; the model is not easily interpreted, and selecting the appropriate parameters can be difficult. Outliers can cause problems, because they are used to define the boundaries between the target variable's values (Hämäläinen & Vinni, 2011).

For an example of the use of SVMs, see Fernández-Delgado, Mucientes, Vázquez-Barreiros, & Lama (2015).

Linear regression. Linear regression, an estimation rather than a classification method, nevertheless works well as a classifier when all attributes are numeric. For example, if passing a course depends on a student's accumulated points, the points can be predicted by linear regression.

In linear regression, the assumption is made that the target variable (e.g., total points) is a linear function of other, mutually independent attributes. The model is very flexible and can work well, even if the actual dependency is only approximately linear or the other attributes are weakly correlated. The reason is that linear regression produces very simple models, which are not as prone to overfitting as more complex models. However, the data should not contain large gaps (empty areas), and should have few outliers (Hämäläinen & Vinni, 2011).

For an example of the use of linear regression, see Jacob et al. (2015).

Other methods that also classify

Data mining methods may be categorized as either supervised or unsupervised. In unsupervised methods, no target variable is identified as such. Instead, the data mining algorithm searches for patterns and structures among all the variables. The classification method that we dealt with in the previous two sections is an example of what is called a supervised method (Larose & Larose, 2015). The target variable was clearly defined as that which we wanted to predict, whether final GPA, course grades, or which students were at risk of dropping out of school.

Another approach to predicting academic success is through unsupervised methods. Through these unsupervised methods, such as clustering, association rule mining, and others, we gather student data of different types: demographic, academic, student activities and others, without specifying any particular target variable. A clustering algorithm, using a pre-determined number of clusters, say three, can then examine the data to see how students should be grouped according to the available data. A study (Harwati & Alfiani, 2015) found that the application of a clustering technique in the scenario just described led to three distinct classes that strongly correlated with the students' performance level. The three groups then were low-performing students, average students, and high-performing ones. For more details, see Harwati & Alfiani (2015).

Clustering then becomes a form of unsupervised classification.

Association rules mining is another popular data mining task. It discovers relationships between attributes in databases, producing if-then statements concerning attribute values (García, Romero, Ventura, de Castro, & Calders, 2011). Association rule mining can be applied in either a supervised or an unsupervised manner. In some scenarios, one may simply be interested in which attributes go together, in which case no target variable would be identified.

Some data sets, however, are naturally structured so that a particular variable fulfills the role of a consequent, and not an antecedent. For example, in the educational domain, we may have a course grade, and a prerequisite course's grade, among other attributes. In this case, association rules could be mined from this data set, where the prerequisite course's grade could represent a possible antecedent, and the course grade could represent the single consequent of interest. In this way, association rules could be used to help classify students' academic performance in a given course, in a supervised learning process (Larose & Larose, 2015). Thus, although association rules are generally used for unsupervised learning, they may also be applied for supervised learning for a classification task. For an example of the use of association rules mining as a classification technique, see Borkar & Rajeswari (2014).

Literature review

Analysis of primary-research articles

The following information was gathered from the 56 primary-research articles that we reviewed for this study:

- The target variable representing what was being predicted
- The predictor variables
- The main EDM method(s) and technique(s) that were used
- The data mining software that was used, if known (WEKA, SPSS, and so forth)

The resulting data are presented in table 3.

The types of target variables that were used in the studies listed in Table 3 are summarized in Table 4. A summary of the predictor variables used in these articles is presented in Table 5. The main methods and techniques used in these studies is shown in Table 6. Finally, the software used to run the various methods and techniques is presented in Table 7.

Table 3. Primary-research articles reviewed in this study

#	Author (year)	Target	Predictors	EDM methods & techniques	Software
1	Al-Barrak & Al-Razgan (2016)	Final GPA	Academic data	Classification: Decision tree	WEKA
2	Buniyamin, Mat, & Arshad (2016)	CGPA	Academic data	Classification: Neuro-fuzzy	Unknown
3	Guo, Zhang, Xu, Shi, & Yang (2016)	Admission exam grade	Academic & other data	Classification: Neural network	Unknown
4	Pruthi & Bhatia (2016)	Job placement	Academic data	Classification: Decision tree; clustering	Unknown
5	Ramanathan et al. (2016)	Job placement	Academic data	Clustering: CIFCM	Unknown
6	Ahmad, Ismail, & Aziz (2015)	GPA	Academic & other data	Classification: Decision tree & others	WEKA
7	Al-Barrak & Al-Razgan (2015)	Course grade	Partial grades	Classification: Naïve Bayes	WEKA
8	Al-Saleem, Al-Kathiry, Al-Osimi, & Badr (2015)	Course grade	Academic data	Classification: Decision tree	WEKA
9	Aziz et al. (2015)	GPA	Academic & other data	Classification: Naïve Bayes	Unknown
10	Fernández-Delgado et al. (2015)	Course grade	Partial grades	Classification: SVM	Unknown
11	Harwati & Alfiani (2015)	Course grade	Academic & other data	Clustering: K-means	SPSS
12	Iam-On & Boongoen (2015)	Drops out or not	Academic & other data	Clustering: Ensemble	Unknown
13	Jacob et al. (2015)	Pass/fail semester	Academic & other data	Classification: Decision tree	WEKA
14	B. Khan, Khiyal, & Khattak (2015)	GPA	Academic data	Classification: Decision tree	WEKA
15	Kolo et al. (2015)	Pass/fail course	Academic & other data	Classification: Decision tree	SPSS
16	López Guarín, Guzmán, & González (2015)	Loss of academic status	Academic & other data	Classification: Naïve Bayes & others	Unknown
17	Rubiano & García (2015)	GPA	Academic & other data	Classification: Decision tree	WEKA
18	Strecht, Cruz, Soares, Mendes-Moreira, & Abreu (2015)	Pass/fail course	Partial grades & other data	Classification: Decision tree & others	Unknown
19	Sweeney et al. (2015)	Course grade	Academic & other data	Matrix factorization	Unknown
20	Zhou, Zheng, & Mou (2015)	Pass/fail course	Academic & other data	Classification: Naïve Bayes	Unknown
21	Ahmed & Elaraby (2014)	Course grade	Partial grades & other data	Classification: Decision tree	Unknown
22	Arsad et al. (2014)	Graduation GPA	Academic data	Classification: Neural network	Unknown
23	Borkar & Rajeswari (2014)	Admission exam grade	Academic & other data	Association rule mining; Classification: Neural network	Unknown
24	Bydžovská & Brandejs (2014)	Student potential	Academic & other data	Association rule mining	Unknown
25	Bydžovská & Popelínský (2014)	Course grade	Academic & other data	Machine learning	WEKA
26	Durairaj & Vijitha (2014)	Course grade	Academic data	Clustering: K-Means; Classification: Naïve Bayes	WEKA
27	Gray et al. (2014)	Pass/fail year	Non-academic data	Classification: SVM & others	Unknown
28	I. A. Khan & Choi (2014)	Win scholarship	Academic & other data	Classification: Decision tree	Unknown
29	Mishra, Kumar, & Gupta (2014)	GPA	Academic & other data	Classification: Decision tree	WEKA, Rapidminer
30	Srečko Natek & Zwilling (2014)	Course grade	Academic & other data	Classification: Decision tree	Excel DM, WEKA
31	Slim, Heileman, Kozlick, & Abdallah (2014)	GPA	Academic data	Classification: Markov networks	Unknown
32	Tekin (2014)	Graduation GPA	Academic data	Classification: Neural networks	Unknown
33	Borkar & Rajeswari (2013)	Admission exam grade	Academic data	Association rule mining	WEKA
34	D. Chen & Elliott (2013)	First year's GPA	Academic data	Classification: Decision tree	SAS EM
35	Hoe et al. (2013)	CGPA	Academic & other data	Classification: CHAID	SPSS, PASW Mod
36	Huang & Fang (2013)	Course grade	Academic data	Linear regression & other models	Unknown
37	Khatwani & Arya (2013)	Pass/fail course	Academic & other data	Classification: Decision tree	Unknown
38	Kongsakun (2013)	Course grade	Academic data	Linear regression; Clustering: K-means	Unknown
39	Márquez-Vera, Cano, Romero, & Ventura (2013)	Pass/fail first year	Academic & other data	Classification: Genetic programming	WEKA
40	Mashiloane & Mchunu (2013)	Pass/fail first year	Academic data	Clasification: Decision tree	WEKA
41	Musso et al. (2013)	GPA	Non-academic data	Classification: Neural network	Unknown
42	Srecko Natek & Zwilling (2013)	Course grade	Academic & other data	Classification: Decision tree	Unknown
43	Ramanathan, Dhanda, & Suresh Kumar (2013)	Pass/fail course	Academic & other data	Classification: Decision tree & Naïve Bayes	WEKA
44	Ramesh, Parkavi, & Ramar (2013)	Course grade	Academic & other data	Classification: Naïve Bayes & others	Unknown
45	Alper & Çataltepe (2012)	Course grade	Academic data	Machine learning	Unknown
46	Bunkar, Singh, Pandya, & Bunkar (2012)	Course grade	Academic data	Classification: Decision tree	Unknown
47	Y. Y. Chen, Mohd Taib, & Che Nordin (2012)	Course grade	Academic & other data	Linear regression	Unknown
48	Drumond, Thai-Nghe, Horváth, & Schmidt-Thieme (2012)	Course grade	Academic data	Matrix factorization	Unknown
49	El-Halees (2012)	Course grade	Academic & other data	Association rule mining & other methods	Unknown
50	Kumar & Vijayalakshmi (2012)	Final GPA	Academic data	Classification: Decision tree & others	WEKA
51	Osmanbegović & Suljić (2012)	Admission exam grade	Academic & other data	Classification: Naïve Bayes & others	WEKA
52	Shovon & Haque (2012)	Course grade	Academic data	Clustering: K-means	Unknown
53	Bhardwaj & Pal (2011a)	Course grade	Academic & other data	Classification: Naïve Bayes	Unknown
54	Bhardwaj & Pal (2011b)	GPA	Academic data	Classification: Decision tree	Unknown
55	Kabra & Bichkar (2011)	Pass/fail first year	Academic & other data	Classification: Decision tree	WEKA
56	Sembiring, Zarlis, Hartama, Ramliana, & Wani (2011)	Final GPA	Academic & other data	Classification: SVM and other methods	Unknown

Table 4. Target variables used in the primary-research articles reviewed

#	Target variable	Count	Frequency %
1	Course grade	20	35.7%
2	Some form of GPA	16	28.6%
3	Pass/fail course, semester, year	10	17.9%
4	Admission exam grade	4	7.1%
5	Job placement	2	3.6%
6	Drops out or not	1	1.8%
7	Wins scholarship	1	1.8%
8	Loss of academic status	1	1.8%
9	Student potential	1	1.8%
	TOTAL	56	100%

Table 5. Predictor variables used in the primary-research articles reviewed

#	Predictor variable	Count	Frequency %
1	Academic & other data	29	51.8%
2	Academic data only	21	37.5%
3	Non-academic data	2	3.6%
4	Partial grades / other data	2	3.6%
5	Partial grades only	2	3.6%
	TOTAL	56	100%

Table 6. Main EDM method used in the primary-research articles reviewed

#	Main EDM method	Count	Frequency %
1	Classification	40	71.4%
2	Clustering	5	8.9%
3	Association rule mining	4	7.1%
4	Linear regression	3	5.4%
5	Machine learning	2	3.6%
6	Matrix factorization	2	3.6%
	TOTAL	56	100%

Table 7. Software used in the primary-research articles reviewed

#	Software	Count	Frequency %
1	WEKA only	16	28.6%
2	WEKA & other software	2	3.6%
3	SPSS alone or with other software	3	5.4%
4	SAS Enterprise Miner	1	1.8%
5	Unknown	34	60.7%
	TOTAL	56	100%

Analysis of review articles

The five review articles listed in Table 2 were examined. All of these articles surveyed the entire EDM field. Instead of just looking into predicting academic performance, they also examined articles dealing with other EDM goals (Romero et al., 2011b), such as:

- Communicating to stakeholders, such as administrators and educators
- Maintaining and improving courses
- Student modeling, i.e., detecting students' states and characteristics such as motivation or learning progress, or certain types of problems such as gaming the system or inefficient exploration or use of resources

Moreover, instead of just looking at the traditional classroom environment, these review articles explored other learning environments and modalities, such as:

- Learning management systems
- Intelligent tutoring systems
- Adaptive and intelligent hypermedia systems
- Test and quiz systems

The key findings of these review articles are listed in Table 8. This table also includes the number of studies surveyed by the review article that related to predicting academic performance, where it was possible to glean this information. It was not possible to isolate either the number of articles or the key findings for a traditional classroom environment only from these review articles.

Conclusions

Predicting students' academic performance continues to be one of the most popular goals for EDM. Judging from the number of primary-research studies on the topic listed for the review articles we surveyed, the trend line shows that the number of articles devoted to this subject has steadily increased. From a local perspective, our anecdotal evidence is that this is a topic of great interest to the academic administrators at the Universidad Técnica del Norte and other Ecuadorian universities.

The EDM methods and techniques that we have explored in this paper, when

Table 8. Review articles and their key findings

#	Author(s) and year	Article count	Key findings in the area of predicting academic performance
1	Thakar et al. (2015)	40	• Classification is the top method for performance prediction • Among classification techniques, decision trees outperformed Bayesian networks • Regarding soft skills, emotional intelligence, self-management, work-life experience are important for employability; academic competence is a weaker predictor • Mismatch between employers' skill requirements and what academia delivers
2	Shahiri et al. (2015)	30	• Most of the researchers have used CGPA and partial grades as predictor attributes • Classification method is the most frequently used EDM method • Among classification techniques, neural networks and decision trees are the two most popular methods to predict performance • Neural networks have the highest prediction accuracy (98%), followed by decision tree (91%), support vector machines (83%), and naïve Bayes (73%)
3	Al-Razgan et al. (2014)	41	• Of all the 11 EDM goals explored by this paper, predicting student performance was addressed by the largest number of articles
4	Peña-Ayala (2014)	46	• Student modeling oriented to represent and anticipate performance is one of the favorite targets of EDM studies • Most of the EDM studies represent the implementation of data mining to explore educational subjects, instead of contributing to extending the data mining field • Most of the EDM studies only apply a small portion of the huge repertoire of data mining tools and techniques
5	Romero et al. (2011a)	15	• The articles on predicting academic performance only looked at the classification method • The three most popular techniques were decision trees, Bayesian networks and neural networks • The accuracy was 72-79% on average; for the best cases, accuracy was 90-94% • The best results are achieved when classifiers can be learned from real data, but in the educational domain the data sets are often too small for accurate learning

properly applied, can achieve high rates of success. In particular, from the analysis of the papers we have reviewed, we conclude that the researchers have been able to successfully predict academic performance using a broad swath of:

- Target variables: various forms of GPA, course grade, pass/fail determinations, and even job placements, winning scholarships, or a student's potential
- Predictor variables: academic and non-academic data, although the preponderance of success occurs with academic-type data, whether alone or in combination with other factors
- EDM methods: although classification has been used in the lion's share of studies, other methods such as clustering, association rule mining, linear regression, machine learning, and matrix factorization have also been successfully tried
- Software packages: WEKA (https://goo.gl/cnqEGq) is clearly the software package of choice; yet, a large minority of researchers did not identify the software they used, without giving a reason. Three of the tools that did get mentioned (WEKA, SPSS, RapidMiner) are the three top tools in terms of functionality mentioned by Slater, Joksimovi, Kovanovic, Baker, and Gasevic (2016).

Nevertheless, these successes notwithstanding, EDM, and data mining in general, are easy to do badly. Researchers may apply inappropriate analyses to data sets that call for a completely different approach, for example, or models may be derived that are built on tenuous assumptions. Therefore, an understanding of the statistical and mathematical model structures underlying the software is required.

Unfortunately, there will continue to be a shortage of talent necessary for organizations to take advantage of the large volumes of data they have collected over the years (Gershkoff, 2015). People with deep expertise in statistics and machine learning are particularly expected to be in short supply. Another caveat is the tendency to rely on software as a substitute for human oversight. Humans need to be involved at every phase of the data mining process. The very power of the formidable data mining algorithms embedded in software packages such as WEKA makes their misuse proportionately more dangerous (Larose & Larose, 2015).

The studies that we have reviewed

represent the experiences of researchers from a broad cross-section of nationalities and perspectives, and in all cases the methods and techniques that we have described have reliably yielded positive results. The expectation then will be that as future studies attempt to replicate in Ecuador what has been done abroad, the high prediction success rates that have been attained elsewhere will also materialize here.

BIBLIOGRAPHIC REFERENCES

Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. Applied Mathematical Sciences, 9(129), 6415–6426. http://doi.org/10.12988/ams.2015.53289

Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. World Journal of Computer Application and Technology, 2(2), 43–47. http://doi.org/10.13189/wjcat.2014.020203

Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting students final GPA using decision trees: A case study. International Journal of Information and Education Technology, 6(7), 528–533. http://doi.org/http://dx.doi.org/10.7763/IJIET.2016.V6.745

Al-Barrak, M. A., & Al-Razgan, M. S. (2015). Predicting students' performance through classification: A case study. Journal of Theoretical and Applied Information Technology, 75(2), 167–175.

Al-Razgan, M. S., Al-Khalifa, A. S., & Al-Khalifa, H. S. (2014). Educational data mining: A systematic review of the published literature 2006-2013. In 1st International Conference on Advanced Data and Information Engineering, DaEng 2013 (Vol. 285 LNEE, pp. 711–719). Kuala Lumpur: Springer Verlag. http://doi.org/10.1007/978-981-4585-18-7_80

Al-Saleem, M., Al-Kathiry, N., Al-Osimi, S., & Badr, G. (2015). Mining educational data to predict students' academic performance. In 11th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM 2015 (Vol. 9166, pp. 403–414). Hamburg: Springer Verlag. http://doi.org/10.1007/978-3-319-21024-7_28

Alper, M. E., & Çataltepe, Z. (2012). Improving course success prediction using ABET course outcomes and grades. In 4th International Conference on Computer Supported Education, CSEDU 2012 (Vol. 2, pp. 222–229). Porto.

Arsad, P. M., Buniyamin, N., & Manan, J. A. (2014). Neural network and linear regression methods for prediction of students' academic achievement. In IEEE 2014 Global Engineering Education Conference (EDUCON) (pp. 916–921). Istanbul: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/EDUCON.2014.6826206

Aziz, A. A., Ismail, N. H., Ahmad, F., & Hassan, H. (2015). A framework for students' academic performance analysis using naïve bayes classifier. Jurnal Teknologi, 75(3), 13–19. http://doi.org/10.11113/jt.v75.5037

Bhardwaj, B. K., & Pal, S. (2011a). Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security (IJCSIS), 9(4).

Bhardwaj, B. K., & Pal, S. (2011b). Mining educational data to analyze students' performance. International Journal of Advanced Computer Science and Applications (IJACSA), 2(6), 63–69. Retrieved from www.ijacsa.thesai.org

Bolaji, A. (2015). A cross-disciplinary systematic literature review on Kanban. University of Oulu. Retrieved from https://goo.gl/Q9CTbp

Borkar, S., & Rajeswari, K. (2013). Predicting students academic performance using education data mining. International Journal of Computer Science and Mobile Computing, 2(7), 273–279.

Borkar, S., & Rajeswari, K. (2014). Attributes selection for predicting students' academic performance using education data mining and artificial neural network. International Journal of Computer Applications, 86(10), 25–29.

Buniyamin, N., Mat, U. Bin, & Arshad, P. M. (2016). Educational data mining for prediction and classification of engineering students achievement. In IEEE 7th International Conference on Engineering Education, ICEED 2015 (pp. 49–53). Kanazawa: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICEED.2015.7451491

Bunkar, K., Singh, U. K., Pandya, B., & Bunkar, R. (2012). Data mining: Prediction for performance improvement of graduate students using classification. In 9th IEEE and IFIP International Conference on Wireless and Optical Communications Networks, WOCN 2012. Indore: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/WOCN.2012.6335530

Bydžovská, H., & Brandejs, M. (2014). Towards student success prediction. In 6th International Conference on Knowledge Discovery and Information Retrieval, KDIR 2014 (pp. 162–169). Rome: INSTICC Press.

Bydžovská, H., & Popelínský, L. (2014). Weak student identification: How technology can help. In L. K.T. & O. R. (Eds.), 13th European Conference on e-Learning, ECEL (2014) (pp. 89–97). Copenhagen: Academic Conferences and Publishing International Limited.

CES. Reglamento de Régimen Académico - Consejo de Educación Superior, Pub. L. No. RPC-SE-13-No.051-2013 (2016). Ecuador. Retrieved from https://goo.gl/szgYGO

Chen, D., & Elliott, G. (2013). Determining key (predictor) course modules for early identification of students at-risk. In 2013 International Conference on Advanced Information Engineering and Education Science, ICAIEES 2013 (pp. 74–79). Beijing: Atlantis Press.

Chen, Y. Y., Mohd Taib, S., & Che Nordin, C. S. (2012). Determinants of student performance in advanced programming course. In 7th International Conference for Internet Technology and Secured Transactions, ICITST 2012 (pp. 304–307). London: Institute of Electrical and Electronics Engineers Inc.

Drumond, L., Thai-Nghe, N., Horváth, T., & Schmidt-Thieme, L. (2012). Factorization techniques for student performance classification and ranking. In 20th Conference on User Modeling, Adaptation, and Personalization, UMAP 2012 (Vol. 872, pp. 1–6). Montreal: Springer.

Durairaj, M., & Vijitha, C. (2014). Educational data mining for prediction of student performance using clustering algorithms. International Journal of Computer Science and Information Technologies (IJCSIT), 5(4), 5987–5991.

El-Halees, A. M. (2012). Mining educational data to improve students' performance: A case study. International Journal of Information and Communication Technology Research, 2(2), 140–146.

Fernández-Delgado, M., Mucientes, M., Vázquez-Barreiros, B., & Lama, M. (2015). Learning analytics for the prediction of the educational objectives achievement. In 44th Annual Frontiers in Education Conference, FIE 2014. Madrid: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/FIE.2014.7044402

García, E., Romero, C., Ventura, S., de Castro, C., & Calders, T. (2011). Association rule mining in learning management systems. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d Baker (Eds.), Handbook of Educational Data Mining (1st ed.). CRC Press.

Gershkoff, A. (2015). How to stem the global shortage of data scientists » TechCrunch. Retrieved from https://goo.gl/XoSCLz

Gray, G., McGuinness, C., & Owende, P. (2014). An application of classification models to predict learner progression in tertiary education. In 2014 4th IEEE International Advance Computing Conference, IACC 2014 (pp. 549–554). Gurgaon: IEEE Computer Society. http://doi.org/10.1109/IAdCC.2014.6779384

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2016). Predicting students performance in educational data mining. In International Symposium on Educational Technology, ISET 2015 (pp. 125–128). Wuhan: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ISET.2015.33

Hämäläinen, W., & Vinni, M. (2011). Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d Baker (Eds.), Handbook of Educational Data Mining (1st ed.). CRC Press.

Harwati, A. P., & Alfiani, F. A. W. (2015). Mapping student's performance based on data mining approach (A case study). Agriculture and Agricultural Science Procedia, 3, 173–177. http://doi.org/10.1016/j.aaspro.2015.01.034

Hoe, A. C. K., Ahmad, M. S., Hooi, T. C., Shanmugam, M., Gunasekaran, S. S., Cob, Z. C., & Ramasamy, A. (2013). Analyzing students records to identify patterns of students' performance. In International Conference on Research and Innovation in Information Systems, ICRIIS (pp. 544–547). Kuala Lumpur: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICRIIS.2013.6716767

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers and Education, 61, 133–145. http://doi.org/10.1016/j.compedu.2012.08.015

Iam-On, N., & Boongoen, T. (2015). Improved student dropout prediction in Thai university using ensemble of mixed-type data clusterings. International Journal of Machine Learning and Cybernetics. http://doi.org/10.1007/s13042-015-0341-x

IESALC. (2006). Informe sobre la educación superior en América Latina y el Caribe 2000-2005 por el Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (1st ed.). Caracas: Editorial Metrópolis C.A.

Jacob, J., Jha, K., Kotak, P., & Puthran, S. (2015). Educational data mining techniques and their applications. In 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015 (pp. 1344–1348). Delhi: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICGCIoT.2015.7380675

Kabra, R. R., & Bichkar, R. S. (2011). Performance prediction of engineering students using decision trees. International Journal of Computer Applications, 36(11), 8–12.

Khan, B., Khiyal, M. S. H., & Khattak, D. M. (2015). Final grade prediction of secondary school student using decision tree. International Journal of Computer Applications, 115(21), 32–36.

Khan, I. A., & Choi, J. T. (2014). An application of educational data mining (EDM) technique for scholarship prediction. International Journal of Software Engineering and Its Applications, 8(12), 31–42. http://doi.org/10.14257/ijseia.2014.8.12.03

Khatwani, S., & Arya, A. (2013). A novel framework for envisaging a learner's performance using decision trees and genetic algorithm. In 3rd International Conference on Computer Communication and Informatics, ICCCI 2013. Coimbatore: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICCCI.2013.6466309

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering EBSE Technical Report EBSE-2007-01. Keele, Staffs, and Durham, UK.

Kolo, K. D., Adepoju, S. A., & Alhassan, J. K. (2015). A decision tree approach for predicting students academic performance. International Journal of Education and Management Engineering, 5, 12–19. http://doi.org/10.5815/ijeme.2015.05.02

Kongsakun, K. (2013). An improved recommendation model using linear regression and clustering for a private university in Thailand. In International Conference on Machine Learning and Cybernetics (pp. 1625–1630). Tianjin: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICMLC.2013.6890859

Kumar, S. A., & Vijayalakshmi, M. N. (2012). Mining of student academic evaluation records in higher education. In 2012 International Conference on Recent Advances in Computing and Software Systems, RACSS 2012 (pp. 67–70). Chennai: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/RACSS.2012.6212699

Larose, D. T., & Larose, C. D. (2015). Data mining and predictive analytics (1st ed.). John Wiley & Sons, Inc.

López Guarín, C. E., Guzmán, E. L., & González, F. A. (2015). A model to predict low academic performance at a specific enrollment using data mining. Revista Iberoamericana de Tecnologías Del Aprendizaje, 10(3), 119–125. http://doi.org/10.1109/RITA.2015.2452632

Márquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence, 38(3), 315–330. http://doi.org/10.1007/s10489-012-

Mashiloane, L., & Mchunu, M. (2013). Mining for marks: A comparison of classification algorithms when predicting academic performance to identify "students at risk." In R. Prasath & T. Kathirvalavakumar (Eds.), First International Conference, Mining Intelligence and Knowledge Exploration (MIKE) 2013 (Vol. Lecture No, pp. 541–552). Tamil Nadu: Springer International Publishing. http://doi.org/10.1007/978-3-319-03844-5_54

Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for performance prediction. In International Conference on Advanced Computing and Communication Technologies, ACCT (pp. 255–262). Rohtak: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ACCT.2014.105

Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. Frontline Learning Research, 1, 42–71. http://doi.org/http://dx.doi.org/10.14786/flr.v1i1.13

Natek, S., & Zwilling, M. (2013). Data mining for small student data set – Knowledge management system for higher education teachers. In Management, Knowledge and Learning International Conference 2013 (pp. 1379–1389). Zadar.

Natek, S., & Zwilling, M. (2014). Student data mining solution-knowledge management system related to higher education institutions. Expert Systems with Applications, 41, 6400–6407. http://doi.org/10.1016/j.eswa.2014.04.024

Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. Economic Review – Journal of Economics and Business, X(1), 3–12.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications, 41(4 PART 1), 1432–1462. http://doi.org/10.1016/j.eswa.2013.08.042

Pruthi, K., & Bhatia, P. (2016). Application of data mining in predicting placement of students. In 1st International Conference on Green Computing and Internet of Things, ICGCIoT 2015 (pp. 528–533). Delhi: Institute of

Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICGCIoT.2015.7380521

Ramanathan, L., Dhanda, S., & Suresh Kumar, D. (2013). Predicting students' performance using modified ID3 algorithm. International Journal of Engineering and Technology, 5(3), 2491–2497.

Ramanathan, L., Geetha, A., Khalid, M., & Swarnalatha, P. (2016). Apply of sum of difference method to predict placement of students' using educational data mining. In 3rd International Conference on Information Systems Design and Intelligent Applications, INDIA 2016 (Vol. 433, pp. 367–377). Visakhapatnam: Springer Verlag. http://doi.org/10.1007/978-81-322-2755-7_39

Ramesh, V., Parkavi, P., & Ramar, K. (2013). Predicting student performance: A statistical and data mining approach. International Journal of Computer Applications, 63(8), 35–39.

Ray, S. (2015). 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python). Retrieved from https://goo.gl/EWd7sx

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12–27. http://doi.org/10.1002/widm.1075

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d (Eds.). (2011a). Handbook of Educational Data Mining (1st ed.). CRC Press.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (2011b). Introduction. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d Baker (Eds.), Handbook of Educational Data Mining (1st ed., pp. 1–6). CRC Press.

Rubiano, S. M. M., & García, J. A. D. (2015). Formulation of a predictive model for academic performance based on students' academic and demographic data. In IEEE 2015 Frontiers in Education Conference, FIE 2015. El Paso: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/FIE.2015.7344047

Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011). Prediction of student academic performance by an application of data mining techniques. In International Conference on Management and Artificial Intelligence (IPEDR) 2011 (Vol. 6, pp. 110–114). Bali: Institute of Electrical and Electronics Engineers Inc.

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. In 3rd Information Systems International Conference, 2015 (Vol. 72, pp. 414–422). Shenzhen: Elsevier. http://doi.org/10.1016/j.procs.2015.12.157

Sheard, J. (2011). Basics of statistical analysis of interactions data from web-based learning environments. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d Baker (Eds.), Handbook of Educational Data Mining (1st ed.). CRC Press.

Shovon, M. H. I., & Haque, M. (2012). Prediction of student academic performance by an application of k-means clustering algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, 2(7), 353–355.

Slater, S., Joksimovi, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2016). Tools for educational data mining: A review. Journal of Educational and Behavioral Statistics, XX(X), 1–22. http://doi.org/10.3102/1076998616666808

Slim, A., Heileman, G. L., Kozlick, J., & Abdallah, C. T. (2014). Employing Markov networks on curriculum graphs to predict student performance. In 13th International Conference on Machine Learning and Applications, ICMLA 2014 (pp. 415–418). Detroit: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICMLA.2014.74

Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A comparative study of classification and regression algorithms for modelling students' academic performance. In 8th International Conference on Educational Data Mining 2015 (pp. 392–395). Madrid.

Sweeney, M., Lester, J., & Rangwala, H. (2015). Next-term student grade prediction. In IEEE International Conference on Big Data, IEEE Big Data 2015 (pp. 970–975). Santa Clara: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/BigData.2015.7363847

Tekin, A. (2014). Early prediction of students' grade point averages at graduation: A data mining approach. Eurasian Journal of Educational Research, (54), 207–226.

Thakar, P., Mehta, A., & Manisha. (2015). Performance analysis and prediction in educational data mining: A research travelogue. International Journal of Computer Applications, 110(15), 60–68.

WCHE. (1998). World declaration on higher education for the twenty-first century: Vision and action. Retrieved from https://goo.gl/B0UTSL

Zhou, Q., Zheng, Y., & Mou, C. (2015). Predicting students' performance of an offline course from their online behaviors. In 5th International Conference on Digital Information and Communication Technology and Its Applications, DICTAP 2015 (pp. 70–73). Beirut: Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/DICTAP.2015.7113173.

Recibido para revisión: 11 octubre 2016
Aceptado para publicación: 28 noviembre 2016