Dropout prediction in e-learning courses through the combination of machine learning techniques

Abstract

In this paper, a dropout prediction method for e-learning courses, based on three popular machine learning techniques and detailed student data, is proposed. The machine learning techniques used are feed-forward neural networks, support vector machines and probabilistic ensemble simplified fuzzy ARTMAP. Since a single technique may fail to accurately classify some e-learning students, whereas another may succeed, three decision schemes, which combine in different ways the results of the three machine learning techniques, were also tested. The method was examined in terms of overall accuracy, sensitivity and precision and its results were found to be significantly better than those reported in relevant literature.

Keywords: distance education and telelearning, e-learning, machine learning, dropout prediction

1. Introduction

During the last few years, e-learning has presented significant growth and is attracting an increasing number of participants. Contemporary web-based courses take advantage of internet capabilities to support and improve the effective approaches of traditional education, while at the same time they offer great innovative possibilities. In addition to assisting individual learners to acquire knowledge in a flexible manner, without the barrier of their physical presence in a classroom, this new form of education also benefits institutions that now have the opportunity to provide more students with their educational services, reducing at the same time their expenditure (Cohen & Nachmias, 2006; Mendling, Neumann, Pinterits & Simon, 2005; Simpson, 2004). The nature of e-learning courses has also attracted a significant number of adult learners who seek to combine higher education and technical level training with demanding work responsibilities (Holder, 2007).

Despite the fact that they appeal to a great number of students, web-based courses demonstrate higher dropout rates than traditional education courses, often surpassing the latter by a 10-20% rate as reported by Carr (2000) and Diaz (2002). Additional studies confirm the above reports and estimate e-learning dropout rates to be 25-40% compared to the 10-20% rates of on-campus courses (Carter, 1996; Doherty, 2006; Frankola, 2001; Parker, 2003; Xenos, 2004).

Nevertheless, student retention rates are among the indicators that universities, policymakers, higher education funding bodies and educators consider as an objective, outcome-based measure of the quality that an educational institution offers. This quality indicator has been recognized and used at international levels including Australia, the European Union, the United States of America and South Africa (Blom & Meyers, 2003). For instance, it has been reported that the US Department of Education is placing increasing significance on higher education retention issues as part of its quality assessment and funding of higher education institutions (Borrego, 2002). Moreover, Doherty (2006) reports that the United States Commission on the Future of Higher Education verified that college and university retention rates were analyzed as part of a review on federal funding. Retention performance indicators are also taken into consideration by the Higher Education Funding Council of the United Kingdom as a means to provide reliable information on the performance of the UK higher educational sector, to allow comparison among individual institutions, to facilitate them to benchmark their own performance and to contribute to the public accountability of higher education (HEFCE, 2007). Student retention rates are also reported to be among the criteria for the evaluation of distance education according to the Council for Regional Accrediting Commissions (Terkla, 2001) and the Higher Learning Commission (HLC, 2007).

This increased emphasis on retention, combined with the high dropout rates that e-learning courses present, makes the reduction of dropout rates fundamental to the acceptance and success of this type of course. One of the key elements in reducing e-learning dropout rates is the accurate and prompt identification of at-risk students. As soon as these students have been identified, instructors will be able to better address their specific needs and take the appropriate actions to reduce their probability to drop the course. The identification of students prone to dropout is also expected to aid the specialized intervention programs that educational institutions incorporate in their strategic plans in order to increase retention.

This paper presents a dropout prediction method to accurately identify dropout-prone students in early stages of the e-learning course. The proposed method takes advantage of the fully electronic characteristics of e-learning courses, which include detailed student activity and progress data, to make its predictions dynamically and adapt them according to the performance and participation levels of each student throughout the course. These data are used to train three machine learning techniques, namely feed-forward neural networks (FFNNs), support vector machines (SVMs) and probabilistic ensemble simplified fuzzy ARTMAP (PESFAM). Then, since a single machine learning technique may fail to predict a number of dropout students, the results of these techniques are combined using three decision schemes. The estimations produced by each machine learning technique as well as those produced by each decision scheme were compared in terms of overall accuracy, sensitivity and precision. Experimental results indicate that combining the outputs of the three techniques leads to a more accurate and prompt identification of dropout students, surpassing all relevant results reported in current literature.

The rest of this paper is structured as follows: Section 2, referring to the materials and methods of this study, firstly presents the findings of related research literature. Then, it provides an introduction to the theoretical basis of the three machine learning techniques used, as well as a description of the three decision schemes examined. Next, a description of the e-learning courses data set upon which the method was applied as well as the details of the specific algorithm implementations used, are provided. The experimental results of the study according to the three performance evaluation criteria used, are reported in section 3. In section 4 the experimental results are discussed and their limitations are presented. Finally, section 5 concludes with the main findings and future work of this study.

2. Materials and Methods

2.1 Related Literature

Several studies focus on the field of dropout prediction in high school, university and online learning levels. These studies can be divided into two categories, according to the type of data they use. The first category includes a number of studies which use time-invariant characteristics to determine the most important variables that lead to the prediction of the dropout status of students. In the second category, there are studies that incorporate time-varying student attributes, which change as the course progresses, to make predictions over student dropout or retention.

2.1.1 Studies using time-invariant student attributes

Firstly, a variety of studies use past time-invariant student data as predictors, to retrieve the most important attributes related to a student's decision to withdraw from a course. The predictors identified by the aforementioned studies are three-fold: related to the demographic characteristics of the students, to their prior academic performance and to other factors that do not belong to the two aforementioned categories. It should also be noted here that studies of this type examine the correlation of the predictors with the student's decision to drop the course independently for each predictor, and their scope does not usually include reporting overall classification rates which would result from the combinative use of the predictors.

A number of these studies use logistic regression analysis to construct and test the hypothesized prediction models. Roblyer, Davis, Mills, Marshall & Pape (2008) use a sixty-item Likert scale instrument to gather data on student characteristics regarding the prediction of student retention and dropout in a virtual school environment. The variables gathered refer to a large student population of 2,162 students and are evaluated through a binary logistic regression analysis. This study reports an overall correct classification rate of 79.3%, which corresponds to the best combination of predictors identified. The predictors which refer to the students' demographic characteristics are age and home computer availability, those that refer to the students' prior academic performance is the self-reported Grade Point Average (GPA) and the predictor that refers to other student-related factors is the school period for working on the virtual course. Although the model examined by the aforementioned study performs well in predicting the overall classification rate, which includes both completers and dropouts, it only achieves a 30.4% accuracy rate in correctly categorizing the dropout students only.

The identification of age as a predictor of dropout likelihood is also supported by the study of Newell (2007). This study uses logistic regression along with Chi-square and bivariate analysis to retrieve student-related predictor variables based on data derived from a very large population of 89,473 students enrolled in online courses, offered by the Georgia Virtual Technical College. Apart from age, the above study also identifies a variety of other demographic elements which include gender, ethnicity and financial aid eligibility, as predictors of successful course completion. Supporting that demographic characteristics can be used as indicative predictors of retention, Zhang, Anderson, Ohland, Carter & Thorndyke (2003) use students' pre-existing demographic and academic data along with multiple logistic regression to make predictions upon the retention of a large population of 57,549 engineering students located in 9 universities. The results of this study indicate that the most significant demographic characteristics to be used as predictors are gender, ethnicity and citizenship. Apart from these elements, the study also concludes that a number of other predictors, which are related to the students' prior academic performance and include the SAT (Scholastic Aptitude Test) math scores, SAT verbal scores (SAT-V) and high school GPA, also play a significant role in students' retention rates. SAT refers to a standardized test used for college admissions in the United States.

Multiple regression analysis is also used by Vare, Dewalt & Dockery (2000) along with various past student data, to also support that prior academic performance is a significant predictor of retention. This study uses student attributes which include high school Grade Point

Ratio (GPR), SAT scores, scores on McCarty's Learning Type Measure (LTM) and other demographic information to retrieve the most significant variables in predicting retention rates in teacher education. This study is based on a data set that consists of 316 students and suggests that the strongest demographic predictors are the level of father's education, while the most significant predictors that are related to the students' prior academic performance are the SAT-V score, the high school GPR, and the thinking/reflection score on the LTM. Another study which focuses on the prior academic performance of the students in order to identify dropout predictors is the study of Allen & Robbins (2008) who use hierarchical logistic regression to test a theoretical model to predict college major persistence over a large data set of 50,000 students. The constructed model comprises three main variables, namely the students' vocational interests, their academic preparation and first-year academic performance. This study also concludes that prior academic performance and interest-major correspondence are critical elements in predicting major persistence.

A study that uses logistic regression but focuses on a different student variable, not related to demographic characteristics or prior academic performance, is the study of Wegner, Flisher, Chikobvu, Lombard & King (2008). In this study, the effectiveness of the attribute of leisure boredom is investigated as a predictor of high school dropout on a population of 281 students. The study concludes that leisure boredom is a significant factor of predicting high school dropout likelihood especially among students aged 14 years and older.

Apart from logistic regression, a variety of alternative methods have also been used along with time-invariant student data to infer the most important predictors of dropping out. Xenos, Pierrakeas & Pintelas (2002) use multivariate methods as well as correlation analysis to examine various student characteristics in relation to student dropout in a technologically oriented university course. The data set of this study is composed of data regarding 1,230 students. This study concludes that dropouts are correlated to a demographic attribute, namely the age of the students, and to attributes related to their prior academic performance, specifically their level of use of technological means and previous experience with computers, as well as their prior education in the field of Informatics. The aforementioned study also reports that the students' gender and degree of specialization in computers were not found to be correlated with their decision to drop the course.

Another method, namely discriminant analysis, is used by Dupin-Bryant (2004), over a survey completed by 464 students, to determine the strongest pre-entry variables to be used in the prediction of learner completion and non-completion in university-level online distance education courses. This study agrees with the previously mentioned study of Xenos, et al. (2002) and reports that prior academic performance in terms of past educational experience and previous computer training are the most helpful criteria in distinguishing between completers and dropouts. Predictive discriminant analysis is also used by Morris, Wu & Finnegan (2005) to develop a classification rule to predict undergraduate students' withdrawal from online courses based on data from 11 student cases. The study reports that a correct classification ratio of 74.5% was achieved using the attributes of financial aid and locus of control. It is also concluded that certain attributes related to prior academic performance, namely the high school GPA and SAT mathematics score can be used as predictors of retention.

Another study which supports prior educational experience as an attribute related to retention is the study of Mendez, Buskirk, Lohr & Haag (2008) that use classification trees and random forests to identify important predictors of persistence in Science and Engineering majors with a relatively large dataset of 2,232 student records. The study concludes that high school and freshman year GPAs are of the highest importance when predicting student persistence. In addition, the method used by this study presents promising results since it offers a better distinction of the important variables than the typically used method of logistic regression. Apart from identifying predictors related to demographic characteristics and prior academic performance, a number of studies focus on other student attributes. Challenging behaviour is examined, along with academic achievement, year level, socioeconomic and family structure variables, by Boon (2008). In this study, structural equation path models are used to predict the dropout status of 1,050 students in North Queensland urban high schools. The results of this study show that challenging behaviour is a stronger predictor than the other variables tested. The satisfaction with the course as well as the academic locus of control that e-learning students demonstrate are examined by Levy (2007). This study uses time-invariant student data to examine whether the two aforementioned attributes affect the decision of students to drop a course. The study is based on a data set of 133 students and the data obtained are analyzed using one-way ANOVA and non-parametric tests. It is concluded that the students' satisfaction with the course plays a significant role in their decision to drop out, while academic locus of control does not have an impact on this decision.

Finally, unlike other studies that mainly use student data, Ukpabi (2004) investigates the effect that institutional-related variables have on student retention. This study uses the techniques of pooled cross-sectional time series and ordinary least squares to formulate a predictive model of retention rates regarding 150 students from 15 institutions of the University of North Carolina. The study concludes that the four most influential factors are headcount enrolment, amount of education and general expenditure on instruction and academic support, the population of the county where the institution is located and the rate of unemployment in the county.

Thus, summarizing, the most frequently reported demographic-related predictors identified include the age, gender and ethnicity of the student, as well as the financial aid received. The predictors related to the prior academic performance of the student include GPA and SAT scores in addition to the student's previous training on technological means and computers, and the relevance of their past educational experience to the course under study. Other factors that do not belong to these two groups are also reported.

2.1.2 Studies using time-varying student attributes

Apart from using only past student data to predict dropout status, a number of studies also use time-varying characteristics, which are gathered as the course progresses, to predict student persistence and dropping out. This type of studies is mainly based on machine learning techniques. Bayesian networks are used by Xenos (2004) to model past educational experience regarding university-level student behaviour, using both past student data and data related to the students' progress through the academic year. The data set used consists of approximately 800 students. The model presented in this study aims in facilitating instructors in their decision making regarding the educational procedure and as part of this, help them identify the causes or the indicators that lead students to drop the course. This study does not report overall classification accuracy results.

Another technique, namely decision trees, is used by Moseley & Mead (2008), on a data set of 3,978 students, for the deduction of rules regarding student attrition in higher educational nursing institutions. This study is based on both time-invariant and time-varying student data related to the students' performance during the course. Taking into consideration only the time-invariant characteristics in the form of demographic data, this study reports that the amount of dropout students correctly predicted was 31%. The proportion of students that the system classified as dropouts and eventually did drop the course was 44%. However, using the time-varying student data obtained during the course, the aforementioned success ratios reached 84% and 70% respectively. The study does not report an overall accuracy rate when time-invariant data are used, but the respective reported rate yielded by the use of time-varying student data reaches 94%. Although the authors do not report whether the prediction is made in an early or late stage of the course, this result is very promising. Each of the aforementioned studies uses a single method to predict the completer or dropout status of a student.

A study which compares three different methods on the same task is the one by Herzog (2006). More specifically, this study examines the predictive accuracy of back-propagation neural networks, rule-induction decision trees and multinomial logistic regression, over the problem of predicting college freshmen retention. The data used in this study, that is, data referring to 23,475 students, include both time-invariant characteristics –in the form of demographic and pre-college data– as well as time-varying attributes which are acquired as the freshmen progress in the first year of their studies. The prediction of the aforementioned methods is made in the middle of the academic year (end of first term) and at its end (end of second term), using both time-invariant and time-varying data. The results indicate that all the examined methods achieve similar correctness ratios of approximately 75% in the middle and 84% at the end of the year. The performance results, using only time-invariant student data, are not reported.

Another study that compares a variety of machine learning techniques to predict students prone to dropping a university class is the one made by Kotsiantis, Pierrakeas & Pintelas (2003). This study examines the predictive accuracy of six techniques, namely neural networks, decision trees, the naive Bayes algorithm, support vector machines, instance-based learning algorithms and logistic regression. The study uses both time-invariant student attributes and data regarding the performance of the students during the course, in the form of plenary meeting attendance and written assignment data. A total of 354 student records were collected for this study. The results that this study reports, using only time-invariant student data, range from 58% to 63%. The respective results reported in early course stages range from 63% to 72%, while the results reported before the middle of the academic year range from 78% to 84%. Based on these results, the study concludes that although it is not possible to select a single best algorithm for the task of dropout prediction, however, the naive Bayes algorithm seems to be the most appropriate, achieving an accuracy ratio that reaches 84% on the data used.

2.1.3 Literature discussion

The aforementioned studies have made significant contributions in the field of dropout prediction. The first type of studies reported in the literature focus on the identification of the most important attributes which are related to a student’s decision to drop a course using past time-invariant student data. As mentioned previously, the scope of this type of studies is to separately examine the correlation of each attribute to the students’ final decision, in order to determine the most representative predictors of dropping out. For this reason, the majority of the aforementioned studies do not report overall classification rates which would result from the combinative use of the predictors. Nevertheless, their contribution in the present study is valuable, as they allowed us to determine which of the available student data should be optimally used as past time-invariant student attributes for the task of dropout prediction.

The second type of reported studies, include both time-varying student characteristics and time-invariant data, for the dropout prediction task, using various methods, mainly machine learning based techniques. Studies of this type are more closely related to the present study, since they predict student dropout or completer status not only a-priori but also throughout the course duration, using data that emerge as students interact with the course. Although these studies are fewer, their contribution is significant as they introduce new insight to the dropout prediction task, enabling instructors to monitor the students' dropout or completer status throughout the course, instead of just in the beginning. However, they present certain shortcomings that the proposed study seeks to amend.

Firstly, not all of the discussed studies report accuracy results or the exact stage of the course that the prediction took place. In addition, for studies that report classification accuracy results, the student attributes used are measured infrequently during the course. Therefore, the predictions are made only on a few predefined dates in the duration of the course, and not frequently, in this way reducing the opportunity to promptly intervene and help at-risk students. These sparse data are the only available as far as classical education is concerned, but especially for e-learning courses, a method can be developed to take advantage of the fully computerized and extensively logged student data. Such log data have been used in relevant literature for various purposes which include obtaining, analyzing and understanding the learner paths and learning behavioural patterns (Chen, Liu, Ou & Liu, 2000; Zaiane, Man & Jiawei, 1998; Hung & Zhang, 2008; Bellaachia, Vommina & Berrada, 2006), eliciting the motivational level of the students towards the course (Cocea & Weibelzahl, 2007) and assessing the performance of learners (Chih-Ming, Yi-Yun & Chao-Yu, 2007). In addition, LMS logs have been used to obtain feedback regarding students' perceptions of the online community (Black, Dawson & Priem, 2008) and to develop visual tools that facilitate instructors in processing the LMS data and in understanding better the social, behavioural and cognitive attributes of learners (Mazza & Dimitrova, 2007; Juan, Daradoumis, Faulin & Xhafa, 2008). Nevertheless, to the best of the authors' knowledge, this study is the first attempt to use log files regarding student progress and behaviour for the specific task of dropout prediction throughout the duration of an e-learning course. Furthermore, although a comprehensive comparison of various machine learning techniques is made by two of the studies that use time-varying data, no combination of these techniques has been examined in order to amend their predictive deficiencies.

Aiming at providing an accurate and prompt means of student dropout prediction, this study focuses on e-learning courses that allow the keeping of more detailed records of students' everyday actions. The developed method uses these detailed data in order to make more frequent and more accurate predictions on the dropout or completer status of each student, in that way enabling instructors to intervene more promptly in order to reduce e-learning dropouts. In addition, in this study three popular machine learning techniques are compared in terms of performance, and combined to overcome individual predictive deficiencies, thus leading to the development of a dropout prediction mechanism, which yields more accurate results than those of the present literature.

2.2. Machine learning techniques in dropout prediction

To predict student dropouts, three popular machine learning techniques were used in this work, namely feed-forward neural networks (FFNN), support vector machines (SVM) and probabilistic ensemble simplified fuzzy ARTMAP (PESFAM). These techniques have been successfully applied to solve various classification problems and function in two phases, the training and the testing phase. During the training phase each technique is presented with a set of example data pairs (X, Y), where X represents the input and Y the respective output of each pair. In this study, Y can receive one of the following values, 0 if a student is a completer or 1 if a student drops the course. Therefore, the dropout prediction task is a two-class classification problem. Each technique then adjusts its internal parameters to infer the mapping implied by the data provided. During the testing phase each technique is presented with data that were not used during training to examine its classification performance. If the technique is found to classify most of the data in the test set correctly, then the training is considered to be successful and the machine learning technique demonstrates generalization capabilities. The following sections briefly describe the characteristics of the three machine learning techniques used.

2.2.1 Feed-forward neural networks

Feed-forward neural networks (FFNN) attempt to mimic the connectivity and learning attributes of the human brain. A typical FFNN, as described by Haykin (1999), consists of various information processing elements called neurons, interconnected so that they do not form a directed cycle. The network neurons are arranged in three types of layers, the input, one or more hidden and the output layers. Neuron connections are called synapses and exist only among neurons of adjacent layers. The FFNN architecture is shown in Fig. 1.

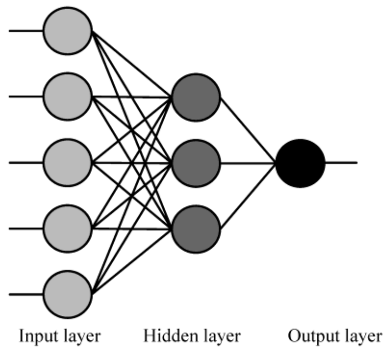

Fig. 1. Feed-forward neural network architecture

Usually, the FFNN is trained using the popular error back-propagation algorithm, which is established by Rumelhart, Hinton & Williams (1986a, b). This algorithm aims at minimizing a cost function, typically defined as the mean square error between the network actual and target output, by adjusting the synaptic weights and neuron biases. To this end, during its training phase, the network is presented with the training set, which consists of examples of an input vector and the corresponding output vector. Next, the input information is forwarded from the input nodes, through the hidden layers, to the output nodes and the error between the desired and the actual response of the network is calculated. This error signal is then propagated backwards adjusting the network weights and biases. This process is repeated for each example in the training set and when the entire training set has been presented to the network, an epoch has elapsed. Several epochs may be needed before the training phase is completed. A popular variation of the error back-propagation algorithm is the Levenberg-Marquardt algorithm (Hagan & Menhaj, 1994). This algorithm has been found to increase the speed convergence and effectiveness of the network training, since it is effective in solving non-linear least squares problems, as in the case of minimizing the cost function of the FFNN.

During its training, however, a FFNN may end up adjusting its weights and biases to represent only the training data and thus lose its ability to generalize to situations not presented during training. This phenomenon is called overfitting and can be avoided by using a separate data set called the validation set. The FFNN parameters are estimated based only on the training set, and the performance of the network is evaluated by computing the mean square error on the validation set. When the network performance deteriorates, it usually means that overfitting has occurred. Then, the training stops and the parameters of the previous network epoch are stored. The training phase can be terminated by reaching to a minimum of the cost function, by meeting the performance goal or by detecting that the validation set produced increasing mean square error.

Finally, as soon as the FFNN training finishes, the network testing phase takes place. During this phase, unseen data are presented to the trained network to evaluate its performance. These data comprise the test set that is disjoint to both the training and the validation data sets.

2.2.2 Support Vector Machines

The support vector machines technique is pioneered by Vapnik (Cortes & Vapnik, 1995; Vapnik, 1995, 1998) and has been successfully applied to many classification problems throughout the literature. This technique attempts to separate two classes of data using a hyperplane defined by support vectors, which are part of the data set.

Through its training, the support vector machines technique searches for the Optimal Separating Hyperplane (OSH), which is the optimal hyperplane that maximizes the margin between the two classes of the training dataset. To make this clearer, Fig. 2 presents an example in 2 dimensions. In this example, several lines can be found to successfully separate the data into two classes. The optimal line, the OSH, is the one which lies half-way in the margin. The term "margin" refers to the sum of the distances *d* of those data that are closest to the line, which are defined as the support vectors.

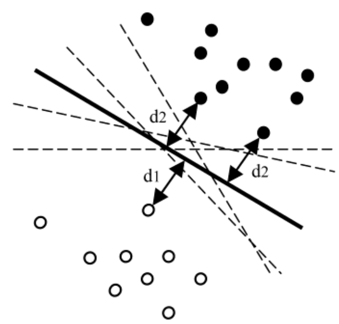

Fig. 2. Support vector machines concept

Usually, since the data in most real-world problems are not linearly separable, the aforementioned technique is generalized to handle non-linearly separable inputs at the cost of accepting a number of misclassifications. To further improve its effectiveness, the SVM technique transforms the data into a feature space of a higher dimension than that of the input before attempting to separate them using a linear discriminator. To this end, it uses a kernel function. A variety of popular kernel functions exist, like linear, polynomial, and radial basis function kernels. In this work, the radial basis function kernel was used.

During its testing phase, the SVM technique transforms the data of the test set into the feature space that it used for its training and then it classifies these data, based on the OSH found at the training phase.

2.2.3 Probabilistic Ensemble Simplified Fuzzy ARTMAP

The last machine learning technique used by this study is the probabilistic ensemble simplified fuzzy ARTMAP (PESFAM). The PESFAM (Loo & Rao, 2005) combines a number of simplified fuzzy ARTMAP (SFAM) modules with a plurality voting strategy based on the findings of Lin, Yacoub, Burns & Simske (2003). The ARTMAP technology is an evolution of the original Adaptive Resonance Theory (ART), introduced in (Carpenter & Grossberg, 1987), into a self-organizing neural network.

The SFAM technique was proposed by Kasuba (1993) and reduced the computational overhead and architecture redundancy of the original fuzzy ARTMAP network (Carpenter, Grossberg, Markuzon, Reynolds & Rosen, 1992). A SFAM network consists of three layers, as depicted in Fig. 3, that is the input, the output and the category layer. Every input node is connected to each node in the output layer and every output layer node points to single category node, which corresponds to one of the classes of the classification problem at hand. All the data to be fed into the SFAM network are firstly transformed into the [0,1] value range.

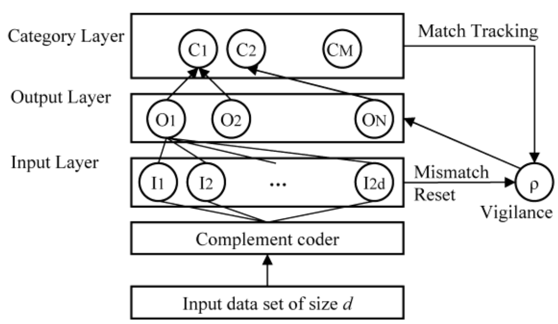

Fig. 3. Simplified Fuzzy ARTMAP architecture

The algorithm of the SFAM technique can be summarized as follows. During its training phase, the SFAM algorithm firstly complements its input data, through the complement coder module. That is, for each feature a in the data set, the algorithm creates its complement, $a'=1-a$. The features, as well as their complements are then encoded together into an input vector. Then, each input vector of the data set is used to calculate an activation value for every output node O, based on the activation function of the network. The node that has the maximum activation value is defined as the winner. If this winner node does not point to the correct category that the data of the input vector actually belong to, then a new output node is created for this input in the output layer, which points to the true category of the input vector. Otherwise, if the winner node points to the correct output category, a process called “match tracking” takes place. Through this process, the SFAM technique examines whether resonance between the current input and the winner output node occurs or not, by comparing a match function to a parameter called vigilance ρ. The term “resonance” states that the current input corresponds to the winner node in the output layer, while at the same time it matches this winner node well enough to be encoded to it. If resonance happens, the output node adjusts its weight, to “learn” the current input. On the other hand, if the winner node does not meet the appropriate granularity represented by the vigilance parameter, its weights are not adjusted, but a new output node is created for the current input vector instead.

During its testing phase, for each of the vectors in the test set, the SFAM technique calculates the activation values of every node *O* in the output layer, to calculate the corresponding winner node. Then, the input vector is classified to the category that this winner node points to.

The SFAM network although effective, may present a serious drawback; it is highly dependent on the order of the pattern presentation during its training. To overcome this, Loo & Rao (2005) propose the probabilistic ensemble SFAM classifier (PESFAM). According to this model, multiple SFAM networks are trained using a different data presentation sequence for each one of them. Their results undergo a voting strategy, called probabilistic plurality voting (Lin, et al., 2003), to enhance the general performance. The plurality voting strategy is basically a method to combine decisions from a group of classifiers. In general, the main motivation for using an ensemble of neural networks is to mitigate the limitations of each constituent network. In Fig. 4, the PESFAM architecture is presented.

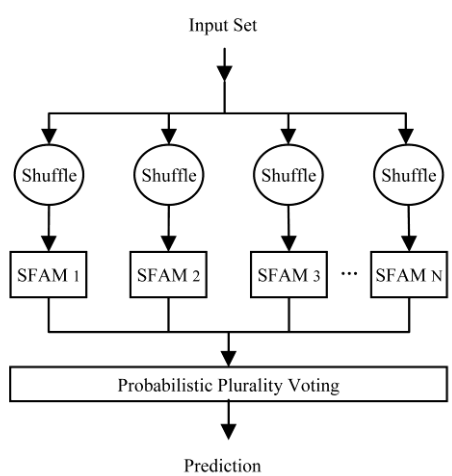

Fig. 4. The probabilistic ensemble SFAM architecture

2.3 Combining the results of the machine learning techniques using decision schemes

Since a single machine learning technique may fail to detect a number of dropout students, while another may succeed, three decision schemes to combine the network results were also examined.

First, the results of the three networks are summed to calculate the dropout level of each individual student in the dataset. The results of each network are binary, where ‘1’ represents the case that a student has been categorized as a dropout and ‘0’ indicates that this student has been identified as a completer. Therefore, the dropout level of a student may receive four distinct integer values ranging from 0 (completer) to 3 (all the networks have categorized this student to be a dropout). As soon as the dropout level of a student has been calculated, this student may be categorized as a dropout or a completer using three different decision schemes which are defined as follows:

Decision Scheme 1: A student is considered to be a dropout if at least one technique has classified this student as such. In other words, a student is categorized as a dropout if his dropout level is greater than or equal to ‘1’, otherwise he is classified to the completer category.

Decision Scheme 2: If this scheme is selected, a student is considered to be a dropout if at least two techniques indicate this student to be a dropout, that is when the student’s dropout level is calculated to be greater than or equal to ‘2’.

Decision Scheme 3: According to this scheme, a student is considered to be a dropout if all techniques agree that this student is a dropout. In this case, the student is identified as a dropout only if his predicted level is equal to ‘3’.

The three decision schemes are illustrated in Fig. 5.

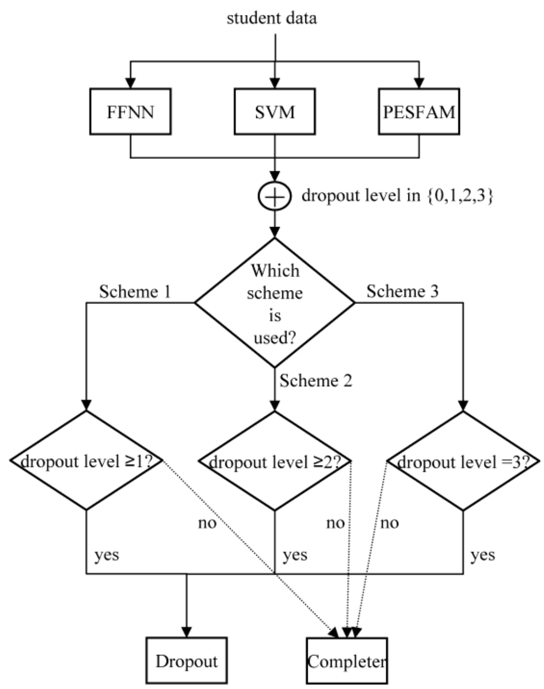

Fig. 5. Overview of the decision schemes used by the proposed method

2.4 Dataset description and algorithm implementation

2.4.1 Dataset description

This study was applied on two introductory level e-learning courses, namely “Computer Networks & Communications” (NET) and “Web Design” (WEB). The courses are provided by the e-learning team of the Multimedia Technology Laboratory of the National Technical University of Athens, Greece (Medialab, 2008).

Each course consists of eight educational sections and is offered twice a year, in the Spring and Fall semesters. During the first seven sections, the educational material of each program is delivered to the students and their knowledge is assessed through testing material which consists of multiple choice tests, to examine the theoretical knowledge that the students acquired, and projects to test the application of this knowledge on practical terms. The final examinations for each course are conducted during the eighth section. The structure of each educational section is predetermined, using the same number of multiple choice tests and projects every year. This allows the proposed method to produce valid results when applied in later semesters.

The NET course includes 4 multiple choice tests, conducted at the end of the 1st, 3rd, 4th and 6th course section, and 4 projects conducted at the end of the 1st, 2nd, 3rd and 5th section.

The WEB course includes 5 multiple choice tests conducted at the end of each of the first 5 sections and 7 projects conducted at the end of each of the first 7 sections.

At the end of the final examinations of each course, an overall grade, measured in the scale of 0 to 100, is calculated for each student, to determine whether the student passes the course or not. Specifically, 60% of the overall grade is calculated using the average scores that the student achieved on multiple choice tests and projects throughout the first seven sections of the course. The remaining 40% of the overall grade is calculated using the average scores achieved in the final multiple choice test and project assignment, at the eighth section of the course. To pass the course, students need to achieve an overall grade of at least 50. The participation of the students in the final examinations is mandatory in order to pass the course; however there are no other rules regarding the projects and multiple choice tests taken throughout the rest of the course duration.

Both courses are introductory level and are targeted towards adults of various educational backgrounds, ranging from high-school graduates to master-degree holders. Nevertheless, students are advised to have basic computer and English language skills, since an important part of the material is delivered in English. The courses are delivered solely through the Moodle open-source Learning Management System (LMS) platform (Moodle, 2008) without the requirement for plenary meetings during the semester.

E-learning students can be divided into four categories:

1. Students that registered but never entered the e-learning course.

2. Students that entered the e-learning course and completed a number of sections but decided to drop out completely.

3. Students who completed some of the course sections but decided to discontinue their studies and repeat the course in a following semester.

4. Students who completed all of the course sections and successfully completed the course.

Only students who belong to the 4th category are considered to be completers. As far as the dropout students are concerned, the NTUA Multimedia Technology Laboratory policy states that once they have dropped the course, they are not entitled to a refund but instead they are automatically re-enrolled in the subsequent semester. This, however, may result in the accumulation of non-active past student registrations in each semester. Since effective dropout prevention measures can be taken by the instructors regarding only active students, this study expands the dropout definition provided by Levy (2007) as follows:

Dropout students of a given semester are those students that choose to discontinue their studies having financial penalties and have accessed the e-learning platform at least once during the semester.

Therefore, only students that belong to categories 2 and 3 are considered as dropouts in this work.

Data used in this study originate from the Spring 2007, Fall 2007 and Spring 2008 courses. The number of the total participating students in the two courses for these semesters is 193, out of which 109 students successfully completed the courses. The total number of the dropout students is 84 thus producing a dropout rate of approximately 44%, a number that confirms similar results from related studies. The analytical data regarding the examined cases of students who completed the course and students who dropped out are presented in table 1.

	NET			**WEB**		
	Spring 2007	Fall 2007	Spring 2008	Spring 2007	Fall 2007	Spring 2008
Completers	27	23	20	14	11	14
Dropouts	23	18	19	9	5	10
Total	130			63		

Table 1. Student completer and dropout cases examined

The experimental procedure consisted of the following steps: first, for each one of the two courses, the data set to be fed into the machine learning networks was extracted from the Learning Management System Database. This dataset consists of both time-invariant and time-varying attributes. The time-invariant attributes were acquired from the registration forms completed by the e-learning students prior to the beginning of the courses. These forms are delivered by the National Technical University of Athens and their questions are predefined. From the available registration data, the attributes that relevant literature identifies as among the most important predictors were then selected. Some important predictors, according to related literature, such as age, were not available from the registration forms, while other identified attributes, like ethnicity, were not applicable to this study, since all participants had the same ethnicity. As stated in the related work section, there are two major categories in time-invariant dropout prediction variables, namely, demographic and prior academic performance related attributes. Firstly, from the demographic attributes proposed by relevant literature, this study incorporates gender, residency (in the place of ethnicity) that receives values capital or province, and working experience in the field of the course, which incorporates the students' past training with technological means and computers. Secondly, as far as prior academic performance is concerned, this study uses the educational level of the students in the value range of basic to PHD degree, as well as their level of fluency in the English language.

The rest of the extracted attributes were time-varying characteristics depicting the students' progress during the courses, as well as their level of engagement with the e-learning procedure, as also supported by relevant literature. More specifically, student progress is measured using their project scores and multiple choice test grades. The participation that students demonstrated during each course section is measured using their activity levels, which are derived from the LMS log files and include the number of posts a student has made to the course forums, the number of times that the student has accessed the platform, the number of times he has viewed the educational material and so on. As a final indicator of the students' engagement to the course, their ability to comply with project deadlines was also measured. The value that this attribute receives is calculated as the total days that have elapsed after the expiration of the project submission deadline until the date the student submits his project. Table 2 presents the attributes used during the experiments and the range of their values.

	Related literature category	**Attribute**	**Value Range**
Time-invariant attributes	Demographic	Gender	male, female
		Residency	Capital, province
		Working experience	≥0 in years
	Prior academic performance	Educational level	basic, intermediate, higher, master degree, PhD degree
		English language literacy	elementary, basic, high, proficient
Time-varying attributes	Multiple choice test grade		0-20
	Project grade		0-100
	Project submission date (days counting from section deadline)		≥0
	Section activity		≥0

Table 2. Student attributes used to train and test the machine learning networks

2.4.2 Algorithm Implementation

The above data set was then divided into two separate sets, namely the training set and the test set, which were used during the experiments. The network training set consisted of the Spring 2007 and Fall 2007 student data, that is 91 students of the NET course and 39 students of the WEB course. The network test set consisted of the Spring 2008 student data, corresponding to

39 students of the NET course and 24 students of the WEB course. Thus, instead of constructing the test set using a random selection of students from every semester, we chose to test the algorithms using data from a separate semester, in particular Spring 2008. This type of test set was chosen to examine whether the aforementioned techniques can predict on a different student population, whose characteristics may differ from the student examples they were trained with. This process simulates how an instructor would train the networks using past available student data and then use the trained networks on a subsequent semester. If this process produces accurate results, it is an indication that the prediction methods can generalize well.

Next, for each course, the training and testing phases of the algorithms took place. During the training phase, the algorithms were trained using Spring 2007 and Fall 2007 course data, extracted from the Moodle LMS database. Initially, the networks were trained using only the demographic characteristics. Then, the training was completed in seven successive steps, during which the algorithms were gradually provided with the attributes of each section up to the current, until all section attributes had been used. For instance, in the first step the algorithms were tested using the attributes of the first section. The second step included the attributes of both the first and the second section and so on.

As soon as the training phase was over, the testing phase took place. During this phase the performance of the three machine learning techniques and the decision schemes was examined using data extracted from the Spring 2008 course. The testing phase also included testing the method, first using the time-invariant attributes and then according to the same seven steps described in the training phase.

The machine learning techniques used in this study were implemented using the Mathworks MATLAB R2007b platform. The precise technique specifications are as follows: As far as the FFNN implementation is concerned, seven X-4-1 FFNNs were constructed for each course. X corresponds to the number of the neurons in the input layer, that is the number of the section attributes available at each step. For instance, in the first section of the WEB course, a 4-4-1 FFNN was used, while in the third section of the same course, a 12-4-1 was used. During the training phase of the FFNN technique, the training set was divided into two disjoint subsets, used as the training and validation sets. More specifically the resulting training set consisted of the 85% of the initial training set, that is 77 and 33 randomly selected students for the NET and WEB courses respectively. The remaining 15%, corresponding to 14 students from the NET course and 6 students from the WEB course, was used as the validation set. This process, as described above, was used to ensure that no overfitting occurred. The use of a validation set is only necessary for the implementation of the FFNN technique, and thus the training set was used as is for the other two techniques. As far as the SVM implementation is concerned, seven SVMs were constructed for each course receiving as input the respective data for each section. The type of kernel used was the radial basis function kernel. Finally, for the PESFAM implementation, seven PESFAMs were constructed for each course receiving at each step the same data as the other two techniques. Each PESFAM combined the outputs of five SFAMs to reach to a decision.

As described above, the output of each technique for a single input student data is binary, 0 corresponding to the estimation that the student is a completer and 1 corresponding to the estimation that the student is a dropout.

3. Experimental Results

In this section, the experimental results of the study are presented. To evaluate the performance that each one of the machine learning techniques and decision schemes had on the test set, three performance criteria were used, which estimate the overall accuracy, sensitivity and precision of the methods. Next, an analytical description of each criterion and the rationale for using it are presented, followed by its respective results.

3.1 Overall accuracy

The overall accuracy criterion is used to measure the total proportion of the students whose final status, completer or dropout, was correctly predicted by a technique. In other words, this criterion measures the number of successfully predicted completers plus the number of successfully predicted dropouts, versus the total number of the students:

$$overall\_accuracy = \frac{correctly\_predicted\_completers + correctly\_predicted\_dropouts}{total\_number\_of\_students}$$

The overall accuracy results of the three machine learning techniques and the decision schemes on the test sets of the NET and WEB courses are depicted in figures 6 and 7. The left vertical axis in these figures presents the number of students that were correctly predicted by each technique, while the right vertical axis depicts the respective rate of correct predictions. The horizontal axis corresponds to the steps in which the algorithms were tested, starting with the demographic characteristics and reaching the seventh section of the course.

Initially, as one may observe, using as input only the demographic characteristics, the techniques were not found to provide significant performance results, on neither course.

Nevertheless, as the courses progressed, the prediction results of all techniques improved significantly, using the student attributes extracted from the LMS database.

As far as the NET course is concerned, no individual network was found to be superior to the others in every course stage. More specifically, feed-forward neural networks were superior during the first section, whereas support vector machines outperformed the other two techniques during the second section and PESFAM performed better in the third section. Throughout the rest of the course sections, two or three of the networks presented the same high results. As far as the decision schemes under examination are concerned, "Decision Scheme 1" was found to produce the best results, since it performed better than or at least equally to all the examined individual networks and decision schemes, in all course stages. More specifically, this scheme managed to achieve 85% overall accuracy on the test set in the first course section and reached a 97% rate in the middle of the course, which remained unchanged until the end of the course. A small fluctuation in the generally increasing percentages achieved by Scheme 1, occurred between section 1, where the scheme successfully predicted 33 out of the total 39 students (85%) of the course, and section 2, where it successfully predicted 32 out of the 39 students (82%). This can be explained by the fact that some students may not always present behaviour consistent to their final status, especially during early course sections, which, in conjunction to the small amount of student data available in these sections, makes the techniques more susceptible to error. Nevertheless, Scheme 1 remained the most successful technique during these sections too.

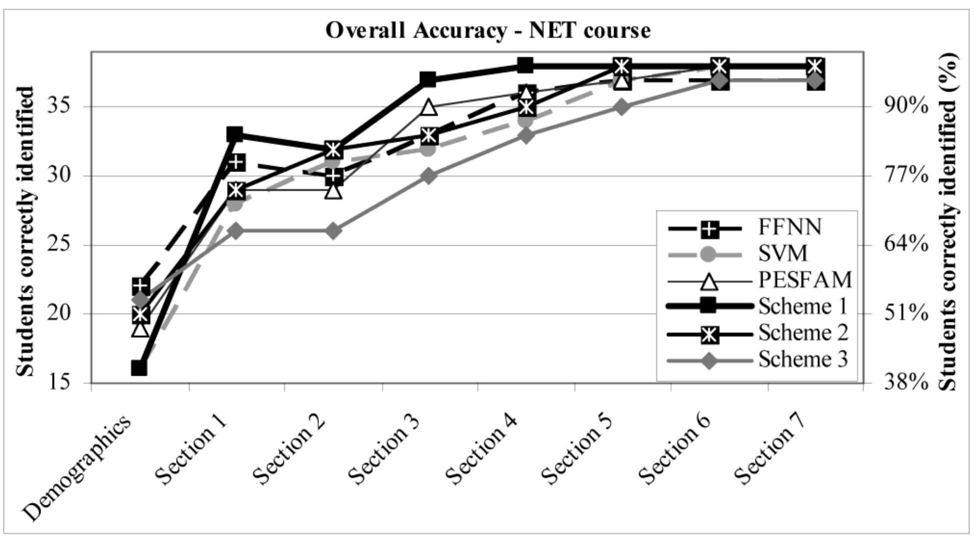

Fig. 6. Overall accuracy results on the 39 students of the NET course

The experimental results on the overall performance of the examined techniques over the total 24 students of the WEB course are depicted in Figure 7. These results also indicate that a single machine learning technique does not provide accurate estimations over every course section. The predictions made by the FFNN technique outperformed or were equal to the predictions of the other two techniques during sections 1, 4, 6 and 7, while the best predictions of the SVM technique were achieved during sections 2, 5, 6 and 7. The PESFAM technique achieved the best results during sections 3, 5, 6 and 7. Similar to the NET experiments, “Decision Scheme 1” was found to consistently achieve the highest overall accuracy results in all course sections, starting from 75% in the first section and reaching 100% in the seventh section of the course.

As in the case of the NET course, a small drop in the performance of a technique might occur in some stages of the course. For instance, the performance of the PESFAM technique drops slightly between sections 3, where it correctly predicted 21 out of the 24 students and 4, where due to an additional misclassification it correctly predicted 20 students. This can also be explained due to changes in student behaviour that led this particular technique into misclassifying them. However, since the three individual techniques function in different ways, this changing student behaviour did not affect negatively the performance of the FFNN and SVM, and thus it did not have an effect on the correct student classification achieved by Decision Scheme 1.

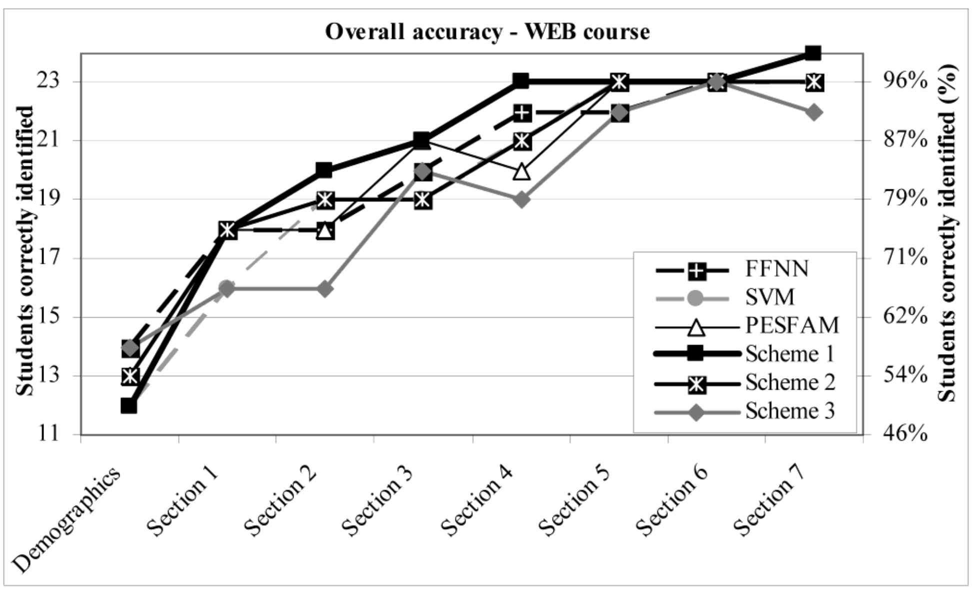

Fig. 7. Overall accuracy results on the 24 students of the WEB course

As one may observe, the implementation of the machine learning techniques on e-learning data, and especially their combination using the “Decision Scheme 1”, resulted in high performance in predicting student completer or dropout status even from the first course sections.

In addition, demographic characteristics were not found to aid in accurately predicting whether a student will drop out or not.

Nevertheless, the overall accuracy rates may not fully reflect the method capabilities over the task of specifically predicting dropout students, since they include predictions regarding both completers and dropouts. To this end, and since this study focuses on predicting dropout students, two additional criteria are used to depict the method performance over dropout prediction in particular.

3.2 Sensitivity

To measure the efficiency that each technique demonstrates in correctly identifying dropout students, the sensitivity criterion is used. This criterion measures the proportion of students that were correctly identified by a technique as dropouts, versus the total number of actual dropout students.

$$sensitivity = \frac{correctly\_predicted\_dropouts}{actual\_dropouts}$$

Figures 8 and 9 present the sensitivity results of the machine learning techniques and the decision schemes on the NET and WEB courses respectively. Similarly to the above, the right axis depicts the sensitivity percentages achieved by each technique and the left axis represents the correctly predicted number of dropout students. In addition, the horizontal axis corresponds to the steps during which the algorithms were tested. In this case also, the demographic characteristics were not found to accurately predict dropout students, since they produced low results for both the NET and the WEB courses.

The experimental results on the NET course indicate that the individual machine learning techniques produced moderate to low sensitivity results in the first two sections, mainly due to the fact that some students decide to drop out later in the course and therefore could not be accurately predicted. However, even in this case, “Decision Scheme 1” achieved satisfactory to high performance results. The sensitivity results of the techniques improved significantly from the 3rd section to the end of the course, mainly as a result of the very accurate predictions made by PESFAM. This method was found to produce the best results regarding sensitivity, followed by the support vector machines technique. Feed-forward neural networks did not perform as well as the other two techniques, except in the first section. As far as this criterion is concerned, “Decision Scheme 1” produced the best results since its sensitivity ranged from 74% during the first two sections, and reached 95% from the 3rd section to the end of the course. These high sensitivity results indicate that this decision scheme correctly predicted the majority of dropouts in the e-learning course, even in the early stages of the course. The results of the other two decision schemes ranged at the same levels as the results of the individual techniques.

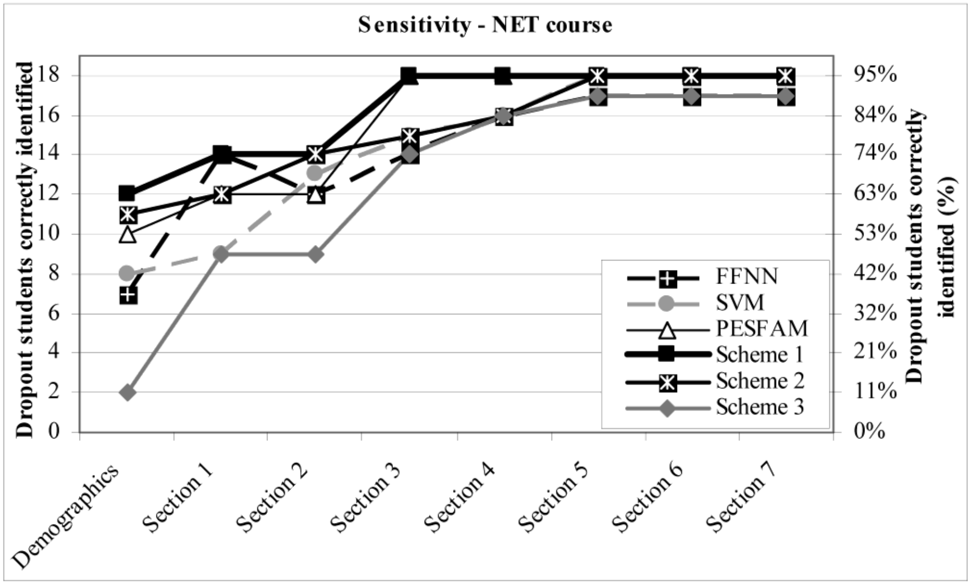

Fig. 8. Sensitivity results on the 19 dropout students of the NET course

Figure 9 depicts the experimental results that the techniques presented over the sensitivity criterion on the WEB course data set. One may observe that during the first two course sections, the accuracy of the three individual machine learning techniques was moderate to low in correctly identifying dropout students. However, the combination of their results using “Decision Scheme 1” produced accurate estimations. This indicates that although an individual machine learning technique may fall into error, a dropout student is unlikely to be misclassified by all three of them. In most cases the three techniques misclassify different students and therefore the use of “Decision Scheme 1” produces a consistently high sensitivity rate even in the early course sections. However, this decision scheme still depends on the predictions made by the individual techniques and thus, in case the latter provide moderate results, then the results of “Decision Scheme 1” also drop, as it can be observed in section 2. In this section, the sudden drop in the performance of “Decision Scheme 1”, from 90% to 70%, is explained by the respective moderate results of the individual techniques and especially the decreased performance of the PESFAM technique. Nevertheless, Scheme 1 still performed 10-20% better than the individual techniques. As the course progresses and more student data are gathered, the three machine learning techniques and the decision schemes demonstrate generally improved results. The high sensitivity rate of “Decision scheme 1” indicates that this particular combination of the three machine learning techniques enhances their individual results and leads to the accurate identification of at-risk students.

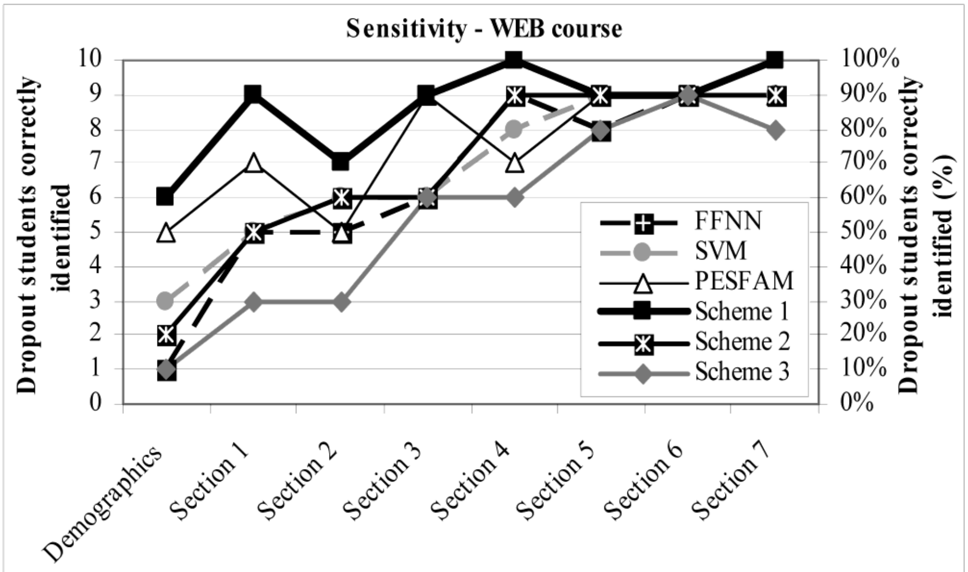

Fig. 9. Sensitivity results on the 10 dropout students of the WEB course

3.3 Precision

The number of students that a technique or decision scheme classifies as dropouts is two-fold: it consists of the students that were correctly identified as dropouts and the students that are actually completers but were erroneously identified by the technique as dropouts. The precision criterion is used to determine the proportion of the students that were actually dropouts among all those that the technique predicted as such. A high precision rate indicates that the examined technique tends to label as dropouts only those students that actually drop the course.

$$precision = \frac{correctly\_predicted\_dropouts}{incorrectly\_predicted\_dropouts + correctly\_predicted\_dropouts}$$

The precision performance results of the networks and the decision schemes on the NET and WEB courses are presented in figures 10 and 11 respectively. In these figures the vertical axis corresponds to the precision rates achieved by a technique, while the horizontal axis corresponds to the steps during which the techniques were tested.

Firstly, as it can be observed, the precision results of the demographic characteristics present similarly low prediction accuracy results, for both courses, as in the case of the other two criteria.

As far as the NET course experimental results are concerned, only a few completers were misclassified by the techniques. In this case also, “Decision Scheme 1” presented the most accurate results, achieving precision rates that ranged from 88% to 100%.The only case that this technique slightly dropped its performance was between section 1, where it correctly predicted 14 out of 19 dropout students and misclassified 1 completer (93%) and section 2, where it predicted the same number of dropout students but misclassified one more completer (88%).

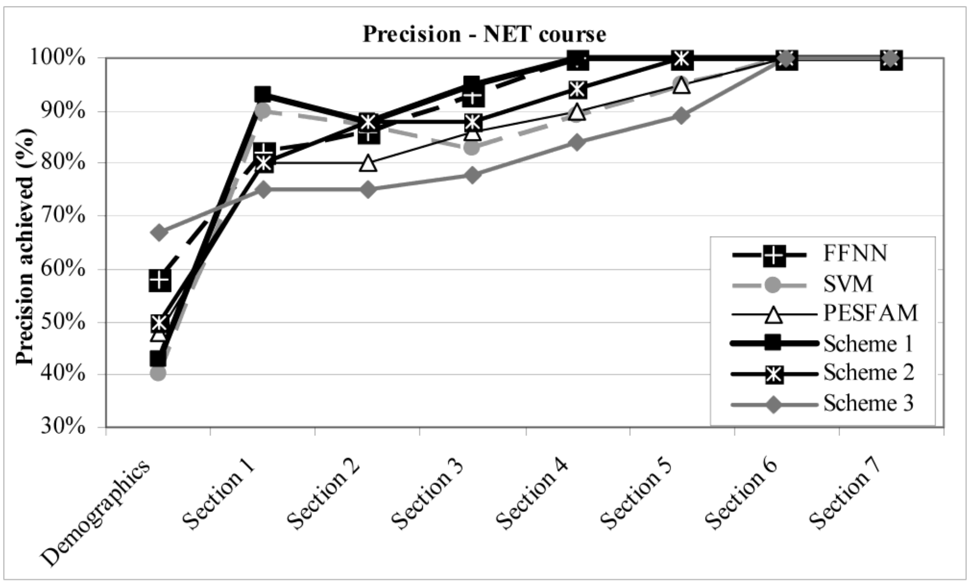

Fig. 10. Precision results on the NET course

The precision results that the examined techniques and decision schemes had on the data of the WEB course are depicted in figure 11. On the one hand, as far as the individual techniques are concerned, the FFNN achieved the highest results in sections 1, 2 and 3 while the SVM and PESFAM techniques achieved moderate results. On the other hand, the combinations of the three techniques during the first three sections, produced results of moderate to high precision. From section 4 until the end of the course, all the techniques and decision schemes were very precise.

In the case of the WEB course, and compared to the other techniques, Decision Scheme 1 was found to yield lower results in sections 1 and 3. This is probably explained by the fact that Decision Scheme 1 tends to be affected more by the completer misclassification errors made by the individual techniques than the other two Decision Schemes. Nevertheless, as indicated by the experimental findings, the precision results of this scheme were equal to or higher than those of the other techniques in 5 out of the 7 sections of the course. It should also be noted that although the three individual techniques misclassified fewer completers during sections 1 and 3, thus achieving higher precision results, they also identified only a small number of dropouts, as indicated by their respective sensitivity results. On the other hand, "Decision scheme 1", misclassified more completers, yet it correctly identified significantly more dropouts. For instance, in section 3 this scheme achieved an 82% precision rate, resulting from the fact that it correctly identified 9 dropout students and incorrectly classified 2 completers as dropouts. In the same section, the FFNN technique misclassified no completers, achieving a precision rate of 100%, but identified only 6 dropouts. Therefore, despite providing lower precision results than the FFNN technique, "Decision scheme 1" is preferable for the purposes of this study. Apart from sections 1 and 3, and as the course progresses, "Decision Scheme 1" achieved high precision results, reducing the number of misclassified completers and correctly categorizing most dropouts.

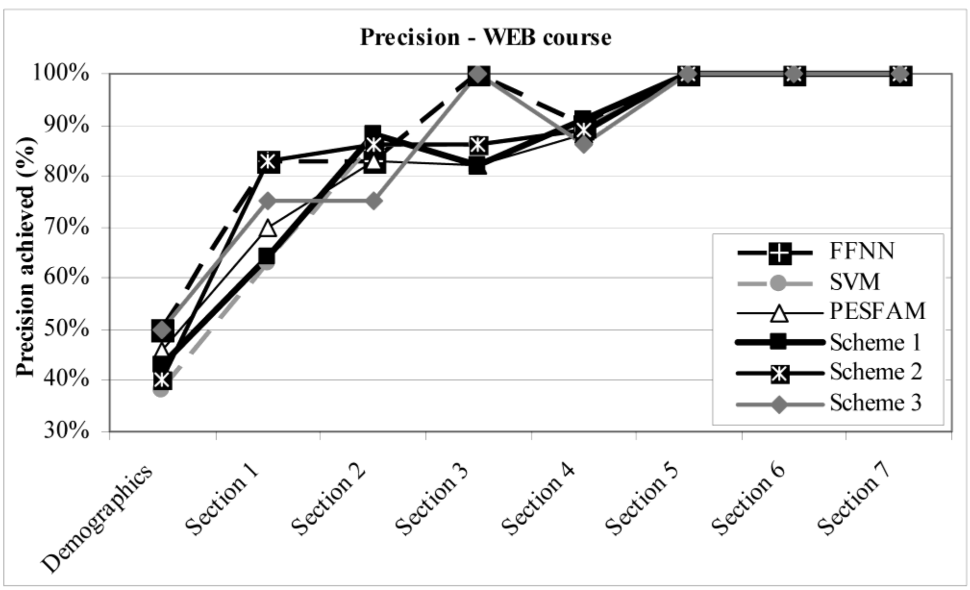

Fig. 11. Precision results on the WEB course

3.4 Method Promptness in Identifying Dropouts

Another significant element of dropout prediction is the promptness of the proposed method, that is, how earlier the method accurately identifies a dropout student before the student actually decides to drop the course. This is important, since a timely dropout prediction is expected to facilitate instructors to promptly intervene and help students to complete their studies.

For this reason, for each dropout student, the last section that this student showed progress or engagement was calculated. This section corresponds to the last section that the student completed a multiple choice test, submitted a project, or had a positive value in the participation attribute. The calculated section was then compared to the section that the proposed method correctly identified the student as a dropout for the first time.

Results show that the average section that the dropout students of the two courses last presented progress or participation was section 4.6 (out of the first 7 sections which comprise the educational course duration), a fact which also shows that dropout students participate for a number of sections averagely before dropping the courses. The average section that the proposed method firstly identified the dropout students was section 1.5. This means that the proposed method identifies dropout-prone students at an average of 3.1 sections earlier than the section that these students last show interest in the courses and thus it is expected to help instructors promptly intervene to aid students at risk.

4. Discussion

In this section, the key features of the proposed method are discussed. Initially, as far as the data used are concerned, this study takes advantage of the fully computerized character of e-learning, using data automatically extracted from the LMS database. This type of data provides instructors with highly detailed student profiles, which include student grades in multiple choice tests and projects, the students' ability to comply with project deadlines and their activity throughout the course. All these attributes were found to be important to the dropout prediction method, since they describe different student cases. For example, although a student may log in to the platform and participate in the forums frequently, this student may get low grades and eventually drop the course. The opposite example can also be found in the dataset, depicting the case of a student who does not actively participate in the course, demonstrating low activity levels, yet completes the course achieving good results. Additionally, as it has been observed in the dataset, a student may show progress by completing multiple choice tests and submitting projects, yet this student has difficulty in completing the course requirements, thus delaying the submission of his projects, and eventually drops the course. Therefore, the use of these detailed student data help towards more accurate dropout predictions.

As an additional observation, time-invariant student data were found to be less accurate predictors of a student's decision to drop out compared to time-varying data, obtained as the course progresses. As shown in section 3, by using time-invariant data the overall accuracy, sensitivity and precision rates achieved were 41-50%, 60-63% and 43%, while higher experimental results were acquired using time-varying attributes, that correspond to 75-82%, 70-74% and 64-88% in early course stages and 97%, 95-100% and 100% in latter stages, respectively. This fact is also supported by the results of related literature studies that, like the present study, use both time-invariant and time-varying attributes on the task of dropout prediction. More specifically, using only time-invariant data Kotsiantis et al. (2003) report 58% - 63% overall classification accuracy, which reaches 84% through the use of time-varying characteristics. Furthermore, Moseley & Mead (2008) report a sensitivity rate of 31% using time-invariant data that reaches 84% when time-varying data are used. However, higher results can also be found in the related literature, in studies that use only time-invariant data. Specifically, Roblyer et al. (2008) report an overall classification rate of 79.3%. However the sensitivity results reported by this study regarding the prediction of dropout students alone are 30.4%. In addition, Morris et al. (2005) report an overall correctness ratio of 74.5%. Therefore, the results of both this work and the ones of the related literature suggest that time-invariant student data can be used to obtain an indicative a-priori student dropout prediction. Then as the course progresses, instructors can benefit from availability of the time-varying e-learning data, to obtain more accurate predictions regarding students' decision to drop the course.

As far as the machine learning techniques are concerned, they performed well on all evaluation criteria for both e-learning courses. Machine learning techniques have the advantage of being data-driven instead of model-driven, that is they do not a-priori assume an explicit relationship model among the data, as model-based linear or nonlinear methods do. Instead, the model structure and the model parameters that they use are derived from the actual dataset of the problem. In addition, to overcome individual technique limitations, in this study we use three decision schemes that combine the results of the machine learning techniques. The experimental results show that Decision Scheme 1 seems to be the most appropriate solution for achieving and maintaining high accuracy, sensitivity and precision results in predicting at-risk students, as this scheme yielded the highest results in almost all course sections for both courses.

The use of the proposed method demonstrates results of higher accuracy than those produced by the related studies mentioned in section 2.1. More specifically, the highest overall accuracy rate reported in the literature, concerning the application of feed-forward neural networks, was approximately 84% (Kotsiantis et al., 2003), while the respective experimental results of the proposed method were 96% (NET course) and 95% (WEB course). Moreover, the highest overall accuracy rate reported, regarding the application of support vector machines, was 83% (Kotsiantis et al., 2003) which is lower than the respective 97% (NET course) and 96% (WEB course) achieved in the present study. The PESFAM technique, which was also found to yield high overall accuracy results, is not used in the relevant literature. The highest overall correct classification results reported in the relevant literature were 94% and were achieved using decision trees by Moseley & Mead (2008). This study also provides results regarding the criteria of sensitivity (84%) and precision (70%). Compared to these results, the proposed method using Decision Scheme 1 is found to be better in all three criteria, achieving an overall accuracy rate ranging from 97% to 100% for the two courses examined, a sensitivity rate of 95-100% and a precision rate of maximum 100%. Summarizing, the above comparison can be used as a qualitative indicator that the use of detailed data can further improve the efficiency of machine learning techniques on the task of dropout prediction, and lead to more accurate identification of dropout students.

The application of the proposed method on other kinds of courses, like blended learning, distance education or classical education, is not examined in the present study; nevertheless, the method may also be applied on other kinds of courses, as well. The results that the method will present in each case are expected to depend on the level of detail of the available student data. However, it should be noted that due to the fact that other kinds of courses either provide less detailed automatic logging of everyday student actions or none at all, these courses are expected to provide sparser and less detailed data compared to e-learning. Therefore, in this case the ability of the method to promptly and accurately identify a dropout-prone student might be reduced.

However, the proposed method also presents a few shortcomings. First of all, the algorithm training phase generally requires a certain amount of time. For instance, the training for a course, such as the ones presented in this study, on a 2.4GHz computer of 512MB RAM is completed in approximately one minute. Nevertheless, this is not considered a major drawback of the method, for the purposes of the specific application. In particular, e-learning instructors need to train the algorithms only once at the beginning of each semester, using data from previous courses. Then, the trained networks can be used to automatically make their predictions on the new students of the current semester. Therefore, training time is not expected to impede the prediction process.

The accuracy results of the method range from moderate to high, and improve as the course progresses. Moderate results, ranging from 75% to 85%, mostly appear in the first section of each course and can be partially explained due to the fact that certain students did not consider dropping the course at this point in time, though this decision was made in later sections. A student is categorized as a dropout by taking into consideration his final status at the end of the course and not the actual time that the student decided to drop out, because this information would be very difficult to retrieve. Therefore, these results do not necessarily render the proposed method less potent.

Finally, the machine learning techniques make their predictions based on data from past courses and therefore depend on the size and quality of the data used for their training. The more indicative examples of the problem they are presented with, the more accurate the predictions they are expected to make. In addition, although they can infer a correct solution based on noisy data, they have difficulty in making correct predictions on data that are contradictory to the ones used for their training. However, as illustrated by the experiments, accurate results can be produced based on a medium-sized educational dataset corresponding to few students, a fact which is expected to facilitate instructors when using the proposed method.

5. Conclusion

This paper presents a method for early and accurate student dropout prediction in e-learning courses. It draws on the detailed student logs, extracted from the Learning Management System that hosts the e-learning courses, to dynamically make its estimations and adapt them to student progress throughout the course.

The method uses three machine learning techniques, namely feed-forward neural networks, support vector machines and probabilistic ensemble simplified fuzzy ARTMAP. To overcome individual technique inaccuracies in identifying dropout students, the method combines their estimations using three different decision schemes.

The three individual machine learning techniques, as well as the three decision schemes were tested using data from two e-learning courses, to determine their overall accuracy, sensitivity and precision. The experimental results showed that the most successful technique in promptly and accurately predicting dropout-prone students is the decision scheme which identifies a student as being a dropout if this student has been determined as such by at least one of the three machine learning techniques. Using these schemes, the proposed method achieved a 75-85% overall student classification rate from the first section of the two courses, to reach a 97-100% rate in the final sections. The results concerning the sensitivity and precision criteria were also high, indicating that the scheme was accurate in both correctly identifying dropouts and avoiding completer misclassifications. Additionally, compared to other studies reported in the literature, the proposed method was found to be more accurate.

The proposed method is expected to facilitate instructors in promptly identifying at-risk students and focusing on their needs, thus increasing e-learning retention rates. In the future, the possible application of the proposed dropout prediction method on other types of courses besides e-learning, such as blended learning, distance and classical education, will also be investigated. Another issue which should also be investigated is the potential of achieving better results using different student attributes. Moreover, more techniques could also be examined in terms of individual and combined performance for dropout prediction. Finally, an issue which should be investigated in the future is the potential of incorporating the proposed method in the retention strategies of educational institutions to help increase student retention.

6. References

Allen, J. and Robbins, S. (2008). Prediction of College Major Persistence Based on Vocational Interests, Academic Preparation, and First-Year Academic Performance. *Research in Higher Education*, 49(1), 62-79.

Bellaachia, A., Vommina, E. and Berrada, B. (2006). Minel: a framework for mining e-learning logs. *In Proceedings of the 5th IASTED international conference on Web-based education*, Puerto Vallarta, Mexico, pp. 259-263.

Black, E. W., Dawson, K. and Priem, J. (2008). Data for free: Using LMS activity logs to measure community in online courses. *The Internet and Higher Education*, 11(2), 65-70.

Blom, K. and Meyers, D. (2003). Quality indicators in vocational education and training: International perspectives Adelaide. National Centre for Vocational Education Research (NCVER).

Boon, H. J. (2008). Risk or Resilience? What Makes a Difference? *The Australian Educational Researcher*, 35(1), 81-102.

Borrego, A. M. (2002). Today’s news: education department will emphasize retention issues in next higher education act, official says. http://chronicle.com/daily/2002/07/2002070102n.htm. Retrieved: July 4, 2008.

Carpenter, G. A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37(1), 54-115.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H. and Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3(5), 698–713.

Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *The Chronicle of Higher Education*, 46(23), A39-A41.

Carter, V. (1996). Do media influence learning. Revisiting the debate in the context of distance education. *Open Learning*, 11(1), 31-40.

Chen, G., Liu, C., Ou, K. and Liu, B. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research*, 23(3), 305-332.

Chih-Ming, C., Yi-Yun, C. and Chao-Yu, L. (2007). Learning Performance Assessment Approach Using Web-Based Learning Portfolios for E-learning Systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6), 1349-1359.

Cocea, M. and Weibelzahl, S. (2007). Cross-System Validation of Engagement Prediction from Log Files. *Creating New Learning Experiences on a Global Scale*, pp. (14-25).

Cohen, A. and Nachmias, R. (2006). A quantitative cost effectiveness model for Web-supported academic instruction. *The Internet and Higher Education* 9(2), 81-90.

Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273-297.

Diaz, D. (2002). Online drop rates revisited. http://technologysource.org/article/online_drop_rates_revisited/. Retrieved: October 2, 2008.

Diaz, D. P. (2000). Comparison of student characteristics, and evaluation of student success, in an online health education course. PhD dissertation. Nova Southeastern University. http://home.earthlink.net/~davidpdiaz/LTS/pdf_docs/dissertn.pdf. Retrieved: October 8, 2008.

Doherty, W. (2006). An analysis of multiple factors affecting retention in Web-based community college courses. *The Internet and Higher Education*, 9(4), 245-255.

Dupin-Bryant, P. A. (2004). Pre-entry Variables Related to Retention in Online Distance Education. *American Journal of Distance Education*, 18(4), 199-206.
Frankola, K. (2001). Why online learners dropout. *Workforce*, 80(10), 53-63.
Hagan, M. T. and Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), 989-993.
Haykin, S. (1999). *Neural Networks A Comprehensive Foundation*. Prentice-Hall.
HEFCE. (2007). Review of performance indicators Outcomes and decisions. Higher Education Funding Council of England.
Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-ΰ-vis regression. *New Directions for Institutional Research*, 2006(131), 17-33.
HLC (2007). Best practices for electronically offered degree and certificate programs. Regional Accrediting Commissions. www.ncahlc.org/download/Best_Pract_DEd.pdf. Retrieved: September 2, 2008.
Holder, B. (2007). An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education online programs. *The Internet and Higher Education*, 10(4), 245-260.
Hung, J. L. and Zhang, K. (2008). Revealing Online Learning Behaviors and Activity Patterns and Making Predictions with Data Mining Techniques in Online Teaching. *Journal of Online Learning and Teaching*, 4(4),
Juan, A. A., Daradoumis, T., Faulin, J. and Xhafa, F. (2008). Developing an Information System for Monitoring Student's Activity in Online Collaborative Learning. *In Proceedings of the Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on*, pp. 270-275.
Kasuba, T. (1993). Simplified Fuzzy ARTMAP. *AI Expert*, 8(11), 18–25.
Kotsiantis, S., Pierrakeas, C. and Pintelas, P. (2003). Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. *Knowledge-Based Intelligent Information and Engineering Systems*, pp. (267-274).
Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48(2), 185-204.
Lin, X., Yacoub, S., Burns, J. and Simske, S. (2003). Performance analysis of pattern classifier combination by plurality voting. *Pattern Recogn. Lett.*, 24(12), 1959-1969.
Loo, C. K. and Rao, M. V. C. (2005). Accurate and Reliable Diagnosis and Classification Using Probabilistic Ensemble Simplified Fuzzy ARTMAP. *IEEE Transactions on Knowledge and Data Engineering*, 17(11), 1598-1593.
Mazza, R. and Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, 65(2), 125-139.
Medialab. (2008). E-Learning Services, Multimedia Technology Laboratory, National
Technological University of Athens. http://elearn.medialab.ntua.gr/. Retrieved: September 23, 2008.
Mendez, G., Buskirk, T. D., Lohr, S. and Haag, S. (2008). Factors Associated With Persistence in Science and Engineering Majors: An Exploratory Study Using Classification Trees and Random Forests. *Journal of Engineering Education*, 97(1),
Mendling, J., Neumann, G., Pinterits, A. and Simon, B. (2005). Revenue Models for E-Learning at Universities. *Wirtschaftsinformatik 2005*, pp. (827-846).
Moodle. (2008). Moodle LMS. http://moodle.org/. Retrieved: September 20, 2008.
Morris, L. V., Wu, S.-S. and Finnegan, C. L. (2005). Predicting Retention in Online General Education Courses. *American Journal of Distance Education*, 19(1), 23-36.
Moseley, L. G. and Mead, D. M. (2008). Predicting who will drop out of nursing courses: A machine learning exercise. *Nurse Education Today*, 28(4), 469-475.
Newell, C. (2007). Learner Characteristics as Predictors of Online Course Completion Among Nontraditional Technical College Students. Ed.D Dissertation. University of Georgia http://www.coe.uga.edu/leap/adminpolicy/dissertations_pdf/2007/newell_2007_edd.pdf. Retrieved: 15 October 2008.
Parker, A. (2003). Identifying predictors of academic persistence in distance education. *USDLA Journal*, 17(1), 55-62.

Roblyer, M. D., Davis, L., Mills, S. C., Marshall, J. and Pape, L. (2008). Toward Practical Procedures for Predicting and Promoting Success in Virtual School Students. *American Journal of Distance Education*, 22(2), 90-109.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986a). Learning internal representations by error propagation. *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, pp. (318-362). MIT Press.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature*, 323(1), 533-536.

Simpson, O. (2004). E-Learning, Democracy, and Social Exclusion: Issues of Access and Retention in the United Kingdom. *Global Perspectives on E-Learning: Rhetoric and Reality*, pp. (89-100). Sage Publications, Inc.

Terkla, D. G. (2001). Competencies, Regional Accreditation, and Distance Education: An Evolving Role? *New Directions for Institutional Research*, 2001(110), 65-81.

Ukpabi, C. V. (2004). Formulating a Prediction Model of Retention Rate in the University of North Carolina System. PhD dissertation. North Carolina State University. http://www.lib.ncsu.edu/theses/available/etd-04262005-144219/. Retrieved: 15 October 2008.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley & Sons.

Vare, J. W., Dewalt, M. W. and Dockery, R. E. (2000). Predicting Student Retention in Teacher Education Programs. Annual Meeting of the American Association of Colleges for Teacher Education, Chicago, U.S.A.

Wegner, L., Flisher, A. J., Chikobvu, P., Lombard, C. and King, G. (2008). Leisure boredom and high school dropout in Cape Town, South Africa. *Journal of Adolescence*, 31(3), 421-431.

Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers & Education*, 43(4), 345-359.

Xenos, M., Pierrakeas, C. and Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education*, 39(4), 361-377.

Zaiane, O. R., Man, X. and Jiawei, H. (1998). Discovering Web access patterns and trends by applying OLAP and data mining technology on Web logs. *In Proceedings of the IEEE International Forum on Research and Technology Advances in Digital Libraries*, pp. 19-29.

Zhang, G., Anderson, T., Ohland, M., Carter, R. and Thorndyke, B. (2003). Identifying Factors Influencing Engineering Student Graduation and Retention: A Longitudinal and Cross-Institutional Study. American Society for Engineering Education Annual Conference, Nashville, Tennessee, U.S.A.