

## Paper 12

### The Use of Classification Techniques to Enrich e-Learning Environments.

Aski, B. A. and Torshizi, H. A. Islamic Azad University, Iran

#### Abstract

In recent years many data mining applications have been developed to mine and classify the learner's records and characteristics in an E-learning environment, in order to help the learners to predict their studying results. In this article we examined and compared the predictions' results of four classification methods which were used to classify and analyze the learner's information and found out that those methods which had used the Simple Bayesian or Decision Tree Algorithms had more accurate results and can be used as useful agents for leading the learners to have better improvements in an E-learning environment.

#### Introduction

These days, thousands of web based e-courses are offered to the e-learning candidates, regardless of their natural ability, skill or powers. If we pay attention to the learner's abilities and based on these abilities find some ways to predict the results of their learning studies, we would be able to help them to choose more related courses in order to have better results ( Beck and Woolf , 2000). Data mining or "knowledge of automatic deciphering of useful patterns from large data collections" can be used to classify the learners' information and to find out the more useful courses for that learners which could be more beneficial to them. In the following sections of this article we have some brief description about: data mining, classification, prediction of learners' studying result and our experiments of using different classification techniques for the learners' studying prediction and finally our conclusion of these experiments, which shows that if we use simple Bayesian or Decision Tree as our classification technique, we would have more precision in the learners' studying prediction.

#### Data mining

The amount of saved data in data warehouses has been rapidly increased. So one needs a useful method for automatic and intelligent organization of large data collection. This point has been led to data mining or Knowledge Discovery in Database (KDD). Data mining is a process to mine and organize data in useful and coherent collections ( Han and Kamber ,2006). Data mining is sometimes used to discover and show some knowledge in an understandable form. The aim of data mining is description and prediction. There are many strategies in data mining which can be led to the prediction. One of them is *classification*.

#### Classification

One of the data mining strategies is data classification. The aim of classification is to separate different data into different Pre-defined classes. Classification is based on available features that leads to new data description and causes a better understanding of each class in a Database or in a Data Warehouse, so classification can prepare a model to describe the proper class for any given data. In other words by using classification, we can predict that which given data would belong to which predefined class. Different statistical techniques are used for classification functions like; Bayesian, Neural Network, Decision Tree and Support Vector Machine.

#### Prediction of learner studying result using classification

The servers which give service to a Learning Management System (LMS) in an E-learning environment can log most of the learners learning behaviours and can also make a large amount of profiles, which can be classified in meaningful and useful classes. It is necessary to predict the learner's studying result in future according to their given abilities. If a teacher can predict learners studying result, he or she will be able to give a suitable learning plan for the learner and will control weak learners studying progress and also will help more talented learners. Romero et al (2003) suggested a methodology for mining data in the system log files to discover proper patterns, for example this pattern can be regarded as a relationship between learners' learning level and his or her grades. Muehlenbrock (2005), showed a system for automatic analysis of learners actions in a web based learning environment by which the learners' future usages will be predicted. In Dutton et al (2001), the final

grades of learners are predicted by neural network .Neural networks have also been used in Bernardete and Alberto (2007) for making models to discover user's behaviour patterns. The outcome of this research showed that if we use user's behaviour reactions, we will predict learners studying result successfully.

Moodle is popular open source software for Learning Management System. This system has some modules for logging user behaviours and characteristics such as user ID, IP, time of user connection to the system, user actions and also user scores. Delgado et al (2006) showed how they could use the Moodle log files of 240 learners to discover knowledge for better management of learning sources.

## Our Experiments

The information of 352 learner of an electronic learning system has been used in our experiments as a collection. The minimum, maximum and average age of the learner in this collection is 24, 46 and 31.2 respectively. Learners of this collection had to do three of four homework during the term. If they could do it; they would be able to take their final exam, otherwise they missed their final exam. Characteristics of these learners are shown in table 1. At the final exam of these learners, 155 students were absent or had been rejected. Table 1 shows learners' characteristics chosen for this experience and their related values:

**Table1: Learners 'Characteristics and their related Values**

<b>Learner Characteristics</b>	<b>Related Values</b>
Sex	Male , Female
Age	24-49
No of Children	0,1,2,3,4
Marriage Status	Single , Married
Computer Knowledge	Yes, No
Working with Computer	Yes, No
First Homework done?	Yes, No
First Exam Result	Excellent, Good , Absent, Rejected
Second Homework done?	Yes, No
Second Exam Result	Excellent, Good , Absent, Rejected
Third Homework done?	Yes, No
Third Exam Result?	Excellent, Good , Absent, Rejected
Fourth Homework done?	Yes, No
Fourth exam Result	Excellent, Good , Absent, Rejected
Final Exam Result	Accepted, Absent or Rejected

Selection of the effective factors for evaluation of learning machine methods, was one of the important part of this experiment .We used the Cross Validation method for 10 learner groups to evaluate the classifier's quality. Ninety percent of the total collected data were used as the training set for learning machine and the other ten percent of the total learning data were used as the test data. The amount of samples that were classified correctly, were attended as a basis for determining its accuracy. This process was done by Weka software for ten times and each time, 10 percent of the learning data were used. When this process was done completely, the average amount of accuracy in each part was considered as the final accuracy of the related method.

### Comparison of the prediction precisions using different statistical techniques

We have used different statistical techniques of classification to predict the learners studying results, based on their final exam results. These techniques were consisting of Decision Tree, Neural Networks, Simple Bayesian, and Support Vector Machine. When we used the Decision Tree as our classification technique, the result of our experiment showed that 282 cases from 352 are correctly classified and the result precision of our prediction is 80.113%.

**Table 2. Results of the prediction using the Decision Tree technique**

Description of related class	Precision(Percent)
Precision of accepted class	0.759
Precision of rejected or absent class	0.843
Recall of rejected or absent class	0.792
Recall of accepted class	0.813
F-Measure of accepted class	0.783
F- Measure of rejected or absent class	0.817
Final prediction accuracy of this method	80.113

When we used the Simple Bayesian technique for classification, the result shows that again 282 cases are correctly classified and the result precision of our prediction is as before (80.113 %).

**Table 3. Results of the prediction using the Simple Bayesian technique**

Description of related class	Precision(Percent)
Precision of Accepted Class	0.746
Precision of rejected or absent class	0.855
Recall of rejected or absent class	0.832
Recall of accepted class	0.777
F-Measure of accepted class	0.787
F- Measure of rejected or absent class	0.814
Final prediction accuracy of this method	80.113

When we used the Support Vector Machine technique for classification, the result shows that 276 cases from 352 are correctly classified and the result precision of our prediction is 78.409%.

**Table 4. Results of the prediction using the Support Vector Machine technique**

Description of related class	Precision(Percent)
Precision of Accepted Class	0.739
Precision of rejected or absent class	0.824
Recall of rejected or absent class	0.787
Recall of accepted class	0.82
F-Measure of accepted class	0.763
F- Measure of rejected or absent class	0.782
Final prediction accuracy of this method	78.409

And finally when we used the K Nearest Neighbour technique for classification, the result shows that 255 cases from 352 are correctly classified and the result precision of our prediction is 72.443%.

**Table 5. Results of the prediction using the K Nearest Neighbour technique**

Description of related class	Precision(Percent)
Precision of Accepted Class	0.679
Precision of rejected or absent class	0.763
Recall of rejected or absent class	0.71
Recall of accepted class	0.736
F-Measure of accepted class	0.694
F- Measure of rejected or absent class	0.749
Final prediction accuracy of this method	72.443

By comparing the final prediction accuracy of the above mentioned tables ,we can see that when we use the Decision Tree or Simple Bayesian techniques for our classifications, we would acquire more precision in our learners' studying result predictions.

**Table 6. Comparison of results**

Statistical technique of the Classification	Prediction Precision
Decision Tree	80.113%
Simple Bayesian	80.113%
Support Vector Machine	78.409%
K Nearest Neighbor	72.443%

### Conclusion

Using the classification method for predicting the learner studying results in an E-learning environment causes the E-learning system to be potentially enriched and to be more beneficial to the users. An intelligent E-learning system tries to analyze the logged information of the learners, and then based on this information, tries to make a model to predict learner studying results and according to this prediction it helps the learner to choose more related courses for their improvement. In this article we described a prediction model based on the classification method for predicting the learners studying results in an E-learning system and we showed the prediction result precisions of this model with four different statistical techniques which were used in our classification model . As the outcome of this experiment we found that if we use the Decision Tree or Simple Bayesian techniques as our classification statistical methods we would have more precise prediction results.

### References

- Beck, J.E., Woolf, B.P. (2000) *High-Level Student Modeling with Machine Learning* in Gauthier, G., et al. (eds.): *Intelligent Tutoring Systems*, ITS 2000. Lecture Notes in Computer Science, Vol. 1839. Springer-Verlag, Berlin Heidelberg New York.
- Bernardete R. and Alberto C. (2007) Behaviour Pattern Mining during the Evaluation Phase in an e-Learning Course, *International Conference on Engineering Education – ICEE 2007*.
- Delgado M. , Calvo F. , Gibaja E., Pegalajar C. and Pérez O. (2006);”*Predicting students’ marks from Moodle logs using neural network models*”, proc. Current Developments in Technology-Assisted Education.
- Dutton,J. ,Dutton ,M., and Perry,J. (2001) Do Online Students Perform as Well as Lecture Students ?”, *Journal of Engineering Education*.
- Han, J. & Kamber M. (2006) , *Data Mining : Concepts and Techniques* , Elsevier Science & Technology Books , ISBN-13: 9781558609013

Muehlenbrock, M. *Automatic action analysis*, in an interactive learning environment., 2005.

Muehlenbrock, M. (2005);” *Automatic Action Analysis in an Interactive Learning Environment*”, In: 12th International Conference on Artificial Intelligence in Education, AIED 2005. July 18, Amsterdam, The Netherlands.

Romero, C., Ventura, S., De Bra, P., De Castro, C (2003) Discovering Prediction Rules, in AHA! Courses. In: *User Modelling Confer*