

## Analyzing E-Learning Systems Using Educational Data Mining Techniques

Anduela Lile

*Epoka University, Tirana, Albania*

*E-mail: [anduela.lile@gmail.com](mailto:anduela.lile@gmail.com)*

---

**Abstract** Recently, Educational Data Mining has become an emerging research field used to extract knowledge and discover patterns from E-learning systems. The educational system in Albania is currently facing a number of issues such as identifying students' needs, personalization of training and predicting the quality of student interactions. Educational Data Mining provides a set of techniques, which can help the educational system to overcome these issues. The objective of this research is to introduce Educational Data Mining, by describing a step-by-step process using a variety of techniques such as Attribute Weighting (Weighting by Information Gain, Relief, Hi-Squared, Uncertainty), Clustering (K-Means), Classification (Tree Induction), Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) in order to achieve the goal to discover useful knowledge from the Moodle LMS. Analyzing mining results enables educational institutions to better allocate resources and organize the learning process in order to improve the learning experience of students as well as increase their profits. The experimental results have shown that the data mining model presented in this research was able to obtain comprehensible and logical feedback from the LMS data describing students' learning behavior patterns. For this work, Rapid Miner (v5.0) and Weka (v3.6.2) data mining tools were used to mine data from the Moodle system, used in "C Programming - CEN112" course taken by Computer Engineering students at Epoka University, during Spring Semester 2009-2010.

**Keywords:** Educational Data mining, E-learning, LMS, moodle, learning patterns

---

### 1. Introduction

#### 1.1 Background

This work describes the application of data mining techniques to the usage data of the Moodle course management system, a case study of Epoka University in Tirana, Albania. The educational system in Albania is currently facing several issues such as identifying students' needs, personalization of training and predicting the quality of student interactions.

Educational Data mining (EDM) provides a set of techniques, which can help the educational system to overcome these issues. EDM uses powerful tools that enable educational institutions to better allocate resources in order to improve the learning experience of students as well as increase their profits. Moodle has been used as a LMS platform for sharing useful information, and knowledge management, enabling the achievement of valuable results (Tirado, Estrada, & Castro 2007, Romero & Ventura 2007). However, Moodle LMS does not cover all teaching and learning aspects since it doesn't provide tools to monitor and evaluate all the activities performed by learners (Zaiane & Luo, 2001; Mazza & Dimitrova, 2007).

Several studies have demonstrated that Data Mining techniques could successfully be incorporated into E-learning environments. The application of data mining techniques and concepts in e-Learning systems helps to support educators to improve the e-Learning environment. Data mining techniques have also been addressed as complementary systems to LMS, and in particular to Moodle, where results are achieved through the use of associates, classifiers, clusters, pattern analyzers, and statistical tools (Romero, Ventura, & García, 2007). The scope of data mining is to discover useful knowledge using a variety of techniques such as prediction, classification, association rule mining, clustering, fuzzy logic, etc. Recently, Educational data mining has become an emerging research field used to extract knowledge and

discover patterns from E-learning systems.

The objective of this work is to introduce Educational Data Mining, by describing a step-by-step process using a variety of techniques such as Attribute Weighting (Weighting by Information Gain, Relief, Hi-Squared, Uncertainty), Clustering (K-Means), Classification (Tree Induction), Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) in order to achieve the goal to discover useful knowledge from the Moodle LMS. Analyzing mining results enables educational institutions to better allocate resources and organize the learning process in order to improve the learning experience of students as well as increase their profits. The mining results can be used by the course instructor to investigate the impact of a number of e-learning activities on the students' learning development.

## 2. Literature Review

### 2.1 A Theory for E-Learning

E-learning system is referred to as Learning Management System (LMS), Course Management System (CMS), Learning Content Management System (LCMS), Managed Learning Environment (MLE), Learning Support System (LSS), Web Based Training System (WBT-System). These systems accumulate a great amount of information that can be further processed in analyzing students and educators' behavior. e-learning Systems offer the facilitation of communication between students and educators, sharing resources, producing content material, preparing assignments, conducting online tests, enabling synchronous learning with forums, chats, news services, etc. Some examples of commercial E-learning systems are Blackboard (2010), TopClass (2010), etc. and some examples of open systems are Moodle (2010), Ilias (2010), Caroline (2010), Dokeos (2010), eFront (2010), Olat (2010) etc.

### 2.2 Educational Data Mining

Due to the large quantities of data that E-learning systems can generate, it is very difficult for educators to analyze them manually. EDM was defined by Baker (Baker, 2010) "as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in". "Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." as defined by The Educational Data Mining community. This research community has started by organizing workshops since 2004, then conducting an annual International Conference on Educational Data Mining. The first International Conference on Educational Data Mining (EDM'08) was held in Montreal, QC, 2008, followed by a second in Cordoba, Spain (July 2009), the third in Pittsburgh, PA, USA (June 2010), and attending the fourth in Eindhoven, the Netherlands), and now already publishing a Journal on Educational Data Mining since October 2009.

An extensive literature review on Educational Data Mining (Romero & Ventura, 2007) providing an overview of the research efforts in the field between 1995 and 2005, review of the many applications of Data Mining to E-learning over the period 1999-2006 (Castro, Vellido, Nebot, & Mugica, 2007); review of the history and current trends of EDM (Baker & Yacef, 2009), review of most relevant studies carried out in this field to date (Romero & Ventura, 2010). Data Mining techniques relation with the fields of Artificial Intelligence (AI) and Machine Learning (ML) have been highlighted in many researches (Sison & Shimura, 1998; Baker, 2000, Ha, Bae, & Park, 2000; Margo, 2004; Xindong, 2004; Tang & McCalla, 2005). Main applications of EDM as defined by Baker (2010) are classified into four main areas: 1) improving student models, that provide detailed information about a student's characteristics; 2) discovering models of the knowledge structure of the domain; 3) studying the pedagogical support provided by learning software; 4) scientific discovery about learning and learners. The first three categories are universal across different fields of data mining, the fourth and fifth categories are particularly related to educational data mining. EDM has the following categories (Romero & Ventura, 2007):

- Statistics and visualization
- Web mining (Clustering, Classification, Outlier detection Association rule mining and Sequential pattern mining)

- Text mining
- According to Baker (2010), the educational data mining has been classified as:
- Prediction (Classification, Regression , Density estimation)
  - Clustering
  - Relationship mining (Association rule mining, Correlation mining, Sequential pattern mining, Causal data mining )
  - Distillation of data for human judgment
  - Discovery with models

The table below introduces a review of the available literature according to different topics for data mining in E-learning. Table 1 includes the following information: reference, Data Mining techniques used (DM techniques), goal description, E-learning actors involved, and type of publication: Journal (J), International Conference (C), or Book Chapter (B).

**Table 1.** *A review of Educational Data Mining literature*

Reference	DM Techniques	Goal Description	E-learning actor	Type of Publication
(Hwang, 1999)	Classification (Fuzzy reasoning)	Network-based intelligent testing and diagnostic system that can be used to guide the students to raise their learning status	Student	J
(Hogo, 2010)	Classification Clustering	Comparative study based on the use of different fuzzy c-means	Student	J
(Matsui, 2003)	Classification	Learning Process Assessment under the e-Learning Environment applying ID3 algorithm	Teacher	C
(Hwang, 2003)	Classification	Applying Concept effect relationship, (CER) model to Diagnosing Student Learning Problems in Science Courses.	Teacher	J
(Kotsiantis Pierrakeas, Pintelas, 2004)	Classification and Prediction	Using Naïve Bayes, kNN, MLPANN, C4.5, Logistic Regression and SVM to Predict	Teacher	J
(Yoo, J., Yoo, S., Lance, Hankins, 2006)	Classification and visualization	Student Progress Monitoring Tool Using Treeview (ADT tree)	Student, Teacher	C
(Muehlenbrock, 2005)	Prediction	Automatic Action Analysis in an Interactive Learning Environment using	Teacher	C
(Wang & Shao, 2004)	Clustering Classification	Time-framed navigation clustering and association mining for Expert Systems	Student	J
(Rodrigo,Anglo, Sugay, Baker, 2008)	Clustering	Unsupervised Clustering to Characterize Learner Behaviors	Student	C
(Tsai, Tsen., & Lin, 2001)	Classification	Integrating Apriori algorithm, fuzzy set theory and inductive learning (AQR algorithm) for Adaptive Learning Environment	Teacher	C
(Jeong, Biswas, Johnson, & Howard, 2010)	Statistical Analysis	Applying hidden Markov model analysis, to	Student	C

		investigate how the high-performing students transitioned through the different phases in the system in contrast with the low-performing student		
(Falakmasir & Habibi, 2010)	Classification Decision tree	Applying C4.5 algorithm in order to rank the students activities based on their impact on the performance of students in final exams.	Student	C
(Rajibussalim, 2010)	Statistical analyses Clustering Classification	Using statistical analyses, clustering and classification techniques to identify learning behaviors	Student	C
(Romero, Ventura, & Garcia) 2007)	Statistics, Visualization, Classification, Clustering, Association Rule Mining	Mining E-learning using statistics, visualization, classification, clustering and association rule mining of Moodle data.	Student	J
(Romero, Ventura, Vasilyeva, & Pechenizkiy, 2010)	Association Rules Mining	Class Association Rules Mining from Students' Test Data	Student	C
(Dekker ,Pechenizkiy, & Vleeshouwers, 2009)	Classification	Case study aimed at predicting the Electrical Engineering (EE) students drop out after the first semester applying decision trees	Student	C
(Romero et al. 2009)	Genetic Algorithms	Grammar guided genetic programming algorithm, G3P-MI, has been applied to predict if the student will fail or pass and identifies activities to promote learning	Student	C
(Amershi & Conati, 2009)	Classification	Framework for user Modeling that uses both unsupervised and supervised clustering to discover and capture effective or ineffective student behaviors	Student	J
(Bidgoli & Punch, 2003)	Classification Genetic Algorithms	Genetic Algorithms for Data Mining Optimization	Student	C
(Talavera & Gaudioso, 2004)	Clustering Classification	Defining similar behavior groups in unstructured collaboration spaces	Student	C

Data Mining Techniques for Associations, Clustering and Classification are introduced in (Aggarwal & Yu, 1999). Surveys of Clustering Algorithms can be found in (Xu & Wunsch, 2005; Berkhin, 2006).

### 2.3 Process of Data Mining in E-Learning

Knowledge Discovery and Data mining (KDD) process is interactive and iterative; a brief description about each process as proposed by (Cios, Pedrycz, & Swiniarski 1998):

- Define the Goal of KDD Process.
- Selection of Target Dataset.

- Data cleaning and preprocessing.
- Data reduction and transformation.
- Define a particular data-mining method.
- Exploratory analysis and hypothesis selection.
- Data mining: Searching for hidden patterns of interest.
- Interpreting mined patterns.
- Acting on the discovered knowledge.

The application of data mining in educational systems has specific requirements, mainly the need to take into account learners' specific behavior, including pedagogical aspects (Romero & Ventura 2007). The application of data mining in E-learning systems can be described as an iterative cycle (Romero, Ventura, & Garcia, 2007) where data mining applications contribute in enhancing learning, and also using mined knowledge for decision making. The E-learning data mining process consists of the same four steps (Romero & Ventura, 2007) in the general data mining process as follows:

1. Collect data. Interaction information is stored in the database of the LMS.
2. Preprocess the data. The data is transformed into an appropriate format.
3. Apply data mining. The data mining algorithms are applied to create and execute the model that discovers the knowledge and patterns of interest. In order to achieve this goal a data mining tool can be used.
4. Interpret, evaluate and deploy the results. The model obtained is interpreted and used by the educator for further analyses. The educator can use the information discovered to make decisions about e-learning system and process.

#### *2.4 Data Mining Tools for Learning Management Systems*

Several data mining tools which assist educator to extract knowledge and discover patterns can be used. Such tools can be generic or specific, commercial or open source. Some Examples of commercial data mining tools are: DBMiner, IBM SPSS Modeler (former SPSS Clementine), DB2 and Intelligent Miner (Kdnuggets, 2011) etc.

Examples of open source tools are Rapid Miner, Weka, Keel (Kdnuggets, 2011) etc. In order to offer a better and more flexible service to educators there have been developed several specific educational data mining tools: Tools for association and classification and text mining – The Mining tool, EPRules, Simulog, Sequential Mining tool, O3R and KAON (Zai'ane & Luo, 2001), MultiStar and CIECoF (Silva & Vieira, 2002); Tools for statistics and visualization-Synergo/ColAT, GISMO, Listen tool TADAEd (Mazza & Milani, 2005) etc

### **3. Research Methods**

#### *3.1 Participants*

In this study, some EDM techniques were applied in the “C Programming - CEN112” course taken by 29 Computer Engineering students at Epoka University, during Spring Semester of 2009-2010. Blended Learning was applied for this course.

#### *3.2 Procedures and Instruments*

Although Moodle LMS offers several reports on the students' activities, they are not flexible enough to satisfy the educators' needs for evaluating their interactions with the system (Mazza & Dimitrova 2007). The research methodology of this study has four components: (1) select and prepare a data set through summarization tables, (2) select data mining algorithms set, (3) perform data-mining experiments, (4) analyze experimental results (Romero & Ventura 2007). The goal was to understand and prepare the required data and apply data mining techniques that allow teachers to detect patterns of use for further evaluation. This work is based on log analysis of Moodle LMS by applying data mining techniques to student's usage of data in order to evaluate web activity and to acquire knowledge about how the students learn on the E-learning environment.

Data mining techniques are used to build relationships among Moodle LMS system components (Uribe-Tirado, A., Melgar-Estrada, L.-M., & Bornacelly-Castro, J.-A. ,2007 ; Romero, Ventura, & Garcia

2007). Rapid Miner (v5.0) and Weka (v3.6.2) data mining tools are used to extract knowledge from the Moodle system. Data for the initial set of experiments in this study was derived from the Moodle logs. These data were analyzed from various levels and perspectives, in order to provide more insight in the overall educational system. From the level of an individual course user activity is considered, accomplishing assignments, quizzes project assignments, participation in forum and chat. This work focus on adopting data mining approaches for solving various educational data mining tasks by applying Attribute Weighting (Weighting by Information Gain, Relief, Hi-Squared, Uncertainty)), Clustering ( K-Means), Classification(Tree Induction) and Association Mining ( Apriori, FPGrowth, Create Association Rule, Generalized Sequential Patterns)

### 3.3 Applying Statistics and Visualization Techniques to Moodle Data

Moodle does not provide visualization tools but it offers log reports. Several graphical system tools can extract tracking from Moodle data and visualize them graphically. Moodle offers some statistical information in some of the modules (grades and quizzes) and has statistical quiz reports which show item analysis for representing the performance of each question for the function of assessment.

### 3.4 Mining Moodle Logs

Web usage mining, based on Moodle log files, tries to discover usage patterns by making use of data mining techniques with the purpose of understanding and better serving the user and the application itself.

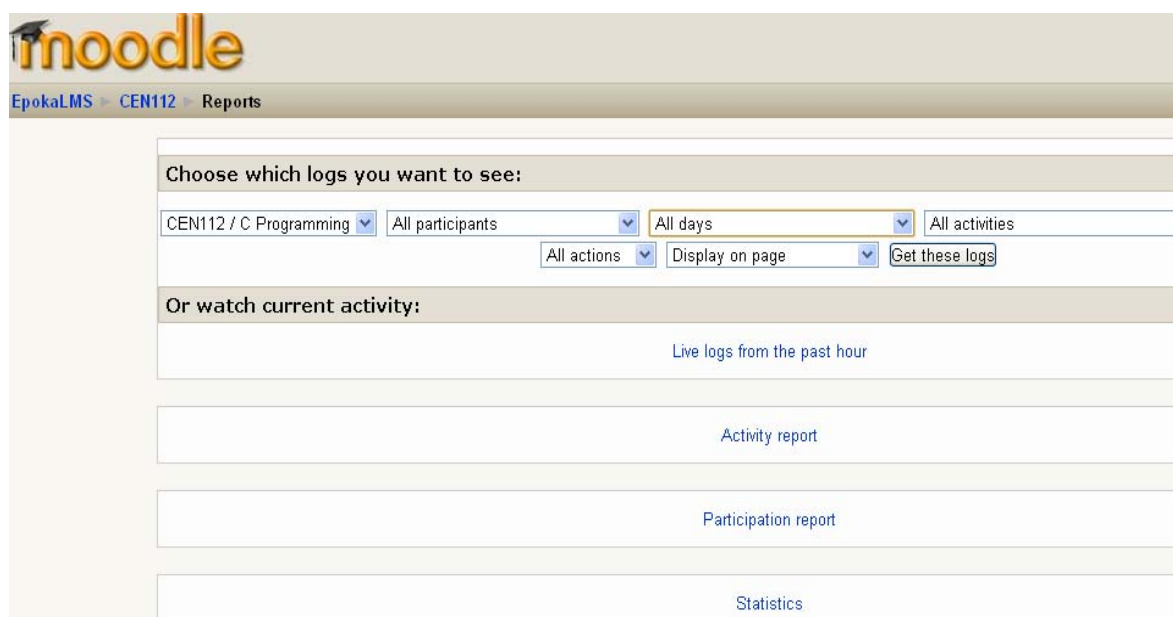
Table 2 represents the summarization table created from the Moodle logs tables such as: mdl\_user, mdl\_user\_students, mdl\_log, mdl\_assignment, mdl\_assignment\_submission, mdl\_chat, mdl\_chat\_user, mdl\_forum, mdl\_forum\_posts, mdl\_forum\_discussion, mdl\_message and mdl\_quiz.

**Table 2.** Summarization table of students' activities

Name	Description
UserName	Name of User
ResourceView	Number of Courseware and Other Supporting Materials Views (action='resource view')
ChatTalk	Number of Chat Messages (action='chat talk')
BlogView	Number of Blog Views (action='blog view')
ForumRead	Number of Forum Reads ( action='view forum')
ForumPost	Number of Forum Posts (action='forum add post')
DiscussionRead	Number of Discussion Reads (action='forum view discussion')
AssignmentView	Number of Assignments Views Participations ( action='assignment view')
AssignmentUpload	Number of Assignment Answer Uploads (action=assignment upload')
QuizScore	Quiz Result (view quiz report)
AssignmentScore	Assignment Result
FinalMark	Final Mark

### 3.5 Ranking Moodle Activities

Moodle stores detailed records of students' activities and the educator can access summarized reports about these activities according to the categories specified by the Moodle system (see Figure 1).



**Figure 1.** *Displaying summarized reports of Student logs in Moodle.*

Although there are several activities such as messaging, forums and text chat to support collaboration of teachers and students, a full use of synchronous collaboration tools in Moodle is still missing. We have to take into consideration that it was the first semester that the Moodle LMS was used at Epoka University. The main activities are classified into: resource view, virtual classroom participation, archive view, assignment view, forum read, forum post, assignment upload, discussion read, and discussion post. The ranking table of students activities was created using Information gain, Relief, Hi-Squared, and Uncertainty (Quinlan 1993, Kononenko 1994) as the main activity evaluation metrics.

### 3.6 Applying Clustering Techniques

In E-learning, clustering has been used for finding clusters of students with similar learning characteristics and to promote group-based collaborative learning and to provide learner diagnosis (Tang & McCalla, 2005). Rapid Miner system has several clustering algorithms available. The K-Means (MacQueen, 1967), has been used here. Clustering techniques apply when the instances of data are to be divided into natural groups. In k-means algorithm (Hartigan 1975; Hartigan & Wong, 1979) clusters are specified in advance prior to application of the algorithm. (Berkhin, 2006) provides a good review of different data mining clustering techniques.

### 3.7 Applying Classification Techniques

In this case, the objective is to classify students into different groups with equal final marks depending on the activities carried out in Moodle. Rapid Miner system has several classification algorithms available. The Tree Induction operator (similar to C4.5 algorithm (Quinlan, 1993) is used to characterize students who passed or failed the course. The goal is to define a set of IF-THEN-ELSE rules from the decision tree that can show interesting information about the classification of the students. The tree induction algorithm in Rapid Miner works as follows: Whenever a new node is created, an attribute is picked to maximize the discriminative power of that node. Different split evaluation criteria can be selected by the user such as: ratio gain from C4.5, information gain from ID3, the Gini impurity measure from CART etc. (gain ratio was used in this case). The educator can use the knowledge discovered by these rules for making decisions about Moodle course activities and deciding to eliminate some activities related to low marks detect in time if they will have learning problems (students classified as FAIL). Decision Trees are able to estimate the importance of each factor and how it affects the model (Apte & Weiss, 1997).

### 3.8 Applying Association Mining Techniques

Rapid Miner and Weka offer several association mining techniques.

### 3.8.1. Applying Association Rule Mining – Apriori in Weka

Apriori algorithm (Agarwal, Imilienski, & Swami, 1993) was used for finding association rules over moodle tables. Apriori is a seminal algorithm (Agarwal & Srikant, 1994) for finding frequent item sets using candidate generation, by applying an iterative approach (level-wise search), where k-item sets are used to explore (k+1) item sets.

### 3.8.2. Applying FPGrowth and Create Association Rule in Rapid Miner

Rapid Miner has several association rule-discovering algorithms available. In the FPGrowth algorithm, the process of finding frequent item sets involves two steps – candidate item set generation and pruning. In Rapid Miner, the process of frequent item set mining can be divided into main operators: first, the generation of frequent item sets and second, the generation of association rules from these sets. For the generation of frequent item sets without candidate generation, we have used the operator *FPGrowth*. The result will be a set of frequent item sets, which served as input for Create Association Rule operator. This operator will create Association Rules, which express regularities that occur often in the learning and teaching process. The interest is restricted to those rules that predict with a high confidence.

### 3.8.3. Applying Sequential Pattern Mining – GSP in Rapid Miner

Sequential pattern mining tries to discover if the presence of a set of items is followed by another item in a time-ordered set of sessions. It was first introduced in the study of customer purchase sequences by (Agrawal & Srikant, 1995). Sequential pattern mining algorithms (Agrawal & Srikant 1995; Ha, Bae, & Park, 2005) discover inter-session patterns. The input set is focused on raw pattern events or episodes and in order to identify the correct interesting sequences, further analyses with field experts should be considered. The following select query can be used in order to select the required data.

```
SELECT time,ip,userid,module,action,url
FROM moodle.mdl_log
WHERE course_id=5
```

## 4. Results

### 4.1 Ranking Results

The results of activity ranking are obtained using Rapid Miner v5.0 (see Figure 2). Table 3 presents the results obtained from applying gain attribute weighting method on the summarized table of students' activities.

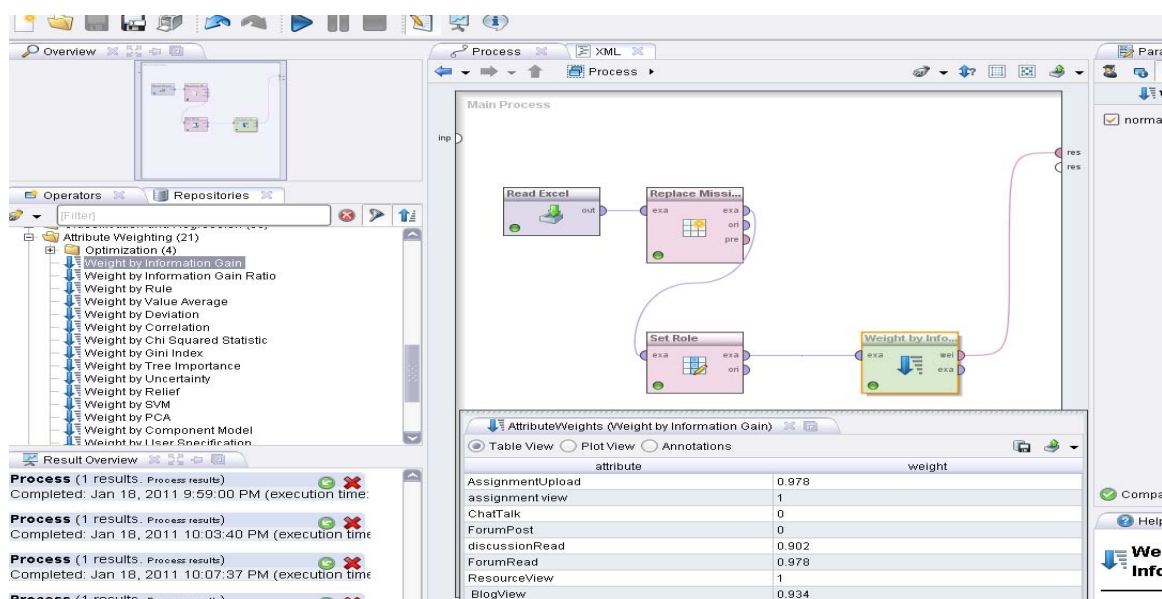


Figure 2. Rapid Miner executing Attribute Weighting Algorithms



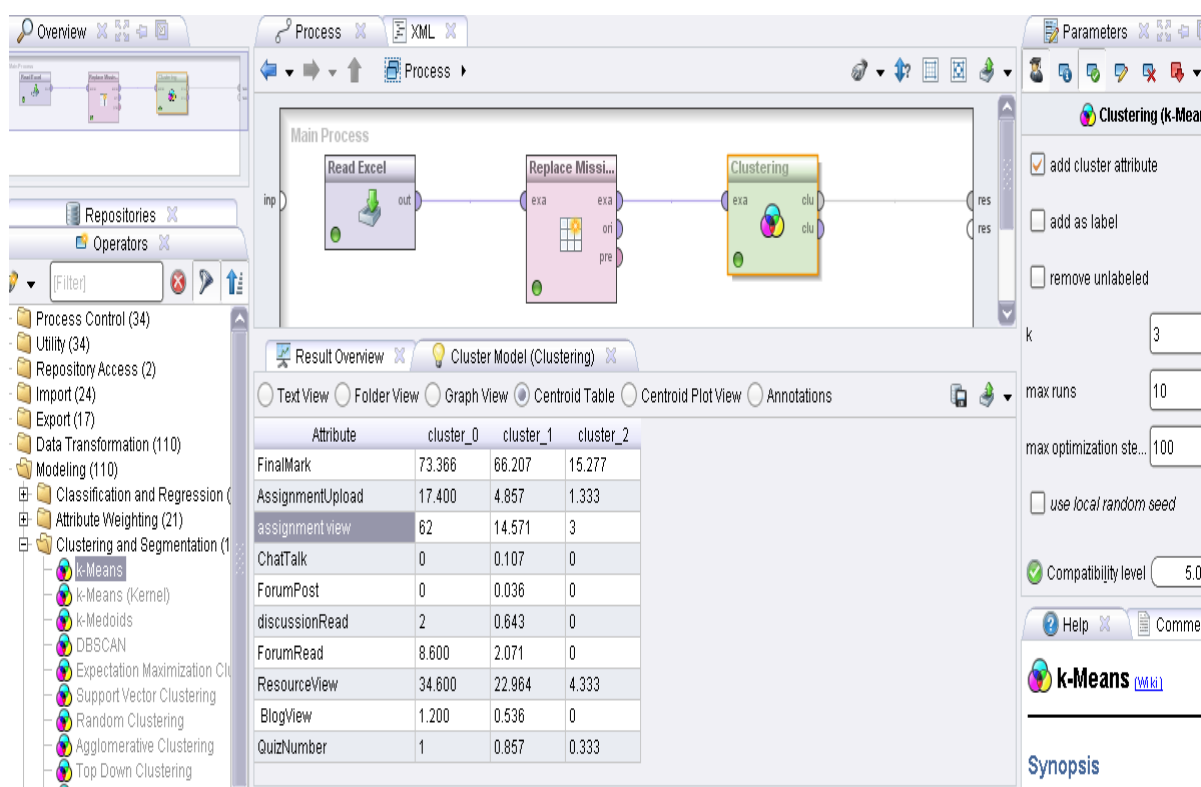
The results show that Resource View has the most dominant role in this ranking while the second place belongs to assignment upload. Results reveal a high weight for Discussion Read and Forum Read but a low weight for Forum Post, which points out that student are interested in such activities but the complementary action (Post) is missing, so the educator should enhance the use of the forum and discussion.

**Table 3.** Attribute Weighting by Information Gain, Relief, Hi-Squared

Activity	Information Gain	Relief	Hi-Squared	Uncertainty
AssignmentUpload	0.98	0.67	0.62	0.82
AssignmentView	1.00	0.78	0.63	0.83
ChatTalk	0.00	0.00	0.00	0.00
ForumPost	0.00	0.00	0.00	0.00
DiscussionRead	0.90	0.69	0.38	0.52
ForumRead	0.98	0.83	0.75	0.69
ResourceView	1.00	1.00	1.00	1.00
BlogView	0.93	0.65	0.37	0.53

#### 4.2 K-Means Results

The educator can use the cluster centroid K-Means results as presented in Table 4 in order to group students into three types of students: very active students (cluster 0), active students (cluster 1) and non-active students (2). This information helps educator to group students for working together in collaborative activities (see Figure 3).



**Figure 3.** Rapid Miner executing K-Means Algorithm for CEN112 Course.

Students were divided in 3 groups based on their activities done in Moodle: Cluster 0 is characterized by most active students in Moodle, with high assignment number, which participated to the online quiz and have a moderately number of discussion and forum read; Cluster 1 is characterized by moderately active students in Moodle, with moderately assignment number, who participated or not to the online quiz; Cluster 2 is characterized by inactive students in Moodle with a low number of actions in the system.

Table 4. Rapid Miner K-Means Centroid Table

Attribute	Cluster 0	Cluster 1	Cluster 2
AssignmentUpload	17.4	4.85	1.33
AssignmentView	62.0	14.57	3.00
ChatTalk	0.00	0.11	0.00
ForumPost	0.00	0.04	0.00
DiscussionRead	2.00	0.64	0.00
ForumRead	8.60	2.07	0.00
ResourceView	34.6	22.96	4.33
BlogView	1.20	0.54	0.00
QuizNumber	1.00	0.86	0.33

4.3 Classification - Tree Induction Results

The Tree Induction algorithm used to predict students' potential performance was focused on using student assessment (quiz score, assignment score and final mark) and Moodle usage (number of hits on Moodle system). Pessimistic pruning was applied to insure that the expected confidence levels obtained for predictions on the training data are similar to actual confidence levels obtained from unseen data (see Figure 4).

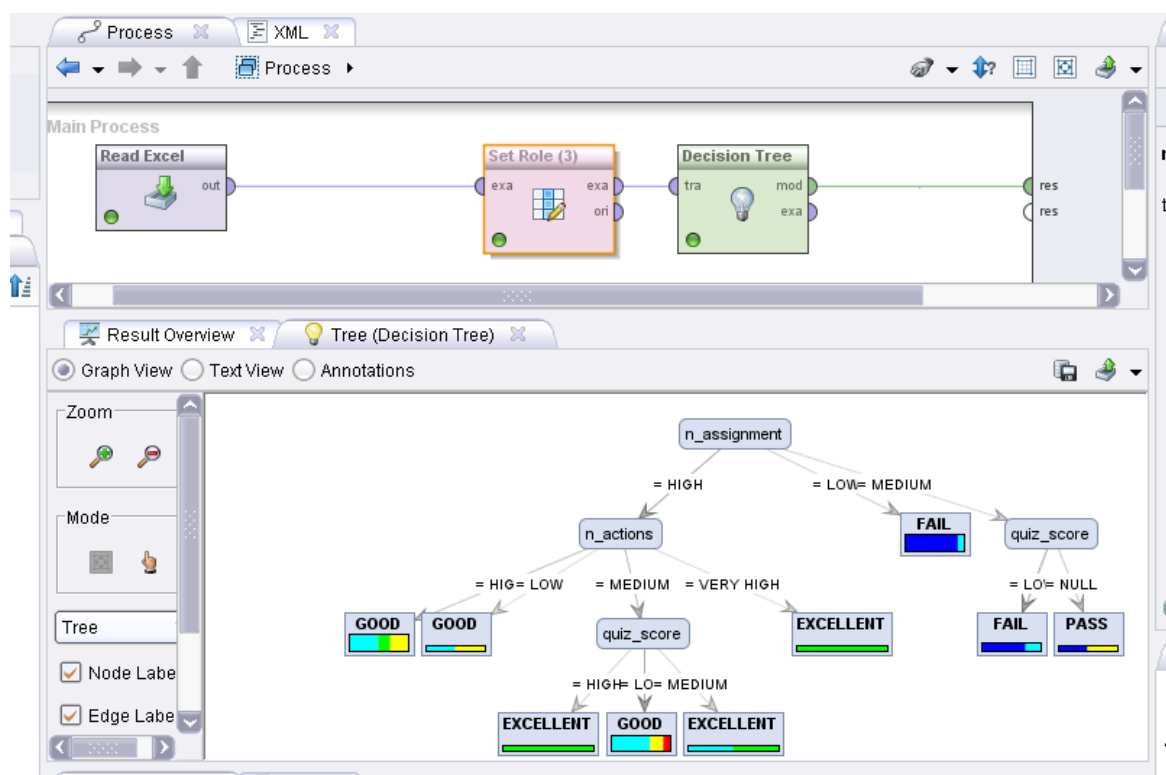


Figure 4 Rapid Miner executing Tree Induction Algorithm

A set of IF-THEN-ELSE rules was generated from the decision tree operator, showing interesting information about the supervised classification of the students. These rules classify at least three main categories of students: students with a low number of assignments are classified as FAIL; students with medium number of assignments quizzes are classified as FAIL or PASS depending on their quiz score, and students with a high number of assignments are classified as FAIL, PASS or EXCELLENT depending on number of actions in the Moodle system etc. The rules generated from the Tree Induction Algorithm operator were:

```

n_assignment=HIGH
| n_actions = HIGH: GOOD {FAIL=0, GOOD=5, EXCELLENT=2, PASS=3, VERY GOOD=0}
| n_actions = LOW: GOOD {FAIL=0, GOOD=1, EXCELLENT=0, PASS=1, VERY GOOD=0}
| n_actions = MEDIUM
| | quiz_score = HIGH: EXCELLENT {FAIL=0, GOOD=0, EXCELLENT=2, PASS=0, VERY GOOD=0}
| | quiz_score = LOW: GOOD {FAIL=0, GOOD=6, EXCELLENT=2, PASS=0, VERY GOOD=1}
| | quiz_score = MEDIUM: EXCELLENT { FAIL=0, GOOD=1, EXCELLENT=2, PASS=0, VERY GOOD=0}
| n_actions = VERY HIGH: EXCELLENT {FAIL=0, GOOD=0, EXCELLENT=2, PASS=0, VERY GOOD=0}
n_assignment = LOW: FAIL {FAIL=9, GOOD=1, EXCELLENT=0, PASS=0, VERY GOOD=0}
n_assignment = MEDIUM
| quiz_score = LOW: FAIL {FAIL=3, GOOD=1, EXCELLENT=0, PASS=1, VERY GOOD=0}
| quiz_score =NULL: PASS { FAIL=1,GOOD=0, EXCELLENT=0, PASS=1, VERY GOOD=0}
    
```

#### 4.4 Association Rule Mining Results

Association Rule Mining (ARM) was applied using two different tools : 1) FPGrowth, 2) Create Association Rule, in Rapid Miner; and the popular Apriori algorithm in Weka. Each rule obtained is accompanied by two meaningful measures, confidence and support, where Confidence measures the percentage of transactions containing X that also contain Y, and support measures the percentage of transactions that contain X or Y.

##### 4.4.1 Frequent Item Set (FPGrowth) Results

The results of FPGrowth (see Figure 5) are a set of frequent item sets which would be used as input for Create Association Rule operator. The most frequent item sets are presented in Table 5.

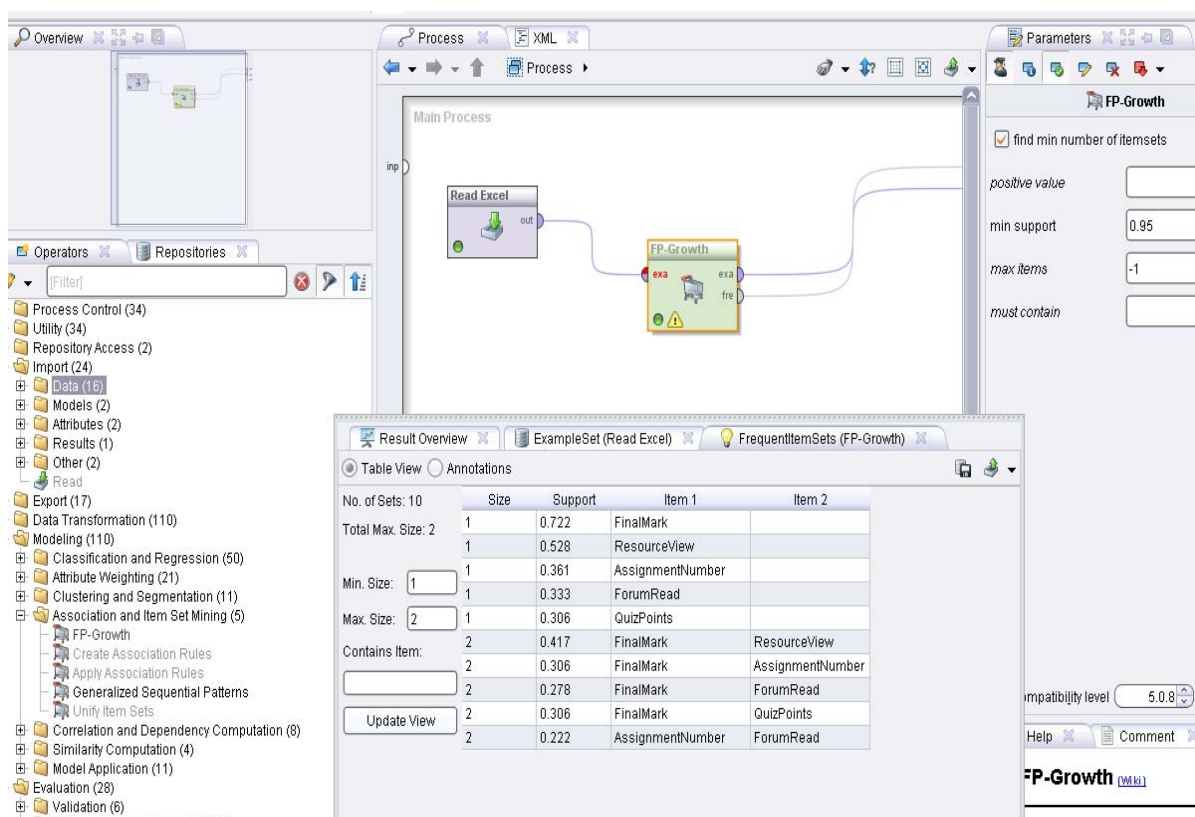


Figure 5. Rapid Miner executing FPGrowth Algorithm

Table 5. Rapid Miner FPGrowth Item Sets

Size	Support	Item1	Item2
1	0.722	FinalMark	
1	0.528	ResourceView	
1	0.361	AssignmentNumber	
1	0.333	ForumRead	
1	0.306	QuizPoints	
2	0.417	FinalMark	ResourceView
2	0.306	FinalMark	AssignmentNumber
2	0.278	FinalMark	ForumRead
2	0.306	FinalMark	QuizPoints
2	0.222	AssignmentNumber	ForumRead

4.4.2 Creating Association Rule Results

The result of running the process of *FPGrowth* operator will be a set of frequent item sets which could be used as input for Create Association Rule operator (see Figure 6). This operator has created several Association Rules, which express regularities that occur often. The most valuable rules are those that predict with a high confidence, presented in Table 6.

Table 6. Rapid Miner Creating Association Rule

Premisiss	Conclusion	Support	Confidence	LaPlace	Gain	g-s	Lift	Conviction
ForumRead	FinalMark	0.28	0.83	0.96	-0.39	0.0 4	0.1 5	1.66
Assignment Score	FinalMark	0.31	0.85	0.96	-0.42	0.0 4	1.1 7	1.80
Quiz Score	FinalMark	0.31	1.0	1.0	-0.31	0.0 8	1.3 8	Infinity

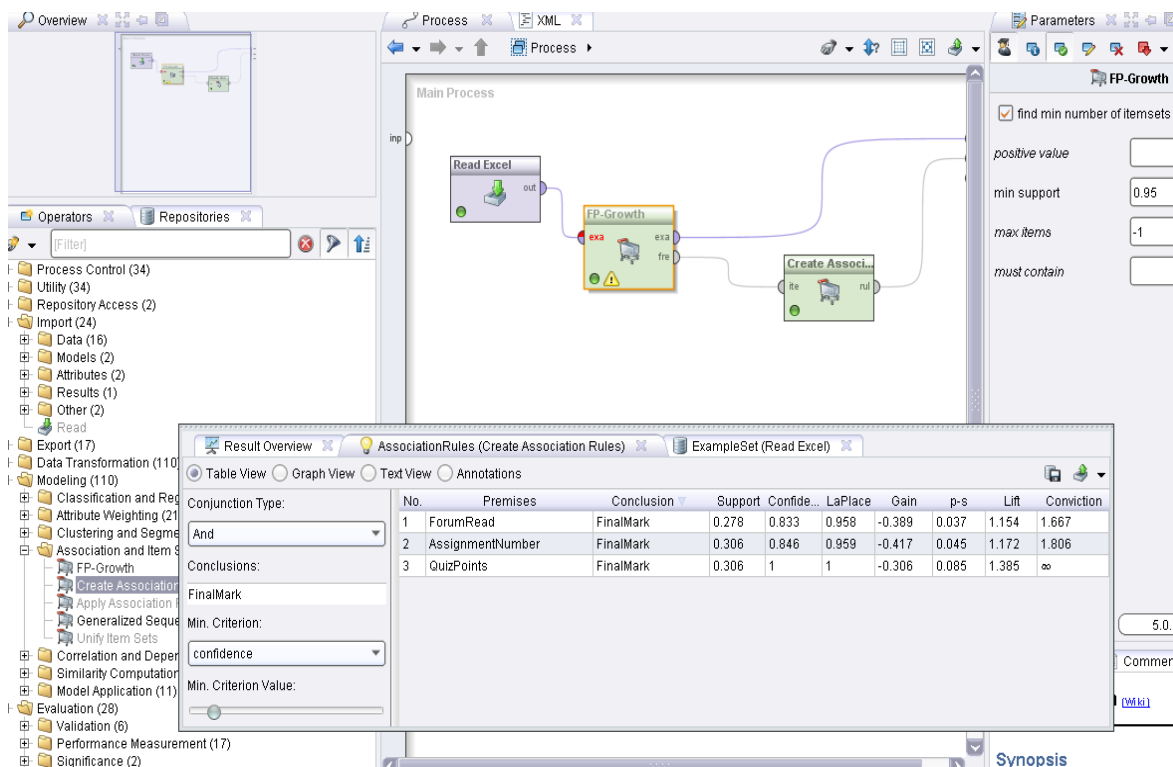


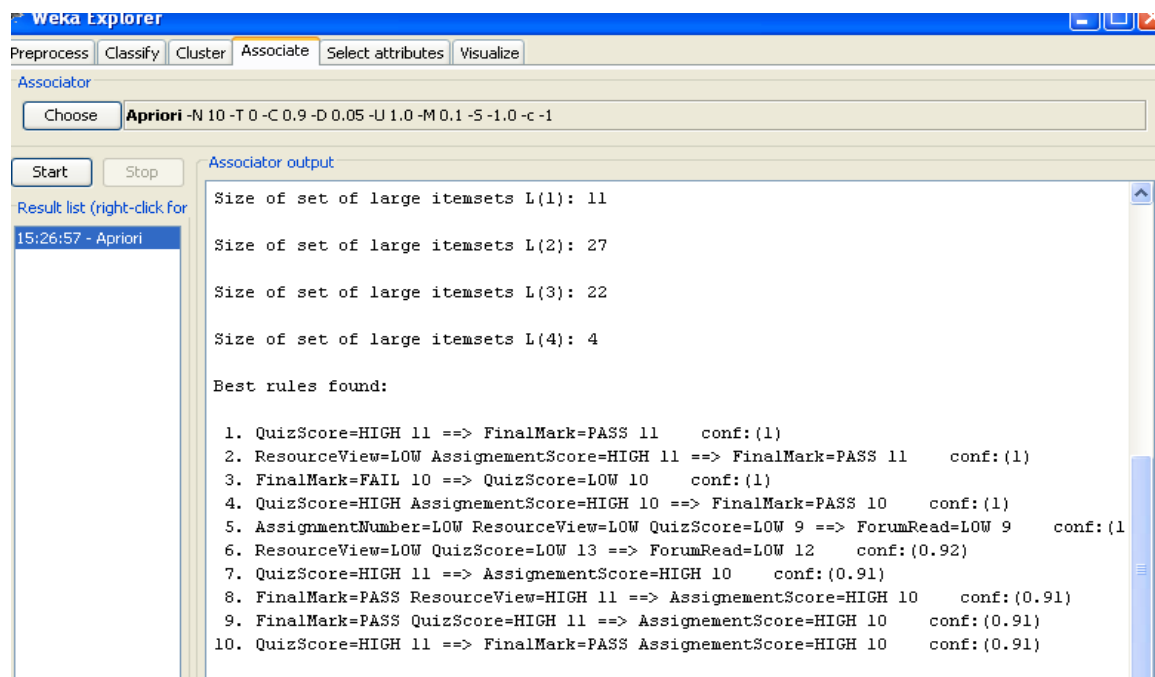
Figure 6. Rapid Miner executing Create Association Rule Algorithm

Summarizing the obtained rules, with support over 0.3 and confidence over 0.8, two main rules were identified concluding that Assignment and Quiz Score serve as predictors for Final Marks.

#### 4.4.3 Apriori Results

Apriori algorithm (Agarwal et al., 1993) was applied for finding association rules over Moodle tables in Weka (see Figure 7). In this research Apriori algorithm was used for identifying attributes characterizing patterns of performance differences between various groups of students.

Figure 7. Weka executing Apriori Algorithm



Best rules found are obtained are described as below:

1. QuizScore=HIGH 11 ==> FinalMark=PASS 11 conf:(1)
2. ResourceView=LOW AssignmentScore=HIGH 11 ==> FinalMark=PASS 11 conf:(1)
3. FinalMark=FAIL 10 ==> QuizScore=LOW 10 conf:(1)
4. QuizScore=HIGH AssignmentScore=HIGH 10 ==> FinalMark=PASS 10 conf:(1)
5. AssignmentNumber=LOW ResourceView=LOW QuizScore=LOW 9 ==> ForumRead=LOW 9 conf:(1)
6. ResourceView=LOW QuizScore=LOW 13 ==> ForumRead=LOW 12 conf:(0.92)
7. QuizScore=HIGH 11 ==> AssignmentScore=HIGH 10 conf:(0.91)
8. FinalMark=PASS ResourceView=HIGH 11 ==> AssignmentScore=HIGH 10 conf:(0.91)
9. FinalMark=PASS QuizScore=HIGH 11 ==> AssignmentScore=HIGH 10 conf:(0.91)
10. QuizScore=HIGH 11 ==> FinalMark=PASS AssignmentScore=HIGH 10 conf:(0.91)

Summarizing the obtained Apriori rules, with support over .3 and confidence over .9, and comparing them with the rules obtained from the two step process in Rapid Miner as described in the previous section, the results set were quite similar by concluding that Assignment and Quiz Score served as predictors for Final Marks.

#### 4.4.5 Generalized Sequential Patterns (GSP) Results

By applying Generalized Sequential Patterns (GSP) to summarization of Moodle logs table from, possible sequence candidates were identified (see Figure 8).

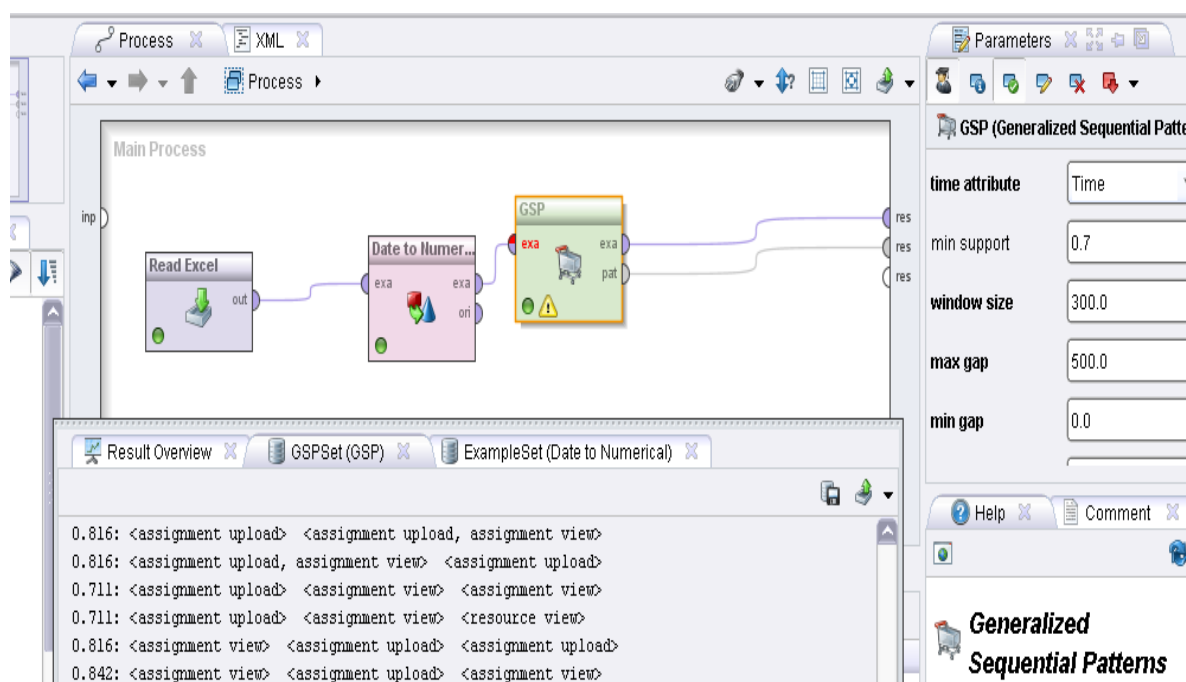


Figure 8. Rapid Miner executing GSP Algorithm

The GSP window width parameter was set to  $w = 5$  minute, because the most interesting items are those that come in a close chronological sequence. GSP operator processed 6866 transactions with minimum support of 0.7 (min support was decreased from 0.9 to 0.7). The results included sequences like: {assignment view, assignment upload, resource view}, {resource view, assignment view, assignment upload} etc. Due to the low amount of data related to collaborative activities (forum, chats etc) the sequential patterns discovered for the CEN112 Course were not very interesting. Interpreting GSP patterns enable how to best organize the educational web space and be able to make suggestions to learners who share similar characteristics (Ha et al., 2000).

## 5. Discussion

Several data mining techniques such as: Attribute Weighting (Weighting by Information Gain, Relief, Hi-Squared, Uncertainty, Clustering (K-Means), Classification (Tree Induction), Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) were applied to Moodle summarization tables.

By applying clustering methods, the goal was to split data set in groups of data points that naturally group together. Student actions were clustered together in order to investigate patterns of students behavior in the Moodle System. KMeans algorithm was used to define clusters, which starts with no prior knowledge about groups in the data.

By using prediction techniques, the goal was to develop a model that can infer predicted variables from predictors' variables. Inductive Decision tree algorithm was selected as classification method. The predicted variable was the categorical variable Final Mark. The target was to define variables that significantly affect in the Final Mark.

By applying ARM, the goal was to discover relationships between variables. FPGrowth, Create Association Rule and APRIORI algorithms were selected as association rule mining techniques. The rules obtained can be explained in the form that if some set of variable values is found, another variable will have a great chance to have a specific value.

## 6. Conclusions

In this research, a data mining model for Moodle data was proposed based on several techniques: Attribute Weighting (Weighting by Information Gain, Relief, Hi-Squared, Uncertainty), Clustering (K-Means), Classification (Tree Induction), Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) was proposed. This educational data mining work allowed identifying and locating information about E-learning processes that need improvements, or those that perform very well and

could be used as good examples. The educational data mining investigated in this research allows analyzing and better understanding the learning and teaching processes by applying data mining techniques. The experimental results have shown that the data mining model presented was able to obtain comprehensible, actionable and logical feedback from the LMS data describing students' learning behavior patterns.

This work concentrated on the overall LMS performance at Epoka University and the mining process of Moodle data. Mining the Moodle data allowed identifying the most effective ways to the teaching process that can be used to enhance the education process. To further test the effectiveness of the proposed model and to increase the generality of this research, more extensive experiments should be conducted by using larger LMS data sets.

## References

- Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In J. B. Bocca, M. Jarke, & C. Zaniolo (Eds.) *Proceedings 20th Int. Conf. Very Large Data Bases, VLDB*, (pp. 487–499). Morgan Kaufmann.
- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. *Data Engineering, International Conference on*, 0, 3–14.
- Agrawal, R., Imilienski, and A. Swami. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, (pp. 207–216).
- Aggarwal, C., & Yu, P. (1999). Data Mining Techniques for Associations, Clustering and Classification. In Zhong, N., & Zhou, L. (Eds.), *Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining*, London, UK: Springer-Verlag.
- Amershi, S., & Cristina Conati, Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments, *Journal of Educational Data Mining*, 2009, 1(1), 18–71.
- Apte, C., & Weiss, S., (1997). Data Mining with Decision Trees and Decision Rules. *Future Generation Computer Systems*, 13, 197–210.
- Baker, M., (2000). The Roles of Models in Artificial Intelligence and Education Research: A Prospective View. *International Journal of Artificial Intelligence in Education*, 11, 122–143.
- Baker, M., (2010). Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education* (3rd edition), vol. 7, pp. 112–118. Oxford, UK: Elsevier.
- Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.) *Grouping Multidimensional Data*, (pp. 25–71). Berlin/Heidelberg: Springer-Verlag.
- Castro, F., Vellido, A., Nebot, A., & Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems. In Kacprzyk, J. et al (Eds), *Applying Data Mining Techniques to e-Learning Problems*, vol. 62, pp. 183–221. Springer Berlin Heidelberg.
- Claroline e-Learning System. Retrieved December, 2010 from <http://www.claroline.net>
- Cios, K. J., Pedrycz, W., & Swinarski, R. W. (1998). *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publisher. Norwell, USA : MA.
- Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Drop Out: A Case Study. In Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (Eds.) *Proceedings of International Conference on Educational Data Mining*, (pp. 41–50), Cordoba, Spain.
- Dokeos e-Learning System. Retrieved December, 2010 from <http://www.dokeos.com>
- Educational Data Mining Research Community. Retrieved February, 2011 from <http://www.educationaldatamining.org/>
- eFront e-Learning System. Retrieved December, 2010 from <http://www.efrontlearning.net>
- Falakmasir, M. H., & Habibi, J. (2010). Using Educational Data Mining Methods to Study the Impact of Virtual Classroom in E-Learning. In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) *Proceedings of the 3rd International Conference on Educational Data Mining* (pp 241–248). Pittsburgh, USA.
- Ha, S.H., Bae, S.M., Park, S.C. (2000). Web Mining for Distance Education. *IEEE International Conference on Management of Innovation and Technology*, 1, 715–719.
- Hartigan, J., (1975). *Clustering Algorithms*. John Wiley & Sons, New York, NY.
- Hartigan, J., & Wong, M., (1979). Algorithm AS136: A k-means clustering algorithm. *Applied Statistics*, 28, 100–108.
- Hogo M. A., (2010, February). Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study, *International Journal of Computer Science and Information Security, IJCSIS*, 7(2), 131–140.
- Hwang, G.J., (1999). A Knowledge-Based System as an Intelligent Learning Advisor on Computer Networks. *J. Systems, Man, and Cybernetics*, 2, 153–158.
- Hwang, G.J., Hsiao, C.L., Tseng, C.R. A. (2003). Computer-Assisted Approach to Diagnosing Student Learning Problems in Science Courses. *Journal of Information Science and Engineering*, 19, 229–248.
- Ilias e-Learning System. Retrieved January, 2011 from <http://www.ilias.de/docu>
- Jeong, H., Biswas G., Johnson J., & Howard L. (2010). Analysis of Productive Learning Behaviors in a Structured Inquiry Cycle Using Hidden Markov Models. In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) *Proceedings*

- of the 3rd International Conference on Educational Data Mining (pp. 81-90). Pittsburgh, USA.
- Kdnuggets - Data Mining Resource, Retrieved February 2011, [www.kdnuggets.com/](http://www.kdnuggets.com/)
- Kononenko, I. (1994). Estimating attributes: analysis and extensions of Relief. *Machine Learning: ECML* (pp.182). Springer Verlag.
- Kotsiantis, S.B., Pierrakeas, C.J., Pintelas, P.E. (2004). Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence*, 18(5), 411-426.
- Margo, H., (2004). Data Mining in the e-Learning Domain. *Computers & Education*, 42(3), 267-287.
- Maull, K. E., Saldivar, M.G., & Sumner T. (2010). Online Curriculum Planning Behavior of Teachers. In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 121-130). Pittsburgh, USA.
- Matsui, T., Okamoto, T. (2003). Knowledge Discovery from Learning History Data and its Effective Use for Learning Process Assessment Under the e-Learning Environment. In Crawford, C., et al. (eds.): *Society for Information Technology and Teacher Education International Conference*.
- Mazza, R., & Milani, C. (2005). Exploring usage analysis in learning systems: Gaining insights from visualisations. In *Workshop on usage analysis in learning systems at 12th international conference on artificial intelligence in education* (pp New York, USA.
- Mazza R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting educators in web-based distance courses. *International Journal of Human-Computer Studies*, 65(2), 125-139.
- Minaei-Bidgoli, B., & Punch, W. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In GECCO (pp. 2252-2263). Chicago, USA: Springer-Verlag.
- Moodle e-Learning System. Retrieved February 2011 from <http://moodle.org>.
- OLAT e-Learning System. Retrieved February 2010 from <http://www.olat.org>.
- Muehlenbrock, M.(2005), Automatic Action Analysis in an Interactive Learning Environment. In *Proceedings of 12th International Conference on Artificial Intelligence in Education*.Amsterdam, The Netherlands:IOS Press.
- Nichols, M. (2003). A theory for eLearning. *Educational Technology Society*, 6(2), 1-10.
- Nugent, R., Ayers, E., & Dean, N. (2009). Conditional Subspace Clustering of Skill Mastery: Identifying Skills that Separate Students. In Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (Eds.) *Proceedings of International Conference on Educational Data Mining*, ( pp. 101-110) , Cordoba, Spain.
- OLAT e-Learning System. Retrieved February 2010 from <http://www.olat.org>.
- Rajibussalim. (2010). Mining Students' Interaction Data from a System that Support Learning by Reflection. In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 249-256). Pittsburgh, USA.
- Rodrigo, M.M.T., Anglo, E.A., Sugay, J.O., Baker, R.S.J.d. (2008). Use of Unsupervised Clustering to Characterize Learner Behaviors and Affective States while Using an Intelligent Tutoring System. In *Proceedings of International Conference on Computers in Education*,(pp. 392-401).
- Romero, C., Ventura, S., García, E. (2007). Data Mining in Course Management Systems: MOODLE Case Study and Tutorial. *Computers and Education* , 51,368-384.
- Romero, C., Ventura, S. (2007). Educational Data Mining: a Survey from 1995 to 2005. *Expert Systems with Applications* , 33(1), 135-146.
- Romero, C., Ventura, S., Vasilyeva, E., Pechenizkiy, M. (2010). Class Association Rules Mining from Students' Test Data. In Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 317-318). Pittsburgh, USA.
- Romero, C., Ventura, S.(2010), Educational Data Mining: A Review of the State-of-the-Art. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 40(6), 601-618.
- Uribe-Tirado, A., Melgar-Estrada, L.-M., & Bornacelly-Castro, J.-A. (2007). Moodle learning management system as a tool for information, documentation, and knowledge management by research groups. *Profesional de la Información*, 16(5), 468-474.
- Silva, D., & Vieira, M.(2002). Using data warehouse and data mining resources for ongoing assessment of distance learning. In *Proceedings of the 2002 IEEE Conference on Advanced Learning Technologies* (pp. 40-45), Kazan, Russia.
- Sison, R., & Shimura, M. ,(1998). Student Modelling and Machine Learning. *International Journal of Artificial Intelligence in Education*, 9,128-158.
- Srikant, R., & Agrawal, R. (1996). Mining Sequential Patterns: Generalizations and Performance Improvements. In P. M. G. Apers, M. Bouzeghoub, & G. Gardarin (Eds.) *Proc. 5th Int. Conf. Extending Database Technology, EDBT*, vol. 1057, (pp. 3-17). Springer-Verlag.
- Srivastava J., Cooley R., Deshpande M., & Tan P. (2000, January) Web Usage Mining: Discovery and Applications of usage Patterns form Web Data, *SIGKDD Explorations*, 1(2), 12-23.
- Talavera, L.,& Gaudioso E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *Workshop on Artificial Intelligence in CSCL. ECAI*. 17-23.
- Tang, T.Y., McCalla, G. , (2005).Smart Recommendation for an Evolving e-Learning System:Architecture and Experiment. *International Journal on e-Learning*, 4(1), 105-129.
- Tsai, C. J., Tseng, S. S., & Lin, C. Y. (2001). A Two-Phase Fuzzy Mining and Learning Algorithm for Adaptive



- Learning Environment. In *ICCS '01: Proceedings of the International Conference on Computational Science-Part II*, (pp. 429–438). London, UK: Springer-Verlag.
- Wang, F., & Shao, H. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications* 27(3), 365-377.
- Wu X., (2004). Data Mining: An Artificial Intelligence Perspective. *The IEEE Intelligent Informatics Bulletin*. 4(2), 23-26.
- Yoo, J., Yoo, S., Lance, C., Hankins, J.(2006). Student Progress Monitoring Tool Using Treeview. In *The 37th Technical Symposium on Computer Science Education*, (pp. 373-377). Houston, USA :ACM Press.
- Zaïane, O. , & Luo J.,(2001). Web usage mining for a better web-based learning Environment. Proceedings of conference on advanced technology for education, Banff, Alberta, , pp. 60-64.
- Xu, R., & Wunsch, D. I. I. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–67.