BUSINESS INTELLIGENCE IN HIGHER EDUCATION
Managing the Relationships with Students

Maria Beatriz Piedade
School of Technology and Management, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal

Maribel Yasmina Santos
Information Systems Department, Algoritmi Research Centre, University of Minho, 4800-058 Guimarães, Portugal

Keywords: Business Intelligence, Customer Relationship Management, Data Mining, Data Warehouse, OLAP, Student Relationship Management.

Abstract: The closely monitoring of the students' academic activities, the evaluation of their academic success and the approximation to their day-by-day academic activities are key factors in the promotion of the student's academic success in higher education institutions. To be possible the implementation of monitoring processes and activities, it is essential the acquisition of knowledge about the students and their academic behaviour. This knowledge supports the decision-making associated with teaching-learning process, enhancing an effective *institution-student* relationship. This paper presents a *Student Relationship Management* (SRM) system that is under development. The *SRM* system supports the *SRM concept* and *practice* and has been implemented using concepts and technologies associated to the *Business Intelligence* systems. To demonstrate the *SRM* system relevance in the process of acquisition of knowledge about the students and in the support of actions and decisions based on such knowledge, an application case carried out in a real context is also presented.

1 INTRODUCTION

Portuguese Higher Education has been characterized by a high rate of students' failure and abandon, mainly in the first year of the graduation courses (statistical results can be consulted in the official *web* page http://www.gpeari.mctes.pt). Although this reality has changed due to several individual or institutional actions, integrated activities need to be proposed and adopted. To be possible, it is necessary the identification of the factors and the measures that need to be monitored in the teaching-learning process and in the student-teacher relationship. One of the activities usually pointed out as crucial to promote the students' success is the closely monitoring of the students' academic activities (Pile and Gonçalves, 2007). Although important, this activity does not take place in many institutions. Among the reasons, we point out the huge number of students with failure in the first graduation year, the huge number of new students in some courses and the work overload of the teaching staff. In the Portuguese institutions, teachers are involved in lecturing, researching and management tasks. To help teachers and students in this complex process, an adequate conceptual and technological support is needed. The conceptual framework and the technological infrastructure are in this work integrated in a *Student Relationship Management* (SRM) *system*. The *SRM system* supports the *SRM concept* and *practice* and is based in a *Business Intelligence* infrastructure. To demonstrate its relevance in the students' knowledge acquisition process, in the early identification of failure situations, in the decision-making support and in the automatic interaction with the students, it is presented an application case that has occurred in a real context. This paper is organized as follows: Section 1 refers the academic failure as the problem that motivates and justifies the *SRM system*; Section 2 includes an overview of the *SRM* principles and concepts; Section 3 describes the *SRM system* architecture; Section 4 describes the application case and the data analysis process. This section also

Beatriz Piedade M. and Yasmina Santos M. (2009).
BUSINESS INTELLIGENCE IN HIGHER EDUCATION - Managing the Relationships with Students.
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, pages 297-302
DOI: 10.5220/0002302602970302

presents the interpretation of the obtained results and their support to interaction activities with the students; Section 5 concludes with the *SRM system* expected advantages and the upcoming future work.

2 CONCEPTUAL FRAMEWORK

The *CRM* (*Customer Relationship Management*) systems are nowadays used in business contexts to support and to manage the relationships with the customers. These systems allow the identification of knowledge about the customers, using the information and business transactions available in the organization databases. Using this *knowledge*, the organization defines the activities and actions that allow maintaining a close and strong relationship with its customers (Payne, 2006). The *SRM system* was inspired in the *CRM* principles, but supports processes and activities concerned with the *teaching-learning* process, mainly activities that allow the monitoring and the supervision of the students academic activities. To exemplify the similarity between the *CRM/SRM* actions, one could compare the actions developed by the customer's manager, that on the scope of the banking activity, alerts the customer when he/she exceeds his/her credit account, and the actions developed by the student's tutor/teacher, that on the scope of the monitoring processes send an alert message to the student when detects he/she misses several lessons. To refer, that the "*Student Relationship Management*" designation was already used in a technological/commercial environment to designate solutions mainly dedicated to support processes related with the students in academic areas (students' management information, courses and lessons management, admissions management, enrolment and registration management) and areas related with available services (communications, marketing, financial aids, accommodation). In this work is proposed a definition of the *SRM concept*, understood as a process based on the students' acquired knowledge, whose main purpose is to keep an effective *student-institution* relationship through the closely monitoring of the students' academic activities. This concept was based on the premise that there exist a strong correlation between the closely monitoring of the students' academic activities and their academic success promotion. The *SRM practice* is understood as a set of activities/actions, which should guarantee the students' individual contact, and an effective, adequate and closely monitoring of his/her academic performance. To validate the *SRM concept* and the set of activities included in the *SRM practice* it was adopted a methodology based on the *Grounded Theory* principles, which included the interviews realization and analysis (Piedade and Santos, 2008).

3 SRM SYSTEM

The SRM system is based on the *SRM concept* and on a set of activities that composes the *SRM practice*. To undertake a S*RM practice* it is necessary: (i) to have adequate, consistent and complete information about the students. This information must be stored in an appropriate data repository, which allows maintaining a single vision of students' data; (ii) the analysis of such data in order to obtain knowledge about the students and their academic behaviour; (iii) the starting of automatic actions whenever specific situations are detected; and (iv) to assess the impact of all the implemented actions over the students. With respect to i) and ii) the *SRM system* envisages the implementation of a *data warehouse* and its exploration using data analysis tools. These structural framework leads that the *SRM system* is implemented using the technological infrastructure that traditionally supports the *Business Intelligence systems* (Negash and Gray, 2003). With respect to iii) and iv) it was identified the relevant indicators and behaviour patterns that characterizes the different situations to supervised and it was implemented the actions to be executed automatically by the different participants in the *SRM practice* (teacher, tutor, course director) and analysed the impact of the carried out actions in the students behaviour and their final results.

The *SRM system* architecture aggregates four main components: (i) The *Data Acquisition and Storage* component, responsible for the storage, in the *data warehouse*, of the students' data; (ii) the *Data Analysis* component, responsible for obtaining knowledge about the students, using appropriate data analysis tools that allows the patterns identification; (iii) the *Interaction* component, responsible for maintaining an adequate and effective relationship with the student. In this interaction is used the knowledge about the student(s) to start a set of automatic actions that reflect the academic situation of an individual or a set of individuals; (iv) the *Assessment* component, responsible for the assessment of all the concretized actions and their impact. The *SRM system* prototype implementation was done using database management tools, *Business Intelligence* tools and *web* development tools. Considering the context in which this project takes place, the selected development tools integrate the *Microsoft* environment. The *SRM system* prototype validation has been done through the execution of a set of application cases, taking place in different Higher Education Institutions (Piedade and Santos, 2009).

4 APPLICATION CASE

The data used in this application case was gathered in a graduate course of a Portuguese Higher Education Institution. This course, in the engineering area, is composed by a set of curricular units. The selected unit, with 70 students, is integrated in the first graduation year. The teaching method adopted is based in a presential component, and on an *e-learning* component supported by the *e-learning* platform available in the institution. The presential component includes different types of classes (theoretical, practical and tutorial orientation). The activities included in the theoretical classes are related with the curricular subjects' presentation and explanation. The activities included in the practical classes are related with exercises and problems solving in a laboratory environment. The activities included in the tutorial orientation classes include students' individual work support and orientation and subject clarification. The *e-learning* component includes activities related with the distribution of relevant information and materials (unit general information, curricular contents, exercises and project guidelines), and, also, communication activities (messages that are automatically or manually sent and discussion forums). The unit assessment includes two distinct methods: the normal assessment period and the exam assessment period. The normal assessment period integrates a written test and a project with individual discussion, with weights of 40% and 60%, respectively, in the final mark. The quantitative *mark* scale comprises values among 0 to 20. In both assessments, 8 is the minimum mark that needs to be obtained by the students in order to be possible the calculus of their final mark. This final mark results from the application of the weights associated with the test and the project and needs to be equal or higher than 10 to guarantee success in the unit. The exam assessment period only comprises a written exam. To pass the unit, the student must obtain a positive mark (value greater or equal then 10). In the exam evaluation, we frequently have students who have failed the normal assessment and/or students that missed the normal assessment. The available data, about each student and his/her involvement in the *teaching-learning* process, was: (i) provided by the institutional academic system (like students' personal information and unit information); (ii) provided by the unit teachers (include students presences in classes, developed activities and the corresponding student marks); (iii) provided by the *e-learning* system (information related with the *student-unit* interaction using the *e-learning* platform).

The analysis of all the available data allowed the identification of the data subset considered in this application case. This data subset includes:

- student number; student registration year; and student phase of admission to higher education. The phase information is only related with first year students and can have the values *first* or *second*. This attribute is also used to identify the students that are repeating the unit (value *rep*), as a consequence of a previous failure; and, worker/full time students information. In order to maintain the students' privacy, all the information that allows his/her identification is ignored or codified.
- unit identification; unit designation; unit curricular year and semester; and associated course.
- class type identification; class type description; class start hour; class duration; and, class week day.
- assiduity rate associated with each student and each class type. The assiduities values were transformed in the following classes: *Low* (< 50%), *Acceptable* (≥ 50 and < 70%) and *High* (≥ 70% and ≤ 100%).
- assessment activity identification; activity description; weight in the final mark, mandatory information; and, marks (obtained by each student). To represent some specific situations, negative values were used. In the project assessment results the value *-1* mean that the student misses the project individual discussion; the value *-3* mean that the student did not implemented the project work. In the unit final results, the value *-1* mean that the student was not submitted to any type of evaluation (test, project or exam); *-2* mean that the student *failed* the unit, but he/she was submitted to any one of the activities includes in the unit assessment (test/ project/exam). The final marks were also classified in qualitative terms, using for that purpose the following attributes: *Satisfactory* (between 10 and 13), *Good* (between 14 and 16) and *VeryGood* (between 17 and 20).
- number of distinct days that each student interacted with the unit using the *e-learning*

platform. In qualitative terms, it was considered that from 0 to 16 corresponds to a *low* interaction, 17 to 32 corresponds to a *reasonable* interaction, 33 to 49 corresponds to an *expressive* interaction and values greater or equal to 50 correspond to a *high* interaction (the distributions of the values was analysed in order to be possible the definition of this limits).

Considering all the relevant data, it was: (i) designed the *data warehouse* model, a multidimensional data model which follows the *constellation schema* (Figure 1); (ii) implemented the *data warehouse,* by fact and dimension tables creation*;* (iii) loaded the operational data to the *data warehouse*. The loading process followed the ETL process steps, in which the relevant data was *extracted* from the source databases, *was cleaned* (when errors in data were detected) and *was transformed* in order to accomplish the format of the target system (the *data warehouse*).

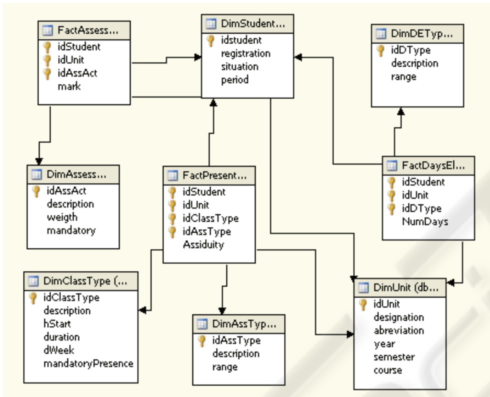

Figure 1: The *Data warehouse* model.

The *data warehouse* exploration has been done using *OLAP* and *data mining* techniques.

In this application case, OLAP cubes were created to analyse the students' results verifying both the *teaching-learning* experiences and the assessment methods influence. To analyze particularly the unit results, and the correlation with the theoretical classes' presences, the unit interaction through the *e-learning* platform and the project work a cube was created. From its analysis it is clear that all the students that did not implement the project (mark value *-3*) or misses the individual project discussion (mark value *-1*) *fail* the unit. It can also be verified that there exists a large number of students with low assiduity to theoretical classes. Many of them are repeating the unit and others are differentiated by the phase of admission to the University. In these cases, only the students that are repeating the unit pass. These students also have a *reasonable* or an *expressive* interaction with the *e-learning* platform. The students that *fail* have few interactions with the unit, existing only one exception, a student (id 15) with *many* interactions (Figure 2).

				Description		
				low		
Dim Student	Phase	Situation	Descriptior	Assiduity	Num Days	Mark
36	rep	full-time	Final	0	19	12
			Project	0	19	8
58	rep	full-time	Final	0	13	-1
			Project	0	13	-3
62	rep	full-time	Final	0	18	10
			Project	0	18	9
64	rep	full-time	Final	0	0	-1
			Project	0	0	-3
68	1phase	full-time	Final	0	1	-1
			Project	0	1	-3
7	rep	worker	Final	8.33	21	-1
			Project	8.33	21	-3
47	rep	full-time	Final	8.33	0	-2
			Project	8.33	0	-3
48	rep	full-time	Final	8.33	37	15
			Project	8.33	37	14
12	2phase	full-time	Final	16.67	8	-1
			Project	16.67	8	-1
15	rep	full-time	Final	16.67	52	-2
			Project	16.67	52	8
29	rep	full-time	Final	16.67	11	-1
			Project	16.67	11	-1
30	rep	full-time	Final	16.67	38	12
			Project	16.67	38	12
33	rep	full-time	Final	16.67	23	11
			Project	16.67	23	8
40	rep	full-time	Final	16.67	10	-2
			Project	16.67	10	-3
13	rep	full-time	Final	25	16	-1

Figure 2: Students data extract, grouped by *Low* assiduity rate.

Another analysis (Figure 3) allows us to verify that many students with *high* assiduity rates, and *expressive* or *many* interactions, *pass* the unit with *good* marks. Students with *acceptable* assiduities and *few* or *reasonable* interactions, *fail* or *pass* the unit with a *sufficient* mark. Preoccupant situations occur with students that go to the University in the second phase, as many of them *fail* the unit. It was verified that some students *fail* and others *pass* the unit with different marks.

				Description					
				acceptable			high		
Dim Student	Phase	Situation	Descriptior	Assiduity	Num Days	Mark	Assiduity	Num Days	Mark
52	1phase	full-time	Final	58.33	43	10			
			Project	58.33	43	8			
8	1phase	full-time	Final	66.67	24	-2			
			Project	66.67	24	8			
18	1phase	full-time	Final	66.67	68	12			
			Project	66.67	68	13			
32	rep	full-time	Final	66.67	32	-1			
			Project	66.67	32	-1			
41	2phase	full-time	Final	66.67	68	-2			
			Project	66.67	68	8			
55	1phase	full-time	Final	66.67	6	-1			
			Project	66.67	6	-1			
57	rep	full-time	Final	66.67	25	10			
			Project	66.67	25	8			
11	1phase	full-time	Final				75	35	-2
			Project				75	35	-1
43	1phase	full-time	Final				75	44	15
			Project				75	44	14
53	2phase	full-time	Final				75	19	-2
			Project				75	19	8
56	1phase	full-time	Final				75	45	15
			Project				75	45	16
60	1phase	full-time	Final				75	73	-2
			Project				75	73	8
2	1phase	full-time	Final				83.33	42	15
			Project				83.33	42	15
20	1phase	full-time	Final				83.33	43	16
			Project				83.33	43	18
21	rep	full-time	Final				83.33	19	14
			Project				83.33	19	14
31	1phase	full-time	Final				83.33	27	15

Figure 3: Students data extract, grouped by *Acceptable* and *High assiduity* rate.

Now our objective is to identify the behaviour of the students differentiating them by marks. The main purpose is to identify the students' profile. Through it, mainly in the case of *fail*, adequate actions can be implemented to minimize the failure rate in future unit editions.

Data mining algorithms were used to identify the students' profile considering the assessment results: *Fail*, *Satisfactory*, *Good*, *Very Good*. The objective is to find a model that describes the predictable attribute, *Mark*, as a function of the input attributes *phase*, *situation*, *numDays* and the assiduity rates for the *theoretical, practical and tutorial* classes. It was followed the traditional steps of the knowledge discovery in databases process: *Data Selection, Data Treatment*, *Data Pre-Processing*, *Data Mining* and *Results Interpretation*. The *Data Selection*, *Data Treatment*, *Data Pre-Processing* steps were supported by the *data warehouse* implementation process. In the *Data Mining* step, it was selected a decision tree algorithm to carry out the classification task previously defined. The obtained model (Figure 4) integrates a set of rules, in a tree form, where each tree node has associated a set of conditions that lead to a specific decision.

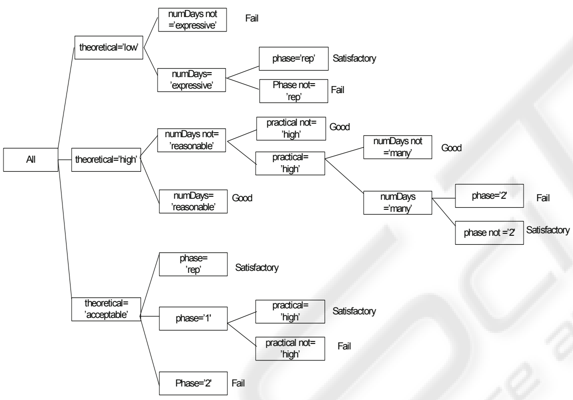

Figure 4: *Data mining* model.

The identified model integrates a set of rules that explicitly describe the students' profile. From the analysis of the rules, it is possible to verify that the *theoretical* attribute is the attribute that more influence the students' profile, followed by the *phase*, *practical* and *numDays* attributes.

The *Fail* profile is associated with:

1. Repeating students with low theoretical assiduity rates and few, reasonable or many interaction with the *e-learning* platform;
2. Students at the first time in the unit, with low theoretical assiduity and an expressive interaction level;
3. Students with the first registration (first phase), with an acceptable theoretical assiduity and with few or acceptable practical assiduity;
4. Students with the first registration (second phase) with an acceptable theoretical assiduity.

The *Satisfactory* profile is associated with:

1. Repeating students with acceptable theoretical assiduities, or low theoretical assiduities, but with an expressive interaction;
2. Students with the first registration in the unit, with acceptable theoretical assiduities and high practical assiduities

The *Good* profile is associated with:

1. Students with high theoretical assiduities.

No rule was obtained to characterize the *Very Good* profile. This is due to the fact that only one student achieved this mark.

The *OLAP* and *data mining* analyses allow us to verify that to decrease the failure profile it is necessary to take special attention to the students, differentiating them by the admission phase.

The students that are at the first time in the unit must be encouraged to go in a regular basis to the different type of classes. These students are, in many cases, influenced by older students that say to them to avoid classes, mainly the theoretical classes. For the second phase students, an additional support must be given as they arrive to the University when half semester has passed. Due to this situation, these students lose the initial curricular contents explanation and the subsequent curricular content comprehension, fact that could help to explain their failure. For these cases, the institution could adopt special activities or procedures, as extra classes or tutorial orientation, providing additional support to the students. It is also necessary, for all the students, to verify the evolution of the project work implementation, motivating the students and providing additional support when necessary. This support can be achieved using tutorial classes. This was not the case of the current edition of the unit, helping to explain the lack of presences and interest in this king of lectures. In what concerns repeating students, it is also necessary to motivate them to go to the presential classes, although in many situations these students have timetable incompatibilities. To overcome this limitation, the institution could adopt the schedule differentiation, like classes in the morning for the first year students and in the afternoon for the second year students. In complement, it is also necessary to motivate the use of *e-learning* platform increasing the interaction of the student with the unit. These different situations need to be evaluated without neglecting the students that present what we could say is a "success profile".

The previous results allow us to conclude that data analysis supports the process of obtaining knowledge about the students and their academic behaviour, through the students' behaviour pattern identification.

The obtained students' information allows the definition of a set of actions (included in the *SRM practice)* to closely follow the students. The main purpose is to decrease the failure rate. The identified actions, to integrate in future editions of the unit, include the presences/interaction automatic monitoring and an adequate support to the project work. The presences/interaction automatic monitoring needs the automatic recording of this data and the subsequent automatic sent of alert messages to the student when deviating behaviours are detected. The main purpose of these actions is to alert the students to both the continuous unit interaction and the daily study as a way of increase their academic success. Between the teachers and the course coordinator/director could also exist periodically changes of information, which state how the semester is ongoing. Based on these reports specific actions can be took, as verifying if the student has any problem.

The implementation of activities, like described above, is always supported by a *web application*. Figure 5 shows an example of the presences report. To refer that the descriptions are in the Portuguese language and neither the student nor the institution is identified.

Figure 5: Application *web* page view.

Next steps, in this project include: i) the integration of additional information in the data analysis process, related with the students and their activities in the unit scope ii) the implementation of other activities that enable a closely monitoring of the students (activities included in the SRM practice); and (iii) the SRM practice assess, through the analysis of its impact in the students behaviour and their final results.

5 CONCLUSIONS AND FUTURE WORK

In Portuguese Higher Education institutions exists a strong budget control and persists a high rate of failure and abandon (mainly in the first graduation year). With the new formative process implementation aligned with the *Bologna Process*, the number of hours of contact between the teachers and the students decreased. This requires high student autonomy in the learning process. In this scenario, it is essential the design and implementation of mechanisms that facilitate the monitoring of the students' academic activities. In this context, we believe that the *SRM concept* and *practice* implementation, supported by the *SRM system*, creates an advantage towards the students success promotion, and, therefore, in the institution success, ensuring an effective *student-institution* relationship. Future work, in this project, includes the fully implementation of the prototype and its validation with more application cases (running in two higher education institutions).

REFERENCES

Negash, S. and Gray, P. 2003. "Business Intelligence." In Ninth Americas Conference on Information Systems.

Payne, A. 2006. Handbook of CRM. Achieving Excellence in Customer Management. Elsevier-BH

Piedade, M. B. and Santos, M. Y. 2008. "Student Relationship Management: Concept, Practice and Technological Support." In IEMC 2008 - IEEE International Engineering Management Conference. June 28 to 30, Estoril, Portugal.

Piedade, M. B. and Santos, M. Y. 2009. "An Application of the Student Relationship Management Concept." In ICEIS 2009 - 11th International Conference on Enterprise Information Systems. May 6 to 10, Milan, Italy.

Pile, M. and Gonçalves, I. 2007. "Programa de Monitorização e Tutorado." In A qualidade em estabelecimentos de ensino superior. Exemplo de boas práticas. Instituto Português da Qualidade. Caparica.