# Mining Retention Rules from Student Transcripts: A Case Study of the programs at a Federal University

Marcelo S. Santos<sup>1</sup>, Liz C. de Santana<sup>1</sup>, Quemuel L. Pereira<sup>1</sup>, Marcos Silva<sup>1</sup>, Daniela B. Claro<sup>1</sup>, Veronica M. C. S. Lima<sup>1</sup>, Vaninha Vieira<sup>1</sup>, Silvana Ribeiro<sup>1</sup>, Ana R. Telles<sup>1</sup>, Denivaldo Lopes<sup>2</sup>

<sup>1</sup>FORMAS - Formalismos e Aplicações Semanticas - Departamento de Ciência da Computação Instituto de Matemática – Universidade Federal da Bahia (UFBA) Av. Adhemar de Barros S/N – 40.170-110 – Salvador – BA – Brazil

<sup>2</sup>Departamento de Engenharia de Eletricidade - Federal University of Maranhão (UFMA) 65080-040 – São Luís – MA – Brazil.

{marceloss,lizcastro,quemuel,massilva}@dcc.ufba.br

{dclaro,cadena,vaninha,silvanar,anaregi}@ufba.br, dlopes@dee.ufma.br

Abstract. Due to the increase in inflows mainly because of REUNI procedures and low completion rates observed at a Federal University, it is necessary to identify which courses may cause the students to remain in their programs longer than expected or even leaving the university before their conclusion. In this work, we suggest that the combinations of courses in a semester may not be appropriate thus causing a student retention. We present a case study on analyzing retention rules in all programs at a Federal University. With the results obtained by mining academic transcripts, we discuss how changes can be made in the flowchart so as to decrease the student retention rates.

# 1. Introduction

Brazilian universities have seen an increasing in the number of inflows, mainly because of REUNI procedures (Restructuring and Expansion of Federal Universities) [REUNI 2012]. In its design and implementation at the Federal University of Bahia (UFBA), there was in 2012 an enrollment of over 32,000 students due to the new programs. Among these students over 10,500 of them are enrolled in evening programs. Indeed, in 2012, 29,171 students were enrolled at the Federal University of Bahia (UFBA), where 3,313 of them were enrolled in BI (Interdisciplinary Bachelors) and 6,102 students were enrolled in evening programs (2,107 refer to an evening BI program). Among the main guidelines of REUNI procedures, it is possible to highlight an increase in the number of places of entry, especially for evening programs and a reducing rate of evasion and occupation of unfilled vacancies. According to [Filho et al. 2010], the success rates expected in the REUNI procedures were about 90% of the inflows. But success rates in undergraduate programs are still much lower than the desired.

After joining the REUNI procedures, UFBA increased by 70% the number of programs. Nowadays UFBA has 112 programs in different shifts, where 100 programs of Linear Progression, 10 of Interdisciplinary Bachelors and 2 programs of Technology. However, this University has not reached yet the proposed goal of raising the graduation rate or reducing the rate of retention and evasion.

In this work we stated that students are retained when they fail a particular course in a previous semester and they cannot follow a certain course in the following semester. Failure to complete the regular period can be due to the retention caused by a particular course, thus damaging the regular flow of the academic student's life. The non regular flow of academic student's life causes loss to the University, especially in programs under implementation [Filho et al. 2010].

According to [Barros and Mendonça 1998], despite the positive effects of student's failure in a course, it is possible to highlight two negative effects related to it: i) it obstructs the regular educational flow, raising the cost of a student from a University's perspective. To the family's point of view, the student spends more than the average time of the program to have a professional position and as a consequence a financial return to aid their family, ii) student's failure has a negative effect on self-esteem and motivation, increasing the likelihood of future failures.

Thus, this project aims to analyze retention rules in all programs at the Federal University of Bahia, highlighting three important points: i) courses that most contribute to student retention ratio; ii) courses offered in the same semester that can contribute to student retention and; iii) which courses retained or not-retained students can be approved or failed. Technically, this project was aimed at mining all programs to analyze each student transcript and each course they followed. Through data mining techniques [Tan et al. 2009], specifically association rules, it was possible to identify some courses which were causing student retention.

The first program analyzed was the Information Systems program [Silva et al. 2013], which relevant results were presented in relation to retention. This first initiative also allowed a better calibration of our algorithms. These results were important to validate our research objectives and go further to analyzing the other programs at our University.

From our preliminary results it was possible to identify a retention ratio on programs such as Oceanography and Computer Science respectively about 79% and 71% of student that were enrolled on the second semester. For example, on Gender and Diversity program, students are 100% retained when enrolled in FCHE28 (Supervised Internship in Gender and Diversity I) course on their sixth semester. After applying our association rules, our results presented an important information regarding the student retention in academic flowchart courses. Thus, it was possible to observe each flowchart and propose some modifications in order to reduce student retention in a given semester.

This work is structured as follow: Section 2 presents our related work; Section 3 depicts our methodology. Section 4 describes our database; Section 5 presents the data analysis. Section 6 presents our experiment using association rules and section 7 discusses our results. Section 8 presents the Threats of Validity and section 9 depicts our conclusion and future work.

# 2. Related Work

Data Mining techniques applied to Educational data have already been used by several authors abroad the World [Romero and Ventura 2010, Baker et al. 2011]. Specifically in the context of retention and evasion, many authors used different techniques and approaches to analyze this problem [Nandeshwar et al. 2011, Rigo et al. 2012]. Silva et al. [Silva et al. 2013] conducted an analysis regarding student retention on the Information Systems course at UFBA. Their work aimed to use association rules to identify which courses can contribute to student retention and on which courses a student tend to be approved or failed. With those information, the authors have suggested to reorganizing such curriculum in order to reduce student retention ratio. Different from our approach, their evaluation was concerned only to a specific program analyzing the frequent association rules in the program. Our results was concerned all programs at the UFBA, analyzing the frequent and infrequent rules (seção 6) of programs and creating a panorama of courses and programs that retain more per semester, thus retrieving such kind of information: which courses have the highest student retention ratio, which semesters have the highest retention rates, which courses most retain on a semester. Moreover, such analysis was conducted even on courses on progress, it means, on courses that no students have been graduated yet.

Authors in [Nandeshwar et al. 2011] conducted a literature review and suggested using different learning methods. They found that student's family related variables, socio-economic status, high GPA (Grade Point Average) and exam grades had influenced over the student retention problem. Zhang et al. [Zhang et al. 2010] assemble a data-warehouse of three systems (library, online learning and academic), and used Naive Bayes in order to notify students on their potential chance to be retained. Yu et al. [Yu et al. 2010] used decision trees and social factors to identify the impact on student retention. Campello et al.[Campello and Lins 2008] analyzed the duration of 6 years student's socio-economic, enrollment exam and student transcripts. They used clustering algorithms to identify retained student profile at the Federal University of Pernambuco. Authors in [Manhães et al. 2011] used ten different data mining algorithms to identify those who tend to evade the course of Engineering of the Polytechnic School of UFRJ. They used as variables the highest GPA and students' grades on a first semester course.

We can observe that most work focus on socio-economic and student profile factors to analyze student retention. Different from our approach, we are interested on the disposal of the courses per semester. Our hypothesis concerns that such course's organization has an impact on student retention.

# 3. Methodology

In several Universities, students are considered retained when they exceed the maximum limit for conclusion of a program.

Due to many new programs in our Federal Universities there are not enough students who have completed their degree, thus making it impossible the usage of this definition as a valid hypothesis of retention.

It is important to note that if a student does not get approval in a course during a semester, this mean that he/she cannot follow other course in the next semester. Thus, the combination of each course in a semester could contributing to the student retention. However, as stated by our hyphotesis, it is possible to reorganize this flowchart since some changes can be made before student's graduation.

In this work, we stated that a student is considered as retained if for each recommended course for a student in a given semester, if the student failed in at least one of the prerequisites for this course, this student will be retained. Considering the heuristic stated above and the characteristics of the algorithm used, a Program must satisfy the following requirements to be analyzed: i) having at least 3 semesters with courses that have prerequisites; ii) the frequency of an association rule must be equal to at least 15 in a database.

The requirements above were needed since: i) programs with few prerequisites does not retain students throughout the semester; ii) Important rules were produced with the analysis of at least one thousand students, and iii) an association rule must have a minimum frequency to be analyzed;

In the program of UFBA, there are some particularities that were highlighted, such as, the same program offered in different shift have different flowchart (eg BSc in Physics, day and night), the same program has several qualifications, where each qualification has different flowchart (eg. Foreign Language, English, French, etc.), the same program with different modalities (bachelor and teaching) have different courses and yet the profile of students who attend the night shift is different from those one who attend during the day. It is therefore important that the shifts, qualifications and modalities are analyzed separately in order to obtain relevant information according to each specific program.

It is important to note that any change in a flowchart creates a new flowchart and so we may have few students belonging to a specific flowchart. In order to minimize this problem we considered that if a change in a flowchart did not imply in a change of prerequisites we can consider the analysis of these flowchart together. However, some programs lacked the minimum requirements so they could not be analyzed, in most cases for the following reasons: a) a program may have few students and so there are few students doing a course, and; b) a curriculum was changed recently causing low number of student linked to the current curriculum.

Thus, of the 85 programs at the UFBA, 70 courses were analyzed. Because of this analysis, it was necessary to evaluate the data base's granularity in order to have a better definition of the criteria for the application of data mining algorithms.

#### 4. Dataset

There are three entities with informations in our dataset: students, courses and programs. A student is enrolled in a programme. This enrollment occours each semester. To complete a programme, a student has to do a set of courses. Moreover, a student must be enrollment in many courses on any given semester. In case of a regular semester, the student may have different remarks, such as: be approved in a course, be failed in a course due to not attending it and be failed due to grade.

The finest grain of data in our dataset is (*student,semester,course*). That is, which course a student has taken in a given semester. Analyzing the data in such grain provides insight of which courses may be more associated to the student being retained. Specifically, this analysis emphasizes the hypothesis that *student retention can be caused by the combination of courses on each semester*. From this point of view, modifying the flowchart would diminish the chances of a student being retained (given that the student follows the recommendation).

After defining the granularity was necessary to transform some data to the utilization of association rules algorithm.

**First transformation**. From the semester the student enrolled in his programme, we can understand each new semester as of being his second, third, forth, semester etc, which can be understood as the student *relative semester* to his year of enrollment. We use the *relative semester* for sub-setting the dataset. Notice that this is different than sub-setting by year of enrollment. Specifically, if we sub-set by semester of enrollment, students of different year of enrollment should not occur within a given sub-set, while in a given sub-set of *relative semester* can contain students of different years of enrollment, so long both students have been long enough on the programme to have in common a second, third semester, etc. We decided to use the *relative semester* sub-setting because it seems intuitive to test our hypothesis that changes can be done to the programme flowchart in order to reduce retention.

**Second transformation**. Lastly, we should be aware that a student may be enrolled on a given course for his second, third or fourth time (maximum amount of times that a student is allowed to fail *each* course). To do so, we transformed each of our variables so that their label indicates the attempt count (e.g. Calculus A\_1, Calculus A\_2 instead of Calculus A for both rows).

**Third transformation**. Over the time, the flowchart of a program can be modified, either their prerequisites or courses code. As the present work aims to examine the student retention according to their course enrollment, the most current flowchart was used and only students of being part of it was selected to be evaluated.

After analyzing the preprocessing done, the Apriori algorithm was carried out in order to generate our association rules for each course by each semester.

# 5. Analyzing of Course Retention

As described, students retention is denoted when they lacks some course prerequisite. Thus, each course prerequisite can be identified as the course that implies on a student retention. In other words, the previous course of a semester, which are the prerequisites of the current courses, were the main cause of student retention in his current semester.

In order to illustrate this situation, our Architecture programme, in its third semester, has ARQ022 (Mathematics II) as a course prerequisite of ENGC021 (Mechanical Structures); and ARQ016 (Atelier I) as a course prerequisite of ARQ024 (Atelier II); Thus, we can state that ARQ022 and ARQ016 are potential courses that can retain students because of their prerequisites. Considering 100 students following the third semester, where 35 has failed due to grade on ARQ002 and 30 have failed due to grade on ARQ016 , we can conclude that ARQ002 retains 35 students and ARQ016 retains 30 students, no matter on what semester these courses are followed. The retention motivation of ARQ002 (the highest retain rate course on the third semester) was 100% failed due to grade, 0% per absence and 0% not enrolled.

In order to obtain these retention rates, an analysis on each semester of each of the 70 programs at UFBA was carried out. If the course does not have any prerequisite defined for a semester, there are no guarantees to affirm that this programs can induce student retentions. Thus, programs that do not have prerequisites on their curriculum grade, according to our proposed heuristic, never retain students.

#### 6. Experiments with Association Rules

We used the Apriori association mining algorithm [Agrawal et al. 1994] to mine our rules. Authors in [Agrawal et al. 1993] define an association rule as follows: An association rule is an implication of the form  $X \to Y$ , where  $X \subset I$ ,  $Y \subset I$  and  $X \cap Y = \emptyset$ . Where  $I = \langle i_1, i_2, ... i_n \rangle$  be a set of literals, called items, and D is a set of transactions, where each transaction T is a set of items such that  $T \subseteq I$ . A transaction T contains X (a set of some items in I) if  $X \subseteq T$ . The rule  $X \to Y$  holds in the transaction set D with confidence *c* if *c*% of transactions in D that contain X also contain Y. The rule  $X \to Y$  has support *s* in the transaction set D if *s*% of transactions in D contain  $X \cup Y$  [Agrawal et al. 1993].

The lift measure allows evaluating the dependency between a rule precedent and its consequent. In general, this measure expresses how much the resulting becomes frequent when the precedent occurs. For association rules  $X \rightarrow Y$ , the lift indicates how much more common it becomes Y when X occurs. The lift is defined as follows: Lift( $X \rightarrow Y$ ) = Conf( $X \rightarrow Y$ )/ Sup(Y) when the lift is greater than 1 means that the rule tends to be significant. Therefore, the lift was set to 1.1 in this work [Brin et al. 1997].

During the rule analysis, we have observed that some programs had high failed rates or high approved rates. In order to better analyze these kinds of results, we have used the rare rules approach[Koh and Rountree 2009]. Association rules are based on the discovery of frequency items in a dataset. However, there exists some relevant information in some dataset that is on respect their non frequency items, called infrequent. Despite the fact that association rules discover frequent and infrequent items, there are some specific algorithms to find these types of rules, called rare rules [Koh and Rountree 2009].

The dataset of rare items are those which appear together in few transactions in a database[Koh and Rountree 2009]. The rare rules have a low support measure and a high confidence, different from frequent rules that have high support and high confidence.

The most simple way to find rare rules is using the Apriori algorithm[Agrawal et al. 1993] within a low support value. It is important to high-light that this simple way does not mean in poor nor infeasible results. The main difference between these algorithms is the number of rare rules generated, i.e. in the simple procedure it generates more rules to be analyzed. The use of rare rules in our approach is due to specialist domain can identify potential rules inside those rare rule set.

Given this, we applied the Apriori algorithm in each semester (support >= 10 %, confidence >= 75 %, lift = 1.1), for each course generating a number of rules. Case no relevant rule was discovered, it starts searching for rare rules (support between 1% and 10%, confidence = 75%, lift = 1.1), due to the following factors: i) courses with new programme flowchart results on few students linked with those charts; ii) low frequency rules, but with high relevance to domain specialists. In order to confirm that this would not bias our results, we used another tool [Webb 2000] to validate our rules and measures used. We have obtained the same results thus revalidating our significant rules obtained.

# 7. Results

From the heuristic retention defined it was possible to investigate student retention throughout the semesters obtaining information, such as: i) courses that retain more students in a semester; ii) courses with the highest retention rate per semester; iii)

courses with a high retention rate in different programs, and; iv) student retention rate per semester.

Within our first analysis it was possible to identify that a course retention increases during the semesters. This implies that students who are retained in a semester tend to continue retained in subsequent semesters. From 20,442 students following a course in the second semester, 7,395 were retained (36%). The retention rates increase in the third semester, where 9,237 students were retained from the total of 18,382 (50%). In the sixth semester, the retention rates increases over 60%, and finally, in the eighth semester, this rates achieve 80% of retained students.

Regarding Programs, when following the second semester, the Programs that most retain students are: Oceanography and Computer science, with respectively 79% and 71% of their students being retained. Considering the third semester, the Veterinary Medicine program retains about 96% and Pharmacy program retains 87%. The Genre and Diversity program retains 100% of their students and it is the most program that retain students on our University. Veterinary Medicine program in the fifth semester, retains 98% of their students. The Veterinary Medicine was the program with the highest retention rates per semester. Table 1 presents a retention ranking per semesters of some programs.

Semester	Program	Retained in a Semester	Total Students per Semester	Retention on a Program
2	Oceanography	610	765	79%
2	Computer Science	1540	2140	71%
3	Veterinary Medicine	4522	4669	96%
3	Pharmacy	2550	2925	87%
4	Decor	340	345	98%
4	Pharmacy	822	876	93%
5	Veterinaire Medicine	3444	3486	98%
6	Gender and Diversity	102	102	100%
7	Geography	376	380	98%
8	Veterinary Medicine	1890	1911	98%

#### Table 1. Programs that most retain in our University

Moreover, it was also important to note the courses which were retaining more programs per semester. Overall, FCHE28 (Supervised Internship in Gender and Diversity I) course from Gender and Diversity program, was the most retained in the sixth semester, within 100% of their students, it means, all students who enroll in their sixth semester could not attend FCHE59 (Supervised Internship in Gender and Diversity II), because they were retained by FCHE28, that is a prerequisite of FCHE59. In early semesters, the course that retains more students in the second semester is MEVA40 (General Chemistry - Analytical and Organic) from the Animal Science program. This course is the prerequisite of MEVA48 (Animal Biochemistry) course and the student retention rate were of 65%. Still on the second semester, the course MATA42 (Discrete Mathematics), course from the Computer Science program and the Information Systems program, retained 53% of their students. Table2 presents other courses and their highest retention rate per semester.

It is important to note that initial courses which are retaining students should be analyzed on advance due to their cascade effect. For instance, if a student was retained in the second semester of the program X, it is impossible to attend the course Y because X is prerequisite of Y, and consequently, Y is prerequisite of Z and so on. In addition, there are courses that are prerequisites of various other courses, such as MATA42. This kind of course needs a special attention, because a high retention rate will result in many student retention.

Semester	Course	Course Retention per Semester	Programs
2	MEVA40 (General Chemistry-Analytical and Organ-	65%	Zootechnology
	ica)		
2	MATA42 (Discrete Mathematics)	51%	Computer Science, Information Systems
3	MATD04 (Data Structure)	83%	Computer Sciente
3	MAT027 (Statistics IV)	80%	Veterinary Medicine
4	MATA58 (Operating Systems)	78%	Computer Science
4	MATA55 (Object-Oriented Programming)	78%	Computer Science
5	FIS005 (Theoretically mechanics General I)	91%	Physical
6	FCHE28 (Supervised Stage on Genre and Diversity)	100%	Genre and Diversity
7	BIOC13 (Zoological Diversity)	82%	Biological Sciences
8	MEVA20 (Final Project II)	98%	Veterinary Medicine

#### Table 2. Courses that most retain

With the usage of association rules it was possible to identify the correlation between courses and student retention. For each analyzed semester, approximately 2,000 rules were generated. The most important rules selected were found on the second, third and fourth semester. The selection and evaluation of the rule set were made by some specialists (Chief of Courses, Program Coordinator, Teachers). Thus it was possible to propose some suggestions regarding the organization of the flowchart of some programs in order to reduce retention rates Below, it is presented some relevant rules:

- 1. In Mathematics License Program, 150 students were enrolled in courses when attending their second semester. For this course this rule was obtained: 19 of the 26 students who failed Euclidean Plane (MATC21) course (second semester) for the first time, also have failed Differential and Integral Calculus (MATC20) course (First Semester) for the second time. There is therefore a tendency of students to fail MATC20, when he abandons MATC21. Thus, in this case, It is suggested that MATC20 is prerequisite for MATC21, aiming to increase the number of students who do not enroll in both courses simultaneously, consequently reducing the failure rate for MATC21.
- 2. In the course of Biological Sciences, there are 230 students enrolled in the second semester. For this course this rule was obtained: from the 41 students who failed in BIO010 (Cell Biology) and approved in BIOB24 (Formulation of Research Problems) and BIOB26 (Origin and Evolution of prokaryotes and eukaryotes), 31 of them were approved in BIOB27 (Introduction to Zoologica diversity). This rule shows that BIO010 (Cell Biology) course has a high failure rate even though there are prerequisite for pursuing it. It is noticed that students are approved in other courses on the second semester, but fails in BIO010. Thus, this course can be offered in subsequent semesters, reducing the failure rate, since in the second semester, students are not yet ripe for studying this level of subject in the second semester. They lack some maturity on the subject.
- 3. In the Chemistry course, there are 300 students enrolled in the second semester. For this course the rule obtained was: from 43 students that failed in QUI138 (Organic Chemical Fundamentals III), 32 of these students were not retained. This course is a prerequisite in later semesters, thus it could soon contribute with retention rates. Moreover, the students who failed are no-retained students, i.e., they were approved on their necessary prerequisites, but still they do not achieve their

approval. Thus, it is possible to suggest that this course be offered in subsequent semesters observing if there is not other necessary prerequisites for pursuing it.

#### 8. Threats of Validity

There are a couple of threats to this work that are important to be highlighted: First, as described in Sections 4 selected the most current curriculum of the courses, which caused a reduction in the data set because some curriculums were changed during a recent period. Second, a heuristic retention was defined to characterize the state of retention student in each semester, but when there are not prerequisites in the semester cannot be classified as retained students. Other heuristics can be used for such cases: minimum workload each semester, minimum of approval for courses in a semester. Third, the analysis of intent of this study looked at data regarding student performance. However, socioeconomics and cultural data can be useful for a better understanding of retention at our University. Fourth, while variables dichotomizations were theoretically validated by a domain expert, the chosen intervals might not work for other programs. Fifth, the lack of more specific course information limits the strength of conclusions. For instance, it is known and observed that students make choices based on which professor is offering the course. Such choices are not taken into account in our work due to privacy reasons.

#### 9. Conclusion and Future Work

In this work, we have analyzed the retention of students in the programs of UFBA from the hypothesis that the combination of courses that students take on each semester may impact in the chance of them to be retained. As we already highlighted over the introduction, the non regular flow of academic student's life cause some prejudices to the university, especially in programs under implementation because they must offer the required courses per semester. Having observed a subset of students that is following the regular recommended flow chart of the programs.

Thus, it was possible to analyze the retention throughout the semesters of programs through the results of students in courses. Therefore, it was possible to identify the rates of retention of the semester of a program and courses that more retain students. In view of the findings observed in the UFBA currently there is a high rate of retention in all semesters and certain courses contribute more to this retention rate.

Therefore, association rules were used to identify correlations between courses and retained students, in order to obtain information about the impact of the flowchart in student retention. Thus it was possible to propose some changes into the flowchart so as to reduce the retention rate of a program.

The analysis of retention is a complex study involving several factors. In this study, student transcripts were used to classify the student retention. In future work, we intend to add other factors that may be leading student retentions, such as socio-economic and cultural. Furthermore, we intend to use probabilistic models, specifically Bayesian networks and Markov chains to identify future probabilities regarding student retention.

# References

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*, volume 22, pages 207–216.

- Agrawal, R., Srikant, R., et al. (1994). Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, volume 1215, pages 487–499.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.
- Barros, R. P. and Mendonça, R. (1998). Consequências da repetência sobre o desempenho educacional. *Projeto de Educação Básica para o Nordeste*, Série Estudos(7).
- Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. In *ACM SIGMOD Record*, volume 26.
- Campello, A. d. V. C. and Lins, L. N. (2008). Metodologia de análise e tratamento da evasão e retenção em cursos de graduação de instituições federais de ensino superior. XXVIII Encontro Nacional de Engenharia De Produção, RJ, 13p.
- Filho, N. d., Mesquita, F., Marinho, M., Lopes, A. A., Lins, E., and Ribeiro, N. e. a. (2010). Memorial da universidade. Technical report, Universidade Federal da Bahia.
- Koh, Y. S. and Rountree, N. (2009). Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection. Information Science Reference - Imprint of: IGI Publishing, Hershey, PA.
- Manhães, L. M. B., da Cruz, S. M. S., Macário Costa, R. J., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In Anais do Simpósio Brasileiro de Informática na Educação, volume 1.
- Nandeshwar, A., Menzies, T., and Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, 38(12):14984–14996.
- REUNI (2012). Reestruturação e expansão das universidades brasileiras.
- Rigo, S. J., Cazella, S. C., and Cambruzzi, W. (2012). Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In *Anais do Workshop de Desafios da Computação Aplicada à Educação*, pages 168–177.
- Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics*, 40(6):601–618.
- Silva, C. V., Santos, M. S., Claro, D. B., Silva, V., Silva, M., Ribeiro, S., Telles, A. R., and Lopes, D. (2013). Mining retention rules from student transcripts: A case study of the information systems programme at a federal university. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 24.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2009). *Introdução a Mineração de Dados*. Ciência Moderna.
- Webb, G. I. (2000). Efficient search for association rules. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.*
- Yu, C. H., DiGangi, S., Jannasch-Pennell, A., and Kaprolet, C. (2010). A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science*, 8(2010):307–325.
- Zhang, Y., Oussena, S., Clark, T., and Kim, H. (2010). Use data mining to improve student retention in higher education a case study. In *ICEIS (1)*, pages 190–197.