Computers in Human Behavior xx (2017) xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

An educational data mining approach to assess students' performance through Self-Regulated Learning behavioral analysis

<omit to review>[*,a], <omit to review>[b], <omit to review>[b,c], <omit to review>[c], <omit to review>[c], <omit to review>[c]

a <omit to review>
b <omit to review>
c <omit to review>

a r t i c l e   i n f o

Article history:
Received xx November xx
Received in revised form
xx February xx
Accepted xx February xx
Available online xx February xx

Keywords:
Xxx
Xxxx
xxxx
Self-regulated learning
xx

a b s t r a c t

The increasing use of the Learning Management Systems (LMSs) is making available an ever-growing volume of data from interactions between teachers and students. By analyzing all these data, it is possible to gain insights into how students self-regulate their learning process and how it impacts their academic performance. This study aimed to develop a model capable of predicting students' academic performance based on indicators of their self-regulated behavior in LMSs. To accomplish this goal, we analyzed behavioral data from a LMS platform used in a public University for distance learning courses, collected during a period of seven years. With these data, we developed, evaluated, and compared predictive models using four algorithms: Decision Tree (CART), Logistic Regression, SVM, and Naïve Bayes. The Logistic Regression model yielded the best results in predicting students' academic performance, being able to do so with an accuracy rate of 0.893 and an area under the ROC curve of 0.9574. Finally, we conceived, implemented, and evaluated a dashboard-like interface intended to present the predictions in a user-friendly way to tutors and teachers, so they could use it as a tool to help monitor their students' learning process.

1. Introduction

Given the evolution in the development of technological artifacts for education, computational environments for teaching and digital contents start to be seen as essential instruments in the teaching process. Also, the use of these online learning platforms is contributing to the expansion of distance learning courses (Wang, Doll, Deng, Park, & Yang, 2013), which in turn generate huge volumes of data as a result of the users' interactions with the system. These data reflect the students' behavior and, if correctly explored, can offer important clues about how students interact and learn (Koedinger, et al., 2010).

However, keeping track of all students' interactions in different communication tools is one of the most exhausting and challenging tasks for teachers nowadays. Given the growing size and heterogeneously of classes and the limited time available, teachers tend to adopt a passive posture, limited to answering messages and grading assignments. Noticing the individuality of each student and guiding his/her learning process becomes, in the distance learning context, a challenge.

In virtual learning environments, the students can count with flexibility of time, space, and features, which also brings the need to be more responsible and autonomous. As a result, students have a greater level of control over its own leaning process (PINTRICH, 1999).

It is also necessary to take into account that each student have different levels of proficiency in establishing self-regulated learning strategies. Those who are more skilled in this ability can control the rhythm and direction of their learning process and also manage a set of characteristics in the learning environment, thus making choices that can aid in their cognitive structure development (BROADBENT and POON, 2015).

On the other hand, many students have difficulties self-regulating and self-monitoring their routine learning tasks. These students, in most cases, come from a past of dependency on teachers and administrative staff for regulation, which is common in traditional pedagogy. Furthermore, the conditioning of these students by traditional face-to-face education makes them more dependent on personalized guidance and direct supervision of teachers in order to manage their assignments.

In this context, recent findings by authors like Kizilcec et al. (2016) state that such socioemotional abilities — like self-regulation and self-monitoring — influence individual's capacity to learn. Such researches aim to understand which abilities affect the learning process, based on each student behavioral profile in the virtual learning platforms.

Such phenomena have been addressed by Cognitive Psychology scholars, which aims to understand the self-regulation learning process. This area focuses in the study of strategies to manage resource, time, effort, environment, interaction, and help-seeking management (PINTRICH, 1999).

In this way, our work aims to answer the following research question: *Is it possible to develop a predictive model capable of detect students' academic performance based on their self-regulatory behavior?*

From this main research question, we derived the following sub-questions:

"What are the behavioral variables recorded in LMS environments that best describe the self-regulated learning strategies?"

"What are the relationships between self-regulation factors and students' academic performance?"

The remaining of this paper is organized as follows: Section 2 describes the background and related work regarding the relationship between *self-regulated learning* (SRL) and academic performance. In Section 3 we present the used methodology. In Section 4 we present the results obtained from the use of educational data mining techniques. In Section 5 we describe the application developed to present the data mining results in a user-friendly way. In Section 6 we discuss the results and its implications and, finally in Section 7 we draw some conclusions and perspectives for future work.

2. Background

2.1. *Self-regulated learning (SRL)*

Self-regulated of learning is a process composed of cognitive abilities that allow the development of strategies in the face of tasks and challenges. (WANG, SHANNON & ROSS, 2013). According to this understanding, the autonomous, active, and planned action of an individual in the teaching-learning process is crucial.

Other researchers define SRL as a process by which individuals, after establishing goals that reflect their expectations, develop strategies to achieve these goals, thus creating the necessary conditions for learning (ZEIDNER, BOEKAERTS & PINTRCH, 2000); (Silva, 2004); (Veiga Simão, 2006).

Gonçalves (2010) believes that students should be explicitly encouraged to adopt an active role in the development of their own skills instead of acting only as spectators during the classes, as is common in traditional education.

Zimmerman (1989) states that a student who is able to self-regulate his/her learning process has an active, metacognitive, motivated, and behavioral posture. Zimmerman also says that there will be inexperienced students who have a low capacity to self-regulate their learning due to the individual's level of sociocognitive development. In the same way, there will be experienced students who have developed their self-regulatory learning skills to a greater extent. Zimmerman (1989) also points out the main differences between students who can and cannot self-regulate their learning process.

In summary, we can consider that self-regulated learning is a process of knowledge acquisition that is active, constructivist, goal-oriented, and guided by three mental functions: metacognition, motivation, and emotions. We can also state that self-regulated learning is a learner's ability to develop the knowledge, the strategic competencies, and the necessary attitude in order to facilitate learning both in school and in other contexts.

2.2. *The relationship between SRL and academic performance*

In this work, we consider *academic performance* as the achievement of the expected result in a task, assignment, exam, or the degree of understanding shown by a student regarding some topic, which is normally quantified in a numerical scale (RICHARDSON, ABRAHAM and BOND, 2012).

A large number studies have been showing a positive correlation between the presence of strong SRL abilities and good academic performance in traditional face-to-face learning (Wang, Doll, Deng, Park, & Yang, 2013); (AGUSTIANI, CAHYAD & MUSA, 2016). These studies focus on metacognitive activity, time management, and effort regulation (RICHARDSON, ABRAHAM & BOND, 2012). Therefore, by understanding students' SRL behavior, teachers can them improve their academic performance.

However, there are few studies dealing with the relationships between SRL and academic performance in distance learning (e-learning) contexts (BROADBENT & POON, 2015). The majority of studies in this topic are prospective ones (ChanLin, 2012); (CHO and SHEN, 2013), followed by experimental (Chang, 2010) and cross-sectional works (Klingsieck, Fries, Horz, & Hofer, 2012); (WANG, SHANNON & ROSS, 2013).

Regarding the prospective studies, Carson (2011) measured the degree of success of a student using the online learning scale LASSI (LLO). The author provides evidence of performance predictability based on the LASSI scale items with overall accuracy of 67%. In another study, Cho and Shen (2013) discuss the relationship between students' individual goals and their academic performance. The results showed a positive correlation between the amount of time students spend on the platform, time management, and their academic performance.

A cross-sectional study that deserves attention is Klingsieck *et al.*, (2012). This work emphasizes the problem of student procrastination and its implications for the performance of academic activities. Although there are studies on student procrastination, there are relatively few studies to understand the effects of procrastination in e-learning environments. Among the studies regarding self-regulated learning, procrastination is still an underexplored topic, which requires more studies focusing on e-learning contexts.

2.3. *Strategies for measuring the self-regulated learning process*

Over the years, several approaches have been developed to measure the self-regulated learning process. However, it remains a challenge to quantify the students' self-regulatory skills during specific learning tasks. Such measurements are usually performed using questionnaires with scaled answers.

Weinstein, Schulte and Palmer (1987) proposed one of the first scales focused on measuring student self-regulation: The Learning and Study Strategies Inventory (LASSI). This scale is a self-report questionnaire with 77 items intended to assess the learning strategies of higher education students. The scale is composed by ten subscales, namely: attitude, motivation, time management, anxiety, concentration, information processing, selecting main ideas, study aids, self-testing, and test strategies.

Later *et al.* (1993) developed the MSLQ scale — Motivated Strategies for Learning Questionnaire. This scale is composed by 81 items and aims to evaluate the role of motivation and the use of learning strategies by students working on a specific course or topic. Regarding the motivation aspect, it is analyzed by observing three components: expectations (self-efficacy and sense of control), values (intrinsic and extrinsic goals, and task value), and affective factors (anxiety about exams). Concerning the use of learning strategies, the MSLQ scale assesses cognitive and metacognitive strategies (repetition, elaboration, organization, critical thinking and metacognition) and resource management strategies (time and place for studying, effort regulation and help-seeking).

Unlike the aforementioned instruments, Zimmerman and Pons (1986) developed a qualitative measurement instrument called SRLIS — Self-Regulated Learning Interview Scale. This approach collects the data by using a structured interview, which analyses 14 strategies: self-assessment, information organization and transformation, goal setting and planning, information seeking, note taking and monitoring, Environment Structuring, self-consequence, mnemonic strategies, seeking help from peers, seeking help from teachers, seeking help from adults, reviewing notes, reviewing previous exams, and reviewing books.

It is important to note that the majority of approaches intended to measure the self-regulated learning process use *ad hoc* data collecting instruments, like questionnaires and observation. However, one drawback of these approaches is that they cannot be used to measure the students' self-regulatory skills in real time.

In distance learning environments, especially virtual learning environments, it is very important to have information about the students' self-regulatory skills in real time. The use of Educational Data Mining (EDM) techniques can play a crucial role in meeting this requirement, since it makes it possible to measure several cognitive and metacognitive abilities without requiring the use of questionnaires or qualitative research approaches. In the next section, we discuss EDM in more detail.

2.4. *Educational Data Mining and SRL*

With the widespread use of computerized systems in schools and universities, the volume of data generated and stored in databases is growing steadily (Rigo, Cambruzzi, Barbosa, & Cazella, 2014). Among the factors that has been contributing to this growth are: the popularization of Information and Communication Technologies, the emergence of virtual learning environments, the consolidation of the distance and blended learning modalities, the use of computerized management systems in educational institutions, virtual communities and its communication tools, and the online content sharing (Cambruzzi, W. L, et al., 2012).

This increasing availability of educational data has provided an understanding of the self-regulatory behavior in learning through the application of Educational Data Mining (EDM) techniques. Several studies have demonstrated the potential of applying EDM techniques in order to discover self-regulatory patterns in behavioral data stored by learning platforms. Some examples of studies using this approach are Schoor and Bannert (2012); Sabourin, Mott and Lester (2012); Bondareva, Conati, et al. (2013), Nussbaumer, Hillemann, et al. (2015); Segedy, Kinnebrew, and Biswas (2015); Sonnenberg and Bannert (2015); and You (2016).

One study that deserves attention is You (2016). In this study, the researcher sought to identify behavioral indicators for self-regulated learning in LMS systems as a way of predicting student success at the end of a given course. The author used linear regression in order to build the models and identify significant variables. However, the study presented limitations regarding the choice of the explanatory variables: only variables related to *number of accesses to the environment*, *time spent visualizing resources*, and *delay in finishing of assignments* were used.

Two other techniques commonly used in EDM are Mining of Educational Processes (SCHOOR & BANNERT, 2012); (SONNENBERG & BANNERT, 2015) and Student Classification (SABOURIN, MOTT, & LESTER, 2012); (BONDAREVA, 2013).

Clustering is also commonly used in EDM. This technique seeks to group students by their profiles, which are identified according to the students' behavior in virtual learning environments. Nos trabalhos de (LAWANTO, 2014) and (COLTHORPE, 2015) The researchers aimed to identify groups of behavioral patterns of self-regulation. It is important to note that a given technique can be considered as part of more than one research area. The clustering technique, for instance, has its origins in Statistics, but researches in Learning Analytics and EDM are further developing it, proposing new algorithms and applications.

The next section presents the research method used in this work and also discusses the EDM and Statistics techniques used to develop the predictive models.

3. Method

The method used in this research followed the CRISP-DM process (Cross Industry Standard Process for Data Mining). CRISP-DM is an industry standard originally developed by NCR Systems Engineering Copenhagen, Daimler Chrysler AG, SPSS Inc. and OHRA (Chapman, et al., 2000).

3.1. *Phases of the CRISP-DM process*

For the purposes of this work, we instantiated the CRISP-DM process to an educational setting, adapting its six phases as follows: understanding of the educational domain to be modeled, understanding of the educational data involved, data preparation for applying EDM techniques, modeling, model evaluation, and implementation of the educational solution. Figure 1 shows these phases and their inter-relations.

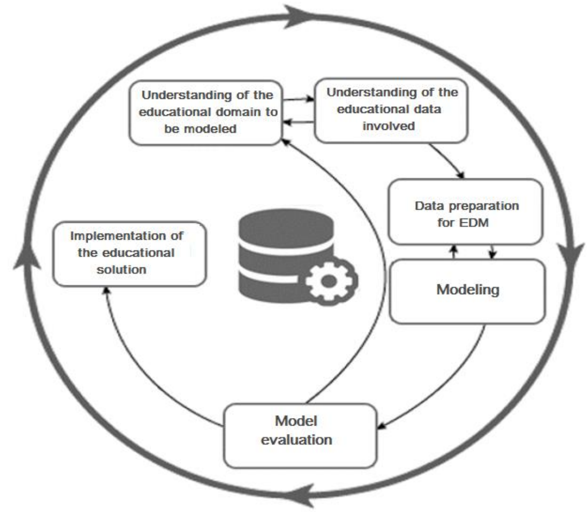

Fig. 1. The CRISP-DM methodology cycle adapted to EDM projects.

In Sheth and Patel (2010), CRISP-DM is presented as a suitable process for educational data mining projects.

The first phase**, understanding of the educational domain to be modeled,** was one of the most important phases in the whole process. In this phase, we identified the needs of the institution that participated in the experiment, following two steps: first, we conducted a literature review about current challenges in distance learning, focusing in the self-regulated learning process and related theoretical models; then, we conducted several interviews with domain specialists from the University of Pernambuco Distance Learning Department (NEAD/UPE) in order to understand how the institution and its students use the Moodle LMS platform.

In the second phase**, understanding of the educational data,** we identified the key behavioral variables available in Moodle's database that could quantify the students' self-regulatory behavior. As criteria for selecting the variables, we used the constructs proposed by Pintrich's theoretical model (1999), namely: (ES) Environment Structuring; (HS) Help-seeking; (TS) Task completion strategies; (TM) Time management; (GS) Goal-setting; and (SA) Self-assessment.

In the third phase**, data preparation for EDM,** we used several SQL scripts to extract the desired variables from the Moodle database. The used dataset contains data from four undergraduate majors, including 202 distinct courses, which took place over a period of 66 semesters, totaling 18.9 gigabytes coming from twelve platforms with different versions (Moodle 2.x to 3.x) and with DBMS systems ranging from *Postgres* to *Mysql*. From this dataset, we selected 32 variables that represented students' actions related to a self-regulated learning behavior. After extracting the desired variables, we preprocessed the dataset by applying data dimensionality reduction procedures, categorization, and cleansing. Besides that, we removed from the dataset courses that used some particular assessment strategy, different from the majority of other courses in the University.

In the fourth phase, modeling, we built a set of classification models intended to predict the students' performance, based on the variables in the dataset that describe their self-regulatory behavior in the Moodle LMS.

Finally, in the fifth phase, evaluation, we used the values of Accuracy, Sensibility, Specificity, and Precision and ROC curve. Shows the results of these measures for each classifier (Logistic Regression, Decision Tree, SVM, and Naive Bayes).

3.2. *Implementation of the educational solution*

As a way of materializing the results of the *modeling* phase, in this phase we developed a dashboard-like software interface in which the professor/tutor could monitor the students that, according to the classification models, tend to show low academic performance due to their self-regulatory behavioral patterns.

The development of this interface followed three steps: first, we created low-fidelity prototypes, then specified a software architecture and developed high-fidelity prototypes, and finally validated the solution with professor and tutors.

4. Experiments and Results

4.1. *Identification and retrieval of SRL behavioral variables*

The first step in the process was to quantify the number of students in each subject. Table 1 shows the number of students by subject in the database used for this research. As we can see, Pedagogy has the greater number of students, followed by Biology.

Table 1
Number of students per subject.

Subject	Frequency	%
Business Administration	2,892	9.57%
Biology	6,526	21.60%
Languages	6,297	20.84%
Pedagogy	14,502	47.99%
Total	**30,217**	**100.00%**

After integrating the datasets made available by the institution, we had unified database with data from 30,217 students. Then, we identified in the available log data the variables that could represent the constructs related to the self-regulation learning process, reaching a set of 33 variables that could be used to measure this process in real time. Table 2 describes in detail these variables.

Table 2
SRL behavioral variables.

	Variable	**Description**
(ES)	var01	Number of different places (IP addresses) from which the student accessed the learning environment.
(HS)	var02	Number of messages sent by the student to Professor(s) using the learning environment.
	var03	Number of messages sent by the student to Tutor(s) using the learning environment.
	var04	Total number of messages sent by the student using the learning environment.
	var05	Total number of messages received by the student in the learning environment.
	var06	Number of threads created by the student in Q&A forums.
	var07	Number of threads in the Q&A forum
	var08	Number of forum threads created by the student that were answered by other students.
	var09	Number of forum threads created by the student that were answered by the Tutor or Professor.
	var10	Number of different classmates to whom the student sent messages in the learning environment.
(TS)	var12	Number of times the section "Contents" (which lists the files describing the course syllabus) was visualized.
	var13	Hour of the day when the student most frequently worked on his/her assignments.
	var14	Time of the day (morning, afternoon, evening, night) in which the student most frequently worked on his/her assignments.
	var16	Number of assignments handed in by a student after the deadline, per course.
	var17	Average time between the moment an assignment is given and the moment the student completes it.
	var18	Number of times a student accesses the forum (*pageviews*)
(SA)	var20	Number of responses to a thread in the forum (denotes the action of reconsidering the opinion on the subject)
	var21	Number of times a student accesses the grades report
	var22	Number of times the student viewed the activities
	var23	Number of views the activity notes
(TM)	var24	Weekly average of the amount of times the student accessed the learning environment.
	var25	Average time between the moment a topic is created in the forum and the moment the student posts his/her first response to this topic.
	var28	Number of session timeouts.
	var31a	Number of times the student accessed the learning environment.
	var31b	Number of distinct days where the student accessed the course in the learning environment
	var31c	Number of distinct days where the student accessed the learning environment.
	var32a	Number of times the student accessed the learning environment by time of the day (mornings).
	var32b	Number of times the student accessed the learning environment by time of the day (afternoons).
	var32c	Number of times the student accessed the learning environment by time of the day (evenings).
	var32d	Number of times the student accessed the learning environment by time of the day (nights).
(GS)	var33	Number of assignments handed in by a student on time, per course.
	var34	Total number of messages posted by a student on forums.
	var35	Number of responses of a Professor for student's questions on forums.

After identifying the target variables listed in Table 2, we wrote a set of SQL scripts in order to extract the values for each of the variables from the database, grouping them by self-regulation construct. During this process, we faced some challenges: changes in Moodle's data model across versions, heterogeneity of SGDBs, lack of a consistent key for uniquely identifying the students, the huge volume of data involved, and syntax divergences in the scripts written for extraction the data — a consequence of the heterogeneity in SGDBs.

4.2. *Extraction of the "student performance" variable*

The institution's strategy for assessing the students' performance consisted in two in-person exams, four assessments through participation on forums, and two WebQuests (assignment).

The "student performance" variable — which we used as basis for the definition of the binary goal in the predictive model — was composed by weighting each of the grades as follows: in-person exam = 5.5, forum participation = 2.0, and e = 2.5. Makeup exams, when present, were also considered in the calculation.

The equation shown below summarizes the rationale behind the "performance" variable composition.

$$Performance = (\mathbf{Exams} * 5.5) + (\mathbf{Forum} * 2.0) + (\mathbf{WebQuest} * 2.5)$$

After defining "student performance" as a continuous variable, we needed to convert it to a "dummy" (binary) variable in order to build the binary classification models. We defined the two categories as: (1) grades below 4, and (0) grades equal or above 4. The first category (1) represents the students who are at risk of failing the course, and the second category (0) represents the students who have a chance of passing the course.

4.3. *Classification models' construction*

In order to build our models, we chose the classification function. According to Da Silva, Peres and Boscarioli (2016) classification is the process of assigning items from a collection to target classes, using a set of pre-classified items as training data.

We used the R[1] programming language to implement four classification models. The first one was based on Logistic Regression, the second one used CART Decision Trees, the third one used Support Vector Machines (SVM), and the fourth one used the Naive Bayes algorithm.

In order to train and test the developed models, we split the dataset in two subsets, as shown in Figure 2. The first subset makes up 70% of the dataset and was used for training, and the second subset makes up 30% of the dataset and was reserved for testing.

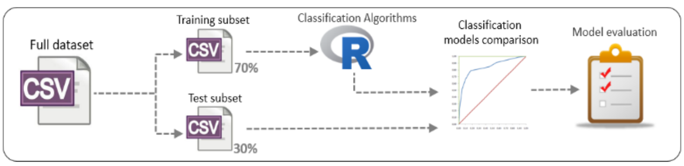

Fig. 2. Training and testing process.

Based on the results obtained by running the models on the test subset, we compared the performance of each one of the classification models we built (*Decision Tree, SVM, Logistic Regression, and Naïve Bayes*). In the next subsection, we describe the measures used for these comparisons.

4.4. *Evaluation of the developed models*

The evaluation process aims to validate the classification models by testing them with a dataset that was not used for training. This is a very important step, in which we can understand each model's characteristics and strengths before deploying them in a production environment (Chapman, et al., 2000). To perform this evaluation, we used the training subset mentioned in the previous section, which contained 11,824 records belonging to "class 0" (students with satisfactory performance) and e 9,330 records belonging to "class 1" (students with unsatisfactory performance).

We took into account five measures in the models' evaluation: Accuracy, Sensibility, Specificity, Pos Pred Value (Precision), and ROC curve. Figure 3 shows the performance of each classifier according to these measures.

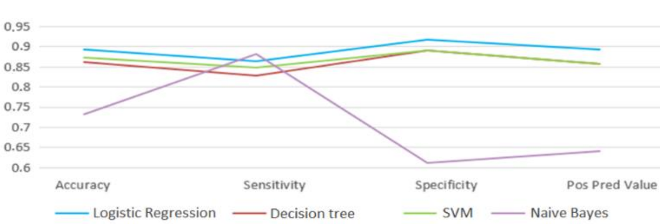

Fig. 3. Comparison of developed classifiers by Accuracy, Sensibility, Specificity, Pos Pred Value (Precision).

In "Accuracy", the Logistic Regression classifier presented the best result (0.893) followed by SVM (0.8717). In "Sensibility" (recall), the Naïve Bayes showed the best result (0.8815), followed by Logistic Regression (0.8633). This was the only measure where the Naïve Bayes classifier was superior to others. In the last two measures, "Specificity" and "Precision", the Logistic Regression model showed the best results: 0.9167 and 0.8921, respectively.

4.5. *Comparing the classifiers using the ROC curve*

In order to choose the best classifier, we used the ROC graph (Fawcett, 2006). This graph helps to visualize the performance of a binary classifier and also denotes the variations in sensibility and specificity for different cut-off values (Fawcett, 2006). Figure 4 shows a ROC graph for the four classifiers developed in this work.

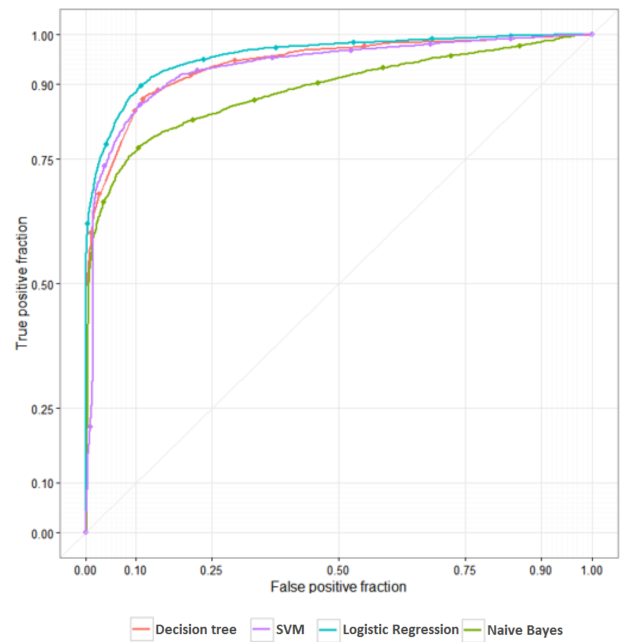

Fig. 4. Comparison between classifiers using the ROC curves.

The ROC curve analysis for the four classifiers yielded the following results: Decision Tree (0.9157), Naive Bayes (0.8850), Logistic Regression (0.9574), and SVM (0.9291). According to this metric, the Logistic Regression classifier can be considered the most powerful out of the four.

These results are in accordance to the studies of Sabourin, Mott and Lester (2012) and Sabourin *et al*. (2012), which showed promising results in the use of Logistic Regression for predicting the performance of students in elementary school. These studies used data from the Crystal Island platform — a game-based learning environment — related to self-regulatory behavior.

Given the good results attained with the Logistic Regression model, in the next section we present a detailed analysis of the modeling process used to build this classifier, as well as the details regarding the validation and parameter adjustments.

4.6. *The Logistic Regression classifier*

Using the Logistic Regression model, we could understand how the chosen binary goal variable ("performance") relate to a set of independent numerical and categorical variables — the ones that describe self-regulated behaviors on students. The Logistic Regression model was formulated following the Equation (1), where $\beta_0, \beta_1, \ldots \beta_n$ are the coefficients that explain the occurrence of a given event.

$$logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \qquad (1)$$

The biggest advantage of logistic regression in relation to other classification algorithms used in this research is the fact that it allows not only the prediction of the occurrence of a particular event of interest, but also the probability of its occurrence, as well as the odds ratios for each of the "self-regulation" variables. This feature helps in explaining the level of influence of each variable in the student's performance.

[1] https://www.r-project.org/

Our logistic classification model was built from a training subset of 21,154 records, which corresponds to 70% of the full dataset. We used the n-fold Cross-Validation technique with 10 folds in order to increase the generalization capacity of the trained model. This technique is widely used for building predictive models (STONE, 1974).

When building a logistic regression model, it is important to analyze the results of Wald's hypothesis test for each coefficient, which indicates how statistically significant it is. If the result of Wald's test is less than 0.05 for a given coefficient, we can consider it useful for composing the final model.

Table 3 shows the results of Wald's test. As we can see, the variables VAR06, VAR07, VAR08, VAR09, VAR13, VAR14, VAR17, and VAR23 are not statistically significant. Therefore, we removed these variables from the final refined model.

Table 3

Statistical significance of each variable before stepwise.

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.5961457  0.0665924  54.002  < 2e-16 ***
VAR01        0.0053145  0.0012980   4.094 4.23e-05 ***
VAR02        0.0971106  0.0272446   3.564 0.000365 ***
VAR03       -0.0140579  0.0045620  -3.082 0.002060 **
VAR04        0.0025958  0.0013660   1.900 0.057381 .
VAR05        0.0024620  0.0006952   3.541 0.000398 ***
VAR06       -0.2819494  0.2156452  -1.307 0.191054
VAR07        0.0582210  0.1272019   0.458 0.647164
VAR08        0.0173138  0.0212034   0.817 0.414182
VAR09        0.0280463  0.0183993   1.524 0.127430
VAR10       -0.0669658  0.0139889  -4.787 1.69e-06 ***
VAR12       -0.0689171  0.0289311  -2.382 0.017214 *
VAR13        0.0008412  0.0055663   0.151 0.879882
VAR14       -0.0393113  0.0371802  -1.057 0.290366
VAR16       -1.0816371  0.0592768 -18.247  < 2e-16 ***
VAR17        0.0013450  0.0017969   0.749 0.454144
VAR18        0.0110846  0.0018030   6.148 7.86e-10 ***
VAR20        0.3991621  0.0787712   5.067 4.03e-07 ***
VAR21        0.1089689  0.0137360   7.933 2.14e-15 ***
VAR22       -0.0190252  0.0096252  -1.977 0.048086 *
VAR23       -0.0165666  0.0211643  -0.783 0.433767
VAR24       -0.1338515  0.0721790  -1.854 0.063677 .
VAR25       -0.0060239  0.0004758 -12.661  < 2e-16 ***
VAR28        0.0013602  0.0007590   1.792 0.073119 .
VAR31        0.0175758  0.0061181   2.873 0.004069 **
VAR31b      -0.0156180  0.0027218  -5.738 9.58e-09 ***
VAR31c      -0.0106327  0.0021396  -4.969 6.72e-07 ***
VAR32a      -0.0224103  0.0056532  -3.964 7.36e-05 ***
VAR32b      -0.0146981  0.0055969  -2.626 0.008636 **
VAR32c      -0.0151420  0.0057026  -2.655 0.007924 **
VAR33       -1.6689937  0.0412373 -40.473  < 2e-16 ***
VAR34       -0.3363416  0.0189809 -17.720  < 2e-16 ***
VAR35        0.2908787  0.1120174   2.597 0.009412 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
```

Among the available approaches for assisting in the selection of the more significant variables for the construction of a model, we chose the Stepwise method. Stepwise is the process of statistical model estimation in which independent variables are added or removed from the model according to their discrimination power regarding the group of explanatory variables (Hair, Black, Babin, Anderson, & Tatham, 2009). Table 4 shows the list of significant variables after applying the stepwise method.

Table 4

Significance of SRL variables after stepwise

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.6022762  0.0659374  54.632  < 2e-16 ***
VAR01        0.0052704  0.0012947   4.071 4.69e-05 ***
VAR02        0.0984237  0.0271719   3.622 0.000292 ***
VAR03       -0.0137607  0.0045174  -3.046 0.002318 **
VAR04        0.0025634  0.0013643   1.879 0.060256 .
VAR05        0.0024759  0.0006843   3.618 0.000296 ***
VAR10       -0.0661682  0.0139603  -4.740 2.14e-06 ***
VAR12       -0.0686481  0.0288113  -2.383 0.017187 *
VAR16       -1.0815550  0.0590862 -18.305  < 2e-16 ***
VAR18        0.0109902  0.0017967   6.117 9.55e-10 ***
VAR20        0.4143110  0.0764235   5.421 5.92e-08 ***
VAR21        0.1057263  0.0133887   7.897 2.86e-15 ***
VAR22       -0.0189974  0.0095872  -1.982 0.047530 *
VAR24       -0.1308503  0.0719442  -1.819 0.068946 .
VAR25       -0.0060779  0.0004704 -12.921  < 2e-16 ***
VAR28        0.0013142  0.0007570   1.736 0.082548 .
VAR31        0.0177954  0.0060928   2.921 0.003492 **
VAR31b      -0.0156337  0.0026765  -5.841 5.18e-09 ***
VAR31c      -0.0108233  0.0021315  -5.078 3.82e-07 ***
VAR32a      -0.0226223  0.0056171  -4.027 5.64e-05 ***
VAR32b      -0.0149143  0.0055568  -2.684 0.007276 **
VAR32c      -0.0154217  0.0056644  -2.723 0.006477 **
VAR33       -1.6658986  0.0384280 -43.351  < 2e-16 ***
VAR34       -0.3320120  0.0178559 -18.594  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

After applying the variable reduction process using the Stepwise method, our final model was comprised of 23 variables. Among them, one belongs to the "Environment Structuring" construct (var01), five belong to the "help-seeking" construct (var02, var03, var04, var05, var10), three to the "task completion strategies" construct (Var12, var16, var18), three to the "self-assessment" construct (var20, var21, var22), nine to the "time management" construct (var24, var25, var28, var31, var31b , Var31c, var32a, var32b, var32c) and two to the "goal-setting" construct (var33, var34). We can notice that the constructs that holds most of the significant variables are "help-seeking" and "time management".

Regarding the "help-seeking" construct, *var02* and *var10* were the most impactful in the students' performance. The first one represents the number of messages sent by the student to his/her Professor (using the virtual learning environment), and the second one represents the number of distinct classmates to whom the student sent messages (virtual learning environment).

Regarding the "time management" construct, *var25* and *var32b* were the most impactful in the students' performance. The first one represents the "average time between the moment a topic is created in the forum and the moment the student posts his/her first response to this topic". In other words, this variable represents the time spent by the student to complete an assignment, starting from the time it is made available in the environment. It is important to note that this variable has a *negative* influence in the student's performance, that is, students that hand in their assignment sooner show better academic performance.

Equation (2) shows our final logistic regression model, which is represented by a *logit* function involving all the significant variables.

$$logit(p_{des.}) = 3.602 + var_1 * 0.005 + \cdots + var_{34} * (-0.332) \qquad (2)$$

After finding the most significant variables and building the final model, we calculated the pseudo-R squared and Cox-Snell's pseudo-R squared statistics for our model. The measures obtained through these statistics describe how much the model fits the research data, in other words, how well the model explains the variability, with the value 1 representing a perfect fit. Our model obtained the values 0.785 and 0.776 for the pseudo-R squared and Cox-Snell's pseudo-R squared statistics, respectively, which indicates that it can be used to build a good classifier.

We also evaluated the performance of our model by using a set of different measures, based upon a test subset containing 9,063 records, which corresponds to 30% of the complete dataset. Table 5 shows the results of this evaluation.

Table 5

Confusion matrix and evaluation metrics.

```
          Reference
Prediction    0    1
         0 4620  550
         1  420 3473

               Accuracy : 0.893
                 95% CI : (0.8864, 0.8993)
    No Information Rate : 0.5561
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7825
 Mcnemar's Test P-Value : 3.444e-05

            Sensitivity : 0.8633
            Specificity : 0.9167
    Pos Pred(Precision) : 0.8921
         Neg Pred Value : 0.8936
             Prevalence : 0.4439
         Detection Rate : 0.3832
```

```
Detection Prevalence : 0.4295
   Balanced Accuracy : 0.8900

    'Positive' Class : 1
```

According to the confusion matrix presented in Table 5, our classifier was able to make a correct prediction in 89.3% of the cases, which can be considered a good performance level according to the thresholds proposed by Faveiro, Belfiore, et al. (2009), who state that a "good model" makes a correct prediction in more than 85% of the cases. Besides accuracy, we also used other four measures to evaluate our model: *Sensibility*, *Specificity*, *Precision*, and *Area under the ROC curve*.

For the **Sensitivity** measure, our model attained the value of 0.8633 (86.33%). This measure is defined as the fraction of positive instances correctly predicted by the classifier, that is, the number of students at risk who were correctly predicted. Thus, we can state that of the 4,023 students at risk in the test subset, the model identified 3,474 correctly.

For the **Specificity** measure, our model attained the value of 0.9167 (91.67%). This measure is defined as the number of negative instances correctly predicted by the classifier, that is, the number of students who were correctly predicted as not being at risk. Thus, we can say that of the 5,040 students in the subset who were in satisfactory performance situation, 4,620 of them were correctly classified by the logistic model.

For the **Precision** measure, our model attained the value of 0.8921 (89,21%). This measure determines the percentage of records correctly identified by the classifier in the positive class. The higher the precision rate, the smaller the number of false positive errors committed by the classifier, that is, the higher the precision the smaller the number of students who are at risk, but the classifier classified them in the "satisfactory performance" class.

For the last measure, the **Area under the ROC curve,** our model attained the value of 0.9574. The ROC curve graph is built based on the rate of true positives and the rate of false positives — false positives are plotted on the x-axis and true positives on y-axis. Figure 5 shows the graph representing the area under the ROC curve for our model.

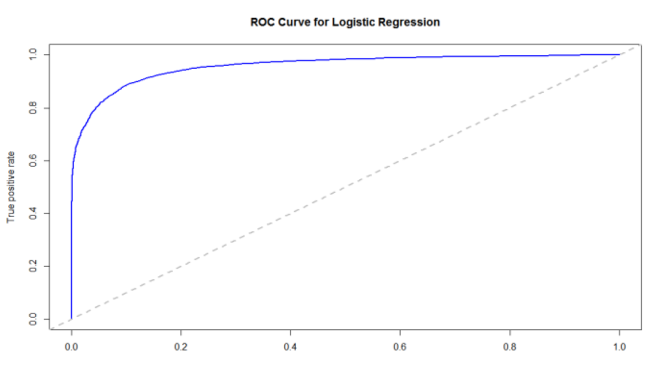

Fig. 5. ROC curve graph for the logistic regression model.

The area under the ROC curve is a portion of the area of the unit square (ROC space), which its values varying from 0 to 1. The larger the area under the curve, the better the overall performance of the classifier. Thus, according to the positive precision rate the logistic regression model was the best one among all the four classifiers used and evaluated in this research.

As a way to make the results of this research usable by teachers and tutors, we designed a dashboard-like GUI where they could monitor their students' self-regulatory skills. In the next section, we present and evaluate the developed interface.

5. Development of the monitoring dashboard

In order to generate a set of alternative software solutions, conceived from the discovery of knowledge in previous phases, we used two techniques: low-fidelity prototyping and high-fidelity prototyping with involvement of the users in the evaluation phase.

Our goal in this phase was to develop a dashboard in which teachers and tutors could monitor the data mining results — specifically predictive modeling — without the need of technical knowledge in the area. The following subsections describe the process of designing, implementing, and validating the solution.

5.1. *Low-fidelity prototypes*

In this phase, we designed use cases and low fidelity prototypes that could be refined throughout the development of the solution. Figure 6 shows the prototyped interface. In this interface, the teacher can select between the interfaces: Data overview and Performance Analysis.

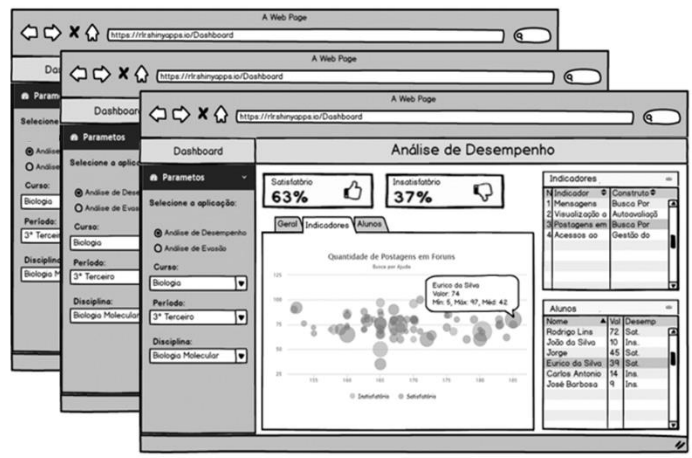

Fig. 6. Low-fidelity prototypes.

Through the low-fidelity prototyping technique we could identify the key features to be implemented. Also, this phase was extremely important since it allowed the users to test, validate, and modify the interfaces before the implementation phase. In this process, we created an interface similar to the one shown in Figure 6 — based on the initial requirements — and refined it through an iterative process.

5.2. *Architecture definition*

Based on the initial requirements and low-fidelity prototypes, we developed the interactive prototype that could be tested by end users. However, we first defined a software architecture to support the development of a high-fidelity prototype. Figure 7 shows this architecture, which is composed by three modules: (1) the Data Collection module, (2) the Intelligence module, and the (3) Interface module.

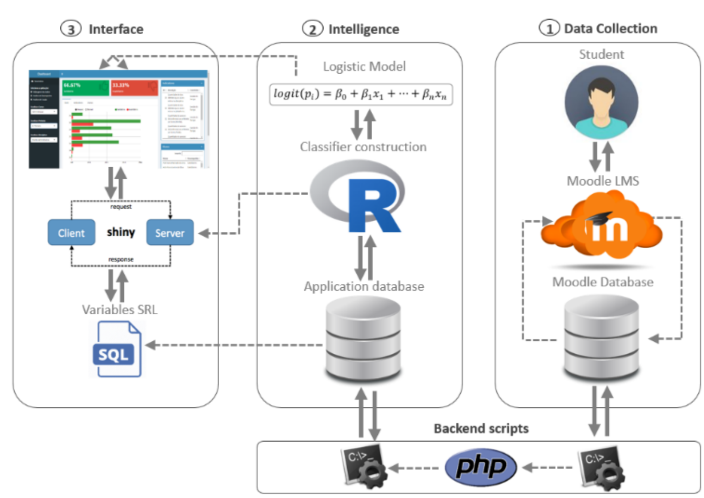

Fig. 7. The proposed software architecture.

The **Data Collection** module is responsible for retrieving the variables related to the self-regulated learning behavior of students in an external Moodle. The collection process runs as a batch job and is executed once a day, collecting the students' behavioral data and sending them to the Intelligence module.

The **Intelligence** module is responsible for executing and updating the predictive model — which was developed taking into account the self-regulation variables and the students' performance in past disciplines. Every time a course finishes and the students' behavioral records and performance indicators can be accessed, this module updates the predictive model using the new available data.

Finally, the **Interface** module is responsible for consuming the results of the intelligence module and presenting them to the user as a dashboard. In the next subsection, we present in detail the interfaces provided by this module.

5.3. *High-fidelity prototype*

In this phase, we implemented the proposed prototypes following the architecture described in Subsection 5.2. We chose the *Shiny* framework to create the interfaces and hosted the application code on Github.

The *Shiny* framework made it possible to code both the predictive model and the graphical interfaces using a single programming language: R. Figure 8 shows the *Data Overview* interface, which offers only descriptive charts, without predictions.

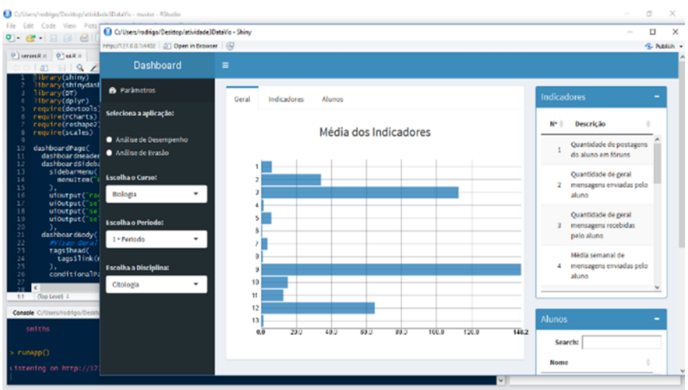

Fig. 8. *Data Overview* screen.

The second interface, *Performance Analysis*, offers predictions about the students' performance; it is composed by three tabs: Overview, Performance Indicators, and Students. Figure 9 shows the "Overview" tab. In this tab, the user can visualize predictions such as "percentage of students with satisfactory performance" and "percentage of students with unsatisfactory performance" for each one of the variables listed in Table 2, grouped by class, semester, and by the constructs defined in Section 2.4 (Environment Structuring, Help-seeking, Task completion strategies, Time management, Goal-setting, and Self-assessment). These predictions are calculated using the model described in Subsection 4.6. When the user hovers the mouse over each bar of the chart, a hint is displayed informing the description of the variable it represents and its minimum, average, and maximum value.

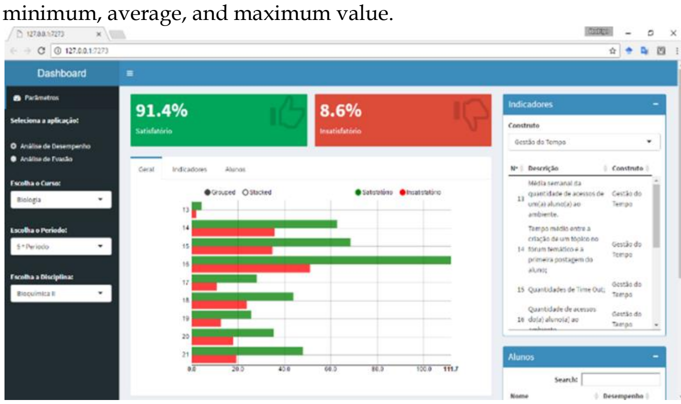

Fig. 9. *Performance Analysis* interface — "overview" tab.

Figure 10 shows how the satisfactory/unsatisfactory percentages are calculated. The classifier was built using the R language and, based on SRL behavioral variables, it predicts to which class a student would likely belong.

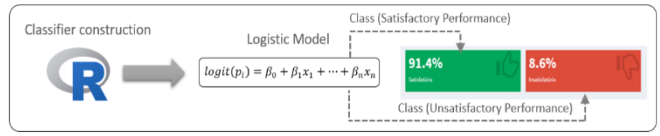

Fig. 10. Workfklow for calculating the satisfactory/unsatisfactory percentages.

Figure 11 shows the "Performance Indicators" tab from the Performance Analysis interface. In this tab, the user can visualize how each student is doing according to each one of the performance indicators, that is, to each variable identified in the logistic regression model as significant for self-regulatory behavior. This tab is responsible for displaying the frequency of all students in the class, selected according to information on a specific indicator. The information is shown by means of the bubble chart. The bubbles are represented in two colors, the red color represents the students who are in an unsatisfactory situation, and the green bubbles represent the students in a satisfactory situation. The diameter of the bubble represents the probability of the individual belonging to a specific class (Satisfactory or Unsatisfactory).

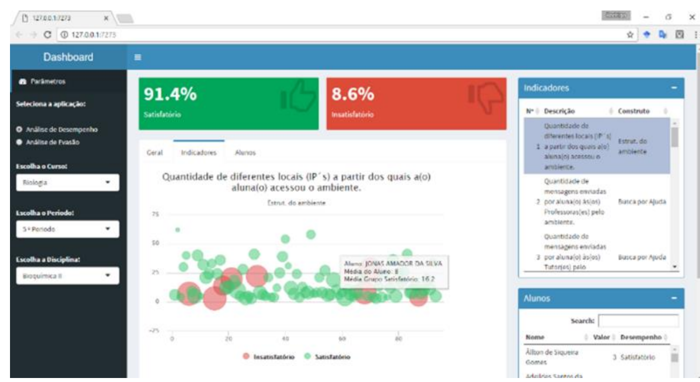

Fig. 11. *Performance Analysis* interface — "Performance Indicators" tab.

When the user hovers the mouse pointer over one of the bubbles, a hint is displayed describing: the student's name, the student's performance level in the selected indicator, and the minimum, average and maximum value in the class for the selected subject.

Finally, Figure 12 shows the "Students" tab from the *Performance Analysis* interface. When the user selects a student in the lower right corner of the screen, the system shows a graph depicting how far the student's performance is from what is considered "satisfactory" by our model in each performance indicator.

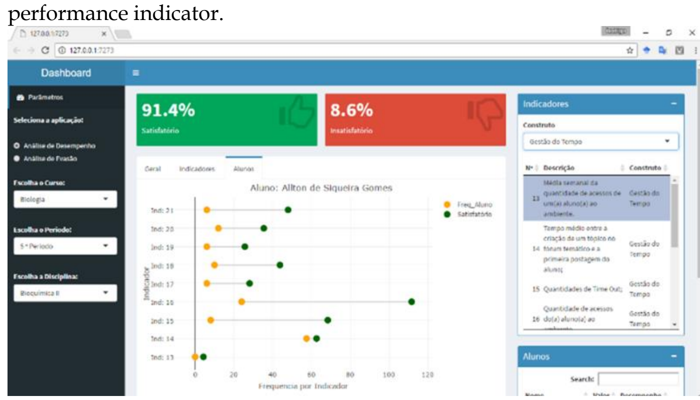

Fig. 12. *Performance Analysis* interface — "Students" tab.

In order to evaluate the user's expectation and experience regarding this prototype, we submitted it to two types of user evaluation: usability and user experience. In the next subsections, we present the results for these tests.

5.4. Usability evaluation

For this experiment, we selected a group of 20 teachers and tutors, using as selection criteria the level of knowledge in the Moodle LMS and years of experience as tutor or teacher in distance education courses (at least one year).

The first step of the test was showing to users the low-fidelity prototypes and its intended features. Subsequently, the users answered to a questionnaire related to their expectations after seeing the prototypes. After that, we presented and let the users interact with the high-fidelity (functional) prototype to the users and asked them to execute six pre-defined tasks. Finally, we applied a second questionnaire that sought to measure some characteristics of the user experience while performing the tasks in the prototype.

In order to analyze the usability of our high-fidelity prototype, we recorded the interaction of each user that participated in the test while performing the six proposed tasks and collected the following variables: number of errors, amount of help requests, task execution time, and number of clicks. In addition to these variables, users' opinions were collected regarding the features of the tool and suggestions for improvement. Figure 13 shows the test performed with teachers from Federal Rural University of Pernambuco.

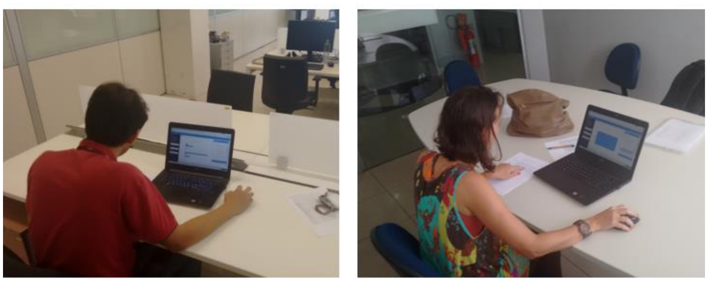

Fig. 13. Teachers performing the tasks using the high-fidelity prototype.

A group of twenty teachers from UPE, UFPE, UNIVASF, and UFRPE universities participated in the tests, with an average age of 41.88 years (SD = 8.84) and an average experience with distance learning courses of 3.97 years (SD = 2.62).

Each test had an average duration of 40 minutes. During the tests, we identified an average of 0.89 errors (SD = 0.79) per activity, an average time of 87.57 seconds (SD = 44.18) for completing each task, an average of 0.84 requests for help (SD = 0.71), and an average of 10.63 clicks (SD = 0.58).

Based on these metrics, we can affirm that the developed solution presented good results in relation to the usability satisfaction dimension. The difficulties encountered were similar among users, invariably affecting the same interface problems. We can consider that the proposed interface presented promising results, considering the positive reaction of users, especially their satisfaction with the ease of use of the tool developed.

As expected, some small issues were also detected, such as: unclear meaning of some labels in interface components and charts, the position of some captions, and how mouse clicks are handled in some components.

We reproduce below some of the feedbacks given by the users involved in the usability tests, which confirm that the proposed solution had a good acceptance and has potential for being deployed in production scale:

> *"The tool will allow individualizing the problem of each student. This allows a greater focus on the difficulty situation of each student. When contacting the student, the tutor has already a report of its specific difficulties, making it possible to give a focused assistance. This is quite different from the current tools that do not allow this type of diagnosis."*

> *"This tool will facilitate the visualization of Moodle data, which teachers cannot understand today."*

> *"We do not have this culture of taking action before things happen. This tool gives us this information, the teacher needs to create the culture of taking action based on this kind of information. This tool should also be used by people involved in the course management."*

5.5. User experience evaluation

To evaluate the user experience, we used the methodology proposed by Hassenzahl (2004), which includes the AttrakDiff instrument. AttrakDiff evaluates a product's usability using three dimensions: (1) Pragmatic Quality (PQ); (2) Hedonic Quality (Stimulation (HQ–S) and Identity (HQ–I)); and (3) Attractiveness (ATT). The PQ dimension refers to the interface's usability, that is, it evaluates if the users were able to reach the objectives of the tasks during the use of the product; the HQ-S dimension evaluates if the product has innovative characteristics, and presents stimulating interactions; The HQ-I dimension evaluates the identification of the user with the interface; finally, the ATT dimension describes the overall value of the product based on the quality of the user's perception (Hassenzahl, 2004).

We used the AttrakDiff questionnaire in two moments during our experiment: after showing to the user the low-fidelity prototypes and after the user interacted with the high-fidelity prototype. The goal was to evaluate if there would be significant differences between the users' expectations and experience regarding the proposed solution.

Figure 14 presents the results of the comparison between the users' expectations and experience. We can observe that our prototype was considered as "desired" both regarding the users' expectations (the orange rectangle) and experience (the blue rectangle). Thus, its pragmatic and hedonic qualities are clearly highlighted in this quadrant, indicating that the product assists, sparks interest, and stimulates users.

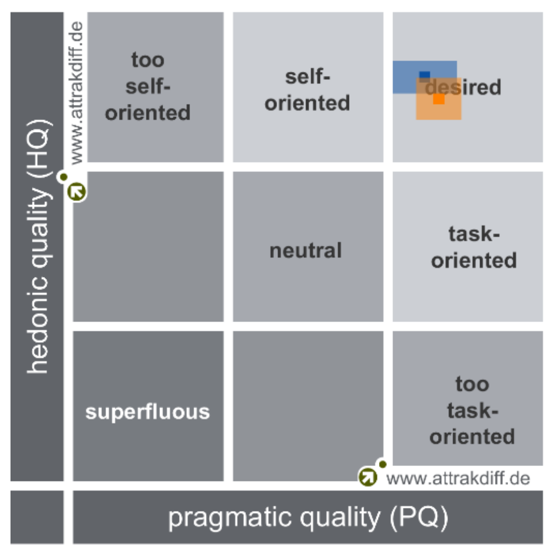

Fig. 14. Average values (by dimension) and confidence rectangles.

Figure 15 shows the average values by dimensions according to the AttrackDiff analysis. Only in the Pragmatic Quality (PQ) dimension the expectation was measured as greater than the user experience. In the other dimensions — Hedonic Quality of Stimulus (HQ-S), Identity (HQ-I), and Attractiveness (ATT) — the user experience always got higher averages.

One important aspect in this analysis is that all points (averages) are "above-average" region (above zero), which means that the proposed interface met both the users' expectation and experience, according to Hassenzahl's (2004) methodology.

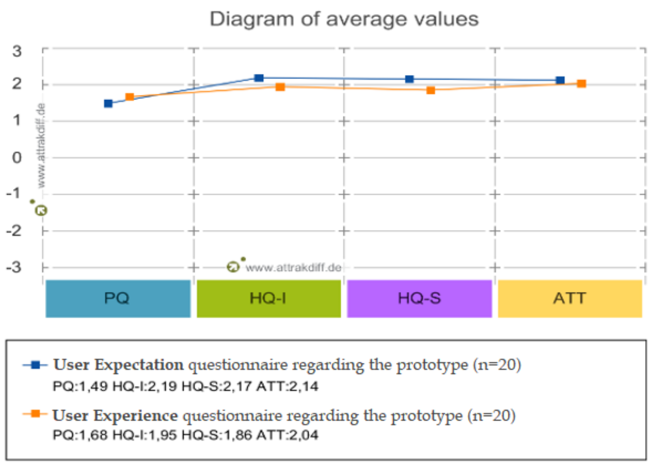

Fig. 15. Average values for each dimension.

In order to have a more detailed view of the experiments' results for each item in the four categories of AttrackDiff's questionnaire, we compared the average values of each item for the user's expectations (first questionnaire) and experience (second questionnaire). Figure 16 shows this comparison. According to this figure, in the Pragmatic Quality category (PQ), four items showed a higher average on "expectation" than "experience", and three the opposite. Regarding the Hedonic Quality of Identity (HQ-I), six items showed a higher average on "expectation" than "experience". In the Hedonic Quality of Stimulus category (HQ-S), all items showed a higher average on "expectation". In the Attractiveness category (ATT), three items showed a higher average on "experience", three on "expectation", and one item had equal averages. However, it is important to mention that all marks on the graph (average values) are situated in the right side, that is, above the mean value in AtrakDiff's scale (zero).

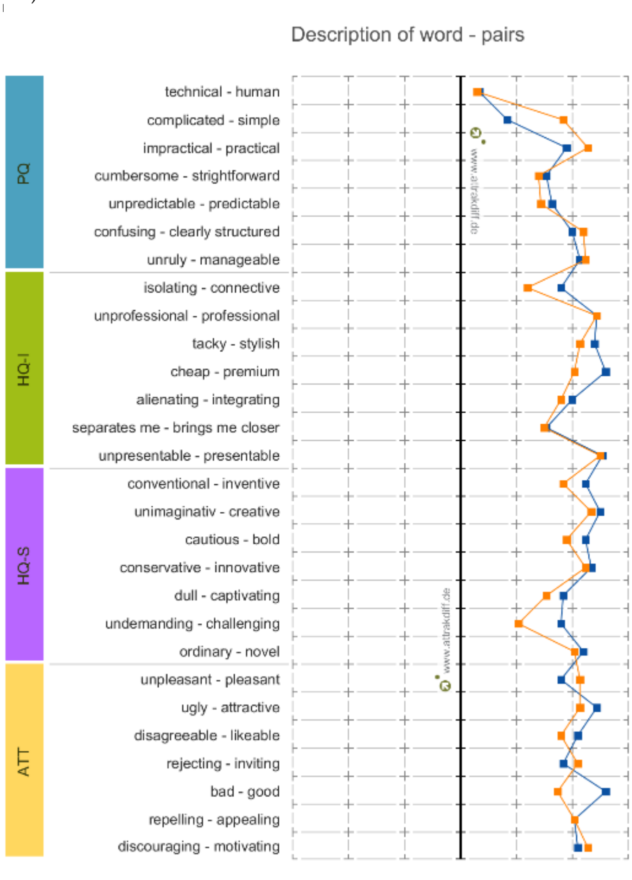

Fig. 16. Comparison between the average results for *experience* and *expectations* in users' responses for each item of the questionnaire.

Although Figure 16 shows some differences between experience and expectations on AttrackDiff's results, visual representations are not enough to draw conclusions in this aspect. To test statistically the hypothesis that there are differences between experience and expectations, we used *Wilcoxon*'s signed-rank test (Corder & Foreman, 2014) with a 5% significance level. Each item of the first sample (the first questionnaire that measured the *expectations* based on a low-fidelity prototype) was paired with the item from the second sample (the second questionnaire that measured the *experience* based on the high-fidelity prototype). The null hypothesis H_0 was that there was no significant difference between the user's expectation and experience, and the alternative hypothesis H_1 was that there was significant difference in this aspect.

The test's results showed a p-value greater than 0.05 for all variable pairs, so we failed to reject the null hypothesis. According to theis result, we can state with a 95% confidence level that there was no significant difference between the user's expectation and experience regarding the proposed tool.

Considering the graph showed in Figure 16 and the results of the hypothesis test, we can conclude that the proposed dashboard was considered satisfactory by the users and has potential to be incorporated in the teaching practice of professors and tutors.

6. Discussion and implications

This section discusses the results achieved in this research and its implications. We address the findings that have enriched the results in relation to the choice of behavioral variables, model development and proposed software solution.

6.1. *Choice of the behavioral self-regulation variables*

According to Rates (2013), almost a third of students drop out of college after their first year. Academic performance is considered one of the most influential factors in this context (ESSA & AYAD, 2012). One strategy to decrease this retention rate is to identify in advance which factors influence the students' performance, allowing teachers to use a variety of strategies to communicate with students at risk and help them overcome these problems.

The first step to increase academic success is identifying students at risk of showing poor performance. With the use of predictive modeling methods, it is possible to identify these students and their behaviors. Predictive analysis allows teachers, tutors, and managers to take preventive measures, since they can identify the actions of potential students in critical situation.

In order to build the predictive models, we used behavioral data extracted from the Moodle platform. The choice of which variables should be chosen was such that it allowed to measure the of self-regulated learning process, defined by theorists as being the process in which individuals develop self-management abilities of their behavioral actions to perform tasks (Veiga Simão, 2006); (Zimmerman B. J., 1989); (Silva, 2004); (Zeidner, Boekaerts, & Pintrch, 2000).

These variables were used as behavioral indicators capable of predicting the academic performance of distance learning students. For the construction of the predictive models we used a training subset of 21,154 records, which corresponded to 70% of the total dataset. This training base was composed of 11,824 students with poor performance levels and 9,330 students in a satisfactory performance situation.

6.2. *Impact of the chosen variables on the performance prediction*

According to James, Witten, et al. (2013) the choice of which variables to use in a predictive model is one of the most important factors to build a good classifier. Usually, the strategies for making this choice try to select the variables that have the greatest predictive power and are close related to the predicted variable.

Initially, the logistic regression model was composed of 32 explanatory variables. After applying the Stepwise method, it was possible to maintain the same level of accuracy with only 23 explanatory variables. These 23 variables were distributed in six self-regulated learning constructs: (ES) Environment Structuring; (HS) Help-seeking;

(TS) Task completion strategies; (TM) Time management; (GS) Goal-setting; and (SA) Self-assessment.

Among the 23 variables selected to build the final model, some had a greater power of influence in the prediction of the students' performance. Variable 33, which was part of the "Goal-setting" construct, was the model's most statistically significant variable. This variable represents the number of assignments delivered on time per student. This behavior was the most significant of all, that is, there is a relationship between how often students hand out their assignments on time and their performance.

The second most significant variable was *var34*, which also belongs to the "Goal-setting" construct. This variable measures the total number of messages posted by a student on forums, and had great influence in the classifier's prediction power.

Finally, *var25* was the third most significant. It is part of the "Time management" construct and measures the average time between the moment a topic is created in the forum and the moment the student posts his/her first response to this topic, that is, how much time passed until the student started to participate in the forum.

Although the three variables presented above were the most significant ones, all the other twenty ones are also important for the complete predictive model, since the its predictive power comes from the composition of all its variables.

6.3. *Performance of the developed model*

The developed model was able to correctly classify 8,093 out of the 9,063 students in the dataset, considering both classes (satisfactory and unsatisfactory performance), which gives an error rate of only 10.7%. Of the 5,040 students who belonged to the "satisfactory performance" class, the model was able to classify 4,640 of them (91.66%) correctly. Of the 4,023 students who belonged to the "unsatisfactory performance" class, the model was able to classify 3,473 of them (86.32%) correctly.

In the context of this work, we consider the "false positive" error — to predict that a student will show satisfactory performance when in fact he/she will show unsatisfactory performance — the most critical case of misclassification. In our experiments, the developed model committed this type of error in 550 cases, which represents only 6.07% of the test data. We consider this error rate quite acceptable, especially because one of the goals of this research was to develop a less specialized predictive model, which could be used on different distance learning courses — instead of a specialized model for a specific course.

6.4. *Development of a software solution*

The proposed software solution was considered innovative by its ability to integrate data visualization and predictive models. This arrangement empowers teachers and tutors, giving them the possibility of monitoring predictions about their students' academic performance, based on self-regulated learning behavioral variables.

In order to build the monitoring dashboards, we used the *Shiny* framework. This choice was motivated by the possibility to use the programming language — the R language — to implement both the predictive model and the interfaces. This strategy contributed to improve the overall learning curve.

Among all the requirements identified during the conception of the proposed software solution, the ease of use was one of the most important ones. Therefore, after the development phase we conducted a series of usability tests in order to identify problems in the interfaces.

Regarding the evaluation of the user's expectation and experience, the results were excellent in both aspects. The product was considered as "desired" by the users in two evaluations — after seeing the low-fidelity prototypes and after interacting with the high-fidelity prototype. Besides that, the tests also indicated that the product assists, sparks interest, and stimulates the users.

7. Conclusion

This paper aimed to identify how the students' self-regulated learning behavior influences their academic performance in a distance learning context.

To accomplish this goal, we used data mining techniques to build a predictive model capable of predicting a student's academic performance based on his/her self-regulatory behavior. The developed model yielded satisfactory results in experiments with real data from 9,063 students.

The contributions of this research are twofold: we present and validate a model for predicting student's performance and also propose a dashboard that teachers and tutors can use to monitor the students' self-regulatory behavior, improve their pedagogical decisions, and help mitigate student dropout and failure.

Taking into account that this work proposed a software solution for distance learning contexts, it is necessary to validate it in real scenarios, that is, in the daily practice of tutors and professors. Besides that, other researchers can use this work as a starting point for testing the applicability of the solution in real classes and verify its robustness in long-term experiments and in other educational institutions, which can generate new requirements for the proposed dashboard.

References

Adesope, O. O., Zhou, M., & Nesbit, J. C. (2015). Achievement Goal Orientations and Self-Reported Study Strategies as Predictors of Online Studying Activities. *Journal of Educational Computing Research*, 436-458.

Agustiani, H., Cahyad, S., & Musa, M. (2016). Self-efficacy and Self-Regulated Learning as Predictors of Students Academic Performance. *The Open Psychology Journal*.

Ali, L., Hatala, M., Gašević, D., & Winne, P. H. (2014). Leveraging MSLQ Data for Predicting Students Achievement Goal Orientations. *Journal of Learning Analytics*, 157-160.

Baker, R. S., Carvalho, A. M., & Isotani, S. (2011, Agosto). Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*.

Barnard, L., Lan, W. Y., To, Y. M., & Paton, V. O. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education*, 1-6.

BONDAREVA, D. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. *In Artificial Intelligence in Education, Springer Berlin Heidelberg*, pp. 229-238.

Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. *In Artificial Intelligence in Education* , 229-238.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*(27), 1-13.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, pp. 1-13.

Cambruzzi, W. L, de Moraes, R, Leithardt, V. R, Mendes, C, Geyer, C. F, da Costa, C. A, & & Rigo, S. J. (2012). Um Modelo para Gerenciamento de Multiplas Trilhas Aplicado a Sistemas de Apoio a Educação. *Simpósio Brasileiro de Informática na Educação.*

Campos, H. (1987). Estatística experimental não-paramétrica. *Esalq. Piracicaba-SP*, 230p.

Carson, A. D. (2011). Predicting student success from the LASSI for learning online (LLO). *Journal of Educational Computing Research*, pp. 399–414.

Chang, M. M. (2010). Effects of self-monitoring on web-based language learner's perfor- mance and motivation. *CALICO Journal*, pp. 298–310.

ChanLin, L. J. (2012). Learning strategies in web-supported collaborative project. *Innovations in Education and Teaching International*, pp. 319–331.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Cho, M. H., & Jonassen, D. (2009). Development of the human interaction

dimension of the Self-Regulated Learning Questionnaire in asynchronous online learning environments. *Educational Psychology*, 117-138.
Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education*, pp. 290–301.
COLTHORPE, K. e. (2015). Know thy student! Combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated learning. *Journal of Learning Analytics*.
Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons.
da Silva, L. A., Peres, S. M., & Boscarioli, C. (2016). *Introdução à Mineração de Dados com Aplicações em R* (1 ed.). Rio de Janeiro: ELSEVIER.
Essa, A., & Ayad, H. (2012). Student success system: Risk analytics and data visualization using ensembles of predictive models. *In The 2nd International conference on learning analytics and knowledge*.
FAVEIRO, L., BELFIORE, P., SILVA, F., & CHAM, B. (2009). Análise de dados: modelagem multivariada para tomada de decisão. *São Paulo: Campus*.
Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 861-874.
Gonçalves, I. C. (2010). Contributos dos Modelos da Auto-Regulação da Aprendizagem para a formação de Alunos e Professores no Ensino Superior.
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados*. Bookman Editora.
Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction*, 319-349.
Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*.
Klingsieck, K. B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance univer- sity setting. *Distance Education*, pp. 295–310.
Koedinger, K. R., Baker, R. S., Cunningham, K. S., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. *Handbook of educational data mining*, 43.
LAWANTO, O. (2014). Self-Regulated Learning Skills and Online Activities between Higher and Lower Performers on a Web-Intensive Undergraduate Engineering Course. *Journal of Educators Online*.
McGaw, B., Peterson, P., & Baker, E. (2010). *International Encyclopedia of Education*. Amsterdam: Elsevier.
Nussbaumer, A., Hillemann, E. C., Gütl, C., & Albert, D. (2015). A competence-based service for supporting self-regulated learning in virtual environments. *Journal of Learning Analytics*, 101-133.
Park, Y., & Lim, K. (2015). Effects of Environmental and Human Constructs on E-Learning Effectiveness in Online University Settings. *Indian Journal of Science and Technology*, 103-109.
PINTRICH, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459-470.
Pintrich, P. R., Smith, D. A., García, R., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement*(53(3)), 801-813.
Rates, F. T. (2013). National Collegiate Retention and Persistence to Degree Rates.
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance. A systematic review and meta-analysis. *Psychological Bulletin*, pp. 353–387.
Rigo, S. J., Cambruzzi, W., Barbosa, J. L., & Cazella, S. C. (2014, março 8). Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na EducaçãoRevista Brasileira de Informática na Educação*.
Romero, C., & Ventura, S. (2013, February). Data mining in education. *WIREs Data Mining Knowl Discov*, pp. 12-27.
Sabourin, J. L., Mott, B. W., & Lester, J. C. (2012). Early Prediction of Student Self-Regulation Strategies by Combining Multiple Models. *International Educational Data Mining Society*.
SABOURIN, J. L., MOTT, B. W., & LESTER, J. C. (2012). Early Prediction of Student Self-Regulation Strategies by Combining Multiple Models. *International Educational Data Mining Society*.
Sabourin, J., Shores, L. R., Mott, B. W., & Lester, J. C. (2012). Predicting student self-regulation strategies in game-based learning environments. *International Conference on Intelligent Tutoring Systems*, 141-150.
Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, 1321-1331.
SCHOOR, C., & BANNERT, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, pp. 1321-1331.
Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. 13-48.
Sheth, J., & Patel, B. (2010). Best practices for adaptation of Data mining techniques in Education Sector. *National Journal of System and Information Technology*, 186.
Silva, A. L. (2004). A auto-regulação na aprendizagem: a demarcação de um campo de estudos e de intervenção. *Aprendizagem auto-regulada pelo estudante: perspectivas psicológicas e educacionais*, 17-39.
Sonnenberg, C., & Bannert, M. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, 72-100.
SONNENBERG, C., & BANNERT, M. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, pp. 72-100.
Stone, M. (1974). Cross-validatory choise and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 111-148.
Veiga Simão, A. M. (2006). Auto-regulação da aprendizagem: um desafio para a formação de professores. *Formação de professores de línguas estrangeiras: reflexões, estudos e experiências*, 192-206.
Wang, C. H., Shannon, D., & Ross, M. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, pp. 302–323.
Wang, J., Doll, W. J., Deng, X., Park, K., & Yang, M. G. (2013). The impact of faculty perceived reconfigurability of learning management systems on effective teaching practices. *Computers & Education, 61*, 146-157.
Weinstein, C. E., Schulte, A., & Palmer, D. R. (1987). The Learning and Study Strategies Inventory. *Clearwater, FL: H & H Publishing*.
You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, 23-30.
Yukselturk, E., & Bulut, S. (2009). Gender differences in self-regulated online learning environment. *Journal of Educational Technology & Society*, 12-22.
Zeidner, M., Boekaerts, P., & Pintrch, P. (2000). Handbook of self-Regulation. *New York: Academic Press*, 13-39.
Zimmerman, B. J. (1989). A social cognitiveview of self-regulated learning. *J. Educ. Psychol*, 329-339.
Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal, 23(4)*, 614-628.