International Journal of Learning, Teaching and Educational Research
Vol. 17, No. 9, pp. 91-101, September 2018
https://doi.org/10.26803/ijlter.17.9.6

The Use of Business Intelligence Tools to Analyze the Influence of Interactivity and Interaction Factors on the Assessment of Distance Students' Performance in Virtual Learning Environments

Ana Ichihara
Faculdade de Tecnologia do Estado de São Paulo
[State of São Paulo Technology University]
São Paulo, Brazil

Omar Nizam
Universidade Presbiteriana Mackenzie
[Presbyterian University Mackenzie]
São Paulo, Brazil

Abstract. This paper aims to improve the practice of distance education, by providing managers with a view of aspects that influence the progression of students. To that end, it analyses "Interactivity and Interaction" factors in Virtual Learning Environments (VLE) communication systems, seeking to understand how these elements influence the performance of distance learning students at the beginner level. The study was carried out using data from a Brazilian distance learning private university, which utilizes a virtual learning environment. The research involved four steps: construction of a business intelligence environment, statistical analytical work, decision trees and clustering techniques to describe data, establish the most relevant variables and identify standards that may support the conclusion.

Keywords: Distance Education; student evaluation; statistics; decision trees; clustering.

1. Introduction

Given its diversity, geographic dimensions and socioeconomic conditions, Brazil presents favorable conditions for the expansion and development of Distance Learning. The same scenario, however, imposes on the managers of educational institutions the challenge of understanding, managing and expanding the distance learning method.

> *"There is still much to expand and experiment, in terms of program offerings, types of courses, methodologies, as well as technological and administrative innovation." (Associação Brasileira de Educação a Distância – ABED [Brazilian Association of Distance Education], 2015)*

Thus, this study seeks to improve the practice of distance learning teaching. The goal is to provide educational managers with an overview of the factors that influence the performance of beginner-level students, aspects which must also be in line with the governmental standard of quality, expressed in "the Critical Success Factors of the Communication System" (Turrioni & Stano, 2016) identified in the Quality Referential in Distance Higher Education of the MEC/SEED (2007).

To this end, the study uses operational data from a Virtual Learning Environment platform of a Private Superior Education Institution, preserving confidentiality regarding the identification of the entity and the students. The data refer to the beginner-level students of four courses. The research applied techniques of Business Intelligence, based on Ichihara and Nizam (2017) studies, as well as a decision tree and clustering approach to achieve the proposed goal.

2. Theoretical References

A survey of the existing publications on the subject identified some works. Among them are the Business Intelligence techniques that can automate the institutional management systems with the purpose of extracting, transforming, analyzing and mining data, facilitating them as evaluation tools (Pascal, Servetto, Mirasson, & Luna, 2017). The BI architecture, as proposed by Pifarré (2015), consists of several components that support the generation of managerial data, such as processes of Extraction, Transformation and Load (ETL), the Dimensional Data Model, and analytical tools of the most varied types of analysis.

In the studies of Ichihara and Nizam (2017), the dimensional data models' metrics were elaborated from the Critical Success Factors of the Communication System, mapped according to the guidelines of the Quality Referential in Distance Higher Education of the MEC/SEED (2007). Other components of this environment are the exploratory data tools such as decision trees and clustering techniques, used in this work to show patterns and variables relevance to the performance of students starting a DL course.

The decision tree extracts knowledge from the dataset and represents it in the shape of a tree and is widely used in distance learning applications. Zacharis (2017) develops a work that adopts a predictive model using the CART algorithm to predict student performance, with data from Moodle related to message exchange, wiki group content creation, interactive quiz, and file opening. Cruz, Duarte and Goldschmidt (2017) use several machine learning algorithms to construct a recognition model adopted in the authentication of VLE users, among which are decision trees.

The learning of a decision tree is supervised and employs an induction process that uses training cases with input and output attributes, classes with labels and a learning algorithm, in the hypothesis space, that seeks a hypothesis that best fits the training data, to generate a classification model that maps independent attributes to the respective class and can be used in descriptive modeling when explaining why a case belongs to a class; or in predictive modeling, when used to classify unlabeled examples (Tan, Steinbach, & Kumar, 2005).

This work uses the decision tree to distinguish variables of relevance to the approval of a student, identified by the classification algorithm, in the face of the production of the more homogeneous results, sorted graphically in the shape of trees and rules, so as to distinguish the different observational classes.

According to Cheng (2017), the clustering task is a widely used method in educational data mining and focuses on good student performance. It is a technique of data analysis and exploration that can be employed to extract intrinsic patterns in a dataset, from the application of algorithms that form groups, which maximizes the similarity of its elements and minimizes similarities between the groups. Corsatea and Walker (2015) use clustering techniques, with a clustering k-means algorithm to compare the student's performance with the use of Moodle. Ramos, Almeida and Nóbrega (2016) compared two clustering methods: hierarchical and non-hierarchical, using Moodle data from student interactions. Preidys and Sakalauskas (2010) use logs of the VLE Blackboard Vista Enterprise platform and the clustering task to analyze the student's style to plan individualized materials and adequate course methods.

In the clustering process, this work uses the relevance variables identified and obtained by the decision tree to establish similar groups of students of each course, and describes the groups of each course, with the best performance.

3. Experiment

3.1. Study Data

This study is carried out using the data provided by a Brazilian private higher education institution, whose courses are provided in the Distance Learning mode, through the Blackboard Virtual Learning Environment platform. According to the criteria and policies established by the university, this work uses proper nomenclature to name the study variables. The data refer to the beginner-level students of four courses, TEH, MAT, BIO and SES, related to the Business, Science, Biology and Humanities areas. The provision of five disciplines characterizes each course, in a sequence (curriculum) 2, 3, 4, 5 and 6, during the semester, lasting one month each. The instructional design of the courses is uniform, with distance and classroom evaluation, whose grades will underpin the final grade.

3.2. Methods

The work was carried out in four steps: 1) construction of a business intelligence environment according to the authors' proposal, defined in studies by Ichihara and Nizam (2017); 2) descriptive statistics of the study variables; 3) analysis of the relevance of the same variables using the decision tree, to identify those that most influence the good performance of the student; and 4) finding similarities using clustering with variables pointed out by decision trees, with characterization of only the best-performing groups, for each course.

4. Results and Discussion

4.1. ETL Process and BI Dimensional Model

The same methodology of Ichihara and Nizam (2017) was used for the development of the processes of Extraction, Transformation and Load (ETL) and construction of the Dimensional Model as per Kimball and Ross (2002), permeating all phases of the proposed methodology: identification and mapping of critical success factors (CSF) of the VLE Blackboard communication system, as per Quality Referential in Distance Higher Education of the SEED/MEC (2007), identification of metrics and tables of the operational database and construction of the dimensional model and processes of ETL.

In this study, this model has of interactivity and interaction elements described by the following study variables: AMTU, number of access to the "Click here to send a message to the tutor" option; AFPA, number of responses sent by the teacher to the student in the forums; AFAP, number of responses sent by the student to the teacher in the forums; AFOR, number of access to forums; ADOC, number of access to documents; AAVI, number of access to notices; AASS, amount of access to tasks and evaluations; and elements of performance (MEFI - student's final grade in the course); characterization of the student (IDADE and REGIÃO, namely, age and region); CURSO [course]; DISCIPLINA [discipline] of the student; and SEQ, discipline sequence in the semester.

4.2. Cleaning and outliers

In the cleaning procedure, cases with access values and final grades equal to zero, and with inconsistent access values and age were eliminated, resulting in the set of examples for each course: MAT: 286; TEH: 1330; SES: 615; and BIO: 127.

4.3. Statistical analysis

In Table 1, position measurements minimum, first quartile, median, mean, third quartile, and maximum show the summary of the variables of the study through quantitative position and dispersion measures to characterize them regarding their variability and distribution. We can assume from the data analysis that, in general, for all courses, the distributions of access variables are positive asymmetries (medians in the midDistance Learninge of the boxplot boxes), indicating that the frequency of observations (students attending a discipline) decreases with the higher number of access. The AASS (access to tasks) variable has the most symmetric dataset (median in the midDistance Learninge of the rectangle).

The volume variability of access follows a pattern of use by type in all courses: the highest access volumes are found in ADOC, ASSS and AAVI; the communications established by the AMTU-type access also follow a pattern, and are more used than those established in the AFPA and AFAP forums, found in all the courses, and are less significant when compared to other types of access;

the final grades (final mean) vary between 0 and 10, with 75% (third quartile) of the data varying between 0 and 5.5, evidencing a significant number of cases of fails; the worst performance is of the BIO course, with 75% of data below 4.0; despite the significant range in the different types of access (ADOC, AFOR, AASS and AAVI), with the exception of the BIO course, 50% of access, box height (defined by the interquartile range) concentrate volumes ranging from 0 to 100. Fifty percent of the dataset between the ages of 25 and 40 represent the MAT, SES and BIO courses. In the TEH course, 50% of the observations record ages between 25 and 35 years; the strongest association, obtained by Spearman's coefficient, between the student's grade (MEFI) and the access variables, occurs with the variable AASS, with values ranging from 0.70 to 0.76 in the courses. However, the association between IDADE [age] and grade (MEFI) is weak but negative; the final grade declines with age. All courses have high fail rates concerning study observations: MAT records 61 cases (21.86%); SES 131 cases, (21.30%); TEH 281 cases, (21.12%) and BIO 22 cases (17.32%).

Table 1: Descriptive statistics of the study variables

Course	Variable	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Standard Deviation
MAT	AFAP	0.0000	0.0000	0.0000	0.3741	0.0000	6.0000	0.9000
	AFPA	0.0000	0.0000	0.0000	0.5245	0.0000	0.0000	0.2900
	AMTU	0.0000	0.0000	0.0000	0.9790	1.0000	21.0000	2.6400
	AFOR	0.0000	2.0000	8.0000	21.4700	30.7500	161.0000	29.7600
	ADOC	0.0000	10.0000	40.5000	62.0500	92.7500	393.0000	66.7700
	AAVI	0.0000	3.0000	13.0000	17.8300	26.0000	109.0000	18.9900
	AASS	0.0000	0.0000	33.5000	41.2000	68.0000	213.0000	42.3700
	MEFI	0.0000	0.0000	2.2500	2.8760	5.0000	10.0000	2.9400
	IDADE	18.0000	24.0000	31.0000	32.0800	38.0000	61.0000	9.1900
TEH	AFAP	0.0000	0.0000	0.0000	0.2098	0.0000	8.0000	0.7700
	AFPA	0.0000	0.0000	0.0000	0.3459	0.0000	3.0000	0.2300
	AMTU	0.0000	0.0000	0.0000	0.9346	0.0000	74.0000	3.8300
	AFOR	0.0000	2.0000	12.0000	24.0700	33.0000	262.0000	32.2100
	ADOC	0.0000	13.0000	41.0000	58.9800	83.0000	587.0000	65.1400
	AAVI	0.0000	3.0000	10.0000	15.3700	22.0000	135.0000	17.2800
	AASS	0.0000	0.0000	35.0000	38.8300	63.0000	228.0000	37.5700
	MEFI	0.0000	0.0000	3.0000	3.1390	5.0000	10.0000	2.7900
	IDADE	17.0000	25.0000	30.0000	31.1500	36.0000	65.0000	8.3800
SES	AFAP	0.0000	0.0000	0.0000	0.1951	0.0000	7.0000	0.7200
	AFPA	0.0000	0.0000	0.0000	0.2114	0.0000	2.0000	0.1500
	AMTU	0.0000	0.0000	0.0000	2.8590	2.0000	76.0000	7.5200
	AFOR	0.0000	1.0000	6.0000	17.1900	23.5000	309.0000	27.0700
	ADOC	0.0000	6.0000	39.0000	57.9500	86.0000	510.0000	63.6600
	AAVI	0.0000	2.0000	9.0000	15.5700	22.0000	127.0000	18.9100
	AASS	0.0000	0.0000	35.0000	37.2400	63.0000	191.0000	36.9900
	MEFI	0.0000	0.0000	3.2120	3.1390	5.0000	10.0000	2.9000
	IDADE	18.0000	29.0000	35.0000	35.5400	41.0000	61.0000	9.1000
BIO	AFAP	0.0000	0.0000	0.0000	0.1575	0.0000	6.0000	0.7100
	AFPA	0.0000	0.0000	0.0000	0.3937	0.0000	1.0000	0.2000
	AMTU	0.0000	0.0000	0.0000	1.8270	1.0000	22.0000	4.0000
	AFOR	0.0000	1.0000	6.0000	17.1900	25.0000	107.0000	23.9700
	ADOC	0.0000	8.5000	25.0000	43.1500	63.5000	266.0000	48.2400
	AAVI	0.0000	5.0000	15.0000	20.0000	26.5000	100.0000	20.4600
	AASS	0.0000	0.0000	28.0000	33.6200	57.5000	142.0000	33.7100
	MEFI	0.0000	0.5000	3.0000	2.9020	4.0000	10.0000	2.6000
	IDADE	18.0000	22.0000	31.0000	30.5200	37.5000	52.0000	9.0600

4.4 Variables of relevance using decision trees

The Machine Learning in R (RPR) RPART (Recursive Partitioning and Regression Trees) package was used to build decision trees. It provides a unified interface for machine learning tasks, such as classification, clustering analysis (Therneau & Terry, 2018) and regression - and implements a modified version of the CART (Recursive Partitioning and Regression Trees) algorithm used in this work.

The dataset of the MAT, SES, BIO and TEH courses is the same already detailed in this work, with seven predictive variables: AFPA, AFAP, AASS, AMTU, AAFOR, AAVI, ADOC and IDADE [age], and a dependent variable, MEFI. The variable SUCESSO [success] was established from this variable, with the following criterion: if MEFI < 6, SUCESSO [success] assumes value 'N'; if MEFI is ≥ 6, SUCESSO [success] assumes value 'S'. The variable SUCESSO [success] is used in the model to predict student approval, 'S' if approved, and 'N' if failed.

Hyperparameters were optimized by cross-validation (Witten & Frank, 2017) with five folds during the construction of the models. Tree size-related parameters, the minimum number of observations to divide, the minimum number of observations in the terminal node and the minimum improved error value for the tree to continue the divisions were adjusted.

The optimal parameters were found by determining the highest accuracy value for the five folds. The test sets consisted of 20% of the total observations of each course. The mean classification accuracy for each tree is 0.7951 for MAT, 0.8081 for TEH, 0.8031 for SES and 0.7729 for BIO. The models are more than 77% accurate and can correctly predict more fail cases. The relevance of predicting each study variable in descending order is achieved by the mean of the relevance obtained in each fold, shown in Figure 1, for each course.

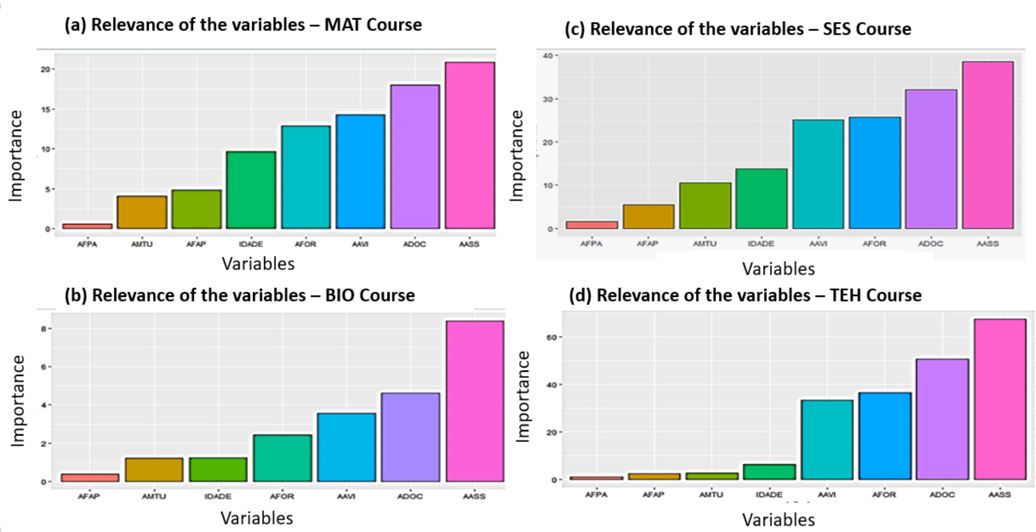

Figure 1: Mean of the relevance obtained in each fold

Concluding from the generated trees, eight predictive attributes were organized by order of decreasing relevance of prediction for modelling. When analyzing the data, we can observe that, in all courses, the number of access to the tasks

(AASS) is always the most important attribute, followed by the number of access to the documents. The third most important attribute is access to the notices (AASS) for the MAT and BIO courses or access to the forums (AFOR) for the TEH and SES courses. The age attribute is less relevant in the TEH course.

4.5. Clustering analyses

The software programs used were the MLR and Cvalid, to obtain internal, external and validation rates (Peres & Lima, 2015). To choose the number of groups (K), its initial values and the clustering algorithm, experiments were made with k varying in the interval 2 and 15, with criteria in Elbow (Leskovec, Rajaraman, & Ullman, 2010), and with the algorithms: hierarchical clustering, Pam and k-means. From the results, we chose k-means with k = 4.

Four models were trained, one for each course, with the respective set of observations of the study, and the variables AASS, ADOC, AFOR, AAVI and MEFI, identified by decision trees. A measure of the quality of clusters obtained by k-means is the relationship (BetweenSS/TotalSS), where WithinSS is the sum of the squares of Euclidean distances between each observation of a group and its centroid and measures the total variation within the cluster (internal cohesion) and should be minimized. The sum of WithinSS of all groups measures the clustering compactness and should be minimized; BetweenSS is the sum of squares of Euclidean distances between clusters and measures the separability between clusters; TotalSS totals the two values: (WithinSS + BetweenSS). Ideally, the ratio (BetweenSS/TotalSS) should be close to 100%. For each trained model, this quality rate is 65.9%, for the MAT course, 62.8% for the SES course; 65.00% for the TEH course; and 78.6% for the BIO course – suggesting groups with moderate intragroup compactness and good intergroup separability.

The Tables 2, 3, 4, 5 and 6 describes the groups of each course, with the best performance.

Table 2: Groups' characterization

Course	Group	Number of Observations of the Group	Number of Observations of the Study / Number of Observations of the Study (%)	Number of observations of the group with student approved	Number of observations of the group with student approved / Total observations of the study with student approved (%)	Number of observations of the group, with student approved / Total observations of the group (%)	Number of students of the group
MAT	2	73	25.5	56	91.8	76.71	33
SES	3	233	37.88	101	77.09	47.63	127
TEH	3	429	32.25	184	69.03	45.22	215
BIO	4	38	71.05	11	50	40.74	26

Table 3: Group characterization, considering discipline and sequence

Course	Do all the subjects in the semester participate in the group?	Number of the sequence-subject that gathers more students	Number of observations of the sequence that gathers more students / Total observations of the group (%)	Number of sequence with more observations of approved students	Did the performance improve throughout the semester?	Number of approved students in the sequence with more of these observations / Total sequence observations
MAT	S	4.5.6	75.34	2 e 5	N	88.88 and 82.35
SES	S	4.5.6	72.1	6	S	66.66
TEH	S	4.5.6	75.05	6	S	72.52
BIO	S	3.4	76.31	6	S	75.5

Table 4: Characterization of students

Course	Majority region, in the group	Number of observations of the group with majority region / Total observations of the group (%)	Majority age groups	Number of observations of the priority age group / Total observations of the group (%)	Number of observations of the group with cancellation or transfer or leave of absence	Number of observations of the group with cancellation or transfer or leave of absence / Total observations of the group (%)
MAT	Southeast	31.5	20-40	80.82	13	28.8
SES	Southeast	47.63	30-40	42.06	51	21.88
TEH	Southeast	50.34	20-30	48.71%	132	30.76
BIO	Southeast	52.63	20-40	47.36%	30	78.94

Table 5: Characterization of students, group quality and important variables

Access preference considering the number of accesses by the student	Number of Interactions	Cohesion and separability BetweenSS / TotalSS	Important variables of the group
AASS, ADOC, AFOR, AAVI	1 to 10	0.52	AASS, ADOC and AAVI
AASS, ADOC, AFOR, AAVI	1 to 10	0.33	ADOC, AASS, AAVI and AFOR
AASS, ADOC, AFOR, AAVI	1 to 10	0.39	ADOC, AASS, AAVI and AFOR
AASS, ADOC, AFOR, AAVI	1 to 10	0.33	AAVI, ADOC, AFOR and AASS

Figure 6: Characterization of access

Course	Type of access	Interquartile Range	Students of the group with at least one access to the resource (%)
MAT	ADOC	32 to 81	100
	AFOR	3 to 42	90.41
	AAVI	8 to 23	95.89
	AASS	46 to 87	100
SES	ADOC	34 to 78	100
	AFOR	3 to 24	87.55
	AAVI	8 to 22	98.71
	AASS	40 to 71	96.99
TEH	ADOC	31 to 64	99.06
	AFOR	6 to 27	89.04
	AAVI	5 to 17	93.47
	AASS	36 to 65	96.06
BIO	ADOC	20 to 46	100
	AFOR	1 to 12	78.94
	AAVI	11 to 26	100
	AASS	37 to 64	100

The groups formed by k-means, guided by the grade, age and access to resources, define standards: in all courses, the Southeast region is the majority region of origin of the students; the age groups vary between courses; however, in all courses, the groups concentrate younger students: in MAT, 20 to 30 years, and 30 to 40 years; in SES, 30 to 40 years; TEH, 20 to 30 years; and BIO, 20 to 30 years and 30 to 40 years; with the exception of BIO, whose rate is very high, the cancellation /leave of absence/transfer rate of the other courses is similar; all course disciplines participate in the respective group; the groups consist of students who attend more than one discipline, that is, the same student is approved in more than one discipline with the same access profile; the number of students is higher in the last three disciplines of the semester, which means they are beginner-level students with more course time. MAT concentrates students in sequences 3 and 4. With the exception of MAT, for all courses, the access preference order, given by the number of accesses, is: access to tasks (AASS), access to documents (ADOC), access to forums (AFOR) and access to notices (AAVI). However, the amount of access considering the study's observations is low; we can observe that the interaction range (sum of AMTU, AFPA and AFAP) ranges from 1 to 10, and records the number of interactions between students and teachers, expressed through the sum of the number of accesses registered in the messages to teachers (AMTU), the amount of teacher response to students in the forums (AFPA) and the amount of responses of students to teachers in the forums (AFAP), and is very low when compared to other accesses (students in the group make at least one access to all resources); analyzing the sequence of disciplines provided, in general, the number of accesses decreases during the semester. Documents are preferred at the beginning of the semester (ADOC) and tasks (AASS) have a uniform usage preference. The access variables are scattered, against the value of their mean, and are the most heterogeneous: AFOR and AAVI; considering the interquartile ranges (50%) of the variables of each group, similarity is found in the number of accesses, and are lower for BIO; as regards the percentage of approval, MAT is the best course, with high percentage of approvals in relation to the total observations of the study and the total number of cases in the group. BIO is the worst group because it only gathers 50% of the approvals of the study; with the exception of MAT, the performance improves during the semester, particularly in the last discipline, that of sequence 6; thus, MAT is a course with a different profile, since it has the highest percentage of approval in sequences 3 and 4; as for the quality of the group, the groups generated have low to moderate cohesion and low separability, and MAT is the best course; concerning the relevance of the variables, except for BIO, the most important for all courses are AASS, ADOC and AAVI (obtained by the decision tree).

5. Conclusion

The adoption of BI tools to analyze the data allows a series of inferences that can assist the managers in the decision processes, aiming at the improved teaching, in general. BI techniques show that tools are complementary and that an in-depth study of the data must be conducted with more than one analysis technique to produce consistent conclusions.

In general, common points are observed in the results. One of them, for example, is that in all courses the task access number (AASS) is always the most important attribute, followed by the number of accesses to the documents. In other words, the students who performed better were those who recorded greater access to tasks and documents. Also, tasks (AASS) have a uniform usage preference throughout the semester.

Access to notices and access to forums, which are other forms of communication provided by VLE platforms, have a lower influence on student performance and greater access volume at the beginning of the analyzed period. This is also true of the documents, which are preferred at the beginning of the semester (ADOC). In general, it is observed that the lack of access by students characterizes low final means. However, the reverse is not true. Finally, we can observe that the variable age has little influence on the performance of the students within the universe of data analyzed.

The conclusions of the analysis found in this work are only a first step toward the construction of a model that, in the future, will provide objective information for the improvement of VLE platform communication systems.

References

Cheng, J. (2017). Data-Mining Research in Education [PDF]. Retrieved from https://arxiv.org/pdf/1703.10117.pdf

Cruz, M. A. S., Duarte, J. S., & Goldschmidt, R. R. (2017). Keystroke Dynamics Applied to Periodic Authentication in Virtual Learning Environments. *Brazilian Journal of Computers in Education, 25*, 36-60. doi: 10.5753/RBIE.2017.25.02.36.

Corsatea, B. M., & Walker, P. F. (2015). *Opportunities for Moodle data and learning intelligence in virtual environments.* Paper presented at the IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), Douai, France. Retrieved from https://www.semanticscholar.org/paper/Opportunities-for-Moodle-data-and-learning-in-Corsatea-Walker/b3d1105498bca47017dbcdc255700d8ce767742e

Ichihara, A. T., & Omar, N. (2017). Proposta de um ambiente para avaliação do sistema de comunicação de cursos EAD baseado em técnicas de business intelligence e em referenciais de qualidade do MEC/SEED [Proposal of an environment for evaluation of the EAD course communication system based on business intelligence techniques and the MEC/SEED quality referentials]. *Minutes of the Ibero-American IADIS conferences 2017 and applied computing 2017, 5*, 45-55. Vilamoura, Algarve, Portugal: IADIS.

Kimball, R., & Ross M. (2002). *The Datawarehouse Toolkit: the complete guide to dimensional metodology*. New York, USA: Wiley Compute Publishers.

Leskovec, J., Rajaraman, J., & Ullman, J. A. (2010). *Mining of Massive Datasets* (2nd ed.). Stanford University, CA: Cambridge University Press. doi.org/10.1017/CBO9781139924801

MEC/SEED - Ministério da Educação. Secretaria de Educação a Distância [Ministry of Education. Secretariat of Distance Education] (2007). Referenciais De Qualidade Para Educação Superior a Distância [Quality Referential in Distance Higher Education] [PDF]. Retrieved from http://portal.mec.gov.br/seed/arquivos/pdf/legislacao/refead1.pdf

Pascal, G., Servetto, D., Mirasson, L. U., & Luna, Y. (2017). Aplicación de Business Intelligence para la toma de decisiones en Instituciones Universitarias.

Implementación de Boletines Estadísticos en la Universidad Nacional de Lomas de Zamora (UNLZ). *Revista Electrónica sobre Tecnología, Educación y Sociedad, 4*(7). Retrived from http://www.ctes.org.mx/index.php/ctes/article/viewFile/659/752

Peres, S. M., & Lima, C. A. M. (2015). Medidas de avaliação de agrupamentos (Clustering) [Evaluation measures of clustering]. Retrieved from http://each.uspnet.usp.br/sarajane/wp-content/uploads/2015/11/avaliacao_clustering.pdf

Pifarré, K. T. (2015). Metodología para la creación de conocimiento en los entornos virtuales de enseñanza mediante herramientas de Business Intelligence. Retrieved from http://hdl.handle.net/10803/385623

Preidys, S., & Sakalauskas, L. (2010). Analysis of students' study activities in virtual learning environments using data mining methods. *In Technological and Economic Development 16*(1), 94-108. doi: 10.3846/tede.2010.06.

Ramos, W. M., Almeida, T. R. de, & Nóbrega, G. (2016). Estudo sobre padrões de participação e avaliação da satisfação discente em disciplina híbrida no ensino superior por meio de mineração de dados [Study on patterns of participation and evaluation of student satisfaction in hybrid discipline in superior education through data mining] [PDF]. Retrieved from http://www.aforges.org/wp-content/uploads/2016 /11/16-Wilsa-Maria-Ramos-et-al_Estudo-sobre-padroes-de-participacao.pdf

Witten, I.H., & Frank E. (2005). Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.). San Francisco, CA: Elsevier. Retrieved from ftp://ftp.ingv.it/pub/manuela.sbarra/Data%20Mining%20Practical%20Machine%20Learning%20Tools%20and%20Techniques%20-%20WEKA.pdf

Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining* (1ª ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. Retrieved from https://www-users.cs.umn.edu/~kumar001/dmbook/dmbook_2ed.jpg

Therneau, E. J., & Terry, M. (2018). An Introduction to Recursive Partitioning Using the RPART Routines [PDF]. Retrieved from
https://cran.r-project.org/web/packages/rpart/ vignettes/longintro.pdf

Turrioni, A. M. S., & Stano, R. C. M. T. (2009). *Critérios de avaliação para a educação a distância [Evaluation criteria for education distance].* Paper presented at the XXII Ibero-American Congress of Distance Education, Manaus, AM, Brazil.

Zacharis, N. Z. (2018). Classification and Regression Trees (CART) for Predictive Modeling in Blended Learning. *I. J. Intelligent Systems and Applications, 3*, 1-9. doi: 10.5815/ijisa.2018.03.01 Retrieved from http://www.mecs-press.org.