Educational Administration: Theory and Practice
2024, 30(6)(s), 155-163
ISSN: 2148-2403
https://kuey.net/

Research Article

Data Mining Of Educational Data In Government Distance Learning

Jalpa N. Gondaliya[1*], Dr. Hiren R. Kavathiya[2]

[1*]Research Scholar, Department of computer science, Atmiya University, Rajkot,Gujarat
[2]Dr. Hiren R. Kavathiya, Department of computer science, Atmiya University, Rajkot,Gujarat

Citation: Jalpa N. Gondaliya, et al. (2024) Data Mining Of Educational Data In Government Distance Learning *Educational Administration: Theory and Practice*, *30*(6)(s) 155-163
Doi: 10.53555/kuey.v30i6(S).5339

ARTICLEINO

ABSTRACT

Classification methods based on decision trees are used to confirm a correlation between students activity patterns in class and their final grades. By facilitating tasks like identifying participant characteristics, doing predictive performance analysis, and recognising learning kinds and patterns, Educational Data Mining (EDM) has proven to be an indispensable tool for enhancing online and distance learning (ODL). There is a significant body of research on the surroundings of universities and colleges presented in the scientific literature. However, the pedagogical paradigm used in these settings shares features with higher-level classes. In this section, we propose the application of EDM techniques for descriptive and predictive identification of interaction patterns in a governmental corporate Virtual Learning Environment (VLE), in the offer of short-term training courses in the instructional modality (with tutoring). Data were analysed regarding the interaction logs of students from two classes of a distance learning course. Classification methods based on decision trees are used to confirm a correlation between students' activity patterns in class and their final grades. Then, through clustering techniques and using the final grades as criteria, the groups of students separated according to the characteristics of interaction with the VLE and the final performance are identified. The results show that the application of EDM techniques can be used in corporate education scenarios, identifying the interaction profiles of students according to the performance obtained at the end of the course.

Keywords: Classification; Clustering; Corporate Distance Learning; Government Schools; Educational Data Mining.

I.INTRODUCTION

The area of Educational Data Mining (EDM) aims to apply computational techniques for the treatment of large masses of data generated in Virtual Learning Environments (VLE). The EDM is based on providing the discovery of knowledge that is relevant, unique and valid, as well as: the identification of patterns among students; the predictive analysis of performance; and the identification of profiles, in order to assist the qualitative management of distance education [Baker et al. 2011].
The work in the area of EDM is highly concentrated in scenarios related to a specific type of institution, the Higher Education Institutions (HEIs). The methodology of offering distance education of HEIs is focused, of course, on the courses that such institutions offer. These works have specific characteristics in relation to the methodology in which teaching is offered, such as: pre-academic information of students, duration of courses, information on economic indicators and variables related to other activities of the institutions
These characteristics, which are common in some studies related to Virtual Learning Environments (VLE) of HEIs are often not present in other types of institutions, such as those of governmental corporate education, i.e., the National School of Public Administration.
Within this specific context – distance education in the corporate government environment – this work focuses on data generated by the interaction of students with the VLE during the offer of a course of great importance in the training of public servants, in the year 2015, the course of Project Management: Theory and Practice.

Data from the logs referring to 698 students of this course were analysed from the perspective of identifying student profiles and potential causes of failure.
Supervised pattern classification is a task that is characterized by organizing objects into predefined classes. This is a systematic approach to building classification models from data sets. There are several techniques that can be used, e.g. decision tree-based classifiers, rule-based classifiers, artificial neural networks, support vector machines, and Bayesian classifiers [Tan et al. 2009].
In educational settings, the prediction of student performance has two distinct contexts for its application: 1) the study of the influence of the attributes of a specific model for the prediction ofa class and 2) prediction of an outcome for an output target class according to the predictive attributes used. It is possible, in this sense, to direct classification techniques for the analysis and prediction of student performance, enabling the identification of patterns that can be monitored as intervention indicators for the improvement of distance education [Baker et al. 2010].

II.RELATED WORKS

Bachhal, P., Ahuja, S., & Gargrish, S. (2021): Data mining is a very important part of the field of education. The main goal of this study is to find out how researchers have used data mining in the past and what is happening with data mining in educational research right now. It talks about how learning analytics and academic data were used with academic data. EDM uses computer methods to analyse data about education in order to answer study questions about education. This paper talks about the most important works done in this field so far. EDM is put into place, and the different user groups, training environments, and data are all set up.
Zhang, J. (2023): The new mode of carrying Internet+ on educational resources, made possible by the expansion of online learning, eliminates some of the inequalities in access to education that have arisen due to geographical and temporal disparities in the past. Therefore, a data-driven intelligence platform for education. The platform's big data centre collects and stores all business data from networks, sensors, and other devices in enormous data storage devices; the software design component collects student data and completes early warning of student plight by employing the multi feature fusion acquisition method. The site features video lessons and career guidance for self-study. Create a series of interest-based plates to complement Holland's job interest exam and better inform today's college students about their career options.
Adekitan, A. I., & Salau, O. (2019): Knowledge gained through machine learning methods aids in improving decision making in higher education, leading to a rise in research studies on educational data mining. “The purpose of this research was to use a data mining model built in Konstanz Information Miner (KNIME) to predict a student's cumulative grade point average (CGPA) after five years of study in a Nigerian university based on their major, year of entry, and GPA after the first three years of study. Six data mining techniques were evaluated, with a best accuracy of 89.15%. Both linear and pure quadratic regression models were used to confirm the conclusion, with R2 values of 0.955 and 0.957, respectively”. This affords the chance to proactively intervene on behalf of students who are at risk of not graduating or of graduating with subpar outcomes.
Almeida et al., (2016) presents a study related to the development of a specific plugin for Moodle, aiming to combat dropout in distance courses. This plugin had as its main functionality the improvement of the communication of the institution with the students, reducing the rates of evasion through the automatic sending of messages according to specific rules defined according to the level of intervention desired.
Coelho et al. (2015) conducted a study of the data related to the information present in Enap's communication channels. The authors identified that the number of calls decreased considerably, taking into account the actions carried out over the years by this school. The authors also presented results of the activity of mining terms, which were recorded in the calls of the communication channels with the institution, which indicated the source of the main problems in relation to the use of Enap's virtual school.
Gottardo et al. (2014)The work carried out by served as the basis for the realization of this article. In this work, the authors used variables related to the level of interaction with the VLE, student-student interaction and bidirectional student-teacher interaction for predictive analysis of the performance of students. The Random Forest and Multilayer Perceptron Network classification models were used.
Baker et al. (2011) presents the possibilities of applying EDM techniques in Brazil. In this paper, the authors demonstrate the various data mining tasks that can be applied to the educational context. The study presents an analysis of several works and demonstrates how this new area of research can contribute to a better understanding of the teaching and learning processes and to the motivation of students who use distance education in the Brazilian education scenario.
It is noticed that all the mentioned works do not simultaneously associate the analysis of interaction variables with the techniques of machine learning applied to the context of corporate education. Hence this work is positioned precisely in this niche, using data related to the variables of interaction with the VLE, with classification techniques from the C4.5 decision tree, the J48 implementation [Quinlan 1993], and clustering with the K-means algorithm [Hartigan and Wong 1979], generating indicators aimed at supporting decision making in scenarios of corporate distance education offering.

3. METHODOLOGY

The CRISP-DM framework organizes data mining projects into six phases: (A) Business Understanding, (B) Data Understanding, (C) Data Pre processing, (D) Modeling, (E) Evaluation, and (F) Implementation, as presented in Fig. 1 [Wirth and Hipp 2000]. “In this representation, the internal arrows indicate the most important and frequent dependencies between the phases. The outer circle symbolizes the cyclical nature of data mining, where lessons learned throughout the process can trigger new questions for project phases”.

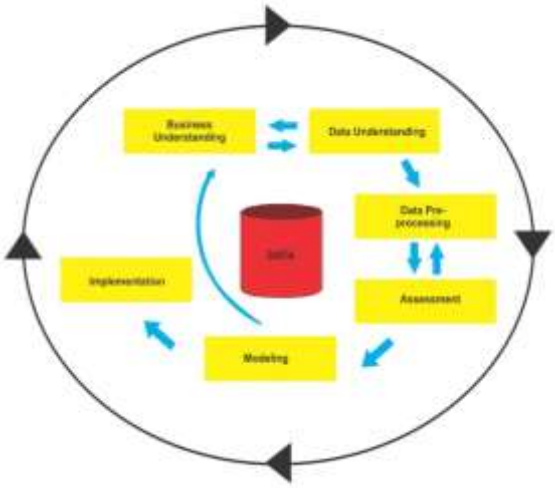

Figure 1. CRISP-DM Framework Steps

The experiment followed the main phases proposed by the CRISP-DM methodology until the evaluation stage. The focus of the application of the framework is as support and for the development of this project, according to the phases described in the following subsections.

3.1 Business Understanding

This is the initial phase where the objectives and targets for data mining must be identified, generating a plan for the project [Wirth and Hipp 2000]. At Enap, the General Coordination of Distance Education (CGEAD) is responsible for offering distance learning courses. The courses are focused on the improvement and training of public servants in india. CGEAD uses the Moodle software (acronym for Modular Object Oriented Dynamic Learning Environment) as a VLE.

During this phase, the course Project Management: Theory and Practice was defined as a target for this work, because, for Enap, it is a strategic course in the training of managers working in the indian public service. Among the courses offered in the distance modality with tutoring, the selected course is the one that presents the largest number of modules for the iteration of students with the VLE, and can be used as a basis for the application of the mining model proposed in this article, in the other courses of this modality that are offered.

Still at this stage, it was possible to identify that the school had a dropout rate of about 20%. In addition, there are no indicators that allow the understanding of these numbers, and it is necessary to enable measures for possible interventions.

3.1 Grasping the Business Landscape

This preliminary stage is crucial for laying out the objectives and goals for data extraction and analysis, thereby forming a strategic outline for the project [Wirth and Hipp 2000]. Within the Indian landscape, IIGNOU's (“Indira Gandhi National Open University) Department of Distance Education” (DDE) is entrusted with the task of administering distance learning courses. The primary focus of these courses is to augment the competency and training of India's civil servants. To aid this, the DDE employs Moodle software, known for its Modular Object Oriented Dynamic Learning Environment, as a digital platform for education.

During this initial stage, the course 'Project Management: Theory and Practice' was chosen as the main subject for this research. For IIGNOU, this course is of strategic significance in enhancing the skills of managers in the Indian civil service. Amongst the myriad of distance learning courses that include tutoring, the selected course stands out due to the maximum number of modules facilitating student engagement with the VLE. This could potentially provide the groundwork for the integration of the data mining model, proposed in this article, across other similar courses.

At this point, it was clear that the institution faced a challenge with a dropout rate hovering around 20%. Furthermore, the lack of any metrics to shed light on these statistics calls for the establishment of strategies for potential corrective actions.

3.2 Understanding of Data

This phase addresses the initial data collection and also serves to familiarize the researchers involved with the project-specific data types [Wirth and Hipp 2000]. In this sense, the target database for the development of the data mining project was identified: the AVA Moodle database, which has approximately 361 native tables of the system. This database has all the records related to the use of the VLE in the realization of the courses offered by Enap during the year 2015.

The dimensions related to the students' data, the characteristics of the courses, the characteristics of the VLE, the log records, the characteristics of the evaluations of the courses and the interaction characteristics of the participants were analyzed.

As can be seen in Table 1, the attributes are related to the level of interaction of students with the Moodle VLE modules. These attributes represent for Enap interaction indicators that will serve as a basis for the taking of splits, in order to improve the quality of the courses that are offered.

Table 1. Descriptive table of VLE interaction attributes

Attribute	Attribute Properties
primeiro_acesso	Time in days that students made the first access to the course in the VLE
count_quiz_view	Number of views to the course assignment module
count_page_view	Number of views to the course support content
count_book_view	Number of views to the course content module
count_forum_view	Number of views of course forums
count_folder_view	Number of views to the course virtual library module
count_questionnaire_view	Number of views in content pinning surveys
count_quiz_submitted	Number of unscored activities submitted
count_forum_upload	Number of messages sent in the course forums
count_assign_view	Number of views to the course's scored activities
count_assign_submitted	Number of scored activities submitted for evaluation
count_questionnaire_submitted	Number of self-assessment activities submitted
nota_final	Final grade obtained by students at the end of the course.

Still in this stage, the final result obtained by the students was defined as the target attribute for the data mining activity. This information is present in attribute not a_final, which records the grade obtained by students in a range of 0 to 100 points.

3.3 Pre-processing of Data

In this phase, the pre-processing of the data is performed for the construction of the data set to be used in the model defined for mining [Wirth and Hipp 2000]. Initially, the discretization activity of the target attribute not a_final was carried out, in which the students' grades were separated into specific categories, according to the classification used by Enap. This resulted in the new attribute CLASS NOTE, with the students distributed among three possibilities:

- 155 students in Class EVA (Dropout), for grades with 0 points obtained;
- 130 students in Class REP (Failed), for grades between 1 and 59 points obtained and
- 413 students in Class APR (Approved), for grades with values between 60 and 100 points obtained.

The data of the por fim attributes were separated in a specific table, which served as the basis for the follow-up of the mining project totaling 698 instances, considered of good quality and without null values.

3.4 Modeling

This is the phase that defines the model “that will be used for data mining, which, in practical terms, involves choosing the specific mining activities: classification, with supervised learning, for predictive performance analysis and clustering, with unsupervised learning, for the identification of information related to student profiles. Both activities used the WEKA software (Waikato Environment for Knowledge Analysis [Hall et al. 2009])”.

In order to validate the model inferred by the classification activity, we used the technique of separating the database into 2/3 for training and 1/3 for tests, i.e. 459 instances for the first and 239 instances for the second. In the case of clustering,all 698 instances were used in the execution of the algorithm.

Prior to the classification task, a variable selection technique was also used, with the objective of reducing the presence of irrelevant and redundant attributes in the data set and thus improving the quality of the results [Silva 2009]. This activity proved to be necessary because it was observed that the use of the entire group of variables, as shown in Table I, had not been presenting acceptable performance in preliminary trials.

Specifically, the CfsSubsetEval algorithm [Hall and Smith 1998]was used with the Best First search technique [Rich and Knight 1991]. After its execution in the WEKA tool, the attributes were selected according to Table 2.

Table 2. Selected attribute group for classification activity

Selected attributes	
count_quiz_view	count_quiz_submitted
count_questionnaire_view	count_questionnaire_submitted
count_assign_submitted	count_forum_view count_forum_uploaded

In Table 2 it is possible to observe that the attributes selected by the algorithms are related to the interaction with the activity modules (quiz and questionnaire) and with the forum module. This selection demonstrates that these are the attributes that have a greater correlation when considering the context of prediction of the final grades.

3.5 Review score

For the first activity, which included the classification from decision trees, the proposed model presented good results to predict the path taken by the students, targeting the final grade class obtained. Fig. 2 shows the results of the execution of the J48 algorithm.

Performance Algorithm J48 - Test Base		
Correctly Classified Instances	209	87,46%
Incorrectly Classified Instances	30	12,55%
Total instances	239	100%

(a) Assertiveness index

```
=== Confusion Matrix ===
   a   b   c   <-- classified as
 126   0   8 |   a = APR
   3  52   3 |   b = EVA
  10   6  31 |   c = REP
```

(b) Confusion matrix

Figure 2. J48 Algorithm Performance

We can see in Fig. 2a that the model correctly classified about 87% of the total instances and incorrectly classified about 13%. Fig. 2b shows the confusion matrix generated by the algorithm. We classified 134 instances for the cl asse APR, being 126 instances correctly classified and 8 errors that were classified as REP. Class EVA had 58 instances classified being 52 correct and 6 with errors, being 3 instances as APR and3 as REP. Finally, the REP class obtained a t otal of 47 instances being 31 correctly classified and 16 with errors being 10 as APR and 6 asEVA.

As for the second activity, Fig. 3 presents the results obtained with the execution of the K-means clustering algorithm, with K = 3, which was defined

considering the number of classes referring to the student's final result — see Section 4.3. The clusters formed by the algorithm indicate the reference values of the iterations with the VLE according to the grade class obtained by the students.

```
kMeans
======
Cluster centroids:
                                Cluster#
Attribute                  Full Data        0        1        2
                             (459)      (279)     (83)     (97)
=============================================================
primeiro_acesso              3.2026     2.6308    2.988    5.0309
count_book_view            279.6471   371.3082  205.9277  79.0825
count_quiz_view             57.4074    78.086   44.3012   9.1443
count_quiz_submitted        11.1983    15.2581   8.4217   1.8969
count_folder_view            11.024    13.3226  10.8675   4.5464
count_page_view             11.6078    15.1326  10.4578   2.4536
count_questionnaire_view     2.0022     3.2509   0.1446        0
count_questionnaire_submitted 1.1176     1.81    0.0964        0
count_assign_view           54.8802    83.9211  20.7108   0.5876
count_assign_submitted       4.5316     6.9534   1.6867        0
count_forun_view           145.6514   196.6595  108.4578  30.7629
count_forun_uploaded        16.6928    22.8315  13.5783   1.701
nota final                      APR        APR      REP      EVA

Clustered Instances
0      134 ( 56%)
1       47 ( 20%)
2       58 ( 24%)
```

Figure 3. Characteristics of the groups generated by the K-means algorithm

The Algorithm grouped the students in the clusters according to the centroids of the variables analysed, being grouped in cluster 0 the students who were approved (APR), in cluster 1 the students who failed (REP) and, finally, grouped in cluster 2 the students who dropped out of the course (EVA). It is noteworthy that the variable final grade was also included in this activity, because the purpose was to analyse the typical profile of students regarding the final result in the course.

4. DISCUSSION

The results obtained with the classification algorithms showed high accuracy values for the model used. The techniques used correctly classified around 87% of the instances belonging to the test set, in relation to the classes of the final result obtained.
From the decision tree model inferred by the J48 algorithm, presented in Fig. 4, it was possible to analyse the level of interaction and the relationship of the variables according to the path of interactions traveled by thelion's. It is observed in the tree a referential base on how the students reached the grades. The algorithm defined the variable count_forum_uploaded, which represents the number of posts sent to the forum, as the central node of the tree, identifying it as the variable of greater expressiveness of the classification task.

From it follow the other nodes of decision inferred by the method, until the final decision as to the class is reached, in the leaves of the tree. Note, for example, that a value above 2 for count_forum_uploaded and above 0 for count_questionnaire_view, i.e. the student has ever accessed the quiz module, already indicates the approval status for the student.

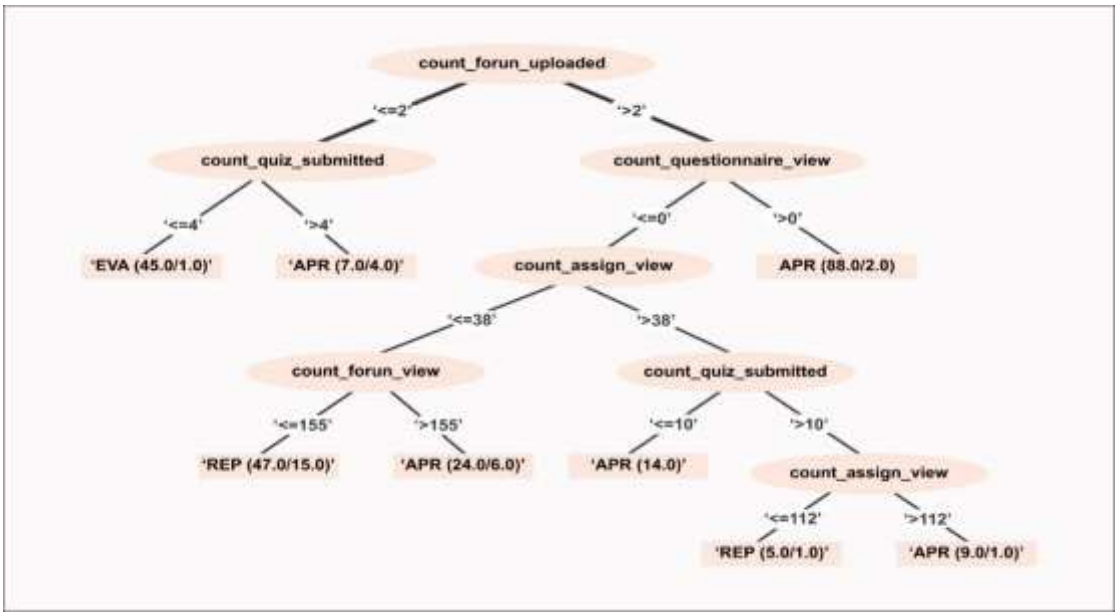

Figure 4: Decision tree generated by the J48 algorithm

For the clustering activity, with thealready mentioned, the K-means algorithm available in the WEKA tool was used. In order to provide more information about the students' profiles, a 2D analysis of variables related to the clustering result was also performed. As shown in Fig. 5a, the K-means algorithm displays the three clusters on the X-axis in full compliance with the three categories of notes arranged on the Y-axis.

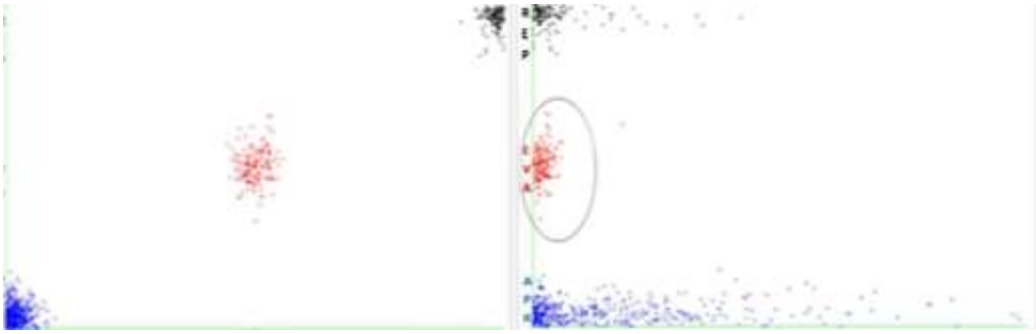
(a) Cluster projection vs. final grade (b) Cluster projection vs. count_assign_view
Figure 5: 2D analysis of K-means results

Fig. 5b highlights the analysis in relation to variable count_assign_view, which represents the number of views to the task module. It is possible to identify the level of interaction of the students who were classified in cluster 1, which represent the final grade EVA (dropout students). A large concentration of instances with the variable count_assign_view close to 0 can be observed, which indicates a low interaction with this module, unlike the other clusters, which have a greater distribution in values.

5. CONCLUSIONS AND FUTURE WORK

The results showed a good predictive potential with about 87% of assertiveness, from decision trees, of the students' performance when the variables associated with the interaction with the Moodle modules used by Enap in its VLE were analysed. “The
clustering provided relevant information in relation to the profile of the students that allows the identification of information that can contribute to the analysis of behaviors that, when observed, enable the planning of specific pedagogical actions with greater effectiveness”.
As a future work, variables related to students' socioeconomic data and/or variables related to students' professional information can be considered. It is also a future perspective to analyse the variables of this work under the particular context of time, i.e. according to the week of progress of the course. It is also a perspective of future work to use the information generated by the clustering activity to analyse the behaviours of students.

REFERENCES

[1]. Bachhal, P., Ahuja, S., & Gargrish, S. (2021, August). Educational data mining: A review. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012022). IOP Publishing.

[2]. Zhang, J. (2023, January). Research on System Design of Educational Curriculum Construction Based on Big Data Platform. In *Application of Big Data, Blockchain, and Internet of Things for Education Informatization: Second EAI International Conference, BigIoT-EDU 2022, Virtual Event, July 29–31, 2022, Proceedings, Part III* (pp. 177-185). Cham: Springer Nature Switzerland.

[3]. Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250.

[4]. de Almeida, L. R., da Costa, J. P. C., de Sousa, R. T., de Freitas, E. P., Canedo, E. D., Prettz, J., ... & Del Galdo, G. (2016, October). Motivating attendee's participation in distance learning via an automatic messaging plugin for the moodle platform. In *2016 IEEE Frontiers in Education Conference (FIE)* (pp. 1-5). IEEE.

[5]. Coelho, V. C. G., Costa, J. P. C. L. d., Souza, D. d. C. R. d., Canedo, E. D., Silva, D. G. e., and Sousa Júnior, R. T. d. Mining of educational data to identify barriers in the use of distance education. In 21st ABED International Congress of Distance Education. ABED, 2015.

[6]. Gottardo, E., Kaestner, C. A. A., and Noronha, R. V. Estimation of academic performance of students: Analysis of the application of data mining techniques in distance courses. Revista Brasileira de Informática na Educação 22 (01): 45, 2014.

[7]. Baker, R. S. J., Isotani, S., and de Carvalho, A. M. J. B. Educational data mining: opportunities for Brazil. Brazilian Journal of Informatics in Education 19 (2), 2011.

[8]. Baker, R. S. J. D., McGaw, B., Peterson, P., and Baker, E. Data mining for education. International encyclopedia of education vol. 7, pp. 112–118, 2010.

[9]. Bresfelean, V. P. Analysis and predictions on students’ behavior using decision trees in weka environment. In Proceedings of the Information Technology Interfaces (ITI). IEEE, pp. 25–28, 2007.

[10]. Aggarwal, C. C., & Zhai, C. X. (2013). Mining text data. (Vol. 9781461432). https://doi.org/10.1007/978-1-4614-3223-4

[11]. Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. Journal of Communication, 54(3), 402–420. doi:10.1111/j.1460-2466.2004. tb02636.x

[12]. Aparicio, M., Bacao, F., & Oliveira, T. (2016). An e-learning theoretical framework. Educational Technology and Society, 19(1), 292–307.

[13]. Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. Journal of Computing in Higher Education, 26(1), 87–122. https://doi.org/10.1007/s12528-013-9077-3

[14]. Bozkurt, A., Akgun-Ozbek, E., Yilmazel, S., Erdogdu, E., Ucar, H., Guler, E., Sezgin, S. Karadeniz, A., Sen-Ersoy, N., GokselCanbek, N., Dincer, G. D., Ari, S., & Aydin, C. H. (2015). Trends in distance education research: A content analysis of journals 2009-2013. International Review of Research in Open and Distance Learning, 16(1), 330–363. https://doi.org/ 10.19173/irrodl.v16i1.1953

[15]. 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. https://doi.org/10.1016/j. iheduc.2015.04.007

[16]. Çakiroğlu, Ü, Kokoç, M., Gökoğlu, S., Öztürk, M., & Erdoğdu, F. (2019). An analysis of the journey of open and distance education: Major concepts and cutoff points in research trends. International Review of Research in Open and Distance Learning, 20(1), 2–20. https://doi.org/10.7202/1057969ar
[17]. Carr, S. (2000). As distance education comes of age. Chronicle of Higher Education, 46(23), A39–A41.
[18]. Diaz, D. P. (2002). Online drop rates revisited. The Technology Source, 2002(1), 93–106.
[19]. Ellis, R. A., Han, F., & Pardo, A. (2018). When does collaboration lead to deeper learning? Renewed definitions of collaboration for engineering students. IEEE Transactions on Learning Technologies, 12(1), 123–132.
[20]. Esfahani, N., Elkhodary, A., & Malek, S. (2013). A learning-based framework for engineering feature-oriented self-adaptive software systems. IEEE Transactions on Software Engineering, 39(11), 1467–1493. https://doi.org/10.1109/TSE. 2013.37
[21]. Fang, J. W., Hwang, G. J., & Chang, C. Y. (2019). Advancement and the foci of investigation of MOOCs and open online courses for language learning: A review of journal publications from 2009 to 2018. Interactive Learning Environments, 1–19. https://doi.org/10.1080/10494820.2019.1703011
[22]. Garrison, D. R. (2011). E-learning in the 21st century: A framework for research and practice (2nd ed.). Taylor & Francis.
[23]. Granić, A. (2011). Usability testing and expert inspections complemented by educational evaluation. Journal of Educational Technology & Society, 14(2), 107–123.
[24]. Gurcan, F. (2018). Major research topics in big data: A literature analysis from 2013 to 2017 using probabilistic topic models. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 1–4.
[25]. Gurcan, F., & Kose, C. (2017). Analysis of software engineering industry needs and trends: Implications for education. International Journal of Engineering Education, 33(4), 1361–1368.
[26]. Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003-2008: A categorical meta-trend analysis using text mining techniques. Journal of Computing in Higher Education, 24(1), 1–17. https://doi.org/10.1007/s12528-011-9044-9
[27]. Hussein, M., Han, J., & Colman, A. (2011). An architecture-based approach to context-aware adaptive software systems, Technical Report, #C3-516_04, Swinburne University of Technology.
[28]. Hwang, G. J., & Fu, Q. K. (2019). Trends in the research design and application of mobile language learning: a review of 2007–2016 publications in selected SSCI journals. Interactive Learning Environments, 27(4), 567–581. https://doi.org/10. 1080/10494820.2018.1486861
[29]. Jeong, H. Y. (2016). UX based adaptive e-learning hypermedia system (U-AEHS): an integrative user model approach. Multimedia Tools and Applications, 75(21), 13193–13209. https://doi.org/10.1007/s11042-016-3292-7
[30]. Jivani, A. G. (2011). A comparative study of stemming algorithms. International Journal of Computer Technology and Applications, 2(6), 1930–1938. https://doi.org/10.1.1.642.7100
[31]. Jung, I. (2011). The dimensions of e-learning quality: From the learner's perspective. Educational Technology Research and Development, 59(4), 445–464. https://doi.org/10.1007/s11423-010-9171-4
[32]. Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning.
[33]. Lin, H. C., & Hwang, G. J. (2019). Research trends of flipped classroom studies for medical courses: A review of journal publications from 2008 to 2017 based on the technology-enhanced learning model. Interactive Learning Environments, 27(8), 1011–1027. https://doi.org/10.1080/10494820.2018.1467462
[34]. Mathew, G., Agrawal, A., & Menzies, T. (2018). Finding trends in software research. IEEE Transactions on Software Engineering, https://doi.org/10.1109/TSE.2018.2870388
[35]. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. Journal of the American Medical Informatics Association, 18(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464
[36]. O'Donnell, E., Lawless, S., Sharp, M., & Wade, V. P. (2015). A review of personalised E-learning. International Journal of Distance Education Technologies, 13(1), 22–47. https://doi.org/10.4018/ijdet.2015010102
[37]. Ogada, K., Mwangi, W., & Cheruiyot, W. (2015). N-gram based text categorization method for improved data mining. Journal of Information Engineering and Applications, 5(8), 35–43.
[38]. Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. IEEE Transactions on Learning Technologies, 10(1), 82–92. https://doi.org/10.1109/TLT.2016.2639508
[39]. Rahimi, E., Van Den Berg, J., & Veen, W. (2015). Facilitating student-driven constructing of learning environments using Web 2.0 personal learning environments. Computers and Education, 81, 235–246. https://doi.org/10.1016/j.compedu. 2014.10.012
[40]. Sangra, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an inclusive definition of e-learning: An approach to the conceptual framework. International Review of Research in Open and Distributed Learning, 13(2), 145–159. https:// doi.org/10.19173/irrodl.v13i2.1161

[41]. Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic N-grams as machine learning features for natural language processing. Expert Systems with Applications, 41(3), 853–860. https://doi.org/
[42]. 10.1016/j.eswa.2013.08.015
[43]. Simonson, M., Schlosser, C., & Orellana, A. (2011). Distance education research: A review of the literature. Journal of
[44]. Computing in Higher Education, 23(2-3), 124. https://doi.org/10.1007/s12528-011-9045-8
[45]. Srivastava, A. N., & Sahami, M. (2009). Text mining: Classification, clustering, and applications. CRC Press.
[46]. Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. Review of Educational Research, 79(2), 625–649. https://doi.org/10.3102/0034654308325896
[47]. Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers and Education, 50(4), 1183–1202. https://doi.org/10. 1016/j.compedu.2006.11.007
[48]. Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Computers and Education, 94, 252–275. https:// doi.org/10.1016/j.compedu.2015.11.008