Learning profiles to assess educational prediction systems

Amal Ben Soussia[1], Célina Treuillier[1], Azim Roussanaly[1], and Anne Boyer[1]

Université de Lorraine, CNRS, LORIA, Campus scientifique 54506
Vandoeuvre-lès-Nancy, France
{firstname,lastname}@loria.fr

Abstract. Distance learning institutions record a high failure and dropout rate every year. This phenomenon is due to several reasons such as the total autonomy of learners and the lack of regular monitoring. Therefore, education stakeholders need a system which enables them the prediction of at-risk learners. This solution is commonly adopted in the state of the art. However, its evaluation is not generic and does not take into account the diversity of learners. In this paper, we propose a complete methodology which objective is a more detailed evaluation of a proposed educational prediction system. This process aims to ensure good performances of the system, regardless of the learning profiles. The proposed methodology combines both the identification of personas existing in a learning context and the evaluation of a prediction system according to it. To meet this challenge, we used a real dataset of k-12 learners enrolled in a french distance education institution.

Keywords: Learning Analytics · Assessment Methodology · Risk Prediction · Learning Profiles · K-12 Learners.

1 Introduction

Nowadays, schools and universities are moving towards online learning due to the generalization of digital infrastructures and learning platforms which allow to better meet the needs of learners. However, this learning modality is facing many challenges, and the most widespread is the high failure rate among learners. This phenomenon is due to many reasons such as the large diversity of student profiles expressing different needs and requiring personalized support [23].

Virtual learning environments (VLE) store learner's online activity. The corresponding data, called learning traces, is very diversified and is used by Learning Analytics (LA) [19]. One aim among others is to provide educational stakeholders with intelligent technology-based solutions to help them in identifying at-risk of failure learners as early as possible. These solutions need to take into consideration all learners behaviors. Therefore, a major issue is: **does a system perform equally with all learners profiles?**

To answer this research question, we propose a methodology which is based on the identification of personas, defined as learners profiles representations [8, 22] and on the evaluation of model's performances for each persona. We illustrate the methodology on a case study (evaluation of a prediction model). To resume, our main contribution relies on a more precise evaluation of educational systems, taking into account the different learning profiles and based on a broad range of metrics. We proceeded according to the following steps:

- Given the disparity of available learning traces, we defined several learning indicators characterizing a learner's behavior. Then, we identified homogeneous groups of learners sharing similar behaviors according to these indicators. These learners groups are finally characterized into personas.
- We reviewed the existing assessment indicators and identified new ones to complete the evaluation.
- We conducted a precise evaluation on a specific use case relying on a weekly prediction approach.

We carried out our experimentation using real data of k-12 learners, enrolled in a French distance learning center (CNED). This institution is characterized by the multi-modality of learning and the total autonomy of its learners.

This paper is organized as follows. The section 2 presents the general methodology and the used dataset. Section 3 and 4 present the first and second steps of the methodology respectively. The results of the evaluation are detailed in Section 5. A general conclusion and several perspectives are given in section 6.

2 Evaluation methodology

In this section, we start by describing the proposed the methodology and its different steps. Then, we present our case study for the experimental part.

2.1 Methodology description

In order to achieve our assessment objective, the methodology is organized around three main steps (See figure 1):

1. **Identification of learner profiles** from learning. The profiles are characterized by personas, containing key information about learners' behaviors.
2. **Run of the prediction system** on the data and measurement of a complete set of metrics, containing both precision metrics, as accuracy, and new time-dependent ones (earliness, stability).
3. **Deeper evaluation of the system** according to the identified learners profiles and the various performance metrics.

2.2 Case study

The case study concerns the k-12 learners enrolled in the physics-chemistry module during the 2017-2018 school year within the French center for distance education (CNED) [2]. It offers a large variety of fully distance courses to numerous

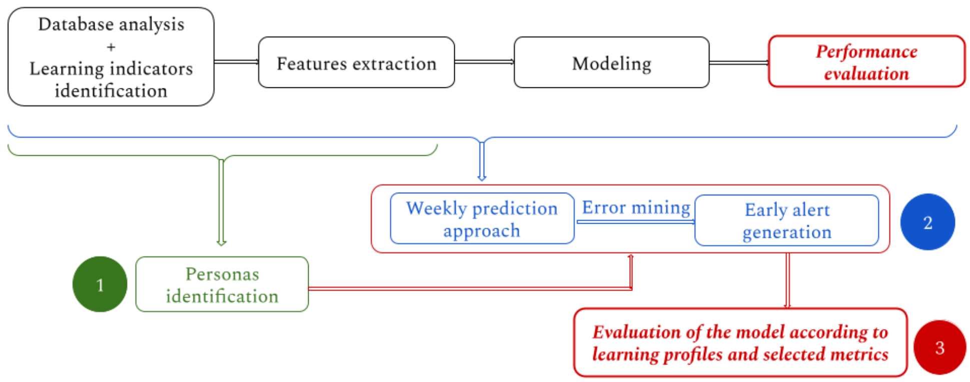

Fig. 1. The in-depth methodology phases.

physically dispersed learners. The courses contents are both available online and in printed papers which gives the learner the freedom to choose the learning mode which suits him/her the best. Given the large number of learners and the specificity of learning, it is highly time consuming for teachers to provide their students with an effective and personalized feedback.

2.3 Data description

The learning traces are collected from two data sources. The first one is the Moodle platform, which generates the *logs* and the interaction traces between the learner and the learning content. The second platform is GAEL, which is a management system where all performance data, including grades, are stored. In CNED, learners don't start the school year at the same time t_0 [2]. We select learners with t_0 between Sept. 1^{st} and Oct. 31^{st}, as they share similar learning paths and characteristics. According to this information, our database gathers learning traces of 639 learners. The learning period of the physics-chemistry course is 300 days, during which 6 exams could be submitted. On average, learners only submit around 4.51 assignments. The average mark on the submitted exams is 13.73. However, if we consider setting the grade of 0 to the unsubmitted assignments, this average is lowered to 10.21.
The bi-modality (digital or paper-based) of the learning makes the study of the dataset difficult. Indeed, learners who use the course exclusively in paper format may not produce any logs and it is therefore not relevant to compare them with active learners on the VLE. In our dataset, we noticed that 37.25% of the population never logged in. To handle this particularity, the dataset was divided into two subdatasets: one containing data about learners who made at least one log on the VLE, and the other one about those who have never logged in. Finally, the learners were classified into 4 classes according to their mean performance:

- **Success** *(C_1)* : average score superior to 12.

– **Medium risk of failure** *(C_2)* : average score between 8 and 12.
– **High risk of failure** *(C_3)* : average score inferior to 8.
– **Drop out** *(C_4)* : at least the two last assignments are not submitted.

The Table 1 summarizes the number of learners from *logs* and *no_logs* subdatasets belonging to each class. The process of identifying learners profiles within these classes is described in the following section.

	logs	*no_logs*
C_1	178	64
C_2	53	29
C_3	17	28
C_4	153	117
Total	**401**	**238**

Table 1. Number of students in each class for each subdataset.

3 Methodology step 1: identification of learner profiles

3.1 State of the art

Learners' behaviors are observed through their online learning traces. In LA, multiple studies exploit this data to compare learners based on various indicators, such as engagement [11], performance [4] or regularity [7]. In our context, learning behaviors need to be described according to a set of indicators, allowing a more detailed characterization of learners [22, 24]. For this reason, we define learners personas corresponding to typical learners identified through Machine Learning classification processes [8]. In one hand, the identification of such personas enables a more precise description of the corpus, especially in terms of learners profiles representation. In another hand, these personas meet the need for an ethical learning analytics implementation [20], and ensure fair support between learners and provide useful tools to the field stakeholders who need to help their learners with equal support [10]. However, the diversity of the available data makes the task tricky: the variety of recorded data does not allow for the same indicators to be computed all the time. The indicators we calculated for the case study are described in the following subsection.

3.2 Study and selection of learning indicators

Learning traces available in the CNED dataset are diverse and contain both logs and performance data. This data was first used to define five absolute indicators (e.g. calculated for each learner):

– **Engagement** reflects the learner's activity on the VLE (logs) [11].

- **Regularity** translates the learner's constancy of connection between the beginning and the end of the course (frequency of connection) [2].
- **Curiosity** expresses the intrinsic motivation of the learner to consult various educational resources (variety of accessed content) [17].
- **Performance** corresponds to learner's scores in the exams.
- **Reactivity** provides information about learner's responsiveness during course-related events (timeliness of the assessments)[7].

To go further, we completed these absolute indicators with a set of **relative indicators**. The average of each indicator is computed for all learners, and the associated relative indicator gives information about the behavior of a specific learner profile comparing to his/her peers (negative or positive difference in relation to the rest of the group). Both types of indicators were used as a basis for the identification of learners profiles, described in the following subsection.

Obviously, engagement, curiosity and regularity (on the VLE) indicators, based on the logs, were not computed for the *no_logs* subdataset as associated learners have never logged in.

3.3 Identification of learners profiles

For each subdataset, the study of learning indicators enables the identification of learners profiles. These profiles correspond to homogeneous subsets of learners, sharing similar behaviors, and are identified through different steps:

- **Data-preprocessing:**
  - *Data normalization:* use of the RobustScaler[1] method (ScikitLearn [15]) to improve the model's performance.
  - *Outliers identification:* Use of the IsolationForest[2] algorithm [14] to set apart the atypical data and increase model's performance. This step is crucial because outliers' atypicity does not allow them to be associated with other students.
- **Identification of homogeneous groups of learners:** k-means algorithm [13] is used to identify homogeneous groups of learners. Results are evaluated with Silhouette analysis [18] and Davies-Bouldin Criterion [9]. We run the algorithm with values from 2 to 15 and selected the one giving the best performance.
- **Description of learners profiles:** each of the identified clusters are then characterized by a size (number of associated students), its proportion in dataset to which it belongs and a set of learning indicators. Outliers are not discarded but are studied individually.

Applying this methodology, we identified 12 outliers among the 639 learners. Each of the remaining learners is associated to one of the 21 identified learners profiles. Some statistics are given in Table 2.

[1] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html

	Logs				No_logs			
	C1	C2	C3	C4	C1	C2	C3	C4
Number of inliers	176	52	16	151	63	28	27	115
Number of outliers	2	1	2	2	1	1	1	2
Optimal value of k	2	2	2	2	3	3	3	4
Silhouette Index	0,44	0,30	0,85	0,28	0,43	0,36	0,36	0,34
Davies-Bouldin Index	1,00	1,45	0,07	1,32	0,85	0,96	0,92	1,04

Table 2. Clustering results by subdatasets and classes.

3.4 Personas: examples

Each persona contains a large variety of information: narrative description of the learning behavior, its proportion in the dataset it belongs to, visual indicators of the risk of failure, and learning modality (See Figure 2).

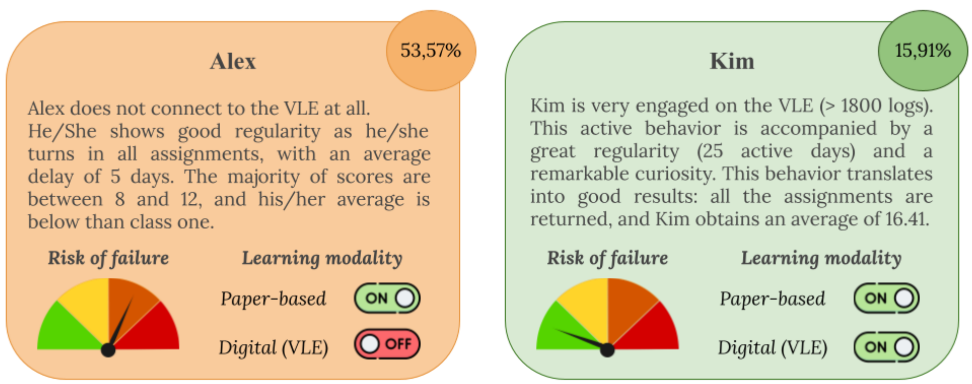

Fig. 2. Example of two personas.

The utility of such personas is threefold. In addition to providing valuable information about learners behaviors, they contribute to the improvement of the results interpretation of a LA system. Finally, they are particularly interesting for our study because they can be used to refine the evaluation of an educational system. The results presented in the section 5 confirm this last point.

4 Methodology step 2: earliness and stability measurements

In addition to the usual performance measures, this section defines new metrics for a deeper evaluation of an educational prediction system. These metrics consider the importance of the temporal evolution of the prediction.

4.1 State of the art

The main objective of the majority of educational prediction systems is the early identification of at-risk of failure or dropout learners. Static and precision Machine Learning (ML) metrics such as accuracy are mainly used to evaluate the performance of educational prediction systems. For example, [5] studied the accuracy of early warning system (EWS) on identifying at-risk students in a real educational setting. The study of [12] aimed to improve the performance of a dropout EWS by evaluating the trained classifiers with both receiver operating characteristic (ROC) curves and precision–recall (PR) curves. [3] compared the performance of a developped EWS on two different subjects based on the accuracy, the true negative rate (TNR) and the true positive rate (TPR) measures. [1] compares the performance of different ML model in analyzing the problems faced by at-risk learners enrolled in online university. This performance assessment is based on accuracy, precision, recall, support and f-score results. The majority of education prediction systems uses static and precision ML metrics for performance evaluation. However, both learning and prediction are time-evolving. Consequently, we need to consider the temporal dimension in the performance measures and illustrate the evolution of the whole process over the learning period. For this aim, we propose new metrics to evaluate the prediction and which the definition is based on the regular tracking of the prediction results.

4.2 Metrics description

Prediction earliness: Researchers work on providing stakeholders with the most accurate prediction results. A common theoretical definition of the early prediction is the right time to identify at risk learners. The earliness of the right prediction depends always on the studied context. We propose to measure the earliest time to predict as accurate as possible the classes of learners. We define the earliness of prediction as the mean time from which we start to correctly predict the learners classes [6]. While defining this measure, we focus on at-risk learners to best respond to the objectives of our study.
Prediction stability: Stability is usually related to small changes in system output when changing the training set [16]. In our context, we are interested in temporal stability referring to the capacity of a classifier to give the same output over time when training the same dataset [21]. We measure temporal stability as the average of the longest sequences of successive right predictions [6].

5 Methodology step 3: in-depth evaluation of a prediction system

This section presents the whole methodology from the prediction system description to the modeling and assessments steps. It ends up by a comparative study based on the obtained results.

5.1 Short description of the prediction model

Our system is based on a weekly prediction model of at-risk of failure or dropout learners. As explained, learners of the cohort are classified into four classes. First, we went through both processes of features extraction and selection. Going through these processes is important to select the activity features most correlated to the learner's final result as well as to minimize noise in the model. Thus, each week w_i, a learner is represented by a vector X composed of features going from f_1 to f_n and the class y to which he belongs to. Each learner belongs to one and only class over the year.

$$X = < f_1,f_2,...,f_n,y >$$

Each feature f_1 to f_n represents one learning activity till the prediction time w_i. For each prediction time w_i, the value of one feature is added to that of prediction time w_{i-1}: we proceed to an accumulation of values. Based on the accuracy results of [2], we use the Random Forest (RF) as a ML model for our system.

5.2 Results

In the first evaluation phase, we divided the test dataset population into two groups *(logs, no_logs)* as explained in the section 2.3. In this experimental part, we report on the results of 3 metrics: *accuracy*, *earliness* and *stability*.

Accuracy analysis The curves of the Figure 3 show a difference in the accuracy between the test dataset of the total population, *logs* and *no_logs* subdatasets. Indeed, we notice that until almost the week 15, classes of learners who belong to the *no_log* group are the best predicted. In fact, the dropout class is the most predictable one and is highly represented in the *no_log* subdataset (cf. Table 1). However, the further we advance in the school year, the more the prediction results of *logs* and *no_logs* converge towards almost the same values.

The figure 4 shows the curves of the evolution of the accuracy of the personas identified in *logs* (A) and *no_logs* (B) subdatasets. To ensure the figure lisibility, we only present the results of one persona per class and by subdataset: personas 1, 4, 5 and 8 were selected for the *logs* subdataset, and personas 10, 13, 16 and 20 were selected for the *no_logs* subdataset. From the different curves, we can clearly notice that personas belonging to the same profile group do not have the same prediction accuracy. Differently from the results shown in Figure 3, even at the end of the learning period, the accuracy curves do not converge towards a same value for all the personas.

Earliness and Stability analysis The Table 3 shows the results of earliness and stability metrics of the test dataset, *logs* and *no_logs* subdatasets. We can notice that both *logs* and *no_logs* have different values for the earliness. Furthermore, we can see from this table that the stability performances of the system

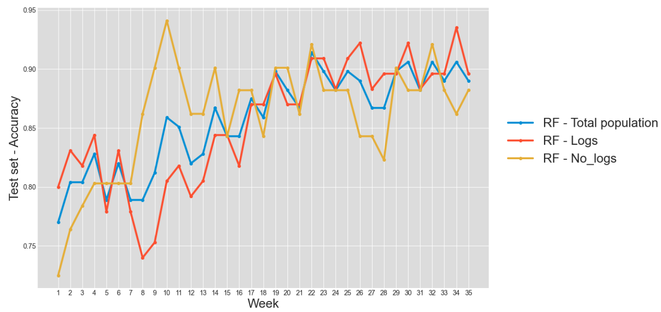

Fig. 3. Accuracy evaluation with total population and the two subdatasets (*logs, no_logs*).

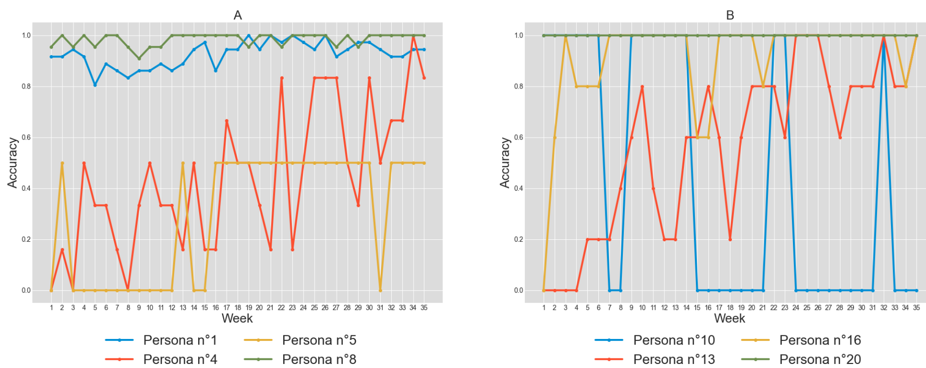

Fig. 4. Accuracy evolution of personas of *logs* (A) and *no_logs* (B) subdatasets.

are different from one profile group to another. In addition, whatever the subdataset is, the algorithm has the same stability and earliness performance for each class. Thus, the dropout class has always the best metrics values, whereas the medium risk class has the worst results.

The Table 4 shows the results of the earliness and stability metrics of each persona belonging to *logs* or *no_logs* subdatasets. We notice that the measures are different from one persona to another. In addition, the difference is even more tangible when it comes to the personas of medium *(C_2, in pink in Table 4)* and high risk learners *(C_3, in yellow in Table 4)*. Due to the lack of pages, we cannot report all the results: we only present a selection in order to illustrate the kind of results that we can provide with the presented methodology.

5.3 Discussion

The previous tables and figures showed that the prediction algorithm out performs globally (up to 93% of accuracy). However, the prediction algorithm doesn't exhibit the same performance with each learner profile. For example,

	Total		*logs*		*no_logs*	
	Earliness	Stability	Earliness	Stability	Earliness	Stability
Dropout	1.01	31.44	1.03	31	1	31.88
High Risk	3	16.55	1	7.5	3.57	19.14
Medium Risk	8.06	6.62	6.77	4.33	8.28	9.57
Success	1.1	28.38	1.12	28.69	1	28.6
Total	2.06	25.35	1.75	26.10	2.35	26.43

Table 3. Earliness and Stability measurement of each class of a profile group.

Subdataset	Persona	Earliness	Stability
logs	1	1.13	28.69
	2	1	26
	3	12.66	1.66
	4	5.5	5.66
	5	1	7.5
	7	1	25.2
	8	1.04	32.31

C_1
C_2
C_3
C_4

Subdataset	Persona	Earliness	Stability
no_logs	9	1	35
	10	1	6
	11	1	35
	12	5	12
	13	8.2	10.8
	14	12	1
	15	6.5	8.5
	16	24	23.2
	18	1	31.6
	19	1	35
	20	1	35
	21	1	33.16

Table 4. Earliness and stability for each persona.

the successful learners and those who dropout are much better predicted than those who are at-risk of failure. In addition, learners who belong to the *log* group are also more accurately predicted. Earliness and stability results show that the algorithm performance is dependent on the learners profiles. In order to provide education stakeholders with accurate and reliable results over time, the prediction system has to take into consideration the different learning profiles existing within a cohort.

6 Conclusion and perspectives

The identified learners profiles, characterized by personas, within our dataset were diversified and confirmed that learners adopt different behaviors and must receive an adapted support. In addition, the prediction model evaluation reveals that the algorithm's performances were not the same for all personas and classes. The obtained results answer our research question and confirm the interest of personas in LA tools assessment. Furthermore, indicators such as earliness and stability, which have been introduced, give information about the confidence that a user can have in the system. Indeed, the usual accuracy metrics are insufficient to evaluate the weekly results of an educational prediction system. It's a reason why, we plan to investigate several research directions relying either on personas or on new refinements in LA assessment. First, we believe that the personas identified in year N could also be used as a basis for evaluating classes in year N+1, assuming that the behaviors observed from one year to the next are similar. This research context deserves attention because it would help to provide quick feedbacks for teachers about their learners' situations. This early information could help them to promptly develop solutions for students considered at risk. Secondly, we wonder how much the separation of the initial dataset according to learning modalities (*logs*, *no_logs*) and classes (C_1, C_2, C_3, C_4) influences the performance of the learning systems, and particularly of the prediction system in our case study. Therefore, it seems interesting to compare the results with different partitions of the dataset. In one hand, it could allow to highlight the key features which are essential to the good functioning of the model. In another hand, this would further improve the explainability by allowing teachers and academics to select the appropriate partition according to their pedagogical objectives. Finally, both indicators presented (earliness and stability) provided additional information about systems' behavior. In that way, a further work on these indicators and especially on their generalization seems to be necessary, so that they can be used in a wider range of areas.

References

1. Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A.A., Abid, M., Bashir, M., Khan, S.U.: Predicting at-risk students at different percentages of course length for early intervention using machine learning models. IEEE Access **9**, 7519–7539 (2021)
2. Amal Ben Soussia, A.R., Boyer, A.: An in-depth methodology to predict at-risk learners. In: European Conference on Technology Enhanced Learning. pp. 193–206. Springer (2021)
3. Anjeela Jokhan, B.S., Singh, S.: Early warning system as a predictor for student performance in higher education blended courses. STUDIES IN HIGHER EDUCATION **Vo.44, Issue 11**, pp.1900–1911 (May 2018)
4. Arnold, K.E., Pistilli, M.D.: Course signals at purdue: Using learning analytics to increase student success. In: Proceedings of the 2nd international conference on learning analytics and knowledge. pp. 267–270 (2012)
5. Bañeres, D., Rodríguez, M.E., Guerrero-Roldán, A.E., Karadeniz, A.: An early warning system to detect at-risk students in online higher education. Applied Sciences **10**(13), 4427 (2020)
6. Ben soussia, A., Labba, C., Roussanaly, A., Boyer, A.: Assess performance prediction systems: Beyond precision indicators. CSEDU (2022)
7. Boroujeni, M.S., Sharma, K., Kidziński, Ł., Lucignano, L., Dillenbourg, P.: How to quantify student's regularity? In: European conference on technology enhanced learning. pp. 277–291. Springer (2016)
8. Brooks, C., Greer, J.: Explaining predictive models to learning specialists using personas. In: Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14. pp. 26–30. ACM Press (2014). https://doi.org/10.1145/2567574.2567612, http://dl.acm.org/citation.cfm?doid=2567574.2567612
9. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence **PAMI-1**(2), 224–227 (1979)

10. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., Santos, O.C., Rodrigo, M.T., Cukurova, M., Bittencourt, I.I., et al.: Ethics of ai in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education pp. 1–23 (2021)
11. Hussain, M., Zhu, W., Zhang, W., Abidi, S.M.R.: Student engagement predictions in an e-learning system and their impact on student course assessment scores. Computational intelligence and neuroscience (2018)
12. Lee, S., Chung, J.Y.: The machine learning-based dropout early warning system for improving the performance of dropout prediction. Applied Sciences **9**(15), 3093 (2019)
13. Likas, A., Vlassis, N., J. Verbeek, J.: The global k-means clustering algorithm. Pattern Recognition pp. 451–461 (2003). https://doi.org/10.1016/S0031-3203(02)00060-2
14. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation Forest. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413–422 (2008). https://doi.org/10.1109/ICDM.2008.17, iSSN: 2374-8486
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research pp. 2825–2830 (2011)
16. Philipp, M., Rusch, T., Hornik, K., Strobl, C.: Measuring the stability of results from supervised statistical learning. Journal of Computational and Graphical Statistics **27**(4), 685–700 (2018)
17. Pluck, G., Johnson, H.: Stimulating curiosity to enhance learning. GESJ: Education Sciences and Psychology (2011)
18. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics pp. 53–65 (1987). https://doi.org/10.1016/0377-0427(87)90125-7
19. Siemens, G., Long, P.: Penetrating the fog: Analytics in learning and education. EDUCAUSE review p. 30 (2011)
20. Slade, S., Prinsloo, P.: Learning Analytics: Ethical Issues and Dilemmas. American Behavioral Scientist **57**(10), 1510–1529 (2013). https://doi.org/10.1177/0002764213479366
21. Teinemaa, I., Dumas, M., Leontjeva, A., Maggi, F.M.: Temporal stability in predictive process monitoring. Data Mining and Knowledge Discovery **32**(5), 1306–1338 (2018)
22. Treuillier, C., Boyer, A.: Identification of class-representative learner personas. Learning Analytics for Smart Learning Environments (LA4LSE) Workshop, at EC-TEL'21 (2021)
23. Xu, D., Jaggars, S.S.: Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. The Journal of Higher Education **85**(5), 633–659 (2014). https://doi.org/10.1080/00221546.2014.11777343
24. Zhang, N., Biswas, G., Dong, Y.: Characterizing students' learning behaviors using unsupervised learning methods. In: International conference on artificial intelligence in education. pp. 430–441. Springer (2017)