# A New Pre-Processing Approach Based on Clustering Users Traces According to their Learning Styles in Moodle LMS

https://doi.org/10.3991/ijet.v18i07.37635

Mohammed Aitdaoud[1,2(✉)], Abdelwahed Namir[1], Mohammed Talbi[2]
[1]LTIM, Department of Computer Science, Faculty of Sciences Ben M'sick, Hassan 2 University of Casablanca, Casablanca, Morocco
[2]ORDIPU, Faculty of Sciences Ben M'sick, Hassan 2 University of Casablanca, Casablanca, Morocco
mohammed.aitdaoud@univh2c.ma

**Abstract—**Nowadays, many Moroccan universities and institutions start offering training and online courses "E-learning". Which accumulate a vast amount of information that is very valuable for analyzing students' behavior and could create a gold mine of educational data. However, handling the vast quantities of data generated daily by the learning management systems (LMS) such as Moodle has become more and more complicated. This massive data can be used to improve decision making and management, which requires a proper extracting and cleaning methods. The purpose of this paper is to suggest a new approach for the preprocessing of the execution traces generated during the interaction of learners with the Moodle LMS and especially the educational content in SCORM format. In this study, we built two experimental corpus with the learning platform Moodle. Using the data generated by the experimental corpus, we applied the Clustering data mining technique as a preprocessing step in our process discovery. Hence, students with similar learning styles or performance levels are grouped together which should help us to build a partial process model (learning process) that is easier to understand for the decision makers.

**Keywords—**educational data mining, clustering, learning styles, higher education, technology enhanced learning, learning management system

## 1 Introduction

In recent years, many educational institutions (universities; private schools;…) has been trying to establish an online presence in teaching and learning, and have begun to offer online courses to their students. Some of these courses are combined with traditional education in what is known as blended learning, while others are conducted entirely online. Hence, Online courses require an environment where they are managed and organized, which in the majority of cases, this task is fulfilled by a Learning Management System also known as LMS. Those latter offer a variety of features to help teachers create, administer, and manage online courses [1]. However, they generally do not take into account the individual differences of learners and treat all learners equally, regardless of their needs and personal characteristics.

Learners play a central role in both traditional learning and online learning. Each learner has individual needs and characteristics such as prior knowledge, cognitive abilities, learning styles, and so on. These individual differences influence the learning process and are the reason why some learners find it easy to learn in a particular course, while others find the same course difficult [2]. E-learning courses usually overlook the learners' individual distinctions and treat them all the same, regardless of their requirements or personal features [3].

Tailoring the educational experience to student preferences and needs is an important goal of Technology Enhanced Learning (TEL). Learning style is one of the individual characteristics that is taken into account, as it encompasses the strategies and preferences used by a student to approach the learning process [4]. For example, personalizing content into students' learning styles has been found to benefit learning in several ways, such as improving satisfaction [5], increasing learning outcomes [6] and reducing the time needed to learn [7].

From a theoretical point of view, integrating students' learning styles into TEL makes learning easier and increases learning efficiency. On the other hand, learners whose learning styles are not supported by the learning environment may encounter problems in the learning process [8].

With the enormous growth of e-learning, several works are currently being done on TEL with the aim of improving the quality of distance education. They are designed to promote learning, that is to say the construction of knowledge in a learner [9].

One of the important features of e-learning systems is the large amount of educational data generated. This problem of information overload highlights the potential danger of getting lost in data. Having the right information at the right time is crucial. Therefore, it is important to identify the methods and models, which can extract reliable and comprehensive knowledge.

In this article, we propose a new approach for the pre-processing of learner traces in the Moodle LMS, which is our main source of data combined with the learning style category. These data will be useful for us to better understand the behavior of learners and the nature of the traces they generate, so that we can transform them and aggregate them to produce indicators of student behavior.

This paper is organized as follows: Section 2 provides background information regarding educational data mining and its related works. Section 3 define our research process, explain the dataset and describes our proposed approach. Section 4 reports the experimental results obtained. Finally, in section 5 we conclude with a summary and outlook for some future work.

## 2 Related works

During the last decade, several works have been done on TEL with the aim of improving the quality of e-learning through the collection and analysis of the LMS data in order to give the users good insight about the interaction traces of learners with the learning system. Thus, several techniques and methods have been used for the traces analysis to create adaptive e-learning systems, such as: Ontology [10], Cloud Computing [11], Trace-Based System [12] and also Data Mining [13].

Researchers have recently started to explore the possibilities of applying data mining techniques to educational data. This research was primarily designed to help educators and tutors better understand their students' behavior and learning style by enabling them to assess student performance and track their learning pathways [14], [15].

Learning styles are increasingly integrated into TEL and a lot of research work is being done in this area [16]. The field of learning styles is complex and has been affected by several aspects that lead to different concepts and viewpoints. There are many learning style models that exist in the literature [17], each offering different descriptions and classifications of learning types [18], [19]. Moreover, several researches have been carried out over the last 50 years on different aspects of these learning style models. To date, no single unique definition of the term learning style has been identified. Hence, Felder [20] defined learning styles as "characteristic strengths and preferences in the way learners take in and process information".

FSLSM was chosen for this study due to its widespread use in the current discourse of e-Learning personalization [2], [21] and its relevance in categorizing diverse learning styles in an e-Learning system [22], [23].

Today's education systems attempt to provide a personalized learning method by building a model of an individual's goals, preferences and knowledge. Data Mining in education systems can be seen as an iterative cycle of training, testing and refinement. [24] introduce us to the application cycle of Data Mining in education systems, and they describe how data, from traditional classrooms and web-based education systems, can be used to extract knowledge. from databases (KDD). EDM can be used to learn the model of the learning process [25] or student modeling [26] by applying data mining techniques that further help educators to make decisions, such as clustering which has been used in several ways, to:

- identify leadership models among students working in teams and rank the behaviors associated with positive and negative outcomes [27];
- group students according to their mistakes [28];
- discover and regroup students in difficulty [29];
- group students according to the accuracy of their answer [30];
- find groups of students who have similar learning characteristics;
- encourage collaborative group learning [31].

That said, clustering can also be done according to various characteristics such as learner interaction patterns, the amount of access to educational content or the type of problems learners face [32], [33].

In our research, we are interested in extracting knowledge from web-based educational data, also known as web mining, to analyse learners similar behaviour through browsing the e-learning website using Data Mining techniques.

# 3 Methods

## 3.1 Data mining process

EDM uses an iterative process of knowledge discovery (see Figure 1). The process begins and takes place in educational environments, such as e-learning systems and web-based adaptive education systems. These educational environments generate large amounts of raw data, including student usage and interaction data, course information, and academic data.

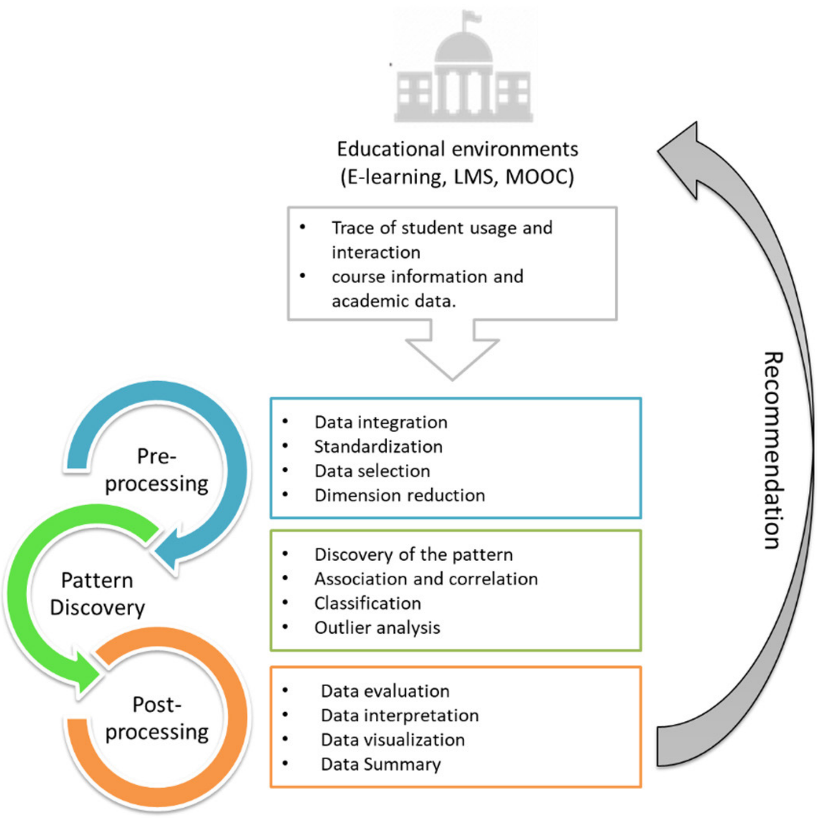



**Fig. 1.** Data mining process

The DM process converts raw data from education systems into useful information that could have a significant impact on educational research and practice. This process is not much different from other fields of data mining application like business, genetics, medicine, etc. because it follows the same steps as the general Knowledge discovery process [34], which is: pre-processing, pattern discovery and post-processing.

1. Pre-processing: it consists of converting the execution traces contained in the source data into the abstraction data necessary for the discovery of models.

2. Pattern discovery: is the step in selecting a paradigm for extracting knowledge such as statistical techniques, Data Mining, and machine learning.
3. Post-processing: In this stage, the knowledge extracted in the previous stage (Pattern discovery) could be visualized and summarized later so that they become simpler [35].

Pre-processing, Pattern discovery as well as Post-processing complete the EDM process as such. The EDM process is not as simple as we have explained. Each step of the EDM process faces a number of challenges and issues in the real world scenario and extracts potentially useful information.

### 3.2 Data collection

The experiments were carried out with the students of the ISIF Master of the Faculty of Sciences Ben M'Sik-Hassan II University of Casablanca. We have established the "Management of Information System" (MSI) course of this training into a SCORM educational content (Moodle) that has been offered in this platform. This course is organized using Multi SCO terminology [36]. Its content is structured in three chapters, each of which is divided into 3 sections and at the end of each section, we find one or more exercises (formative assessment). Each chapter ends, finally, with an evaluation (summative assessment).

The experiments were carried out during the third semester of the second year of the Master. Our experimental analysis corpus contains 27 students, including 14 males and 9 females, 4 of whom did not participate in the online training, as it is complementary and not compulsory.

The platform logs, which we collected after this period, are 2130 entries. After the selection and cleaning steps, we obtained 1239 valid entries as a CSV file.

### 3.3 Attribute selection

In order to better understand what is happening in our learning process and to more effectively assess our learning situation, we apply the analysis process of interaction traces to derive all kinds of indicators. As [37] already explained, trace analysis allows the control and regulation of learning activity.

The Moodle platform has a modest log visualization system. Log files can be filtered by course, participant, day and activity and can be viewed or saved to files in the following formats: text format (TXT), open document format for office applications (q) or Microsoft Excel file (XLS). We therefore proceeded to extract the attributes directly from the database.

After a first pre-processing phase, the data is structured and grouped together in the form of a list of indicators that will allow us to compare the students with each other more easily. Aggregate variables can be grouped into four main types:

1. variables related to working time,
2. the variables linked to the user (description, notes, interactions),
3. the variables linked to the work session,
4. variables related to learning styles.

The indicators that we have calculated and that are of interest to the teacher, were aggregated from student interaction traces via SQL queries using the open source data integration tool Talend Open Studio and by creating Jobs that allow us the aggregation of several variables. Table 1 gives us an example of the variables related to working time.

**Table 1.** Variables related to working time

| Attribute | Type | Description | Moodle Extraction Method and Nomenclature |
|---|---|---|---|
| Time | Data time | Periods between view of resources and next different action | For an action i ($act_i$) and i ∈ N<br>$\Delta t = \text{time}(act_{i+1}) - \text{time}(act_i)$ |
| Dur_course | Second | Sum of user interaction periods in a course | For a course C of i actions and Δt(i) is the duration of an action i for the same user U<br>$\text{Duration(C)} = \sum_{i=1}^{i} \Delta t(act_i)$ |
| Dur_day | String | The time of day the session takes place | For an action i ($act_i$)<br>if (time ($act_i$)) > 03) Dur_day = "morning"<br>if (time ($act_i$)) > 12) Dur_day = "afternoon"<br>if (time ($act_i$)) > 20) Dur_day = "night" |
| Dur_week | String | The time of the week the session takes place | For an action i ($act_i$)<br>No(time($act_i$)) : Returns the week number<br>if (No(time ($act_i$)) =6 or No(time ($act_i$)) =7)<br>Dur_day = "weekend"<br>else<br>Dur_day = "business day" |

To perform our analysis, in addition to the preprocessed data, we used attributes that describe the personal characteristics, results, and interactions of different learners. This information is collected from the student database and integrated into our database (see Table 2).

**Table 2.** Variables related to the user

| Attribute | Type | Description | Moodle Extraction Method and Nomenclature |
|---|---|---|---|
| Gender | String | The gender of the user | Gender : Female or Male |
| Grade | Number (float) | The learner's grade in the course | Direct extraction from the platform database |
| Ch(**n**^ième^)_grade | Number (float) | The learner's mark in the nth chapter of the course | Direct extraction from the platform database |
| Total | Number (float) | The learner's final grade in the module | Extraction from the APOGEE database |
| Action | Number (integer) | The number of user actions in a course | For a course C of i actions and (Count(i) is the number of actions For the same user U<br>Action (U) = Count($act_i$) |
| Chat | Boolean | Chat participation | 0 : false \| 1 : true |
| Forum | Boolean | Forum participation | 0 : false \| 1 : true |

In the following Table 3, we present statistics on user sessions in aggregate way.

**Table 3.** Variables related to the work session

| Attribute | Type | Description | Moodle Extraction Method and Nomenclature |
|---|---|---|---|
| Nb_session | Number | The number of a user's sessions | $\text{Nb\_session} = \sum_{i=1}^{i} S(U)$ |
| Dur_moy_ss | Second | The overall duration of a session | For a session S of i actions and Δt(i) is the duration of an action i $\text{Dur\_moy\_ss(S)} = \sum_{i=1}^{i} \Delta t(act_i)$ |
| Nb_act_moy_ss | Number | The average number of user session actions | Count ($act_i$) where $act_i$ {i€ [1…i]} is the set of actions of the session S |

The last type of variables, which we will present and those related to the learning styles of the students, will allow us to identify how a student likes to learn using the ILS questionnaire.

**Table 4.** Variables related to learning styles

| Attribute | Type | Description | Moodle Extraction Method and Nomenclature |
|---|---|---|---|
| Active | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Reflective | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Sensitive | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Intuitive | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Visual | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Verbal | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Sequential | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |
| Global | Number | [1: balanced, 2: moderate, 3: strong] | Using the ILS Questionnaire |

For the various attribute selection operations presented in this section, we used the Talend Open studio tool. The latter allowed us to create the Jobs necessary to calculate our indicators.

### 3.4 Clustering approach

Clustering or grouping, also called unsupervised classification or segmentation, is a statistical methodology that groups similar objects into clusters or segments. This procedure groups similar objects into clusters (homogeneous packets). In our process of analyzing execution traces, we propose an approach to use clustering as a processing task to improve the extraction process and, at the same time, optimize both performance/finesse and comprehensibility/size of the model obtained. The traditional approach uses all of the data in the log to uncover a model of a student's learning process, which can generate models that are too large and complex to be used or analyzed by an instructor. These patterns are generally referred to as spaghetti patterns [38].

Our approach is to first apply clustering as a pre-processing step for process mining. By grouping students with similar learning styles or performance levels, we will be able to build partial process models that are easier to understand. The proposed approch (see Figure 2) uses the clustering technique as a processing step according to two methods, the first according to the performance and the second according to the learning style:

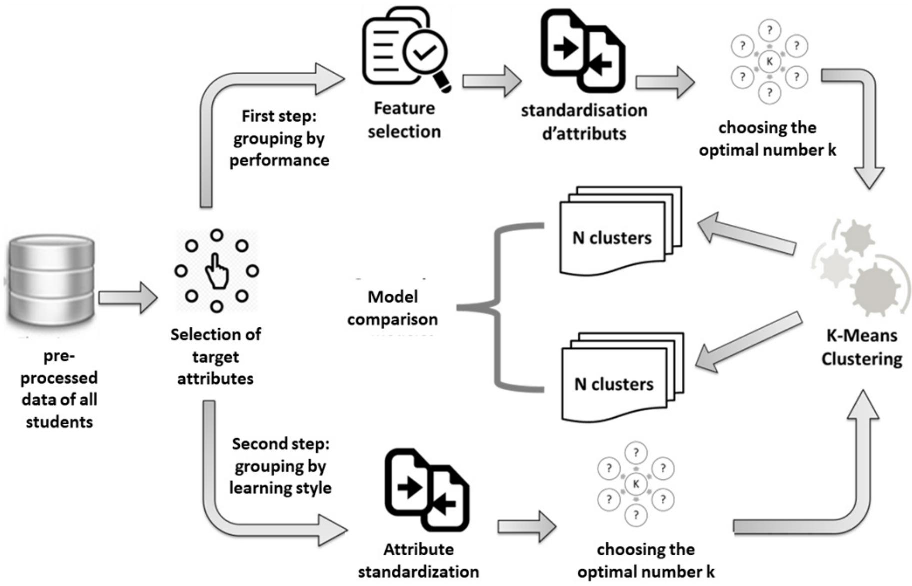



**Fig. 2.** Representation of the proposed clustering approach

We used K-Means as the clustering algorithm for our experiment, because it is easier to understand graphically and statistically, since each cluster is represented by its center of gravity calculated through the average of all of its points. As we mentioned earlier, in most clustering algorithms and in particular in the K-Means algorithm, the selection of the number of clusters is very subjective. This is because clustering is an exploratory analysis method, aimed at uncovering new interesting patterns and relationships, and there are no criteria to optimize and use to compare multiple alternatives.

As such, there are different ways to select the number of clusters (manual or automatic). We have chosen to use the EM (Expectation-Maximization) algorithm which automatically calculates the optimal number of clusters via a particular heuristic or metric. The EM algorithm offers a probabilistic distribution allowing it to be used as a criterion for measuring similarity to characterize the data and determine the number of K clusters.

Before applying our clustering algorithms, we performed the standardization or normalization of the “Feature Scaling” variables. Clustering attributes must be at measurable scales, otherwise, the attribute with a higher value will dominate the distance calculation. Thus, standardization is a very important step commonly performed in cluster analysis.

Many algorithms allow variables to be automatically standardized at similar scales. We therefore used one of the common methods which is the standard score. This method consists of calculating the Z-score of each value in the field. It indicates how much standard deviation is above or below an average value of a certain score [39]. The Z-score converts the score so that the mean is 0.0 and the standard deviation is 1.0, and its formula is:

$$\dot{x} = \frac{x - \bar{x}}{\sigma} \qquad \begin{array}{l} \bar{x} : \text{mean} \\ \sigma : \text{standard deviation} \end{array}$$

Finally, after applying distance-based clustering via the K-Means algorithm, we compared the two models generated in each of the clustering methods.

## 4 Experimental results and discussion

### 4.1 First method: Clustering approach application according to performance

The first method consists in creating clusters of similar student profiles according to the performance indicators of the training path calculated in section (3.3). The first criterion corresponds to the number of student actions in the course, the second represents the number of sessions per user, the third represents the length of time the student has spent in the course, the fourth concerns the grade obtained in the global module, and finally, the fifth concerns the grade obtained in the course. Note that the last three represent the grades of the three chapters of the course (Ch1_grade, Ch2_grade, Ch3_grade) as presented in Table 5.

**Table 5.** Variables selected for clustering analysis (performance)

| Attribute | Type | Description |
|---|---|---|
| Action | Number | The number of user actions in a course |
| Nb_session | Number | The number of user sessions |
| Dur_course | Second | The overall time spent in the course |
| Total | Numerical | Overall student grade |
| Grade | Numerical | Overall score of the student in the platform |
| Ch1_grade | Numerical | Student grade in the 1st chapter of the course |
| Ch2_grade | Numerical | Student’s grade in the 2nd chapter of the course |
| Ch3_grade | Numerical | Student’s grade in the 3rd chapter of the course |

In the clustering process [40], the Feature Selection phase is a very subtle, but important where we can improve the performance of machine learning algorithms by allowing these algorithms to focus on the features that actually contain the most predictive power.

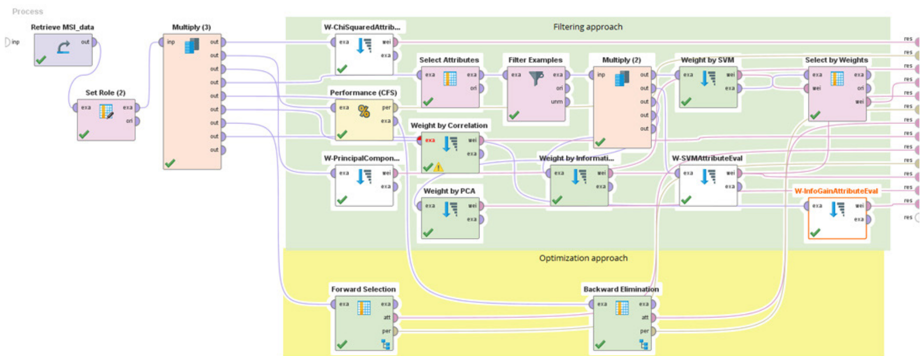



**Fig. 3.** RapidMiner process for feature selection

In this step, we applied a set of selection algorithms in order to have the most significant attributes for the clustering analysis and to define the most discriminating factors of the learner profiles. Two variable selection approaches were used to identify the attributes that could be the most descriptive of a learner profile (see Figure 3). The first is the filtering approach, which is independent of the learning task and which is essentially based on a weighting scheme. The second is the performance optimization approach which is a more integrated optimization approach.

The RapidMiner software offers a multitude of features and algorithms for attribute selection and classification and integrates several Weka algorithms. To investigate attribute weights and their importance in determining learner profiles, we used:

- the attribute weighting techniques: W-ChiSquaredAttributeEval, W-Principal Components, W-SVMAttributeEval, Weight by Correlation, Weight by SVM, Weight by Informatin Gain, Weight by PCA, W-InfoGainAttributeEval;
- and performance optimization techniques: Forward Selection, Backward Elimination.

We then compared the similar attributes most cited by the algorithms of each approach, then from these attributes, we selected those that were chosen by at least four of these algorithms.

After the selection of variables and their standardization, we analyzed these attributes by applying Expectation and Maximization algorithm, then we started this analysis with K-Means. Thus, applying the EM algorithm allowed us to estimate K to be 2.

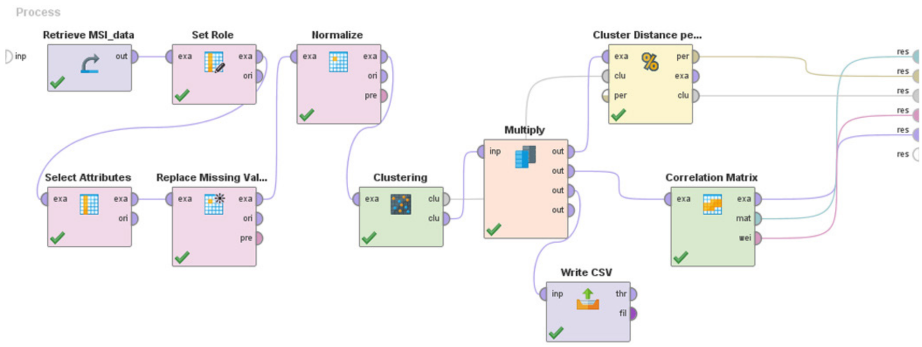
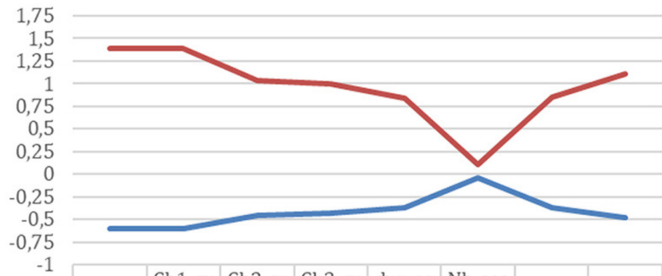



**Fig. 4.** RapidMiner process for clustering via K-Means

In addition to this experiment, we have confirmed this proposition (k = 2) through the use of two algorithms which are based on a visual representation to select the optimal number of clusters (see Figure 4): the dendrogram of the hierarchical cluster and the graphic silhouette analysis [41].



| | Grade | Ch1_grade | Ch2_grade | Ch3_grade | dur_course | Nb_session | Action | Total |
|---|---|---|---|---|---|---|---|---|
| Cluster_0 | -0,605 | -0,605 | -0,454 | -0,437 | -0,368 | -0,046 | -0,37 | -0,483 |
| Cluster_1 | 1,384 | 1,383 | 1,037 | 0,999 | 0,841 | 0,106 | 0,846 | 1,104 |

**Fig. 5.** Table and view of centroids for clustering via K-Means (Performance)

Clustering algorithms provide a highly interpretable result model of the average of the values of each centroid in the cluster (Figure 5). The centroid represents the most typical student or prototype of a group, and it does not necessarily describe a given case in that group.

We can conclude, from Figure 5, that our learners are distributed according to this set of criteria into two groups (clusters):

1. Cluster_P0 (cluster_0): 16 students (13 passed and 3 failed)
2. Cluster_P1 (cluster_1): 7 students (all 7 passed)

As shown by the average values of the different variables, the majority of students passed the module. The first group (in which most of the students were successful) spent less time in following the different parts of the course that they consulted, this minimum duration allows us to say that this group is less engaged in the follow-up of this content, and by extension, it is less efficient.

The second group (in which all the students passed) obtained higher values than the first group in terms of time (Dur_course) and interaction (Action). In addition, the number of sessions (NB_session) and the score (Grade) of this group are higher compared to the other group (Cluster_0). Thus, we can conclude that this group has a fairly high motivation for follow-up (working time could be a measure of learner attention or engagement) [42]. In addition, this motivation enabled these learners to follow the content better and to perform well.

### 4.2 Second method: Clustering approach according to learning style

This method consists of creating clusters of similar student profiles according to the learning style indicators of each student calculated in section (3.3). We have chosen a set of eight numeric attributes that represent the four dimensions of the FSLSM learning style (see Table 4).

After selecting the attributes, we present the results of applying K-Means clustering on the data that describes the learner's learning style. Then, we make a comparison between these results and those obtained in the first method (cf. section 3.4).

| Attribute | cluster_0 | cluster_1 |
|---|---|---|
| Active | 0.205 | -0.319 |
| Reflective | -0.176 | 0.273 |
| Sensitive | -0.481 | 0.748 |
| Intuitive | 0.361 | -0.562 |
| Visual | 0.652 | -1.014 |
| Verbal | -0.701 | 1.091 |
| Sequential | -0.290 | 0.451 |
| Global | 0.350 | -0.545 |

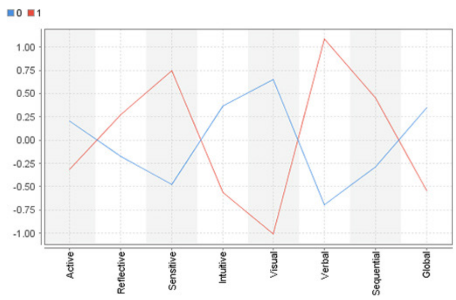



**Fig. 6.** Table and view of centroids for clustering via K-Means (Learning style)

Applying the K-Means algorithm on the attributes of the learning style generates two clusters with the following characteristics (Figure 6):

1. Cluster_S0 (cluster_0): 14 students (12 passed and 2 failed)
2. Cluster_S1 (cluster_1): 9 students (8 passed and 1 failed)

Compared to the results of the first method, the first cluster (Cluster_0) keeps roughly the same structure and composition as those found previously in the first method. However, the cluster lost three learners who join the second group. In addition, we notice that the learners grouped in this cluster score higher values compared to the centroids of the attributes of the learning styles (active, intuitive, visual and global).

The second cluster contains the same seven learners of the cluster obtained from the first method, in addition to 3 other learners who are added to the group (Cluster_1). Thus, a learner of the last three who failed is added to the group. As a result, we have deduced that this group is no longer homogeneous with respect to the performance criterion. The students in this cluster are characterized by high values relative to the centroids of the attributes of learning styles (reflective, sensitive, verbal and sequential).

According to FSLSM, each learner has a preference for each of its four dimensions (active/reflective, sensitive/intuitive, visual/verbal, sequential/global), in fact:

- Active learners prefer to learn in groups and trying things out manually. In contrast, reflective learners prefer to learn individually and by reflecting on the learning material;
- Sensitive learners prefer to learn concrete materials such as data and facts. They also prefer solving practical exercises. On the other hand, intuitive learners like challenges and theoretical concepts;
- Visual learners remember what they saw better through pictures or videos, while verbal learners are more effective when the lesson is presented in text form.
- Sequential learners prefer to learn by linear methods with a linear increase in complexity. On the other hand, global learners tend to perform less well when using partial knowledge.

As we mentioned before, our SCORM course consists of three chapters, accompanied by exercises and a summative assessment at the end of each chapter. Said chapters are structured in a linear form and are based primarily on text. Thus, we can see that the course structure favors the learners with the preferences (reflective, sensitive, verbal and sequential) of the FSLSM.

According to the results obtained, we notice that the models, generated in the first and the second method, are more or less similar and are made up almost of the same students. The students, composing the second cluster (cluster_1) in both models, showed a strategic achievement or approach based on the linearity of course-related actions, and had a good score in the online course. In contrast, the students belonging to cluster_0 were likely to adopt a less structured approach and they are less engaged in the course.

Finally, we can see that the linear structure and the textual form of the course play an important role in the score and motivation of the learners. Indeed, according to [43], the learner's performance is influenced by his learning style, which explains the similarity of the models of our two grouping methods.

## 5 Conclusion

Through this study, our objective was to observe the relevance and the opportunities of an approach centered on the cluster analysis of the learning processes of the students according to their styles and traces, whose final goal is the personalization of the e-learning systems. This study presents the basic work of our research project by providing an overview of our approach to data preprocessing. To that end, we explored the indicators generated from our two case studies using descriptive and inferential statistical methods. However, it should be noted that our first sample does not seem large enough. Thus, we exposed an innovative approach to data preprocessing based on one of the Data Mining methods "Clustering" in order to extract, transform and load the execution traces of learners from their interactions in the e-learning platform Moodle.

The next step of our project aims at exploiting the results of our approach. In particular, our main objective is to propose a Process Mining analysis scenario of a SCORM course in order to provide the users/actors of the platform with information on the actual course of learning and to detect deviations as well as bottlenecks in the learning process, which will facilitate decision-making about the effectiveness of the design and organization of the content offered.

## 6 References

[1] M. J. Rosenberg, *E-learning: Strategies for delivering knowledge in the digital age*, vol. 9. McGraw-Hill New York, 2001.

[2] S. Graf, *Adaptivity in learning management systems focussing on learning styles*. na, 2007.

[3] S. L. Hoskins and J. C. Van Hooff, "Motivation and ability: Which students use online learning and what influence does it have on their achievement?", *British Journal of Educational Technology*, vol. 36, no. 2, p. 177–192, 2005. https://doi.org/10.1111/j.1467-8535.2005.00451.x

[4] R. M. Felder, L. K. Silverman, and others, "Learning and teaching styles in engineering education", *Engineering Education*, vol. 78, no. 7, p. 674–681, 1988.

[5] E. Popescu, "Adaptation provisioning with respect to learning styles in a web-based educational system: An experimental study", *Journal of Computer Assisted Learning*, vol. 26, no. 4, p. 243–257, 2010. https://doi.org/10.1111/j.1365-2729.2010.00364.x

[6] N. Bajraktarevic, W. Hall, and P. Fullick, "Incorporating learning styles in hypermedia environment: Empirical evaluation", in *Proceedings of the workshop on adaptive hypermedia and adaptive web-based systems*, 2003, p. 41–52.

[7] S. Graf, C. H. Lan, T.-C. Liu, and others, "Investigations about the effects and effectiveness of adaptivity for students with different learning styles", in *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on*, 2009, p. 415–419. https://doi.org/10.1109/ICALT.2009.135

[8] N. Bousbia, "Analyse des traces de navigation des apprenants dans un environnement de formation dans une perspective de détection automatique des styles d'apprentissage", Paris 6, 2011.

[9] P. Tchounikine, "Précis de recherche en ingénierie des EIAH", 2009.

[10] B. Bouihi and M. Bahaj, "Ontology and rule-based recommender system for E-learning applications", *International Journal of Emerging Technologies in Learning*, vol. 14, no. 15, 2019. https://doi.org/10.3991/ijet.v14i15.10566

[11] S. Ouahabi, A. Eddaoui, and H. Labriji, "Cloud computing for improving e-learning", *International Journal of Science and Advanced Technology*, 2013.

[12] B. Mohssine, A. Mohammed, N. Abdelwahed, and T. Mohammed, "Adaptive Help System Based on Learners 'Digital Traces' and Learning Styles", *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 10, p. 288–294, 2021. https://doi.org/10.3991/ijet.v16i10.19839

[13] I. Jugo, B. Kovačić, and V. Slavuj, "Increasing the adaptivity of an intelligent tutoring system with educational data mining: A system overview", *International Journal of Emerging Technologies in Learning*, vol. 11, no. 3, 2016. https://doi.org/10.3991/ijet.v11i03.5103
[14] N. Bousbia, "Analyse des traces de navigation des apprenants dans un environnement de formation dans une perspective de détection automatique des styles d'apprentissage", Paris 6, 2011.
[15] M. P. P. Liyanage, L. G. KS, and M. Hirakawa, "Detecting learning styles in learning management systems using data mining", *Journal of Information Processing*, vol. 24, no. 4, p. 740–749, 2016. https://doi.org/10.2197/ipsjjip.24.740
[16] M. Aitdaoud, "Standardized modeling learners to enhance the learning service in the ILE", *New Trends and Issues Proceedings on Humanities and Social Sciences*, vol. 2, no. 11, p. 118–129, 2017. https://doi.org/10.18844/prosoc.v2i11.1913
[17] F. Coffield, D. Moseley, E. Hall, and K. Ecclestone, *Should we be using learning styles? What research has to say to practice*. London: Learning and skills research centre, 2004.
[18] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.
[19] I. B. Myers, "The Myers-Briggs Type Indicator: Manual (1962).", 1962. https://doi.org/10.1037/14404-000
[20] R. M. Felder, "Matters of style", *ASEE Prism*, vol. 6, no. 4, p. 18–23, 1996.
[21] B. Ciloglugil, "Adaptivity based on felder-silverman learning styles model in e-learning systems", in *4th International Symposium on Innovative Technologies in Engineering and Science (ISITES2016) 3–5 Nov 2016 Alanya/Antalya-Turkey*, 2016.
[22] L. Jegatha Deborah, R. Baskaran, and A. Kannan, "Learning styles assessment and theoretical origin in an E-learning scenario: A survey", *Artificial Intelligence Review*, vol. 42, no. 4, p. 801–819, 2014. https://doi.org/10.1007/s10462-012-9344-0
[23] J. Kuljis and F. Liu, "A Comparison of Learning Style Theories on the Suitability for elearning", *Web Technologies, Applications, and Services*, vol. 2005, p. 191–197, 2005.
[24] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005", *Expert Systems with Applications*, vol. 33, no. 1, p. 135–146, 2007. https://doi.org/10.1016/j.eswa.2006.04.005
[25] S. V. Kolekar, R. M. Pai, and M. P. MM, "Prediction of learner's profile based on learning styles in adaptive E-learning system", *International Journal of Emerging Technologies in Learning*, vol. 12, no. 6, 2017. https://doi.org/10.3991/ijet.v12i06.6579
[26] R. R. Maaliw III, "Classification of learning styles in virtual learning environment using data mining: A basis for adaptive course design", *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 7, p. 56–61, 2016.
[27] D. Perera, J. Kay, K. Yacef, and I. Koprinska, "Mining learners' traces from an online collaboration tool", in *Workshop of Educational Data Mining (AIED'07)*, 2007.
[28] A. Merceron and K. Yacef, "Tada-ed for educational data mining", *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, vol. 7, no. 1, p. 267–287, 2005.
[29] M. Pechenizkiy, T. Calders, E. Vasilyeva, and P. De Bra, "Mining the student assessment data: Lessons drawn from a small scale case study", in *Educational Data Mining 2008*, 2008.
[30] T. Y. Tang and G. McCalla, "Smart recommendation for an evolving e-learning system: Architecture and experiment", *International Journal on elearning*, vol. 4, no. 1, p. 105, 2005.
[31] E. Popescu, "Adaptation provisioning with respect to learning styles in a web-based educational system: An experimental study", *Journal of Computer Assisted Learning*, vol. 26, no. 4, p. 243–257, 2010. https://doi.org/10.1111/j.1365-2729.2010.00364.x

[32] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions", *JEDM|Journal of Educational Data Mining*, vol. 1, no. 1, p. 3–17, 2009.
[33] S. Rajper, N. A. Shaikh, Z. A. Shaikh, and G. A. Mallah, "Automatic detection of learning styles on learning management systems using data mining technique", *Indian Journal of Science and Technology*, vol. 9, no. 15, p. 1–5, 2016. https://doi.org/10.17485/ijst/2016/v9i15/85959
[34] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware authors", *User Modeling and User-Adapted Interaction*, vol. 14, no. 5, p. 425–464, 2004. https://doi.org/10.1007/s11257-004-7961-2
[35] R. Tamilselvi, B. Sivasakthi, and R. Kavitha, "An efficient preprocessing and postprocessing techniques in data mining", *Intl. J. Res. Comput. Appl. Rob*, vol. 3, p. 80–85, 2015.
[36] C. Bur, C. Heyvaerts, and S. Poinsart, "Documentation logicielle Opale – SCORM", 1 décembre 2011. [Online]. Disponible sur: https://archives.scenari.org/projects/opale/doc/opale-3.3/userGuide/co/0813-publications-scorm.html. [Consulté le: 30 juillet 2019]
[37] J.-M. Heraud, J.-C. Marty, L. France, and T. Carron, "Helping the interpretation of web logs: Application to learning scenario improvement", in *Proc. Workshop Usage Analysis in Learning Systems*, 2005.
[38] W. M. van der Aalst, "Analyzing 'Spaghetti Processes'", in *Process Mining*, Springer, 2011, p. 301–317. https://doi.org/10.1007/978-3-642-19345-3_12
[39] E. Kreyszig, "Applied mathematics", *Hoboken, NJ: John Wiley & Sons*, 1979.
[40] G. W. Milligan and M. C. Cooper, "Methodology review: Clustering methods", *Applied Psychological Measurement*, vol. 11, no. 4, p. 329–354, 1987. https://doi.org/10.1177/014662168701100401
[41] G. Forney, "Generalized minimum distance decoding", *IEEE Transactions on Information Theory*, vol. 12, no. 2, p. 125–131, 1966. https://doi.org/10.1109/TIT.1966.1053873
[42] M. Cocea and S. Weibelzahl, "Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems", in *International Conference on User Modeling*, 2007, p. 197–206. https://doi.org/10.1007/978-3-540-73078-1_23
[43] S. Graf and K. Kinshuk, "Providing adaptive courses in learning management systems with respect to learning styles", in *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2007, p. 2576–2583.

## 7 Authors

**Mohammed Aitdaoud** is a Ph.D. in Educational Data Analysis from the Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, Morocco. He is currently a professor at the same faculty. Also, he is a member of the Laboratory of Technologies for Information and Modeling (LTIM) and also a member of the Observatory of Research in Didactics and University Pedagogy (ORDIPU). His research interests include educational technologies, distance training engineering, techniques of training, information systems, Educational Data Mining, and Educational Process Mining (email: mohammed.aitdaoud@univh2c.ma).

**Abdelwahed Namir** is a Ph.D. in Numerical Methods for Engineers from Mohammadia School of Engineers (EMI Rabat) Morocco. He is also the Head of the Laboratory of Technological Information and Modelisation (LTIM) Faculty of Sciences Ben M'Sik, University Hassan II – Casablanca, Morocco. His research interests include Mathematics, Computers, and modeling.

**Mohammed Talbi** is a Ph.D. in Sciences and Processes of Analysis from the University Pierre et Marie Curie of Paris. He is currently the Dean of the Faculty of Sciences Ben M'Sik at Hassan II University, B.P 7955 Sidi Othmane, Casablanca, Morocco, and the Director of the Observatory of Research in Didactics and University Pedagogy (ORDIPU) since 2014. He is an Expert in the fields of teaching and research on educational technologies, engineering of distance training, techniques of training, and information systems. He is the author of several national and international awards and has accumulated more than 30 years of scientific productions.

Article submitted 2022-12-24. Resubmitted 2023-02-04. Final acceptance 2023-02-08. Final version published as submitted by the authors.