Explainable AI for Unsupervised Machine Learning: A Proposed Scheme Applied to a Case Study with Science Teachers

Yael Feldman-Maggor[1][a], Tanya Nazaretsky[2][b] and Giora Alexandron[1][c]

[1]*Department of Science Teaching, Weizmann Institute of Science, Rehovot, Israel*

[2]*School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland*

fl

Keywords: Explainable Artificial Intelligence, Clustering, Personalized Learning.

Abstract: Explainable Artificial Intelligence (XAI) seeks to render Artificial Intelligence (AI) models transparent and comprehensible, potentially increasing trust and confidence in AI recommendations. This research explores the realm of XAI within unsupervised educational machine learning, a relatively under-explored topic within Learning Analytics (LA). It introduces an XAI framework designed to elucidate clustering-based personalized recommendations for educators. Our approach involves a two-step validation: computational verification followed by domain-specific evaluation concerning its impact on teachers' AI acceptance. Through interviews with K-12 educators, we identified key themes in teachers' attitudes toward the explanations. The main contribution of this paper is a new XAI scheme for unsupervised educational machine-learning decision-support systems. The second is shedding light on the subjective nature of educators' interpretation of XAI schemes and visualizations.

1 INTRODUCTION

In recent years, the rapid growth of AI has led to its widespread application in various fields, including education. Personalized learning systems, in particular, have gained widespread interest, and mounting evidence suggests that they are highly effective in enhancing learning outcomes (Khosravi et al., 2022). Although AI applications are increasingly used in education, recent research indicates that human factors like trust can negatively influence educators' readiness to embrace AI tools (Cukurova et al., 2020; Nazaretsky et al., 2022a). One reason for this reluctance is that AI is often experienced as a 'black box,' with users not understanding how and why the algorithm reaches specific results (Rudin, 2019). XAI addresses the problem of 'black boxes' by proposing algorithmic approaches aimed at increasing transparency of AI-powered systems by specifying the explanations in terms of factors that influenced the system's decision-making process at varying degrees of detail (e.g., for entire system behavior or for a particular prediction). These explanations are critical in guaranteeing algorithmic fairness, identifying potential issues or biases in the training data, and verifying that the algorithms function as intended (Gilpin et al., 2018). In particular, AI-powered educational technologies are a critical context where XAI is necessary to assist educators and students in comprehending how AI works, determining how it may affect them, assessing its trustworthiness, and ensuring that ethical concerns are properly addressed. Indeed, the XAI field in education has rapidly expanded in the last ten years. However, most XAI methods currently studied in education are designed for *supervised* machine learning (ML), mainly focusing on deep neural network architectures (Fiok et al., 2022; Swamy et al., 2023). In fact, we are familiar with only one educational XAI framework for unsupervised ML, FUMA(Conati et al., 2021). FUMA can identify clusters of user behaviors mapped into different learning outcomes and predict when a new student is not learning well early during the interaction. However, this framework has several shortcomings, including being overly complex (Khosravi et al., 2022), calling for more research on XAI for educational unsupervised ML. We note that this is in line with recent voices outside the LA and AI in education space, which marked XAI for unsupervised ML as a topic that should receive more attention from the AI research community

[a] https://orcid.org/ 0000-0002-0456-6664
[b] https://orcid.org/0000-0003-1343-0627
[c] https://orcid.org/0000-0003-2676-6912

Feldman-Maggor, Y., Nazaretsky, T. and Alexandron, G.
Explainable AI for Unsupervised Machine Learning: A Proposed Scheme Applied to a Case Study with Science Teachers.
DOI: 10.5220/0012687000003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 436-444
ISBN: 978-989-758-697-2; ISSN: 2184-5026

(Morichetta et al., 2019; Kauffmann et al., 2022).

In contrast to the supervised approach, where the models are trained using labeled data and can be evaluated based on their ability to predict a known output, unsupervised methods aim to identify unknown patterns in the data (Kauffmann et al., 2022). Thus, not only is the model decision process difficult to explain but the semantic meaning of the revealed patterns should also be interpreted, posing considerable XAI challenges due to the sophisticated nature of the statistical methods and the complex and abstract relations being revealed. One example is cluster analysis, commonly used for grouping students based on their learning behavior. Clustering algorithms can group data points based on their similarity (Dasgupta et al., 2020). However, the criteria for similarity are often challenging to understand and explain. Cluster analysis is a frequent application of unsupervised ML in education (Marras et al., 2021; Pereira et al., 2020), stressing the need for targeted research on its XAI aspects.

In this research, we aim to take a step toward closing this gap. First, we present a novel XAI scheme to explain the cluster analysis results of student assessment data. Next, we exemplify how this scheme can be applied to a common real-world teaching task. In this task, the teacher's goal is to adjust the instruction to the student's state of conceptual understanding, using an AI-powered tool that visualizes to the teacher the knowledge profiles of the students in her class and maps the students into these profiles. The profiles and the mapping are based on cluster analysis of student responses to multi-dimensional, interactive assessment instruments. The algorithmic process is done under the hood and remains a 'black box' for the teachers, raising the need to explain its decision-making logic in an understandable way for the end-users. To this end, we ask the following research questions (RQs):

RQ1. If and how ML and XAI methods can be employed to create meaningful explanations for automatically detected profiles of student knowledge?

RQ2. Do teachers find the above-mentioned explanations useful for understanding the detected knowledge profiles?

In what follows, we start with a theoretical background. Then, we present the XAI scheme and validate it computationally and qualitatively with teachers.

2 BACKGROUND

XAI is studied in various application domains, each with its own objectives and research traditions, posing challenges for establishing common design and evaluation methodologies (Mohseni et al., 2021). In the context of XAI in education, among the key factors that define the research and design space are the various types of stakeholders (teachers, students, developers, etc.), machinery to be explained (supervised, unsupervised, deep learning, NLP, etc.), XAI frameworks (SHAP, LIME, etc.,), contexts (K12, higher education), and theoretical frameworks (Khosravi et al., 2022; Cukurova et al., 2020; Swamy et al., 2022).

2.1 Explainable AI Terminology and Methods

The ability to correctly interpret and explain an ML model output is essential for both the model developers and its end-users (Mohseni et al., 2021). It engenders appropriate user trust, provides insight into how a model may be improved, and supports understanding of its internal logic (Qin et al., 2020). In general, the concept of explainability in ML covers various aspects that contribute to the transparency and comprehensibility of models. This includes explaining details of the data being used and performance metrics (Liao et al., 2020). One way to classify explanations is as either *local* or *global* (Setzu et al., 2021). Roughly, a global explanation is a type of explanation that describes how the overall ML model works. Model visualization and decision rules are examples of explanations falling into this category (Mohseni et al., 2021). In contrast, local explanations aim to explain the relationship between specific input-output pairs, the reasoning behind the results of an individual user query, or a particular input instance (Mohseni et al., 2021). The local explanation is considered less overwhelming for novices and can be suited for investigating edge cases for the model or debugging data (Mohseni et al., 2021).

Two well-established local explanation frameworks are LIME (Local Interpretable Model-Agnostic Explanations) (Ribeiro et al., 2016) and SHAP (Lundberg et al., 2020). LIME explains the prediction of a certain input by sampling its neighboring inputs and learning a sparse linear model based on the predictions of these neighbors; features with significant coefficients in the linear model are then considered important for that input's prediction (Ribeiro et al., 2016). The original SHAP method and its variation, Kernel SHAP, are based on the game-theoretic concept of Shapley values, which assigns a contribution value to each feature in a prediction, considering all possible combinations. Kernel SHAP (Lundberg et al., 2020) is a model-agnostic method that approximates SHAP values by sampling a subset of the possible combinations of features, and it is computationally faster than the original SHAP and can handle larger datasets and more complex models. Applying this approach, (Lundberg et al., 2020). They suggested SHAP TreeExplainer - an explanation method for tree-based models that enables the tractable computation of optimal local explanations. They proposed to utilize SHAP interaction values as a richer type of local explanation. Thus, TreeExplainer can uncover significant patterns that might otherwise be missed by considering interaction effects.

2.2 XAI for Cluster Analysis in Education

Cluster analysis has been applied in numerous educational studies to forecast the probability of students succeeding, failing, or dropping out of a course based on their academic performance and online behavior (Gabbay and Cohen, 2022; Käser et al., 2013; Klingler et al., 2016; Mojarad et al., 2018), predicting the specific skills and competencies that students have acquired or are yet to develop, and classifying students into groups based on their knowledge profile (Asif et al., 2017; Nazaretsky et al., 2022a; Klingler et al., 2016). However, many clustering algorithms lead to cluster assignments that are difficult to explain. XAI research has started to explore this space (Swamy et al., 2023; Dasgupta et al., 2020). Moreover, as opposed to supervised ML, there are still no standard methods. Since clustering is a common and standard method in educational research, studying and developing XAI for clustering in education is of practical importance and scientific interest.

2.3 Interactive AI-Powered Dashboards

Many institutions have started to adopt AI-powered solutions that help educators make data-driven decisions using LA dashboards (Ahn et al., 2019; Michaeli et al., 2020). Such dashboards can assist teachers in various tasks (Baker, 2016). However, the provided analytics should be understandable and actionable to be effectively used. In the educational context, the information should incorporate the pedagogical logic behind the information presented. As such, it should be properly adapted to the specific learning and teaching goals, the needs of the different end-users (e.g., instructional designers, instructors, and students), and various learning contexts (e.g., individual students, student groups, or the entire class) (Dillenbourg et al., 2011; Nazaretsky et al., 2022a).

In the case of AI-based LA dashboards, it is thus essential to incorporate XAI features that enable the users to understand the pedagogic rationale behind the ML analysis. In order to address the gaps mentioned above, the proposed research aims to propose an XAI scheme for cluster analysis in education, validate it computationally, and experiment with its value through a user study with teachers.

3 THE XAI SCHEME

Rationale. A common goal of cluster analysis in education is classifying the learners into groups with similar knowledge profiles (Gabbay and Cohen, 2022; Klingler et al., 2016) based on their responses to collections of educational items. The results of such clustering analysis are inherently difficult to interpret, thus requiring explicit user-focused explanations. However, there is a lack of effective XAI schemes for cluster analysis (Bandyapadhyay et al., 2023). Feature importance is a key approach for explaining classifiers' output Section 2. In our XAI scheme, we employ the feature importance approach to explain cluster analysis , with the features being *educational items*. So, the important features in our context are educational items that are most significant in distinguishing between resulting clusters (hereafter referred to as *item importance*). Such item importance-based explanations have two main desired properties: first, such item importance-based explanations are interpretable by teachers – the typical end-users of cluster-analysis applications. A working assumption underlying our approach is that the teachers are either familiar with the items or can analyze them to extract the underlying skills. Second, item importance can be computed for the output of a specific cluster analysis – a clustering instance – in an automated fashion.

The Scheme. Our XAI schema consists of three steps. *Step 1* is the cluster analysis into knowledge profiles and *Step 2* and *3* provide the 'explanation layer' for this analysis. Below, we explain each step in detail.

1. Cluster analysis of student response data into clusters, each representing a knowledge profile.
2. Building the item importance-based explanations:
   (a) Building a labeled dataset with the students serving as samples, the items as features, and the student-to-cluster mapping as the labels. So, a sample in this dataset comprises a student's scored response vector (correct/incorrect

score to all educational items) and the student's cluster as the label.

(b) Training an explainable classifier on the data (e.g., *Random Forest*).

3. Analyzing the classifier's feature importance to identify the items discriminating between the clusters.

This scheme results in a set of items and their relative importance to the classification. The proposed scheme should be evaluated concerning two main aspects. First, such item importance-based analysis provides statistically meaningful results. Second, it is useful for teachers. In the following sections, we first apply the proposed XAI scheme to real student data and validate its output quantitatively. Next, we evaluate its usability with teachers who are real users of the learning platform PETEL.

4 RESULTS: APPLYING THE SCHEME TO REAL STUDENT RESPONSE DATA

4.1 Quantitative Evaluation

Below we describe the context, data, results of applying the scheme to these data, and the quantitative analysis of the method's output. This analysis addresses RQ1:

Context, Population and Data. The scheme presented in Section 3 was applied to four high-school-level digital science learning activities in chemistry and physics (two in each topic), each containing 13 to 23 auto-graded, multiple-skill assessment items around a specific curricular topic. The four activities were administered to 216 to 1572 high-school students majoring in physics/chemistry between 2020 and 2023 as part of the regular teaching of the relevant subjects. Response matrices containing student-scored responses to the activities were mined from the learning platform PETEL.

Below we present in detail the results of applying the scheme to one of the chemistry activities dealing with Stoichiometry. This activity included 13 items that appeared in matriculation exams in chemistry.

Step 1. Cluster Analysis of Student Response Data. This stage followed the procedure and ideas presented in (Asif et al., 2017; Nazaretsky et al., 2022b). The proposed clustering method assumes that students tend to perform similarly on items requiring the same skills and competencies. Its aim is to divide the students into clusters (knowledge profiles) based on their responses to an instrument containing a set of 13 auto-graded interactive assessment items (marked as correct/incorrect). When multiple attempts were allowed, the score was computed as a correct-on-first attempt to reduce the possible effect of guessing or gaming the system (Baker et al., 2008; Ruiperez-Valiente et al., 2017). The data was represented as a response matrix, with items as columns and students as rows. The data was pre-processed according to the guidelines recommended in (Feldman-Maggor et al., 2021). Fully empty rows (students who did not submit the activity) were removed from the dataset. For partially incomplete rows, 'Missing Not at Random' was assumed and treated as incorrect response (Nazaretsky et al., 2019).

The clustering was done using a K-means clustering algorithm with Euclidean distance (Jain, 2010), and the optimal number of clusters was determined using the Weighted Gap Statistics method, which suitability for educational data was validated in (Din et al., 2023). Results from this stage were presented on the platform PETEL through an interactive dashboard called Grouper (Nazaretsky et al., 2022b) that enables teachers to observe the clusters, the mapping of students into them (Fig.1) and[1], and the status of each item with respect to each cluster (Fig.2). Demonstrated on the chemistry activity on Stoichiometry, Step 1 identified four clusters. Fig.2 presents the representative response pattern of each profile. For a certain Item-Group cell, Green means that more than 70% of students assigned to the cluster were correct on that item, Red means that less then 50% were correct, and Yellow denotes the between cases.

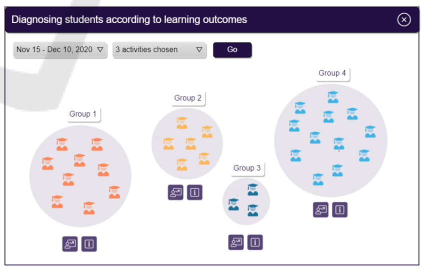

Figure 1: Dashboards for Clusters of Students with Similar Profiles.

Step 2. Building the Item Importance-Based Explanations. In Step 2, a supervised dataset was created and randomly divided into training and test sets (70% and 30%, respectively). At this stage, we trained a Random Forest (RF) algorithm[2] to create predictive models. The decision to use RF was based on the large body of research demonstrating its

[1] Grouper presentation

[2] RandomForest, sklearn python package

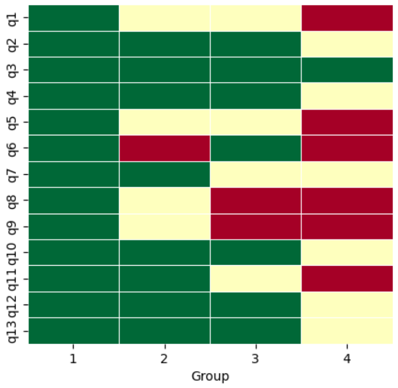

Figure 2: Dashboards Displaying the Typical Response Matrix for Each Cluster.

high accuracy and the usability of its feature importance analysis. 100 RF models were trained, yielding $\mu_{accuracy} = 0.922$ and $sd_{accuracy} = 0.012$ on the test set, indicating the stability of the fitted RF models. The resulting clusters and student response vectors were used to construct the labeled dataset. We recall that XAI scheme treats the items as the *features* and the student-to-cluster mapping as the *labels*.

Step 3. Using Feature Importance to Identify the Profile Explanation. The resulting RF models were used to generate two types of knowledge profile explanations: global explanations and per-cluster explanations.

First, we applied the Feature Importance algorithm to the RF model to analyze the contribution of each feature (item) to the classification of students into clusters. Using the RF models and SHAP TreeExplainer algorithm, feature importance was computed for the entire model as described in step 2 and shown in Fig 3. In this step, we focused on the individual explanation of each cluster (Fig.4).

In the original, unsupervised task, we interpret this as the relative importance of each item in distinguishing between clusters. We refer to this as a global explanation, as the values represent item importance with respect to the *entire* clustering model. As noted in Section 2, global explanations are more geared toward understanding the model as a whole, which in our case can be interpreted as “which are the most important items in the activity.” However, we anticipated that such explanations would be challenging for teachers to translate into actionable insights that can be used to assign learning activities based on the needs of the specific knowledge profile of students who belong to the same cluster.

Thus, we focused on explanations that unpack the knowledge profile of the students in each cluster (we refer to this as a profile explanation). We used the SHAP TreeExplainer algorithm to build such explanations that analyze which items were most important for the RF classifier (Lundberg et al., 2020) to assign a specific student to a specific cluster. To assess the agreement on the ordering of the features (in terms of feature importance) between the data samples classified into the same clusters and the different clusters, we followed the procedure proposed in (Swamy et al., 2022; Swamy et al., 2023). We used Spearman’s Rank-Order Correlation to identify the rank correlation (statistical dependence between the rankings) between every two pairs of samples (students) (Spearman, 1987).

The results of Spearman’s Rank-Order correlation tests are presented in Fig.5. The four dark squares on the diagonal correspond to the four clusters and indicate a high Spearman’s p-value inside the detected profiles. Interestingly, additional dark rectangles indicate a high correlation between two pairs of groups, namely, Groups 1 and 3 and Groups 2 and 4. Indeed, the most important items for Groups 1 and 3 are Items 9, 6, 8, and 5. However, in the case of Group 1, the relevant cells are green (Fig.2), indicating that the students of the group were correct on these items, while for Group 3, the cells of Items 9, 8, and 5 are colored red (Fig.2), indicating that the students of Group 3 were incorrect on these items.

Spearman’s Rank-Order correlation validates the feature importance, ranks the item importance for each data point (student), and examines the similarity of the ranking between data points mapped to the same cluster. We observed high consistency between the item importance ranking for data points belonging to the same cluster across all clusters, providing additional validation to the stability of the per-profile item importance profile. This validation step was also applied to the three other instruments, yielding similar proof.

Eventually, the resulting item importance scores are interpreted as explaining the specific assignment of a student into a certain knowledge profile, providing the items that are the most important for distinguishing between assigning the student to a certain cluster vs. into the neighbor ones.

4.2 Qualitative Evaluation: User-Study with Pilot Teachers

In the context of RQ2, we employed a qualitative approach to evaluate teachers’ attitudes toward the explanations rendered by the proposed XAI scheme. We were especially interested in evaluating teachers’ opinions about the effectiveness of the information to assist them in designing follow-up activities that ad-

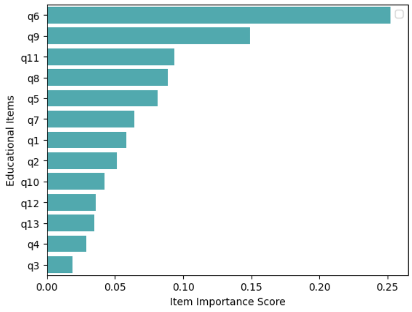

Figure 3: Feature importance for the entire RF model.

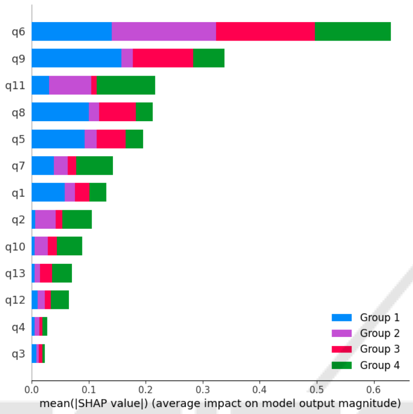

Figure 4: Per-cluster feature importance.

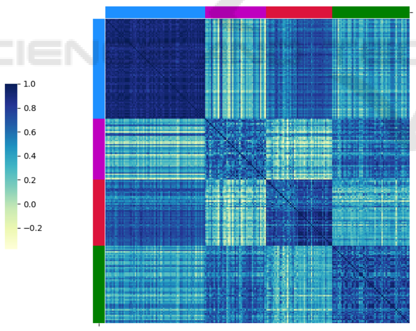

Figure 5: Spearman's Rank-Order correlation between all the pairs of data points. The data points ordered by the corresponding groups and represented by colors as in Figure 4.

dress the needs of each knowledge profile. To this end, we conducted semi-structured interviews with twelve teachers. During the interviews, we first presented the following scenario/task as a context for the interview:

1. Identify students' difficulties based on their performance in an interactive learning activity.
2. Divide students into groups of students having similar knowledge profiles with respect to the skills and competencies that underlie the activity.
3. Propose a learning sequence for each group based on its knowledge profile.

Next, we presented to the participants the underlying learning activity and the corresponding automated division into knowledge profiles. We used the learning analytic dashboard (Fig.1, Fig.2) to visualize the resulting knowledge profiles. Finally, we presented to the participants the explanations for the knowledge profiles. We presented the XAI results using two mock-ups: one containing global explanations (item importance for the entire model) and one presenting profile explanations (item importance for each cluster). For each mock-up, we asked the teachers to share their perspectives on the potential advantages and disadvantages of including the item importance analytics in the interactive dashboard. Each interview lasted 45-60 minutes and was recorded and transcribed by the first author. The interview analysis proceeded in the following manner. In the first stage, the first author identified and coded three primary themes that emerged in the interviews. These themes were formed inductively in a "bottom-up" manner (Braun and Clarke, 2006). In the second stage, to validate the themes, the second author analyzed the interviews using the themes identified by the first author. After reaching almost a full consensus regarding the interview data, the two authors met to discuss and reach a complete agreement. Due to space limitations, we omit the full results of the qualitative analysis. Instead, we focus only on the examples of teachers' opinions about incorporating the XAI analytics into the cluster dashboard. We used the interviews to assess the XAI scheme and examine whether and how we should implement the feature importance for the entire model and the knowledge profiles in the interactive dashboard. The twelve interviews resulted in nineteen citations, as summarized below, according to the following three themes:

1. Interviewees accepted the explanation and visualization: Feature Importance Algorithm: For example, "It adds information for me because the questions are very goal-oriented, and then I know exactly what I need to improve" (6 interviewees). Profile explanations: For example, "To me, it's important because when developing an assignment, you don't know the important questions. It gives more information. I can tell that question 5, for example, is critical. The form of the presentation in the table is good. For me, it's good".(5 interviewees).
2. Interviewees accepted the explanation but not its visualization. Feature Importance Algorithm: For example, "I would present the important general questions elsewhere, next to the table with the

differences between the groups"(2 interviewees). Profile explanations: For example, "I would not present it in this way, but provide a clarification; this first question is a question that most of the cluster students answered correctly, and the rest of the clusters did not answer correctly. I would prefer this explanation to be in text and not graphical". (3 interviewees).

3. Indecisive - Interviewees could not decide whether they accepted the explanation and visualization or not. Feature Importance Algorithm: For example, "I have to try it myself to trust it".(3 interviewees).

   Profile explanations: No support for this was found in the interviews.

As evident from the interviews, there was no clear consensus among the teachers on this matter. As we advance our research, we aim to investigate deeper into teachers' understanding to determine whether it impacts their acceptance of feature importance. Another factor that we intend to further explore is the impact of the task (e.g., developing learning materials vs. formative assessment) on teacher acceptance and use of item importance analytics.

5 DISCUSSION AND CONCLUSION

In recent years, the importance of AI in educational technologies has grown substantially. One of the primary barriers to its acceptance is the perception of AI as a 'black box', as both end-users and developers often find the pathways to algorithmic outcomes opaque and non-transparent. Previous studies have highlighted stakeholders' challenges due to their inability to understand how AI algorithms reached particular results (Rudin, 2019). Explaining and understanding these AI decisions have thus become critical for their acceptance in educational contexts (Khosravi et al., 2022; Qin et al., 2020). While XAI aims to increase the transparency of AI-driven systems by providing detailed insights into the decision-making process, there is a notable gap in the literature concerning its application in education, especially for clustering, which is a key unsupervised ML method. While some research studied this area, it mainly focuses on bioinformatics (Rider et al., 2010), leaving the educational sector without standardized XAI techniques for clustering. This is despite the extensive use of cluster analysis in various educational research scenarios (Gabbay and Cohen, 2022; Klingler et al., 2016; Mojarad et al., 2018).

With this in mind, the study first addressed the following research question, RQ1: If and how ML and XAI methods can be employed to create meaningful explanations for automatically detected profiles of student knowledge? We developed an XAI scheme tailored to unsupervised ML to address this. This scheme introduces an 'explainability layer' for the knowledge profiles automatically computed from student responses to educational items. In short, the crux of this scheme is treating the cluster analysis output – assignment of students to clusters, as a supervised learning task, with the assessment items as features, student response vectors (true/false on each item) as inputs, and the mapping of response vectors (students) to clusters as labels. Applying feature importance techniques (SHAP TreeExplainer) to this modeling computes the item importance, which can be interpreted as the contribution of each item to the overall clustering (global explanations) as well as to specific student-cluster pair (profile explanation). We first assessed this scheme computationally, on actual student response data received from four instruments in Physics and Chemistry. For all instruments, the application of the scheme was statistically robust with respect to agreement on the ordering of the features (in terms of feature importance) between the data samples classified into the same clusters and the different clusters. Delving into the results of one instrument, analysis of the scores conducted by a content expert indicated that they unpack the key factors that determine the clustering and specific student-to-cluster mappings.

After computationally assessing the scheme, we conducted a user study with teachers to address RQ2: Do teachers find the explanations useful for understanding the detected knowledge profiles? The pilot's goal was to understand the scheme's comprehensibility and practical value for teachers. By investigating teachers' acceptance of the item importance scores, yielding both global and profile explanations, and presented through two mock-up dashboards, we identified three main attitudes among the teachers: 1) accept the explanations, 2) accept the explanations but do not accept their visualization, and 3) indecisive. Most of our interviewees were assorted to the first or second attitude. Surprisingly, only three out of twelve teachers were indecisive regarding whether they should accept the *global* explanations, and *none of them* was indecisive regarding the *profile* explanations. Concluding, the pilot provided a clear indication that the teachers found the explanations insightful and largely comprehensible. Yet, there were a few indications of confusion among the teachers. Interpreting these in the light of the results of (Swamy et al., 2023), who reported disagreement and possible confusion among educators (in Higher Education) regarding applying LIME and SHAP to explain the classification of student assessment data. Synthesizing the results, we hypothesize that the *visualization* of the feature importance analytics plays an important role in their comprehensibility. Thus, and since the interpretation of explainability schemes may be subjective and complex, it may be useful to allow users to navigate and choose between alternative forms of visualization, as is common in some other algorithmic-oriented used interfaces (Sadeh, 2008).

In summary, a pilot computational and user study conducted on the XAI scheme for unsupervised educational ML that was proposed in this research yielded promising results. The main limitation of the research is the small sample in terms of the amount of teachers that participated, the amount of instruments, and the domains from which they were taken. Thus, applying this scheme to additional teacher populations and to instruments from other domains will help to understand its generalizability. In addition, the present work devoted little attention to the visual aspects of the analytics, but the way analytics are presented to teachers needs further exploration through user-interface studies, including exploring the contribution of making educational dashboards customizable by their end users, an approach that rendered positive outcomes in other domains (Kuznetsova et al., 2021). To conclude, the suggested XAI scheme provides a promising pathway for adding transparency and explainability to LA applications that rely on unsupervised learning algorithms. Research-wise, it underlines the different nature of explanations that are required for supervised and unsupervised educational ML, the latter being mostly dominated by applications of cluster analysis – the algorithmic bedrock underlying our learning analytic dashboard. This makes the proposed scheme applicable to a wide range of LA solutions.

REFERENCES

Ahn, J., Campos, F., Hays, M., and DiGiacomo, D. (2019). Designing in context: Reaching beyond usability in learning analytics dashboard design. *Journal of Learning Analytics*, 6(2):70–85.

Asif, R., Merceron, A., Ali, S. A., and Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113:177–194.

Baker, R. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26:600–614.

Baker, R., Corbett, A. T., and Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Intelligent Tutoring Systems: 9th International Conference*, pages 406–415.

Bandyapadhyay, S., Fomin, F. V., Golovach, P. A., Lochet, W., Purohit, N., and Simonov, K. (2023). How to find a good explanation for clustering? *Artificial Intelligence*, page 103948.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Conati, C., Barral, O., Putnam, V., and Rieger, L. (2021). Toward personalized xai: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298.

Cukurova, M., Zhou, Q., Spikol, D., and Landolfi, L. (2020). Modelling collaborative problem-solving competence with transparent learning analytics: is video data enough? In *LAK20: 10th international conference on learning analytics and knowledge*, pages 270–275.

Dasgupta, S., Frost, N., Moshkovitz, M., and Rashtchian, C. (2020). Explainable k-means and k-medians clustering.

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., Cuendet, S., and Kaplan, F. (2011). Classroom orchestration: The third circle of usability.

Din, B., Nazaretsky, T., Feldman-Maggor, Y., and Alexandron, G. (2023). Automated identification and validation of the optimal number of knowledge profiles in student response data. In *Proceedings of the 16th International Conference on Educational Data Mining*, pages 458—-465.

Feldman-Maggor, Y., Barhoom, S., Blonder, R., and Tuvi-Arad, I. (2021). Behind the scenes of educational data mining. *Education and Information Technologies*, 26(2):1455–1470.

Fiok, K., Farahani, F. V., Karwowski, W., and Ahram, T. (2022). Explainable artificial intelligence for education and training. *The Journal of Defense Modeling and Simulation*, 19(2):133–144.

Gabbay, H. and Cohen, A. (2022). Investigating the effect of automated feedback on learning behavior in moocs for programming. In *Proceedings of the 15th International Conference on Educational Data Mining*, pages 376–383.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666.

Käser, T., Busetto, A. G., Solenthaler, B., Kohn, J., von Aster, M., and Gross, M. (2013). Cluster-based prediction of mathematical learning patterns. In *Artificial Intelligence in Education: 16th International Conference*, pages 389–399.

Kauffmann, J., Esders, M., Ruff, L., Montavon, G., Samek, W., and Müller, K.-R. (2022). From clustering to cluster explanations via neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., and Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3:100074.

Klingler, S., Käser, T., Solenthaler, B., and Gross, M. (2016). Temporally coherent clustering of student data. *International Educational Data Mining Society*.

Kuznetsova, M., Frits, M. L., Dulgarian, S., Iannaccone, C., Mort, E., Bates, D. W., and Salmasian, H. (2021). An analysis of the structure and content of dashboards used to monitor patient safety in the inpatient setting. *JAMIA open*, 4(4):ooab096.

Liao, Q. V., Gruen, D., and Miller, S. (2020). Questioning the ai: informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–15.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S. I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2:56–67.

Marras, M., Vignoud, J. T. T., Kaser, T., et al. (2021). Can feature predictive power generalize? benchmarking early predictors of student success across flipped and online courses. In *Proceedings of the 14th International Conference on Educational Data Mining*, pages 150–160.

Michaeli, S., Kroparo, D., and Hershkovitz, A. (2020). Teachers' use of education dashboards and professional growth. *International Review of Research in Open and Distributed Learning*, 21(4):61–78.

Mohseni, S., Zarei, N., and Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Transactions on Interactive Intelligent Systems*, 11.

Mojarad, S., Essa, A., Mojarad, S., and Baker, R. S. (2018). Data-driven learner profiling based on clustering student behaviors: learning consistency, pace and effort. In *Intelligent Tutoring Systems: 14th International Conference*, pages 130–139.

Morichetta, A., Casas, P., and Mellia, M. (2019). Explain-it: Towards explainable ai for unsupervised network traffic analysis. pages 22–28.

Nazaretsky, T., Ariely, M., Cukurova, M., and Alexandron, G. (2022a). Teachers' Trust in AI-powered Educational Technology and a Professional Development Program to Improve It. *British Journal of Educational Technology*, 53(4):914–931.

Nazaretsky, T., Bar, C., Walter, M., and Alexandron, G. (2022b). Empowering Teachers with AI: Co-designing a Learning Analytics Tool for Personalized Instruction in the Science Classroom. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, pages 1–12.

Nazaretsky, T., Hershkovitz, S., and Alexandron, G. (2019). Kappa learning: A new item-similarity method for clustering educational items from response data. In *Proceedings of the 12th International Conference on Educational Data Mining*, pages 129–138.

Pereira, F. D., Oliveira, E. H., Oliveira, D. B., Cristea, A. I., Carvalho, L. S., Fonseca, S. C., Toda, A., and Isotani, S. (2020). Using learning analytics in the amazonas: understanding students' behaviour in introductory programming. *British Journal of Educational Technology*, 51:955–972.

Qin, F., Li, K., and Yan, J. (2020). Understanding user trust in artificial intelligence-based educational systems: Evidence from china. *British Journal of Educational Technology*, 51(5):1693–1710.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. pages 1135–1144.

Rider, A., Siwo, G., Emrich, S., Ferdig, M., and Habtom, N. V. C. (2010). A supervised learning approach to the unsupervised clustering of genes. In *2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 322–328.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215.

Ruiperez-Valiente, J. A., Munoz-Merino, P. J., Alexandron, G., and Pritchard, D. E. (2017). Using machine learning to detect 'multiple-account'cheating and analyze the influence of student and problem features. *IEEE transactions on learning technologies*, 12(1):112–122.

Sadeh, T. (2008). User experience in the library: a case study. *New library world*, 109(1/2):7–24.

Setzu, M., Guidotti, R., Monreale, A., Turini, F., Pedreschi, D., and Giannotti, F. (2021). Glocalx - from local to global explanations of black box ai models. *Artificial Intelligence*, 294.

Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4):441–471.

Swamy, V., Du, S., Marras, M., and Kaser, T. (2023). Trusting the explainers: teacher validation of explainable artificial intelligence for course design. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 345–356.

Swamy, V., Radmehr, B., Krco, N., Marras, M., and Käser, T. (2022). Evaluating the explainers: Black-box explainable machine learning for student success prediction in MOOCs. In *Proceedings of the 15th International Conference on Educational Data Mining*, pages 98–109.