Discovering Unusual Study Patterns Using Anomaly Detection and XAI

Elena Tiukhova
KU Leuven
LIRIS
elena.tiukhova@kuleuven.be

Pavani Vemuri
KU Leuven
LIRIS
pavani.vemuri@kuleuven.be

María Óskarsdóttir
Reykjavík University
Department of Computer Science
mariaoskars@ru.is

Stephan Poelmans
KU Leuven
LIRIS
stephan.poelmans@kuleuven.be

Bart Baesens
KU Leuven
LIRIS
bart.baesens@kuleuven.be

Monique Snoeck
KU Leuven
LIRIS
monique.snoeck@kuleuven.be

Abstract

Learning Analytics (LA) has been leveraged as a tool to analyze and improve educational processes by informing its stakeholders. LA for student profiling focuses on discovering learning patterns and trends based on diverse features extracted from trace data. Prior studies have used classical clustering methods to group students and understand the study patterns of each cluster. However, variations within the clusters are still large making it difficult to draw concrete insights into the relation between study behaviors and learning outcomes. In this work, we leverage anomaly detection and eXplainable AI techniques to distinguish between normal and abnormal study patterns and to possibly discover unexpected patterns that are not apparent from clustering alone. We perform external validation to check the generalizability and compare the insights on study patterns from our method to be at par with insights gained from previous studies.

Keywords: Learning Analytics, Anomaly Detection, Isolation Forest, XAI, SHAP

1. Introduction

Recent developments in machine learning bring new opportunities allowing practitioners to discover patterns and unknown relationships in different domains, including education and learning (Alyahyan and Düştegör, 2020). The field of Learning Analytics (LA) is a growing yet relatively young field that uses data processing technologies and data generated in educational contexts to understand learners' needs and facilitate decision-making in educational institutions (Okoye et al., 2020). Technological breakthroughs made technology-based education possible and led to the emergence of blended learning (BL) that complements instructor-led training with other electronic activities (Bersin, 2004). The data obtained from a blended course together with unsupervised approaches can be leveraged to perform LA for BL. In recent years, several higher education institutions are facing problems of attrition and high duration in obtaining graduation. To mitigate these problems it is important for teachers to understand study behaviors in order to guide them and motivate them to reach their learning goals successfully.

To understand how students study, it has been common practice to look for how patterns emerge from trace data and how different students cluster together. The discovery of patterns in student behaviors using classical techniques like clustering has been used by Lust et al. (2013) to identify subgroups of learning approaches. Examples also include profiling students based on their questions to instructors (Harrak et al., 2018). Research has shown that high-scoring students perform significantly different preparatory learning activities than lowest-scoring students and that students' profiles are consistent across courses (Mejia-Domenzain et al., 2022).

Newer techniques are constantly being borrowed from other domains into LA. This is also true specifically when looking for patterns in trace data. In this regard, methods from the sequence mining and process mining domains allow extracting meaningful insights from time-ordered logs, where learning is seen as sequential, to discover the real process of learning (López-Pernas et al., 2021; Saqr et al., 2023). Even though these are not scrutinized as a part of the current study, these analytical techniques are gaining momentum among educational researchers in pattern finding and student profiling.

Many studies suggest that different learning behavior leads to different outcomes, i.e., active participation in

URI: https://hdl.handle.net/10125/106555
978-0-9981331-7-1

a course correlates with a student passing (Jovanović et al., 2017; N. Li et al., 2015; Sher et al., 2020). Hence, we could presume that being extremely active results in higher success while frequent procrastination leads to a higher probability of failure. When profiling students based on their behaviors extracted from trace data, Harrak et al., 2018 suggests that we can still see some variability in both learning activities and the outcome scores within different groups. As a result, there are students who cannot be easily profiled, as their behavior does not map onto the expected combination of study pattern and outcome. Hence, the current study proposes to leverage anomaly detection (AD) techniques aimed at finding patterns that do not conform to expectations. It is important to not only find students with extreme behavioral patterns but also explain why they are considered extreme. eXplainable AI (XAI) techniques have the potential to help understand the model and use it for improving education.

The goal of this paper is to add to the existing body of LA research by applying AD and XAI techniques in an innovative and creative way for discovering extreme behavioral patterns and providing insights into how they relate to academic success. By doing so, the following contributions will be made. We train an isolation forest (iForest) model to detect unusual study patterns in a single BL course context using a strongly theory-grounded self-regulated learning (SRL) featurization. We also adapt features from consumer consumption research and align them with SRL. The predictions are explained by using SHapley Additive exPlanations (SHAP) for decision trees separately for passing/failing groups of students marked as anomalous. [1] Next, the results obtained from AD and XAI models are used to generate insights that are not apparent from clustering alone. External validation is performed to check generalizability. Lastly, we reflect on how these techniques can be leveraged to be useful for stakeholders.

The remainder of this paper is structured as follows. Section 2 gives an overview of the applications of AD and XAI for LA. Section 3 describes the data, feature engineering, and the analytical techniques used. The results, discussion, and external validation are presented in Section 4. Section 5 concludes the research, discusses the limitations, and looks toward future research.

2. Related Work

Anomaly Detection Anomaly detection (AD) is used to recognize unexpected patterns with techniques such as traditional statistical models, machine learning, and deep learning (Pang et al., 2021). AD presumes that anomalous observations are a minority and that anomalous behavior is rarely known beforehand which is in line with some of the issues that LA research has tackled already and hopes to address.

One of such applications is cheating detection. Cheating detection is particularly vital in the context of Massive Open Online Courses (MOOCs) where it is easier to cheat due to the anonymity (King et al., 2009). AD models trained on a known set of cheaters can be used to detect anomalous behavior such as Copying Using Multiple Accounts and unauthorized collaboration (Alexandron et al., 2019). Applications of AD for cheating has become even more applicable since COVID-19 which changed the learning patterns of students and possibly caused the evolution of abnormal study behavior. In particular, switching to online education and remote assessments has increased the risk of cheating (Otero et al., 2021). Normally, it is expected that students with similar learning activities during the semester perform similarly on assignments or exams. AD can be used to identify those who do not conform to the expectations and achieving unexpected grades can be seen as cheating (Otero et al., 2021).

AD techniques can be applied for mining unexpected patterns in the course evaluation process to investigate deficiencies in the evaluation systems (Pabalkar et al., 2023). Looking for anomalies can also be context driven: a recent review focuses on educational anomaly analytics from a methodological standpoint and categorizes methods into predicting course failure, dropout, difficulty in graduating and in employment after graduation (Guo et al., 2022). In contrast, our research takes on an approach of looking for unexpected patterns in study behavior within a course, and those patterns are not necessarily unfavorable.

XAI in Education Nowadays, there is a rising demand for AI models to be not only accurate but transparent and understandable. These requirements led to the emergence of the eXplainable Artificial Intelligence (XAI) field concerned about making decision-making performed by models understandable for humans (Molnar, 2022). Explanations generated by XAI vary in scope and can be local or global depending on whether they focus on a single prediction or explain an entire model (Molnar, 2022). XAI techniques can also be model-agnostic or model-specific and vary in complexity, expressive power, translucency, and portability (Robnik-Šikonja and Bohanec, 2018).

The LA domain is not an exception with a rising interest in using XAI to increase trust in AI-powered

[1] The code and online appendices of the paper can be found at https://github.com/tiu-elena/XAD_LA

education systems. For example, XAI is used to explain student success prediction models in order to investigate the agreement between explanations generated by different techniques as well as to discover the common patterns revealed by the explainers (Swamy et al., 2022). In particular, XAI models do not generally agree with each other on feature importance; hence, choosing a particular technique has a great influence on the patterns revealed (Swamy et al., 2022). Explaining the reasoning behind deep learning knowledge tracing (DLKT) models is another application of XAI in education (Lu et al., 2020). Applying layer-wise propagation as a tool to explain predictions by mapping output on input features is proved to be feasible to interpret DLKT model's prediction results (Lu et al., 2020). Another application of XAI is facilitating early student-at-risk prediction models: by generating explanations using the SHAP framework applied to the models built at different stages of a course's progression, the instructors can obtain the most important features associated with students' performance and intervene in a timely manner (Adnan et al., 2022).

The use of AI in education raises concerns about transparency and fairness with XAI aimed at increasing trust in AI applications. XAI-ED, a framework for education, addresses unique characteristics of applying XAI in educational contexts and serves as a tool to study, design and develop educational AI tools (Khosravi et al., 2022). It raises concerns about the need for accurate, actionable, and personalized explanations, human-centered design, and evaluation of XAI impact in education (Khosravi et al., 2022).

Using XAI techniques in AD can provide more insight into a model's reasoning behind labeling an observation as anomalous, making it more transparent than just receiving an anomaly alert (Tallón-Ballesteros and Chen, 2020). Explainable Anomaly detection (XAD) is defined as "the extraction of relevant knowledge from an AD model concerning the relationships either contained in the data or learned by the model" (Z. Li et al., 2022). XAD is a relatively new and underexplored research area with great potential to increase trust in AD applications.

3. Methodology

An overview of the methodology is given in Figure 1. The data and its preprocessing is described in Sections 3.1 and 3.2, respectively. Clustering and subsequent profiling are described in Section 3.3. AD and applying the SHAP framework are presented in Sections 3.4 and 3.5, respectively.

3.1. Data and Context

The *Accountancy* course is one of the mandatory courses for first-year bachelor's students offered in four different study programs at the Faculty of Economics and Business (FEB) in the first semester at KU Leuven university. Each year approximately 700+ students of FEB follow this course. The current dataset was sourced from the Academic Year 2018-2019 (AY1819). The instructor or didactic team teaching the course has autonomy over how the course is taught, the pedagogy employed and the toolset used in the LMS. The course is offered as a BL course with course material provided via the LMS. Practical sessions are conducted once a week with a monitor where students work individually and ask for help if needed. Students' first exam attempt grades are used to measure their performance, with a written end-of-semester exam consisting of multiple choice and open questions. The primary datasets were extracted from the foundation technologies of a Higher education Institution: logs from the LMS and grades from the student information system (SIS). Due to strict GDPR guidelines in Europe, other variables like socio-demographic data are not sourced from the SIS.

3.2. Feature Engineering

Theory-based Learning Indicators A review by Ahmad et al., 2022 has revealed that more recent research on study features is grounded in literature, specifically in self-regulated learning (SRL). In line with this evolution, we ground our study in the constructivist model of SRL by Winne and Hadwin, 1998. Learners act as active agents creating learning artifacts by evaluating learning material, tools, and tactics through meta-cognitive monitoring, influenced by various internal (e.g., motivation, prior knowledge) and external (e.g., teachers' roles, course requirements) conditions. In several studies the relevance of the SRL has been well-established in classroom-based, blended, and online learning contexts (Khan and Ghosh, 2021; Rasheed et al., 2020). SRL has been used together with clustering techniques for studying learning patterns (Lust et al., 2013; Sher et al., 2020). With the goal of identifying anomalous patterns, our study incorporates established SRL indicators from the existing literature.

Existing SRL Indicators As the current study is looking for anomalous study patterns, it is only logical to incorporate features that capture high-level learning actions by aggregating original low-level granular events captured by an LMS platform. To this end, we incorporate the features as proposed by Jovanović et al.,

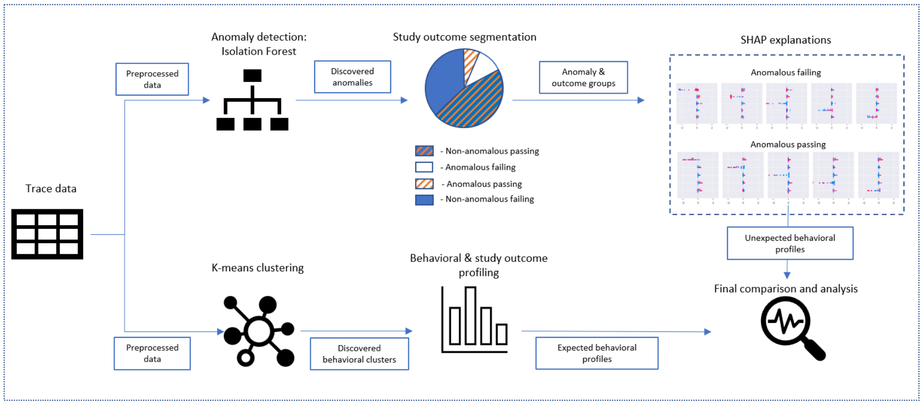


Figure 1: Methodology

2021 where a set of SRL features were derived from the LMS log data based on well-established related studies (See their paper for the complete list of indicators and the list of previous studies that utilized them). Study grades are classified into a binary Pass/Fail.

The study by Jovanović et al., 2021 proposes capturing student study behavior on the levels of activity and study regularity. Additionally, for both levels, the overall and learning action-specific patterns are captured resulting in four sets of features: overall level of activity, learning action-specific level of activity, overall study regularity, and learning action-specific study regularity. Within the specific level of activity, four different specific types of learning actions are further delineated: forum contribution, forum consumption, access to learning materials, and access to the main course page. Due to the unavailability of granular event data for forum consumption and contribution, the original features measuring the proportion of active days/weeks of forum consumption/contribution cannot be calculated. Hence, we express forum consumption as a proportion of total posts read relative to the total number of posts available on a course's discussion forum. Forum contribution is expressed as the total number of posts written on a discussion forum.

Next, the sequence of learning actions performed within a time frame of a maximum of eight hours is defined as a session, and session-related features are also constructed. Cases of multiple events per timestamp which occur due to the idiosyncrasies of the LMS (for example when a directory is opened, event logs are created for opening all sub-directories) are preprocessed. As the courses have weekly periodicity, the weekly indicators are also considered. In earlier findings and according to Jovanović et al., 2021, higher values of entropy correspond to spreading studying more uniformly over time resulting in higher regularity. Hence a more intuitive name 'Constancy' is used to replace entropy. We generate constancy of total clicks and session length; including daily and weekly constancy of clicks on the course main page and its material. These listed indicators/features so far did not seem to capture all aspects of SRL and how a student moderates himself through the semester, especially with respect to how activity in different sessions is captured.

Borrowing New Indicators from Other Domains As learning unfolds with time and returning to the study material in line with the SRL theory, we also propose to capture additional student self-regulatory behaviors. Therefore, we drew some parallels with customer consumption analysis in finance by treating students as customers consuming the learning materials. Features were adopted from customer consumption patterns like exploration, exploitation, and plasticity across time (Singh et al., 2015). These traits are usually proxied by diversity, loyalty, and regularity metrics within customer consumption research. Applying to the current context and aligned to the SRL theory, these metrics will correspond to 'Uniformity', 'Bingeing', and 'Regularity'. See Table 1 for definitions and the online appendix (link in Section 1) for the equations.

Data Preprocessing We drop students who did not participate in the final examination and who did not have corresponding activity on the LMS platform. Missing values for the features using Shannon's entropy calculation are imputed with the maximal possible entropy as the absence of studying represents a "missing not at random" situation where missing data cases are systematically related to the unobserved data. We assume here that the study regularity when no studying is actually happening is uniform which can be represented by a maximal possible entropy value.

Table 1: SRL Features used in this study based on Jovanović et al., 2021 and Singh et al., 2015

Feature	Description
Constancy of clicks	Entropy calculated based on the probabilities estimated as the proportion of the number of learning actions per session relative to the total number of learning actions across all the sessions
Constancy of session length	Entropy calculated based on the probabilities estimated as the proportion of a session's length relative to the total sessions length across all the sessions
Median difference between active days	Median calculated over the time distances between consecutive active days during the course.
Median number of actions per session	Median calculated over all the sessions' total learning actions counts
Median number of active days per week	Median calculated over all the weeks with active days during the course
Median session duration	Median calculated over all the sessions' duration during the course, seconds
Proportion of active days	Total number of active days relative to the duration of the course in days
Proportion of first day of week activity	Median calculated over all the weeks for the proportion of the learning actions performed on the first day of the week relative to the total number of actions performed in this week
Proportion of posts read	Total number of posts read on the forum relative to the total number of posts available on a discussion forum
Proportion of weeks first-day activity	Total number of active weeks relative to the duration of the course in weeks
Total number of created posts	Total number of posts written on discussion forum during the course's duration
Total number of sessions	Total number of sessions of non-zero duration
Total sessions duration	Sum over all the sessions duration during the course, seconds
Uniformity of sessions	Measuring the extent to which students spread their sessions over time, specifically a semester.
Bingeing of sessions	Spread of sessions across different bins (spatial or temporal) with higher values corresponding to having most of the transactions within top-N of the bins. It expresses concentrating the study efforts in a small number of time slots.
Regularity of sessions	Regularity measures the differences between behavioral patterns over shorter and longer periods of studying

The same goes for uniformity, bingeing, and regularity metrics that are imputed with the maximal possible value of 1. Furthermore, we performed a correlation study on the features (see Figures 1 and 2 in the online appendix) and drop the features we find highly correlated to each other, thus arriving at 16 features as listed in Table 1. In particular, specific level of activity and regularity features have been dropped. Finally, min-max normalization is used to transform the data for the clustering algorithm.

3.3. Clustering

According to Ahmad et al., 2022, clustering is commonly used for student profiling based on behavioral features. Navarro and Ger, 2018 identified the K-means clustering algorithm as a superior technique for profiling studies when compared to other commonly utilized clustering algorithms. Hence, we use K-means for its simplicity in implementation, scalability to large datasets, and its popularity within the LA community. First, we evaluate the clustering tendency using the Hopkins test (Hopkins and Skellam, 1954). Next, we search for the optimal solution from two to ten clusters using 1000 restarts and a scree plot with the number of clusters against the sum of squared residuals to find the optimal number of clusters. The Silhouette coefficient is used to evaluate the quality of the clustering solution (Rousseeuw, 1987).

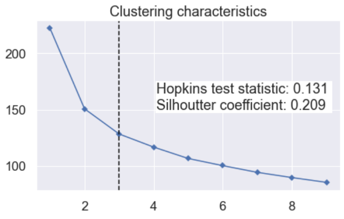


Figure 2: Clustering characteristics: vertical line represents an elbow point - the optimal no. of clusters

3.4. Isolation Forest

iForest is chosen for AD as it consistently demonstrated good performance in terms of detection and speed (Tiukhova et al., 2022). The model uses isolation to identify unusual data points rather than creating a profile of normal instances (Liu et al., 2008). It consists of a group of isolation trees that separate data points by randomly partitioning a subset of the data until each point is isolated. Anomalies require fewer partitions and therefore have shorter path lengths. iForest is implemented using the *scikit-learn* Python library, and the hyperparameters are set as in the original paper by Liu et al., 2008 (100 isolation trees, subsampling size of 256).

3.5. SHAP TreeExplainer

Shapley values originated from game theory and can be used to estimate the contribution of each feature to a prediction by considering a feature as a "player" and the prediction as a "payout" that can be distributed among "players" (Molnar, 2022). However, the exact computation of Shapley values is expensive, so the model-agnostic SHapley Additive exPlanations (SHAP) method was proposed as an approximation (Lundberg and Lee, 2017). A general idea of SHAP is to approximate the original model output $f(x)$ as a sum of feature effects ϕ_i: $g(z') = \phi_0 + \Sigma_{i=1}^{M}\phi_i z'_i$ where $z\prime \in \{0,1\}^M$, M is the number of simplified input features, and $\phi_i \in \mathbb{R}$. Each feature contribution is estimated as the average marginal contribution for each feature over all possible feature subsets (Equation 1).

$$\phi_i(f,x) = \sum_{R \in \mathcal{R}} \frac{1}{M!}\left[f_x(P_i^R \cup i) - f_x(P_i^R)\right] \quad (1)$$

where $\mathcal{R}$ is a set of all feature orderings, P_i^R is a set of feature orderings coming before feature i in the ordering $\mathcal{R}$, M is the number of input features.

TreeExplainer algorithm, an extension of SHAP, can calculate exact Shapley values in polynomial time for tree-based models by collapsing feature subsets into calculations specific to each leaf in a tree by relying only on path coverage information stored in the model (Lundberg et al., 2020). SHAP TreeExplainer is capable of incorporating both interventional and observational expectations feature perturbation techniques, and the choice of the technique is application dependent. "True to the data" approach favors observational expectations while "true to the model" approach requires interventional expectations (Chen et al., 2020). To discover unexpected study patterns based on the trace data, being "true to the data" is more desirable while explaining the model. Hence, we opt for a path-dependent (using observational expectations) TreeExplainer algorithm and generate SHAP interaction values and take into account both main and possible interaction effects.

For iForest, the value explained by TreeExplainer is the average split path length of an observation, which is lower for anomalies than for non-anomalies. Negative SHAP values correspond to a positive contribution to the anomaly score, and vice versa. We use the official implementation of the TreeExplainer algorithm using tree path-dependent approach (Lundberg et al., 2020).

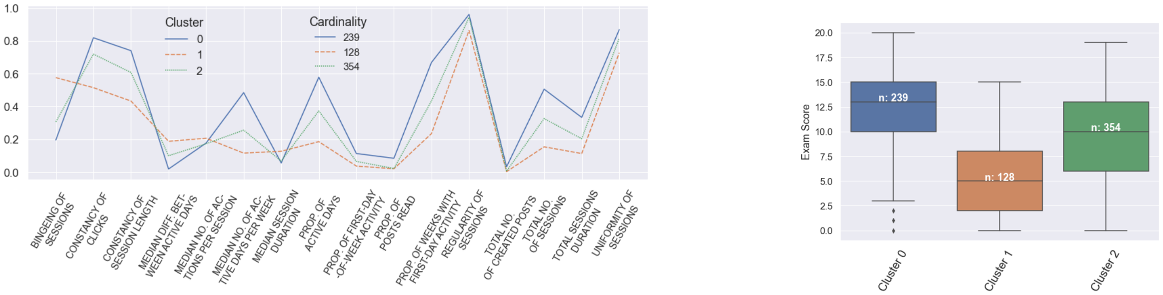
4. Results and Discussion

4.1. Clustering

The characteristics of the clustering solution are presented in Figure 2. The Hopkins test statistic value is lower than the reference value of 0.3 meaning that clustering makes sense. By inspecting the scree plot of SSE values, we locate an elbow corresponding to three clusters. The Silhouette coefficient is positive meaning that an observation belonging to a particular cluster can be well associated with this cluster.

To examine the variations in study indicators across clusters, we analyze the mean value of each indicator within each cluster, along with the cluster's cardinality. Figure 3a provides insights into these differences. Cluster#0 consists of highly active (n=239) students who exhibit high constancy in terms of clicks and session length, engage in a high number of sessions with longer durations, have low bingeing tendencies, and show a high number of active days per week, with smaller intervals between days. In contrast, cluster#1 holds (n=128) students with low levels of activity across the aforementioned behavioral indicators. Cluster#2 represents a majority of students (n=354) with moderate levels of activity, falling between the characteristics of cluster#0 and cluster#1. The revealed clusters align with those of Lust et al., 2013 and Sher et al., 2020. The former identified 3 clusters characterized by different tool-choice (in LMS) behaviors: the non, selective, and intensive users while the latter identified 3 distinct study patterns among learners to be highly or incrementally consistent or completely inconsistent. Nevertheless, for some features like median number of actions per session, total number of created posts, we don't really see a clear difference between clusters.

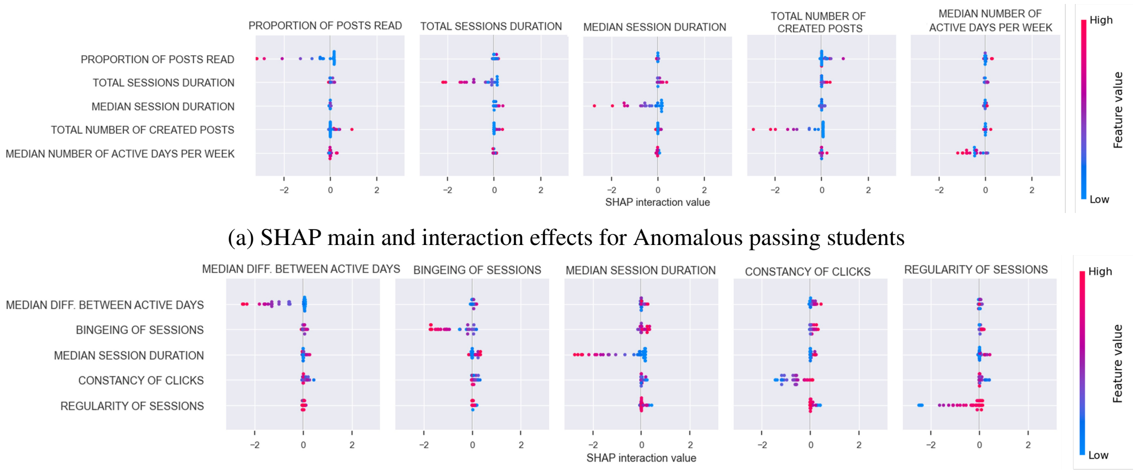
Figure 3b presents the relationship between exam scores and clusters using boxplots. Notably, in cluster#0, the majority of students (75%) pass the exam, whereas an even larger majority of students in cluster#1 fail. The students from cluster#2 exhibit both passing and failing outcomes with no clear tendency. These findings suggest that students who are actively involved in the course are more likely to pass, which comes as no surprise as previous studies also revealed significant associations between identified patterns of activity and consistency with student's academic performance (N. Li et al., 2015; Sher et al., 2020). However, the boxplot reveals the presence of students who cannot be easily profiled solely through clustering, as they display unexpected combinations of study patterns and outcomes which has been also identified as an issue for some courses investigated by Harrak et al., 2018.



(a) Mean values per indicators per cluster

(b) Boxplot: clusters vs. Exam grades

Figure 3: Clustering solution



(a) SHAP main and interaction effects for Anomalous passing students

(b) SHAP main and interaction effects for Anomalous failing students

Figure 4: SHAP summary plots. Each point represents one student and its color corresponds to the feature value (pink: higher, blue: lower). The features are displayed in the order of importance from top to bottom.

Noticeably, highly active students who still fail and inactive students who pass represent such combinations. Our objective is to identify these hidden patterns through the use of AD and XAI techniques.

4.2. Anomaly Detection and XAI

After applying AD and segmenting the student body based on study outcomes and the clusters obtained from applying K-means, the distribution of students is depicted in Table 2. We observe that the proportion of passing and failing students remains relatively consistent within both anomalous and non-anomalous groups with slightly more failing students in the anomalous group. This suggests that anomalous behavior is not correlated with either passing or failing.

As mentioned in Section 3.5, SHAP TreeExplainer algorithm is used to explain generated anomalies. Figure 4 displays SHAP summary plots depicting both main and interaction effects. Based on the intuition of explaining path lengths of the iForest model, higher negative SHAP values in summary plots correspond to a higher propensity of being anomalous. The plots on the diagonal of Figures 4a and 4b show the main effects of feature contributions for the anomalous passing and anomalous failing groups, respectively. The off-diagonal plots in both Figure 4a and 4b represent the interaction effects between top-contributing features: they are significantly smaller than main effects and can be neglected. Notice that the top contributing features

Table 2: Data distribution

Outcome	Anomalous?	Cluster 0 Active	Cluster 1 Inactive	Cluster 2 Moderate	Total
Passing	Yes	17	7	1	25
	No	167	12	179	358
Failing	Yes	3	32	2	37
	No	52	77	172	301

differ for the anomalous passing and failing groups.

The top five contributing features for the anomalous passing group are related to discussion forum behavior and overall level of activity (Figure 4a). Higher values of the proportion of posts read and the total number of posts written are associated with higher anomaly scores. The same goes for the features representing total and median session duration as well as the median number of active days per week. The contribution of most of these indicators (median session duration, proportion of posts read and total number of created posts) to study outcomes was not apparent from the clustering solution presented in Figure 3. Hence, active contribution and consumption of discussion forums are associated with positive study outcomes. It is well established in literature that regular contributions to the discussion forum is associated with higher grades (Conijn et al., 2016; Jovanović et al., 2021)

For the anomalous failing group, two contributing indicators represent regularity of study. In particular, higher values of differences between active days and bingeing and low values of constancy and regularity are associated with higher abnormality. Having sessions with longer duration also positively contributes to the likelihood of being anomalous. These insights were mostly apparent from the clustering solution alone where the mean deviations for these indicators were large enough between the clusters. AD enriched with SHAP explanations confirms these patterns for abnormal cases as well. The magnitude of these features may explain their contribution to abnormality, as extremely high or low values for these indicators are unexpected, despite their consistent impact on academic performance. Therefore, both moderate and extreme deviations from regular study habits are linked to negative study outcomes. Similar to our finding, regularity of study is also an important indicator across several courses as indicated in Jovanović et al., 2021. Hence these findings from SHAP on feature importances are at par with findings in literature establishing AD paired with SHAP as a valid method for investigating student profiles.

Table 2 also indicates that anomalous students are mostly located in either cluster#0 or cluster#1, i.e., among highly active and highly inactive students, respectively. Surprisingly, anomalies that belong to highly active cluster are mostly passing students. Hence, the anomalous patterns do not explain the unexpected combinations of behavior and study outcome but rather give additional insights into extreme patterns of an inactive student passing or an active student failing.

4.3. Validation across Time and Courses

Validation is performed across (1) courses on the data of another first-year bachelor course *Macroeconomics* offered in the second semester of AY1819, and (2) time on the data of Accountancy course of the academic year 2019-2020 (AY1920). In these courses, the number of students enrolled and participated in the first exam attempt is 722 and 710 respectively. We perform the validation of our main findings, by checking whether the combination of the iForest model and the SHAP TreeExplainer algorithm yields insights not apparent from the clustering solution. The figures with the clustering solution as well as SHAP interaction summary plots can be found on the online appendices. Similar to that of the Accountancy course in AY1819, the clustering solution reveals groups of active, moderate, and inactive students for both these courses. SHAP explanations exhibit similar patterns with higher discussion forum activity associated with passing the course and low regularity of study associated with failing. Hence, we confirm our findings on unexpected study patterns for the Accountancy course in AY1819 and deem this method generalizable.

5. Conclusion

Online education and LMSs generate large amounts of data, which can be used by AI systems for improving learning. This paper used SRL indicators engineered from trace study data together with AD and XAI to identify extreme forms of behavior and determine how they relate to study outcomes. The iForest model was used to detect anomalous behavior with subsequently applying SHAP TreeExplainer algorithm to explain detected anomalies and identify unusual learning patterns for pass and fail students separately.

We show that the combination of iForest and SHAP can provide us with insights not apparent from clustering alone when it comes to study patterns of passing students. However, these insights do not describe unexpected combinations of being inactive and still passing but rather give insights into extreme (extra) activity. More research is needed to reveal these patterns and could be possibly achieved by using local SHAP values that explain each individual case. Hence, clustering still stays important in profiling students as it generates useful insights into how students' self-regulation in terms of the overall level of activity and study regularity relates to study success.

Finally, we illustrated that AD and XAI can be used to give recommendations for promoting behavior leading to favorable study outcomes. We highlight the importance of using the discussion forum as well as spreading studying uniformly throughout the semester for achieving positive study outcomes. We demonstrated how the combination of AD with XAI can be used to analyze study behavior and discover patterns that occur rarely in the case of students with different levels of success. Our approach can be helpful for instructors who are looking for ways of giving study advice or improving course design. The insights obtained from applying iForest together with SHAP provide a higher degree of transparency and explanation and increase trust in the model's predictions.

The research has some limitations. These include the selected variables potentially not being representative enough. The course included in external validation are from the same disciple using similar Instructional Design. Additionally, we explored a limited set of AD and XAI techniques. Future research could address these limitations by analysing a larger course set and studying other ways to engineer features and comparing a wider range of AD and XAI techniques.

References

Adnan, M., Uddin, M. I., Khan, E., Alharithi, F. S., Amin, S., & Alzahrani, A. A. (2022). Earliest Possible Global and Local Interpretation of Students' Performance in Virtual Learning Environment by Leveraging Explainable AI. *IEEE Access*, *10*, 129843–129864.

Ahmad, A., Schneider, J., Griffiths, D., Biedermann, D., Schiffner, D., Greller, W., & Drachsler, H. (2022). Connecting the dots – a literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*.

Alexandron, G., Ruipérez-Valiente, J. A., & Pritchard, D. (2019). Towards a general purpose anomaly detection method to identify cheaters in massive open online courses.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Jour. of Educational Technology in Higher Education*, *17*(1), 1–21.

Bersin, J. (2004). *The blended learning book: Best practices, proven methodologies, and lessons learned*. John Wiley & Sons.

Chen, H., Janizek, J. D., Lundberg, S., & Lee, S.-I. (2020). True to the model or true to the data? *arXiv preprint arXiv:2006.16234*.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29.

Guo, T., Bai, X., Tian, X., Firmin, S., & Xia, F. (2022). Educational anomaly analytics: Features, methods, and challenges. *Frontiers in big Data*, *4*, 124.

Harrak, F., Bouchet, F., Luengo, V., & Gillois, P. (2018). Profiling students from their questions in a blended learning environment. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 102–110.

Hopkins, B., & Skellam, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, *18*(2), 213–227.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., Mirriahi, N., et al. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(4), 74–85.

Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, *172*, 104251.

Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and information technologies*, *26*, 205–240.

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

King, C. G., Guyette Jr, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. *Journal of Educators Online*, *6*(1), n1.

Li, N., Kidziński, Ł., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Eds.), *Design for teaching and learning in a networked world* (pp. 197–210). Springer International Publishing.

Li, Z., Zhu, Y., & van Leeuwen, M. (2022). A survey on explainable anomaly detection. *arXiv preprint arXiv:2210.06959*.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *2008 eighth IEEE international conference on data mining*, 413–422.

López-Pernas, S., Saqr, M., & Viberg, O. (2021). Putting it all together: Combining la methods and data sources to understand students' approaches to learning programming. *Sustainability*, *13*(9).

Lu, Y., Wang, D., Meng, Q., & Chen, P. (2020). Towards interpretable deep learning models for knowledge tracing. *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*, 185–190.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, *2*(1), 2522–5839.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc.

Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5).

Mejia-Domenzain, P., Marras, M., Giang, C., & Käser, T. (2022). Identifying and comparing multi-dimensional student profiles across flipped classrooms. *International Conference on Artificial Intelligence in Education*, 90–102.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). https : / / christophm . github . io / interpretable-ml-book

Navarro, A. A. M., & Ger, P. M. (2018). Comparison of clustering algorithms for la with educational datasets. *IJIMAI*, *5*(2), 9–16.

Okoye, K., Nganji, J. T., & Hosseini, S. (2020). Learning analytics for educational innovation: A systematic mapping study of early indicators and success factors. *International Journal of Computer Information Systems and Industrial Management Applications*, *12*, 138–154.

Otero, J., Sánchez, L., Junco, L. A., & Couso, I. (2021). Analysis of students' online interactions in the covid era from the perspective of anomaly detection. *Computational Intelligence in Security for Information Systems Conference*, 305–314.

Pabalkar, V., Chanda, R., & Vaidya, A. (2023). Anomaly detection in the course evaluation process. In *Computer vision and robotics: Proceedings of cvr 2022* (pp. 85–102). Springer.

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, *54*, 1–38.

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & Education*, *144*, 103701.

Robnik-Šikonja, M., & Bohanec, M. (2018). Perturbation based explanations of prediction models. In *Human and machine learning* (pp. 159–175). Springer.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65.

Saqr, M., López-Pernas, S., Jovanović, J., & Gašević, D. (2023). Intense, turbulent, or wallowing in the mire: A longitudinal study of cross-course online tactics, strategies, and trajectories. *The Internet and Higher Education*, *57*, 100902.

Sher, V., Hatala, M., & Gašević, D. (2020). Analyzing the consistency in within-activity learning patterns in blended learning. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 1–10.

Singh, V. K., Bozkaya, B., & Pentland, A. (2015). Money Walks: Implicit mobility behavior and financial well-being. *PLOS ONE*, *10*(8), 1–17.

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

Tallón-Ballesteros, A., & Chen, C. (2020). Explainable AI: Using Shapley value to explain complex anomaly detection ML-based systems. *Machine learning and artificial intelligence*, *332*, 152.

Tiukhova, E., Reusens, M., Baesens, B., & Snoeck, M. (2022). Benchmarking conventional outlier detection methods. *International Conference on Research Challenges in Information Science*, 597–613.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice.* (pp. 277–304). Lawrence Erlbaum Associates Publishers.