BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 3. 2026

UDC 004.85:004.056.5:37.018.43
DOI: 10.62660/bcstu/3.2026.22

ISSN: 2306-4412
E-ISSN: 2708-6070

Detecting anomalies in LMS user behaviour using machine learning

Vitalii Yanishevskyi*
Postgraduate Student
Private Higher Education Establishment "European University"
03115, 16V Akademik Vernadskyi Blvd., Kyiv, Ukraine
https://orcid.org/0009-0001-5774-4778

Abstract. The study aimed to develop a theoretical-analytical model for the interpretation of atypical digital user traces in learning management systems, covering capabilities of machine learning, structure of LMS logs, and the limits of pedagogical interpretation of algorithmic signals. The methodological framework encompassed an analysis of standards and specifications for digital learning environments, event-based descriptions of learning activity, the quality of software systems, information security, privacy, and decision-making based on artificial intelligence. The study utilised theoretical analysis, a classification approach, feature-based generalisation, comparative analysis, structural-logical modelling and critical generalisation. The study established that users' digital traces are not a homogeneous array of technical records, but reflect temporal, navigational, resource-related, task-related, testing, communicative and technical interactions with the learning environment. The study demonstrated that anomalies may manifest as temporal, navigational, resource-related, task-related, test-related, communicative, technical and mixed deviations. The study determined that machine analysis requires the transformation of activity logs into features related to login frequency, session duration, transition sequences, task completion, test attempts, communicative activity and technical access parameters. The study established that the choice of machine learning methods depends on the type of anomaly, the availability of labelled data, the structure of events, and the need for interpretability. The sequence of steps from activity logs to data cleaning, refinement of the context of learning resources, feature extraction, determination of a typical profile, selection of an algorithmic approach, detection of an anomaly signal, and its pedagogical or administrative interpretation is demonstrated. The study also argued that an anomaly signal cannot automatically be interpreted as evidence of a violation, as its significance depends on data quality, user history, course structure, technical access conditions, and human verification. The practical significance of the results includes possible use of the proposed scheme to build systems for analytical support of learning, where an anomalous signal is considered as a basis for contextual verification

Keywords: digital traces; session duration; isolation-based methods; detection of outliers; sequence models

INTRODUCTION

The research relevance is determined by the fact that Learning Management Systems (LMS) have become a vital digital environment for organising the educational process, in which learning outcomes and digital traces of user behaviour are recorded: logins to the system, session duration, viewing of materials, sequence of transitions, completion of tasks, test activity and time intervals between actions. Such data has traditionally been used to assess activity or predict success, however, with the growing volume of LMS logs, the task of detecting anomalies in user behaviour is becoming increasingly relevant. This refers to atypical temporal, sequential, task-related and communicative patterns, which may indicate a risk of academic failure, automated

Article's History: Received: 01.05.2026; Revised: 28.08.2026; Accepted: 21.09.2026; Published: 07.10.2026

Suggested Citation:
Yanishevskyi, M. (2026). Detecting anomalies in LMS user behaviour using machine learning. *Bulletin of Cherkasy State Technological University*, 31(3), 22-37. doi: 10.62660/bcstu/3.2026.22.

*Corresponding author (v.yanishevskyi@hotmail.com)

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

activity, academic dishonesty, loss of engagement, or technical glitches. In this context, Machine Learning (ML) serves as the theoretical and algorithmic foundation for building behavioural profiling models, detecting deviations and generating interpretable signals. At the same time, an anomaly in an educational environment is not always a negative phenomenon, as it may reflect an individual learning trajectory, a change in the user's strategy, or technical access conditions. Therefore, detection of anomalies in the behaviour of LMS users requires not only algorithmic analysis of logs, but also a theoretical distinction between types of anomalies, indicators of digital behaviour and the limits of pedagogical interpretation of the results.

The study by M. Tsiutsiura & A. Kovalenko (2024) examined the evaluation of anomaly detection algorithms using machine learning methods. The study emphasised that the effectiveness of anomaly detection depends on the choice of algorithm, the quality of the input data, and the model's ability to distinguish atypical objects from normal observations. M. Lakhno (2026) analysed hybrid methods for processing digital traces to detect atypical behaviour in information and education systems. In this work, users' digital traces are considered as sequences of actions that can be used for machine analysis of behavioural deviations. T. Guo *et al.* (2022) systematised educational anomaly analytics through its characteristics, methods and challenges of its application. In the paper, an anomaly is presented not merely as a statistical outlier, but as a multi-level phenomenon linked to behaviour, data, learning outcomes and the characteristics of the digital educational environment. L. Jiang & N. Bosch (2022) investigated the detection of anomalies in students' online learning activity using self-supervised machine learning. The authors have shown that this approach is suitable for analysing behavioural sequences in situations where pre-labelled data on normal and abnormal activity is limited or absent. J.R. Rico-Juan *et al.* (2024) analysed the relationship between LMS usage profiles, student personality traits and learning outcomes using machine learning methods. The study highlights that a user's interaction profile with the LMS can be used as a source of indicators for assessing learning behaviour and predicting its outcomes.

Á. Hernández-García *et al.* (2024) investigated the relationship between LMS interactions and academic performance using the Learning Cycle approach. Results showed that individual digital actions should be interpreted not in isolation, but in relation to the logic of the learning cycle. M. Winter *et al.* (2024) examined behavioural trace data in an online learning environment as indicators of student engagement. The study demonstrated that digital traces are context-dependent, as the same number or frequency of actions may reflect learning activity differently depending on the course structure and the way the user interacts with the materials. T.-T. Goh (2026) investigated LMS log analytics through the role of consistency and continuity of behavioural engagement in academic success. In the study, regularity, repetition and consistency of activity are considered as substantial characteristics of digital behaviour, which can be used to form a user's behavioural profile. L. Aunimo *et al.* (2025) combined LMS data with the student register to analyse the factors influencing the duration of study. This approach demonstrates that interpretation of LMS data may require comparing the platform's internal logs with broader educational indicators, in particular academic status and learning pace. M. Khan *et al.* (2023) used machine learning models to predict student success based on LMS activity logs. The study confirmed that LMS activity logs can be converted into a set of features for building predictive and classification models in a digital learning environment. A synthesis of the cited sources provides grounds for identifying a research niche.

Existing studies examine general anomaly detection algorithms, educational anomaly analytics, or the prediction of academic success, engagement and the risk of failure based on LMS data. At the same time, there is insufficient systematic integration of theoretical explanations of behavioural anomalies among LMS users with the classification of digital traces, the description of atypical activity, the selection of ML method sets, and the definition of the limits of interpretation of algorithmic results. Furthermore, an anomaly in an LMS should not automatically be equated with a violation, as it may be the result of an individual learning trajectory, technical access conditions, a change in educational strategy, or the specific features of the course organisation. The study aimed to develop a theoretical-analytical framework for interpreting atypical digital traces of LMS users, incorporating the structure of learning logs, the capabilities of machine learning methods, and the limitations of pedagogical and administrative interpretation of algorithmic signals. To achieve this aim, the following tasks have been identified: to reveal the essence of the LMS as a source of behavioural educational data; to systematise the types of anomalies and indicators that can be used for machine analysis of LMS logs; to summarise the capabilities of the main groups of ML approaches and to formulate a theoretical framework for interpreting anomalies in a digital learning environment.

MATERIALS AND METHODS

The research was based on international standards, technical specifications and software documentation governing the presentation of educational data, the quality of software systems, information security, user privacy, as well as approaches to the management of artificial intelligence systems and the detection of anomalies in data. To interpret LMS environments as

structured digital platforms, the ISO/IEC 40180:2017 standard (2017) is used, which defines general principles for the description, management and evaluation of e-learning systems. It was used to formalise LMS as an environment for collecting and organising digital traces of learning activity. To unify the description of learning resources and ensure the correct interpretation of metadata, ISO/IEC 19788-1:2024 (2024) was applied, which defines the structure of metadata for learning objects. This covered the context in which users' digital traces are generated. To model learning activity as a sequence of events, the Calliper Analytics specifications (1EdTech Consortium, Inc., n.d.) and IEEE 9274.1.1-2023 (2023) were used. These were used as they provide an event-based model for recording user interactions with the learning environment, which was substantial for analysing temporal and behavioural anomalies.

Quality requirements for software systems within which anomaly detection modules may be implemented are based on ISO/IEC 25010:2023 (2023). This defines reliability, performance and security characteristics for LMS analytical components. Issues regarding the protection of LMS logs and user privacy are regulated by the standards ISO/IEC 27001:2022 (2022) and ISO/IEC 29100:2024 (2024), which define the principles of information security management, data minimisation, access control and the pseudonymisation of behavioural records. Their use is determined by the need for the ethical and secure processing of educational data. Management requirements for artificial intelligence systems were addressed following ISO/IEC 42001:2023 (2023), which identified principles for the responsible use of AI/ML models in the context of educational analytics. To distinguish between anomaly detection tasks, official documentation from Scikit-Learn Developers (n.d.) on novelty detection and outlier detection, which defines the differences between the detection of new, anomalous and statistically rare observations in the data, was used.

Theoretical analysis was applied to interpret the digital traces of LMS users as a multi-level source of behavioural data. LMS logs were considered as a set of events reflecting the user's temporal, navigational, performance-related, testing, communicative and technical interactions with the learning environment. Symbolic generalisation was used to identify parameters suitable for machine analysis, in particular: the frequency of logins to the system, the duration and variability of sessions, intervals between actions, sequence of transitions, number of material views, proportion of completed tasks, test activity and level of communication. A classification approach was applied to develop a typology of anomalies in LMS data based on the following criteria: source of deviation, type of digital trace, nature of the disruption to the event sequence, temporal or outcome parameter, as well as the connection to learning resources or technical conditions. The following anomalies were identified: temporal, navigational, resource-related, task performance anomalies, test-related, communicative, technical, and mixed anomalies.

A comparative analysis was used to match groups of machine learning methods with types of anomalies. The study examined supervised, unsupervised, semi-supervised, and self-supervised approaches, as well as clustering, outlier detection, isolation-based methods, one-class classification, autoencoders and sequence models based on criteria such as data labelling requirements, feature type, ability to handle time series, and the interpretability of results. Structural-logical modelling was applied to construct the process of detecting anomalies in LMS data as a sequence of stages: log collection, data cleaning, normalisation of timestamps, feature extraction, construction of a behaviour profile, application of ML models, detection of anomalous signals and their subsequent interpretation. Critical generalisation is used to define the limits of interpreting the results of algorithmic anomaly detection, since atypical behaviour may be caused not only by deviations, but also by individual learning trajectories, technical access conditions, changes in interaction strategies or the specific features of the organisation of the learning process.

RESULTS

LMS users' digital footprints as a source of behavioural data

An analysis of the digital footprints of LMS users has revealed that their analytical value is determined not only by the fact that an event has been recorded, but also by the source of its formation, the type of action, the connection to the learning resource, time parameters, position in the course sequence and the ability to detect atypical activity. Based on an event-based description of learning activity, metadata of learning resources and technical parameters of interaction with the platform, digital traces were grouped by origin, function and potential value for machine analysis. This comparison showed that LMS logs do not form a homogeneous array of technical records, but reflect different levels of learning behaviour in a digital environment. Some sets of traces record the temporal organisation of activity; others – navigation within the course, interaction with learning resources, task completion, test behaviour, communicative participation or technical access conditions. Consequently, LMS digital traces should be viewed not as isolated events, but as interrelated behavioural indicators, the analytical significance of which depends on the type of action, the learning object, the timestamp, the sequence of interactions, and the course context. Generalised groups of LMS user digital traces and their analytical interpretation are presented in Table 1.

Table 1. Groups of digital footprints left by LMS users and their analytical interpretation

Digital footprints	Examples of LMS events	Key fixing parameters	Potential value for anomaly detection
Traces of time	Login, logout, session start and end, pauses between actions	Timestamp, session duration, interval between events, circadian rhythm of activity	Can detect sessions that are too short or excessively long, night-time activity, sudden changes in rhythm, or prolonged breaks
Navigation tracks	Navigating between modules, opening pages, returning to resources, skipping course elements	Sequence of events, user journey, number of clicks, frequency of the journey	Can detect erratic navigation, the bypassing of key resources, or unusually rapid completion of sections
Resource traces	Viewing lectures, opening files, downloading materials, revisiting items	Type of learning resource, number of views, engagement time, access status	May indicate a superficial reading, a sharp decline in interest, or excessive repetition of the same actions
Evidence of task completion	Viewing the assignment, starting completion, uploading the answer, editing, submitting work	Status, start and end times, number of edits, adherence to deadlines	These features can detect bulk uploading of answers, delayed activity, or completion without prior engagement with the material
Test marks	Start of the test, answer to a question, change of answer, end of the attempt, score achieved	Trial number, response time, test duration, result, pattern of response changes	Can flag instances where a test is completed too quickly, an unusually high number of attempts, a sudden improvement in results, or same answers are given repeatedly
Communicative traces	Forum posts, replies, comments, messages, participation in group discussions	Type of communicative event, frequency of interaction, addressee, response time	May indicate social isolation, sharp decline in communication, excessive activity, or a mismatch between academic and social engagement
Technical traces	Device changes, IP address changes, repeated requests, access errors, session interruptions	Device/IP metadata, error code, session stability, request repeatability	Can differentiate between behavioural anomalies and technical glitches, unstable connections or automated requests

Source: ompiled by the author based on ISO/IEC 40180:2017 (2017), IEEE 9274.1.1-2023 (2023), ISO/IEC 19788-1:2024 (2024), 1EdTech Consortium, Inc. (n.d.)

An analysis of Table 1 showed that LMS digital traces perform various analytical functions in detecting anomalous behaviour. Their value is determined not only by the number of recorded events, but also by the combination of event type, timestamp, learning object, execution status and location of action within the overall sequence of interactions. Temporal traces form an impression of the rhythm and consistency of activity, navigational - regarding the course route, resource traces – regarding the depth of engagement with content, task completion traces – regarding the organisation of independent work, assessment – regarding the nature of assessment activities, communication – regarding the level of educational interaction, and technical – regarding the conditions of access to the system. This means that a potential anomaly should be identified not by a single log entry, but by a deviation in the relationship between several parameters: session duration, sequence of events, resource type, interaction frequency, action outcome, and technical context. This approach makes it possible to avoid oversimplified interpretations of LMS logs. For example, a short session in itself is not sufficient grounds for concluding that activity is superficial, as it may be related to revisiting a resource that has already been processed, a technical interruption, or the deliberate execution of a specific action. Similarly, a high frequency of test attempts may indicate both self-assessment during learning and a risky pattern of passing the test.

The analytical significance of the groups of digital traces presented is determined by their connection to various aspects of the user's learning behaviour in the LMS. Following ISO/IEC 40180:2017 (2017), the quality of a digital learning environment is linked not only to the availability of resources, but also with the organisation of access, support for interaction and the possibility of analytical monitoring of learning. Therefore, time traces reflect the regularity, intensity and temporal organisation of learning activities, whilst navigation traces characterise the logic of course progression and adherence to the intended learning trajectory. Resource traces acquire analytical significance in connection with the metadata of learning objects, since the number of views, reopening of material or access status must be interpreted covering the type of resource, its function and its place within the course structure (ISO/IEC 19788-1:2024, 2024).

Task execution traces and test traces show the relationship between user activity, the completion of learning activities and the result of an assessment or self-assessment task. It is advisable to analyse them as event records, in which the agent, type of interaction, learning object, timestamp, context and outcome of the event are key (1EdTech Consortium, Inc., n.d.;

IEEE 9274.1.1-2023, 2023). Communicative traces reflect the user's social and learning-communicative engagement, including the frequency of responses, participation in forums and the nature of interaction with other course participants. Technical traces record the conditions of access to the LMS, session stability, changes in device or IP address and access errors, which distinguish actual behavioural deviations from technical glitches. Thus, LMS digital traces acquire the significance of behavioural data only when analysed as event sequences, linked to the educational context, rather than as isolated technical records.

Classification criteria and types of anomalies in LMS user behaviour

Classification of behavioural anomalies in the LMS showed that these anomalies cannot be reduced to a single type of "atypical activity", but are formed at various levels of the user's digital interaction with the learning environment. Their manifestation depends on the group of events in which the deviation occurs: in the temporal structure of activity, the course progression path, work with learning resources, task completion, test behaviour, communication, or technical access parameters. For this reason, an anomaly should be considered not merely as a deviation in the value of an individual metric, but also in terms of the nature of the disruption to the expected sequence of events, frequency of interactions, duration of actions, effectiveness or technical context. The typology of anomalies made it possible to distinguish the main manifestations of atypical activity among LMS users according to groups of digital traces: temporal, navigation, resource, task completion, test, communication, technical and mixed. This is relevant for the further application of ML methods, as different types of anomalies require different features, varying depths of contextual analysis, and varying levels of pedagogical validation. A generalised typology of behavioural anomalies among LMS users is presented in Table 2.

Table 2. Typology of behavioural anomalies among LMS users

Type of anomaly	Manifestation in LMS events	Possible analytical signal
Temporal anomaly	Unusual session duration, night-time activity, abrupt interruptions, sudden spikes or drops in activity	Deviations from the usual rhythm of academic work, changes in the frequency of engagement with the course
Navigation anomaly	Inconsistent unlocking of modules, skipping of key resources, chaotic transitions, and unusually rapid progression through sections	Deviations from the expected navigation path and the logic of the learning trajectory
Resource anomaly	Failure to review required materials, excessive repetition of the same findings, and a sharp decline in the use of resources	Discrepancy between access to learning content and the expected level of engagement within the course
Anomaly in task completion	Large volume of tasks in a short space of time, completion without prior engagement, a sudden change in presentation style	An atypical completion pattern, or a breakdown in the link between preparatory activity and the outcome
Test anomaly	Completing the test too quickly, an unusually high number of attempts, a repetitive pattern of answers, a sudden improvement in the result	Deviations in assessment attempts, response times, number of attempts or trends in results
Communication disorder	Sudden drop-in forum activity, an excessive number of posts, generic replies, sudden change in communication style	Discrepancy between communication events and the typical level of socio-educational participation
Technical anomaly	Frequent connection errors, changes to the device or IP (Internet Protocol), repeated requests, session interruptions, and unfinished actions	Atypical device/IP metadata pattern or signs of unstable interaction with the system
Mixed anomaly	Combination of time, navigation, test, performance and technical deviations	Complex outlier signal that occurs simultaneously across several groups of LMS events

Source: compiled by the author based on IEEE 9274.1.1-2023 (2023), 1EdTech Consortium, Inc. (n.d.), Scikit-Learn Developers (n.d.)

Analysis of Table 2 showed that behavioural anomalies in the LMS are of varying nature and cannot be reduced to a single activity indicator. Temporal anomalies indicate a change in the rhythm of learning, navigational – deviations from the expected course route, resource – atypical interaction with learning objects, task performance anomalies – disruption of the link between preparatory activity and the presentation of results, testing – characteristics of assessment or self-assessment activity, communicative – changes in socio-educational participation, technical – conditions of access and system stability. Mixed anomalies are more complex to analyse, as they combine several types of deviations and require not only algorithmic detection, but also contextual verification. This approach is consistent with the event-driven logic of Calliper Analytics (1EdTech Consortium, Inc., n.d.) and IEEE 9274.1.1-2023 (2023), according to which learning activity is described in terms of user action, learning object, timestamp, context, and interaction outcome.

Limitations in the interpretation of anomalies are related to the fact that an atypical LMS event does not

always have the same pedagogical or administrative significance. A time anomaly may be due to an individual schedule, technical issues, time zone differences, or last-minute preparations for a deadline. A navigation anomaly does not always indicate a problem, as the user may be returning to familiar material or working through the course selectively. A resource anomaly may be linked to the use of offline materials, pre-downloading of files, or alternative sources of preparation. An anomaly in task completion requires verification of deadlines, task structure, permitted attempts, and the possibility of group or deferred completion. A testing anomaly is not automatic proof of academic misconduct, as it may be the result of practice, retakes, or prior knowledge of the material.

A communicative anomaly depends on the course format, the role of forums in learning, the lecturer's requirements, and individual participation style. Technical anomalies require distinguishing between a malfunction, an unstable connection, or a change of device and actual behavioural deviation. A mixed anomaly carries the greatest risk of misinterpretation if it is not considered in the context of the course, the user's history and the quality of the LMS logs. The resulting typology confirms that the identification of anomalies in the LMS should not be based on the "normal-abnormal" principle, but on the principle of determining the nature of the deviation and its possible origin. For example, completing a test too quickly may be a sign of risky behaviour, but without considering the number of previous attempts, the difficulty of the test, preparation time and the user's activity history, it does not provide sufficient grounds for an unambiguous conclusion. In this sense, the Scikit-Learn Developers (n.d.) on detecting outliers and anomalies highlight the distinction between a statistically atypical observation and its meaningful interpretation. Therefore, the anomaly typology serves as an intermediary between event-based LMS logs and the selection of appropriate ML approaches: it determines which groups of features should be formed, which algorithmic methods may be appropriate, and which constraints must be addressed when interpreting the results.

Compilation of a feature set of LMS logs for machine-based analysis of anomalous behaviour

Analysis of LMS logs suitable for machine-based analysis of anomalous behaviour has shown that deviations in user activity are not determined by the mere occurrence of an event, but manifest themselves through changes in its parameters: frequency, duration, sequence, effectiveness and connection to the learning object. Therefore, to detect atypical activity, it is not sufficient to consider only the number of logins or the overall level of activity within the course. The feature set should encompass temporal, navigational, resource, task completion, test, communicative, technical and composite parameters, which together form the behavioural profile of an LMS user. In machine analysis, such features serve as an intermediate level between raw LMS events and algorithmic anomaly detection. An event records what the user has done, whereas a feature reflects how this action relates to the typical learning rhythm, course structure, expected sequence of interactions and the user's previous behaviour. Therefore, identical events may have different analytical significance depending on the context: for example, reopening a learning resource may indicate either in-depth study of the material, difficulties in interpretation, or automated repetition of actions. The main features of LMS logs suitable for machine analysis of anomalous behaviour are presented in Table 3.

Table 3. Features of LMS logs for machine-based analysis of anomalous behaviour

Group of characteristics	Specific indicators for analysis	Source of formation in LMS logs	Analytical significance for ML analysis
Temporal markers	Frequency of logins, average and maximum session duration, time between actions, daily distribution of activity	Timestamp, session duration, intervals between LMS events	Determine the pace, consistency and timing of learning activities
Navigational landmarks	Number of clicks, order in which modules are opened, route repetition, frequency of returns to previous pages	Event sequence, navigation path, transitions between learning objects	Illustrate the logic of the course progression and how the user's actions align with the structure of the learning pathway
Resource-related characteristics	Number of page views, time spent on the site, download frequency, and repeat visits	Learning object metadata, resource access event, access status	Reflect the intensity and depth of engagement with the learning content
Indications of task completion	Task start time, response submission time, number of edits, deadline compliance, completion status	Assignment event, completion status, submission timestamp	Describe characteristics of independent study and the relationship between preparatory work and the outcome
Test criteria	Number of attempts, test duration, response time, changes to answers, score trends	Assessment attempt, response time, score change, answer pattern	Used for analysis of user behaviour in control and self-assessment activities

Continued Table 3.

Group of characteristics	Specific indicators for analysis	Source of formation in LMS logs	Analytical significance for ML analysis
Communicative features	Number of messages, response rate, response time, participation in forums, targeted engagement	Communication event, forum post, reply event, message timestamp	Reflect level of social and academic engagement and interaction with other course participants
Technical characteristics	Device type, IP changes, repeated requests, access errors, session interruptions, unfinished actions	Device/IP metadata, error event, request pattern, session interruption	Incorproate technical context of interaction with the LMS and distinguish it from behavioural manifestations
Complex features	Combination of frequency, duration, sequence, effectiveness and technical context	Aggregated LMS logs, combined event sequences, user profile	Create a multidimensional behavioural profile to identify complex behavioural issues

Source: compiled by the author based on IEEE 9274.1.1-2023 (2023), ISO/IEC 19788-1:2024 (2024), 1EdTech Consortium, Inc. (n.d.), Scikit-Learn Developers (n.d.)

An analysis of Table 3 showed that features for the machine analysis of LMS logs should be derived not only from quantitative indicators of activity, but from a combination of event-based, temporal, contextual and outcome-based parameters. Temporal and navigational features can describe the rhythm and route of learning behaviour, resource and task-related - the nature of work with learning objects, test – the specifics of assessment activity, communicative – socio-educational participation, and technical – conditions of access to the system. Comprehensive indicators are of the greatest importance for detecting complex anomalies, as it is precisely the combination of several deviations that categorise a random atypical action from a persistent behavioural pattern. This approach to the feature-based description of LMS logs clarifies the link between the digital footprint and algorithmic analysis. A raw log entry records only the fact of an event, whereas a feature transforms this event into a measurable parameter, suitable for comparison with an individual, group, or course profile. Therefore, the quality of features directly influences the ability of ML methods to detect anomalies without oversimplified or erroneous interpretation. For example, a high number of views of the material may be an indicator of interest, difficulties in comprehension or automated repetition of actions; therefore, it only acquires analytical significance in conjunction with session duration, the sequence of transitions, the outcome of the task, and the overall context of the course.

Suitability of machine learning methods for detecting different types of anomalies

A comparison of groups of ML methods based on their suitability for detecting anomalies in the behaviour of LMS users showed that no single approach can be considered universal for all types of behavioural deviations. The effectiveness of the method depends on the presence or absence of labelled data, the nature of LMS logs, the type of anomaly, the temporal or sequential structure of events, requirements for interpretability and the acceptable level of false positives. Consequently, supervised learning is best suited to known and labelled instances of atypical behaviour, unsupervised learning to the detection of hidden outliers and non-standard activity clusters, semi-supervised learning to the construction of a normal behaviour profile, self-supervised learning to the analysis of sequences of training actions, and autoencoders and sequence models with complex multidimensional or temporal patterns. A generalised comparison of groups of ML methods is presented in Table 4.

Table 4. Suitability of machine learning methods for detecting anomalies in the behaviour of LMS users

Group of ML methods	Need for data structuring	Most suitable types of anomalies	Analytical advantages
Supervised learning	Requires pre-defined examples of normal and abnormal behaviour	Known behavioural patterns relating to testing, performance, technical aspects or risk	Can be used to create classifiers for predefined types of anomalies and evaluate recognition quality using formal metrics
Unsupervised learning	No pre-structuring required	Unknown temporal, navigational, resource-related and mixed anomalies	Can detect hidden clusters of atypical activity, outliers and non-standard behavioural patterns
Semi-supervised learning	Primarily uses data on normal behaviour	Deviation from the individual or group profile of normal activity	Suitable for situations where anomalous cases are rare or insufficiently documented
Self-supervised learning	Does not require full manual labelling; generates training signals directly from the data	Sequential, navigational, temporal and mixed anomalies	Effective for analysing event sequences, navigation paths and complex behavioural trajectories

Continued Table 4.

Group of ML methods	Need for data structuring	Most suitable types of anomalies	Analytical advantages
Clustering	No pre-structuring required	Group differences in activity, varying course completion patterns, and atypical user clusters	Can identify behavioural groups and users who do not belong to typical clusters
Outlier detection	Usually requires no markup or uses minimal information about the standard	Isolated or rare deviations in time, test, technical and resource characteristics	Specifically designed to identify outliers in multidimensional data
Isolation-based methods	Do not require marking	Unusual combinations of characteristics, atypical technical or mixed patterns	perform well with multidimensional data and can identify objects that can be easily distinguished from the bulk of the observations
One-class classification	Requires data describing predominantly normal behaviour	Deviations from the typical user, group or course profile	Can define the boundaries of normal behaviour and record observations outside those boundaries
Autoencoders	Require a sufficient amount of data to train the model to recognise typical patterns	Complex multidimensional, performance-based, test-based and mixed anomalies	Identify anomalies due to increased reconstruction error, which is beneficial in complex behavioural profiles
Sequence models	Consistent LMS logs and a sufficient number of events are required	Temporal, navigational, testing and communication anomalies related to the sequence of actions	Can be used to analyse not only events, but also the place in the user's behavioural trajectory

Source: compiled by the author based on ISO/IEC 25010:2023 (2023), ISO/IEC 42001:2023 (2023), Scikit-Learn Developers (n.d.)

An analysis of Table 4 showed that the choice of ML method for detecting anomalies in the LMS should not be determined by the general popularity of the algorithm, but by the compatibility between the type of anomaly, the data structure, the presence of labels, the quality of features and the requirements for interpreting the result. Teacher-guided training methods are suitable for known and formalised deviation scenarios, however they are of limited use for new or rare anomalies, require high-quality labelling and may reproduce errors from previous interpretations. Unsupervised learning methods, clustering, outlier detection and isolation-based methods are better suited to the task of identifying unknown anomalies; however they require additional contextual interpretation, since statistical unusualness, a small cluster or the distance of an object in the feature space do not always have pedagogical or administrative significance.

The Scikit-Learn Developers (n.d.) on novelty and outlier detection provides grounds for distinguishing between the technical detection of atypical observations and the subsequent interpretation of their meaning. Semi-supervised learning and single-class classification can be used to establish the boundaries of normal user behaviour, a group or a course, but their quality depends on the correctness of the definition of "normal". A change in the course structure, the teaching strategy or an individual's pace of work may be mistakenly perceived as an anomaly if the model does not incorporate the context of educational interaction. Self-supervised learning, autoencoders and sequence models are suitable for complex sequential and multidimensional LMS logs, however they require a sufficient volume of high-quality data, accurate recording of event sequences and additional explanation of the causes of abnormal signals. ISO/IEC 25010:2023 (2023) focuses the analysis on the quality of the software system, in particular on reliability, usability, security and maintainability. ISO/IEC 42001:2023 (2023) establishes a management framework for AI/ ML solutions, in which controllability, accountability, transparency and risk assessment are key.

The comparison confirmed that the detection of anomalies in the behaviour of LMS users should be based on a context-dependent choice of algorithmic approach. For temporal and navigational deviations, methods capable of working with sequences, activity patterns and the order of actions are substantial. For test and task performance anomalies, methods that take into account the combination of time, number of attempts, the dynamics of results, and the relationship between preparatory activity and task performance are suitable. For technical and mixed anomalies, approaches for detecting outliers in multidimensional features are suitable, but their results must be verified taking into account the quality of LMS logs, session stability, technical access conditions and course structure. Therefore, the ML method in LMS analytics should not serve as a final decision-making tool, but rather serve as a tool for identifying an anomaly that subsequently requires pedagogical or administrative review.

Sequence of the algorithmic detection of anomalies in LMS user behaviour

A review of the stages involved in working with LMS behavioural data has shown that anomaly detection is

not simply a matter of directly applying an ML algorithm to the event log. The effectiveness of such a process is determined by a sequential transition from the initial recording of learning activity to the formation of an interpreted anomaly signal. At the initial level, LMS logs contain events such as logins, resource views, transitions between modules, task completion, test attempts, communicative interaction, and technical access. However, these records vary in completeness, temporal structure, event density and dependence on the context of the learning resource, therefore before algorithmic analysis they require cleaning, normalisation of timestamps, removal of duplicates and alignment with course metadata. The study established that it is the feature extraction stage that transforms the technical log into a behavioural profile: login frequency, session duration, intervals between actions, sequence of transitions, number of test attempts, task completion status and communicative activity acquire analytical significance only in relation to an individual or group user profile. The subsequent choice of ML method depends on the type of anomaly, the presence of annotations, the structure of features, and the need for interpretability of the result. Therefore, the detection of anomalies in an LMS should be structured as a multi-stage process, in which the algorithmic signal is not a conclusion but an interim result requiring pedagogical or administrative verification. A generalised structural-logical diagram of anomaly detection in LMS user behaviour using ML is shown in Figure 1.

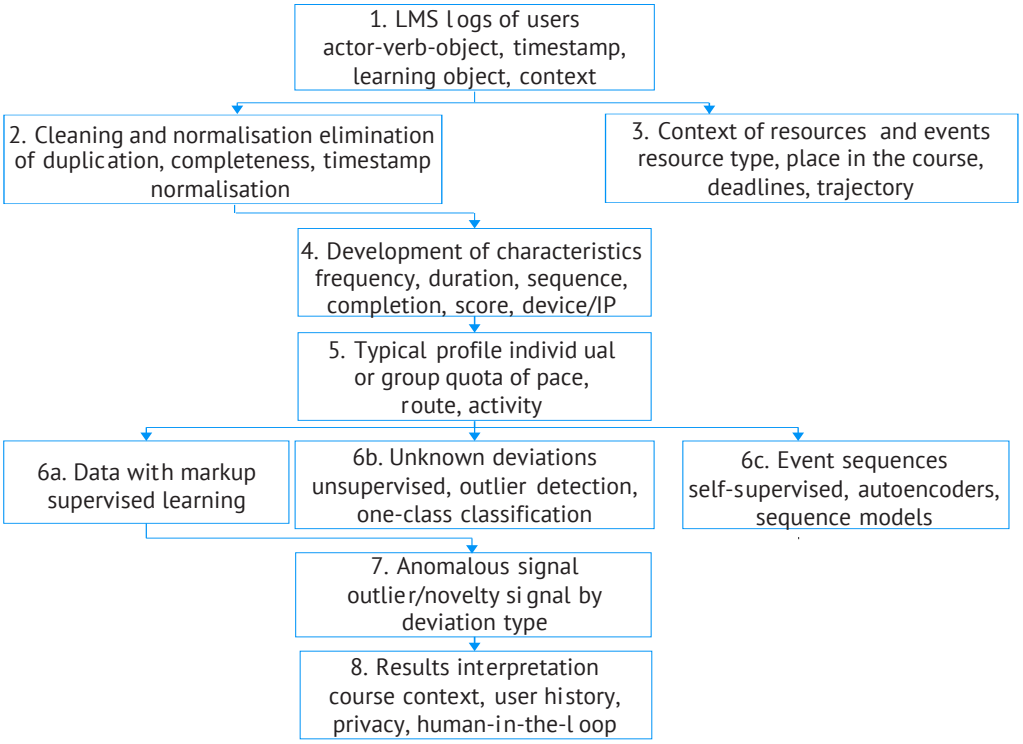


Figure 1. Structural-logical diagram for detecting anomalies in the behaviour of LMS users using machine learning

Source: compiled by the author based on ISO/IEC 40180:2017 (2017), ISO/IEC 27001:2022 (2022), IEEE 9274.1.1-2023 (2023), ISO/IEC 25010:2023 (2023), ISO/IEC 42001:2023 (2023), ISO/IEC 29100:2024 (2024), ISO/IEC 19788-1:2024 (2024), 1EdTech Consortium, Inc. (n.d.)

Diagram shown in Figure 1 illustrates the sequence of steps involved in converting raw LMS logs into an interpreted anomaly signal. Initially, events such as logins, resource views, module transitions, task execution, test attempts, communication and technical access are filtered, normalised and aligned with the course context. This makes it possible to filter out incomplete records, duplicates, technical glitches and incorrect timestamps from behavioural manifestations suitable for further analysis. In the next stage, event records are converted into indicators of behavioural activity: frequency of actions, session duration, sequence of transitions, task completion status, performance trends and device metadata or IP addresses. Based on these features, an individual or group profile is formed, against which deviations are assessed. The choice of ML method depends on the type of anomaly, the data structure, the availability of labelling, and the need for interpretability of the result. The result of algorithmic processing is a signal of atypicality or novelty, which cannot automatically be interpreted as evidence of a breach, failure or academic misconduct. It must be assessed based on

the course structure, user history, log quality, technical access conditions and privacy requirements. Therefore, the accuracy of anomaly detection in the LMS is determined not only by the choice of ML algorithm, but also by the completeness of the data, the quality of pre-processing, the relevance of features, and human verification of the algorithmic signal.

Limits of interpreting algorithmically detected anomalies

An assessment of the interpretability of algorithmically detected anomalies has shown that the results of ML analysis in LMS cannot be regarded as sufficient evidence of a violation, academic misconduct or a user's academic failure. An anomaly signal detects a deviation from the expected profile, but does not explain the cause of this deviation without taking into account the course context, the quality of LMS logs, technical access conditions, the user's previous history and the pedagogical context. Therefore, interpretation of anomalies must be based on criteria that distinguish statistical atypicality from pedagogically significant risk, technical failure or an acceptable individual learning trajectory. Generalised criteria for interpreting algorithmically detected anomalies in the LMS are presented in Table 5.

Table 5. Criteria for interpreting algorithmically detected anomalies in the LMS

Interpretation criterion	Analytical significance	Risk of misinterpretation	Verification method
Completeness of LMS logs	Determines whether enough events have been recorded to assess a behavioural pattern	Incomplete records may provide a false impression of inactivity, a sudden change in activity, or the omission of learning stages	Checking for missed events, duplicates, technical interruptions and the accuracy of session recordings
Data quality and consistency	Shows how accurately the timestamp, event type, user identifier, learning object and completion status have been recorded	Incorrect timestamps or identification errors can turn a technical glitch into a "behavioural" anomaly	Normalisation of timestamps, verification of event formats, comparison with course metadata
Context of educational resource	Can interpret the activity following type of material, its complexity, whether it is compulsory, and its place within the course	Brief interaction with a simple or re-visited resource may be mistakenly perceived as superficial learning	Mapping the event to the resource type, its function, complexity and the sequence of the course
User's personal story	Can compare current activity with previous personal profile	Atypical behaviour for the group may be consistent and normal for a particular user	Analysis of previous sessions, usual work routine, individual pace and track record
Group or course profile	Can determine whether a user deviates from the typical patterns of a group or course	Group "norm" may not be relevant to users with a different pace, schedule or learning strategy	Comparison with the group's aggregated patterns, but without automatically equating deviations with violations
Frequency of the abnormal signal	Distinguishes a single random event from a consistent behavioural pattern	One-off anomaly may be the result of chance, a technical error or situational activity	Analysis of signal recurrence across different sessions, modules, task types or time periods
Technical access requirements	Verifies whether the issue is related to the device, network, browser, IP changes or access errors	Session interruptions, repeated requests or a change of device may be mistakenly interpreted as risky behaviour	Verification of device/IP metadata, access errors, session stability and the recurrence of technical faults
Privacy and access control	Defines the limits of permissible use of users' behavioural data	Excessive data collection or data granularity may violate the principles of data minimisation, purpose limitation and confidentiality	Anonymisation or pseudonymisation of records, restriction of access, data minimisation and specification of the purpose of processing
Interpretability of an algorithmic signal	Indicates whether it is possible to determine which factors contributed to the detection of the anomaly	Opaque ML result may cause an unfounded educational or administrative decision	Use of interpretable features, explanation of parameter contributions, and human signal verification
Human verification	Ensures final alignment of the algorithmic signal with the educational context	Automated decision made without verification could cause false accusations or inappropriate intervention	Review by a lecturer, administrator or designated person, covering the context of the course and the user's history

Source: compiled by the author based on ISO/IEC 27001:2022 (2022), ISO/IEC 42001:2023 (2023), ISO/IEC 25010:2023 (2023), ISO/IEC 29100:2024 (2024)

An analysis of Table 5 showed that the limits of interpretation of algorithmically detected anomalies are determined not only by the properties of the ML method, but also by data quality, the context of the training resource, the stability of the behavioural profile, technical access conditions, and rules for handling personal and behavioural data. At the data level, critical factors include the completeness of LMS logs, the accuracy of timestamps, the consistency of event records and the presence of metadata for learning objects. At the interpretation level, the key factors are the repeatability of an anomalous signal, the comparison of individual and group profiles, the explainability of the result, and human verification. At the data management level, privacy, access control, data minimisation, and preventing automated decision-making without contextual verification remain substantial.

Thus, an anomaly detected algorithmically in the LMS has the status of an indicator of atypical behaviour, rather than a definitive conclusion about user behaviour. The same signal may have different meanings depending on the course structure, the complexity of the learning resource, the user's previous activity, technical operating conditions, and the organisation of the educational process. For example, completing a test quickly may indicate a suspicious pattern, but it may also be the result of a repeat attempt, good preparation or the simplicity of the task; a sudden change in IP address may be a technical feature of the connection, rather than a sign of a breach. Therefore, the correct interpretation of anomalies requires a combination of algorithmic signals, contextual data, and human-in-the-loop verification, in which the final value of the deviation is determined not automatically, but by verifying its causes, repeatability and pedagogical relevance.

DISCUSSION

The results obtained showed that the detection of anomalies in the behaviour of users of learning management systems did not amount to automatic prediction of academic performance or a formal search for statistical outliers, but depended on the sequential relationship between digital traces, indicators, the type of behavioural deviation, a set of machine learning methods, and contextual interpretation of the signal. In the meta-analysis by K. Fahd *et al.* (2022), the application of machine learning in higher education was associated with the assessment of academic performance, the risk of failure, and student dropout. This was consistent with the findings in that user activity logs were considered as a source of early analytical signals, but at the same time the results clarified a broader interpretation: an anomaly in the learning management system did not necessarily imply low academic performance or a risk of expulsion, but could reflect technical access conditions, a change in learning strategy or an individual pace of work.

M. Adnan *et al.* (2021) conducted identification of at-risk students at various stages of the course, which highlighted the significance of early intervention based on partial data regarding academic engagement. This was consistent with the conclusion regarding the importance of temporal features, session duration, intervals between actions and changes in activity patterns, however the findings emphasised that an early signal should undergo a contextual review before any pedagogical or administrative decision. In the study by A. Gonzalez-Nucamendi *et al.* (2023), predictive analytics was used to identify students at risk of dropping out, due to the link between academic performance and behavioural data. This approach aligned with a typology of anomalies, where performance and task deviations were considered as potential risk signals; however, the results did not equate these signals with a definitive conclusion regarding the user's failure. In the study by M.N. Alsubaie (2023), the prediction of student success on an online platform was linked to improving the quality of digital training and the use of machine learning to support management decisions. This was consistent with the view that the analytical component in an educational environment should be assessed not only on the basis of its ability to identify deviations, but also on the quality of the software system, the validity of the features, and the suitability of the results for educational interpretation. N. Abuzinadah *et al.* (2023) examined the use of convolutional features and machine learning to predict academic performance based on Moodle data. This was consistent with findings regarding the need to transform raw activity logs into a feature space, where significance was attached not only to individual events, but also their frequency, duration, sequence, effectiveness and connection to the learning object.

In the study by C. Thangavel *et al.* (2024), student activity in Moodle was assessed using machine learning as an indicator of engagement in the learning process. The results obtained were partly consistent with this approach, but clarified that activity alone could not be a sufficient explanation for user behaviour, since a high or low frequency of actions could have different meanings depending on the course structure, the type of task, the user's previous history and technical access conditions. In the study by M. Yağcı (2022), educational data analytics was used to predict academic performance using various machine learning algorithms. This was consistent with the comparative matrix of method groups obtained, which showed that the choice of algorithmic approach should depend not on its general popularity, but on the type of anomaly, the data structure, the presence of annotations and the requirements for the interpretability of the result. S. Dass *et al.* (2021) predicted dropout rates in a self-directed massive open online course using a random forest model. This was consistent with the finding regarding the importance of time and resource characteristics in environments

where the user regulates the pace of the course themselves, but at the same time the results demonstrated the need to distinguish an acceptable individual trajectory from a genuinely risky or anomalous pattern.

Particularly relevant to the findings were studies in which learning behaviour was examined through Moodle logs, sequences of events, and the temporal dimension of interaction. K. Dobashi *et al.* (2022) classified learning patterns based on Moodle logs and visualised browsing processes using a time-cross-sectional representation. This was consistent with findings where navigational, temporal and resource traces were presented as interrelated elements of a behavioural profile, rather than as isolated log events. D. Rotelli & A. Monreale (2023) analysed the processing and interpretation of Moodle logs, considering their temporal dimension. This approach directly corresponded to the conclusion reached regarding the need to normalise timestamps, verify the completeness of records, account for intervals between actions, and avoid simplistic interpretations of short or interrupted sessions. M.R. Borna *et al.* (2024) investigated click data using artificial intelligence to predict academic success and assess learning. This was consistent with the definition of digital traces as event sequences, in which clickstream behaviour only acquired analytical significance after being correlated with the type of learning resource, the sequence of transitions, and the outcome of the action. C. Kaensar & W. Wongnin (2023) analysed and predicted student performance based on Moodle log data using machine learning methods. This was consistent with the findings regarding features for the machine analysis of activity logs, but the established results broadened the focus from predicting student success to identifying various types of anomalies: temporal, navigational, resource-related, test-related, communicative, technical and mixed.

H.A. Althibyani (2024) conducted a comprehensive analysis of predicting student success in massive open online courses using machine learning models. This approach was consistent with the conclusion regarding the significance of digital behaviour in large online environments, but the results highlighted that, in the case of anomalies, it was not only predictive metrics that were substantial, but also the interpretability of the signal, the repeatability of the deviation, and the possibility of human verification. D. Vergara *et al.* (2024) demonstrated in a bibliometric review the growing role of artificial intelligence in learning management systems. This was consistent with the conclusion that analytical solutions in such systems should be viewed not merely as computational tools, but as components of a digital educational environment requiring management of quality, privacy and interpretative accountability. P. Ngulube & M.M. Ncube (2025) considered learning analytics as a means of improving the user experience in learning management systems. This was consistent with the findings regarding the limits of interpreting anomalies, as the algorithmically detected signal was not to be used as a basis for punitive or automated disciplinary action, but rather to support a more accurate interpretation of user behaviour within the learning environment. H. Li & Z. Liu (2025) analysed student behaviour and academic performance within an intelligent educational system based on multi-source data. This was consistent with the conclusion that interpreting an anomaly signal required comparing activity logs with the user's individual history, group profile, course context, and technical access parameters. Y. Liu *et al.* (2023) predicted student performance using clickstream data and machine learning. This confirmed the importance of event and navigation features; however, the results clarified that clickstream data could not be interpreted solely as a quantitative measure of activity, as the sequence of actions, route repetition, and deviations from the expected trajectory were no less significant than the number of clicks itself.

B. Alnasyan *et al.* (2024) demonstrated in a systematic review the potential of deep learning techniques for predicting success in virtual learning environments. This was consistent with the results obtained for autoencoders and sequence models, which were identified as promising for complex multidimensional or sequential activity logs, but at the same time such methods exacerbated the problem of interpretability and the risk of opaque pedagogical interpretation. M. Riestra-González *et al.* (2021) analysed massive learning management system logs for early prediction of academic performance regardless of the course. This was consistent with the conclusion regarding the importance of log completeness, event aggregation and the construction of a group or course profile, but the results also highlighted that generalising the profile required caution, as the group norm might be irrelevant to a user with an individual pace or a non-standard learning strategy. M.M. Tamada *et al.* (2021) predicted student success based on Moodle logs within an engineering education context. This was consistent with the conclusion regarding the suitability of Moodle logs for machine analysis; however, the established typology of anomalies showed that the same logs can be used not only for predicting results, but also to detect atypical temporal, navigational, testing, communicative and technical patterns.

A comparison with cited studies demonstrated that the results obtained supported the current trend of moving away from simple prediction of academic performance towards a broader analysis of users' digital behaviour in learning environments. In the studies examined, activity logs, Moodle data, clickstream events, engagement metrics and machine learning models were most frequently used to predict academic performance, the risk of leaving education or learning activity. In this context the results obtained highlighted another area of application for such data: detecting anomalies

required not only the construction of a predictive model, but also a typology of deviations, a feature-based description of logs, the selection of a set of methods according to the type of anomaly and mandatory contextual interpretation. Therefore, the detection of anomalies in the behaviour of users of learning management systems should be regarded as a multi-level process, in which machine learning detects a signal of atypical behaviour, and its significance is determined by data quality, course structure, user history, technical access conditions, privacy and human verification.

CONCLUSIONS

As part of this study, a theoretical framework has been developed for an approach to detecting anomalies in the behaviour of LMS users using machine learning, based on the comparison of digital traces, types of behavioural deviations, LMS features - logs, ML groups – methods and criteria for interpreting algorithmically detected signals. The study demonstrated that LMS digital traces are not a homogeneous array of technical records, but reflect temporal, navigational, resource-related, task-related, test, communicative and technical interactions of the user with the digital educational environment. Their analytical significance is determined not by a single event, but by the relationship between the type of action, the timestamp, the learning object, the completion status, the sequence of interactions and the course context. This typology established that anomalies in an LMS can manifest as temporal, navigational, resource-related, task-related, test-related, communicative, technical, and mixed deviations, and none of these manifestations should automatically be equated with misconduct or academic dishonesty. The study determined that machine analysis of LMS logs requires the conversion of events into features that encompass the frequency of logins, session duration, intervals between actions, sequence of transitions, number of resource views, task completion status, test attempts, performance trends, communication activity and technical access parameters.

A comparison of groups of ML methods showed that supervised learning is suitable for known and labelled outlier scenarios, whilst unsupervised learning, clustering, outlier detection and isolation-based methods – for identifying unknown anomalies, semi-supervised learning and one-class classification – for detecting deviations from a normal profile, and self-supervised learning, autoencoders and sequence models – for complex sequential and multidimensional patterns. The structural-logical diagram specified that anomaly detection should follow the sequence "LMS-logic data cleaning and normalisation context of training resources and events feature extraction typical profile", after which the choice of ML approach depends on the nature of the data: labelled data is suited to supervised learning, unknown deviations to unsupervised learning, outlier detection or one-class classification, whilst event sequences point towards self-supervised learning, autoencoders or sequence models. The final stages involve recording the anomalous signal and its pedagogical or administrative interpretation. The study established that an anomaly detected algorithmically has the status of an indicator of atypicality, rather than a definitive conclusion about user behaviour. A correct interpretation of this requires consideration of the completeness and quality of logs, the metadata of learning resources, signal repeatability, individual user history, group profile, technical access conditions, privacy, result interpretability, and human-in-the-loop verification. Further research should conduct an empirical validation of the proposed framework using real LMS logs, comparing the accuracy, interpretability and stability of different groups of ML methods for specific types of behavioural anomalies.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None

REFERENCES

[1] 1EdTech Consortium, Inc. (n.d.). *Caliper analytics*. Retrieved from https://www.1edtech.org/standards/caliper.

[2] Abuzinadah, N., Umer, M., Ishaq, A., Al Hejaili, A., Alsubai, S., Eshmawi, A.A., Mohamed, A., & Ashraf, I. (2023). Role of convolutional features and machine learning for predicting student academic performance from MOODLE data. *PLoS ONE*, 18(11), article number e0293061. doi: 10.1371/journal.pone.0293061.

[3] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A.A., Abid, M., Bashir, M., & Khan, S.U. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access,* 9, 7519-7539. doi: 10.1109/ACCESS.2021.3049446.

[4] Alnasyan, B., Basheri, M., & Alassafi, M. (2024). The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence,* 6, article number 100231. doi: 10.1016/j.caeai.2024.100231.

[5] Alsubaie, M.N. (2023). Predicting student performance using machine learning to enhance the quality assurance of online training via Maharat platform. *Alexandria Engineering Journal*, 69, 323-339. doi: 10.1016/j.aej.2023.02.004.

[6] Althibyani, H.A. (2024). Predicting student success in MOOCs: A comprehensive analysis using machine learning models. *PeerJ Computer Science*, 10, article number e2221. doi: 10.7717/peerj-cs.2221.
[7] Aunimo, L., Kauttonen, J., Vahtola, M., & Huttunen, S. (2025). Combining data from a LMS and a student register for exploring factors affecting study duration. *Journal of Computing in Higher Education*, 37, 1032-1055. doi: 10.1007/s12528-024-09414-4.
[8] Borna, M.R., Saadat, H., Hojjati, A.T., & Akbari, E. (2024). Analyzing click data with AI: Implications for student performance prediction and learning assessment. *Frontiers in Education,* 9, article number 1421479. doi: 10.3389/feduc.2024.1421479.
[9] Dass, S., Gary, K., & Cunningham, J. (2021). Predicting student dropout in self-paced MOOC course using random forest model. *Information*, 12(11), article number 476. doi: 10.3390/info12110476.
[10] Dobashi, K., Ho, C.P., Fulford, C.P., Lin, M.F.G., & Higa, C. (2022). Learning pattern classification using moodle logs and the visualization of browsing processes by time-series cross-section. *Computers and Education: Artificial Intelligence,* 3, article number 100105. doi: 10.1016/j.caeai.2022.100105.
[11] Fahd, K., Venkatraman, S., Miah, S.J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, 27, 3743-3775. doi: 10.1007/s10639-021-10741-7.
[12] Goh, T.-T. (2026). Learning management system log analytics: The role of persistence and consistency of engagement behaviour on academic success. *Journal of Computers in Education*, 13, 283-306. doi: 10.1007/s40692-025-00358-x.
[13] Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., & García-Castelán, R.M.G. (2023). Predictive analytics study to determine undergraduate students at risk of dropout. *Frontiers in Education,* 8, article number 1244686. doi: 10.3389/feduc.2023.1244686.
[14] Guo, T., Bai, X., Tian, X., Firmin, S., & Xia, F. (2022). Educational anomaly analytics: Features, methods, and challenges. *Frontiers in Big Data*, 4, article number 811840. doi: 10.3389/fdata.2021.811840.
[15] Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A learning cycle approach. *Computers in Human Behavior,* 155, article number 108183. doi: 10.1016/j.chb.2024.108183.
[16] IEEE 9274.1.1-2023. (2023). *IEEE standard for learning technology – JavaScript Object Notation (JSON) Data Model Format and Representational State Transfer (RESTful) web service for learner experience data tracking and access*. Retrieved from https://standards.ieee.org/ieee/9274.1.1/7321/.
[17] ISO/IEC 19788-1:2024. (2024). *Information technology for learning, education and training – metadata for learning resources. Part 1: Framework*. Retrieved from https://www.iso.org/standard/81950.html.
[18] ISO/IEC 25010:2023. (2023). *Systems and software engineering – systems and software Quality Requirements and Evaluation (SQuaRE) – product quality model*. Retrieved from https://www.iso.org/standard/78176.html.
[19] ISO/IEC 27001:2022. (2022). *Information security, cybersecurity and privacy protection – information security management systems – requirements*. Retrieved from https://www.iso.org/standard/27001.
[20] ISO/IEC 29100:2024. (2024). *Information technology – security techniques – privacy framework*. Retrievev from https://www.iso.org/standard/85938.html.
[21] ISO/IEC 40180:2017. (2017). *Information technology – quality for learning, education and training – fundamentals and reference framework*. Retrieved from https://www.iso.org/standard/62825.html.
[22] ISO/IEC 42001:2023. (2023). *Information technology – artificial intelligence – management system*. Retrieved from https://www.iso.org/standard/42001.
[23] Jiang, L., & Bosch, N. (2022). Mining and assessing anomalies in students' online learning activities with self-supervised machine learning. In A. Mitrovic & N. Bosch (Eds.), *Proceedings of the 15th international conference on educational data mining* (pp. 549-554). Durham: International Educational Data Mining Society. doi: 10.5281/zenodo.6852947.
[24] Kaensar, C., & Wongnin, W. (2023). Analysis and prediction of student performance based on Moodle log data using machine learning techniques. *International Journal of Emerging Technologies in Learning*, 18(10), 184-203. doi: 10.3991/ijet.v18i10.35841.
[25] Khan, M., Naz, S., Khan, Y., Zafar, M., Khan, M., & Pau, G. (2023). Utilizing machine learning models to predict student performance from LMS activity logs. *IEEE Access*, 11, 86953-86962. doi: 10.1109/ACCESS.2023.3305276.
[26] Lakhno, M. (2026). Experimental evaluation of the effectiveness of hybrid methods for digital footprint analysis in detecting atypical behavior in information and educational systems. *Electronic Professional Scientific Journal "Cybersecurity: Education, Science, Technique",* 4(32), 119-126. doi: 10.28925/2663-4023.2026.32.1109.
[27] Li, H., & Liu, Z. (2025). An intelligent educational system: Analyzing student behavior and academic performance using multi-source data. *Electronics*, 14(16), article number 3328. doi: 10.3390/electronics14163328.

[28] Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., & Wei, Y. (2023). Predicting student performance using clickstream data and machine learning. *Education Sciences*, 13(1), article number 17. doi: 10.3390/educsci13010017.
[29] Ngulube, P., & Ncube, M.M. (2025). Leveraging learning analytics to improve the user experience of learning management systems in higher education institutions. *Information,* 16(5), article number 419. doi: 10.3390/info16050419.
[30] Rico-Juan, J.R., Cachero, C., & Macia, H. (2024). Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques. *Applied Intelligence*, 54, 6175-6197. doi: 10.1007/s10489-024-05483-1.
[31] Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education,* 163, article number 104108. doi: 10.1016/j.compedu.2020.104108.
[32] Rotelli, D., & Monreale, A. (2023). Processing and understanding Moodle log data and their temporal dimension. *Journal of Learning Analytics,* 10(2), 126-141. doi: 10.18608/jla.2023.7867.
[33] Scikit-Learn Developers. (n.d.). *2.7. Novelty and outlier detection*. Retrieved from https://scikit-learn.org/stable/modules/outlier_detection.html.
[34] Tamada, M.M., Giusti, R., & de Magalhães Netto, J.F. (2021). Predicting student performance based on logs in moodle LMS. In *2021 IEEE frontiers in education conference* (pp. 1-8). Lincoln: IEEE. doi: 10.1109/FIE49875.2021.9637274.
[35] Thangavel, C., Valliammai, S.E., Amritha, P.P., Chandran, K., Chowdhury, S., Nguyen, T.T., Bao, Q.B., Tran, D.-T., Tran, D.-N., & Trang, D.Q. (2024). Students' activeness measure in Moodle Learning Management System using machine learning. *Journal of Applied Engineering and Technological Science,* 6(1), 744-753. doi: 10.37385/jaets.v6i1.6128.
[36] Tsiutsiura, M., & Kovalenko, A. (2024). Evaluation of anomaly detection algorithms using machine learning methods. *Management of Development of Complex Systems*, 58, 80-85. doi: 10.32347/2412-9933.2024.58.80-85.
[37] Vergara, D., Lampropoulos, G., Antón-Sancho, Á., & Fernández-Arias, P. (2024). Impact of artificial intelligence on learning management systems: A bibliometric review. *Multimodal Technologies and Interaction*, 8(9), article number 75. doi: 10.3390/mti8090075.
[38] Winter, M., *et al*. (2024). Behavioral trace data in an online learning environment as indicators of learning engagement in university students. *Frontiers in Psychology,* 15, article number 1396881. doi: 10.3389/fpsyg.2024.1396881.
[39] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9, article number 11. doi: 10.1186/s40561-022-00192-z.

Виявлення аномалій у поведінці користувачів LMS засобами машинного навчання

Віталій Янішевський
Аспірант
Приватний вищий навчальний заклад «Європейський університет»
03115, бульв. Академіка Вернадського, 16В, м. Київ, Україна
https://orcid.org/0009-0001-5774-4778

Анотація. Метою дослідження було розробити теоретико-аналітичну модель інтерпретації нетипових цифрових слідів користувачів у системах управління навчанням із урахуванням можливостей машинного навчання, структури LMS-логів і меж педагогічного тлумачення алгоритмічних сигналів. Методологічна база охоплювала аналіз стандартів і специфікацій цифрового навчального середовища, подієвого опису навчальної активності, якості програмних систем, інформаційної безпеки, приватності та управління рішеннями на основі штучного інтелекту. У дослідженні використано теоретичний аналіз, класифікаційний підхід, ознакове узагальнення, порівняльний аналіз, структурно-логічне моделювання та критичне узагальнення. У результаті встановлено, що цифрові сліди користувачів не є однорідним масивом технічних записів, а відображають часову, навігаційну, ресурсну, завданнєву, тестову, комунікативну й технічну взаємодію з навчальним середовищем. Показано, що аномалії можуть проявлятися як часові, навігаційні, ресурсні, завданнєві, тестові, комунікативні, технічні та змішані відхилення. Визначено, що машинний аналіз потребує перетворення журналів активності на ознаки, пов'язані з частотою входів, тривалістю сесій, послідовністю переходів, виконанням завдань, тестовими спробами, комунікативною активністю та технічними параметрами доступу. Встановлено, що вибір групи методів машинного навчання залежить від типу аномалії, наявності розмічених даних, структури подій і потреби в інтерпретованості. Показано послідовність переходу від журналів активності до очищення даних, уточнення контексту навчальних ресурсів, формування ознак, визначення типового профілю, вибору алгоритмічного підходу, фіксації аномального сигналу та його педагогічної або адміністративної інтерпретації. Окремо обґрунтовано, що аномальний сигнал не може автоматично трактуватися як доказ порушення, оскільки його значення залежить від якості даних, історії користувача, структури курсу, технічних умов доступу та людської перевірки. Практична значимість результатів полягає у можливості використання запропонованої схеми для побудови систем аналітичного супроводу навчання, де аномальний сигнал розглядається як підстава для контекстної перевірки

Ключові слова: цифрові сліди; тривалість сесії; методи на основі ізоляції; виявлення нетипових спостережень; моделі послідовностей