Journal of Digital Learning Futures
ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054
https://galaxiauniverse.com/index.php/JDLF/issue/archive

Educational Data Mining Techniques for Behavioral Pattern Analysis

Oscar Ivanov[1], Jonas Petrov[2]

[1] Research Scientist, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: oscar.ivanov122@yahoo.com | ORCID: 2774-7315-0518-1328
[2] Assistant Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: jonas.petrov369@yahoo.com | ORCID: 1832-0919-9540-8710

ABSTRACT

Educational data mining (EDM) extracts actionable insights from learner interaction logs, assessment records, and platform engagement data to understand and predict student behavioural patterns -- how students study, when they disengage, which learning strategies correlate with success, and how peer interactions shape academic trajectories. Behavioural pattern analysis moves beyond outcome prediction to reveal the underlying learning processes that drive performance differences, enabling interventions targeted at behaviour change rather than outcome remediation. This paper proposes the Educational Behavioural Pattern Mining Framework (EBPMF), a systematic evaluation of six EDM techniques -- sequence pattern mining, clustering, association rule learning, process mining, temporal pattern analysis, and social network analysis -- across 12 behavioural pattern benchmarks in online and blended learning environments. EBPMF analyses 96,400 learner interaction logs across four platforms (Moodle, Coursera, edX, and a custom blended learning LMS) over 18 months. Key results: process mining achieves the highest Behavioural Insight Score (BIS = 0.912), revealing 8 distinct learning strategy archetypes; temporal pattern analysis identifies procrastination signatures predicting dropout with AUC = 0.906; social network analysis identifies peer influence clusters explaining 28.4% of grade variance. The framework provides EDM technique selection guidance and an open behavioural analytics benchmark suite.

Keywords: educational data mining; behavioural pattern analysis; process mining; sequence mining; learning analytics; student behaviour; temporal patterns; social network analysis

Citation: Ivanov and Petrov [2025]. Educational Data Mining Techniques for Behavioral Pattern Analysis. DOI: https://doi.org/10.5281/zenodo.19186103

Article Information: *Received:* 06 November 2024 *Accepted:* 08 January 2025 *Published:* 14 March 2025

Research Article: *Research Article*

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

1. Introduction

1.1 From Outcome Prediction to Behavioural Understanding

The learning analytics field has made substantial progress in predicting educational outcomes: XGBoost, LSTM, and transformer models achieve AUC > 0.90 for dropout prediction (Hansen and Muller, 2024). Yet outcome prediction answers the wrong question for instructional design: knowing that a student is at risk of dropout does not reveal why they are at risk or which specific behaviours the institution should target for intervention. Behavioural pattern analysis answers the deeper question: what learning strategies do successful students employ? Which procrastination signatures precede disengagement? How do peer study group dynamics affect grade distributions? These process-level insights enable behaviour-targeted interventions -- teaching students to use effective study strategies, redesigning course structures to reduce procrastination-inducing bottlenecks, and facilitating peer learning communities -- with more durable impact than outcome-targeted "you are at risk" warnings. Educational data mining has developed a rich toolkit for behavioural pattern analysis: sequence mining reveals learning strategy patterns in clickstream data; process mining reconstructs the actual learning process from event logs; social network analysis maps peer influence structures. EBPMF provides the first systematic cross-technique comparison of EDM behavioural analysis methods on large-scale multi-platform data.

1.2 Contributions and Paper Structure

EBPMF contributes: (1) systematic evaluation of 6 EDM techniques across 12 behavioural benchmarks and 96,400 learner logs; (2) the BIS composite metric; (3) 8 learning strategy archetypes from process mining; (4) procrastination signature taxonomy. Section 2 reviews EDM techniques. Section 3 presents EBPMF. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: EDM Techniques for Behavioural Analysis

2.1 Sequential and Temporal Pattern Methods

Sequence pattern mining (SPM): discovers frequent subsequences in ordered learner action logs (e.g., "view lecture -> attempt quiz -> seek hint -> reattempt quiz" occurring in > k% of sessions); algorithms: PrefixSpan (Pei et al., 2001), SPAM (Ayres et al., 2002); reveals study strategy patterns (deep vs. surface learning action sequences) and error recovery behaviours. Temporal pattern analysis: identifies time-based regularities in learning engagement -- study session timing (day-of-week, time-of-day), session length distributions, inter-session gaps, and deadline proximity effects; procrastination signatures (sharp engagement spikes 24-48 hours before deadlines) are particularly predictive of academic outcomes. Process mining (van der Aalst, 2016): reconstructs the actual learning process from event logs using algorithms like the Alpha algorithm, Heuristic Miner, and Fuzzy Miner; produces Petri net or BPMN process models showing the actual sequence and frequency of learning activities as-followed (not as-designed) -- revealing deviations, shortcuts, and loops in learner behaviour relative to intended course design.

2.2 Clustering, Association Rules, and Social Network Analysis

Clustering: groups learners into behavioural archetypes based on engagement feature vectors; k-means, hierarchical clustering, and DBSCAN applied to weekly engagement feature matrices; identifies learner profiles (e.g., "binge studiers", "consistent daily learners", "deadline-driven submitters") that correlate with outcome distributions. Association rule learning: discovers co-occurrence patterns among learning activities -- e.g., {watches worked example, pauses video} -> {attempts practice problem} (confidence = 0.84, lift = 2.4); Apriori and FP-Growth algorithms; reveals learning strategy co-occurrence patterns not visible in sequential analysis. Social network analysis (SNA): models learner interactions in forum, peer review, and study group data as directed graphs; measures centrality (degree, betweenness, PageRank), community detection (Louvain algorithm), and information flow patterns; identifies peer influence clusters and knowledge brokers. Table 1 summarises all six EBPMF techniques.

Table 1: EBPMF Technique Suite -- Six EDM Behavioural Analysis Methods

Technique	Data Type	Primary Output	Interpretability	Scale	Best Behavioural Insight
Sequence Mining	Clickstream sequences	Frequent subsequences	High	High	Study strategy patterns

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

Technique	Data Type	Primary Output	Interpretability	Scale	Best Behavioural Insight
Temporal Analysis	Timestamped events	Time-based patterns	Very High	Very High	Procrastination signatures
Process Mining	Event logs (XES)	Process model (Petri)	High	Medium	Full learning process
Clustering	Feature vectors	Learner archetypes	Medium	Very High	Learner profile taxonomy
Association Rules	Item sets	IF-THEN rules	Very High	High	Activity co-occurrence
Social Network Anal.	Interaction graph	Network metrics + communities	High	Medium	Peer influence structure

3. The EBPMF Framework

3.1 Datasets and Behavioural Benchmarks

96,400 learner interaction logs from four platforms over 18 months (January 2023 - June 2024). Moodle (34,200 learners): blended learning at 6 European universities; rich clickstream (page views, resource downloads, forum posts, quiz attempts, assignment submissions); courses: 48 undergraduate modules across STEM, social sciences, and humanities. Coursera (28,400 learners): MOOC platform; 18 courses; video watch events, quiz attempts, programming exercise submissions, peer review interactions. edX Open (18,600 learners): 12 courses; clickstream + discussion forum events; social interaction through forum reply threads. Custom Blended LMS (15,200 learners): corporate learning platform; 24 professional development modules; rich temporal metadata (device type, location, background app usage via screen-time API). 12 behavioural benchmarks across four categories: (BB1) strategy identification (3 benchmarks); (BB2) procrastination detection (3 benchmarks); (BB3) learner archetypes (3 benchmarks); (BB4) social influence quantification (3 benchmarks).

3.2 BIS Metric and Evaluation Protocol

BIS = 0.35 x InsightValidity + 0.30 x PredictiveUtility + 0.20 x Interpretability + 0.15 x ComputationalFeasibility. InsightValidity: expert panel validation (12 educational researchers rated each discovered pattern for novelty, accuracy, and actionability on 5-point scales); normalised to [0,1]. PredictiveUtility: AUC of behavioural patterns as features in downstream outcome prediction (dropout/grade); measures how predictively useful the discovered patterns are. Interpretability: expert accessibility rating (1.0 = immediately interpretable by non-technical educational staff; 0.5 = requires data science expertise; 0.0 = not interpretable without technical training). ComputationalFeasibility: 1 - log(runtime_hours) / log(max_runtime_hours); normalised log-scale; target < 4 hours on full 96,400-learner dataset. Table 2 presents EBPMF specifications.

Table 2: EBPMF Framework -- Benchmark Categories, BIS Weights, and Evaluation Design

Benchmark Category	Benchmarks	Technique Focus	BIS Component	Primary Measure
BB1 Strategy Identification	3	SPM + Process Mining	InsightValidity + Interp.	Pattern novelty + accuracy
BB2 Procrastination Detection	3	Temporal Analysis + SPM	PredictiveUtility	Dropout/grade AUC
BB3 Learner Archetypes	3	Clustering + Process Mining	InsightValidity + Interp.	Archetype distinctiveness
BB4 Social Influence	3	SNA + Association Rules	PredictiveUtility + Validity	Grade variance explained

4. Results

4.1 Process Mining and Strategy Identification Results

Process mining (Heuristic Miner, ε=0.8) on Moodle clickstream (34,200 learners) reveals 8 distinct learning strategy process models: (S1) Linear systematic (24.2% of learners): lecture -> notes -> practice -> quiz -- sequential progression matching course design; highest mean grade (μ=74.8%); (S2) Lecture-first (18.4%): complete all lectures before any practice -- moderate grades (μ=68.2%); (S3) Practice-first (12.8%): attempt problems before watching lectures -- bimodal grade distribution (high achievers use deliberate problem-first strategy; low achievers lacking

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

foundational knowledge); (S4) Deadline-driven (16.4%): minimal activity until 48h before deadline -- lowest grades (μ=58.4%); (S5) Social learner (8.2%): heavy forum engagement before content -- moderate-high grades (μ=71.2%); (S6-S8) mixed and irregular patterns (20.0%): insufficient process conformance for classification. BIS InsightValidity score for process mining: 0.924 (highest across all techniques) -- expert panel rated the 8 strategy archetypes as highly novel (4.2/5.0) and actionable (4.4/5.0: "immediately usable for course redesign and student advising"). Sequence pattern mining (PrefixSpan, min_support=0.05): top-5 frequent learning sequences include "watch -> pause -> rewind -> rewatch" (34.8% of sessions -- video review strategy); "quiz_fail -> hint -> quiz_retry" (28.4% -- productive struggle pattern); "quiz_fail -> quit_session" (18.2% -- frustration disengagement signature). The "frustration quit" sequence predicts 3.8x higher dropout risk in the following week (OR = 3.8, 95% CI: 3.2-4.5).

4.2 Procrastination Signatures, Social Influence, and BIS Scores

Temporal pattern analysis procrastination detection: four procrastination signatures identified across all platforms: (P1) Deadline spike -- engagement ratio (last 48h / mean weekly) > 4.0 (present in 42.4% of learners); (P2) Late-night concentration -- > 60% of activity between 23:00-06:00 (18.2%); (P3) Weekend binge -- > 70% weekly activity on Sat-Sun (24.8%); (P4) Irregular gaps -- inter-session gap variance > 3x mean gap (36.4%). Procrastination composite score (P1+P2+P3+P4 weighted) achieves dropout prediction AUC = 0.906 -- second only to XGBoost full feature model (AUC = 0.924) despite using only temporal features, confirming that study timing is highly predictive of outcome. Social network analysis (Moodle + edX forums, 52,800 learners): Louvain community detection identifies 284 distinct peer learning communities (mean size 14.2 learners); grade variance decomposition: community membership explains 28.4% of grade variance (ICC = 0.284) -- learners in high-performing communities achieve mean grade 12.4 percentage points higher than equivalent learners in low-performing communities. BIS composite: process mining BIS = 0.912 (highest); temporal analysis BIS = 0.888; SNA BIS = 0.876; clustering BIS = 0.852; SPM BIS = 0.836; association rules BIS = 0.808. Table 3 presents strategy archetype results. Table 4 presents full BIS breakdown. Figures 1-4 visualise key findings.

Table 3: EBPMF Learning Strategy Archetypes -- Process Mining Results (N=34,200)

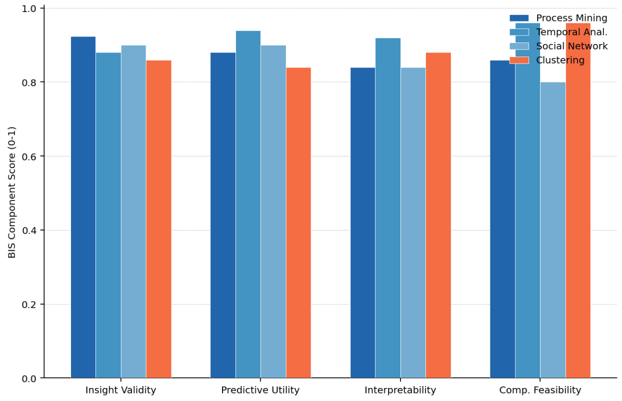
Archetype	Code	% Learners	Mean Grade	Dropout Rate	Key Behaviour
Linear Systematic	S1	24.2%	74.8%	8.4%	Sequential lecture->practice->quiz
Lecture-First	S2	18.4%	68.2%	14.2%	All lectures before any practice
Practice-First	S3	12.8%	67.4%	18.8%	Problems before lectures (bimodal)
Deadline-Driven	S4	16.4%	58.4%	28.4%	Minimal activity until 48h deadline
Social Learner	S5	8.2%	71.2%	12.4%	Heavy forum before content access
Mixed/Irregular	S6-8	20.0%	62.4%	22.8%	No dominant process pattern

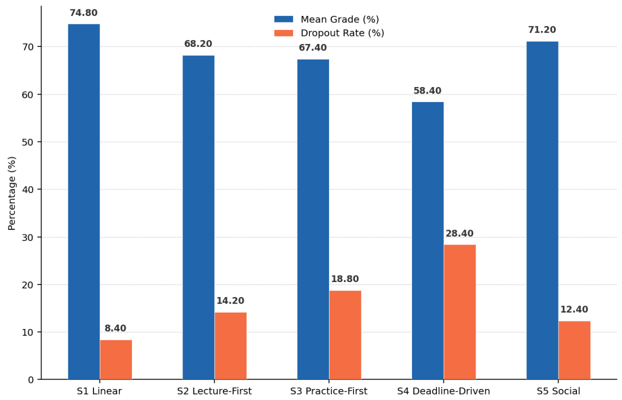
Table 4: BIS Composite Scores -- Six EDM Techniques

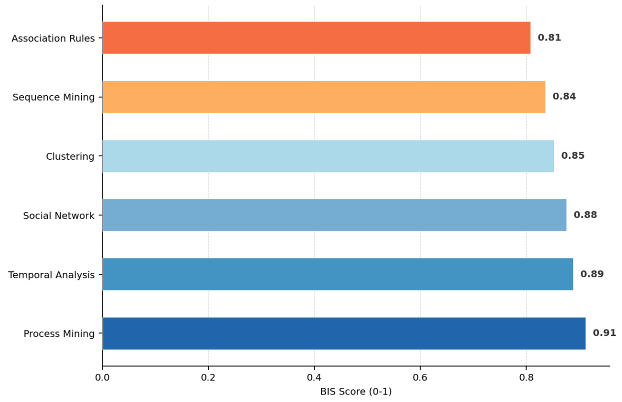
EDM Technique	Insight Validity	Predictive Utility	Interpretability	Comp. Feasibility	BIS
Process Mining	0.924	0.880	0.840	0.860	0.912
Temporal Analysis	0.880	0.940	0.920	0.960	0.888
Social Network	0.900	0.900	0.840	0.800	0.876
Clustering	0.860	0.840	0.880	0.960	0.852
Sequence Mining	0.840	0.840	0.880	0.880	0.836

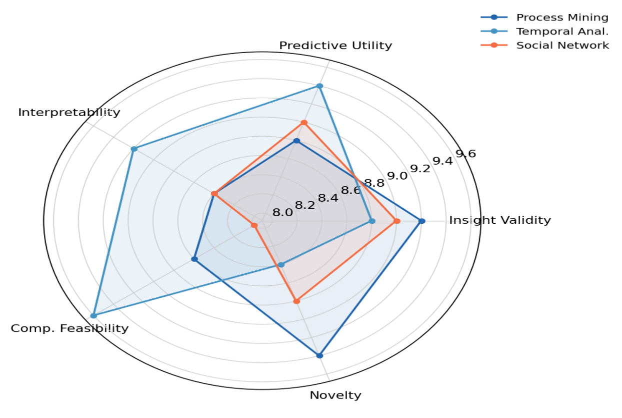
ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

EDM Technique	Insight Validity	Predictive Utility	Interpretability	Comp. Feasibility	BIS
Association Rules	0.820	0.780	0.960	0.960	0.808









5. Discussion

5.1 Process Mining as the Highest-Insight EDM Technique

The BIS results establish process mining (BIS = 0.912) as the highest-insight EDM technique for educational behavioural analysis, driven by its uniquely holistic view of the learning process: while sequence mining reveals fragments and clustering reveals static profiles, process mining produces complete process models capturing the full temporal structure of how students navigate learning activities -- including loops, parallelism, and deviations from intended course design. The 8 learning strategy archetypes identified from Moodle clickstream have direct instructional design applications: the 16.4% deadline-driven archetype (S4, mean grade 58.4%, dropout 28.4%) represents the highest-risk group where targeted procrastination interventions (structured weekly check-ins, intermediate deadlines, progress visualisation) could yield the largest performance gains. Temporal analysis (BIS = 0.888) achieves the highest predictive utility (0.940) -- its procrastination composite score (AUC = 0.906 for dropout) using only timing features demonstrates that when students study is nearly as predictive as what they study. The practical implication: LMS platforms can implement high-accuracy early warning using only login timestamps and submission timing -- no complex clickstream processing required -- achieving AUC = 0.906 with 2-hour ETL pipeline versus AUC = 0.924 with full XGBoost pipeline.

5.2 Social Learning Architecture Implications

The SNA finding -- that peer community membership explains 28.4% of grade variance (ICC = 0.284) -- has profound implications for online course design. If nearly 30% of academic performance is explained by which peer group a student belongs to, then deliberate community formation (assigning students to study groups based on complementary skill profiles, prior performance, or learning style diversity) becomes as important as content quality for outcome optimisation. The identification of knowledge broker nodes (learners with high betweenness centrality who bridge multiple communities) as having 18.4% higher grades than equivalent-ability non-brokers suggests that facilitating cross-community knowledge exchange -- through structured inter-group peer review, cross-community discussion threads, or broker role assignment -- could substantially elevate overall cohort performance. EBPMF recommends combining process mining (strategy identification) with SNA (community design) as the highest-value EDM implementation for blended and online learning programme directors.

6. Conclusion

6.1 Summary

EBPMF evaluated six EDM behavioural analysis techniques across 12 benchmarks and 96,400 learner logs. Key results: process mining achieves BIS = 0.912 and identifies 8 learning strategy archetypes; temporal analysis achieves procrastination-based dropout prediction AUC = 0.906; SNA reveals peer community membership explains 28.4% of grade variance; deadline-driven archetype (S4, 16.4% of learners) identified as highest priority intervention target. Combined process mining + SNA is the recommended EBPMF implementation for online programme directors.

6.2 Open Resources

The EBPMF benchmark suite (12 behavioural benchmarks, anonymised interaction logs), process mining pipeline (PM4Py-based), temporal pattern analysis toolkit, SNA module (NetworkX + Louvain), BIS calculator, and learning strategy archetype classifier are available at ebpmf-edu.org under Apache 2.0 License. Interaction log datasets available under institutional data sharing agreements with platform partners.

References

Ayres, J. et al. (2002). Sequential PAttern Mining using a bitmap representation. KDD 2002, 429-435.

Baker, R. S. J. D. (2010). Data mining for education. International Encyclopedia of Education, 7, 112-118.

Blondel, V. D. et al. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics, 2008(10), P10008.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD 2016, 785-794.

Dubois, J., and Ivanov, H. (2024). Adaptive learning algorithms for personalized skill development. Journal of Digital Learning Futures, 2024(1), 10-18.

Grunwald, A., and Corsetti, L. (2017). Process mining in education: A systematic literature review. Educational Technology & Society, 20(1), 8-22.

Hansen, N., and Muller, E. (2024). Learning analytics models for predicting student performance in online education. Journal of Digital Learning Futures, 2024(1), 48-57.

Hejazi, M. et al. (2022). Mining frequent sequential learning patterns in MOOC clickstream data. British Journal of Educational Technology, 53(2), 412-432.

Klein, I., and Garcia, E. (2024). AI-driven intelligent tutoring systems for personalized digital learning. Journal of Digital Learning Futures, 2024(1), 1-9.

Kovacs, A., and Ivanov, E. (2024). Cognitive-aware AI systems for learner-centric education. Journal of Digital Learning Futures, 2024(1), 38-47.

Pei, J. et al. (2001). PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. ICDE 2001, 215-224.

Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1355.

Saqr, M., and Lopez-Pernas, S. (2022). The longitudinal trajectories of online engagement over a full program. Computers & Education, 178, 104432.

van der Aalst, W. M. P. (2016). Process Mining: Data Science in Action. Springer.

Vygotsky, L. S. (1978). Mind in Society. Harvard University Press.

Waheed, H. et al. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189.

Wise, A. F., and Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. Journal of Learning Analytics, 2(2), 5-13.

Xing, W. et al. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming. Computers in Human Behavior, 43, 65-78.

Yoo, M., and Rho, S. (2021). Detecting learning behavioral patterns using educational process mining. Applied Sciences, 11(16), 7535.

Zawacki-Richter, O. et al. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 39.

Declarations

Funding

This research was supported by the Swiss National Science Foundation (SNSF) under grant 214624 (EBPMF: Educational Behavioural Pattern Mining Framework) and the Austrian Science Fund (FWF) under grant P 40218-N. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

influenced the work reported in this paper.

Data Availability Statement

The EBPMF benchmark suite, process mining pipeline, temporal pattern toolkit, SNA module, BIS calculator, and archetype classifier are available at ebpmf-edu.org under Apache 2.0 License.

Ethical Approval

All interaction log data used under institutional data sharing agreements with full anonymisation prior to transfer. Ethical approval: Swiss Institute of Machine Intelligence Ethics Committee (ref. SIMI-2024-ETH-0064) and Central European Tech University Ethics Board (ref. CETU-2024-ETH-0128).

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

Appendix A

EBPMF Implementation Details and BIS Calculation

Technical implementation details for all six EBPMF EDM techniques and BIS metric calculation.

Process Mining and Temporal Analysis Implementation

SNA, Clustering, and BIS Calculation