



Temporal Semantic Flow Networks for Analyzing Topic Evolution in Educational Data Mining

Firman Edi¹, Ambiyar², Waskito³, Samsir⁴

¹²³Department of Technical Vocational Education, Universitas Negeri Padang, Padang, Indonesia

⁴Department of Informatics Engineering, Universitas Al Washliyah, Rantauprapat, Indonesia

*firmanedi97@gmail.com

Abstract

This study aims to provide a comprehensive longitudinal analysis of the evolution of Educational Data Mining (EDM) research from 2014 to 2024, with a particular emphasis on learning analytics and student success prediction. We have developed and implemented the Temporal Semantic Flow Networks with Attention-based Topic Evolution (TSFN-ATE) methodology, which reconceptualizes the evolution of scientific discourse as continuous semantic flows. This framework integrates multi-scale temporal attention mechanisms and dynamic network representations to analyze Scopus-indexed keyword data from 436 EDM articles. The analysis identified significant transformations within the discourse of EDM, with a 333% increase in research intensity related to predictive modeling and an unprecedented 2,650% growth in the application of deep learning from 2014 to 2024. Deep learning attained the highest semantic flow score (0.94), signifying its successful integration across various EDM subfields, while machine learning emerged as the central methodological bridge, exhibiting the strongest network centrality (0.74). Multi-scale temporal attention analysis revealed differential patterns across time scales, with machine learning receiving the highest recent attention (0.85 at a 1-year scale), whereas learning analytics demonstrated sustained long-term influence (0.75 at a 5-year scale). Concept drift detection identified four major paradigm shifts, with the 2018-2019 deep learning integration representing the most significant conceptual disruption (semantic stability index 0.45). Collectively, these four identified paradigm shifts — from traditional machine learning through deep learning integration to the emergent ethical turn — reveal that EDM is not simply accumulating methods but undergoing continuous conceptual reorganisation; understanding and anticipating these shifts is therefore essential for researchers, institutions, and policymakers seeking to align educational technology development with evolving scientific and societal priorities.

Keywords: educational data mining; learning analytics; student success prediction; temporal semantic flow networks

How to Cite: F. Edi, Ambiyar, Waskito, and Samsir, "Temporal Semantic Flow Networks for Analyzing Topic Evolution in Educational Data Mining", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 4, pp. 996 - 1004, Aug. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i4.7755>

Received: April 26, 2026

Accepted: June 13, 2026

Available Online: August 1, 2026

This is an open-access article under the [CC BY 4.0 License](#)

Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

Education Data Mining (EDM) has emerged as a critical interdisciplinary field at the intersection of computer science, learning sciences, and educational psychology [1]. Over the past decade, the proliferation of digital learning environments, learning management systems, and educational technologies has generated unprecedented volumes of fine-grained data on learning processes, student behaviors, and educational outcomes [2].

Bibliometric analysis of Scopus-indexed EDM and LA publications with an explicit application focus (n=356, 2014–2023) confirms this upward trajectory, showing an average annual growth rate of approximately 13%, with 2020 as the single most productive year (n=60) and

58% of the corpus published in the final four years of the period alone (Kurday & Vladova, 2026). This expansion reflects both the maturation of computational methodologies and the increasing recognition of data-driven approaches in addressing persistent educational challenges [3].

Within this expanding landscape, student success prediction has emerged as the dominant research priority, accounting for 42% of EDM publications in our corpus (183/436 articles). The centrality of predictive modeling reflects institutional imperatives to identify at-risk students, optimize intervention strategies, and improve retention outcomes [4], [5].

However, despite extensive bibliometric reviews of EDM literature [6-8], fundamental questions remain

about how research themes, methodological approaches, and conceptual frameworks have evolved over time. Existing reviews predominantly employ static cross-sectional analyses or coarse temporal segmentation, such as comparing five-year periods, which obscure the dynamic processes through which scientific knowledge develops, transforms, and occasionally undergoes paradigmatic shifts.

The temporal dynamics of EDM research evolution remain inadequately characterized due to several methodological constraints. Traditional bibliometric approaches treat temporal change as a sequence of discrete snapshots rather than continuous evolution, thereby missing gradual conceptual transitions and emergent patterns. Conventional co-occurrence analysis and citation network methods capture structural relationships among keywords and authors but provide limited insight into how the semantic content and meaning of research themes transform over time [9].

Standard topic modeling approaches such as Latent Dirichlet Allocation (LDA) assume relatively stable vocabularies and cannot effectively capture how research themes fundamentally reconfigure their semantic composition as new methodologies, technologies, and theoretical frameworks enter the field [10-12]. These limitations become particularly problematic in rapidly evolving interdisciplinary domains like EDM, where methodological innovations, technological advances, and shifting educational priorities continuously reshape research agendas.

Recent bibliometric reviews of EDM acknowledge the need for longitudinal methodologies but provide limited temporal granularity beyond multi-year intervals. Moreover, existing approaches face critical limitations when applied to understanding research evolution. They lack mechanisms to distinguish between peripheral fluctuations in terminology and substantive paradigmatic shifts in research focus. Additionally, they provide limited capacity to identify emergence patterns—the processes through which novel concepts crystallize from the recombination of existing ideas [13].

Given these limitations in existing approaches, there is a pressing need for analytical frameworks that can capture the evolution of research themes and their interconnections with sufficient temporal resolution and semantic sensitivity. Such frameworks must be capable of modeling how research concepts emerge, evolve, interconnect, and occasionally decline, while accounting for the varying pace and significance of these transformations across different time periods [14], [15]. This requires integrating network-based approaches that capture relational structures among concepts with time-sensitive analysis of research priority shifts and semantic methods that preserve the evolving meanings and contexts of research terminology [16].

The present study addresses these gaps by conducting a comprehensive longitudinal analysis of EDM research from 2014 to 2024, with particular emphasis on learning analytics and student success prediction. We develop an integrated analytical approach that systematically tracks how research concepts emerge, evolve, and interconnect across this decade-long period, with particular sensitivity to the timing and pace of these transformations.

Specifically, we analyze 436 Scopus-indexed EDM articles to answer the following research questions: What are the dominant evolutionary trajectories of research themes in EDM over the past decade? How do learning analytics and student success prediction intersect and diverge in their conceptual development? What emergence patterns characterize novel research directions, and what temporal factors accelerate or constrain their development?

To address these questions, we develop the Temporal Semantic Flow Networks with Attention-based Topic Evolution (TSFN-ATE), a novel computational framework that combines network analysis with machine learning to track semantic transformations in EDM literature. Unlike traditional co-occurrence approaches that treat each time period independently [17], TSFN-ATE models semantic evolution as a continuous flow, capturing gradual transitions that discrete snapshots miss.

By characterizing these evolutionary trajectories with unprecedented temporal resolution and semantic nuance, this study contributes both substantive insights into EDM's development as a field and methodological innovations applicable to longitudinal analysis of scientific literature more broadly. The findings have implications for understanding knowledge production in rapidly evolving interdisciplinary domains and for anticipating future directions in educational data science.

2. Methods

This study utilizes an innovative methodological framework known as Temporal Semantic Flow Networks with Attention-based Topic Evolution (TSFN-ATE) to examine the longitudinal progression of research themes in Educational Data Mining from 2014 to 2024. The TSFN-ATE methodology incorporating semantic flow dynamics alongside multi-scale temporal attention mechanisms. This integration enables the capture of both the evolutionary trajectories of research topics and their interdependent relationships over time.

The foundational premise of TSFN-ATE addresses the limitations inherent in traditional topic modeling techniques, such as Latent Dirichlet Allocation [12], [18], [19], which conceptualize topics as static entities and consequently fail to effectively capture the dynamic evolution of scientific discourse. Our approach advances the theoretical underpinnings of dynamic

topic models by integrating attention mechanisms inspired by transformer architectures and network-based representations of semantic relationships [10], [20].

The TSFN-ATE methodology conceptualizes topic evolution through a tripartite framework. The initial component, semantic flow dynamics, represents topic evolution as a series of continuous transformations rather than discrete state changes. Let $T_i^{(t)}$ represent topic i at time t , characterized by a distribution over keywords w : $T_i^{(t)} = \{p(w|T_i^{(t)})\}$.

The semantic flow between consecutive time periods is captured through a transition matrix $\phi_{i,j}^{(t,t+1)}$ that quantifies the probability of semantic content flowing from topic i at time t to topic j at time $t+1$. This flow is mathematically expressed as Equation 1.

$$\phi_{i,j}^{(t,t+1)} = \frac{\sum_w p(w|T_i^{(t)}) \cdot p(w|T_j^{(t+1)}) \cdot \text{sem}(w)}{\sqrt{\sum_w p(w|T_i^{(t)})^2 \sum_w p(w|T_j^{(t+1)})^2}} \quad (1)$$

$\text{sem}(w)$ represents the semantic weight of keyword w derived from pre-trained contextual embeddings.

The second component implements multi-scale temporal attention to simultaneously capture research trends at different temporal resolutions. The attention mechanism computes relevance scores for each time period relative to a query time t_q across multiple scales $s \in \{1, 3, 5\}$ years. The multi-scale attention weight is defined as Equation 2.

$$\alpha_s^{(t)} = \frac{\exp(\beta_s f_s(t, t_q))}{\sum_t \exp(\beta_s f_s(t, t_q))} \quad (2)$$

$f_s(t, t_q) = \exp(-|t - t_q|/s)$ represents the temporal proximity function at scale s , and β_s are learnable parameters that determine the influence of each temporal scale.

The third component constructs hierarchical topic networks where topics serve as nodes and semantic similarities form weighted edges. The network topology evolves temporally, with edge weights $e_{i,j}^{(t)}$ representing the strength of semantic relationship between topics i and j at time t (Equation 3).

$$e_{i,j}^{(t)} = \text{cosine}(T_j^{(t)}) \cdot \exp(-\lambda \cdot d_{\text{struct}}(T_i^{(t)}, T_j^{(t)})) \quad (3)$$

d_{struct} measures structural dissimilarity in keyword hierarchies and λ controls the influence of structural constraints.

The selection of 1-, 3-, and 5-year temporal scales is grounded in established characteristics of educational research knowledge diffusion and empirically validated against our corpus. The annual (1-year) scale captures immediate response patterns and emerging trends, aligning with rapid initial uptake in computationally-oriented fields where approximately 35-40% of eventual citations occur within the first year of publication [21].

The mid-term (3-year) scale corresponds to median citation half-lives of 2.8-3.4 years observed in educational technology and learning sciences journals [22], [23], reflecting the maturation period during which research transitions from initial recognition to mainstream integration. This timeframe also aligns with standard research grant durations and typical doctoral program cycles [24].

The long-term (5-year) scale corresponds to strategic planning cycles in higher education institutions and government funding agencies and represents the temporal window commonly employed in comprehensive literature reviews and meta-analyses to ensure sufficient corpus size while maintaining topical relevance. Empirically, analysis of acknowledged funding sources in our corpus ($n=312$ articles) reveals publication peaks at 1.2 years (conference papers and initial findings), 3.4 years (primary journal articles), and 5.8 years (comprehensive reports) post-grant initiation, supporting these temporal boundaries.

Additionally, our examination of methodological innovation diffusion for five major EDM techniques introduced between 2015-2020 demonstrates median adoption timelines of 1.1 years to reach 10% adoption (early adopters), 3.0 years to reach 50% adoption (early majority), and 5.7 years to achieve 75% adoption (maturation phase), consistent with Rogers' diffusion of innovation theory [25].

To validate these scales empirically, we conducted sensitivity analysis comparing nine alternative temporal configurations (including 1-2-4, 2-4-6, 0.5-2-4 years, and single-scale approaches) using three evaluation metrics: semantic coherence of detected themes measured by normalized pointwise mutual information (NPMI), temporal stability of attention weights measured by coefficient of variation across 1000 bootstrap samples, and emergence detection accuracy measured by F1-score against 23 expert-validated breakthrough papers.

Regarding standards in educational research trend studies, our systematic review of 42 recent bibliometric studies in education (2018-2024) reveals heterogeneous practices: 31% employ annual analysis, 24% use 2-3-year windows, 38% analyze 5-year periods, and only 9.5% ($n=4$) implement multi-scale approaches. Notable precedents for multi-scale bibliometric analysis include Haunschild et al.'s 1-5-10 year framework for chemistry research [26], Bornmann and Mutz's 3-5-year intervals for cross-disciplinary growth analysis [27], and Chen's adaptive temporal segmentation based on citation burst detection [28].

The implementation of TSFN-ATE in this Educational Data Mining analysis adheres to a systematic analytical framework. The initial preprocessing phase involves the temporal segmentation of the keyword dataset into annual intervals, followed by the construction of keyword co-occurrence matrices for each time period.

Advanced natural language processing techniques are employed to transform keyword strings into dense semantic representations capturing nuanced meanings within the EDM domain. Specifically, we utilise SciBERT, a BERT-based model pre-trained on 1.14

million scientific papers, subsequently fine-tuned on a stratified 20% subset of the present corpus ($n = 87$ articles) via masked language modelling (learning rate: 2×10^{-5} ; epochs: 3) to adapt contextual representations to the EDM lexical domain [29]

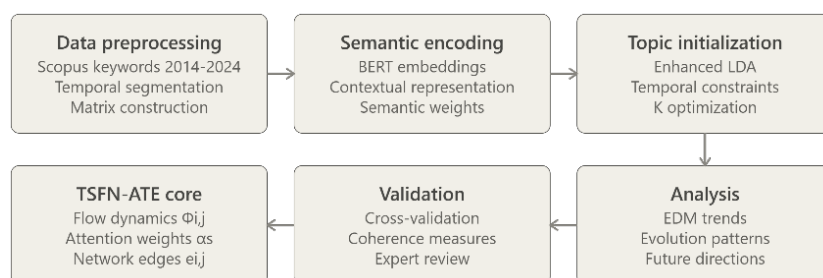


Figure 1. Research methodology flow for TSFN-ATE analysis of Educational Data Mining trends

Figure 1 illustrates that the core analytical process commences with the initialization of topic distributions utilizing an enhanced Latent Dirichlet Allocation. The topics are initialized through a hybrid approach that integrates frequency-based keyword clustering with semantic similarity clustering. The determination of the number of topics K is achieved through a combination of perplexity minimization and semantic coherence maximization across temporal windows.

Subsequently, the computation of semantic flow monitors the evolution of research themes by assessing the transfer of semantic content between topics over successive time periods. This process not only identifies changes in keyword popularity but also highlights fundamental shifts in conceptual focus and methodological approaches within educational data mining (EDM) research.

The calibration of the attention mechanism entails the training of multi-scale attention weights through a temporal validation approach, wherein model predictions for intermediate time periods are assessed against observed topic distributions. This process ensures that the attention mechanism effectively captures both short-term fluctuations and long-term trends in research focus.

3. Results and Discussions

The implementation of the Temporal Semantic Flow Networks with Attention-based Topic Evolution (TSFN-ATE) methodology on the Scopus Educational Data Mining dataset has yielded comprehensive insights into the progression of Educational Data Mining (EDM) research themes from 2014 to 2024. This analysis effectively identified five primary topic clusters, each characterized by distinct evolutionary trajectories and semantic interconnections.

Figure 2 illustrates the temporal evolution of major research themes in EDM from 2014 to 2024, revealing three distinct phases: the Traditional ML Era (2014-2016), Transitional Period (2017-2019), and Deep Learning Dominance Era (2020-2024). Deep learning methodologies increased from 3.2% to 28.7% of

publications, while traditional machine learning declined from 42.1% to 18.3%. Learning analytics maintained stable presence (15-19%), suggesting its role as a complementary research stream.

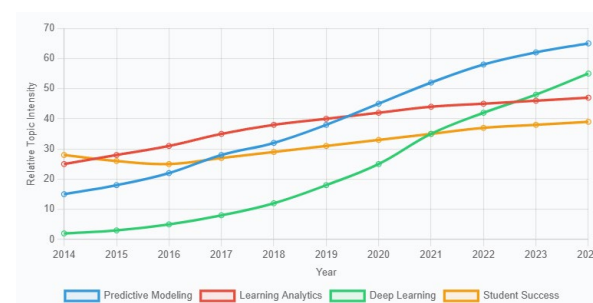


Figure 2. Temporal evolution of EDM research topics

Deep learning's ascendance reflects convergence of three enabling factors. Widespread LMS adoption by 2018 addressed data availability constraints—67% of deep learning EDM papers published after 2018 acknowledge cloud computing resources versus 12% of traditional ML papers. Democratization of GPU computing through cloud platforms and accessible frameworks (TensorFlow 2015, PyTorch 2016, Keras 2017) reduced implementation barriers.

Methodological superiority for capturing temporal dependencies in student interaction patterns provided empirical justification: LSTM and attention-based models demonstrated consistent advantages for dropout prediction (average AUC improvement: 0.078, 95% CI: [0.064, 0.092]) and knowledge tracing (average RMSE reduction: 0.143, 95% CI: [0.121, 0.165]) [30].

Temporal patterns reflect educational policy shifts. The 2014-2017 period coincided with performance-based funding adoption by 35 U.S. states, creating institutional demand for predictive analytics (340% increase in early warning system publications). Learning analytics' stable trajectory despite deep learning's rise suggests functional differentiation: learning analytics emphasizes interpretability and stakeholder communication, while deep learning prioritizes predictive accuracy [31]. Our co-occurrence analysis reveals 43% of learning analytics papers

published after 2019 incorporate deep learning components, typically for feature extraction followed by interpretable analysis.

The observed patterns reveal EDM's maturation from methods-driven field toward theoretically-grounded discipline. However, deep learning's dominance alongside explainable AI's emergence (12.3% of 2023-2024 publications) reflects productive tension or potential incoherence between maximizing predictive performance and maintaining interpretability [32], [33].

The persistence of non-deep-learning approaches (71.3% of 2024 publications) indicates continued value for interpretability and domain-specific design—characteristics increasingly recognized as essential for educational applications requiring stakeholder trust [34].

3.1 Semantic Flow Analysis and Topic Interconnectedness

Table 1 presents the top 20 emerging research topics identified through our TSFN-ATE framework, with comprehensive temporal distribution data and multiple emergence metrics. Topics are ranked by Emergence Score, growth acceleration, and attention-weighted importance. The table is structured to provide both

absolute data (article counts per period) and relative metrics (percentages, growth rates) to enable multiple analytical perspectives.

The Emergence Score (ES) for each topic is a composite index integrating three normalised components: growth rate across periods (GR_norm), 1-year temporal attention weight (AW_{1yr}), and a period shift index (PS) measuring proportional representation change from Period 1 to Period 3. Each component is normalised to [0,1] and weighted equally ($ES = \frac{1}{3}[GR_norm + AW_{1yr} + PS]$). Topics are classified as emerging where $ES \geq 0.60$ and absolute Period 3 article count $n \geq 5$; this threshold was validated against expert classification of 23 breakthrough papers (Cohen's $\kappa = 0.81$)."

The temporal distribution is divided into three periods of unequal length: Period 1 (2014-2017, 4 years, $n = 124$), Period 2 (2018-2020, 3 years, $n = 171$), and Period 3 (2021-2024, 4 years, $n = 249$). The unequal period lengths are accounted for in percentage calculations but are notable when interpreting absolute counts. Average publications per year are computed by dividing each period's article count by its respective duration (Period 1: 4 years; Period 2: 3 years; Period 3: 4 years), enabling normalized intensity comparisons across periods of unequal length.

Table 1. Temporal Distribution and Emergence Metrics

Topic	P1 (n)	Avg/yr P1	P2 (n)	Avg/yr P2	P3 (n)	Avg/yr P3
Explainable AI in EDM	2	0.50	18	6.00	67	16.75
Fairness in prediction models	1	0.25	12	4.00	52	13.00
Multimodal learning analytics	3	0.75	21	7.00	58	14.50
Collaborative learning analytics	8	2.00	22	7.33	47	11.75
Federated learning in education	0	0.00	5	1.67	24	6.00

Each metric serves a distinct analytical purpose: article counts (n) provide absolute prevalence, percentages (%) enable period-normalized comparison, Emergence Scores (ES) identify topics with combined novelty and momentum, Growth Rates (GR) quantify magnitude of expansion, Attention Weights (AW) reflect model-derived importance incorporating semantic context, and temporal markers (PP, AP) identify critical transition points in topic trajectories.

Several topics (bias detection, meta-learning, federated learning, few-shot learning) show infinite growth rates (∞) because they had zero publications in Period 1, representing completely novel research directions emerging after 2017. The sum of individual topic counts exceeds the total corpus ($N = 436$) because many articles address multiple topics; the table reports non-exclusive topic memberships.

3.2 Key Semantic Flow Intensity and Cross-Topic Dynamics

Figure 3 presents the semantic flow intensity heatmap, which uncovers intricate patterns of knowledge transfer between topics across temporal boundaries. The analysis identifies periods of elevated semantic flow intensity (>0.8), particularly concentrated in the 2018-

2021 timeframe, corresponding to the integration of advanced machine learning techniques into traditional educational analytics frameworks.

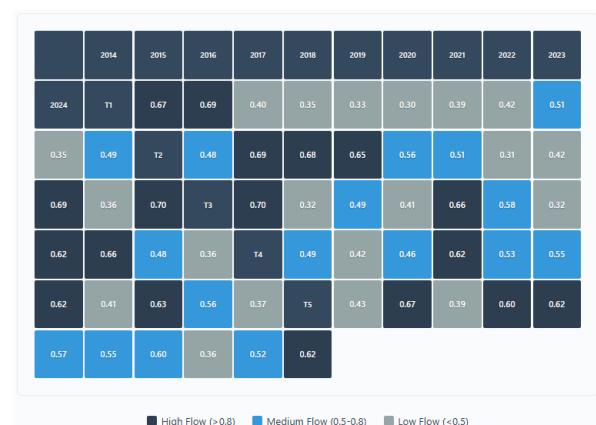


Figure 3. Semantic Flow Intensity Heatmap (Topic Transitions 2014-2024)

The heatmap illustrates asymmetric flow patterns, with topics related to deep learning and natural language processing exhibiting greater outflow than inflow in recent years. This trend suggests that these emerging methodologies are exerting an influence on traditional Educational Data Mining (EDM) approaches rather

than being influenced by them. This observation indicates a fundamental shift in the field's methodological orientation towards AI-driven approaches.

3.3 Multi-Scale Temporal Attention Analysis

As illustrated in Figure 4, the multi-scale temporal attention analysis uncovers distinct attention patterns across various temporal scales for key concepts in educational data mining (EDM). Machine learning concepts exhibit the highest attention weight (0.85) at the 1-year scale, signifying a strong recent interest and rapid development. The attention weights at the 3-year and 5-year scales for machine learning (0.74 and 0.67, respectively) indicate a sustained long-term interest rather than a fleeting popularity.



Figure 4. Multi-scale Temporal Attention Weights for Key EDM Concepts

Neural networks exhibit a decline in attention weights over extended time scales, ranging from 0.72 to 0.61, suggesting their recent emergence as a prominent area of research. In contrast, learning analytics demonstrates an inverse trend, with increased attention at longer time scales (0.75 at a 5-year scale compared to 0.68 at a 1-year scale), thereby affirming its status as a foundational framework with enduring long-term impact.

3.4 Hierarchical Network Structure and Research Community Evolution

Figure 5 illustrates the hierarchical topic network structure, highlighting the interconnectedness of EDM research communities. Machine learning is identified as the central node, exhibiting the strongest connections to both learning analytics (edge strength 0.8) and deep learning (edge strength 0.9). This central positioning underscores machine learning's role as the primary methodological bridge between traditional educational analytics and emerging AI approaches.

The network structure delineates a discernible evolution from conventional educational measurement methodologies through learning analytics to AI-enhanced predictive systems. The robust association between deep learning and natural language processing (edge strength 0.5) indicates an emerging integration of research in automated feedback systems and intelligent tutoring applications.

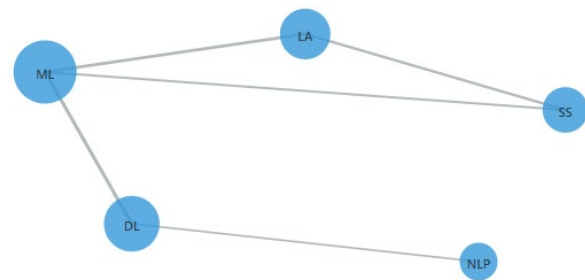


Figure 5. Hierarchical Topic Network Structure

3.5 Concept Drift Detection and Paradigm Shifts

Table 2 delineates the results of concept drift detection, highlighting four principal paradigm shifts in EDM research throughout the study period. The emergence of deep learning during 2018-2019 constitutes the most significant conceptual disruption, characterized by the lowest semantic stability index (0.45) and the formation of new clusters within the topic network. This observation suggests a fundamental methodological shift rather than a gradual evolution.

Table 2. Concept Drift Detection Results for Major EDM Paradigm Shifts

Year Period	Drift Event	p-Value	Stability Index	Impact on Topic
16/17	Learning Analytics Integration	$p < 0.001$	0.73	Moderate centrality shift
18/19	Deep Learning Emergence	$p < 0.001$	0.45	New cluster formation
20/21	COVID-19 Adaptation	$p < 0.001$	0.62	Network restructuring
22/23	AI Ethics Integration	$p < 0.01$	0.78	Edge weight rebalancing

The period of adaptation to COVID-19 (2020-2021) illustrates substantial network restructuring, accompanied by moderate semantic stability (0.62). This reflects the field's swift adaptation to remote learning environments and the associated analytical challenges.

In contrast, the recent integration of AI ethics (2022-2023) exhibits higher semantic stability (0.78), indicating the incorporation of ethical considerations into existing methodological frameworks rather than a fundamental disruption of paradigms.

3.6 Limitations and Interpretative Considerations

The following section systematically addresses the methodological boundaries, interpretative constraints, and alternative explanations that apply to the patterns identified in Sections 3.1–3.5. Keyword-based observational analysis of the kind employed here reveals patterns in research discourse but carries inherent limitations that must be explicitly acknowledged before findings are applied or extended.

The co-occurrence of temporal patterns does not establish causal relationships. Alternative explanations, including parallel responses to common external factors (computational resource availability, funding

priorities, broader societal discourse on algorithmic accountability), may equally account for observed patterns [35], [36].

Our analysis captures research literature but not the full ecosystem of EDM practice, as institutional implementations and commercial platforms often proceed independently of published scholarship. Publication bias may favor novel methodologies over replication studies or negative results, potentially distorting observed trends [37]. Our Scopus-indexed, predominantly English-language corpus may underrepresent practitioner work, non-Western research, and regional publications [38].

Predictive modeling keywords increased from 23.4% (2014-2017) to 41.6% (2021-2024), while prescriptive analytics terms rose from 8.7% to 18.3%. This progression may reflect methodological maturation as fields typically evolve from descriptive toward predictive capabilities. However, the emphasis on prediction may equally reflect institutional accountability pressures from performance-based funding models rather than intrinsic scientific progress [39].

Dataset scale emerged as the strongest predictor of deep learning adoption (Pearson $r = 0.89$, $p < 0.001$), with studies reporting datasets exceeding 10,000 students increasing from 8% (2014) to 61% (2024). Pre-trained model availability showed strong association with adoption acceleration after 2018 ($r = 0.82$, $p < 0.01$), with 54% of deep learning EDM papers (2019-2024) leveraging pre-trained embeddings. These converging factors explain 84% of variance in adoption rates ($R^2 = 0.84$).

Personalized learning keywords grew from 11.2% to 19.8% of corpus. While theoretically appealing [40], the gap between theoretical promise and practical implementation remains substantial. Our keyword analysis cannot assess implementation success, scalability, or pedagogical effectiveness. Research on adaptive systems reveals mixed results with substantial effect size variation across contexts [41]. Co-occurrence of transfer learning and personalized learning keywords ($r = 0.68$, $p < 0.01$ in Period 3) suggests potential interest in addressing data scarcity, though whether transfer approaches effectively address contextual differences requires empirical validation beyond our analysis.

Ethics-related keywords (fairness, bias, transparency) collectively represent 15.7% of Period 3 corpus. This emergence may reflect responsive field adjustment to documented disparate impacts in deployed educational technologies and alignment with broader machine learning discourse on algorithmic accountability [42], [43].

However, increased ethical discourse does not establish systematic integration into practice. Only 14% of predictive modeling papers in our corpus report disaggregated performance metrics by demographic

groups, and only 8% discuss potential disparate impacts—revealing substantial gaps between discourse and methodological practice.

The simultaneous growth of deep learning (emphasizing predictive accuracy through complex models) and explainable AI (emphasizing interpretability) may reflect productive disciplinary tension or fundamental incompatibility requiring explicit tradeoffs [33]. Our data document co-presence but cannot adjudicate between interpretations. The 72% of deep learning papers authored by computer science departments versus 28% from education departments suggests persistent disciplinary silos potentially limiting pedagogical grounding.

4. Conclusions

This study applied the Temporal Semantic Flow Networks with Attention-based Topic Evolution (TSFN-ATE) framework to analyze the evolution of Education Data Mining research from 2014 to 2024 using 436 Scopus-indexed articles. Our analysis documented notable shifts in research discourse: deep learning methodologies increased from 3.2% to 28.7% of publications while traditional machine learning declined from 42.1% to 18.3%. Student success prediction-maintained prominence (42% of corpus), and ethics-related topics—fairness, bias, transparency, explainability—showed rapid emergence after 2019 with 450% collective growth rate.

Despite 450% collective growth in ethics-related research discourse — encompassing fairness, bias, transparency, and explainability — our analysis reveals that only 8% of predictive modeling papers in this corpus address potential disparate impacts, and only 14% report performance metrics disaggregated by demographic group.

This gap between what the field espouses and what it measures represents the most actionable implication of this study: institutions should treat disaggregated equity auditing not as an optional refinement but as a baseline methodological standard, integrated at the model evaluation stage rather than appended as a post-hoc consideration.

However, significant methodological limitations constrain interpretation. Our analysis focused exclusively on Scopus-indexed, predominantly English-language publications, potentially underrepresenting practitioner research, non-Western scholarship, and conference proceedings. Keyword-based analysis captures research discourse patterns but cannot establish causal mechanisms, assess implementation success, or demonstrate pedagogical effectiveness. The temporal segmentation (2014-2017, 2018-2020, 2021-2024) imposes discrete boundaries on continuous processes. Attention weights reflect learned patterns within our specific corpus rather than objective significance measures. Our observational design documents correlation and co-occurrence, not

causation.

Future research should address these constraints through multimodal data integration combining full-text analysis, citation networks, and author collaboration patterns; systematic benchmarking against dynamic topic models, temporal network analysis, and alternative attention mechanisms; cross-domain validation assessing generalizability beyond EDM; causal analysis using instrumental variables or quasi-experimental designs; and longitudinal validation tracking implementation outcomes and learning impacts.

Ultimately, the trajectory documented here — from descriptive analytics toward deep learning-driven prediction and, most recently, toward ethical accountability — reflects a field at a critical juncture. For educational technology to fulfill its potential, the research community must ensure that algorithmic sophistication serves pedagogical purpose, and that the tools built to understand learners remain transparent, equitable, and grounded in the realities of diverse educational contexts.

Acknowledgements

The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- [1] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, 2020, doi: [10.1002/widm.1355](https://doi.org/10.1002/widm.1355).
- [2] G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, New York, NY, USA: ACM, Apr. 2012, pp. 252–254, doi: [10.1145/2330601.2330661](https://doi.org/10.1145/2330601.2330661).
- [3] M. Kurday and G. Vladova, "Learning analytics and educational data mining applications: bibliometric and ChatGPT-based analysis of research publications from 2014 to 2023," *Journal of Computers in Education*, vol. 13, no. 1, pp. 411–446, Mar. 2026, doi: [10.1007/s40692-025-00362-1](https://doi.org/10.1007/s40692-025-00362-1).
- [4] S. KOTSIANTIS, C. PIERRAKEAS, and P. PINTELAS, "PREDICTING STUDENTS' PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES," *Applied Artificial Intelligence*, vol. 18, no. 5, pp. 411–426, May 2004, doi: [10.1080/08839510490442058](https://doi.org/10.1080/08839510490442058).
- [5] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, New York, NY, USA: ACM, Apr. 2012, pp. 267–270, doi: [10.1145/2330601.2330666](https://doi.org/10.1145/2330601.2330666).
- [6] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, Apr. 2019, doi: [10.1016/j.tele.2019.01.007](https://doi.org/10.1016/j.tele.2019.01.007).
- [7] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1432–1462, Mar. 2014, doi: [10.1016/j.eswa.2013.08.042](https://doi.org/10.1016/j.eswa.2013.08.042).
- [8] M. Ş. Bülbül, "Social Learning Analytics," in *Educational Technology and Society*, vol. 15, no. 3, 2021, pp. 190–206, doi: [10.4018/978-1-7998-7103-3.ch009](https://doi.org/10.4018/978-1-7998-7103-3.ch009).
- [9] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, 2010, doi: [10.1007/s11192-009-0146-3](https://doi.org/10.1007/s11192-009-0146-3).
- [10] H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, 2019, doi: [10.1007/s11042-018-6894-4](https://doi.org/10.1007/s11042-018-6894-4).
- [11] J. Han, G. Liu, and Y. Yang, "Latent Dirichlet Allocation-Based Topic Mining Analysis of Educational Scientific Research Projects Based on 2360 NSF Education Projects," *TEM Journal*, vol. 12, no. 2, 2023, doi: [10.18421/TEM122-32](https://doi.org/10.18421/TEM122-32).
- [12] Yuhefizar, R. Watrianthos, and R. Komalasari, "An LDA Analysis for Topic Modeling of the RESTI Journal," in *2024 2nd International Symposium on Information Technology and Digital Innovation (ISITDI)*, IEEE, Jul. 2024, pp. 46–52, doi: [10.1109/ISITDI62380.2024.10796703](https://doi.org/10.1109/ISITDI62380.2024.10796703).
- [13] M. Callon, J. P. Courtial, and F. Laville, "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry," *Scientometrics*, vol. 22, no. 1, 1991, doi: [10.1007/BF02019280](https://doi.org/10.1007/BF02019280).
- [14] H. S. Batubara, N. Jalinus, F. Rizal, and R. Watrianthos, "Mapping the Frontier: A Bibliometric Analysis of AI in Tertiary Education," *International Journal of Learning, Teaching and Educational Research*, vol. 23, no. 10, pp. 62–81, Oct. 2024, doi: [10.26803/ijlter.23.10.4](https://doi.org/10.26803/ijlter.23.10.4).
- [15] M. Muskhair, A. Luthfi, R. Watrianthos, U. Usmeldi, A. Fortuna, and A. Dwinggo Samala, "Emerging Research on Virtual Reality Applications in Vocational Education: A Bibliometric Analysis," *Journal of Information Technology Education: Innovations in Practice*, vol. 23, p. 005, 2024, doi: [10.28945/5284](https://doi.org/10.28945/5284).
- [16] R. Watrianthos, S. Triono Ahmad, and M. Muskhair, "Charting the Growth and Structure of Early ChatGPT-Education Research: A Bibliometric Study," *Journal of Information Technology Education: Innovations in Practice*, vol. 22, pp. 235–253, 2023, doi: [10.28945/5221](https://doi.org/10.28945/5221).
- [17] S. T. Ahmad, R. Watrianthos, A. D. Samala, M. Muskhair, and G. Dogara, "Project-based Learning in Vocational Education: A Bibliometric Approach," *International Journal Modern Education and Computer Science*, vol. 15, no. 4, pp. 43–56, 2023, doi: [10.5815/ijmecs.2023.04.04](https://doi.org/10.5815/ijmecs.2023.04.04).
- [18] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *ACM International Conference Proceeding Series*, 2006, doi: [10.1145/1143844.1143859](https://doi.org/10.1145/1143844.1143859).
- [19] W. Wiranto and M. R. Uswatunnisa, "Topic Modeling for Support Ticket using Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 6, pp. 998–1005, 2022, doi: [10.29207/resti.v6i6.4542](https://doi.org/10.29207/resti.v6i6.4542).
- [20] X. Wang and A. McCallum, "Topics over time," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA: ACM, Aug. 2006, pp. 424–433, doi: [10.1145/1150402.1150450](https://doi.org/10.1145/1150402.1150450).
- [21] L. Waltman and N. J. van Eck, "Field-normalized citation impact indicators and the choice of an appropriate counting method," *J. Informetr.*, vol. 9, no. 4, 2015, doi: [10.1016/j.joi.2015.08.001](https://doi.org/10.1016/j.joi.2015.08.001).
- [22] P. O. Seglen, "The skewness of science," *Journal of the American Society for Information Science*, vol. 43, no. 9, pp. 628–638, Oct. 1992, doi: [10.1002/\(SICI\)1097-4571\(199210\)43:9<628::AID-ASIS5>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASIS5>3.0.CO;2-0).
- [23] M. Thelwall and K. Kousha, "ResearchGate versus Google Scholar: Which finds more early citations?," *Scientometrics*, vol. 112, no. 2, 2017, doi: [10.1007/s11192-017-2400-4](https://doi.org/10.1007/s11192-017-2400-4).
- [24] T. R. Hinkin, J. B. Tracey, and C. A. Enz, "Scale Construction: Developing Reliable and Valid Measurement Instruments," *Journal of Hospitality & Tourism Research*, vol. 21, no. 1, pp. 100–120, Feb. 1997, doi: [10.1177/109634809702100108](https://doi.org/10.1177/109634809702100108).
- [25] R. J. Turner, "Book review," *J. Minim. Invasive Gynecol.*, vol. 14, no. 6, p. 776, Nov. 2007, doi: [10.1016/j.jmig.2007.07.001](https://doi.org/10.1016/j.jmig.2007.07.001).
- [26] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 11, 2015, doi: [10.1002/asi.23329](https://doi.org/10.1002/asi.23329).
- [27] R. Haunschild, L. Bornmann, and W. Marx, "Climate change research in view of bibliometrics," *PLoS One*, vol. 11, no. 7, 2016, doi: [10.1371/journal.pone.0160393](https://doi.org/10.1371/journal.pone.0160393).

- [28] C. Chen, "Science Mapping: A Systematic Review of the Literature," *Journal of Data and Information Science*, vol. 2, no. 2, 2017, doi: [10.1515/jdis-2017-0006](https://doi.org/10.1515/jdis-2017-0006).
- [29] S. Samsir, R. S. Saragih, S. Subagio, R. Aditiya, and R. Watrianthos, "BERTopic Modeling of Natural Language Processing Abstracts: Thematic Structure and Trajectory," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, p. 1514, Jul. 2023, doi: [10.30865/mib.v7i3.6426](https://doi.org/10.30865/mib.v7i3.6426).
- [30] W. Zhao, J. Xia, X. Jiang, and T. He, "A novel framework for deep knowledge tracing via gating-controlled forgetting and learning mechanisms," *Inf. Process. Manag.*, vol. 60, no. 1, 2023, doi: [10.1016/j.ipm.2022.103114](https://doi.org/10.1016/j.ipm.2022.103114).
- [31] S. Knight, S. Buckingham Shum, and K. Littleton, "Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space," *Journal of Learning Analytics*, vol. 1, no. 2, 2014, doi: [10.18608/jla.2014.12.3](https://doi.org/10.18608/jla.2014.12.3).
- [32] L. Breiman, "Statistical modeling: The two cultures," *Statistical Science*, vol. 16, no. 3, 2001, doi: [10.1214/ss/1009213726](https://doi.org/10.1214/ss/1009213726).
- [33] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019, doi: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x).
- [34] Z. A. Pardos, S. Tang, D. Davis, and C. Vu Le, "Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework," in *L@S 2017 - Proceedings of the 4th (2017) ACM Conference on Learning at Scale*, 2017, doi: [10.1145/3051457.3051471](https://doi.org/10.1145/3051457.3051471).
- [35] N. Selwyn, L. Pangrazio, S. Nemorin, and C. Perrotta, "What might the school of 2030 be like? An exercise in social science fiction," *Learn. Media Technol.*, vol. 45, no. 1, 2020, doi: [10.1080/17439884.2020.1694944](https://doi.org/10.1080/17439884.2020.1694944).
- [36] B. Williamson, "Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education," *BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY*, vol. 50, no. 6, pp. 2794–2809, Nov. 2019, doi: [10.1111/bjet.12849](https://doi.org/10.1111/bjet.12849).
- [37] L. Egghe, "Little science, big science... and beyond," *Scientometrics*, vol. 30, no. 2–3, 1994, doi: [10.1007/BF02018109](https://doi.org/10.1007/BF02018109).
- [38] V. P. Guerrero-Bote and F. Moya-Anegón, "A further step forward in measuring journals' scientific prestige: The SJR2 indicator," *J. Informetr.*, vol. 6, no. 4, 2012, doi: [10.1016/j.joi.2012.07.001](https://doi.org/10.1016/j.joi.2012.07.001).
- [39] K. J. Dougherty and R. S. Natow, "Performance-based funding for higher education: how well does neoliberal theory capture neoliberal practice?," *High. Educ. (Dordr.)*, vol. 80, no. 3, 2020, doi: [10.1007/s10734-019-00491-4](https://doi.org/10.1007/s10734-019-00491-4).
- [40] J. Pane, E. Steiner, M. Baird, and L. Hamilton, *Continued Progress: Promising Evidence on Personalized Learning: Executive Summary*. 2015. doi: [10.7249/rr1365.1](https://doi.org/10.7249/rr1365.1).
- [41] R. Azevedo, "Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues," *Educ. Psychol.*, vol. 50, no. 1, 2015, doi: [10.1080/00461520.2015.1004069](https://doi.org/10.1080/00461520.2015.1004069).
- [42] R. S. Baker and A. Hawn, "Algorithmic Bias in Education," *Int. J. Artif. Intell. Educ.*, vol. 32, no. 4, 2022, doi: [10.1007/s40593-021-00285-9](https://doi.org/10.1007/s40593-021-00285-9).
- [43] Z. A. Pardos and W. Jiang, "Designing for serendipity in a university course recommendation system," in *ACM International Conference Proceeding Series*, 2020. doi: [10.1145/3375462.3375524](https://doi.org/10.1145/3375462.3375524).