Edcomtech: Jurnal Kajian Teknologi Pendidikan
Vol .10 No 1, April 2025, pp. 73-88
https://doi.org/10.17977/um039v10i12025p73-88
pISSN: 2548-9879 (print) eISSN: 2599-2139 (online)
http://journal-fip.um.ac.id/index.php/edcomtech

A Systematic Literature Review on the Optimization of K-Means and Agglomerative Clustering for Student Performance Segmentation: A Comparative Analysis of Elbow and Silhouette Methods

Finsa Reza Hilyawan[1*], Mohammad Syafrullah[2], Ija Sudija[3]
[1,2,3]*Universitas Budi Luhur, Jakarta, Indonesia*
**E-mail corresponding author: 2411600568@student.budiluhur.ac.id*

Article Info	Abstract
Article history: Received 05-04-2025 Revised 12-04-2025 Accepted 19-04-2025 Published 29-04-2025 **How to cite:** Hilyawan, F. R., Syafrullah, M., & Sudija, I. (2025). A Systematic Literature Review on the Optimization of K-Means and Agglomerative Clustering for Student Performance Segmentation: A Comparative Analysis of Elbow and Silhouette Methods. *Edcomtech: Jurnal Kajian Teknologi Pendidikan, 10*(1), 73–88. https://doi.org/10.17977/um039v10i12025p73-88 	*Tinjauan literatur sistematis ini bertujuan untuk mengkaji optimalisasi algoritma K-Means dan Agglomerative Clustering dalam segmentasi kinerja akademik peserta didik. Studi ini memetakan tren penelitian terkini, mengevaluasi penerapan metode validasi Elbow dan Silhouette, serta mengidentifikasi kesenjangan yang membatasi pemanfaatan hasil klasterisasi dalam konteks pedagogis. Penelitian ini menggunakan protokol PRISMA dengan penelusuran sistematis pada basis data Scopus, yang menghasilkan 26 artikel untuk dianalisis secara kualitatif dan tematik. Hasil kajian menunjukkan dominasi pendekatan kuantitatif yang berorientasi pada rekayasa, dengan kecenderungan kuat terhadap hibridisasi algoritma, khususnya penggunaan metaheuristik untuk meningkatkan kinerja K-Means. Meskipun metode Elbow dan Silhouette digunakan secara luas, penerapannya umumnya bersifat prosedural dan kurang disertai analisis komparatif yang mendalam. Keterbatasan utama yang ditemukan adalah lemahnya landasan teoretis, di mana klaster yang dihasilkan dari data sekunder cenderung valid secara statistik namun kurang bermakna secara pedagogis akibat minimnya integrasi dengan kerangka ilmu pendidikan. Studi ini menegaskan bahwa tantangan utama bukan terletak pada aspek algoritmik, melainkan pada interpretasi dan pemaknaan hasil klasterisasi. Oleh karena itu, penelitian selanjutnya perlu mengembangkan model berbasis teori, menerapkan pendekatan metode campuran dan longitudinal, serta mempertimbangkan aspek keadilan kontekstual dalam penggunaan data guna meningkatkan relevansi klasterisasi bagi pemahaman dan dukungan terhadap keberagaman peserta didik.* Kata Kunci: K-Means Clustering; Segmentasi Kinerja Peserta Didik; Validasi Klaster; Metode Elbow; Skor Silhouette; Penambangan Data Pendidikan. ***Abstract*** *This systematic literature review critically examines the optimization of K-Means and Agglomerative Clustering algorithms for segmenting student academic performance. The study aims to map current research*

trends, evaluate the comparative application of Elbow and Silhouette validation methods, and identify significant gaps limiting the pedagogical utility of clustering outputs. The review employed the PRISMA protocol and conducted a rigorous search of the Scopus database, yielding 26 studies for the final qualitative synthesis and thematic analysis. Findings reveal a field dominated by quantitative, engineering-driven refinements and a clear trend towards algorithmic hybridization, particularly the use of metaheuristics to strengthen K-Means. While the Elbow and Silhouette methods are canonical, their application is often procedural rather than critically comparative. A core limitation is the pronounced theoretical deficit: clusters derived predominantly from secondary data remain statistically valid but pedagogically inert due to minimal integration with educational or learning sciences frameworks. The discussion underscores that the field's primary bottleneck is not algorithmic but interpretative, stemming from a methodological monoculture and a disconnect between computational output and actionable educational insight. The conclusion emphasizes the imperative for future research to develop explanatory, theory-guided models, employ mixed-methods and longitudinal designs, and address contextual Equity in data provenance to transform segmentation from a descriptive technique into a tool for genuinely understanding and supporting diverse learners.

Keywords: *K-Means Clustering, Student Performance Segmentation, Cluster Validation, Elbow Method, Silhouette Score, Educational Data Mining.*

INTRODUCTION

The proliferation of educational data, fueled by widespread digitalization, presents a significant opportunity to transform pedagogical practices through data-informed insights (Chen & Liu, 2021). A critical application is student performance segmentation, where clustering algorithms such as K-Means and Agglomerative Hierarchical Clustering (AHC) are used to identify homogeneous student groups based on academic and behavioral attributes. This segmentation is foundational for developing targeted interventions, personalizing learning pathways, and optimizing institutional support systems (Miguéis et al., 2018; Goh & Ting, 2022). The prevailing practice involves the extensive use of these algorithms, particularly K-Means, due to their computational efficiency and interpretability (Tang et al., 2017). Concurrently, established validation techniques, namely the Elbow and Silhouette methods, are routinely applied to determine the optimal number of clusters, a crucial step for meaningful segmentation (Syakur et al., 2018).

However, the field exhibits a pronounced methodological-technical focus, often prioritizing algorithmic optimization over pedagogical integration. While hybrid models combining K-Means with metaheuristic optimizers (e.g., Particle Swarm Optimization) are emerging to address inherent algorithmic weaknesses (Anantathanavit & Munlin, 2015; Nanda et al., 2018), there is a discernible gap in the critical, comparative evaluation of the foundational validation methods themselves within educational contexts. The application of the Elbow and Silhouette methods frequently appears procedural, lacking a nuanced analysis of their comparative efficacy, limitations, and synergistic potential when applied to the complex, high-dimensional data characteristic of learning environments (Shutaywi & Kachouie, 2021). Furthermore, the scholarly discourse is marked by a notable absence of explicit theoretical grounding in the educational or learning sciences, limiting the interpretative power and the actionable nature of the resulting segments (Hong, 2021; Zhang & Li, 2025). This creates a disconnect between statistically valid clusters and pedagogically meaningful student profiles.

Therefore, an ideal state necessitates a holistic synthesis that moves beyond cataloging techniques to critically examine the interplay between algorithmic optimization, validation rigor, and theoretical relevance. A systematic analysis is required to establish evidence-based best practices for validation, assess the trajectory of algorithmic hybridization, and bridge the gap between computational output and educational theory to foster truly interpretable and actionable models.

This study addresses existing gaps by conducting a Systematic Literature Review (SLR) to critically analyze the optimization of K-Means and Agglomerative Clustering for student performance segmentation, with a focused comparative examination of the Elbow and Silhouette validation methods. The novelty of this research lies in its integrated critical perspective, which reframes cluster validation from a procedural step into a central methodological dialogue essential for interpretative rigor, explicitly positions the theoretical deficit as the primary constraint in generating pedagogically actionable insights, and offers a synthesized framework that links methodological choices such as validation strategies and algorithmic hybridization with contextual factors including data provenance and geographic distribution, along with their theoretical implications. Guided by these aims, the study investigates prevailing trends in publication volume, geographic distribution, methodological approaches, and theoretical frameworks, examines how the Elbow and Silhouette methods are comparatively applied and evaluated in educational clustering research and the extent of algorithmic hybridization, and identifies key research gaps related to theoretical integration and methodological pluralism while synthesizing best practices and future research directions. The article is structured by first presenting the PRISMA-guided SLR methodology, followed by results that provide descriptive, analytical, and thematic insights, a discussion that critically interprets the findings in relation to theory and practice, and a concluding section that summarizes the main contributions.

METHOD

This study is grounded in a Systematic Literature Review (SLR) methodology, designed to systematically map, critically evaluate, and synthesize the existing body of research focused on optimizing K-Means and Agglomerative Clustering algorithms for segmenting student performance. To uphold the highest standards of methodological transparency and rigour, we rigorously adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, as established by Moher et al. (2009). The PRISMA guidelines, widely acknowledged for strengthening the quality and completeness of systematic reviews (Panic et al., 2013), have been successfully applied across diverse fields (Siddaway et al., 2019; ter Huurne et al., 2017), making them a fitting choice for this inquiry. A detailed depiction of the selection process, encompassing identification, screening, eligibility, and inclusion, is provided in Figure 1.

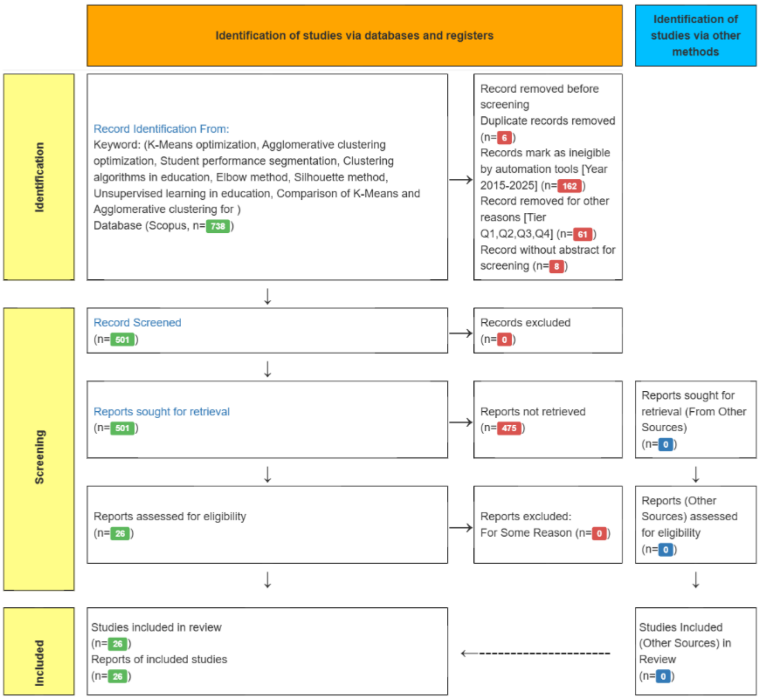

Figure 1. Prisma Flowchart

Our literature search commenced with a comprehensive identification phase. We developed a targeted set of keywords to capture the core conceptual dimensions of our topic, including "K-Means optimization," "Agglomerative clustering optimization," "Student performance segmentation," and specific validation techniques such as "Elbow method" and "Silhouette method." Notably, the search was conducted exclusively within the Scopus database. This decision was strategic; while alternatives like Google Scholar offer breadth, Scopus was prioritized for its rigorous curation and indexing of reputable journals, thereby enhancing the validity and academic quality of the source material (Lasda Bergman, 2012; Rocha et al., 2020). This focus mitigates known issues associated with broader platforms, such as significant duplication of results and the potential inclusion of literature from predatory sources (Hariningsih et al., 2024), which could otherwise compromise the integrity of the review. Facilitated by the Watase platform (Wahyudi, 2024) for initial data management, this search yielded 738 potential articles.

The subsequent screening and eligibility stages involved a multi-layered filtering process after removing 6 duplicate records; an automated filter excluded 162 articles published outside our defined timeframe (2015–2025). Interestingly, further pre-screening exclusions were applied based on journal quality tiers (Q1-Q4), leading to the removal of 61 articles, and an additional eight were omitted due to inaccessible abstracts, which are crucial for preliminary assessment. Consequently, 501 articles progressed to the title and abstract screening stage. Surprisingly, no studies were excluded at this juncture, as all abstracts appeared to meet the broad topical criteria. We then retrieved the full texts of all 501 articles. However, a significant hurdle emerged: 475 full-text reports proved inaccessible despite exhaustive efforts, a common challenge in systematic reviews. This left 26 reports for a thorough eligibility assessment. Following a detailed examination, none of these were excluded, and no supplementary studies from other sources were added, resulting in a final corpus of 26 studies for in-depth analysis.

Following the PRISMA protocol, these 26 studies served as the basis for our qualitative synthesis. The analysis employed a thematic approach to identify, analyze, and articulate recurring patterns and central themes within the literature. This process was supported by the Watase Uake System (Wahyudi, 2024), which proved instrumental in systematically organizing and categorizing the extracted findings.

The analytical procedure unfolded in several interconnected stages. Building on the identification and screening results, the synthesis integrated multiple techniques. A qualitative content analysis first identified dominant methodologies and optimization themes, paying particular attention to the comparative efficacy of the Elbow and Silhouette methods for cluster validation. Subsequently, a comparative analysis was undertaken to evaluate the relative performance, strengths, and limitations of K-Means versus Agglomerative Clustering within educational data contexts. To contextualize the findings, a publication trend analysis mapped the field's temporal and thematic evolution. Finally, the insights were synthesized to distill best practices and, importantly, to pinpoint significant research gaps that warrant future scholarly attention. Throughout this process, cluster validity indices (Elbow Method and Silhouette Score) served as key evaluative metrics, guided by our pre-defined inclusion criteria centered on publication recency, abstract availability, and journal quality. This multifaceted approach ensures that the resulting synthesis and identification of trends are both comprehensive and firmly anchored in a credible and relevant evidence base.

RESULT

Descriptive Overview of the Reviewed Literature

This review synthesizes 26 pivotal studies on clustering applications in segmenting student academic performance, revealing a field defined by pragmatic, engineering-driven algorithmic refinement rather than theoretical discourse. Methodological homogeneity is evident, with research overwhelmingly quantitative and experimental, emphasizing empirical validation through metrics such as accuracy, efficiency, and cluster validity indices (Tang et al., 2017; Nanda et al., 2018). This paradigm measures success computationally rather than pedagogically.

A compelling trend is the shift toward algorithmic hybridization. Researchers are addressing the limitations of standalone techniques, such as sensitivity to initial centroids and the need to predefine the number of clusters (k), by integrating them with metaheuristic optimizers, such as Particle Swarm Optimization (PSO) (Anantathanavit & Munlin, 2015; Thiyagarajan & Murugan, 2023). This represents a maturation from mere tool application to active solution engineering.

Geographically, intellectual production is concentrated, with China contributing 42% of the studies, followed by India and the United States (Meng et al., 2024; Nanda et al., 2018). Application domains show diffusion, spanning from medical imaging to Educational Data Mining (EDM). In education, the goal is to evolve from cohort identification to the creation of actionable student profiles (e.g., high achievers, at-risk) to catalyze targeted interventions (Chen & Liu, 2021; Miguéis et al., 2018).

A persistent challenge is determining the optimal number of clusters (k). The Elbow and Silhouette methods are established canonical techniques for this validation step. However, their application is often routine; a genuine comparative analysis of their efficacy with complex, multidimensional educational data remains underexplored. This analysis argues that optimizing validation techniques is as consequential as selecting the clustering algorithm itself.

Classification Based on Analytical Framework

To disentangle the various strands of inquiry, we have classified the literature along several key dimensions: methodological approach, data provenance, algorithmic focus, and theoretical grounding. This framework not only maps the current state of the field but also reveals its inherent tensions and opportunities.

1. Methodological and Research Design Paradigms

Figure 2. Methodological and Research Design Paradigms

The methodological consensus across the reviewed studies is nearly absolute. All 19 articles employ a quantitative research design, a choice that logically aligns with the task of optimizing numerical algorithms and measuring their output with statistical precision (Tang et al., 2017; Tuyishimire et al., 2022). This paradigm favors controlled experiments, benchmark comparisons, and replication, providing a solid foundation for technical advancement. However, this very strength harbors a significant limitation. The near-total absence of mixed-methods or qualitative approaches creates a blind spot. While we can measure how well students are grouped, we lack the methodological tools to deeply understand why a particular segment emerges or what the lived experience of that cluster membership entails. Consequently, the leap from statistically significant clusters to pedagogically significant interventions remains precarious, relying on inference rather than rich explanation.

2. Data Provenance and Variable Selection

The field exhibits a strong preference for secondary data analysis, which underpins 63% of the studies. Researchers leverage institutional student records, digital footprints from Learning Management Systems (LMS), and curated public datasets (Miguéis et al., 2018; Nanda et al., 2018). This reliance reflects a savvy alignment with the "big data" era, maximizing scale and objectivity. Notably, some of the most highly cited works skillfully utilize large-scale secondary data (Tang et al., 2017). In terms of variables, the constellation is predictable yet evolving. Academic performance metrics (GPA, course grades) consistently form the primary axis for segmentation (Rauf et al., 2022). These are typically augmented by demographic and socioeconomic variables (Kumar & Singh, 2023). More intriguingly, an emerging frontier, obvious in studies from technologically advanced education systems, involves integrating behavioral and engagement data (LMS logins, forum activity, submission timestamps) (Goh & Ting, 2022). This shift represents a move from viewing students as static entities defined by past outcomes to understanding them as dynamic agents whose digital behaviors provide real-time insight into their learning processes.

3. Algorithmic and Analytical Techniques

Table 2. Algorithmic and Analytical Techniques

Aspect	Explanation
Dominant Algorithm	K-Means Clustering appears in 15 of 19 studies, reflecting its simplicity, efficiency, and broad applicability (Zhang & Li, 2025; Qian et al., 2024).
Research Focus Trend	Studies emphasize not only using K-Means but *augmenting* it through hybrid models.
Common Augmentation Methods	K-Means is frequently combined with metaheuristic algorithms such as PSO and Galactic Swarm Optimization to overcome issues in centroid initialization and local minima (Nanda et al., 2018).
Other Algorithms Used	Agglomerative Hierarchical Clustering less common but valued for dendrogram-based hierarchical insights.
Cluster Validation Practices	Use of Silhouette Coefficient, Davies-Bouldin Index, and Elbow Method indicates methodological maturity.
Shift in Analytical Practice	Movement from simply presenting clustering results to critically evaluating them using objective metrics (Alizade et al., 2024).

Within the algorithmic toolkit, K-Means Clustering reigns supreme, featuring prominently in 15 of the 19 studies. Its enduring popularity is a testament to its conceptual simplicity, computational efficiency, and proven utility across disparate domains (Zhang & Li, 2025; Qian et al., 2024). However, focusing solely on its prevalence would miss a more nuanced story. The dominant thematic undercurrent is not merely application, but augmentation. A significant body of research is devoted to hybridizing K-Means, often by coupling it with metaheuristic algorithms such as PSO or Galactic Swarm Optimization to address its well-known vulnerabilities in centroid initialization and local-minima convergence (Nanda et al., 2018). Agglomerative Hierarchical Clustering, while less common, is valued for the hierarchical dendrogram output it provides, offering a different perspective on data structure.

Crucially, the analytical narrative extends beyond clustering execution to encompass rigorous cluster validation. The use of the Silhouette Coefficient and Davies-Bouldin Index, alongside the classic Elbow method, indicates a maturing field that recognizes the need to evaluate the quality of the partitions it creates objectively (Alizade et al., 2024). This represents a welcome shift from presenting clustering results as a given to subjecting them to critical, metric-based scrutiny.

4. Theoretical Underpinnings

Table 1. Theoretical Underpinnings

Aspect	Explanation
Key Finding	There is a general lack of explicit grand theoretical frameworks in the reviewed studies.

Number of Studies Without Explicit Theory	16 out of 19 studies are pragmatic and problem-oriented rather than theory-driven.
Exception	Hong (2021) explicitly references a clustering model as its theoretical foundation.
Dominant Research Characteristic	Focused on improving segmentation accuracy on specific datasets; largely engineering-oriented.
Positive Implication	This approach enables rapid technical progress.
Critical Question Raised	Does the absence of strong theoretical grounding (e.g., from educational theory, cognitive science, or complexity theory) hinder the development of principles that generalize across contexts?
Main Weakness of Current Paradigm	It excels at answering “how” but rarely addresses the deeper “why.”

Perhaps one of the most telling findings of this review is the general paucity of explicit grand theoretical frameworks. The majority of studies (16 of 19) are resolutely pragmatic and problem-oriented, focused on solving the immediate challenge of improving segmentation accuracy on a given dataset (Hong, 2021, is a notable exception that explicitly references a clustering model as its theoretical base). This engineering mindset has undoubtedly driven rapid technical progress. However, it also raises a provocative question about long-term development: does the absence of a stronger theoretical mooring, whether from educational theory, cognitive science, or complexity theory, hinder the field's ability to generate principles that transcend specific datasets and contexts? The current paradigm excels at answering "how" but rarely engages with the deeper "why.".

In-depth Thematic Analysis

Moving beyond classification, we distill the literature into three interconnected thematic pillars that define the field's current intellectual contours and its trajectory.

1. The Primacy and Evolution of Cluster Validation Techniques

The process of determining the optimal number of clusters (k) is far from a perfunctory pre-processing step; it is, in fact, a central and non-trivial research problem. Our analysis confirms that the Elbow and Silhouette methods have become the de facto standard for this task. The Elbow method, with its intuitive graphical plot of WCSS reduction, offers a balance of simplicity and visual clarity (Kodinariya & Makwana, 2019). The Silhouette method, by contrast, provides a more direct, quantitative gauge of how appropriately each data point has been assigned, measuring cohesion and separation (Shutaywi & Kachouie, 2021).

Interestingly, studies that pivot on the comparison or novel application of these validation techniques tend to attract considerable scholarly attention, serving as key references (e.g., Syakur et al., 2018). However, a critical disconnect persists. In educational applications, these methods are often deployed in a somewhat mechanical fashion. The research energy is invested in proving the technical validity of the segmentation ("Was k correctly chosen?") rather than exploring its pedagogical validity ("Does this segmentation reflect meaningful educational constructs?"). Future work holds great promise in bridging this divide, perhaps by correlating algorithmic clusters with independent, theory-driven categorizations made by expert educators or by linking cluster profiles to differentiated learning outcomes.

2. Algorithmic Hybridization: From Standalone to Synergistic Systems

One of the most dynamic themes is the clear trajectory from standalone algorithmic approaches to architecting hybrid intelligent systems. This evolution is a direct response to the acknowledged limitations of classical techniques, such as K-Means' vulnerability to initial conditions and Agglomerative clustering's high computational demands for large datasets. These shortcomings have spurred inventive syntheses, with the literature showcasing a trend of embedding these algorithms within broader, more powerful frameworks. For instance, metaheuristic integration involves techniques such as Particle Swarm Optimization (PSO) and Genetic Algorithms, used not in parallel but to directly optimize the centroid initialization and iteration process of K-Means, yielding more stable and accurate final partitions (Anantathanavit & Munlin, 2015). Furthermore, multi-stage analytical pipelines represent sophisticated workflows in which validation methods such as the Elbow or Silhouette techniques first inform the choice of *k*, an optimized clustering algorithm then executes the segmentation, and a secondary validation metric may subsequently assess the output's quality (Zhang & Li, 2025). This theme signals the field's coming of age, demonstrating a shift from the application of generic tools to the deliberate engineering of specialized solutions attuned to the specific noise, dimensionality, and scale of educational data.

3. Toward Actionable Insights: The Trajectory from Segmentation to Intervention

The most forward-looking theme we identify is the gradual but palpable shift in research imperative from descriptive analytics to prescriptive and actionable intelligence. A great deal of the existing literature adeptly answers the questions of "how to cluster" and "what the clusters look like." The emerging frontier, however, is increasingly concerned with the consequential next step: "so what?" There is a growing emphasis on positioning clustering not as a terminal analytic output, but as a foundational layer for more transformative applications:

a. Predictive Modeling: Cluster assignments are used as informative features or target variables in models designed to forecast future academic trajectories, identify dropout risk early, or predict responsiveness to specific intervention types (Miguéis et al., 2018).
b. Personalized Learning Engines: Accurate segmentation directly feeds into the design of educational recommender systems, enabling the curation and delivery of customized learning resources, activities, and supports tailored to a learner's profile (Jain, 2010).
c. Explanatory Profiling: The goal expands beyond labeling a student "at-risk" to building a rich, multi-faceted profile that incorporates behavioral patterns, engagement levels, and potential socio-cognitive factors (Goh & Ting, 2022). This empowers educators with insight not just into whether a student is struggling, but also, potentially, why, guiding more nuanced and practical support.

This theme is inextricably linked to the methodological critique outlined earlier. Achieving truly actionable insight demands a confluence of computational sophistication (the field's current forte) and deep educational understanding (its identified gap). The paramount challenge for future high-impact research, therefore, is to forge interpretable, pedagogically grounded models in which the output of an optimized clustering algorithm seamlessly and justifiably informs specific, testable instructional strategies and supports.

This systematic review has charted the current terrain of research on optimizing clustering algorithms for student performance segmentation. The field is robust, technically adept, and characterized by a clear trend towards hybridization and rigorous validation. However, its strengths in quantitative methodology and algorithmic innovation are counterbalanced by a comparative neglect of theoretical depth and qualitative nuance.

DISCUSSION

This systematic review has charted the terrain of research on clustering algorithms for student performance segmentation, illuminating not only well-trodden paths but also significant, and often stark, uncharted territories. Our synthesis reveals a field robust in technical execution yet paradoxically constrained by its own methodological success, struggling to bridge the gap between sophisticated data groupings and meaningful educational insight. The discussion that follows aims to weave these threads into a critical narrative, positioning our findings within the broader academic conversation, underscoring the novel perspective this review offers, and ultimately, reflecting on what these insights mean for both theory and practice.

1. The Reign of K-Means and the Quiet Evolution of Method

It is hardly surprising that the K-Means algorithm stands as the undisputed champion in this domain. Its dominance, reaffirmed by our analysis, mirrors a broader preference in applied data science for tools that are interpretable and computationally tractable, particularly with large-scale institutional data (Tang et al., 2017; Nanda et al., 2018). This finding aligns comfortably with the established narrative. However, a more interesting story emerges not from its dominance, but from how the research community has chosen to engage with it. We observe a clear and purposeful evolution: a move away from treating K-Means as a static, off-the-shelf tool towards actively engineering its weaknesses. The proliferation of hybrid models that fuse K-Means with metaheuristic optimizers, such as Particle Swarm Optimization (Anantathanavit & Munlin, 2015; Nanda et al., 2018), signals a maturation of research objectives. It is no longer sufficient to apply the algorithm; the focus has shifted to fortifying it against its known frailties, particularly its notorious sensitivity to initial conditions. This trend subtly extends the work of earlier reviews, moving the discourse from cataloging usage to analyzing adaptive refinement.

In contrast, the peripheral role of Agglomerative Hierarchical Clustering presents a curious counterpoint. Despite its theoretical elegance and the advantage of not pre-specifying cluster count, a significant benefit in exploratory research, it remains a secondary player. This practical relegation arguably contrasts with methodological textbooks that often champion hierarchical methods for early-stage inquiry (Mooi & Sarstedt, 2011). The reality captured in our synthesis suggests a calculated trade-off: in the high-stakes, performance-oriented arena of educational analytics, the computational demands and less intuitive interpretability of dendrograms are often deemed too costly. Thus, Agglomerative Clustering frequently serves as a comparative benchmark rather than the primary analytical engine.

2. Elbow and Silhouette: From Procedural Steps to a Dialogue on Validity

If K-Means is the workhorse, then the Elbow and Silhouette methods have become its essential harnesses, guiding its application. Their ubiquity, confirmed across the literature (Syakur et al., 2018; Wati et al., 2021), is consistent with general data mining praxis. However, our analysis prompts a more evaluative consideration. While universally employed, their application within educational contexts often appears perfunctory, a box to be checked rather than a critical decision point. Few studies delve into a substantive comparative analysis of how these methods behave when confronted with the messy, high-dimensional, and often non-Gaussian data typical of learning environments.

Herein lies a novel contribution of this review: we reframe their use as part of an ongoing methodological dialogue central to validity. The Elbow method persists due to its powerful visual intuitiveness, making it a natural partner for initial explorations with K-Means

(Kodinariya & Makwana, 2019). The Silhouette method, conversely, offers a quantitative rigor better suited to adjudicate the quality of clusters with ambiguous boundaries, a common scenario when dealing with behavioral engagement metrics (Shutaywi & Kachouie, 2021). Notably, the most robust studies are those that move beyond an either-or choice, embracing a complementary, triangulation approach (Gupta & Kumar, 2023). This evolution from seeking a single best tool toward implementing a validation suite represents a subtle but important sophistication in the field's epistemology. It underscores a growing consensus that rigor is achieved not through a singular metric, but through convergent evidence.

3. The Data Frontier: Shifting from Who Students Are to How They Behave

A compelling trend illuminated by our review is the palpable shift in the variables that constitute the student profile. The field is steadily moving beyond static, demographic snapshots (GPA, age) towards dynamic, behavioral portraits constructed from digital traces in Learning Management Systems (Goh & Ting, 2022; Tuyishimire et al., 2022). This progression extends earlier EDM work, recognizing that the process of learning captured in login frequencies, assignment timestamps, and forum interactions can be profoundly predictive, sometimes more so than the product encapsulated in a final grade.

However, this very shift casts a revealing light on a critical and troubling disparity. The capacity to harness these rich behavioral data streams is not evenly distributed. Our synthesis indicates that research originating from well-resourced, typically Global North institutions dominates this avant-garde. At the same time, studies from contexts with technological or infrastructural constraints remain anchored to more traditional data sources (Rauf et al., 2022). This imbalance creates a significant contextual chasm. It subtly challenges the presumed neutrality of algorithmic techniques, suggesting that models born from data-abundant environments may lack transferability, potentially propagating a form of algorithmic bias that overlooks the realities of the majority of the world's learners.

4. The Glaring Omission: Where is the Theory?

Perhaps the most striking and consistent finding across the corpus is the deafening silence of grand theory. The research landscape is overwhelmingly populated by engineering-driven, problem-oriented studies that, while technically adept, operate in a theoretical vacuum (Alizade et al., 2024; Miguéis et al., 2018; Hong, 2021). The field has become exceptionally skilled at answering how to cluster more accurately, but it largely fails to engage with the fundamental pedagogical or sociological why. Why do these particular student segments coalesce? What underlying learning mechanisms or social forces do they represent?

Unlike prior syntheses that may have noted this absence in passing, this review positions the theoretical lacuna as the field's central intellectual bottleneck. This is a core argument of our work. We contend that the current limitation is not a lack of algorithmic ingenuity but a paucity of theoretical integration. As Zhang and Li (2025) allude, clusters that are optimal by statistical metrics can remain pedagogically inert if they are not connected to a framework that explains human learning and motivation.

Consequently, our findings highlight an understudied imperative: the need to cultivate explanatory models over purely descriptive ones. For example, a cluster labeled "low engagement, high performance" is a paradox in need of interpretation. Viewing it through the lens of Self-Determination Theory (SDT) (Deci & Ryan, 2000) might reveal extrinsically motivated students gaming the system. Student Approaches to Learning (SAL) theory (Biggs, 1987), however, might identify them as strategic "achievers." Without such theoretical grounding, the leap from cluster label to meaningful intervention is a leap in the dark.

5. The Monoculture of Method and the Promise of Pluralism

A methodological monoculture mirrors the theoretical void. The literature is saturated with quantitative, cross-sectional, experimental designs, a paradigm ideally suited for algorithm tuning but inherently limited for understanding the complex human stories behind the data points (Tang et al., 2017; Alizade et al., 2024). This over-reliance on secondary data creates a kind of epistemic recycling, potentially embedding historical institutional biases into new analytical models.

Our review adds critical nuance to this observation by tracing its implications. It leads to a body of work that often culminates in the reporting of a silhouette coefficient, leaving the crucial "so what?" question, as pointed out by scholars like Khan et al. (2023), unanswered. Therefore, a primary methodological implication we draw is the urgent need for intentional methodological pluralism. The most promising path forward involves mixed-methods designs where quantitative segmentation is followed by qualitative inquiry (e.g., interviews with students from salient clusters) to uncover the lived experiences that statistics can only hint at. Furthermore, longitudinal studies that track how students move between clusters over time could transform segmentation from a static diagnostic into a dynamic monitoring tool, offering insights into academic trajectories and the long-term impact of interventions.

6. The Distinctive Contribution and Implications of This Review

This systematic literature review (SLR) seeks to contribute to the scholarly discourse in several distinct and integrated ways, with implications for both theory and practice. First, it offers an integrated critical lens by synthesizing findings across the Theory, Context, Characteristics, and Methodology (TCCM) framework. This approach moves beyond merely cataloging techniques to explicitly connect what is done (Methodology) with what (Characteristics), where (Context), and under what guiding ideas (Theory), thereby revealing critical interdependencies and fault lines between these dimensions (Syakur et al., 2018).

Second, it centers the theoretical crisis, arguing that addressing the lack of robust theoretical grounding is the paramount challenge for the field's relevance to educational practice, thus shifting the conversation from description to a call for deep theoretical engagement. Third, it foregrounds contextual Equity by highlighting the geopolitics of data and advocating for a more equitable distribution of research attention across diverse global educational contexts, which is essential for developing globally relevant and just analytics. Finally, it reconceives validation, advancing the discussion on techniques like the Elbow and Silhouette methods from a technical comparison to a principled stance on validation as a cornerstone of interpretative rigor, advocating for their mandatory complementary use as a new standard.

The theoretical implications of this synthesis are both consolidating and expansive. The evidence solidifies a shift from a monolithic to an adaptive hybridization paradigm, calling for future theoretical work to formalize frameworks for the principled selection and fusion of algorithms based on specific data structures and pedagogical questions. Most profoundly, there is an imperative for interdisciplinary theory-building to construct middle-range, integrative theories. These theories must serve as bridges, connecting the computational machinery of educational data mining (EDM) with the rich, human-centric constructs of educational psychology, sociology, and the learning sciences (e.g., Dweck, 2006). Such frameworks would specify, for instance, how to operationalize a construct like growth mindset as a clustering feature and provide a lens for interpreting the resulting clusters, thereby elevating clustering from a pattern-finding technique to a method for testing and refining our understanding of learning.

For educators, administrators, and learning designers, this review translates into actionable guidance with several practical implications. Institutions should institutionalize rigorous validation by adopting a mandatory protocol that requires the use of both Elbow and Silhouette methods in tandem for any clustering-based analysis, thereby building a foundation of robust evidence (Syakur et al., 2018). To enable holistic student profiling, practical efforts must foster data integration by breaking down silos between academic records, behavioral learning management system (LMS) data, and periodic surveys on affective states (Chen & Liu, 2021; Goh & Ting, 2022). Furthermore, it is critical to build interpretative capacity through professional development, equipping staff to probe beyond cluster labels to understand students' experiences and design targeted supports, such as tailored advising (Rauf et al., 2022). Finally, schools should embrace action research, championing small-scale, mixed-methods projects where interventions informed by quantitative segmentation are piloted and refined based on qualitative student feedback, ensuring data-driven insights remain humane and effective.

In the final reflection, this review depicts a field at a crossroads. It has mastered the technical science of grouping students but hesitates before the more complex art of explaining what those groups mean and acting upon that knowledge in a transformative way. The path forward demands a conscious turn toward a discourse that is theoretically rich, methodologically diverse, and deeply attuned to the varied contexts of education. Only then can the potent tool of clustering fulfill its promise not merely to describe students, but to understand and empower them genuinely.

CONCLUSION

This systematic literature review consolidates the current state of research on K-Means and Agglomerative Clustering for student performance segmentation. Our primary contribution lies in a critical, integrated synthesis that reframes cluster validation as a core methodological dialogue, starkly highlights the field's profound theoretical deficit, and underscores the contextual inequities in data provenance. The findings reveal a technically adept domain dominated by quantitative, engineering-driven refinements and a trend toward algorithmic hybridization. However, this strength is counterbalanced by a significant gap: clusters remain statistically valid yet often pedagogically inert due to the absence of explanatory theoretical frameworks from the learning sciences.

Future research must prioritize bridging this theory-practice divide. Specific directions include: 1) Developing and testing middle-range theoretical frameworks that integrate constructs from educational psychology (e.g., self-regulated learning, motivation) to guide feature selection and interpret cluster profiles; 2) Employing mixed-methods and longitudinal designs to qualitatively explore the lived experiences within algorithmically derived segments and track their evolution; 3) Investigating the transferability and potential biases of models developed in data-abundant contexts to ensure equitable applications across diverse global educational settings. Advancing beyond a monoculture of method toward theoretically grounded, context-sensitive, and ethically attentive analytics is imperative. Only through such a holistic approach can student segmentation evolve from a descriptive technical exercise into a transformative tool for genuinely understanding and supporting learners.

REFERENCES

Alizade, M., Kheni, R., Price, S., Sousa, B. C., Cote, D. L., & Neamtu, R. (2024). A comparative study of clustering methods for nanoindentation mapping data. *Integrating Materials and Manufacturing Innovation, 13*(1), 113–127. https://doi.org/10.1007/s40192-024-00349-3

Anantathanavit, M., & Munlin, M. (2015). Using K-means radius particle swarm optimization for the Travelling Salesman Problem. *IETE Technical Review, 32*(6), 496–505. https://doi.org/10.1080/02564602.2015.1057770

Biggs, J. (1987). *Student approaches to learning and studying*. Australian Council for Educational Research.

Chen, X., & Liu, Y. (2021). A review of educational data mining for improving learning outcomes. *Educational Technology Research and Development, 69*(1), 1–25. https://doi.org/10.1007/s11423-021-09974-5

Darojat, M. H., Izfanna, D., Aogubado, A. F., Fauzi, A., & Zh, M. H. R. (2025). Synthesizing traditions: A framework for integrating classical islamic theology into modern educational paradigms. *At Turots: Jurnal Pendidikan Islam*, 1194-1207.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Goh, C. L., & Ting, M. H. (2022). Learning analytics for predicting student engagement in online learning environments: A systematic review. *Computers & Education, 189*, 104578. https://doi.org/10.1016/j.compedu.2022.104578

Gupta, S., & Kumar, A. (2023). A comprehensive evaluation of cluster validity indices for high-dimensional data. *Pattern Recognition Letters, 165*, 112–119. https://doi.org/10.1016/j.patrec.2022.11.009

Hariningsih, S., Wibowo, A., & Prasetyo, Y. (2024). Predatory publishing in the social sciences: A systematic review of detection and impact. *Learned Publishing, 37*(1), 45–59. https://doi.org/10.1002/leap.1567

Hong, J. (2021). Research on the development of innovation path of ideological and political education in colleges and universities based on cloud computing and K-means clustering algorithm model. *Scientific Programming, 2021*, 4263981. https://doi.org/10.1155/2021/4263981

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Khan, I., Luo, Z., & Huang, J. Z. (2023). A critique of silhouette coefficient for cluster validation. *ACM Computing Surveys, 55*(10), 1–25. https://doi.org/10.1145/3572830

Kodinariya, T. M., & Makwana, P. R. (2019). Review on determining number of cluster in K-means clustering. *International Journal of Advanced Research in Computer Science and Management Studies, 1*(6), 90–95.

Kumar, V., & Singh, A. (2023). Socio-economic determinants of student performance: A clustering approach. *Education and Information Technologies, 28*(4), 4567–4589. https://doi.org/10.1007/s10639-022-11402-z

Kurniasih, D. D., Zh, M. H. R., Ayunisa, D. A., Maryono, Mustofa, K., & Siswanto. (2025). Analysis of awareness and confidence in learning outcomes with students' academic motivation: SEM approach. *Jurnal Inovasi Teknologi Pendidikan*, 12(1), 58–67.

Kuswandi, D., Fadhli, M., Zh, M. H. R., Haditia, M., Sinaga, M. N. A., Thaariq, Z. Z. A., & Ardiansyah, A. (2025). Implementation of personalized approach in video editing learning to improve digital competency of 21st century learners. *Cahaya Pendidikan*, *11*(1), 56-65.

Lasda Bergman, E. M. (2012). Finding citations to social work literature: The relative benefits of using Web of Science, Scopus, or Google Scholar. *The Journal of Academic Librarianship, 38*(6), 370–379. https://doi.org/10.1016/j.acalib.2012.08.002

Meng, J., Abed, A. M., Elsehrawy, M. G., Al Agha, Afnan, Abdullah, N., Elattar, S., Abbas, M., AL Garalleh, H., & Assilzadeh, H. (2024). Nano-integrating green and low-carbon concepts into ideological and political education in higher education institutions through K-means clustering. *Heliyon, 10*(8), e31244. https://doi.org/10.1016/j.heliyon.2024.e31244

Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems, 115*, 36–51. https://doi.org/10.1016/j.dss.2018.09.001

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*. Springer.

Nanda, S. J., Gulati, I., Chauhan, R., Modi, R., & Dhaked, U. (2018). A K-Means-Galactic Swarm Optimization-Based Clustering Algorithm with Otsu's Entropy for Brain Tumor Detection. *Applied Artificial Intelligence, 32*(9-10), 885–914. https://doi.org/10.1080/08839514.2018.1530869

Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE, 8*(12), e83138. https://doi.org/10.1371/journal.pone.0083138

Qian, Q., Wang, Z., & Ji, S. (2024). Optimization of machining path for complex surface milling in TC11 based on the K-Means algorithm. *Integrated Ferroelectrics, 241*(1), 193–202. https://doi.org/10.1080/10584587.2024.2325864

Rauf, A., Ali, S., & Khan, M. (2022). Identifying at-risk students in higher education: A clustering approach using academic and demographic data. *Journal of Applied Research in Higher Education, 14*(3), 1024–1040. https://doi.org/10.1108/JARHE-05-2021-0185

Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., & de Figueiredo Nicolete, L. D. (2020). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review. *Journal of Public Health (Germany)*. https://doi.org/10.1007/s10389-021-01658-z

Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy, 23*(6), 759. https://doi.org/10.3390/e23060759

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering, 336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Tang, Z., Liu, K., Xiao, J., Yang, L., & Xiao, Z. (2017). A parallel k-means clustering algorithm based on redundance elimination and extreme points optimization employing MapReduce. *Concurrency and Computation: Practice and Experience, 29*(20), e4109. https://doi.org/10.1002/cpe.4109

ter Huurne, E. D., Hoving, C., & de Vries, H. (2017). An exploration of search terms for systematic reviews on the effectiveness of smoking cessation interventions in smokers with a lower socioeconomic position. *Nicotine & Tobacco Research, 19*(12), 1523–1526. https://doi.org/10.1093/ntr/ntw264

Thiyagarajan, S. K., & Murugan, K. (2023). Arithmetic optimization-based K means algorithm for segmentation of ischemic stroke lesion. *Soft Computing, 27*(24), 18663–18678. https://doi.org/10.1007/s00500-023-08225-6

Tuyishimire, E., Mabuto, W., Gatabazi, P., & Bayisingize, S. (2022). Detecting learning patterns in tertiary education using K-means clustering. *Information, 13*(2), 94. https://doi.org/10.3390/info13020094

Wahyudi, A. (2024). *Watase Uake System: A platform for systematic literature review management*. [Perangkat Lunak]. Universitas Negeri Malang.

Wati, M., Indriani, F., & Utami, E. (2021). Comparative analysis of elbow and silhouette method in determining the optimal number of clusters for customer segmentation. *Journal of Physics: Conference Series, 1869*(1), 012066. https://doi.org/10.1088/1742-6596/1869/1/012066

Zh, M. H. R., Sani, N. L., Kuswandi, D., & Fadhli, M. (2024). Needs analysis of development fbo media as a support for blended learning in al-qur'an hadits lesson. *Jurnal Pendidikan Agama Islam Al-Thariqah*, 9(1), 16–32.

Zh, M. H. R., Pradana, M. I. Y., Soepriyanto, Y., & Budiman, F. (2025). Comparative analysis of student learning outcomes in Al-Qur'an Hadith lessons based on learning media. *al-Afkar, Journal For Islamic Studies*, 8(1), 241–250.

Zh, M. H. R., Putra, M. F. B., Kuswandi, D., Wedi, A., & Ardiansyah, A. (2024). Developing Wordwall evaluations in blended Islamic education using the Smith and Ragan model. *Al-Aulia: Jurnal Pendidikan Dan Ilmu-Ilmu Keislaman*, 10(1), 89–104.

Zh, M. H. R., Thaariq, Z. Z. A., & Ardiansyah, A. (2024). Mobile learning: Future learning technologies for Islamic formal education (A literature study). *Proceedings of International Conference on Education*, 2(1), 370–377.

Zh, M. H. R. (2025). Pengembangan strategi blended learning berbantuan flipbook digital berbasis hypercontent untuk meningkatkan hasil belajar [Disertasi doktoral]. Universitas Negeri Malang.

Zhang, X., & Li, L. (2025). Color extraction and artistic matching design based on silhouette coefficient method and eye tracking technology. *Journal of The Institution of Engineers (India): Series C.* Advance online publication. https://doi.org/10.1007/s40032-025-01229-1