Revista Eletrônica de Iniciação Científica em Computação, 2026, 24:1
ISSN: 3085-8461 • doi: 10.5753/reic.2026.8468

Research Paper

A Cluster-based Framework for Enrollment Forecasting in Brazilian Schools

Davi Vieira L. Correia [Universidade Federal de Alagoas (UFAL) | *davi.correia@nees.ufal.br*]
Lenardo Chaves e Silva [Universidade Federal Rural do Semi-Árido (UFERSA) | *lenardo.chaves@ufersa.br*]
Rafael de Amorim Silva [Universidade Federal de Alagoas (UFAL) | *rafael.amorim@nees.ufal.br*]
Breno J. D. da Costa [Universidade Federal de Alagoas (UFAL) | *breno.duarte@nees.ufal.br*]
Romildo S. Escarpini Filho [Universidade Federal de Alagoas (UFAL) | *romildo.escarpini@nees.ufal.br*]
Dalgoberto M. Pinho Júnior [Universidade Federal de Alagoas (UFAL) | *dalgoberto.pinho@nees.ufal.br*]
Nicholas Cruz [Universidade Federal de Alagoas (UFAL) | *nicholas.cruz@nees.ufal.br*]
Bruno Almeida Pimentel [Universidade Federal de Alagoas (UFAL) | *bruno.pimentel@nees.ufal.br*]

Núcleo de Excelência em Tecnologias Sociais (NEES), Universidade Federal de Alagoas (UFAL), Maceió, AL, Brazil.

Abstract. Forecasting enrollments for policies such as the Brazilian National Textbook Program (PNLD) is hindered by traditional global predictive models that ignore school heterogeneity and often produce algorithmic invisibility in vulnerable contexts. The study proposes a Context-Aware Forecasting Framework based on split modeling strategies, using Direct Bipartite Graphs and Clustering Techniques. Experimental results show that context-specific feature engineering enables drastic dimensionality reduction—compressing 20 features within "Cluster 1" into a single PCA signal and reducing "Cluster 2" to 20 relevant attributes, representing a 94% dimensionality reduction in 81.25% of the tuples — while maintaining comparable predictive performance.

Keywords: Enrollment Forecasting, Bipartite Graphs, Clustering, Algorithmic Invisibility, PNLD, Context-Aware Modeling

Received: 15 June 2026 • **Accepted:** 15 June 2026 • **Published:** 10 July 2026

1 Introduction

The Brazilian National Textbook Program (PNLD) represents one of the most impactful public policies in the Education Global South. Consolidated into a massive logistical operation, it serves over 35 million students in public basic education [Zambon and Terrazzan, 2013]. To ensure the equitable distribution of educational resources, the National Education Development Fund (FNDE) relies heavily on historical data from the annual School Census to forecast student enrollments. However, accurately predicting these numbers across a country of continental dimensions—spanning over 8.5 million km² and encompassing profound regional inequalities — remains a complex challenge. While this study integrates a broader research agenda on educational logistics, the modeling architecture, feature engineering, and empirical validation presented in this paper were developed as the core focus of the first author's undergraduate research.

Despite these structural heterogeneities, current predictive models and administrative practices largely treat schools as uniform data points within a continuous feature space [Yang *et al.*, 2020; Abideen *et al.*, 2023]. This "one-size-fits-all" approach generates a critical gap. When global feature selection strategies are applied to an unequal and topologically disjunct educational system, algorithms inherently prioritize attributes prevalent in well-resourced schools. Consequently, the sparse, yet defining, features of vulnerable institutions are overshadowed, generating a phenomenon of "algorithmic invisibility" in marginalized contexts.

This paper synthesizes and expands upon the advancements of an ongoing research project dedicated to mitigating these inequalities through Context-Aware Enrollment Forecasting. In previous stages of this research, we addressed the structural heterogeneity of Brazilian schools by modeling educational data as a bipartite directed graph. This topological approach successfully revealed the existence of two distinct latent clusters within the public network [Correia *et al.*, 2025]. Subsequent semantic profiling characterized these groups as a "Structural Elite" — defined by a dense, indivisible package of infrastructure — and a "Restricted Reality" — encompassing the vast majority of tuples (81.25%), characterized by fragmented resources and specific legal modalities (e.g., correctional facilities and socio-educational centers). In this context, a tuple refers to a unique pair of School and Teaching Stage, to be further discussed in the section 3.

Building upon these foundational discoveries, this work advances from structural diagnosis to architectural proposition and empirical exploration. We propose a Context-Aware Enrollment Forecasting Framework based on a split-modeling strategy. Furthermore, we conduct a preliminary empirical Proof of Concept (PoC) to test the immediate effects of this split. Rather than falling into confirmation bias by exclusively seeking accuracy maximization, our evaluation reveals that the primary immediate benefit of contextual clustering lies in enabling targeted dimensionality reduction and exposing the need for higher-level predictive strategies to manage the newly clustered contexts.

The main contributions of this paper are:

- **Synthesis of Structural and Semantic Disparities:** Consolidating previous findings on bipartite graph clustering and feature consistency decay to expose the structural bias of global forecasting models in unequal educational environments.

- **The Context-Aware Framework:** Proposing an architectural shift in digital government data pipelines, replacing the global "one-size-fits-all" assumption, and ensuring adaptation feasibility with a cluster-specific modeling approach.
- **Dimensionality Reduction:** Demonstrating empirically that the cluster-based feature selection was paramount for translating 20 features into a single proxy for cluster 1 and a 17.2x dimensionality reduction in cluster 2 (81.25% of the tuples), without significant performance loss, highlighting the opportunity for meta-level algorithmic selection.

The remainder of this paper is organized as follows: Section 2 discusses related work; Section 3 recaps the methodology used for structural discovery and semantic profiling; Section 4 details the dichotomy between the identified school clusters; Section 5 presents the proposed Context-Aware Framework, its empirical benchmarking, and the discussion on asymmetric dimensionality reduction versus meta-level modeling; finally, Section 6 highlights the impacts of this research and outlines future work.

2 Related Work

The challenge of student enrollment forecasting sits at the intersection of educational data mining (EDM) and public policy logistics. Following the state of practice, we classify the literature into three categories to discuss their strengths, limitations, and how they differ from our proposed framework.

The dominant approach in the literature (Group A) focuses on maximizing numerical accuracy by using global models. Studies such as Yang *et al*. [2020] and Abideen *et al*. [2023] employ Time Series and Machine Learning techniques to forecast enrollment based on historical curves. Despite these models achieving high performance metrics (e.g., low RMSE) in homogeneous environments where historical trends are stable, they adopt a "context-blind" strategy, treating all schools as uniform data points. As noted by Baker [2019] and Silva *et al*. [2025a], such "one-size-fits-all" approaches fail to capture local nuances. In the Brazilian context, this means a riverside school in the Amazon is modeled with the same feature assumptions as an urban technical institute, leading to the "algorithmic invisibility" of vulnerable contexts.

Conversely, a significant body of work (Group B) focuses on characterizing the educational context. In Brazil, Maia *et al*. [2021] and reports from INEP/MEC extensively document the "infrastructure debt" and regional inequalities. Internationally, Phakathi [2015] highlights the logistical failures in textbook delivery in South Africa due to poor context management. These studies provide deep sociological insight and accurately describe the "Restricted Reality" of the Global South. However, they are primarily descriptive, not predictive. While they diagnose the inequality, they do not translate these insights into operational algorithms for supply chain optimization. Therefore, the context remains a static statistic rather than a dynamic variable in the forecasting pipeline.

Our proposal differs from the related work by bridging these two disconnected domains. While clustering techniques are used for forecasting in other sectors (Group C) like energy [Laurinec and Lucka, 2018] or retail [Hwang *et al*., 2025], unlike Group A, we reject the global feature space, proving that high-accuracy modeling requires split strategies. Unlike Group B, we do not stop at describing inequality; we operationalize it. By using Bipartite Graph Modeling [Asratian *et al*., 1998] to detect clusters and Semantic Profiling to define them, we propose a framework that uses context to drive the prediction, ensuring that the specificities of the "Restricted Reality" are mathematically preserved in the PNLD logistics. While recent advances by previous studies [Silva *et al*., 2025a] demonstrated the superiority of Machine Learning over traditional methods for the PNLD, they operated under a global modeling assumption.

In this work, we argue that the next leap in government decision-making requires not just better algorithms, but a context-aware architecture that recognizes the structural dichotomy of the Brazilian educational system.

3 Methodology

This section synthesizes the methodological foundations and the analytical findings from our preliminary studies. By integrating graph theory, unsupervised learning, and centrality metrics, we move from the structural detection of school clusters to a semantic understanding of their distinct predictive regimes. Figure 1 illustrates the methodological workflow applied to this study.

To address the high dimensionality and heterogeneity of the Brazilian public school system, we worked with a representative sample of 2500 schools. Our initial approach modeled the data as a bipartite directed graph. Let $G = (V, A)$ be a bipartite graph where the vertex set is partitioned into two disjoint sets: V_1 representing the (School, Teaching Stage) tuples, and V_2 representing the predictive attributes (features) consumed by those tuples.

We quantified the contextual similarity between tuples using the Hamming distance metric [Hamming, 1950], which is particularly suitable for measuring dissimilarity in binary-encoded vectors (presence or absence of an attribute), over a predictive report with the features combination for each tuple, filtered by a Spearman correlation 0.6 threshold [Spearman, 1904; Silva *et al*., 2025b]. Following an exploratory analysis with Hierarchical Agglomerative Clustering (HAC), we applied the Partitioning Around Medoids (PAM) algorithm. Evaluated via the Silhouette Score, the model robustly identified an optimal partition of $K = 2$ distinct clusters within the national educational dataset [Correia *et al*., 2025].

While the clustering algorithm successfully separated the dataset topologically, understanding the drivers of this separation required a semantic analysis. We induced two subgraphs, G_1 and G_2, corresponding to the clusters, and calculated the In-Degree Centrality [Freeman, 1978] for every feature node to rank their relevance within each specific group. This approach revealed a profound socioeconomic and administrative dichotomy.

4 Results: The Structural Dichotomy

The application of the in-degree centrality ranking strategy revealed a sharp contrast in the composition of the two identified subgroups, proving that the distance between them is not merely quantitative, but semantic. To unpack this structural

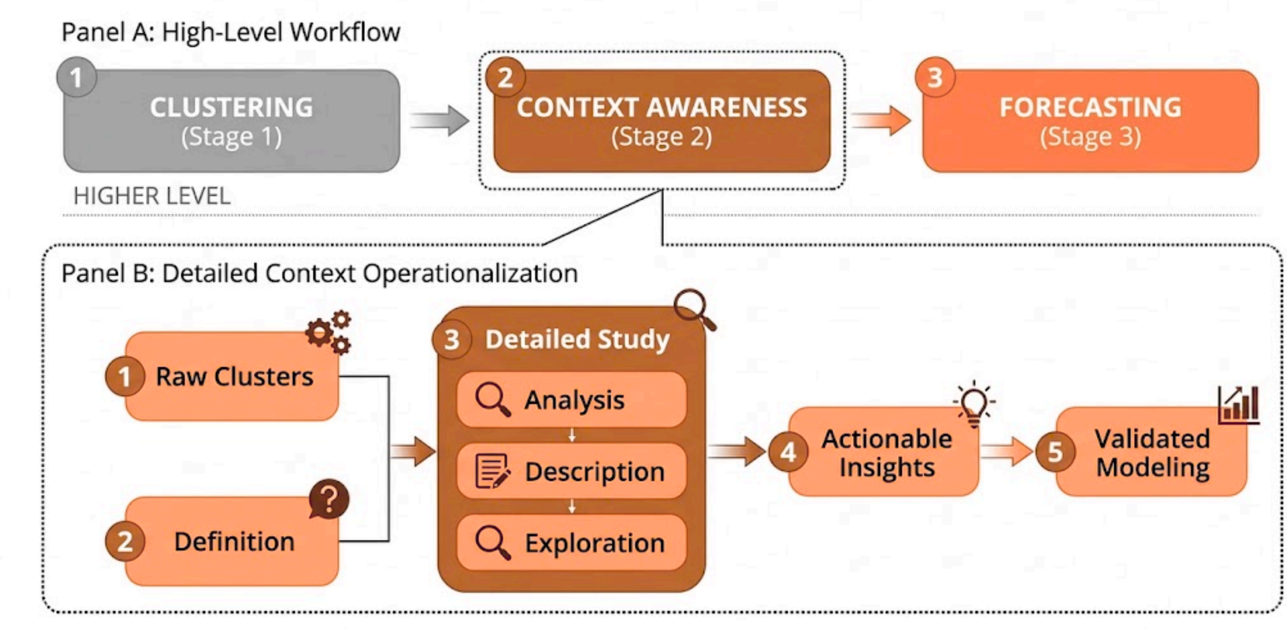

Figure 1. Methodological workflow.

dichotomy, the following sections will first detail the distinct semantic profiles of the clusters—contrasting the ubiquitous infrastructure of an "Elite" group with the fragmented reality of a "Restricted" majority. Building upon this diagnosis, we will then present a Proof of Concept (PoC) for our Context-Aware Forecasting Framework, demonstrating how a split-modeling architecture can address these unique predictive regimes.

4.1 Clusters Semantic Profiling

Comprising a smaller portion of the schools but consuming approximately 43.6% of the feature nodes, the Elite Cluster ($G1$ subgraph) represents the upper echelon of public infrastructure. The semantic analysis reveals a pattern of "ubiquitous completeness". The highest-ranked attributes include specialized amenities such as dance studios, speech therapists, and swimming pools. If a school belongs to this cluster, it likely possesses an "Indivisible Infrastructure Package", reflecting high administrative stability and data density.

In contrast, Restricted Cluster ($G2$ subgraph) encompasses the vast majority of tuples (81.25%) but operates with a significantly sparser feature set (averaging only 25 features consumed). This group is defined by its constraints rather than its assets. Top-ranked features shift dramatically to specific legal modalities, such as prison units (correctional education), socio-educational centers, and Distance Learning (EAD). This highly fragmented structure represents the vulnerable "long tail" of the Brazilian educational system.

The fundamental dichotomy is mathematically formalized through the Feature Consistency Decay analysis, as observed in Figure 2. When plotting the in-degree frequencies, Cluster 1 exhibits a flat curve, confirming high collinearity and artificial administrative homogeneity. Conversely, Cluster 2 displays a rapid Zipfian decay [Zipf, 1949], reflecting organic resource scarcity and fragmentation. Consequently, applying global feature selection to this disjunct system prioritizes the dense attributes of the elite, mathematically separating and suppressing the sparse signals of vulnerable institutions, generating algorithmic invisibility.

4.2 The Context-Aware Forecasting Framework: A Proof of Concept

Given the structural bias inherent in global modeling, we propose a Context-Aware Enrollment Forecasting Framework based on a split-modeling architecture. To validate this approach, we conducted a preliminary empirical Proof of Concept (PoC), with the parameters described as follows in Table 1. To comprehend the magnitude of the forecasting challenge, it is essential to detail the underlying complexity of the data source. The empirical validation leverages the "MEC-PROSPECCAO" dataset, which consists of microdata from the Brazilian government's school census encompassing 218,598 schools. This dataset is highly dimensional, featuring 340 binary and categorical attributes that characterize both the schools and their specific teaching stages. The structural heterogeneity addressed by our framework is deeply rooted in this data granularity, as the system registers 37 distinct types of teaching stages, spanning from K-12 education to technical and special modalities. Because each institution operates with a specific combination of these stages, the dataset is fragmented across 1,514,844 individual records, with 93.44% representing valid school-stage combinations. The vast scale and inherent sparsity of this data repository—dominated primarily by binary attributes — highlight the computational and logistical hurdles that render traditional global predictive modeling inappropriate. This dataset was provided by the Ministry of Education of Brazil (MEC) [Silva *et al.*, 2025b].

A baseline scenario (Scenario A) was established by

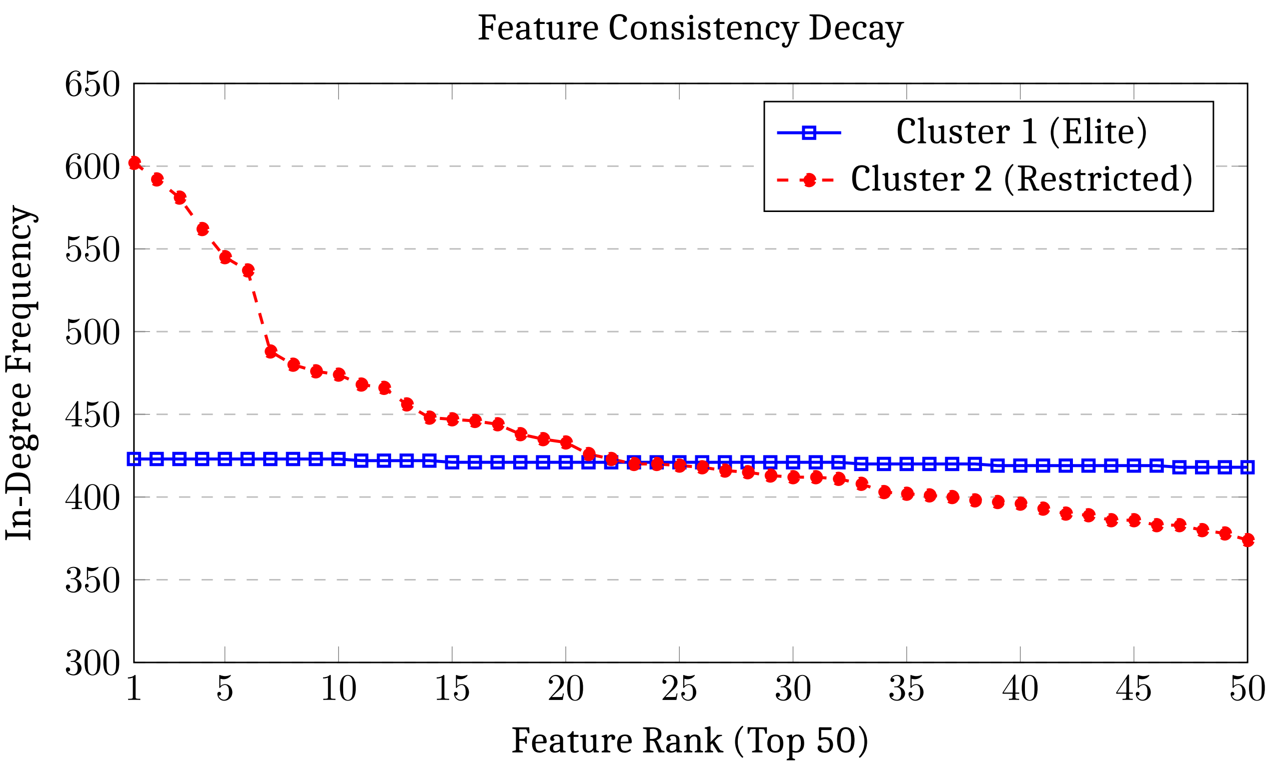

Figure 2. Feature Decay Comparison Between Clusters

Table 1. Summary of the main parameters used for model configuration and cross-validation.

Component	Parameter	Value
Random Forest Regressor	`n_estimators`	100
	`random_state`	42
	`n_jobs`	-1
Cross-Validation (K-Fold)	`n_splits`	5
	`shuffle`	`True`
	`random_state`	42

training a standard Machine Learning algorithm (Random Forest) globally across all schools using the unified feature space — 344 total variables, with 340 characteristic attributes of schools and their respective teaching stages and 4 descriptive variables. This configuration represents the conventional "one-size-fits-all" modeling paradigm commonly adopted in large-scale educational forecasting systems.

We then compared this baseline with our proposed split-modeling approach (Scenario B), in which independent models were trained for Cluster 1 and Cluster 2 using cluster-specific feature engineering strategies. The same learning algorithm was used in all scenarios to isolate the effect of context-aware feature design from algorithmic bias. The evaluation focused on predictive accuracy (MAE and RMSE) and feature space consumption, enabling the assessment of predictive efficiency.

For Cluster 1, dimensionality reduction was performed using Principal Component Analysis (PCA) [Hotelling, 1933]. The first principal component was used as a synthetic signal representing the dominant structural pattern across the original variables. This component alone explained 81.63% of the variance in the selected attributes, allowing the model to compress a large portion of the predictive signal into a single latent variable while reducing feature redundancy.

In contrast, Cluster 2 exhibited a sparse and structurally fragmented feature space. Instead of applying PCA, we adopted a ranking-based feature selection strategy, selecting the top 20 attributes according to the in-degree feature ranking within the cluster, explained in Section 3. This choice reflects the structural characteristics of these schools: unlike Cluster 1, they do not exhibit a large set of stable structural variables in Figure 2. In the global model, these observations were forced to process hundreds of features originating from other educational contexts, resulting in an unnecessarily high-dimensional feature space with limited predictive relevance. Tables 2 and 3 summarize the experimental results.

To further investigate whether the clusters represent genuinely distinct predictive regimes, we conducted a counterfactual experiment (Scenario C). In this scenario, we intentionally treated both clusters using the same reduced feature configuration, combining the PCA signal with the 20 ranked attributes from Cluster 2, resulting in a 21-feature representation. If the clusters were semantically similar, this uniform reduction should preserve predictive performance.

However, the results show a substantial degradation in accuracy, with a sharp increase in both MAE and RMSE. This experiment provides evidence that the clusters differ not only structurally but also in terms of their predictive representation requirements. In other words, dimensionality reduction strategies that are effective for one context may be detrimental when transferred to another, reinforcing the need for context-aware modeling strategies rather than uniform feature compression.

Despite the modest differences in predictive error, the

Table 2. Fairness Comparison Between Forecasting Errors

Scenario	Target	Features	Cluster 1 Errors		Cluster 2 Errors	
			MAE	RMSE	MAE	RMSE
A (Baseline)	Global	344	31.62	69.32	28.52	65.63
B (Proposed)	Split Models	324 / 20	29.44	64.17	29.01	65.66

Table 3. Model Efficiency: accuracy and feature space consumption comparison between Global and Split-Modeling approaches.

Scenario	Model Target - Approach	Features	MAE	RMSE
A (Baseline)	Global - All	344	29.13	66.72
B (Proposed)	Cluster 1 - PCA	324	29.44	64.17
B (Proposed)	Cluster 2 - Ranking	20	29.01	65.66
C (Counterproof)	Cluster 1 - PCA + Cluster 2 Ranking	21	56.92	95.18

most immediate benefit observed in the PoC was a significant gain in feature space efficiency through asymmetric dimensionality reduction. Guided by the Feature Consistency Decay analysis, we aggressively reduced the feature space for Cluster 1 without substantial loss of predictive signal, eliminating large amounts of redundant information. More importantly, the split-modeling strategy prevented the global model from forcing the majority of observations (81.25%, corresponding to Cluster 2) to process hundreds of features that are largely irrelevant to their context. This targeted reduction demonstrates the framework's potential to substantially decrease memory usage, computational latency, and energy consumption, aligning with Green AI principles while preserving the sparse but critical signals associated with marginalized educational contexts. This represents a reduction of approximately 94% in the feature space required for Cluster 2, while maintaining comparable predictive accuracy.

5 Contributions, Impact and Student Role

The application of Bipartite Directed Graphs to represent the relationship between schools and predictive attributes in a nationwide educational logistics system is an important contribution of this research. Other contributions include the identification and semantic characterization of two structural school contexts within the Brazilian public education network, termed the *Structural Elite* and the *Restricted Reality*, and the formalization of the *Feature Consistency Decay* metric, which analytically exposes how global machine learning models obscure context-specific predictive signals, a phenomenon we describe as *Algorithmic Invisibility*. Finally, we empirically demonstrated that cluster-specific feature engineering enables drastic dimensionality reduction while preserving predictive performance. In particular, Cluster 1 was compressed into a single latent PCA signal explaining 81.63% of the variance, while Cluster 2 achieved a 94% reduction in dimensionality (344 to 20 features) across 81.25% of the observations.

The main practical implication of this research is the modeling of structural heterogeneity across schools by the proposed framework contributes to more equitable forecasting strategies and resource allocation policies, potentially impacting the distribution planning for over 35 million students in the Brazilian public education system, particularly within the Brazilian National Textbook Program (PNLD), managed by the National Education Development Fund (FNDE). By avoiding high-dimensional processing, the proposed framework reduces memory usage, computational overhead, and energy consumption, aligning with Green AI principles and promoting more sustainable computational practices in government-scale machine learning systems. This result highlights the importance of context-aware modeling for building scalable and environmentally responsible AI systems in public administration.

Finally, in terms of impact for the scientific community, the theoretical architecture and structural foundations of this research were peer-reviewed, methodologically validated, and accepted in high-impact international venues, such as the 18th annual International Conference of Education, Research and Innovation (ICERI 2025, Qualis B3) [Correia *et al.*, 2025] and the 27th Annual International Conference on Digital Government Research (DG.O 2026, Qualis A2 / CORE B – in press).

Crucially, regarding the required delineation of authorship for this CTIC submission: while this study integrates a macro-project in partnership with the Ministry of Education, the undergraduate student acted as the primary driver for the technical execution. Relying on the essential project management provided by the senior co-authors, the first author, with academic supervision, designed the Context-Aware Framework's architecture, spearheaded the feature engineering process, executed the empirical baselines, and performed the dimensionality reduction analysis. This academic growth and maturity is further evidenced by the student's primary authorship in the aforementioned foundational studies.

6 Conclusion

Beyond predictive accuracy, the findings of our study highlight the importance of designing context-aware data architectures for public sector machine learning. By preventing unnecessary high-dimensional processing and preserving context-specific signals, the proposed framework contributes to more equitable and computationally efficient forecasting pipelines.

While this study utilized Random Forest as a robust proof-of-concept to validate the clustering architecture, we acknowledge that the absence of comparative experiments with other machine learning models introduces a degree of uncertainty regarding predictive optimality. Furthermore, a direct empirical comparison with existing forecasting methods from the literature was not conducted. It is important to note, however, that such benchmarking is inherently constrained by the unprecedented scale and exclusivity of the dataset provided by the FNDE. The unique heterogeneity of the Brazilian public education network, coupled with its consolidated public policy, positions this research at the global vanguard of educational logistics, lacking direct equivalent counterparts in the open literature. To address these limitations, future work will expand the experimental baseline to explore specialized predictive regimes for each cluster—including boosting frameworks (e.g., XGBoost, LightGBM)—and reproduce standard global modeling approaches from related studies to serve as internal baselines for our exclusive dataset, initiating a broader meta-learning discussion. Additionally, scaling the framework to the full national School Census dataset represents a crucial step toward operational deployment in government decision-making systems.

Declarations

Acknowledgements

This research was supported by the National Fund for Education Development (FNDE) under the Term of Decentralized Execution (TED) 12244, and we thank all the members of the Center for Excellence in Social Technologies (NEES) involved.

Funding

This research was supported by the National Fund for Education Development (FNDE) under the Term of Decentralized Execution (TED) 12244.

Authors' Contributions

DVLC contributed to the conception of this study, designed the Context-Aware Framework's architecture, led the feature engineering process, executed the empirical baselines, and performed the dimensionality reduction analysis. LCS, RAS, BJDC, RSEF, DMPJ, NC, and BAP contributed to project management, academic supervision, and methodological guidance. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The dataset analyzed during the current study was provided by the Brazilian Ministry of Education (MEC) under restricted-access terms and is therefore not publicly available; aggregated derivatives may be made available from the corresponding author upon reasonable request.

Further relevant information

References

Abideen, Z. u. *et al.* (2023). Analysis of enrollment criteria in secondary schools using machine learning and data mining approach. *Electronics*, 12(3):694. DOI: 10.3390/electronics12030694.

Asratian, A. S., Denley, T. M. J., and Häggkvist, R. (1998). *Bipartite Graphs and Their Applications*. Cambridge University Press.

Baker, R. S. (2019). Challenges for the future of educational data mining: The baker learning analytics prizes. *Journal of Educational Data Mining*, 11(1):1–17.

Correia, D. V. L. *et al.* (2025). Uncovering contextual clusters in brazilian public education. In *Proceedings of the 18th Annual International Conference of Education, Research and Innovation (ICERI 2025)*.

Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239. DOI: 10.1016/0378-8733(78)90021-7.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441.

Hwang, S., Lee, Y., Jeon, B.-K., and Oh, S. H. (2025). Sales forecasting for new products using homogeneity-based clustering and ensemble method. *Electronics*, 14(3):520. DOI: 10.3390/electronics14030520.

Laurinec, P. and Lucka, M. (2018). Clustering-based forecasting method for individual consumers electricity load using time series representations. *Open Computer Science*, 8(1):38–50. DOI: 10.1515/comp-2018-0006.

Maia, J. *et al.* (2021). Assessing the educational performance of different brazilian school cycles using data science methods. *PLOS ONE*, 16(3):1–14. DOI: 10.1371/journal.pone.0248525.

Phakathi, E. N. (2015). The management of learning and teaching support materials in public schools: a comparative study. Master's thesis, University of KwaZulu-Natal, Edgewood, South Africa.

Silva, L., de Cabral, L., Santos Júnior, J., Santos, L., Oliveira, T., Costa, B., Lobo, J., Pinho Júnior, D., Cruz, N., de Silva, R., and Pimentel, B. (2025a). Forecasting student enrollments in brazilian schools for equitable and efficient education resource allocation. *Conference on Digital Government Research*, 26. DOI: 10.59490/dgo.2025.939.

Silva, L. *et al.* (2025b). A computational approach to feature selection and enrollment forecasting in brazilian schools. *Proceedings of the 11th International Conference on Artificial Intelligence and Applications*. DOI: 10.5121/csit.2025.151910.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Yang, S. *et al.* (2020). Student enrollment and teacher statistics forecasting based on time-series analysis. *Computational Intelligence and Neuroscience*, 2020:1246920. DOI: 10.1155/2020/1246920.

Zambon, L. and Terrazzan, E. A. (2013). Policies of didactic material in brazil: Organization of procedures for selection of textbooks in basic education public schools. *Revista Brasileira de Estudos Pedagógicos*, 94(237):585–602.

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology.