XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Combining Clustering and Association Rule Mining in Educational Data: Patterns in Diagnostic Assessments

Luann F. Almeida[1], Jefferson O. Andrade[1], Karin S. Komati[1]

[1]*Graduate program in Applied Computing (PPComp)*
Instituto Federal do Espírito Santo (IFES), Campus Serra
Av. dos Sabiás, 330 - Morada de Laranjeiras, 29166-630, Serra-ES, Brazil
luannfa@gmail.com, {jefferson.andrade, kkomati}@ifes.edu.br

Abstract. This study explores the application of pattern mining and unsupervised learning techniques to analyze student performance data from large-scale educational assessments in Brazilian public schools. The dataset comprises detailed binary records of learning descriptors associated with students from the final year of secondary education. The proposed approach begins with the FP-Growth algorithm to extract initial association rules from the full dataset. These rules are then transformed into numerical vectors and clustered using HDBSCAN to identify student groups based on rule similarity. FP-Growth is subsequently reapplied within each cluster to generate localized association rules, enhancing interpretability and pedagogical relevance. Results suggest that this sequential strategy improves the contextual coherence and clarity of the discovered patterns when compared to global rule extraction. This work contributes to the field of Educational Data Mining by demonstrating how combining rule mining and clustering in multiple stages can reveal differentiated learning profiles and support more targeted educational interventions.

Keywords: Educational Data Mining; Student Performance; Unsupervised Learning; Brazilian Education.

1 Introduction

Educational data analysis has become a strategy to support evidence-based public policies and pedagogical practices [1]. The growth of databases originating from external assessments enables the identification of performance patterns and the guidance of interventions within the school environment. In this context, approaches based on machine learning and data mining have been employed, particularly due to their capacity to reveal relationships between variables that influence teaching and learning processes [2].

The ability to capture the diversity of learning profiles and the contextual nuances influencing student performance is essential. The complexity of educational data demands strategies that not only predict outcomes but also facilitate the comprehension of underlying factors in an interpretable manner, thereby supporting pedagogical interventions tailored to specific student group needs [3].

This paper investigates the combined application of clustering and rule extraction techniques on a real educational dataset, aiming to generate interpretable insights into student performance. The proposed approach integrates the FP-Growth (Frequent Pattern Growth) algorithm [4] with HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm [5]. Initially, FP-Growth extracts association rules from the full dataset, which are then transformed into numerical vectors and clustered by HDBSCAN to identify student groups based on rule similarity. Subsequently, FP-Growth is reapplied within each identified cluster to generate localized association rules, enhancing their contextual relevance and pedagogical interpretability and evaluating the effectiveness of this combined application in identifying interpretable performance patterns in diagnostic educational assessments.

The analysis focuses on data from the 2023 Diagnostic Assessment, a large-scale evaluation administered within the Espírito Santo public school system as part of the Espírito Santo Basic Education Assessment System (SICAEB) [6, 7]. This dataset specifically comprises records of learning descriptors associated with students from the final year of secondary education. The structure of the article is as follows: Section 2 presents related works. Section 3 describes the dataset and the methodological approach. Section 4 reports and discusses the results obtained. Finally, Section 5 concludes the article and outlines directions for future research.

2 Related Work

Rahman et al. [8] proposed an educational data mining (EDM) framework to support programming education, utilizing problem-solving data from the Aizu Online Judge platform. Their methodological structure involved the collection and preprocessing of submission logs and scores , followed by the vectorization of these data with derived attributes such as solution accuracy and test results. Students were grouped using the MK-means algorithm, with cluster number selection via the Elbow method. Subsequently, the FP-Growth algorithm was applied individually within each cluster to extract patterns and association rules. This approach led to the generation of interpretable patterns and recommendations for pedagogical purposes. The methodological framework employed in the present work establishes a direct dialogue with this model, particularly concerning the segmented rule extraction following the vectorial organization of the data.

This study by Rani et al. [9] addresses the challenge faced by higher education students in selecting specialization courses amidst numerous options, aiming to provide data-driven guidance for their study plans. Employing a hybrid data mining approach, the research utilizes the K-Means algorithm to cluster student grade data into groups, identifying how many students achieved good or poor grades in various subjects. Subsequently, the FP-Growth algorithm is applied to these clustered values to uncover appearance patterns, which serve as a guideline for students in determining suitable skill courses for future semesters. The study processed data from 16 subjects and 15 student attributes from Universitas Putra Indonesia YPTK Padang. Key findings indicate that a specialization theme in "business information systems" is more frequently chosen by students compared to other themes, and the identified patterns can inform students and academic supervisors during study plan compilation. The authors highlight the effectiveness of K-Means for accurate data clustering and FP-Growth for fast pattern processing.

In a recent approach, Rahman et al. [10] applied unsupervised machine learning techniques to tailor educational content to individual student weaknesses. Their work combined K-means clustering with association rule algorithms (Apriori and FP-Growth) to analyze data from students performing below a 60% threshold , aiming to generate targeted recommendations and identify problematic topics. Although the primary focus was the construction of a recommendation system in online learning environments , the use of clusters to support the visualization of weaknesses and the extraction of educational patterns aligns with this study's methodological framework. Both works share the strategy of associating profiles with frequent pattern sets, albeit in distinct contexts, reinforcing the potential of combining clustering and rule mining for constructing data-driven pedagogical interventions.

Thus, the reviewed works reinforce the viability and scientific relevance of adopting hybrid strategies based on clustering and rule extraction in educational data analysis. The present study expands this scope by applying such an approach to a real large-scale diagnostic assessment dataset, aiming to refine the understanding of performance patterns among basic education students.

3 Materials and methods

3.1 SICAEB Database

Large-scale educational assessments constitute instruments for verifying adherence to the learning objectives established by the Brazilian Common National Curriculum Base (BNCC) [11]. In this context, the Espírito Santo State Department of Education (SEDU/ES), in partnership with the Center for Public Policy and Education Assessment (CAEd), linked to the Federal University of Juiz de Fora (UFJF), implements the Espírito Santo Basic Education Assessment System (SICAEB). SICAEB comprises various instruments designed to systematically monitor the teaching and learning process within the state. Among these instruments, the Diagnostic Assessment is notable for its pedagogical and formative nature, intended for direct use by teachers and school administrators.

Diagnostic assessments are developed based on the BNCC and the Espírito Santo Curriculum [12], through reference matrices that define skill descriptors. These matrices function as technical-pedagogical instruments guiding item development, specifying the skills to be evaluated at each stage and knowledge area. Descriptors represent a didactic-evaluative transposition of general BNCC skills into a specific and measurable format suitable for standardized external assessments. For example, skill `EM13LGG105` can be broken down into descriptors addressing the identification of discursive strategies, language analysis, and interpretation of multisemiotic elements. These descriptors organize knowledge into areas such as Languages, Mathematics, Natural Sciences, and Human Sciences, enabling the objective segmentation and evaluation of specific competencies. This structure ensures curricular fidelity and facilitates the generation of systematic and comparable data, essential for diagnostic, formative, and research purposes.

The initial dataset from the 2023 Diagnostic Assessment contained 943,149 records and 69 attributes. However, to ensure pedagogical coherence and analytical focus, only students from the 3rd year of secondary education were considered for analysis. The original dataset comprises variables organized into functional groups:

Assessment identification (6 variables); Geographical and institutional data (10 variables); Class information (8 variables); Student identification and profile (6 variables); Test configuration (10 variables); Application status and control (6 variables); Overall student performance (6 variables); Data per skill descriptor (6 variables) and Data per applied item, including answer key, correct, and incorrect responses (6 variables).

3.2 Preprocessing dataset

The preprocessing pipeline of the dataset begins with (1) loading the dataset, followed by (2) an anonymization step to comply with LGPD regulations (LGPD – Law nº 13.709/2018) [13]. This procedure ensures the scientific integrity and ethical use of the information, guaranteeing participant confidentiality and compliance with legal and ethical research principles. Subsequently, (3) specific columns are selected from the anonymized data. This refined dataset is then (4) transformed into a matrix format. Finally, (5) the prepared dataset is ready for processing by the FP-Growth and HDBSCAN algorithms. Figure 1 illustrates this five-stage preprocessing pipeline for the educational dataset.

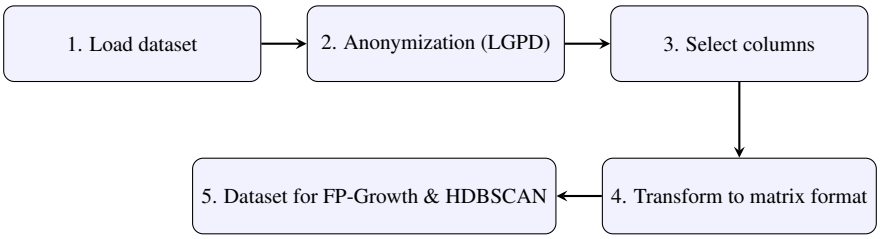


Figure 1. Preprocessing pipeline of the educational dataset.

Among the 69 available columns, only three were used to compose the final dataset: the student identifier column (`CD_ALUNO`), the descriptor codes column (`NM_HABILIDADE`), and the column with the respective correct or incorrect responses (`DC_HABILIDADE_ACERTO`). These variables were structured in a concatenated format, with each student having up to five distinct records, one for each test booklet administered. A treatment process was performed on these columns, which unfolded the concatenated values into multiple rows, so that each row represented an individual association between student and descriptor.

Next, the data were unified per student, consolidating all solved descriptors into a single record per student. This transformation resulted in a binary matrix, where each column represents a descriptor and each cell indicates a correct (1) or incorrect (0) response. The dataset structure follows a tabular format where each row represents a student, and each column corresponds to a skill descriptor. Descriptors are identified by codes such as `D060_P` and `D117_G`, where the numeric prefix is accompanied by a suffix indicating the area of knowledge: P (Languages), M (Mathematics), CN (Natural Sciences), H (History), G (Geography), and B (Biology). Each variable assumes a binary value: 1 for correct and 0 for incorrect.

After the filtering, anonymization, and restructuring procedures, the final dataset used for analysis comprised 24,703 records and 89 attributes. These records specifically represent learning descriptors linked to students in the final year of secondary education.

3.3 Methods

The proposed approach was conducted in three sequential stages: initial extraction of association rules via FP-Growth, vectorial clustering of rules via HDBSCAN, and reapplication of the FP-Growth algorithm for identifying group-specific patterns.

The FP-Growth algorithm is an efficient method for generating frequent item sets without the need for candidate set generation [4]. This technique employs a “divide-and-conquer” strategy and requires only two scans of the database to identify frequent patterns. A key advantage of FP-Growth is its use of a compact data structure, known as an FP-Tree, which stores critical and compressed information on frequent patterns. This tree-based structure allows for faster database analysis compared to other association mining techniques. By eliminating the costly and tedious candidate generation process inherent in Apriori-like algorithms, FP-Growth offers reduced memory utilization and execution time, making it suitable for large and medium-sized datasets.

To filter the rules extracted by FP-Growth, three parameters were considered: support, which indicates the proportion of records in which the antecedent-consequent set occurs; confidence, which represents the conditional probability of the consequent occurring given the antecedent; and lift, which quantifies the relevance of the asso-

ciation by comparing the observed confidence with that expected under statistical independence. Minimum values adopted for these parameters are: support $\geq 0{,}1$, confidence $\geq 0{,}7$ and lift $> 1{,}2$ [14].

The HDBSCAN is a density-based clustering algorithm that enhances DBSCAN by establishing a hierarchical representation of the clusters [5]. This algorithm does not require every data point to be assigned to a cluster, as it identifies dense clusters and considers unassigned points as outliers or noise. While the original DBSCAN may encounter issues when cluster density varies within a dataset , HDBSCAN overcomes this limitation by identifying clusters of any density. It also offers the advantage of effectively finding distinct groups and detecting outliers. For HDBSCAN, a minimum cluster size parameter of 10 was adopted. Rules not allocated to any group were labeled as noise (label -1).

Finally, the FP-Growth algorithm was reapplied independently within each cluster identified by HDBSCAN, using the same support, confidence, and lift parameters. This step aimed to identify more specific and cohesive rules within each grouping, allowing for the analysis of internal patterns with greater granularity and, consequently, contributing to a refined pedagogical interpretation.

4 Results and Discussion

The sequential application of FP-Growth, HDBSCAN, and FP-Growth identified association patterns among skill descriptors, segmenting the rules into four clusters with distinct characteristics. Table 1 presents the main association rules extracted per cluster, ordered in descending manner by their lift values. Each rule expresses an (*antecedente* $\rightarrow$ *consequente*) relationship, accompanied by metrics indicating its statistical relevance: confidence, lift, and support.

Table 1. Top-3 association rules per cluster with highest *lift* values

Cluster	Rule (antecedent $\rightarrow$ consequent)	Confidence	Lift	Support
0	D061_CN, D068_G $\rightarrow$ D185_H	0.832	1.52	0.141
	D077_F $\rightarrow$ D068_G	0.812	1.49	0.137
	D117_G, D061_CN $\rightarrow$ D060_CN	0.798	1.48	0.134
1	D185_H, D117_G $\rightarrow$ D077_F	0.846	1.62	0.163
	D060_CN, D117_G $\rightarrow$ D068_G	0.824	1.59	0.151
	D077_F $\rightarrow$ D060_CN	0.803	1.54	0.148
2	D077_F, D061_CN, D117_G $\rightarrow$ D060_CN	0.934	1.50	0.163
	D117_G, D068_G $\rightarrow$ D185_H	0.897	2.10	0.181
	D061_CN, D060_CN $\rightarrow$ D077_F	0.921	1.72	0.176
3	D117_G $\rightarrow$ D068_G	0.845	1.58	0.149
	D068_G, D077_F $\rightarrow$ D117_G	0.816	1.54	0.140
	D185_H, D060_CN $\rightarrow$ D077_F	0.801	1.50	0.132

The association rules extracted and presented in Table 1 encompass the following descriptors:

- **D060_CN**: Identify different energy sources;
- **D061_CN**: Identify reversible and irreversible transformations of matter;
- **D068_G**: Recognize the main natural landscapes of Brazil;
- **D077_F**: Correspond a system of first-degree polynomial equations with two variables to a graphical representation;
- **D117_G**: Recognize the importance of cartography for the organization of geographical space;
- **D185_H**: Understand relevant historical processes.

The characteristics and patterns identified within each of the four clusters generated by the grouping process are presented below. A detailed analysis of each cluster allows for a deeper understanding of skill associations and their underlying pedagogical profiles.

1. Cluster 0: the rules revealed associations between History and Natural Sciences topics, highlighting descriptors D061_CN, D068_G, and D185_H. Some rules included Mathematics elements (D077_F), indicating specific interdisciplinary application. This cluster's pedagogical profile aligns with sequential and disciplinary approaches.
2. Cluster 1: Presents rules connecting History, Geography, Natural Sciences, and Mathematics. Rules with higher lift indicate that performance in topics such as cartography (D117_G) relates to proficiency in History (D185_H), Chemistry (D060_CN), and Algebra (D077_F) content. The student profile suggests an ability to integrate interdisciplinary knowledge, which supports contextualized teaching strategies.
3. Cluster 2: is the highest number of rules and highest average metrics, this cluster is characterized by dense associations among content from multiple disciplines: History, Geography, Biology, Natural Sciences, and Mathematics. High confidence and lift values indicate deterministic patterns. This group requires attention to prerequisites, as difficulties in one skill impact other areas.
4. Cluster 3: exhibits patterns similar to Cluster 1, with associations between Human and Exact Sciences. The recurrence of high-confidence rules involving descriptor D117_G (cartography) as a consequent reinforces its central role. Students in this cluster may benefit from integrated and interdisciplinary pedagogical practices.

From an Educational Data Mining (EDM) standpoint, the lift metric quantifies the degree to which the joint occurrence of antecedent and consequent exceeds the frequency expected under statistical independence. A lift value greater than 1 indicates that the presence of the antecedent increases the likelihood of the consequent beyond what would be expected by chance, thereby characterizing a dependency structure. Association rules with high lift values capture non-trivial dependencies among learning descriptors, functioning as reliable indicators of conditional performance patterns. These rules provide a basis for designing evidence-driven pedagogical interventions aimed at addressing prerequisite relationships and supporting personalized learning trajectories.

All clusters presented relevant association rules with high confidence and lift values, supporting the robustness of the identified patterns. The observed differences among clusters reflect distinct learning profiles and organizational structures:

- **Cluster 0**: predominantly intra-disciplinary structure, with sequential content progression.
- **Clusters 1 and 3**: integration across subject areas, enabling interdisciplinary instructional strategies.
- **Cluster 2**: explicit interdependence among multiple disciplines, requiring prerequisite-driven interventions.

4.1 Cluster 2

Cluster 2 is the most expressive among the identified groupings. It exhibits the highest lift (1.7584) and confidence (0.8040) metrics. This performance indicates that the association patterns in this group exhibit greater determinism compared to the other clusters, revealing structural interdependencies among the assessed skills. Figure 2 presents the distribution of lift values within each cluster, while Figure 3 shows the dispersion of rule confidence.

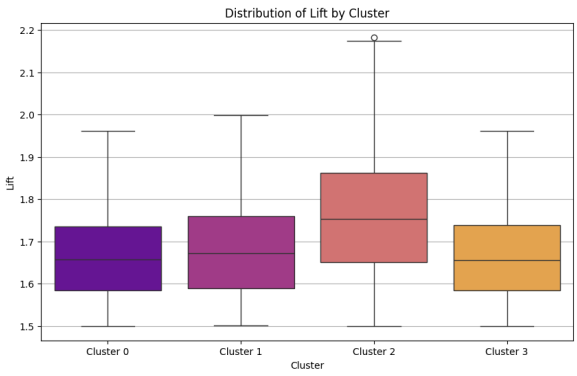


Figure 2. Boxplot of lift values per cluster.

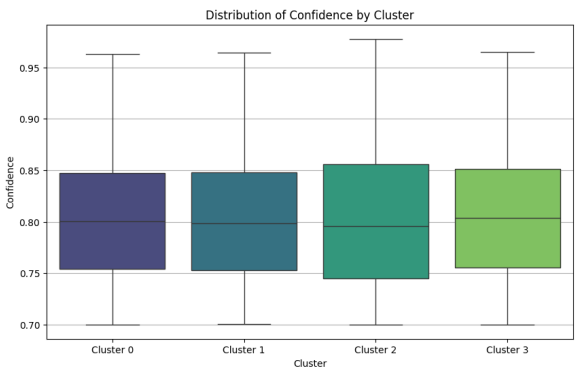


Figure 3. Boxplot of confidence values per cluster.

One of the most relevant association rules identified in Cluster 2 links three prerequisite skills to a target learning outcome. The antecedent includes the ability to represent problem situations using systems of first-degree polynomial equations (D077_F), to distinguish between reversible and irreversible physical and chemical changes (D061_CN), and to understand the role of cartography in representing geographic space (D117_G). The consequent refers to the ability to recognize and categorize different energy sources (D060_CN). This rule is characterized by the following statistical metrics: confidence of 0.934 and lift of 1.50. These values indicate a strong and non-

trivial association, suggesting that students who master the three antecedent skills have a 93.4% probability of also mastering the target skill. This pattern exemplifies a robust instance of interdisciplinary predictive modeling.

From a pedagogical perspective, this rule highlights a sequence of interrelated competencies that span Mathematics, Natural Sciences, and Geography. Its practical relevance lies in supporting diagnostic assessment and guiding intervention strategies: when a student exhibits difficulties with the concept of energy sources (D060_CN), their performance in algebraic modeling (D077_F), matter transformation classification (D061_CN), and cartographic reasoning (D117_G) should be reviewed to identify potential prerequisite gaps and plan appropriate remediation.

4.2 Global x Cluster

To illustrate the benefits of applying the FP-Growth algorithm after segmenting the dataset with HDBSCAN, Table 2 compares the main association rules extracted globally (i.e., over the entire dataset) with those identified within the resulting clusters.

Table 2. Comparison between main rules of global FP-Growth application and HDBSCAN Clusters.

Type	**Rule (antecedent → consequent)**	**Confidence**	**Lift**	**Support**
Global	D110_H → D061_CN	0.835	1.27	0.555
Global	D077_F → D061_CN	0.872	1.33	0.576
Global	D061_CN → D110_H	0.846	1.27	0.555
Cluster	{D185_H, D061_CN} → {D077_F, D110_H}	0.850	1.53	0.432
Cluster	{D110_H, D061_CN} → {D185_H, D077_F}	0.772	1.53	0.432
Cluster	{D060_CN, D185_H} → {D110_H, D061_CN}	0.895	1.60	0.434

D110_H: Analyze the impacts of the Industrial Revolution on the contemporary world.

Although the confidence and lift values are comparable across both approaches, the rules extracted within the clusters exhibit a more complex structure. Unlike the global rules, which primarily involve single-item antecedents and consequents, the cluster-based rules frequently involve itemsets—combinations of multiple descriptors—on both sides of the association. This shift indicates that local patterns capture multi-skill relationships that are not evident in the global analysis.

The use of HDBSCAN prior to pattern mining enables the identification of thematic groupings of students based on performance similarity. When FP-Growth is applied within these clusters, it reveals associations that reflect the internal logic of each group, leading to more specific and pedagogically informative rules. For example, rules with antecedents combining descriptors from different subject areas may highlight interdisciplinary learning paths or prerequisite chains that are typical of a given subgroup.

These results suggest that the main advantage of the combined HDBSCAN and FP-Growth approach lies not in the improvement of individual metrics, but in the increased expressiveness and contextual relevance of the discovered patterns. Such an approach is particularly useful in educational data mining scenarios that require the interpretation of diverse learning profiles and the design of targeted instructional strategies.

5 Conclusions

This study demonstrated the feasibility of combining FP-Growth and HDBSCAN for analyzing educational data from the 2023 Diagnostic Assessment in the public school system of Espírito Santo. The proposed approach enabled the extraction of meaningful association rules between skill descriptors and their segmentation into cohesive clusters with distinct learning profiles, supporting a more contextualized understanding of student performance.

The high-lift rules uncovered in each cluster indicate non-trivial relationships among skills and offer practical applications such as grouping students with similar performance profiles, informing the development of personalized instructional materials, anticipating future difficulties based on early skill mastery, and prioritizing pedagogical interventions using metrics like confidence, lift, and support.

While confidence and lift values were similar across global and clustered analyses, the rules extracted within clusters were more complex, involving multi-descriptor antecedents and consequents. This structure revealed latent patterns—particularly in Cluster 2, where strong interdependencies suggested prerequisite chains requiring

targeted pedagogical attention.

The use of clustering improved the specificity and interpretability of the results by avoiding statistical dilution and aligning patterns with learner subgroups. From an Educational Data Mining (EDM) perspective, this hybrid strategy proved effective in uncovering hidden structures and informing data-driven educational planning.

As future work, we propose expanding the dataset to include multiple assessment cycles and different grade levels, maintaining the focus on performance at the skill level. Additionally, we suggest integrating the extracted association rules into supervised machine learning models to explore their predictive potential for estimating future assessment outcomes. This hybrid approach could contribute to the development of intelligent systems capable of anticipating difficulties, personalizing learning trajectories, and supporting educational decision-making based on demonstrated student knowledge.

Acknowledgements. The authors would like to thank FAPES/UnAC (No. FAPES 1228/2022 P 2022-CD0RQ, No. SIAFEM 2022-CD0RQ) for the financial support provided through the Sistema UniversidaES; FAPES for project nº 1023/2022 P:2022-8TZV6 and CNPq for project nº407742/2022-0. Professor Komati thanks CNPq for the DT-2 grant (nº 302726/2023-3).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain. Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, vol. 28, n. 1, pp. 905–971, 2023.

[2] W. Chango, J. A. Lara, R. Cerezo, and C. Romero. A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, n. 4, pp. e1458, 2022.

[3] W.-C. Choi, C.-T. Lam, and A. J. Mendes. A systematic literature review on performance prediction in learning programming using educational data mining. In *2023 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. IEEE, 2023.

[4] M. Patil and T. Patil. Apriori algorithm against fp growth algorithm: a comparative study of data mining algorithms. In *Proceedings of the 7th International Conference on Innovations and Research in Technology and Engineering (ICIRTE-2022), organized by VPPCOE & VA, Mumbai-22, INDIA*, 2022.

[5] G. Stewart and M. Al-Khassaweneh. An implementation of the hdbscan* clustering algorithm. *Applied Sciences*, vol. 12, n. 5, pp. 2405, 2022.

[6] ES. Espírito Santo. Secretaria de Estado da Educação. Avaliação diagnóstica. `https://avaliacaoemonitoramentoespiritosanto.caeddigital.net/`. Acesso em: 23 de maio de 2025, 2023.

[7] CAED. Centro de Políticas Públicas e Avaliação da Educação. Avaliações educacionais. `https://www.caed.ufjf.br`. Acesso em: 23 de maio de 2025, 2023.

[8] M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran, and K. Nakamura. Educational data mining to support programming learning using problem-solving data. *Electronics*, vol. 10, n. 24, pp. 3091, 2021.

[9] M. Muhammad. Determination of student subjects in higher education using hybrid data mining method with the k-means algorithm and fp growth. *International Journal of Artificial Intelligence Research*, 2021.

[10] S. I. Rahman, S. Ahmed, T. A. Fariha, A. Mohammad, M. N. M. Haque, S. Chellappan, and J. Noor. Unsupervised machine learning approach for tailoring educational content to individual student weaknesses. *High-Confidence Computing*, vol. 4, n. 4, pp. 100228, 2024.

[11] BRASIL. Ministério da Educação. Base nacional comum curricular, 2018.

[12] ES. Espírito Santo. Secretaria de Estado da Educação. Currículo do espírito santo – etapa ensino médio. `https://curriculo.sedu.es.gov.br`. Acesso em: 23 de maio de 2025, 2021.

[13] BRASIL. Lei geral de proteção de dados pessoais – lgpd. lei nº 13.709, de 14 de agosto de 2018. `https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`. Acesso em: 4 jun. 2025, 2018.

[14] B. Mudumba and M. F. Kabir. Mine-first association rule mining: An integration of independent frequent patterns in distributed environments. *Decision Analytics Journal*, vol. 10, pp. 100434, 2024.