REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

Volúmen 1, Número 4 (2024) Pág. 1-16

Optimization of Inclusive Educational Programs through Statistical Modeling and Data Mining

Optimización de Programas Educativos Inclusivos mediante Modelación Estadística y Minería de Datos

Mayuri Viviana Pico Gordón
https://orcid.org/0009-0006-7838-6232
mayuri_pico@hotmail.com
Escuela superior Politecnica de Chimborazo
Ecuador

Ecuador Abstract.- This article presents an analysis of the optimization of inclusive education programs using advanced statistical modeling and data mining techniques. The objective is to improve educational equity by identifying factors that influence academic performance and evaluating the effectiveness of educational interventions. Models such as Random Forest and Gradient Boosting are used to predict educational outcomes, showing moderate performance, with Gradient Boosting slightly superior. Key factors identified include prior academic performance, available resources, and the absence of special educational needs (SEN). The importance of integrating advanced statistical and analytical methods with ethical and contextual considerations to ensure inclusive and sustainable education policies is highlighted. The study concludes that these approaches allow for a better understanding of the impact of educational variables and support informed decision-making.

Keywords: *education, data, mining, models, optimization.*

1. Introducción

Optimizing the evaluation of inclusive education programs using advanced statistical modeling and data mining techniques is a critical field of study that aims to improve educational equity for students of diverse abilities and backgrounds [1]. Inclusive education strives to ensure that all students receive equitable learning opportunities, but often faces challenges such as a lack of resources, a shortage of specialized training for educators, and diverse student needs.

These challenges require robust evaluation methods to assess the effectiveness of inclusive practices and inform educational policy and practice. Recent advances in statistical modeling and data mining provide powerful tools for analyzing large and complex educational datasets [2]. By employing methodologies such as multilevel modeling, Bayesian approaches, and machine learning techniques, researchers and policymakers can gain valuable insights into student achievement, identify effective teaching strategies, and predict outcomes for diverse student populations. These methods enable a nuanced understanding of the factors that influence educational success, ultimately leading to informed decision-making that promotes inclusion and equity in education. However, the integration of these advanced techniques also raises significant ethical considerations, including data privacy and potential biases in algorithmic decision-making [3].

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

The reliance on proprietary datasets and the digital divide can exacerbate existing inequalities, highlighting the importance of ethical practices and equitable access in educational assessment [4]. Furthermore, the effectiveness of these models can vary across different educational contexts, requiring ongoing refinement and validation to ensure their applicability in diverse settings.

Therefore, optimizing the evaluation of inclusive education programs through advanced statistical modeling and data mining techniques is crucial for improving educational outcomes [5]. As the field evolves, it is essential to balance the use of sophisticated analytical tools with ethical considerations and a commitment to inclusivity, ensuring that all students benefit from advances in educational assessment and policy.

Inclusive education aims to provide equitable learning opportunities for all students, regardless of their abilities or disabilities. However, this educational approach faces several challenges that can hinder its effectiveness. Key issues include a lack of specialized training for teachers, insufficient resources, and the need for adequate classroom support systems to accommodate diverse learning needs [6].

Understanding these challenges is essential to developing effective solutions and improving the overall inclusive education environment. The factors that influence inclusive education go beyond classroom dynamics; they also encompass student demographics, curricula, and teaching quality. Research indicates that background data such as parental education, family income, and household records can significantly improve predictive models of educational achievement [7].

These findings suggest that a comprehensive understanding of the educational landscape is crucial for addressing the various dimensions of inclusive education. In recent studies, various methodologies have been employed to analyze the challenges and efficiencies of education systems. Techniques such as statistical modeling and data mining have emerged as valuable tools for extracting insights from large data sets. These methodologies not only help identify factors that contribute to successful inclusive education but also assist in the assessment of instructional efficiency and educational outcomes [8].

By leveraging advanced data analysis techniques, educators and policymakers can gain a clearer understanding of the effectiveness of inclusive education programs and make informed decisions that promote equity in education.

Therefore, addressing the multifaceted problems associated with inclusive education is crucial for the continued development and optimization of educational practices and systems. Therefore, addressing the multifaceted problems associated with inclusive education is crucial for the continued development and optimization of educational practices and systems [9].

Key Concepts

Statistical Modeling in Education

Statistical modeling plays a critical role in understanding and evaluating educational programs, particularly in the context of inclusive education. It refers to the process of using mathematical frameworks to identify relationships within data sets relevant to educational settings, such as student achievement metrics and demographic information [10].

Statistical models are constructed to provide insight into behavioral patterns among students, enabling educators and administrators to make data-driven decisions [11].

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

These models often incorporate both random and non-random variables, allowing for a nuanced interpretation of educational outcomes. They can also highlight variations in student performance, which is essential for tailoring educational strategies to the needs of diverse learners [12].

Data Mining Techniques in Educational Evaluation

Educational Data Mining (EDM) is a growing field that focuses on developing methods to extract meaningful information from large educational data sets [13]. Various techniques such as classification, clustering, and regression are used to identify hidden relationships and trends between students and educational processes.

By leveraging these data mining techniques, educators can improve their understanding of student needs and increase program effectiveness. Furthermore, EDM facilitates predictive analytics, which can predict student outcomes and identify at-risk students, enabling proactive interventions.

Integration of Statistical Methods and Data Mining

The integration of advanced statistical models and data mining techniques offers powerful tools for evaluating inclusive education programs. For example, the use of multilevel models can help evaluate the effectiveness of interventions across different student populations, taking into account variations in individual learning experiences [14].

This comprehensive approach not only helps identify effective practices but also supports the ongoing refinement of educational strategies in response to new insights derived from data [15].

Assessment Frameworks

Psychometric Models in Assessment

In the context of evaluating inclusive education programs, psychometric models play a crucial role in enabling the measurement of student learning and program effectiveness. Standard psychometric models, such as Item Response Modeling (IRM), provide a framework for comparing test takers across different tests and estimating item parameters based on diverse groups of students [16].

These models facilitate the analysis of how well educational assessments reflect students' knowledge and skills, thereby improving the evaluation of instructional effectiveness [17].

Assessment Triangle

The assessment process is guided by the "assessment triangle," which consists of three interconnected elements: cognition, observation, and interpretation. Cognition refers to the constructs or learning objectives that assessments seek to measure. Observation involves collecting data through assessment tasks, such as the grades assigned by teachers on various assignments [18].

Finally, interpretation refers to the statistical methods used to analyze the collected data, often employing measurement models to draw meaningful inferences about student learning and program effectiveness. This triadic relationship ensures that assessments are well coordinated and produce valid conclusions about student achievement [19].

Bayesian Approaches

Recent advances in statistical methodology, particularly Bayesian modeling, have begun to influence educational assessment practices. These approaches allow for more nuanced

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

interpretations of student data by incorporating diagnostic indices into measurement models. For example, the use of multidimensional Item Response Models allows educators to go beyond individual summary statistics, thus providing a richer view of student achievement [20].

Bayesian methods also facilitate the integration of diverse sources of evidence, improving the overall robustness of assessments in inclusive educational contexts [21].

Integration of Educational Data Mining

Educational Data Mining (EDM) techniques are increasingly being used to effectively evaluate instructional programs. EDM provides stakeholders—educators, students, organizations, and researchers—with tools to analyze large data sets, leading to improved teaching methods, individualized learning experiences, and optimized resource allocation within educational institutions [22].

By using advanced statistical modeling and data mining techniques, evaluators can gain useful insights from complex educational data, thereby improving the effectiveness of inclusive education programs [23].

Advanced Statistical Modeling Techniques

Advanced statistical modeling techniques play a crucial role in optimizing the evaluation of inclusive education programs. These methods allow researchers to analyze complex data structures and relationships, providing a deeper understanding of educational outcomes [24].

Multilevel Modeling

Multilevel modeling (MLM) is an advanced statistical technique that allows researchers to partition variance across different levels of analysis [25], such as students nested within schools. This flexibility helps accommodate the hierarchical nature of educational data, improving the accuracy of findings. MLM is particularly useful for assessing the impact of various factors on student achievement while controlling for contextual influences, such as socioeconomic status (SES).

Bayesian Methods

Bayesian methods have emerged as a powerful approach in statistical modeling, particularly for educational assessments. These methods allow for the incorporation of prior knowledge and the continuous updating of estimates as new data become available. This adaptability is beneficial for managing the uncertainties inherent in educational data, such as variations in student achievement across contexts [26].

Bayesian inference networks, for example, allow researchers to model complex interrelationships between variables, improving the interpretability of data related to educational outcomes [27].

Item Response Theory

Item Response Theory (IRT) offers a robust framework for understanding how different test items perform in diverse populations. It allows for the comparison of test-taker performance on different assessments by predicting item properties based on test-taker characteristics [28].

This method is particularly valuable in inclusive educational settings, as it can reveal how diverse learners interact with assessment tools, facilitating more equitable assessment practices [29].

Machine Learning Techniques

Machine learning techniques, including supervised and unsupervised learning, provide innovative tools for analyzing educational data. Supervised learning algorithms, such as

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

regression and decision trees, predict outcomes based on historical data, while unsupervised methods, such as clustering, identify patterns within data sets without pre-labeled outcomes [12].

The use of these algorithms in policy analysis allows for more nuanced evaluations of educational programs, enabling policymakers to tailor interventions based on identified trends and relationships [30].

Quantile Regression

Quantile regression is another advanced technique that assesses how the relationship between variables differs at various points in the outcome distribution, rather than focusing solely on the mean [31].

This approach is particularly relevant in educational research, as it helps identify how different levels of school funding relate to achievement among various demographic groups of students, including those from disadvantaged backgrounds. This approach is particularly relevant in educational research, as it helps identify how different levels of school funding relate to achievement among diverse student demographics, including those from disadvantaged backgrounds [32].

Data Mining Techniques in Education

Educational data mining (EDM) is an emerging field that focuses on developing methods for analyzing large volumes of data from educational settings to extract meaningful insights and improve decision-making processes [33].

Several advanced data mining techniques are employed in education, including classification, clustering, regression, and association rule mining, each with distinct purposes for understanding student behavior and academic performance [34].

Classification Techniques

Classification is a supervised learning method used to categorize data into predefined classes. In educational contexts, classification algorithms help predict student outcomes based on historical data. Common algorithms include decision trees, support vector machines (SVMs), and neural networks. Decision trees, in particular, are popular due to their interpretability and ability to provide "if-then" rules that can guide educators in making informed decisions about student interventions and resource allocation [7].

For example, researchers have used decision trees to model student achievement, effectively predicting academic success based on various predictors, such as attendance and prior grades [35].

Predictive Modeling

Predictive modeling integrates various data mining techniques to predict future student performance. By using algorithms such as k-nearest neighbors (kNN) and regression analysis, educators can build models that anticipate academic challenges and opportunities for intervention. These predictive models not only help identify at-risk students but also allow educational institutions to allocate resources more effectively and improve overall academic performance [36].

Clustering Methods

Clustering is another fundamental technique employed in EDM that groups students based on shared characteristics. By identifying clusters, educators can discern different learning styles and behavioral patterns among students, enabling the development of personalized

ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

educational frameworks. This method can be instrumental in improving community education and fostering adaptive learning environments [37].

Clustering techniques allow instructors to categorize students, facilitating targeted interventions and personalized learning experiences [38].

Regression Analysis

Regression analysis is often used to model the relationships between various educational variables and student achievement [39]. It allows educators to predict continuous outcomes, such as test scores or graduation rates, based on a combination of independent variables. This technique is beneficial for understanding the impact of different factors on student success and for making data-driven decisions regarding curricular adjustments and resource allocation.

Association Rule Mining

Association rule mining is a critical component of EDM, focused on uncovering interesting relationships between variables within large datasets [40].

This technique helps identify correlations between student behaviors, such as online activity, and academic outcomes, allowing educators to gain insights into effective teaching strategies and student engagement. For example, association rules can reveal patterns in students' interactions with online resources and how these behaviors correlate with their final grades, providing useful information for improving educational outcomes [41].

2. Materials and Methods

2.1 Statistical Models

Exploratory Analysis

Exploratory analysis was conducted using advanced data mining techniques to identify patterns and relationships in the data. Clustering algorithms, such as K-means and DBSCAN, were applied to segment students into groups based on their academic performance and emotional well-being.

In addition, the Apriori algorithm was used for association rule analysis, which allowed for the detection of significant relationships between variables such as academic performance, available resources, and teacher training. To simplify the data structure and preserve maximum variability, principal component analysis (PCA) was employed.

Hierarchical Multilevel Models

HLMs were implemented to analyze variability between students and schools, considering the nested structure of the data (students within schools). These models incorporated random effects to capture differences between school contexts and assessed the impact of school factors and educational programs on students' academic and socioemotional outcomes.

Machine Learning

Two machine learning approaches were used to model and predict educational outcomes:

Random Forest: This ensemble model was used to predict academic performance and emotional well-being, identifying the most relevant features by assessing variable importance.

Gradient Boosting Machine: This boosting model optimized predictions through an iterative process that minimized prediction error, highlighting critical factors through feature importance analysis.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

Both models were evaluated using performance metrics such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE).

Validation and Metrics

To ensure the robustness of the models, the data were divided into training (80%) and test (20%) sets. In addition, cross-validation was applied to assess the generalization of the models. Evaluation metrics included (R^2), MAE, and RMSE, which allowed for an accurate measurement of the predictive performance and explanatory power of the models.

Integration of Methods

This methodological approach combines traditional statistical techniques with modern machine learning tools, offering a comprehensive perspective on the factors that influence students' educational success and socio-emotional development. The results obtained provide a solid basis for data-driven decision-making aimed at optimizing inclusive education programs.

2.2 Data Used

Description of the Simulated Database

The simulated educational database integrates multiple dimensions of the educational context, structured into four main components:

Student Characteristics: Unique student identifier, age (6-18 years), gender (M/F), socioeconomic status (Low/Medium/High), type of SEN (None/Physical/Intellectual/ASD/Sensory), prior academic performance, level of family support, and access to health services.

School Context Characteristics: Unique school identifier, institution size (Small/Medium/Large), location (Urban/Rural), teacher training level, and the institution's inclusive culture.

Educational Program Characteristics: Type of intervention (Curricular Adaptations/Personalized Support/Teacher Training), program duration (3-24 months), program intensity (1-20 hours per week), and available resources.

Educational Outcomes: Math achievement (1-10), Language achievement (1-10), Emotional Well-being (1-5), and Class Participation (1-5).

The simulated database consisted of a set of 10,000 records, incorporating realistic probability distributions based on current educational literature. Categorical variables are encoded using one-hot encoding for use in statistical and machine learning models. The hierarchical structure of the data reflects educational reality, with students nested within specific institutions.

This simulated database provided a detailed and structured representation of the educational context, allowing for the analysis of the interactions between individual, contextual, and institutional factors on educational success and socioemotional development.

3. Results

Performance of Predictive Models

The machine learning models used in this study showed a moderate level of accuracy in predicting academic outcomes. The Random Forest model achieved a coefficient of determination (R^2) of 0.13, with a mean absolute error (MAE) of 0.94 and a root mean square error (RMSE) of 1.16. The Gradient Boosting model, on the other hand, showed slightly better performance, with an R^2 of 0.16, a MAE of 0.92, and a RMSE of 1.14. This improvement suggests

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

that Gradient Boosting has a greater ability to identify complex patterns within the dataset (Figure 1).

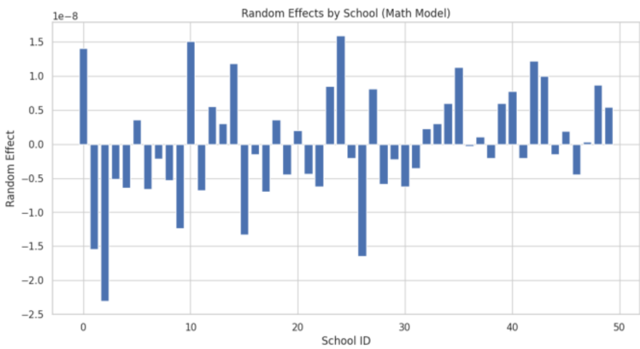

Fig 1. Random effects models.

Critical Factors Identified by Random Forest

Through the analysis of variable importance using the Random Forest model (Figure 2), three levels of influence on academic outcomes were identified.

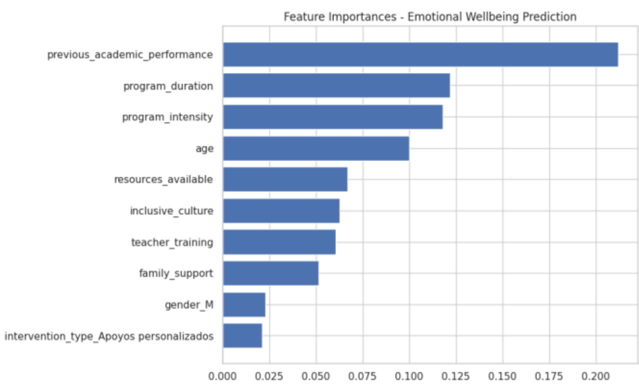

Fig 2. Analysis of the importance of variables performed with the Random Forest model.

First Level of Importance: Prior academic performance (19.3%), program duration (11.4%), program intensity (10.9%), and available resources (10.9%). Second Level of Importance: Student age (9.0%), inclusive culture (5.7%), family support (5.7%), and teacher training (5.5%), and additional factors: absence of Special Educational Needs (SEN) (5.2%), male gender (2.0%), and urban location (2.0%).

Critical Factors Identified by Gradient Boosting

The Gradient Boosting model yielded a distinct profile in terms of the relative relevance of the factors, clearly highlighting some key elements (Figure 3).

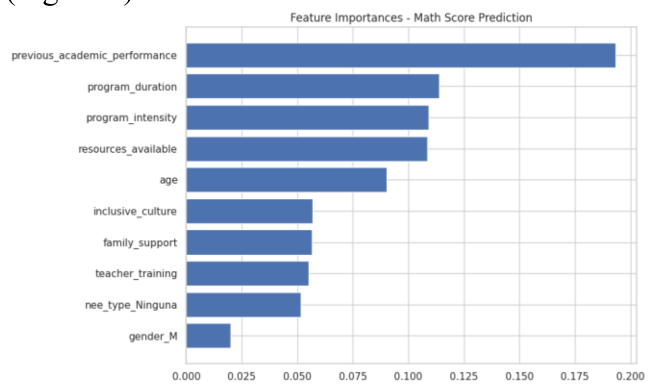

Fig 3. Analysis of the importance of variables performed with the gradient boosting model.

First Level of Importance: available resources (52.5%), absence of special educational needs (24.8%), and prior academic performance (11.7%). Second Level of Importance: program intensity (2.1%), program duration (1.6%), student age (1.3%), and family support (1.3%), as well as Less Relevant Contextual Factors: inclusive culture (0.9%), teacher training (0.8%), and school size (0.3–0.4%).

The results reflect significant differences in the way the two models weight the various factors. Gradient Boosting particularly highlights the availability of resources and the absence of special educational needs as the most influential aspects, underscoring the importance of adequate infrastructure and specialized support to maximize academic success.

On the other hand, Random Forest presents a more balanced distribution (Figure 4) of the relevant variables, placing particular emphasis on the student's prior performance and the structural characteristics of the educational program, such as its duration and intensity.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

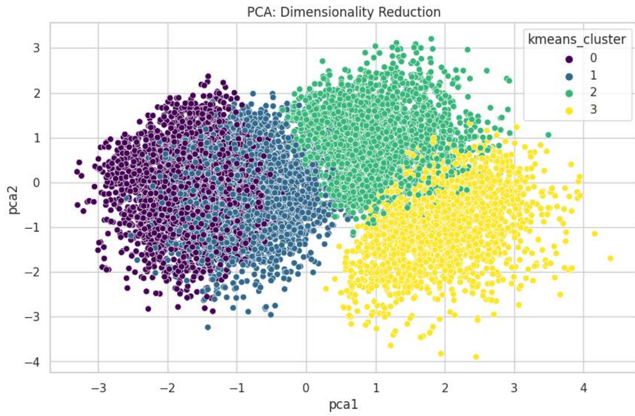

Fig 4. Divergence between models analyzed using PCA.

The divergence observed between the two models (Figure 5) suggests that there are different interactive dynamics between the variables, which warrants a deeper analysis to better understand these relationships and their real impact on school performance.

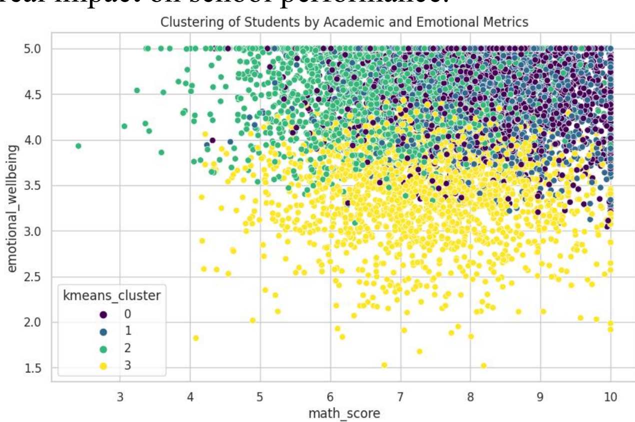

Fig 5. Divergence between models analyzed using cluster analysis.

4. Discussion

The predictive models used in this study show a moderate level of accuracy in predicting academic outcomes, which is consistent with previous findings in the field of educational data mining (EDM), where the complexity and variability inherent in educational data limits absolute predictive power [1][26]. The Random Forest model achieved a coefficient of determination (R^2) of 0.13, with a mean absolute error (MAE) of 0.94 and a root mean square error (RMSE) of 1.16. Gradient Boosting, on the other hand, showed slightly better performance, with an R^2 of 0.16, a MAE of 0.92, and a RMSE of 1.14. This improvement suggests that Gradient Boosting is better able to identify complex patterns within the dataset, possibly due to its iterative and weighted approach to error correction [42].

Variable importance analysis using Random Forest revealed three levels of influence on academic outcomes. First, prior academic performance (19.3%), program duration (11.4%), program intensity (10.9%), and available resources (10.9%) stood out. These findings are consistent with previous research that has indicated the relevance of academic history as a robust predictor of academic success [43], as well as the importance of structural factors of the educational program, such as its curricular design and course load [44].

The student's age (9.0%) was found to be in second place, followed by inclusive culture (5.7%), family support (5.7%), and teacher training (5.5%). These factors reflect contextual aspects that indirectly influence academic performance and are consistent with studies that highlight the positive impact of inclusive school environments and family support on educational development [45]. Finally, additional factors identified included the absence of special educational needs (SEN) (5.2%), followed by male gender (2.0%), and urban location (2.0%).

On the other hand, Gradient Boosting showed a distinct profile in terms of the relative importance of factors. First, it clearly highlighted the availability of resources (52.5%), followed by the absence of SEN (24.8%) and previous academic performance (11.7%). This result underscores the critical importance of having adequate infrastructure and specialized support to maximize academic success, especially in inclusive contexts where

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

barriers to learning can be multiple and heterogeneous [3][15][16]. Factors in second place of importance included program intensity (2.1%), program duration (1.6%), student age (1.3%), and family support (1.3%). Finally, among the least relevant contextual factors were inclusive culture (0.9%), teacher training (0.8%), and school size (0.3–0.4%). The divergence observed between the two models suggests that there are different interactive dynamics between the variables, which warrants further analysis to better understand these relationships and their real impact on academic performance [46].

While Random Forest offers a more balanced distribution of relevant variables, Gradient Boosting seems to focus more on those factors that have a direct and strong relationship with academic outcomes, particularly those related to institutional resources and the presence or absence of special educational needs.

These findings also reflect a recurring tension in the field of educational data mining: how to balance predictive accuracy with contextual interpretability. "Data-driven decision-making requires not only statistically sound models but also a deep understanding of the educational contexts in which they operate" [47]. This implies that, while Gradient Boosting may offer better technical metrics, Random Forest provides a broader view of factors that could be useful for designing multi-causal educational interventions.

5. Conclusions

The results of this analysis reflect the complexity of the factors that influence academic performance and highlight the usefulness of machine learning models—particularly Random Forest and Gradient Boosting—as tools for identifying predictive patterns and prioritizing key variables in educational settings. Both models showed moderate predictive performance, with Gradient Boosting showing slight superiority, suggesting its greater ability to capture nonlinear and complex relationships between variables.

From the perspective of the relative importance of factors, a clear divergence was observed between the two models. While Random Forest offered a more balanced distribution that includes cognitive, contextual, and structural aspects, Gradient Boosting predominantly highlighted the availability of institutional resources and the absence of special educational needs as the most influential determinants. This finding highlights the importance of ensuring adequate material conditions and specialized support within the framework of effective inclusive education.

The methodological difference between the two approaches underscores the importance of considering multiple analytical perspectives when designing educational policies or pedagogical interventions. While Gradient Boosting may be preferable from a technical perspective due to its greater statistical fit, Random Forest provides a more holistic view that allows for a better understanding of the different levels of influence on academic success.

References:

[1] Rao, Y. S. N., & Chen, C. J. (2024). Bibliometric insights into data mining in education research: A decade in review. *Contemporary Educational Technology*, *16*(2), ep502. http://dx.doi.org/10.30935/cedtech/14333

[2] Cerezo, R., Lara, J. A., Azevedo, R., & Romero, C. (2024). Reviewing the differences between learning analytics and educational data mining: Towards educational data science. *Computers in Human Behavior*, *154*, 108155. https://doi.org/10.1016/j.chb.2024.108155

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

[3] Yin, S., Li, H., Teng, L., Laghari, A. A., & Estrela, V. V. (2024). Attribute-based multiparty searchable encryption model for privacy protection of text data. *Multimedia Tools and Applications*, *83*(15), 45881-45902. http://dx.doi.org/10.1007/s11042-023-16818-4

[4] Arriola Navarrete, Ó. (2023). La Brecha Digital En La Revolución Industrial 4.0 Oportunidad Y Reto Para Las Bibliotecas. *Revista Interamericana de Bibliotecología*, *46*(3). http://dx.doi.org/10.17533/udea.rib.v46n3e345719

[5] Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, *28*(1), 905-971. http://dx.doi.org/10.1007/s10639-022-11152-y

[6] Mirakhur, Z., Chen, C., & Schwarz, S. (2022). A case study in teaching inclusive teaching. To Improve the Academy: A Journal of Educational Development, 41(2). https://doi.org/10.3998/tia.386

[7] yuang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. Frontiers in Psychology, 12, 698490. https://doi.org/10.3389/fpsyg.2021.698490

[8] Ampadu, Y. B. (2023). Handling big data in education: A review of educational data mining techniques for specific educational problems. AI, Computer Science and Robotics Technology, 2. https://doi.org/10.5772/acrt.17

[9] Wray, E., Sharma, U., & Subban, P. (2022). Factors influencing teacher self-efficacy for inclusive education: A systematic literature review. *Teaching and Teacher Education*, *117*, 103800. https://doi.org/10.1016/j.tate.2022.103800

[10] Gao, P., Li, J., & Liu, S. (2021). An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. *Mobile Networks and Applications*, *26*(5), 2123-2126. http://dx.doi.org/10.1007/s11036-021-01777-7

[11] Zapata-Cardona, L., & Martínez-Castro, C. A. (2023). Statistical modeling in teacher education. *Mathematical Thinking and Learning*, *25*(1), 64-78. http://dx.doi.org/10.1080/10986065.2021.1922859

[12] Shero, J. A., & Hart, S. A. (2022). Methodological decisions and their impacts on the perceived relations between school funding and educational achievement. Frontiers in Education, 7, 1043471. https://doi.org/10.3389/feduc.2022.1043471

[13] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11. https://doi.org/10.1186/s40561-022-00192-z

[14] Baek, C., & Doleck, T. (2023). Educational data mining versus learning analytics: A review of publications from 2015 to 2019. *Interactive Learning Environments*, *31*(6), 3828-3850. http://dx.doi.org/10.1080/10494820.2021.1943689

[15] Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: a systematic review. *Applied Computational Intelligence and Soft Computing*, *2022*(1), 8924028. http://dx.doi.org/10.1155/2022/8924028

[16] Chen, Y., Schönlieb, C. B., Liò, P., Leiner, T., Dragotti, P. L., Wang, G., ... & Yang, G. (2022).

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

AI-based reconstruction for fast MRI—A systematic review and meta-analysis. *Proceedings of the IEEE*, *110*(2), 224-245. http://dx.doi.org/10.1109/JPROC.2022.3141367

[17] Cao, C., Cui, Z. X., Wang, Y., Liu, S., Chen, T., Zheng, H., ... & Zhu, Y. (2024). High-frequency space diffusion model for accelerated MRI. *IEEE Transactions on Medical Imaging*, *43*(5), 1853-1865. http://dx.doi.org/10.48550/arXiv.2208.05481

[18] Yu, H. (2024). Research on the urban green transportation development level evaluation based on the triangular pythagorean fuzzy multiple attribute decision making. *Journal of Intelligent & Fuzzy Systems*, *46*(3), 6279-6297. http://dx.doi.org/10.3233/JIFS-232579

[19] Zhan, J., Deng, J., Xu, Z., & Martínez, L. (2023). A three-way decision methodology with regret theory via triangular fuzzy numbers in incomplete multiscale decision information systems. *IEEE Transactions on Fuzzy Systems*, *31*(8), 2773-2787. https://ieeexplore.ieee.org/document/10018542

[20] Jyothi, P., & Singh, A. R. (2023). Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. *Artificial intelligence review*, *56*(4), 2923-2969. http://dx.doi.org/10.1007/s10462-022-10245-x

[21] Lye, A., Cicirello, A., & Patelli, E. (2021). Sampling methods for solving Bayesian model updating problems: A tutorial. *Mechanical Systems and Signal Processing*, *159*, 107760. http://dx.doi.org/10.1016/j.ymssp.2021.107760

[22] Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence*, *35*(13), 983-1021. http://dx.doi.org/10.1080/08839514.2021.1922847

[23] Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(4), e1458. http://dx.doi.org/10.1002/widm.1458

[24] Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, *2*, 100016. https://doi.org/10.1016/j.caeai.2021.100016

[25] Lin, H. M., Wu, J. Y., Liang, J. C., Lee, Y. H., Huang, P. C., Kwok, O. M., & Tsai, C. C. (2023). A review of using multilevel modeling in e-learning research. *Computers & Education*, *198*, 104762. http://dx.doi.org/10.1016/j.compedu.2023.104762

[26] Gallardo-Gomez, D., del Pozo-Cruz, J., Noetel, M., Alvarez-Barbosa, F., Alfonso-Rosa, R. M., & del Pozo Cruz, B. (2022). Optimal dose and type of exercise to improve cognitive function in older adults: A systematic review and bayesian model-based network meta-analysis of RCTs. *Ageing research reviews*, *76*, 101591. https://doi.org/10.1016/j.arr.2022.101591

[27] Moradi, M. (2023). A Bayesian Model and Bayesian Classification on the Data Obtained from Children's Educational Activity in the IoT Environment. *Transactions on Machine Intelligence*, *6*(3), 126-136. https://doi.org/10.47176/TMI.2023.126

[28] Hermita, N., Putra, Z. H., Alim, J. A., Wijaya, T. T., Anggoro, S., & Diniya, D. (2022). Elementary Teachers' Perceptions on Genially Learning Media Using Item Response Theory

ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

(IRT). *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 1-20. http://dx.doi.org/10.23917/ijolae.v4i1.14757

[29] Spann, D. J., Cicero, D. C., Straub, K. T., Pellegrini, A. M., & Kerns, J. G. (2024). Examining measures of schizotypy for gender and racial bias using item response theory and differential item functioning. *Schizophrenia Research*, *272*, 120-127. http://dx.doi.org/10.1016/j.schres.2024.08.015

[30] Chen, W., Shen, Z., Pan, Y., Tan, K., & Wang, C. (2024). Applying machine learning algorithm to optimize personalized education recommendation system. *Journal of Theory and Practice of Engineering Science*, *4*(01), 101-108. https://doi.org/10.53469/jtpes.2024.04(01).14

[31] Peng, L. (2021). Quantile regression for survival data. *Annual review of statistics and its application*, *8*(1), 413-437. https://doi.org/10.1146/annurev-statistics-042720-020233

[32] Liu, F., Umair, M., & Gao, J. (2023). Assessing oil price volatility co-movement with stock market volatility through quantile regression approach. *Resources Policy*, *81*, 103375. http://dx.doi.org/10.1016/j.resourpol.2023.103375

[33] Roslan, M. B., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (iJET)*, *17*(5), 147-179. http://dx.doi.org/10.3991/ijet.v17i05.27685

[34] Ijaz, S., Safdar, T., & Sanaullah, M. (2020). Educational data mining: A review and analysis of student's academic performance. In I. S. Bajwa, T. Sibalija, & D. N. A. Jawawi (Eds.), Intelligent Technologies and Applications (pp. 510–523). Springer. https://doi.org/10.1007/978-981-15-5232-8_44

[35] Yu, J. (2021). Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods. *International Journal of Emerging Technologies in Learning*, *15*(5). http://dx.doi.org/10.3991/ijet.v16i05.20297

[36] Kang, S. (2021). K-nearest neighbor learning with graph neural networks. *Mathematics*, *9*(8), 830. http://dx.doi.org/10.3390/math9080830

[37] Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*, *23*(4), 485. http://dx.doi.org/10.3390/e23040485

[38] Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, *110*, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817

[39] Gupta, M., & Kaul, S. (2024). Ai in inclusive education: A systematic review of opportunities and challenges in the indian context. MIER Journal of Educational Studies Trends and Practices, 429–461. https://doi.org/10.52634/mier/2024/v14/i2/2702

[40] Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. Encyclopedia, 4(4), 1644–1664. https://doi.org/10.3390/encyclopedia4040108

[41] Hikmawati, E., Maulidevi, N. U., & Surendro, K. (2021). Minimum threshold determination method based on dataset characteristics in

ISSN: 3103-1420
DOI: https://doi.org/10.70577/chdc6n20

association rule mining. *Journal of Big Data*, *8*, 1-17. http://dx.doi.org/10.21203/rs.3.rs-728509/v1

[42] Ali, Z. A., Abduljabbar, Z. H., Tahir, H. A., Sallow, A. B., & Almufti, S. M. (2023). eXtreme gradient boosting algorithm with machine learning: A review. *Academic Journal of Nawroz University*, *12*(2), 320-334. http://dx.doi.org/10.25007/ajnu.v12n2a1612

[43] Staneviciene, E., Gudoniene, D., Punys, V., & Kukstys, A. (2024). A case study on the data mining-based prediction of students' performance for effective and sustainable e-learning. Sustainability, 16(23), 10442. https://doi.org/10.3390/su162310442

[44] Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. J. S. (2018). Data mining techniques applied in educational environments: Literature Review. Digital Education Review, 235–266. https://doi.org/10.1344/der.2018.33.235-266

[45] Joaquina De Borba, N., Adriana Santin Ferreira, V., Palhares Dos Santos, T., & Carvalho, S. (2024). Inclusive education: Strategies and impact on contemporary society. Revista Gênero e Interdisciplinaridade, 5(03), 182–191. https://doi.org/10.51249/gei.v5i03.2073

[46] Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artificial Intelligence Review*, *55*(8), 6321-6344. http://dx.doi.org/10.1007/s10462-022-10155-y

[47] Jiang, Y., Huang, Y., & Li, Q. (2024). Data-driven instructional decision-making: Framework, instrument, and process. World Journal of Educational Research, 11(3), p72. https://doi.org/10.22158/wjer.v11n3p72

Contribution of Individual Authors to the Development of a Scientific Article (Ghostwriting Policy)

All authors participated equally in the development of the article.

Sources of Funding for the Research Presented in the Scientific Article or for the Scientific Article Itself

No funding was received for this study.

Conflicts of Interest

The authors declare no conflicts of interest relevant to the content of this article.