World Journal of Information Systems 3(2) 44-58
DOI: https://doi.org/10.17013/wjis.v3i2.60

Article

Influence of Curriculum Design on the incoming student profile in informatics programs through Data Mining

Victor Dario Sosa Jauregui [1]*, Vanessa Maribel Choque Soto [1]*

[1]Department of Informatics, Universidad Nacional de San Antonio Abad del Cusco
Cusco, Perú
Email Id: victor.sosa@unsaac.edu.pe, vanessa.choque@unsaac.edu.pe

(Article history: Received December 12, 2025; Received in revised form January 11, 2026; Accepted February 17, 2026; Available online June 15, 2026)

Abstract: It is important to ensure that students meet the admission profile for a professional program, as this prepares them to face future challenges. This is particularly evident in informatics programs, which require a profile oriented toward abstraction, complex problem-solving, and other related skills. This research aims to assess whether incoming students meet the expected profile and to determine whether curriculum design influences their consolidation. To achieve this, an evaluation model was developed based on KDD (Knowledge Discovery in Databases) or Data Mining techniques, applied to student records from an informatics program. The analysis identified patterns indicating that curriculum design has an impact on the fulfillment of the incoming student profile. Therefore, the findings of this study suggest that a well-structured curriculum is a key factor in ensuring that the admission profile is met.

Keywords: Curriculum design, Data mining, Higher education, Incoming student profile, Informatics.

1. Introduction

The incoming student profile plays a vital role in shaping future professionals, not only in informatics programs but also across most higher education fields. This profile reflects students' initial skills and knowledge, and it also helps identify potential academic risk factors or early dropout indicators [1].

According to the National System for the Evaluation, Accreditation, and Certification of Educational Quality (SINEACE), academic programs are expected to define and promote adherence to the incoming student profile. This alignment is designed to facilitate the transition of new students into university life, increasing their chances of retention and graduation [2].

International organizations further emphasize that academic programs should implement admission policies based on clear criteria aligned with the desired incoming profile.

Additionally, they advocate for the planning and execution of student leveling strategies to ensure that all entrants begin their studies with a strong foundational base. These practices are particularly important in maintaining quality standards in fields such as Computing, Informatics, Engineering, Technology, and related disciplines [3].

To ensure that graduates meet the expected professional outcomes, it is essential to define an admission profile that reflects the institution's context and includes relevant skills, knowledge, and personal attributes students are expected to bring into higher education [4].

Curriculum design, on the other hand, is inherently dynamic, often undergoing revisions every three to five years [5]. Given this, it becomes necessary to assess whether the curriculum design effectively aligns with and leverages the characteristics of incoming students.

It is also worth noting that in Peru and much of Latin America, admission and selection processes have traditionally relied on university entrance exams and rankings [6], based on the assumption that high-performing candidates are more likely to succeed in higher education. However, once students enter university, little effort has been made to evaluate whether they truly meet the defined incoming profile of their chosen program.

This study aims to explore how curriculum design influences the consolidation of the incoming student profile in informatics and related programs. Specifically, it examines student performance during the first three academic semesters to determine whether the expected profile is met within a given curricular structure. This analysis seeks to provide insights that support student retention and academic success, while also addressing the broader issues discussed above.

This paper is organized as follows: Section 2 reviews related work; Section 3 presents the materials and methods used; Section 4 discusses the findings related to curriculum design; and Section 5 offers conclusions and final recommendations

2. Related Work

This section reviews key studies and relevant literature related to the evaluation of incoming student profiles in higher education, particularly within informatics and engineering programs in Latin America. Several important considerations should be noted:

- Informatics programs in Latin America are often integrated within broader engineering curricula.
- The selected studies primarily focus on first-year students and their transition into higher education.
- A variety of analytical and predictive techniques have been employed to explore the challenges associated with student profiles.

One foundational study characterizes incoming students in engineering programs as recent high school graduates, mostly male, with average to above-average academic performance and a predominantly scientific-humanistic educational background [4].

Another line of research stresses the need to revise the desired profile to attract students with creative and critical thinking skills. The lack of emphasis on creativity is highlighted, alongside a call to encourage the development of more innovative future engineers [7].

Further studies underscore the critical importance of meeting the defined admission profile during the first year. A strong correlation has been observed between students' resilience, engagement, and academic performance—factors that are directly tied to student retention and early dropout rates [8].

In an effort to predict academic success, one study examined the relationship between students' entry profiles and first-year academic performance. Key variables included high school GPA, SAT scores, and study habits (attitudes and time management). Using logistic regression, the study sought to identify predictors of future success in engineering education [9].

Shifting focus toward student retention, another investigation argues that identifying at-risk students is only the first step. The study advocates for a broader redesign of institutional policies to incorporate socio-economic factors. Machine learning techniques were proposed to deliver more personalized support to vulnerable students, recognizing that individual circumstances significantly affect academic persistence [10].

Building on predictive modeling, a study employing Bayesian Belief Networks (BBN) demonstrated the ability to forecast first-year dropout risk with an accuracy of 84%. The research suggests that Decision Support Systems (DSS) can be enhanced by selecting relevant features and eliminating redundant predictors, thus achieving better interpretability without sacrificing performance [11].

In a different approach, a comprehensive framework based on a Fuzzy Inference Model was developed to evaluate applicant competencies against program entry requirements. The system, implemented in the Matlab SDK, included both an assessment module and the fuzzy model itself. The results showed significant improvements in key performance indicators: efficiency (80.3%), resource utilization (45.7%), and user satisfaction (25.7%) [12].

Expanding further into the technological domain, recent research has leveraged artificial intelligence (AI) to transform human resource practices in candidate profiling. AI-driven models were applied to assess the fit between job positions and applicant profiles. Findings indicated improvements in the speed and accuracy of candidate selection, with reductions in both operational costs and human error during recruitment processes [13].

3. Materials and Methods

This study aims to assess the extent to which incoming students meet the expected entry profile in informatics programs and to examine whether curriculum design plays a significant role in shaping or reinforcing that profile. To address this objective, the research is guided by the following key questions:

- Is there a relationship between curriculum design in informatics programs and the consolidation of the incoming student profile?
- Do variations in curriculum design influence the achievement of the expected profile among incoming students in informatics programs?
- What patterns can be uncovered between curriculum design and profile fulfillment through the application of data mining techniques?

To explore these questions, the study adopts a data-driven analytical approach, applying knowledge discovery methods to educational records. By leveraging data mining, the research seeks to uncover hidden trends and associations that may inform improvements in curricular strategies and student onboarding practices.

To achieve the primary objective of this study, an exploratory research design was adopted, supported by data analysis techniques rooted in Knowledge Discovery in Databases (KDD), commonly referred to as data mining. The exploratory nature of descriptive data mining focuses on uncovering patterns, correlations, trends, clusters, and trajectories that reveal meaningful relationships within large datasets [14].

The dataset used in this research consists of academic records from undergraduate students enrolled in an informatics program. Spanning approximately a decade, the dataset was originally organized in separate spreadsheets containing information on student profiles, grades, and grade point averages. This rich dataset served as the foundation for identifying variations in curriculum design—such as changes in course structures and credit distributions—while also enabling the tracking of students' academic progress over time.

The methodology for implementing data mining varies across the literature, with different authors proposing distinct conceptual models. Some sources equate data mining with the broader KDD process, while others define it as a critical stage within the multi-step knowledge discovery framework. For this study, the latter interpretation was adopted, treating data mining as a key component within the larger KDD pipeline [15].

Figure 1 illustrates the process model followed to construct and implement the data mining approach used in this investigation.

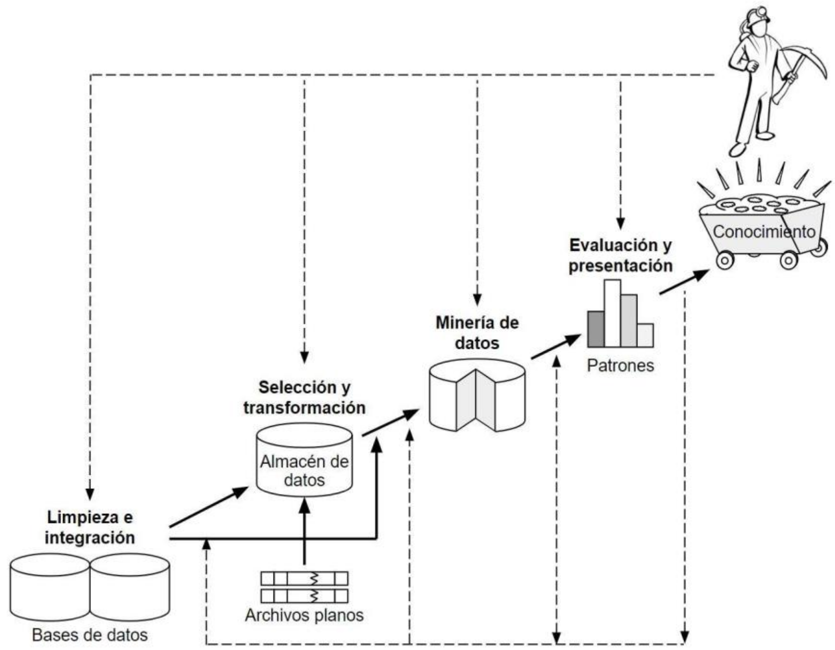


Figure 1 – Knowledge Discovery Process

Regarding the methodologies employed, the dataset was subjected to a structured analysis process using specialized data mining tools. Within the data mining framework, the clustering technique was selected due to its descriptive nature and its effectiveness in identifying natural groupings within the data [14]. To enhance the depth and contextual understanding of the case, complementary research methods were also applied, including document analysis, direct observation, and semi-structured interviews.

The study followed the stages defined by the Knowledge Discovery in Databases (KDD) process, which served as the methodological backbone of the investigation.

Initially, the data were obtained from the institution in raw form, without any prior preprocessing. During the data preparation phase, an ETL (Extract, Transform, Load) process was executed. This step involved data cleaning, transformation, and selection to correct inconsistencies and standardize the dataset, ultimately consolidating it into a structured and relevant form suitable for analysis [15]. As a result, a curated and analyzable dataset was made available for the data mining phase.

During the data mining stage, an evaluation model was developed to assess the alignment of students with the expected incoming profile [16]. Figure 2 presents the proposed evaluation model, which assesses profile fulfillment within the context of varying curriculum designs.

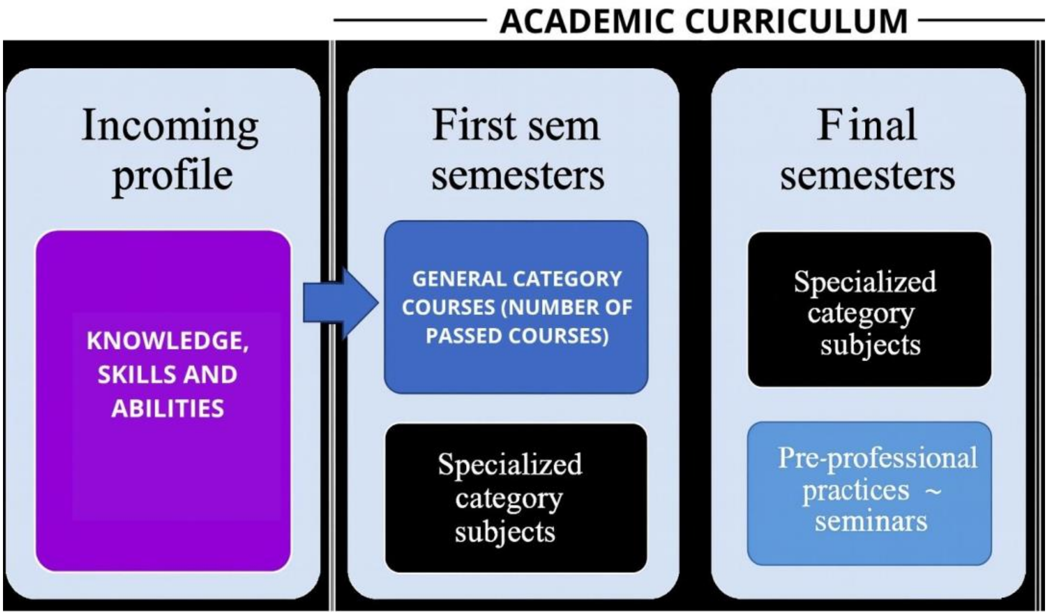


Figure 2. Proposed model for evaluating the incoming student profile

The analytical approach employed in this study relied on clustering, a descriptive data mining technique commonly used to explore and understand underlying structures in datasets before the application of predictive models [17]. Clustering was chosen for its ability to reveal meaningful patterns and natural groupings within the data. Multiple algorithms were applied to ensure the robustness of the findings, allowing for the validation and reinforcement of observed results through consistent pattern detection across different computational approaches.

4. Results

The following section presents the individual results for each curriculum design, along with a consolidated summary of findings across the four curricular structures.

4.1. Results by Curriculum Design

As outlined in the model presented in Figure 2, the evaluation process begins with the defined incoming student profile comprising knowledge, skills, abilities, and desirable generic competencies, formulated and officially published by the academic program itself. Using this profile as a reference, a mapping was conducted to align the intended competencies with corresponding formative courses across each curriculum design. This mapping is presented in Table 1.

It is important to note that formative courses are typically delivered within the first three academic semesters. Consequently, the analysis of student performance was limited to this timeframe. This focus is further supported by related studies, which identify the third semester as a critical point marked by increased rates of student attrition.

Table 1. Mapping of Incoming Student Profile Characteristics to Courses and Categories

Profile Attribute	Course	Category
Numerical reasoning and comprehension skills	Mathematics I	EFG (General Training Component)
Verbal comprehension skills	Communication Workshop Seminar	EFG (General Training Component)
Ability to identify models, perform calculations, formulate and test hypotheses through deduction, induction, and abstraction	Study and Research Techniques	EFG (General Training Component)
Development of logical-mathematical intelligence	Mathematics I	CB (Basic Component)
Development of logical-mathematical intelligence	Mathematics II	CB (Basic Component)
Linguistic intelligence: ability to understand the order and meaning of words in reading, speaking, listening, and writing effectively; communication and teamwork skills	Oral and Written Communication Workshop	CGH (Humanistic General Culture Component)

Based on the criteria previously outlined, a refined analytical dataset—referred to as the sensitive view—was constructed for data mining purposes. This dataset was filtered to include student performance results specifically within the general education and foundational training categories. These categories, identified by codes such as OCG, FCG, CGM, CB, EFG, and EBE, represent core academic domains expected to be well-developed in incoming students for successful performance during the early semesters of study.

Following the preparation of the sensitive view, the data were subjected to clustering analysis to uncover emergent patterns. It is important to note that data processing was conducted separately for each curriculum design, as course structures and grading policies differed across versions. Figure 3 displays the output of the model applied to the first curriculum design, using the K-means clustering algorithm to group students based on their academic performance indicators.

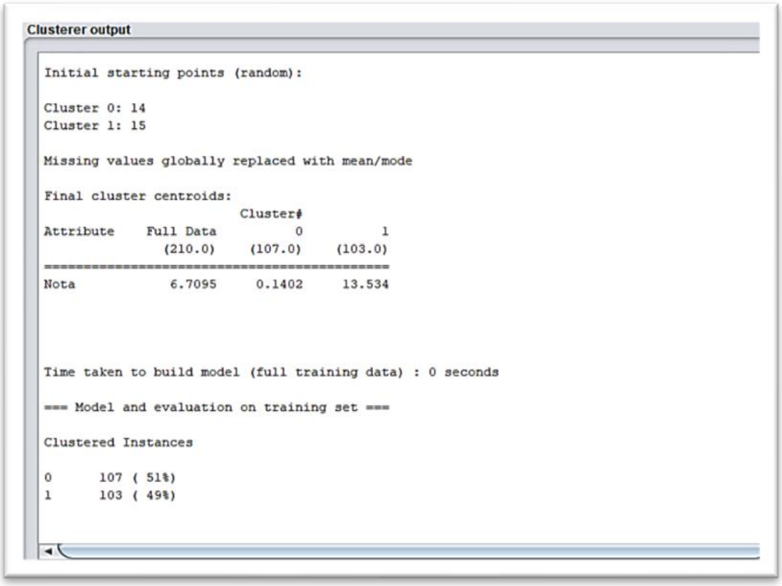


Figure 3. Results for Curriculum Design 1

Figure 4 presents a scatter plot illustrating the performance indicator, specifically the 'Grade' field. Two distinct clusters can be observed, each defined by its centroid. Cluster 0 accounts for approximately 51% of the data and is characterized by an average grade close to 0, indicating unsatisfactory or failing performance. In contrast, Cluster 1 comprises around 49% of the dataset, with an average grade near 13, reflecting a passing level of academic achievement. These results highlight a clear bifurcation in student performance during the early semesters, as captured through the clustering analysis.

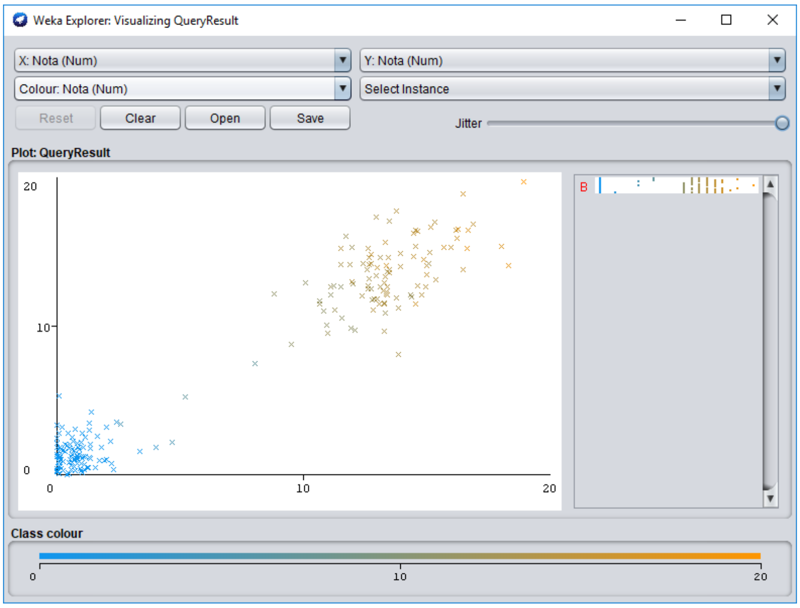


Figure 4. Scatter Plot for Curriculum Design 1

The results for the second curriculum design (Curriculum 2) are presented in Figure 5. As in the previous case, the K-means clustering algorithm was applied, with particular emphasis on the performance indicator Grade and its distribution patterns.

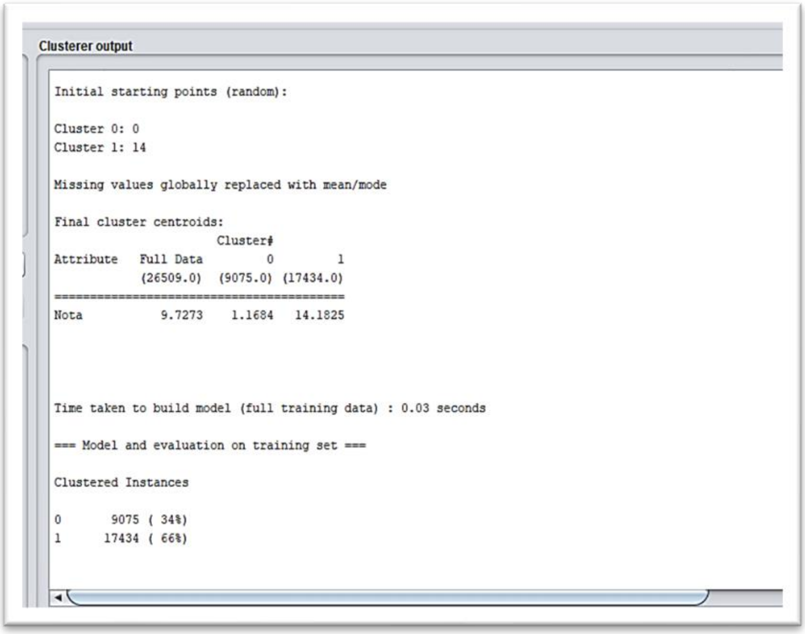


Figure 5. Results for Curriculum Design 2

Figure 6 illustrates the formation of two distinct clusters derived from the analysis. Cluster 0 encompasses approximately 34% of the dataset and is characterized by an average grade close to 1, indicating unsatisfactory academic performance. In contrast, Cluster 1 represents roughly 66% of the data and exhibits an average grade of 14, suggesting a satisfactory or passing level of achievement. These clustering results provide further insight into the distribution of student performance within the second curriculum design.

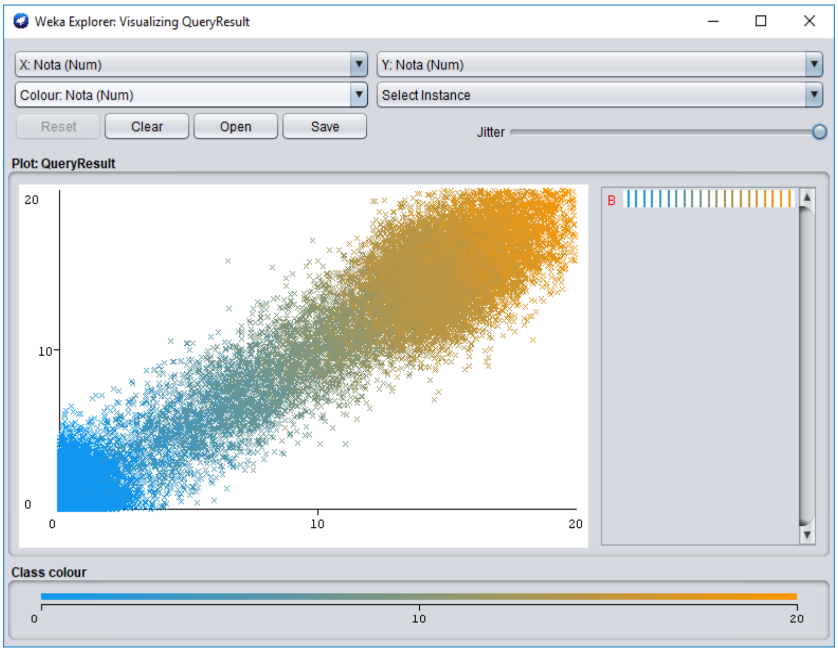
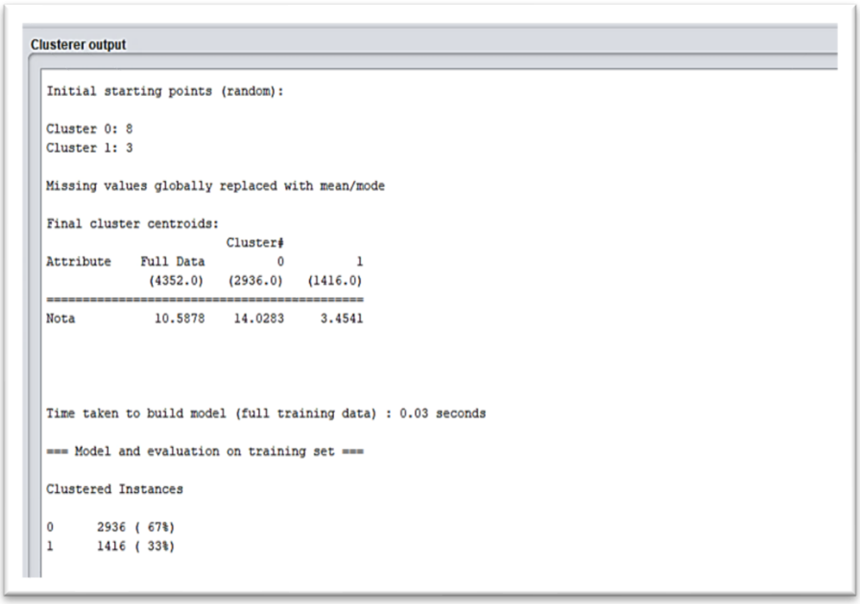


Figure 6. Results for Curriculum Design 2

Figures 7 and 8 present the results corresponding to Curriculum Design 3. The analysis reveals the formation of two clusters: Cluster 0, which accounts for approximately 67% of the data, and Cluster 1, representing the remaining 33%. Cluster 0 is characterized by an average grade close to 14, indicating satisfactory academic performance. In contrast, Cluster 1 displays an average grade near 3, reflecting a failing performance. These results highlight a clear segmentation in student outcomes under this particular curriculum design.



```
Clusterer output

Initial starting points (random):

Cluster 0: 8
Cluster 1: 3

Missing values globally replaced with mean/mode

Final cluster centroids:
                         Cluster#
Attribute    Full Data          0          1
              (4352.0)   (2936.0)   (1416.0)
============================================
Nota           10.5878    14.0283     3.4541



Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      2936 ( 67%)
1      1416 ( 33%)
```

Figure 7. Results for Curriculum Design 3

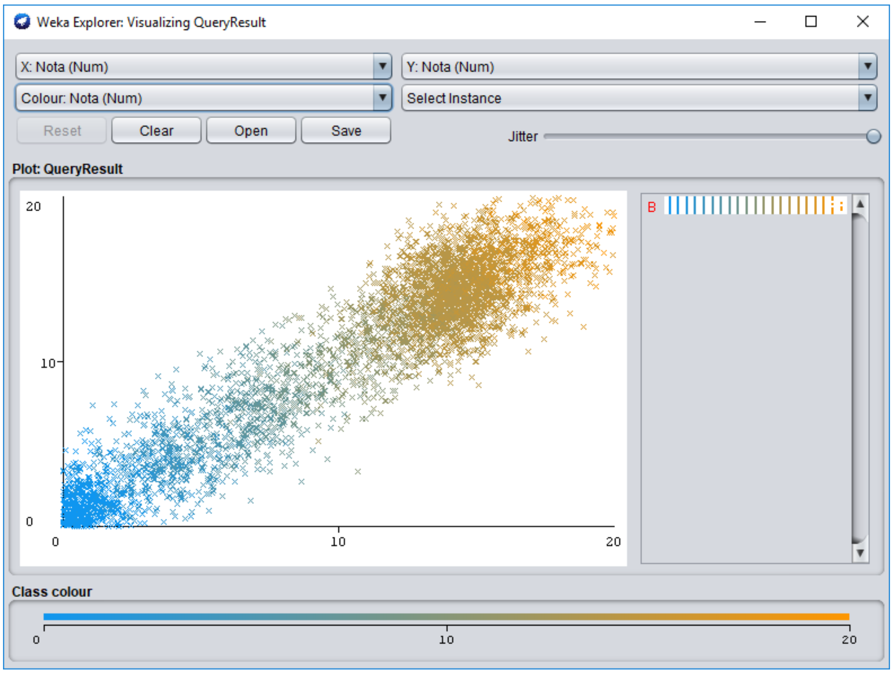


Figure 8. Graph of the curriculum design of Curriculum 3

Figure 9 shows the results corresponding to the design of the last curriculum of the Dataset analyzed.

```
Clusterer output

Initial starting points (random):

Cluster 0: 9
Cluster 1: 15

Missing values globally replaced with mean/mode

Final cluster centroids:
                          Cluster#
Attribute     Full Data          0           1
               (1792.0)    (456.0)    (1336.0)
==============================================
Nota            11.8677     3.2368     14.8136



Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       456 ( 25%)
1      1336 ( 75%)
```

Figure 9. Curricular design results of Curriculum 4

Figure 10 displays the clustering results, revealing two distinct groups. Cluster 0 comprises approximately 25% of the dataset and is associated with an average grade near 3, indicating a trend toward failing performance. In contrast, Cluster 1 includes about 75% of the data and shows an average grade close to 14, suggesting a generally satisfactory academic outcome. These findings underscore a significant performance gap within the student population under this curriculum design.

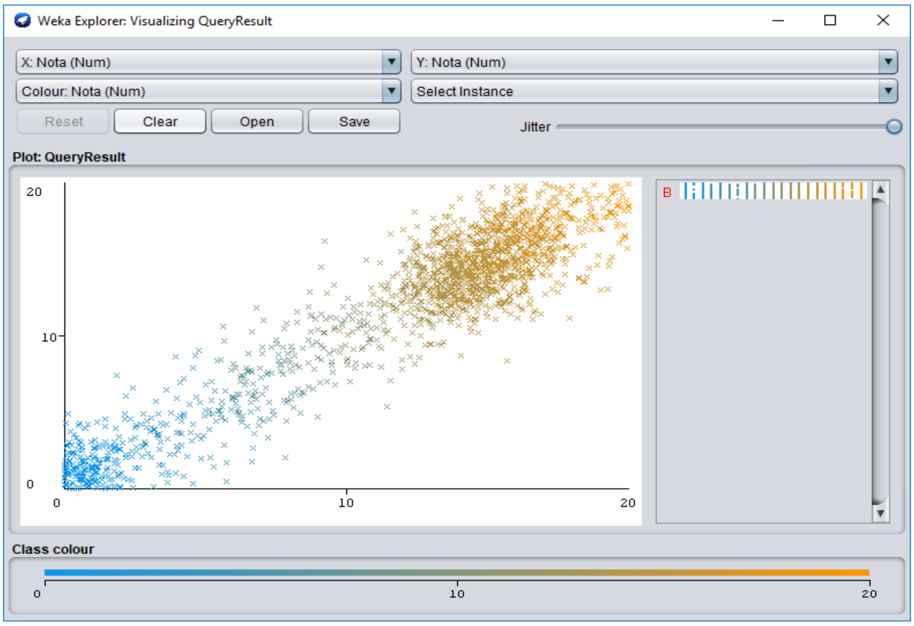


Figure 10. Curricular design graph of Curriculum 4

4.2. Results summary

Table 2 below shows the consolidated results according to curricular design. There is a trend towards the consolidation of the profile of the entrant according to the curricular design, being these indicators being more optimal in the curricular design "Curriculum 4", which shows a higher number of students passing the course.

Table 2. Summary of Results by Curriculum Design

Curriculum Design	Cluster 0 Centroid	Cluster 1 Centroid	Passed (%)	Failed (%)
Curriculum 1	0.1402	13.534	49%	51%
Curriculum 2	1.1684	14.1825	66%	34%
Curriculum 3	14.0283	3.4541	67%	33%
Curriculum 4	3.2368	14.8136	75%	25%

To enhance the above visualization, the bar chart in Figure 11 was created to highlight this trend.

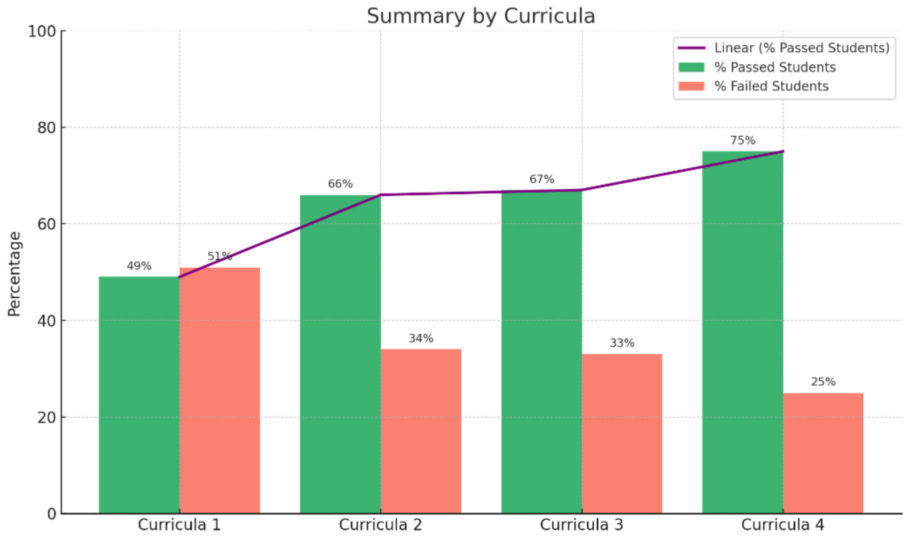


Figure 11. Summary chart by study plans

5. Conclusions

The results of this research show progressive growth percentages of passing grades according to different curricular designs. These results support the importance of curricular design for the consolidation of the entrant profile in computer science programs. The study shows that curricular design can guarantee that students strengthen and consolidate the entry profile in the first semester.

These findings highlight the need for academic programs and institutions to regularly review and adjust their curricular design, thus contributing to the academic performance of students. It also helps to prevent early academic desertion.

In comparison to previous research on the profile of the entrant, this research focused on evaluating, through related subjects, the capabilities that the entrant student should have in his or her first semesters of training. These can be summarized in 4 competencies: numerical and verbal competencies, analytical and scientific thinking, logical and mathematical intelligence, and linguistic intelligence.

In this research, we sought to describe the behavior of student performance during the first three semesters, in order to identify whether they meet the necessary entry profile for good performance in a specific curricular design. While other studies evaluated the first year of studies to predict future academic success. Both studies have in common the importance of evaluating the entry profile; a student who meets the entry profile is likely to achieve future student outcomes, which in turn will lead him/her to meet the graduate competencies and graduate objectives.

It is recommended that the comprehensive evaluation of the entry profile be expanded to include socioeconomic aspects, such as family income level, access to educational resources, family and social environment, among others. The inclusion of these aspects would provide a more complete understanding of the entrant profile.

It is recommended that this research be extended to the labor arena, exploring how data mining could contribute to the evaluation of candidate profiles in the labor market. This would be beneficial for both professionals in search of opportunities and employers in candidate selection.

The limited availability of data and the scarcity of information from academic computer science programs at other universities may affect the generalization of the model to broader contexts.

References

[1] A. Llauró, D. Fonseca, S. Romero, M. Aláez, J. T. Lucas, and M. M. Felipe, "Identification and comparison of the main variables affecting early university dropout rates according to knowledge area and institution," Heliyon, vol. 9, no. 6, p. e17435, Jun. 2023, doi: 10.1016/j.heliyon.2023.e17435.

[2] SINEACE, "Explicación de estándares del modelo de acreditación de programas de estudios de educación superior universitaria," 2018, Accessed: Sep. 12, 2023. [Online]. Available: www.sineace.gob.pe

[3] ICACIT, "Acreditación en Computación," https://www.icacit.org.pe/web/archivos/2024_ICACIT_Criterios.pdf. Accessed: Sep. 12, 2023. [Online]. Available: https://www.icacit.org.pe/web/es/acreditacion/criterios-de-acreditacion/acreditacion-en-computacion

[4] A. Suazo-Schwencke, P. Blanc, F. Salgado, A. Suazo-Schwencke, P. Blanc, and F. Salgado, "Definition of entry profiles in engineering from the characteristics of the students," Ingeniare. Revista chilena de ingeniería, vol. 30, no. 4, pp. 803–820, Dec. 2022, doi: 10.4067/S0718-33052022000400803.

[5] Ministerio de Educacion, "Ley Universitaria 30220," https://www.gob.pe/institucion/minedu/normas-legales/118482-30220.

[6] I. García Ojalvo, J. Galarza López, R. Sepúlveda Lima, I. García Ojalvo, J. Galarza López, and R. Sepúlveda Lima, "El proceso de ingreso a la educación superior. Los sistemas informáticos como herramientas para su ejecución," Revista Cubana de Educación Superior, vol. 39, no. 3, 2020, Accessed: Sep. 12, 2023. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0257-43142020000300018&lng=es&nrm=iso&tlng=pt

[7] S. A. Atwood and J. E. Pretz, "Creativity as a Factor in Persistence and Academic Achievement of Engineering Undergraduates," Journal of Engineering Education, vol. 105, no. 4, pp. 540–559, Oct. 2016, doi: 10.1002/JEE.20130.

[8] J. C. Ayala and G. Manzano, "Academic performance of first-year university students: the influence of resilience and engagement," Higher Education Research & Development, vol. 37, no. 7, pp. 1321–1335, Nov. 2018, doi: 10.1080/07294360.2018.1502258.

[9] H. Taylor et al., “How the Entry Profiles and Early Study Habits Are Related to First-Year Academic Performance in Engineering Programs,” Sustainability 2022, Vol. 14, Page 15400, vol. 14, no. 22, p. 15400, Nov. 2022, doi: 10.3390/SU142215400.

[10] D. Delen, B. Davazdahemami, and E. Rasouli Dezfouli, “Predicting and Mitigating Freshmen Student Attrition: A Local-Explainable Machine Learning Framework,” Information Systems Frontiers, pp. 1–22, Apr. 2023, doi: 10.1007/S10796-023-10397-3/FIGURES/8.

[11] D. Delen, K. Topuz, and E. Eryarsoy, “Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition,” Eur J Oper Res, vol. 281, no. 3, pp. 575–587, Mar. 2020, doi: 10.1016/J.EJOR.2019.03.037.

[12] O. L. Loaiza, N. Saboya, J. Bustamante, and J. J. Soria, “Framework Based on a Fuzzy Inference Model for the Selection of University Program Applicants,” Advances in Intelligent Systems and Computing, vol. 984, pp. 131–140, 2019, doi: 10.1007/978-3-030-19807-7_14/COVER.

[13] G. M. Sridevi and S. Kamala Suganthi, “Artificial Intelligence-Based Job Applicant Profile Quality Assessment and Relevance Ranking Using Clusters in Talent Acquisition Process,” pp. 461–472, 2021, doi: 10.1007/978-981-33-4604-8_37.

[14] P.-Ning. Tan, Michael. Steinbach, Vipin. Kumar, and Anuj. Karpatne, Introduction to Data Mining eBook: Global Edition. Pearson Education Limited, 2019.

[15] J. Han, J. Pei, and H. Tong, Data mining, concepts and techniques., vol. 4. 2022.

[16] V. Sosa, “Modelo para la evaluación del perfil del ingresante universitario utilizando técnicas de Data Mining,” 2019, Universidad Nacional de San Agustín de Arequipa. Accessed: Sep. 13, 2023. [Online]. Available: http://repositorio.unsa.edu.pe/handle/UNSA/11023

[17] C. Aggarwal and C. Reddy, Data Clustering: Algorithms and Applications, 1st ed. 2014.