Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 12 Issue 9, September - 2025

Failure Rates and Graduation of Industrial Engineering Students by Cohort Cohort

[1]Alicia Cortés Fernández, [2]Ma. Elizabeth Montiel Huerta, [3]Alejandra Torres López, [3]María Inés Hernández Díaz
[1]Department of Systems and Computing, [2] Department of Economic-Administrative Sciences,
[3]Department of Industrial Engineering
Tecnológico Nacional de México / Instituto Tecnológico de Apizaco, Tlaxcala, México

***Abstract*— This study analyzes the academic performance of Industrial Engineering students at the Tecnológico Nacional de México/Instituto Tecnológico de Apizaco, focusing on failure rates by generational cohort and degree completion, using a descriptive quantitative approach. The data were normalized using StandardScaler, and K-Means clustering was applied with k=3 and random_state=42, supported by the elbow method. Dimensionality reduction through principal component analysis facilitated visualization of each cluster in two dimensions. The first cluster groups courses with high failure rates, highlighting a significant increase in the January-June 2018 cohort. The second cluster includes subjects with medium failure rates without extreme patterns. The third cluster shows low to moderate failure levels, with stable averages ranging from thirteen to twenty-one percent, without significant spikes.**

The graduation analysis revealed that in the January-June 2017 cohort, sixty percent of graduates were female and forty percent male. In the August-December 2017 cohort, the female proportion increased to sixty-eight percent while the male proportion decreased to thirty-two percent. These findings allow identification of trends and areas for improvement in the Industrial Engineering educational process.

***Keywords*— Cluster; Generational cohort; Enrollment; Course failure; Graduation**

I. INTRODUCTION

The Instituto Tecnológico de Apizaco offers the Industrial Engineering program based on the current Study Plan IIND-2010-227, which comprises 47 courses within a Generic Structure and those included in the “Quality and Manufacturing 4.0” Specialty Module, totaling 5 courses. The duration to complete the curriculum up to the professional internship is 10 to 12 semesters. The grading scale ranges from 0 (zero) to 100 (one hundred) in any evaluation opportunity, and the minimum passing grade for a course is 70 (seventy) [1].

The Integral Graduation Guidelines of Tecnológico Nacional de México [2], establish that integral graduation validates the competencies (knowledge, skills, and attitudes) the student acquired and developed during their professional education. Integral graduation can be attained through the following types of projects: a) Professional Internship, Research and/or Technological Development Project, Integrative Project, Productive Project, Technological Innovation Project, Entrepreneurship Project, Dual Education Integral Project, Internship, Thesis or Dissertation; and b) obtaining a Satisfactory or Outstanding Performance Certificate in the General Undergraduate Exam (EGEL) from the National Center for Higher Education Evaluation (CENEVAL).

Academic performance is a key indicator in evaluating educational quality, as it reflects not only the level of acquired learning but also the effectiveness of pedagogical processes implemented by an institution. Specifically, within the context of Tecnológico Nacional de México / Instituto Tecnológico de Apizaco, ongoing evaluation of the learning outcomes of Industrial Engineering students is essential to ensure educational processes are effective and meet environmental demands. The primary objective of this study is to examine the academic performance of students in this program, emphasizing two critical variables: course failure and degree completion. These variables reflect students’ academic success level and help identify potential areas for improving internal training processes.

Course failure not only reflects students’ academic difficulties but also may affect their retention and motivation within the program [3]. Conversely, degree completion represents the successful culmination of academic training, and its success rate is an important indicator of the educational quality delivered by the institution [4]; [5].

The employed methodology is quantitative with a descriptive focus [16]; [17]; [18], allowing detailed analysis of failure rates and graduation by generational cohorts. This segmentation is essential since each cohort faces different social, economic, and academic contexts that can significantly influence their educational outcomes. By comparing student cohorts across different years, behavioral patterns can be identified that might otherwise remain unnoticed. These patterns may relate to internal factors such as academic load, student profile, and teaching-learning methodologies or external factors like demographic differences among students.

II. Literature Review

In the current context of Higher Education, no institution is exempt from facing challenges associated with failure rates and low academic performance. This persistent and multifactorial issue has prompted Higher Education Institutions to undertake systematic research processes aimed at identifying factors influencing student performance.

From pedagogical, sociocultural, institutional, and technological perspectives, various indicators have been explored to gain a deeper understanding of the structural and contextual causes underlying these phenomena [7]; [8]. Key areas of analysis include students' admission conditions and prior academic trajectories, the relevance of curricular models, teaching quality, access to learning resources, and the impact of academic support strategies. Furthermore, the use of analytical tools and digital platforms has expanded the capacity for monitoring and evaluation, generating empirical evidence to guide decision-making for continuous improvement [9].

Academic failure occurs when a student enrolled in courses or learning units at an educational institution receives from an instructor or evaluation committee [6]; [10], a grade insufficient to pass those courses according to the institution's established evaluation criteria. This reflects poor academic performance due to failure to meet regulatory requirements.

Several authors have conducted studies on failure at the university level. For example, one study analyzed quantitative data from the Academic Leveling Scholarship program, focusing on a sample of 250 first-year students at a public university in Chile [11]. The study employed advanced statistical techniques, such as clustering and decision trees, to explore institutional databases seeking useful information to develop an early warning system. Findings revealed that scores obtained in the specific mathematics section of the university entrance exam were key variables differentiating students by risk of dropout, offering valuable evidence for the university to implement strategies aimed at improving student retention.

Separately, in applying data mining to develop classification and prediction models using the CRISP-DM methodology, three specific problems were addressed: dropout prediction, failure prediction by academic cycle, and failure prediction by course. The models achieved accuracy rates above 73%, constituting a valuable tool for the University of Cuenca in Ecuador [12].

In another study [13], strategies were proposed to minimize course failure rates among students in the Computer Engineering program and thereby increase the program's terminal efficiency. Proposed interventions included specialized faculty tutoring, psychological support, and motivational talks.

Additionally, regarding students experiencing academic delay due to course failure at a Mexican public university [14], the analysis focused on students failing three or more courses (*n3*). It was concluded that the highest proportion of students with *n3* delay were enrolled in Engineering and Technology programs, followed by Natural and Exact Sciences, and the Common Core courses taught during the first university year in several areas.

A study on language policy and its relationship with graduation and learning loss rates at the Universidad Autónoma del Estado de México (UAEMÉX) found that expanding graduation modalities did not significantly increase the graduation rate. It also identified a lack of compensation between terminal efficiency percentages and graduation rates by cohort [15].

III. Basic Idea, Background and Methodology

A quantitative methodology with a descriptive approach and clustering technique was employed using the Python programming language to organize, analyze, and clearly present the quantitative data of each generational cohort. Additionally, data processing related to academic degree attainment was performed using statistical assistants in Microsoft Excel 2016 spreadsheets.

A. *Clustering technique for failure*

The process of this methodology focuses on four basic steps [16]; [17]: data selection, cleaning, preparation, and transformation. Each phase is described in Fig. 1:

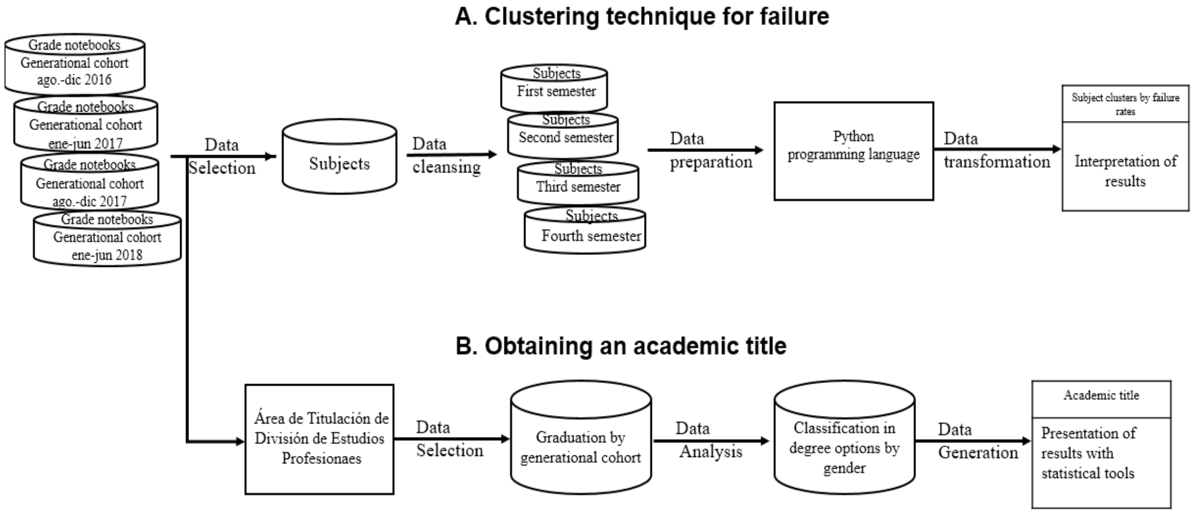

Fig. 1. Methodology of the project under study. (A). Clustering technique for falure. (B). Obtaining an academic title.

Data selection.

This initial stage is crucial for the precise identification of data subject to analysis, ensuring that results are reliable and meaningful. Within this quantitative approach, information was collected directly from individual academic records, known as kárdex, for each generational cohort. This phase involved a thorough study of students who failed courses, focusing exclusively on cases relevant to the study's objective.

The kárdex serves as an indispensable primary source because it provides a detailed overview of each student's academic performance. Reviewing these records yields key data, such as failed courses and the generational cohorts in which these failures occurred details that delineate problematic areas.

Data cleansing.

Once the selection stage is completed, the variables of interest to serve as the foundation for the analysis are determined. These variables include factors such as the failure rate per course and the frequency of failure by generational cohort.

After data collection, to ensure reliability and readiness for analysis, the following steps are carried out:

- Removal of irrelevant data. If inconsistencies are detected based on data behavior, such data must be eliminated.
- Data labeling. It is essential to structure the data properly and facilitate analysis by reorganizing the information to ensure compatibility.

This processing begins with data labeling, a step that ensures clarity and organization throughout the entire analysis. Using Microsoft Excel, data records are labeled based on the failure rates corresponding to each generational cohort for the courses involved. It is important to note that the highest failure rates occur during the first four semesters of each cohort. These labels serve as identifiers that enable coherent structuring and classification of the information, as shown in TABLE 1, where data are grouped according to assigned labels such as generational cohort, courses, and failure rate values.

TABLE I. Índices de reprobación por asignatura y por cohorte generacional.

	Cohort			
Subject Name	**Ago a Dic 2016**	**Ene a Jun 2017**	**Ago a Dic 2017**	**Ene a Jun 2018**
Fundamentos de Investigación	41%	83%	36%	12%
Taller de Ética	16%	8%	20%	6%
Cálculo Diferencial	12%	42%	28%	50%
Taller de herramientas Intelectuales	14%	16%	23%	26%
Química	21%	28%	28%	17%
Dibujo Industrial	14%	21%	42%	8%
Electricidad y Electrónica industrial	14%	42%	19%	13%
Propiedad de los materiales	26%	27%	23%	8%
Cálculo Integral	18%	68%	22%	24%
Probabilidad y Estadística	18%	18%	25%	26%
Análisis de la realidad Nacional	12%	12%	6%	17%
Taller de Liderazgo	29%	12%	20%	16%
Procesos de fabricación	12%	4%	11%	25%
Metrología y normalización	24%	21%	5%	18%
Álgebra Lineal	9%	47%	21%	50%
Cálculo Vectorial	17%	31%	22%	46%
Economía	17%	8%	9%	18%
Estadística Inferencial 1	32%	13%	14%	50%
Estudio del trabajo 1	22%	37%	49%	17%
Higiene y seguridad industrial	12%	8%	25%	32%
Física	6%	23%	0%	25%
Algoritmos y Lenguajes de programación	1%	19%	23%	13%

Data preparation

A fundamental stage in computational analysis processes involves the structured transformation of data using specialized platforms that enable proper and efficient processing [18]. This phase not only entails the technical conversion of data into formats compatible with analytical models but also their organization into sets that share similar properties, which is essential to ensure semantic coherence and the validity of the obtained results.

The success of this stage depends on the accurate identification of patterns, relevant attributes, and clustering criteria aligned with the study's objectives. To achieve this, the following is considered:

Within-Cluster Sum of Squares (WCSS), a metric used in clustering analysis algorithms such as k-means. Its implementation in Python allows optimizing the selection of the appropriate number of clusters. Effective visualizations of clustering, as shown in Fig. 2, indicate a break at number three, leading to the selection of three clusters for this study. [19]; [21].

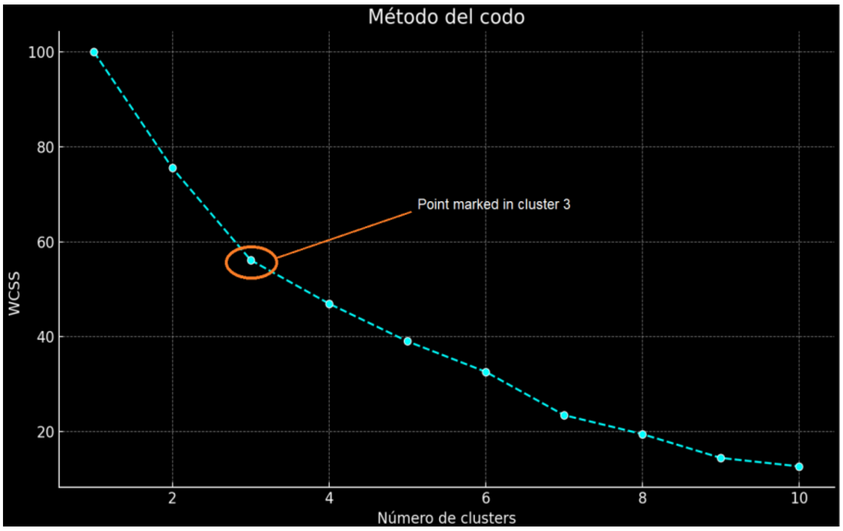

Fig. 2. Elbow method to determine K=3 (number of clusters).

Data normalization or standardization constitutes a fundamental preprocessing step for statistical analysis using clustering algorithms (Fig. 3). This procedure is especially essential when dealing with datasets that include associated percentage values from various subjects. Its application ensures comparability and homogeneity of the data, which is crucial to avoid biases and distortions in the statistical analysis [22].

Asignatura	2016-2	2017-1	2017-2	2018-1
Fundamentos de Investigación	0.41	0.83	0.36	0.12
Taller de Ética	0.16	0.08	0.20	0.06
Cálculo Diferencial	0.12	0.42	0.28	0.50
Taller de herramientas Intelectuales	0.14	0.16	0.23	0.26
Química	0.21	0.28	0.28	0.17
Dibujo Industrial	0.14	0.21	0.42	0.08
Electricidad y Electrónica industrial	0.14	0.42	0.19	0.13
Propiedad de los materiales	0.26	0.27	0.23	0.08
Cálculo Integral	0.18	0.68	0.22	0.24
Probabilidad y Estadística	0.18	0.18	0.25	0.26
Análisis de la realidad Nacional	0.12	0.12	0.06	0.17
Taller de Liderazgo	0.29	0.12	0.20	0.16
Metrología y normalización	0.24	0.21	0.05	0.18
Álgebra Lineal	0.09	0.47	0.21	0.50
Cálculo Vectorial	0.17	0.31	0.22	0.46
Economía	0.17	0.08	0.09	0.18
Estadistíca Inferencial 1	0.32	0.13	0.14	0.50
Estudio del trabajo 1	0.22	0.37	0.49	0.17
Higiene y seguridad industrial	0.12	0.08	0.25	0.32
Física	0.06	0.23	0.00	0.25
Algoritmos y Lenguajes de programación	0.01	0.19	0.23	0.13
Investigación de operaciones 1	0.25	0.08	0.42	0.78
Estadística inferencial 2	0.08	0.08	0.30	0.41
Estudio del trabajo 2	0.16	0.08	0.05	0.25

Fig. 3. Normalized Data (standardscaler python).

Previously normalized data ensure that the results obtained through the k-means algorithm are comparable, consistent, and relevant, thereby facilitating an appropriate and reliable interpretation in the analysis [20].

Data transformation

The statistical transformation of data using a clustering algorithm in the Python programming language is a fundamental step to maximize the potential of descriptive analysis. Once the data are organized and prepared, the necessary techniques are applied, allowing the variables to be adapted into more manageable or relevant formats for clustering. With the transformed data, analysis is conducted using clustering algorithms such as k-means or hierarchical methods to reveal hidden patterns and subgroups with common characteristics.

Resulting data

The results are summarized using scatter plots, which facilitate the visualization and understanding of the underlying data structure. The interpretation of these clustering plots allows raw data to be transformed into valuable information to support informed and strategic decision-making.

B. Obtaining an academic title

First, data from each generational cohort are collected. These data are organized into clear categories such as unique student identifiers, gender, cohort membership, and graduation status. To facilitate handling, the data are structured in a database platform such as Excel.

Data cleaning and preparation are crucial to ensure quality. This involves correcting inconsistent values, removing duplicates, and verifying that all records contain the necessary fields, such as gender and generational cohort.

IV. RESULTS

A. Clustering technique for failure

The implementation of the K-Means clustering algorithm was carried out as part of an analysis of failure rates across four academic semesters with the highest incidence. This process involved the use of several specialized libraries to perform the fundamental stages of data analysis, starting with the loading of the collected data and continuing with its cleaning and preliminary processing. These initial steps ensured the quality and relevance of the information used, preparing the data for clustering.

The K-Means method was employed to partition the data into three clusters, defined based on common characteristics related to the failure rates of the subjects analyzed. This algorithm stood out for its ability to classify subjects according to similar patterns, facilitating the identification of recurring trends among groups of courses. The resulting clusters provided a structured view of academic performance in relation to critical subjects.

The analysis results are presented in Fig. 4, where each cluster is distinguished by different colors, offering a clear and visually understandable representation of the groupings. This approach not only allowed the identification of subjects with the highest failure rates within each cluster but also provided valuable information to design targeted intervention strategies aimed at improving students' academic performance.

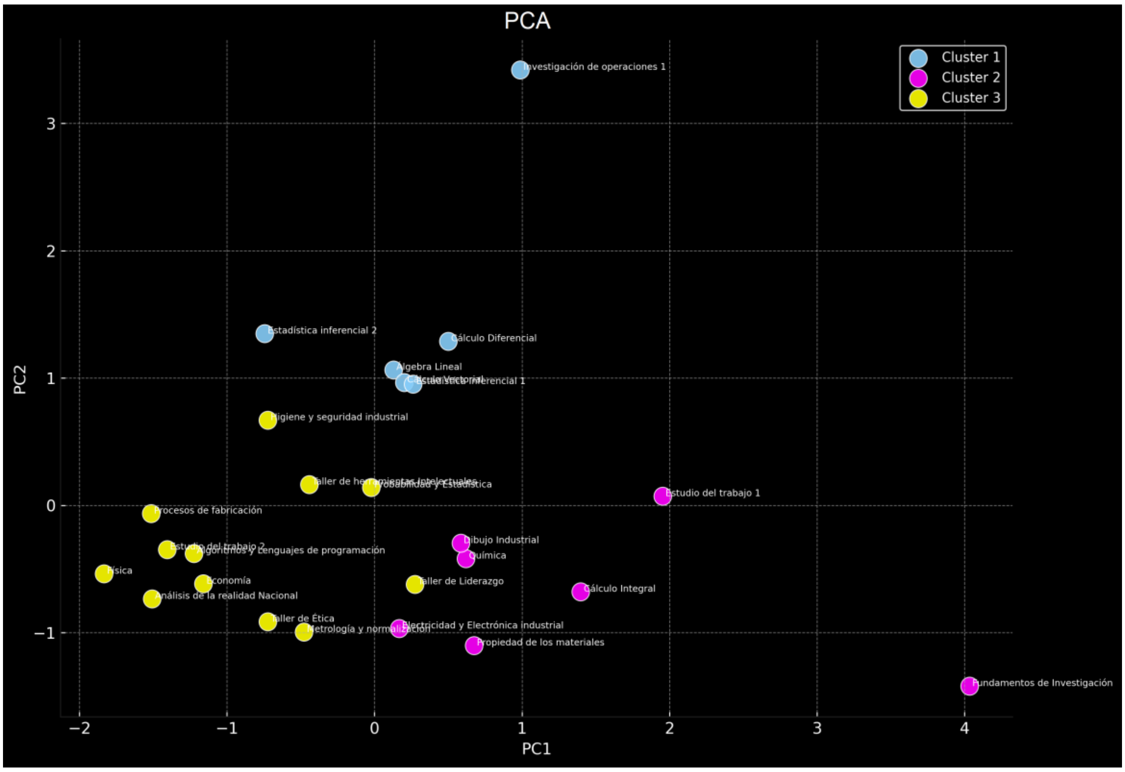

Fig. 4. Subject Clusters Based on Failure Rates

The resulting description of each cluster allows for a detailed understanding of the academic behavior of the grouped subjects.

Cluster 1 (Blue) groups subjects with high failure rates sustained over the four analyzed semesters. This cluster exhibited an atypical behavior during the Jan-Jun 2018 cohort, characterized by a significantly elevated failure average (52.5%). This group of subjects shows thematic concentration. Notable examples include: Operations Research I, with a failure rate of 78% in the Jan-Jun 2018 cohort, and Inferential Statistics I, with a 50% failure rate in the same period. The observed pattern suggests a punctual increase in failure rates exclusive to that semester, without continuity in subsequent cycles. This statistical anomaly highlights a specific focus area within the curriculum, with relevant implications for institutional monitoring and academic management.

The cluster's characterization provides empirical evidence for designing targeted follow-up and improvement strategies, as well as for systematic review of critical indicators in high-complexity subjects.

Cluster 2 (Fuchsia) analysis reveals a significant behavior characterized by a notable peak in failure rates, reaching an average of approximately 43.7%. This cluster groups subjects that experienced spikes in failure rates in specific cases such as: Research Fundamentals with 83%, and Integral Calculus with 68% failure rates. It is worth noting that the following generational cohort of 2018 shows considerably reduced averages around 14.1%, suggesting that the observed issue was temporary and possibly linked to specific factors in previous cohorts.

Cluster 3 (Yellow) is characterized by low to moderate failure rates, with relatively stable averages ranging approximately between 13% and 21%. This group includes subjects that do not exhibit significant peaks in any single cohort. Although the urgency for intervention is lower compared to other clusters, continuous monitoring is recommended to ensure educational quality and prevent potential future increases in failure rates. This preventive approach is key to sustaining academic performance in these courses and ensuring an optimal learning experience for students.

B. Obtaining an academic title

The graduation rate of Industrial Engineering students is a key indicator of progress and success within student generational cohorts. Once the data are organized, metrics such as the percentage of graduated students by cohort and by gender are calculated. These metrics are essential to understand graduation trends and enable more detailed comparisons between male and female students.

In Fig. 5, students' academic achievements can be visualized, allowing the identification of patterns associated with gender and the different generational cohorts studied.

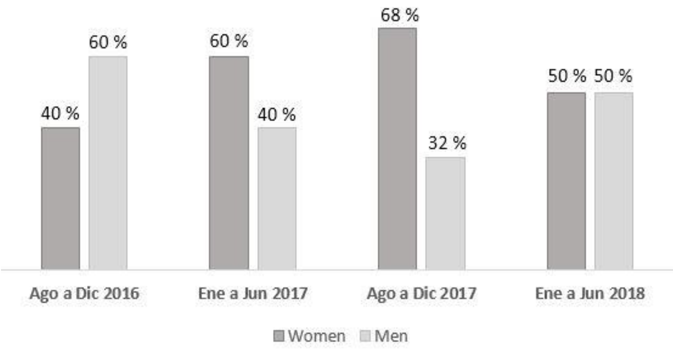

Fig. 5. Obtaining an academic title

The analysis of the graduation distribution presented in Fig. 5 reveals a significant transformation in the participation of women and men in completing higher education studies.

At the close of the Aug-Dec 2016 cohort, the proportion of graduates showed a male predominance, with 60% men compared to 40% women.

For the Jan-Jun 2017 cohort, the situation changed significantly, with higher female representation. During this period, women accounted for 60% of the graduates, while men represented 40%, indicating progress toward educational equity.

In the Aug-Dec 2017 generational cohort, the upward trend in female participation in degree completion was reaffirmed, with women representing 68%, while the percentage of male graduates decreased to 32%. These data reflect a greater female presence in completing higher education.

Finally, in the Jan-Jun 2018 cohort, the graduation distribution reached equilibrium, with 50% women and 50% men.

The graduation outcomes for these cohorts underwent significant evolution. The reduction of the gender gap in graduation reflects the strengthening of equality in the academic environment, allowing more students, regardless of gender, to complete their education and access better professional opportunities.

V. Conclusion

The study provided a comprehensive perspective on the academic performance and achievements of students in the Industrial Engineering program at the Tecnológico Nacional de México/Instituto Tecnológico de Apizaco, generating a deep and enriching analysis focused on two fundamental aspects: the failure rates of subjects in each cohort presented and the graduation rates by generational cohort.

By employing quantitative descriptive techniques and clustering technology implemented in the Python programming language, key patterns were identified that offer a clear view of the students' academic progress.

Regarding failure rates, the analysis allowed for identifying subjects with the greatest difficulties for students in the cohorts from Aug-Dec 2016, Jan-Jun 2017, Aug-Dec 2017, and Jan-Jun 2018. These results highlighted specific areas requiring attention to improve academic performance, such as strengthening pedagogical strategies or implementing support programs for critical subjects. The use of clustering emphasized groups of students with similar academic behaviors, facilitating the identification of recurrent trends in failure rates across different generational cohorts.

Concerning graduation rates, the analysis of the percentage of graduates by gender and generational cohort revealed significant differences, providing valuable data to evaluate the effectiveness of the educational program over time. Both cohorts with outstanding performance and those with lower graduation rates were observed, highlighting the importance of designing specific strategies to support students during the final stages of their academic training.

References

[1] Tecnológico Nacional de México. *Normateca del TecNM.* (20 de 12 de 2024). Obtenido de https://www.tecnm.mx/?vista=Normateca

[2] Tecnológico Nacional de México. *Normateca* (17 de 01 de 2025). Obtenido de https://www.tecnm.mx/?vista=Normateca

[3] Ordóñez Urgilés, Milton Xavier, and César Emiliano Noboa Cisneros. Identificación de los factores de deserción de una carrera universitaria en una institución de educación superior, usando técnicas de minería de datos. MS thesis. 2018.

[4] Báez, Irma Yazmín Hernández, et al. Análisis de índices de aprovechamiento, reprobación y deserción escolar mediante minería de datos. ANFEI Digital, 2019,

[5] Barragán Silva, Homero Vinicio. Una aplicación web para el apoyo de toma de decisiones en el proceso de Titulación de la Carrera de Software-Espoch. (2023).

[6] Lagunas, Javier Rodríguez, and Juan Manuel Manuel Hernández Vázquez. *La deserción escolar universitaria en México*. La experiencia de la Universidad Autónoma Metropolitana, Campus Iztapalapa/The university scholastic desertion in Mexico. The experience of the Universidad Autónoma Metropolitana Campus Iztapalapa." Actualidades Investigativas en educación (2008).

[7] Ramírez-Noriega, Alan, et al. Un modelo de minería de datos para predecir la deserción escolar en la carrera de Ingeniería de Software A data mining model to predict school dropout in the Software Engineering career. Obtenido el 4 de enero del 2025.

[8] López, Maricela Quintana, et al. Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la

Deserción Escolar. Advances in Computing Science (2013): 13-23.

[9] Marcillo Pincay, Joselyn Mercedes. Visualización de información y minería de datos para analisis de patrones en los proyectos de titulación de la facultad de ciencias técnicas en la unesum. BS thesis. Jipijapa-Unesum, 2024.

[10] Urbina-Nájera, Argelia Berenice; Camino-Hampshire, José Carlos and Cruz-Barbosa, R. *Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa*. RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa (2020).

[11] Contreras, C. (2021). Determinación de variables predictivas de deserción inicial para generar un sistema de alerta temprana. Análisis sobre una muestra de estudiantes beneficiarios de la beca de nivelación académica en una universidad pública en Chile. Calidad en la educación, (2021):12-45.

[12] Loja Rodas, Cesar Gabriel. Aplicación de técnicas de minería de datos en el contexto del rendimiento académico en la Universidad de Cuenca. (2019).

[13] Contreras Contreras, Ghiordy Ferney, et al. Development Methodology of Techniques for Data Clustering Using Machine Learning.Tecnura (2022): 42-58.

[14] Mendiola, José Luis Aguirre, et al. *Análisis de deserción escolar con minería de datos*. Res. Comput. Sci. (2015): 71-82.

[15] Nava Gómez, Guadalupe Nancy, and Rosa María Alemán Martínez. Estudio sobre la política lingüística, índice de titulación y pérdida del aprendizaje durante la pandemia de COVID-19 en la Universidad Autónoma del Estado de México. Revista Educación. (2023): 172-200.

[16] Hernández Orallo, José. *Introducción a la Minería de Datos*. (2004).

[17] Perez Gutierrez, Boris Rainiero. Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. Revista UIS Ingenierías. (2020): 193-204.

[18] Marcillo Pincay, Joselyn Mercedes. Visualización de información y minería de datos para analisis de patrones en los proyectos de titulación de la facultad de ciencias técnicas en la unesum. BS thesis. Jipijapa-Unesum, 2024.

[19] Ruiz, M. D. (2024). Clustering. Universidad de Granada.

[20] Casaverde, J. P. G., & Reategui, M. S. A. H. R. Predicción del rendimiento académico mediante algoritmos de machine learning en la Institución Educativa. Moyobamba. Obtenido el 18 de enero del 2025.

[21] UNIR, "Clustering: qué es y cuál es su aplicación en Big Data," 29-abr-2024. [En línea]. Disponible en: https://www.unir.net/revista/ingenieria/clustering/

[22] Stuart, Alan y Keith Ord. Teoría avanzada de la estadística de Kendall, teoría de la distribución . John Wiley & Sons, 2010.