Data-Driven Admission Strategies: A Case Study of the Colombian School of Engineering Julio Garavito

Angie Lorena Duran Chaparro*

Department of Information Technology, Julio Garavito Colombian School of Engineering, Bogota, Colombia

ABSTRACT

This study explores the optimization of the admission process at the Colombian School of Engineering Julio Garavito through data analysis. By integrating advanced tools like power BI, python flask and data mining techniques, a model has been developed to improve student selection and retention, enabling more informed and effective decisions. The results indicate that applying data science not only facilitates efficient data management but also promotes more inclusive and equitable educational policies. This pioneering approach in the field of higher education in Colombia offers a framework for future research and technological developments, highlighting the vital importance of data science in improving educational and administrative processes.

Keywords: University admission; Data analysis; Academic performance; Data science in education; Educational innovation; Equity in higher education; Admission policies

INTRODUCTION

Currently, educational institutions face the constant challenge of optimizing their admission and student retention processes. The Colombian School of Engineering Julio Garavito, aware of this reality, has embarked on an innovative project focused on the exhaustive analysis of student data. This approach allows for a deeper understanding of various factors such as academic performance, socioeconomic characteristics and student origins, providing a comprehensive view that contributes to more informed and effective decision-making [1].

In this project, the creation of innovative technological tools has been prioritized to strengthen and sustain data management at the Colombian School of Engineering Julio Garavito in the long term. A central element is the development of a power BI dashboard, designed to provide executives and administrators with an intuitive platform for analyzing statistical data and visualizing key information on student performance and trends. This dashboard, enriched with precise and up-to-date data, has been the product of a collaborative effort, including weekly meetings with university executives and administrators, where requirements were refined and functionalities adapted to their specific needs.

Additionally, an excel file cleaning assistant has been implemented, essential for improving the operational efficiency of data management. Moreover, as part of the initiative to make information accessible, a virtual assistant in the form of a chat based on a pre-existing model developed by mixtral has been adapted. Although the chat is in its initial implementation phase, it has the potential to improve immediate and direct interaction with users requiring specific information. These tools, along with the data dictionary and user manuals, ensure that both the data and applications are not only used correctly but also fully understood. These collective innovations address not only current needs but also establish the foundations for future adaptations and improvements in data management [2].

The document will present a detailed analysis of the obtained results and the developed technological tools, discussing how data interpretation can positively influence decision-making within the university. The findings of this study are expected to contribute not only to improving the admission process of the Colombian School of Engineering Julio Garavito but also to providing a framework for other areas of the university to explore and leverage the power of data analysis in their own administrative and educational processes.

State of the art

Currently, data analysis in the educational field has gained significant relevance, especially in optimizing admission and academic evaluation processes in Colombia. Studies applying data mining techniques to evaluate socioeconomic and demographic factors impacting performance in standardized exams, such as Saber Pro, highlight the importance of considering individual student characteristics to improve educational predictive models, emphasizing the effectiveness of different models and data mining techniques in predicting academic outcomes.

On the other hand, the integration of data science in education is not limited to the national scope, but global trends are also observed where data mining supports decision-making in university admission systems. In this context, models have been developed that not only predict academic performance but also help institutions formulate more inclusive and equitable policies.

In Colombia, the application of data science has allowed the formulation of more informed admission policies in universities, using criteria beyond previous academic performance, as evidenced in the study. This study suggests that data science facilitates more efficient resource management and improves strategic decision-making in educational institutions [3].

LITERATURE REVIEW

This research adopts the KDD (Knowledge Discovery in Databases) methodology to analyze the effectiveness of admission policies at the Colombian School of Engineering Julio Garavito, with a particular emphasis on improving student selection and retention through exhaustive statistical analysis.

Data selection

The data used comes from various administrative records of the institution, meticulously selected to cover different dimensions of the admission process and student academic performance. The main datasets include:

Academic Bible: Compiles detailed information per academic period on courses, professors and students, covering data from course identification to demographic and academic information of students [4].

Escolaris: Contains school enrollment records prior to the implementation of the "Enlace" tool, covering details from enrollment date to Saber 11 scores.

Enlace: Represents the evolution of enrollment records, offering a detailed history starting from the first semester of 2021, with more complete and structured data.

Dropout: Data including the causes of academic and non-academic dropout.

Enrolled in other universities: Information on students enrolled in other educational institutions.

Knowledge exams: Data on the results of exams measuring specific student competencies.

Data preprocessing and analysis

Preprocessing was carried out using python, utilizing libraries like pandas for data cleaning, handling missing values and normalization. Data from different sources were consolidated into a uniform format to facilitate comparative and temporal analyses.

Data analysis was conducted by integrating power BI for interactive visualizations and trend explorations over time. The focus was on identifying trends and patterns to assess the effectiveness of admission policies and student retention. Exploratory data analysis was used to identify key characteristics and relationships in the data that could indicate areas for improvement in admission processes [5].

RESULTS

Academic dynamics: Correlations, performance and student dropout

The study addressed various aspects related to education in different academic periods, highlighting patterns in score correlations, social and geographical distributions, academic performance and school dropout through a quantitative and analytical approach.

Regarding score correlations, it was found that during the 2021-1 academic period, there was a strong correlation between scores in different areas evaluated in the Saber 11 exam, with values exceeding 90%. This observation suggests a significant relationship between good performance in a specific area and overall good performance. Although this trend showed a slight decrease in the 2022-2 period, with correlations above 85%, the idea that solid exam results could imply good performance in all areas remained. However, in periods like 2021-2 and 2022-1, correlations varied, with some dropping to 77%, still indicating strong connections between certain knowledge areas.

In terms of social and geographical distribution, the analysis revealed that most students came from socioeconomic stratum 3. However, there was a notable increase in the enrollment of students from stratum 2 during the 2022-2 period, balancing the participation of these two strata. Geographically, Bogotá was consolidated as the main destination for students, followed by Cundinamarca, Boyacá, Meta and Tolima, indicating patterns of student preference and mobility.

Academic performance showed an initial trend of grade concentration around the value of 4, with gradual dispersion as students progressed in their careers. This pattern was particularly notable in specific regions such as Huila, Casanare and Meta, as well as in cities like Chía, Cota and Sogamoso, where high grade concentration was maintained, highlighting outstanding and possibly homogeneous academic performance among students from these areas [6].

School dropout, analyzed for both academic and non-academic causes, presented significant figures, especially for non-academic dropouts, which reached a maximum of 97% in the 2020-1 period and remained above 45% in subsequent periods. This

finding underscores the importance of considering external factors to the academic field in student retention strategies. Dropout varied by career, with increases observed in some disciplines and decreases in others, suggesting the influence of specific program factors on student retention.

Recommendations for a more inclusive and effective admission and academic monitoring policy

The data analysis conducted in the study on admission policy and academic performance at the Colombian School of Engineering Julio Garavito yielded significant results suggesting strategic changes to improve inclusivity and effectiveness in the selection and student monitoring process. It highlights the need to adopt a more inclusive admission policy, proposing adjustments to the Saber 11 score cut-off point. The recommendation is based on the increase of students coming from both cities near Bogotá and distant departments, pointing towards the elimination of the restriction based solely on Bogotá scores. It is suggested to establish a general cut-off score that promotes equal opportunities for all applicants, regardless of their geographical location.

The study also revealed a positive correlation between the Saber 11 score and academic performance in foundational subjects, validating its use as an admission criterion. However, it is advised to consider the overall Saber 11 score rather than focusing on specific areas, due to the existing correlation between the different areas evaluated in the exam.

Another key recommendation is the evaluation of students' academic backgrounds in the admission process. It was observed that those with a previous focus in areas such as mathematics, systems or who possess a technical degree or prior university experience, show a more solid preparation and robust academic foundation [7].

Regarding access to advanced subjects, it is recommended not to limit it exclusively to the Saber 11 score. Instead, it is proposed to adapt the knowledge exam to the specific topics of the university curriculum, ensuring that students have the necessary knowledge for their academic advancement.

The importance of students' psychological and emotional well-being, especially those entering directly from high school, was highlighted. Therefore, it is suggested to implement support programs, such as mentorship and career guidance, to provide necessary support during their transition to higher education.

Finally, the implementation of continuous monitoring strategies of students' academic performance throughout their career is proposed. This includes the use of predictive models to identify potential dropout risks and the provision of timely support to ensure academic success. Continuous monitoring of academic performance can allow early intervention in case of difficulties, ensuring the student maintains a consistent and satisfactory academic performance [8].

DISCUSSION

The results obtained in the study reflect the importance of data analysis in improving the admission and student retention

processes at the Colombian School of Engineering Julio Garavito. The information derived from the analysis of academic data, correlations and patterns provides a solid foundation for making informed decisions that promote inclusivity and equity in the admission process [9].

The proposed recommendations aim to optimize the selection process and ensure that students admitted to the institution have the necessary tools and support to succeed academically. In addition, the implementation of continuous monitoring strategies and predictive models can help anticipate and address potential dropout risks, providing timely support to students who need it. This study demonstrates that the integration of data science in educational decision-making can have a significant impact on the quality and effectiveness of university policies. The adoption of innovative tools and methodologies for data analysis, such as power BI and predictive models, facilitates a comprehensive understanding of academic and socio-demographic dynamics, enabling institutions to implement more inclusive and equitable policies.

The conclusions drawn from this research can serve as a reference for other educational institutions seeking to improve their admission and student retention processes. By applying similar methodologies and tools, it is possible to achieve a deeper understanding of the factors that influence academic performance and student success, ultimately contributing to the development of more effective and inclusive educational policies [10].

CONCLUSIONS

This research highlights the importance of integrating data science into the admission and student retention processes at the Colombian School of Engineering Julio Garavito. The results obtained from the analysis of academic data, correlations and patterns provide a solid foundation for making informed decisions that promote inclusivity and equity in the admission process.

The proposed recommendations aim to optimize the selection process and ensure that students admitted to the institution have the necessary tools and support to succeed academically. In addition, the implementation of continuous monitoring strategies and predictive models can help anticipate and address potential dropout risks, providing timely support to students who need it. This study demonstrates that the integration of data science in educational decision-making can have a significant impact on the quality and effectiveness of university policies. The adoption of innovative tools and methodologies for data analysis, such as Power BI and predictive models, facilitates a comprehensive understanding of academic and socio-demographic dynamics, enabling institutions to implement more inclusive and equitable policies.

The conclusions drawn from this research can serve as a reference for other educational institutions seeking to improve their admission and student retention processes. By applying similar methodologies and tools, it is possible to achieve a deeper understanding of the factors that influence academic performance and student success, ultimately contributing to the

development of more effective and inclusive educational policies.

REFERENCES

1. Goethe EV, Colina CM. Taking advantage of diversity within the classroom. J Chem Educ. 2018;95(2):189-192.

2. Helal S, Li J, Liu L, Ebrahimie E, Dawson S, Murray DJ, et al. Predicting academic performance by considering student heterogeneity. Knowl Based Syst. 2018;161:134-146.

3. Mengash HA. Using data mining techniques to predict student performance to support decision making in university admission systems. IEEE Access. 2020;8:55462-55470.

4. Yudono MA, Faris RM, de Wibowo A, Sidik M, Sembiring F, Aji SF. Fuzzy decision support system for ABC University student admission selection. InInternational Conference on Economics, Management and Accounting, 2022, pp: 230-237. Atlantis Press.

5. Marcinkowski F, Kieslich K, Starke C, Lunich M. Implications of AI (un-) fairness in higher education admissions: The effects of perceived AI (un-) fairness on exit, voice and organizational reputation. InProceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp: 122-130.

6. Castillo AM. Inclusive education and legal framework in Colombia. A look from otherness and differences. Psychological Thesis. 2020;15(2):1-32.

7. Torres NI, Buitrago D. Is there inequality in career choice in Colombia?: An analysis of some sociodemographic factors. Soc Educ Hist. 2020;9(2):128-153.

8. Fissalma H, Ajie H. Design and development of student admission system in the state university of Jakarta. KnE Soc Sci. 2019:92-101.

9. Johora FT, Anindita A, Islam N, Islam M, Hasan M. Centralized data driven decision making system for Bangladeshi University admission. InScience and Information Conference, Cham: Springer International Publishing, 2022, pp: 316-330.

10. Shen W, Luo L, Luo L, Zhang L, Zhu T. A data-driven newsvendor model for elective-emergency admission control under uncertain inpatient bed capacity. J Evid Based Med. 2024;17(1):78-85.