Article

# Data-Driven Admission Strategies: A Case Study of the Escuela Colombiana de Ingeniería Julio Garavito

Angie Duran *

*Article*

# Data-Driven Admission Strategies: A Case Study of the Escuela Colombiana de Ingeniería Julio Garavito

**Angie Lorena Duran Chaparro**

Bogotá, Escuela Colombiana de Ingeniería Julio Garavito, Colombia; angieduran953@gmail.com

**Abstract:** This study explores the optimization of the admission process at the Escuela Colombiana de Ingeniería Julio Garavito through data anal- ysis. 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 poli- cies. This pioneering approach in the field of higher education in Colombia offers a frame- work for future research and technological de- velopments, 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

## 1. Introduction

Currently, educational institutions face the constant challenge of optimizing their admission and student retention processes. The Escuela Colombiana de In- geniería Julio Garavito, aware of this reality, has em- barked on an innovative project focused on the ex- haustive analysis of student data. This approach al- lows for a deeper understanding of various factors such as academic performance, socioeconomic char- acteristics, and student origins, providing a compre- hensive view that contributes to more informed and effective decision-making.

In this project, the creation of innovative techno- logical tools has been prioritized to strengthen and sustain data management at the Escuela Colombiana de Ingeniería Julio Garavito in the long term. A cen- tral element is the development of a Power BI dash- board, designed to provide executives and adminis- trators with an intuitive platform for analyzing statis- tical data and visualizing key information on student performance and trends. This dashboard, enriched with precise and up-to-date data, has been the prod- uct of a collaborative effort, including weekly meet- ings 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 oper- ational 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 implemen- tation phase, it has the potential to improve imme- diate and direct interaction with users requiring spe- cific 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.

The document will present a detailed analysis of the obtained results and the developed technologi- cal tools, discussing how data interpretation can pos- itively influence decision-making within the univer- sity. The findings of this study are expected to con- tribute not only to improving the admission process of the Escuela Colombiana de Ingeniería Julio Gar- avito but also to providing a framework for other ar- eas of the university to explore and leverage the power of data analysis in their own administrative and ed- ucational processes.

## 2. State of the Art

Currently, data analysis in the educational field has gained significant relevance, especially in optimiz- ing 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 educa- tional predictive models, emphasizing the effective- ness of different models and data mining techniques in predicting academic outcomes [**Jimenez2018**].

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 sys- tems. In this context, models have been developed that not only predict academic performance but also help institutions formulate more inclusive and equi- table policies [**fairadmission2023**].

In Colombia, the application of data science has allowed the formulation of more informed admission policies in universities, using criteria beyond previ- ous academic performance, as evidenced in the study [**Melendez2020UniversidadHormigas**]. This study suggests that data science facilitates more ef- ficient resource management and improves strategic decision-making in educational institutions.

## 3. Methodology

This research adopts the KDD (Knowledge Discovery in Databases) methodology to analyze the effective- ness of admission policies at the Escuela Colombiana de Ingeniería Julio Garavito, with a particular em- phasis on improving student selection and retention through exhaustive statistical analysis.

### 3.1. Data Selection

The data used comes from various administrative records of the institution, meticulously selected to cover different dimensions of the admission pro- cess 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 identifica- tion to demographic and academic information of students.

- 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 institu- tions.

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

### 3.2. Data Preprocessing and Analysis

Preprocessing was carried out using Python, utiliz- ing libraries like Pandas for data cleaning, handling missing values, and normalization. Data from differ- ent 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 explo- rations over time. The focus was on identifying trends and patterns to assess the effectiveness of ad- mission policies and student retention. Exploratory data analysis was used to identify key

characteristics and relationships in the data that could indicate ar- eas for improvement in admission processes.

## 4. Results

### 4.1. Academic Dynamics: Correla- tions, Performance, and Student Dropout

The study addressed various aspects related to ed- ucation in different academic periods, highlighting patterns in score correlations, social and geograph- ical distributions, academic performance, and school dropout through a quantitative and analytical ap- proach.

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 Saber11 exam, with values exceed- ing 90%. This observation suggests a significant rela- tionship 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 re- sults could imply good performance in all areas re- mained. However, in periods like 2021-2 and 2022-1, correlations varied, with some dropping to 77%, still indicating strong connections between certain knowl- edge 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, Bo- gotá 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 ca- reers. This pattern was particularly notable in spe- cific regions such as Huila, Casanare, and Meta, as well as in cities like Chía, Cota, and Sogamoso, where high grade concentration was maintained, highlight- ing outstanding and possibly homogeneous academic performance among students from these areas.

**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 under- scores 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.

### 4.2. Recommendations for a More In- clusive and Effective Admission and Academic Monitoring Policy

The data analysis conducted in the study on admis- sion policy and academic performance at the Escuela Colombiana de Ingeniería Julio Garavito yielded sig- nificant results suggesting strategic changes to im- prove 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 Saber11 score cut-off point. The recommendation is based on the increase of students coming from both cities near Bogotá and distant de- partments, pointing towards the elimination of the restriction based solely on Bogotá scores. It is sug- gested to establish a general cut-off score that pro- motes equal opportunities for all applicants, regard- less of their geographical location.

The study also revealed a positive correlation be- tween the Saber11 score and academic performance in foundational subjects, validating its use as an ad- mission criterion. However, it is advised to consider the overall Saber11 score rather than focusing on spe- cific 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 previ- ous focus in areas such as mathematics, systems, or who possess a technical degree or prior university ex- perience, show a more solid preparation and robust academic foundation.

Regarding access to advanced subjects, it is rec- ommended not to limit it exclusively to the Saber11 score. Instead, it is proposed to adapt the knowl- edge exam to the specific topics of the

university cur- riculum, ensuring that students have the necessary knowledge for their academic advancement.

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

Finally, the implementation of continuous moni- toring 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.

## 5. Discussion

The results obtained in the study reflect the impor- tance of data analysis in improving the admission and student retention processes at the Escuela Colom- biana de Ingeniería Julio Garavito. The information derived from the analysis of academic data, correla- tions, and patterns provides 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 ad- mitted to the institution have the necessary tools and support to succeed academically. In addition, the im- plementation of continuous monitoring strategies and predictive models can help anticipate and address po- tential dropout risks, providing timely support to stu- dents 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 seek- ing 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 educa- tional policies.

## 6. Conclusions

This research highlights the importance of integrat- ing data science into the admission and student re- tention processes at the Escuela Colombiana de In- geniería Julio Garavito. The results obtained from the analysis of academic data, correlations, and pat- terns 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 ad- mitted to the institution have the necessary tools and support to succeed academically. In addition, the im- plementation of continuous monitoring strategies and predictive models can help anticipate and address po- tential dropout risks, providing timely support to stu- dents 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 seek- ing 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 educa- tional policies.

**Online Resources**

As part of this research, several digital tools and on- line resources were developed to facilitate data analy- sis and provide access to relevant project information. Below are the links to these resources, including the GitHub code repository and the project's associated web pages:

1.  **Data Dictionary:** https://diccionario-d e-datos---eci.glitch.me/

This link leads to the data dictionary used in the research, providing clear and detailed definitions of each analyzed variable.

2.  **User Manual - Tool Maintenance:** https://manual-de-usuario-mantenimiento-de-la-herramienta.glitch.me/

Here you can find the user manual for the main- tenance of the developed tools, offering detailed instructions for their proper use and updates.

3.  **Practical User Guide:** https://guia-de-u suario.glitch.me/

Provides a practical guide for end-users on how to interact with the tools and systems developed during the research.

4.  **GitHub Code:** https://github.com/Angie Duran953/Tesis_Pol-ticas_de_admisi-n_E CI

This link directs to the GitHub repository where the source code of all the tools and IT systems developed as part of this project is hosted.

5.  **Main Page on Google Sites:** https://site s.google.com/view/tesiiis/p%C3%A1gina-p rincipal

Provides access to the project's main page hosted on Google Sites.

6.  **Excel File Cleaner:** http://angieduran65.p ythonanywhere.com/

Online tool for cleaning and preparing Excel files before analysis.

7.  **Virtual Assistant:** https://fxo.io/m/web-r obust-7496

Link to the virtual assistant developed to facil- itate interaction and access to relevant project information.

8.  **PowerBI Dashboard:** https://app.powerb i.com/groups/me/reports/ff5d2191-4bd3-4 a0d-aa84-41764e7c3467/ReportSection?ct id=50640584-2a40-4216-a84b-9b3ee0f3f6cf&experience=power-bi&bookmarkGuid=df8 871f4-9223-436c-9f78-bb9e944df092

Access to the PowerBI dashboard used for the visualization and analysis of the data collected during the research.

These resources are designed to support both fu- ture researchers and professionals interested in ex- ploring the findings and methodologies of this study further.

**References**

Durán Chaparro, A. L. (2024). Más Allá de la Admisión Tradicional: Un Enfoque Personalizado para la Escuela Colombiana de Ingeniería Julio Garavito (Tesis de Maestría, Escuela Colombiana de Ingeniería Julio Garavito). Repositorio Insti- tucional de la Escuela Colombiana de Ingeniería Julio Garavito.

Mistral AI. (2023, December 11). Mixtral of Ex- perts. Recuperado de https://mistral.ai/new s/mixtral-of-experts/Nuffield Foundation, "Fair Admission to Universi- ties in England: Improving Policy and Practice," Informe de un proyecto de investigación finan- ciado por la Nuffield Foundation, 2017/18.

Jiménez, J. (2018). Minería de datos educativos: análisis de los factores económicos, sociales y de- mográficos que influyen en el desempeño de las Pruebas Saber-Pro en estudiantes de ingeniería en Antioquia. Recuperado de: http://hdl.hand le.net/20.500.11912/4317.

P. Meléndez, L. Ramírez, J. S. Escobar, "Una universidad de hormigas: formulación de políti- cas de admisión en un programa universitario," en XVIII Congreso Latinoamericano Encuentro Colombiano y Capítulo Brasilero de Dinámica de Sistemas, 2020.

McKinsey and Company. Diversity Matters - LatAm - 2022. https://diversitymatters-lat am-2022.mckinsey.com/, Accedido en 01/2024.

E. V. Goethe and C. M. Colina, "Taking Advan- tage of Diversity within the Classroom," Journal of Chemical Education, 2017.

Nuffield Foundation, "Fair Admission to Universi- ties in England: Improving Policy and Practice," Informe de un proyecto de investigación finan- ciado por la Nuffield Foundation, 2017/18.

S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, Q. Long, "Predicting academic per- formance by considering student heterogeneity," Knowledge-Based Systems, vol. 161, pp. 134-146, Dec. 2018.

H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support De- cision Making in University Admission Systems," IEEE Access, recibido el 14 de febrero de 2020, aceptado el 11 de marzo de 2020, fecha de pub- licación 19 de marzo de 2020, fecha de la versión actual 30 de marzo de 2020. DOI: 10.1109/AC- CESS.

Mistral AI, "Mixtral of Experts," [Fecha de Pub- licación: 11 de diciembre de 2023]. URL: https://mistral.ai/news/mixtral-of-experts/.

Universidad Icesi, "Página Principal," disponible en: https://www.icesi.edu.co/es/, accedido en Mayo 2024.

Universidad del Rosario, "Página Principal," disponible en: https://urosario.edu.co/, ac- cedido en Mayo 2024.

Universidad de Antioquia, "Página Principal," disponible en: https://www.udea.edu.co/, ac- cedido en Mayo 2024.

Universidad Nacional de Colombia, "Página Principal," disponible en: https://unal.edu.c o/, accedido en Mayo 2024.

Universidad de los Andes, "Página Principal," disponible en: https://uniandes.edu.co/, ac- cedido en Mayo 2024.

M. A. S. Yudono, R. M. Faris, A. De Wibowo, M. Sidik, F. Sembiring, S. F. Aji, "Fuzzy De- cision Support System for ABC University Stu- dent Admission Selection," en Advances in Eco- nomics Business and Management Research, vol. 207, Proceedings of the International Confer- ence on Economics Management and Accounting (ICEMAC 2021), 2021.

P. T. T. Nga, N. T. Luong, T. H. Thang, T. D. Quy, "AN APPROACH FOR BUILDING A CHATBOT SYSTEM FOR THE ADMISSION PROCESS OF DA LAT UNIVERSITY," en TNU Journal of Science and Technology, vol. 227, no. 14, pp. 23-32, Agosto 2022, DOI: 10.34238/tnu-jst.6056.

F. Marcinkowski, K. Kieslich, C. Starke, M. Lünich, "Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organiza- tional reputation," en FAT* '20: Proceedings of the 2020 Conference on Fairness, Account- ability, and Transparency, pp. 122-130, DOI: 10.1145/3351095.3372867, Enero 2020.

P. Meléndez, L. Ramírez, J. S. Escobar, "Una universidad de hormigas: formulación de políti- cas de admisión en un programa universitario," en XVIII Congreso Latinoamericano Encuentro Colombiano y Capítulo Brasilero de Dinámica de Sistemas, 2020.

A. Heryati, M. I. Herdiansyah, "The Applica- tion of Data Mining by using K-Means Cluster- ing Method in Determining New Students' Ad- mission Promotion Strategy," en International Journal of Engineering and Advanced Technol- ogy (IJEAT), vol. 9, no. 3, Febrero 2020, DOI: 10.35940/ijeat.C5414.029320.

C. Bedoya-Dorado, G. Murillo-Vargas, C. H. González-Campo, "Gestión universitaria en tiem- pos de pandemia por COVID-19: análisis del sec- tor de la educación superior en Colombia," en Es- tudios Gerenciales, vol. 37, no. 159, pp. [inserte rango de páginas], Abril-Junio 2021, Epub 27 Julio 2021, DOI: 10.18046/j.estger.2021.159.4409.

A. M. García Castillo, "Educación inclusiva y marco legal en Colombia. Una mirada desde la alteridad y las diferencias," en Tesis Psicológica, vol. 15, no. 2, pp. 72–93, Julio-Diciembre 2020, DOI: 10.37511/tesis.v15n2a4.

Universidad EAFIT, "Página Principal," disponible en: https://www.eafit.edu.co/, accedido en Mayo 2024.

Pontificia Universidad Javeriana, "Página Prin- cipal," disponible en: https://www.javeriana. edu.co/, accedido en Mayo 2024.

Escuela Colombiana de Ingeniería Julio Gar- avito, "Página Principal," disponible en: https://www.escuelaing.edu.co/, accedido en Mayo 2024.

Universidad Distrital Francisco José de Caldas, "Inicio," disponible en: https://www.udistrit al.edu.co/inicio, accedido en Mayo 2024.

H. Zhang, "An Analysis of The Chinese College Admission System," PhD thesis, The University of Edinburgh, 2009.

N. Izquierdo Torres, D. Buitrago, "¿Hay De- sigualdad en la Elección de Carrera en Colom- bia?: Un Análisis de Algunos Factores Sociode- mográficos," en Social and Education History, vol. 9, no. 2, pp. 128-153, Junio 2020, DOI: 10.17583/hse.2020.4303.

H. Fissalma, H. Ajie, "Design and Devel- opment of Student Admission System in the State University of Jakarta," en 3rd UNJ In- ternational Conference on Technical and Voca- tional Education and Training 2018, KnE So- cial Sciences, pp. 92–101, Marzo 2019, DOI: 10.18502/kss.v3i12.4076.

StartUs Insights, "9 Natural Language Process- ing Trends in 2023," [Accessed: 11 de diciembre de 2023]. URL: https://www.startus-insight s.com/innovators-guide/natural-languag e-processing-trends/.

National Association for College Admission Counseling, "State of College Admission 2004-05," en National Association for College Admission Counseling, 2005.

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.