





# Analyzing and Predicting Colombian Undergrads Performance on Saber-Pro Test: A Data Science Approach

Eugenia Arrieta Rodríguez<sup>1</sup>, Paula María Almonacid<sup>2</sup>, Santiago Cortés<sup>3</sup>, Rafael Deaguas<sup>4</sup>, Nohora Diaz<sup>5</sup>, and Maria Paula Aroca<sup>6</sup>

<sup>1</sup> Universidad del Sinú Cartagena Elias Bechara Zainum, Cartagena, Colombia  
[investigacionsistemas@unisinucartagena.edu.co](mailto:investigacionsistemas@unisinucartagena.edu.co)

<sup>2</sup> Universidad EAFIT, Medellín, Colombia  
[fpalmona1@eafit.edu.co](mailto:fpalmona1@eafit.edu.co)

<sup>3</sup> Universidad Antonio Narino, Bogotá, Colombia  
[santiago.cortes@uan.edu.co](mailto:santiago.cortes@uan.edu.co)

<sup>4</sup> KNAUF, Bogotá, Colombia  
[Aguasr@knauf.com](mailto:Aguasr@knauf.com)

<sup>5</sup> iFood, Bogotá, Colombia  
[nohora.diaz@ifood.com.co](mailto:nohora.diaz@ifood.com.co)

<sup>6</sup> Data Science IU Bloomington, Bloomington, USA  
<http://www.unisinucartagena.edu.co>,

<https://www.eafit.edu.co/docentes-investigadores/Paginas/paula-almonacid.aspx>, <https://www.uan.edu.co>, <http://www.knauf.com.co>,  
<http://www.ifood.com.co>, <https://online-distance.ncsu.edu>

**Abstract.** In this paper, we present an analytic tool solution for universities and the Colombian government, which allows them to identify the main factors that impact the performance of undergraduate students and also to predict the performance of the future cohorts of students. This solution consists of an interactive dashboard that visualizes two different types of analysis: in the first place, the descriptive statistics of variables related to the different entities implicated in the education process such as students, universities, and the state governments; and in the second place, the solution helps to visualize the results of different predictive models.

**Keywords:** Saber-Pro test · Spatial models · Machine learning

## 1 Introduction

[8] highlights that few governments have been able to simultaneously increase coverage and quality in education. Following evidence suggests that the rapid expansion of professional programs in Colombia deteriorated their quality [5].

---

Supported by ZettaNet Digital Transformation and Correlation One x.

© Springer Nature Switzerland AG 2021

K. Saeed and J. Dvorský (Eds.): CISIM 2021, LNCS 12883, pp. 81–99, 2021.

[https://doi.org/10.1007/978-3-030-84340-3\\_7](https://doi.org/10.1007/978-3-030-84340-3_7)

This could be argued, since the standard state exams, such as the Saber tests, allow an evaluation of the progress in the quality of education and permit documenting socioeconomic and spatial differentials [2,3].

In Colombia, the Saber-Pro standardized test was developed to evaluate the quality of higher education. It assesses math, reading, citizenship, writing and English skills amongst undergraduates at every higher education institution. The scores of this test become publicly available each year, with the idea that institutions can identify vulnerable populations and adjust their education policies. But raw data alone is not enough for institutions to gain insights.

This problem is exacerbated for lower-income institutions that don't have the resources to hire teams to analyze these raw scores. Academic performance is a fundamental pillar for achieving the long-awaited educational quality. The Saber tests not only allow the evaluation of quality in education but also offer the possibility of adjusting educational policy in both public and private institutions.

Therefore, we built an interactive dashboard that allows universities and other institutions to easily explore which factors matter the most to student performance on the Saber-Pro. In addition, our application lets institutions visualize the results of predictive models to explore how changing some of these factors could affect scores in future exams. With our application, institutions will easily gain insights into which student populations would benefit from targeted interventions.

We also expect to contribute because we are linking the same students who took the standard state tests at the end of high school and undergraduate education.

Some studies mention the importance of reading comprehension and its impact on performance in the Saber Pro tests. The objective of this study was to correlate the total and reading comprehension scores in the Saber Pro tests (2011-I) versus the scores obtained in tests that evaluate working memory, verbal intelligence and general intelligence. This work shows that working and general intelligence can be used as estimators of performance in the Saber Pro tests. Another work in which the efficiency of education programs is evaluated shows the importance of environmental factors, suggesting that, although many educational institutions have a margin to improve their levels of efficiency, they could be restricted by the influence of environmental factors. Socioeconomic status of the students [7].

Moreover, we mention this work that applies the CRISP-DM data mining methodology for the construction of 3 analytical models to study the results obtained in the Saber-Pro tests of engineering students in Antioquia (Colombia). As a result of this work, it is indicated that the most relevant variables are: the number of dependents, teaching method, whether the home is permanent, the academic nature of the institution and economic facilities such as having a micro gas oven and a motorcycle [4].

Finally, mention is made of the work carried out in 2014 by researchers from the Universidad del Rosario, who propose a methodological model for the development of a scientific study within the School of Administration of the Universidad del Rosario that allows developing strategies for improvement in admission policies and pedagogical structuring of the Business Administration [6].

After a literature review and consultation with experts, Colombia seems to be the only country in the world where it is possible, making Colombia a leader in the area of education quality evaluation.

Additionally, the spatiality of the data is taken into account. According to [5] the fact of not considering it could lead to wrong conclusions of regression analysis, and specification problems emerge in the models when the spatial dependence is present in the data. By means of the SAR model, we explore the spatial dependence of the phenomena of study. After this brief introduction, we present the structure of the document: in the first place, we describe the different characteristics of the solution for each of the application panels. After that, we present the sources of information, then the process of data wrangling, finally, the architecture and the final thoughts and conclusions.

## 2 Materials and Methods

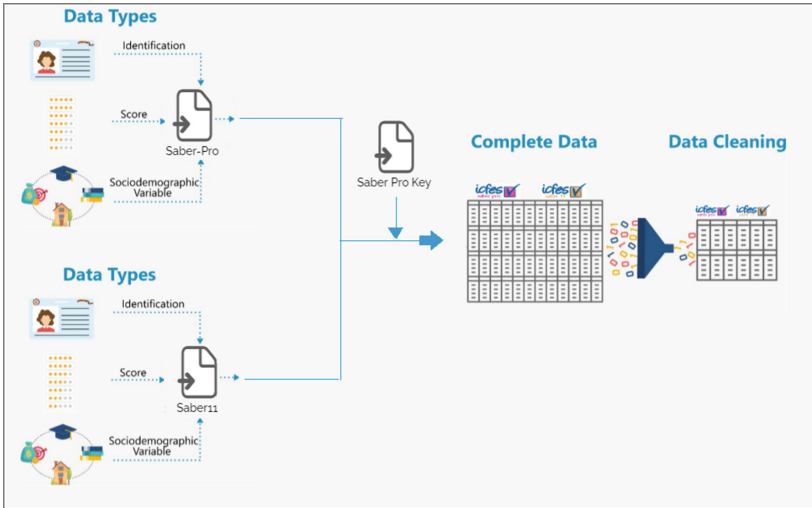
The steps followed to develop this research basically consisted of obtaining the necessary databases SaberPro and Saber11, as well as in the proper pre-processing of the data, the analysis of the relevant variables, and their modeling and prediction. For this, the standard analysis procedure was followed, dividing the database into a training part and a validation part and comparing the prediction precision metrics.

### 2.1 DataSet Construction

We explored three main public datasets, gathered from [datos.gov.co](https://datos.gov.co) [1]:

- Saber-Pro dataset: this is a student-level dataset of the higher education level students, with 986k+rows per year and 132 variables, which contain the student's test scores from 2016 to 2019, as well as a variety of variables related to socio-demographic.
- Saber11 dataset: this is a student-level dataset of high school education level students, with 91 variables containing student's test scores, as well as a variety of variables with socio-demographic, financial, academic and, geographic information for each student.
- Saber Pro Key database: this dataset contains observations that store key that allows us to merge the Saber 11 and Saber PRO databases for students that have taken both tests.
- Additional Data: we compile data regarding department-level and municipality-level sociodemographic and economic variables from the National Administrative Department of Colombia-DANE.

The process of data wrangling and cleaning could be summarized in the following steps: Dataset Appending, Dataset Merging, and Dataset cleaning (See Fig. 1).



**Fig. 1.** Data wrangling and cleaning process

We decided to focus our solution on the Generic Test of Saber PRO. We found information available from 2006, and we decided to limit the analysis using data from 2016 to 2019. Four years of information are enough to see the trends and changes by analysis period (Show Table 1).

**Table 1.** Table captions should be placed above the tables.

Year	Rows	Columns
2016	242.629	99
2017	245.593	106
2018	237.112	106
2019	260.756	105

Additionally, we use the data that we called *Llave*, as is the data that can join the information of Saber 11 with Saber Pro. This Data contains information of the Student ID that presented the test from 2006 and has the following structure:

## 2.2 Dataset Appending

When we append the different years of information, we found some of the following processes fix:

Some years had the information in lowercase, not all. We decided to leave them in lowercase before appending the data. Some variables as the Student key, does not have the same name. We rename the ones necessary to have just one column for the Student Key. Because of the number of rows Saber 11, we decided to first merge the data with the key before appending it.

After Appending the four years of Saber PRO we got a dataset of 986.090 and 132 rows.

## 2.3 Dataset Merging

Using the data *Llave* was possible to merge Saber PRO and Saber 11 (Show Fig. 2). Having a final data of 1.045.290 rows as a same student can present different any of the two tests more than once. For example, this students gave the test more than 3 times:

EK201730077444	5
EK201830224318	4
EK201950226991	4
EK201630049777	4
EK201830102855	4

**Fig. 2.** Keys

If all the columns are included in the merge, it is obtained a dataset of 434 columns. That is why it was decided to continue with the exploratory data analysis adding just the variables considered important.

From the 1.045.290 rows related to Saber Pro, there are 414.701 of them with Saber 11. It means 40% of the data.

Some of the reasons why was not possible to obtain more information are:

- There was just possible to use the information from 2006. There could be students that present the Saber Pro exam from 2016 to 2019 and Saber 11 before 2006, that ones don't have the information available.
- Information from 2010 to 2012 in Saber 11 was not possible to merge with the data *Llave*. We contacted the owner of the data (ICFES) without getting any answer.

We will validate during future analysis that the data available is still enough to find some patterns that help us to suggest possible actions that the stakeholders interested in this project can do.

### 2.4 Dataset Cleaning

In this step, the data is cleaned before the analysis. The first thing we did was verify whether or not our variables contained null values.

Figure 3 below shows an example of academic variables with their number of null values for each of them. A large amount of missing values in some of them led us to removing most of the variables, since they were missing in over 80% of the test-takers. However, we made an effort to find other variables that approximated these. For example, we had a large number of missing values in *estu\_prestigioinstitucion*, *estu\_instporcostomatricula* and *inst\_porprestigio*. These are variables that measure the tuition level at the university and the university ranking. Thankfully we have a student-level variable that measures how much they pay for tuition, and that is correlated to the variables we had to remove. We followed a similar process for all variables in each of the variable categories.

estu_consecutivo_x	242629
mod_razona_cuantitat_punt	29
mod_lectura_critica_punt	29
mod_competen_ciudada_punt	29
mod_ingles_punt	105
mod_comuni_escrita_punt	11039
punt_global_x	242657
estu_nucleo_pregrado	0
inst_cod_institucion	28
inst_nombre_institucion	0
inst_caracter_academico	0
inst_origen	0
estu_razoninstituto	981752
estu_prestigioinstitucion	1017347
estu_instporprestigio	847988
estu_instporcostomatricula	847988
estu_instporunicaqueofrece	847988
estu_instporoportunidades	847988
estu_instporamigosestudiando	847988
estu_instporotrarazon	847988
estu_cole_termino	66832
cole_valor_pension	884439
cole_nombre_sede	687502
cole_calendario	687502
cole_genero	687502
cole_naturaleza	687502
cole_bilingue	687873
cole_jornada	687502
cole_caracter	687502

Fig. 3. Null values in academic variables

Additionally, we had to clean a large number of variables. For example, many of our categorical variables were duplicated because the values had accents, trailing spaces, numbers and other symbols. Misnamed and duplicate categories were identified and replaced by using string replacement operations. See Fig. 4.

Finally, some of our categorical variables had large numbers of categories, so we created new variables where we renamed these into a smaller number

of categories. For example, a variable identifying the undergrad major of the student with 58 categories, was re-grouped into a new “school” variable with 11 categories.

	estu_consecutivo_11	estu_consecutivo_PRO
0	SABER1120082416622	EK201210000523
1	SABER1120082072880	EK201210000528
2	SABER1120062195960	EK201210000532
3	SABER1120072075157	EK201210000533
4	SB11201020158376	EK201210000535

**Fig. 4.** Key Join

Even when the key has different structures per year, as it is seen in the registers 1 and 4, this data makes it possible to join the information available.

As the data *Llave* is available from 2006, We decided to use Saber 11 from 2006 to 2019. Each year has more than 400.000 rows and from 53 to 150 columns depending on the year. There are two different data sets per year, which is a total of 28 datasets for this period of analysis.

Finally, we handle missing values and exclude variables that did not aggregate value to our models and create new variables such as the age, the time Letter elapses between the year in which a high school student presented the Saber11 test and the year the student presented the Saber Pro test, and if the department where the student lives is the department where the student study or presented the test. It is essential to mention that we found a large proportion of missing values for the year 2016 in the Saber Pro data, therefore we decided to exclude from our sample of interest.

In this sense, we consolidated one database with the following clean and preprocessed CSV,

- Saber11
- Saber-Pro
- Basic Education Indicators
- State and city locations

### 3 Exploratory Data Analysis (EDA)

As we are dealing with a large amount of information about socioeconomic and academics aspects of Colombian undergraduate students (as well as their corresponding SABER, and SABER PRO, test results, and this for different years and from all Colombian regions), we distributed the different categories of variables between each of the members of the team. Each team member was responsible to break down the data into summarized CSV files according to its corresponding category; manipulate the data to create more subcategories from the existing columns; find the biggest players in the different subcategories. And find the main trends in the data. See Fig. 5.

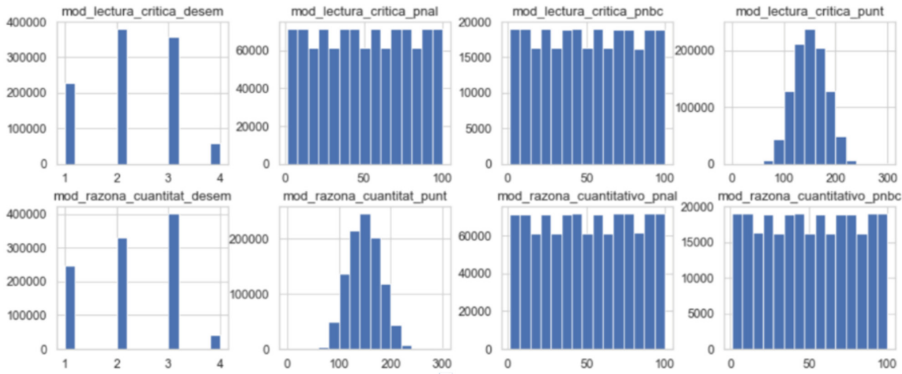


Fig. 5. Exploration features

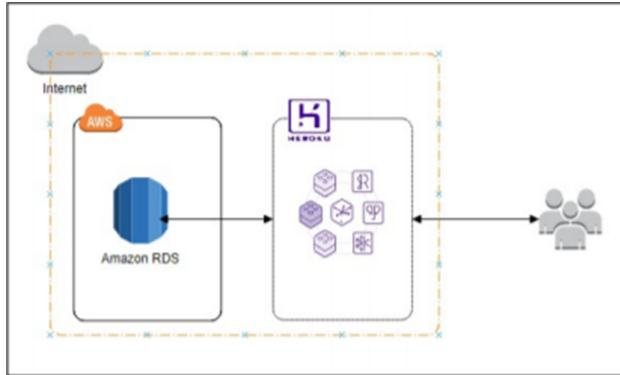
## 4 Construction Dashboard

We propose an innovative application that allows different agents who participate in the education process to easily and quickly access various statistics related to students’ performance, universities, academic programs, and departments (states). This information enables the analysis and prediction of variables of interest.

### 4.1 Architecture

Figure 6 presents the architecture of the proposed solution. Specifically, it shows the main elements used for the Front and Back-End, the application, and its connections. Additionally, we named the technologies used hosted on AWS cloud.

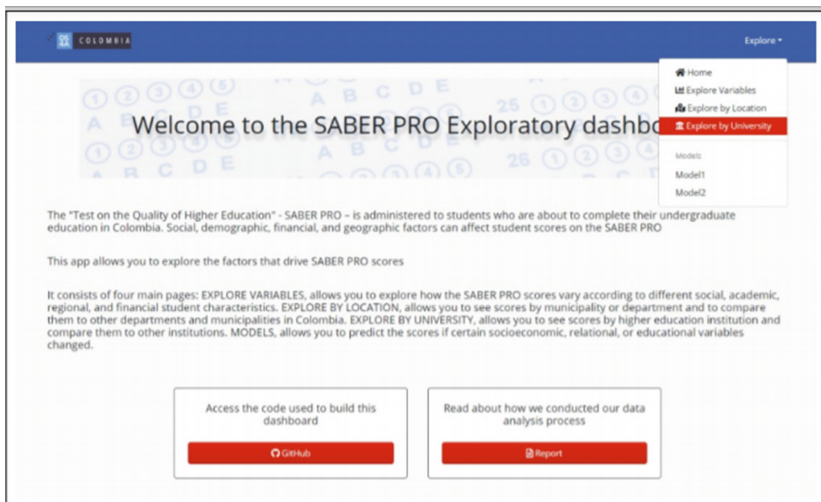




**Fig. 6.** The architecture of software

## 4.2 General Panel

In the initial panel, we find a suggestive and organized page that allows access to panel possibilities such as the panel related to the University, the panel related to geographic location, and a panel related to the models and predictions of results for the panel students. This panel also includes links to access to the GitHub and documentation of the Project. This application is open to the public to be complemented and improved in favor of developing the education sector in the country. See Fig. 7.



**Fig. 7.** General panel layout

### 4.3 University Panel

In the Universities panel, it is possible to compare student performance in the different competencies that are evaluated in the Saber Pro tests. Universities or academic programs can make such comparisons. All this is done through interactive graphics with interesting designs. View Fig. 8.

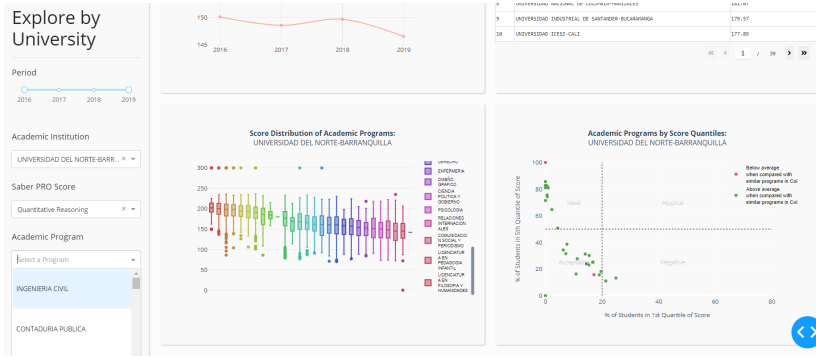


Fig. 8. University panel layout 1

These comparisons allow us to establish which universities and programs should include improvements in their academic processes. When identified, they could be the center for government aid and programs to improve the education sector for the entire country. View Fig. 9.



Fig. 9. University panel layout 2

#### 4.4 Location Panel

Generating insights by location will allow that each state will be able to make their own decisions and to understand their context in order to improve their social and economic goals (Fig. 10).

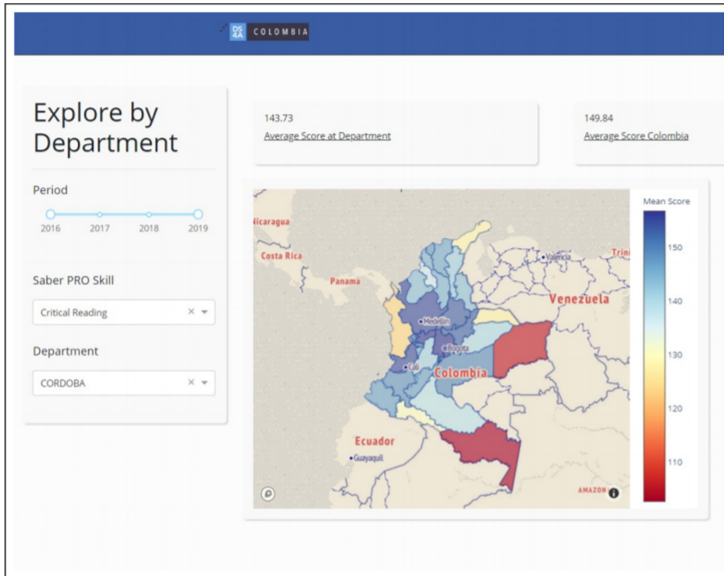


Fig. 10. Location panel layout

#### 4.5 Predict Scores Panel

The objective of this panel is to provide students who aspire to take the Saber-Pro tests, the possibility of obtaining a prediction about their future test scores with precision which could vary in a range between 88.88% and 93.6%. For this, the student must complete the information required on the page, which has to do with the variables identified by the model as the main factors that determine student performance on the tests. Show Fig. 11.

Fig. 11. Predict Scores panel layout

### 4.6 Which Factors Matter

Students will be able to observe the impact of their characteristic variables on their score value of the tests. In this way, the student would have the chance to address her weaknesses and eventually could her education quality (Fig. 12).

Fig. 12. Which factors matter panel layout, first part

## 5 Results

To attain the main factors that explain the undergrad's performance on the Saber-Pro tests, we employ different model approaches such as Linear Regression, Random Forest, XG Boosting, and spatial models.

We began exploring the most relevant variables that explain the student's academic performance with linear regression, a very simple approach for supervised learning. We found linear regression a useful tool for predicting our quantitative response variable, the test scores in the Saber-Pro tests, based on some important explaining variables. This model helped us as a tool for feature selection and as a first approximation to variable relationships. Our first objective is to identify evidence of an association between the students' test scores in the Saber-Pro and variables, in which the literature has not been conclusive, such as gender, the parent's education, the social strata, or the fee of tuition, which could be a proxy of the quality of the University.

According to our preliminary results, we found evidence of the association between the students test scores and, in order of importance, their scores on critical Reading (a positive relationship), the gender (men tend to have higher scores than women), the parents' education (the higher the level of education of the parents the higher the scores obtained by the students), the tuition fee of the University (the higher the tuition the higher the scores), the social strata (a positive relationship), and the socio-political region (the central region of the country, especially the capital, Bogotá tend to present higher students scores than the students from other regions of the country).

Concerning the spatial factors affecting the students' performance, we carry out a geospatial clustering analysis taking as units of the spatial analysis the departments of the country. Those are 32 areas in which the government is administrative and politically divided. We selected them as a unit of study because, in general terms, these are a good proxy of the slightly cultural and environmental differences that are present in Colombia.

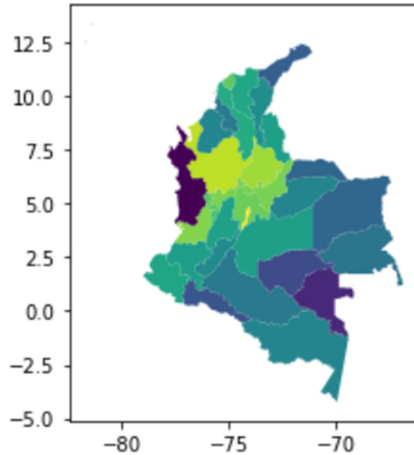
This analysis aims to find similarities in groups or "clusters" in terms of spatial and non-spatial variables, specifically related to the educative level of the high school graduates. We are also interested in detecting spatial patrons regarding education levels in the country, allowing for beneficial externalities between departments. Among the spatial variables included in the geospatial clustering process are the latitude, longitude, and department shape. On the other hand, regarding the education variables, we analyze the number of the students, the student's results on Saber-Pro tests in the subjects of Quantitative thinking, English, Critical Reading, Citizenship competencies, and finally, the global scores.

Figure 13 we present some of the results of the spatial clusterization of the Saber-Pro tests by the different departments of the country. We represent the clusters by the different shades and colors. As the lighter colors reflect better results, it can easily identify the clusters of departments where the students' performance was better. Among these are the departments of the central region such as Cundinamarca, Antioquia, Santander, Valle del Cauca, Boyacá, Caldas,

Risaralda, and Armenia. After these departments, we can find the departments situated in the coastal region and the south of the country according to the level of performance. Finally, it is important to mention that the departments with deficient performance are Chocó, Vaupés, Guanía, Putumayo, Vichada y Guajira, which are departments located in the periphery region of the country. It is also worth highlighting that these patterns found coincided with the test results related to critical reading, quantitative reasoning, English, and citizenship skills.

Analyzing the results obtained through this geospatial clustering, the following can be concluded:

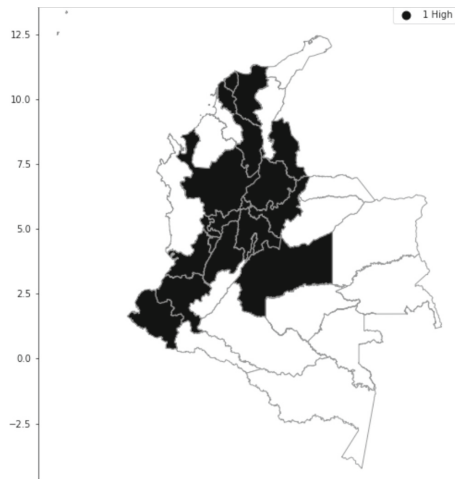
- 1) Colombia is divided into geographic regions with different levels of quality in education.
- 2) The central or Andean region and some departments of the Atlantic coast and the south of the country present a higher level of education quality.
- 3) The geographic areas of the country that present higher levels of education also show higher levels of development, economic growth, and presence of the state.
- 4) Through the spatial and educational structure identified thanks to spatial clustering, it is possible to locate neighboring regions for each department, which could result in beneficial externalities between departments. Considering that the activities carried out in a particular department may affect the decisions made in neighboring departments.



**Fig. 13.** Global score results of Saber-Pro test by departments. Similar colors represent the clusters of the departments. Lighter-colored clusters represent better scores in the overall results of Saber-Pro tests.

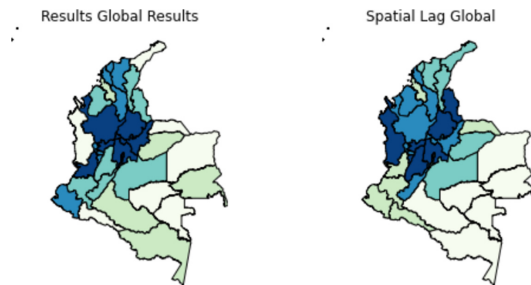
Visual inspection of the map pattern for the test results allows us to search for spatial structure. If the spatial distribution of the tests was random, then we

should not see any clustering of similar values on the map. However, our visual system is drawn the lighter hues (higher performance on the tests). View Fig. 14.



**Fig. 14.** Departments that perform above the average of the global score on the Saber-Pro tests

This visualization tool is not just reserved for the performance of the students on the tests. We can also visualize the distribution all over the country of other important education-related variables such as the rate of the approval, the rate of drop-out, the number of students, etc. Therefore, this type of tool allows central and state government entities to focus and prioritize their development plans towards the most disadvantaged departments and the departments that could most impact the development of their neighbors through externalities and reciprocities. See Fig. 15.



**Fig. 15.** Neighborhood or peer departments. A neighboring department can have direct and indirect effects on its neighbors or peers, impacting its development through various externalities and/or multiplier effects

In this sense, our solution contributes to effectively identify which department represents a neighbor or peer in terms of its similarity in the quality of education found and measured by the results of the Saber tests. In such a way, it is possible for government entities to generate impacts or externalities that positively influence the different clusters and contribute to improving the quality of education throughout the country in general. Finally, but not least, we present some of the results of the Random Forest Regression Model, in Fig. 16. According to these, the variables that impact the performance of the undergrads on the Saber-Pro test are those related to the background and the abilities of the individual on academic competencies, evaluated in Saber11 tests.

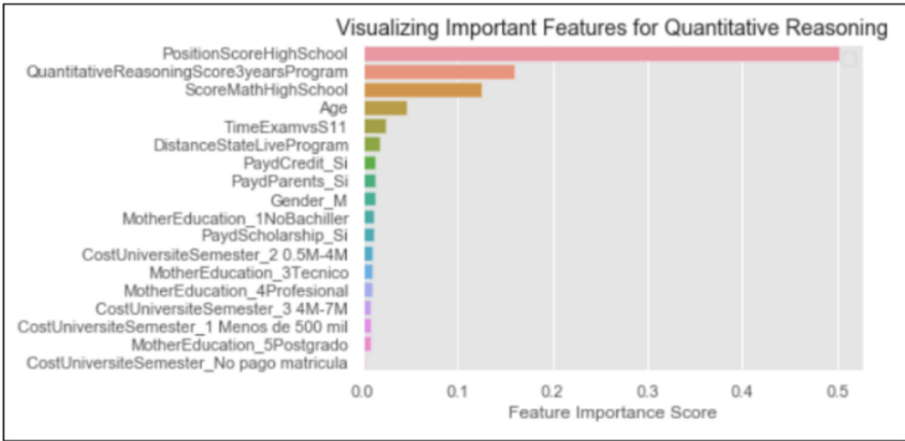


Fig. 16. Importance of the variables included in the Quantitative Reasoning Models

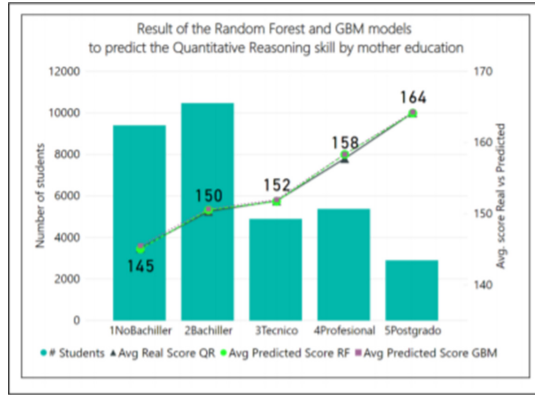
Other variables such as age, mother’s education, the cost of the University, as a proxy for the quality of the education provided by this entity; the interval of time elapsed for a student between the presentation of the exam Saber 11 and Saber-Pro; the cluster of the states (or departments as are known in Colombia), the distance between the place of study and the residence and how the studies were paid (for instance, if the studies were paid by credit, by their parents, or by a scholarship), have a significant impact on the performance of students on the Saber-Pro tests.

On the other hand, it is important to highlight the great predictability of both the Random Forest and spatial models for our case. These models achieved prediction precision in a range between 88% and 93.6%. After tuning the parameters, using cross-validation that helped to reduce the chances of overfitting and keep similar and good results between the train and the test dataset.

To implement an additional validation procedure, we compare the results of the XGBM and the Random Forest for each of the variables. By performing this

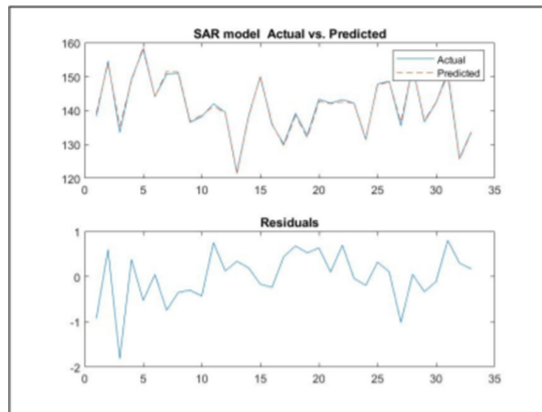


process, we evaluated the performance of the model on the whole distribution of whole variables. In Fig. 17 can compare the behavior of actual observations versus the prediction score for quantitative reasoning tests results.



**Fig. 17.** Comparison of the performance of the two models, Random Forest Regression and Gradient Boosting Machine by mother education

Therefore students, universities, and government entities could get a very good idea of the future performance of potential higher education students if we maintain similar conditions. In the final solution, it has been included the results of the Spatial model and the Random Forest Regression Model results related to the Global score and the five skills evaluated (Fig. 18).



**Fig. 18.** This figure presents the prediction performance of the spatial model versus the real observations.

## 6 Conclusions

The objective of this research aims to contribute to the development of the strategy designed by the National Ministry to improve the quality of education. In order that all children have the same opportunities to acquire knowledge, develop the skills and values necessary to live, live together, be productive and continue learning throughout life.

It is important to assess the potential and qualities of educational evaluation to improve educational processes at a general and private level to achieve beneficially goals. We expect to contribute considering the fact that our educational quality evaluation solution allows us to establish education, by means of holistic perspective of the phenomena of education, by means of the use of two tests (the standard state tests at the end of high school, Saber11, and undergraduate education, Saber-Pro, together).

Identifying the characteristics with the greatest impact on test results could serve universities and government entities to formulate policies that reinforce positive impact characteristics and improve those with negative impact. This tool would serve as an early warning for effective policy design.

The most relevant variables associated with the undergrads' performance on the Saber-Pro test are those related to the background and the abilities of the individual on academic skills, especially those related to quantitative reasoning. The other variables that have more impact on the students' performance on the Saber-Pro tests, according to our models are: age, mother's education, the cost of the University (as a proxy for the quality of the education provided by this entity), the interval of time elapsed for a student between the presentation of the exam Saber11 and sober-pro, the cluster of the state, the distance between the place of study and the residence, and finally the way in which students paid the studies (for instance, if the studies were paid by credit, by their parents, or by a scholarship).

Additionally, the spatiality of the data is taking into account. The fact of not considering it could lead to wrong conclusions of regression analysis, and specification problems. By the means of the SAR model we explore the spatial dependence of the phenomena under study. According to the results obtained by the spatial model, the quality of education in Colombia presents different levels according to geographic location. In general, the performance of higher education students measured by the Saber-Pro tests is superior in the central region and part of the north of the country. In contrast, the periphery regions present low levels of performance on average. On the other hand, the spatial model allows us to identify who would be the neighbors or peers whose development could impact through indirect effects to departments that need to reinforce the quality of their higher education.

## References

1. Acceso a bases de datos y diccionarios - portal icfes. <https://www.icfes.gov.co/investigadores-y-estudiantes-posgrado/acceso-a-bases-de-datos>
2. Barrera Osorio, F., Maldonado, D., Rodríguez Orgales, C.: Calidad de la educación básica y media en Colombia: diagnóstico y propuestas, Technical report, Universidad del Rosario (2012)
3. Camacho, A., Messina, J., Uribe Barrera, J.: The expansion of higher education in Colombia: bad students or bad programs? Documento CEDE (2017–13) (2017)
4. Carrascal, A.I.O., Giraldo, J.J.: Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas saber-pro. *Revista Politécnica* **15**(29), 128–140 (2019)
5. Castro-Ávila, M., Ruiz-Linares, J., Guzmán-Patiño, F.: Cruce de las pruebas nacionales saber 11 y saber pro en antioquia, Colombia: una aproximación desde la regresión geográficamente ponderada (gwr). *Revista Colombiana de Educación* **74**, 63–79 (2018)
6. Garzón González, M.A., Vega Bonilla, S.: Modelo estadístico de correlación de resultados examen Saber 11 (ICFES) y Saber Pro (ECAES). Ph.D. thesis, Universidad del Rosario (2014)
7. Melo-Becerra, L.A., Ramos-Forero, J.E., Hernández-Santamaría, P.O.: La educación superior en Colombia: situación actual y análisis de eficiencia. *Revista Desarrollo y sociedad* **78**, 59–111 (2017)
8. Saavedra, J.E.: Resource constraints and educational attainment in developing countries: Colombia 1945–2005. *J. Dev. Econ.* **99**(1), 80–91 (2012)