

## Exploration of a Classroom Management Model for Decision-Making in Higher Education (first part)

Leonardo Emiro Contreras Bravo<sup>1</sup>, José Ignacio Rodríguez Molano<sup>2</sup>, Héctor Javier Fuentes López<sup>3</sup>

*Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.*

### Abstract

A current problem in Colombian universities is the generation of large volumes of data derived from the iterative learning process of students and teachers. Said data is stored and rarely used with the purpose of improving the academic performance of the student or give academic counseling. Afterwards, an UML-based model is proposed as well as a prototype of the academic performance prediction system in Python, focused in two aspects. The first one consists on giving clarity on the state of the educational process in the Distrital University (case study) and the second one is an attempt to develop the way in which the prediction of the academic performance would occur based on Machine Learning tools that would allow the system (data) to learn and predict results.

**Keywords:** Student performance, Educational data mining, Analysis of data, Big data model, Engineering, Education.

### INTRODUCTION

One of the current trends in universities is to manage ways of improving the learning process of students through technological tools available online and LMS systems. These have paved the way for large volumes of data that are generated by the iterative work of students, teachers and institutions. The data is stored and rarely used to improve the academic performance of students, the academic counseling that guides them during their studies and also reduces the desertion rate, among other actions that can be carried out based in data information [1].

Initially, it is important to differentiate a complex system from a complicated system since it is common to confuse both terms. There are philosophers such as [2] that define the complexity as the impossibility to simplify something, making it nearly impossible to give clarity to the identities and their causalities, where disorders and uncertainties cloud the phenomena. Complexity is understood nowadays as the impossibility to consider particular aspects of a phenomenon or situation based on a specific discipline by handling data or information as stated in [3], summarizing several authors. The phenomena that take place in the real world cannot be solved from a single perspective or discipline, nor categorized under one label. They are a group of components that can interact between them and require the interdisciplinarity of several sciences. In contrast, a complex system needs an approach from different points of view over time and developing an investigative process. These sciences or disciplines are integrated as needed in said process. This concept of interdisciplinarity differs from the concept of

multidisciplinarity by the fact that the second one requires an expert per discipline, to solve the problem with their own theoretical contribution. The first concept (interdisciplinarity) is understood as the integration of different approaches of the problem from several disciplines in order to enclose a problem. This means that studying a complex system requires interdisciplinarity.

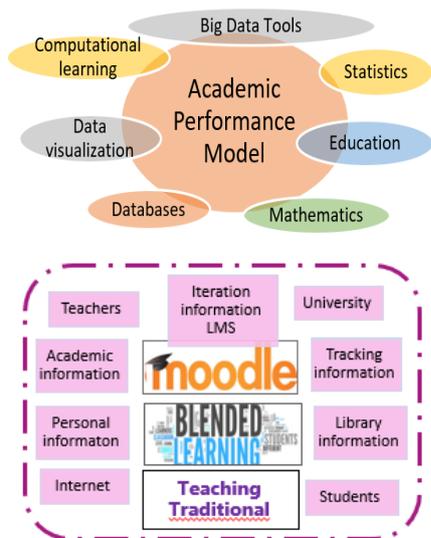
One of the most important factors in the efficiency of the educational process is the retention parameter, defined as the difference between the number of students that being studying in the first semester and the graduates per year [4]. Hence, the academic performance is the main indicator of the student's success or failure as well as of the multivariate factors that affect his possibility of dropping out of a higher education program [5]. Therefore, this determination has led to controversy since there is no definitive theory on the methodology used to measure or assess said dropping rate. Being multidimensional, the academic performance depends on multiple aspects such as the objectives of the teacher, the institution, the student, etc. [6].

Data analytics is an emerging field, and in Colombian education, there is no culture of using and analyzing student-generated data in the learning processes to determine his influence on academic performance, desertion rate and graduation.

### COMPLEXITY OF THE ACADEMIC PERFORMANCE MANAGEMENT SYSTEM

#### Regarding the disciplines

To approach the problem with a possible solution, an integral overview is required which translates into a model that integrates a series of data from different agents (students, teachers, university, etc.) to predict academic aspects such as performance through data analysis tools that can nurture decision-making in higher education. The solution must come from an integral vision from different fields or disciplines, which includes Computer Engineering, Mathematics, Computational Learning, Big Data tools, visualization tools, databases and education tools (Figure 1). Hence, it is not a straight line that needs to be followed to solve the problem of academic performance. It is an axiological path with positive and negative values.



**Figure 1.** (Left) Interdisciplinarity focus, (Right) Agents of the complex system

### Regarding data

The analysis and prediction system requires input data from more than one source related to the academic system of the Universidad Distrital – Colombia. The data should be acquired through the interaction of the student (data producer) with different activities established in the Moodle platform (LMS). Some examples of this activity include the participation in Table 1. Hence, although most of the variables serving as inputs of the system are quantitative and only a few are qualitative (activity in social networks, for instance), the student should not become a being whose behavior can be measured numerically. On the contrary, the model seeks the interpretation of numerical results (change from quantitative to qualitative) in order to establish a plan that can identify other qualities of the student, as well as attempting to improve his academic actions

exercises, writing a message in the forum, reading a document, registering a click on the mouse and the time spent by a student in answering a question. Raw data is required from the platforms used in universities related to admission processes as well as open data from state education systems to feed the academic analytics. This could lead to affirm that the data is the sap that permeates the entire complex system of this work. This is an analogy to the complex system of plants, since it contributes to new functions as it moves to a superior level. One data does not generate knowledge to take actions while a dataset can lead to taking different actions (Figure 1 Right side).

### Regarding the variables

This system that needs to be designed involves interrelated variables to try to predict the academic performance and/or student desertion as well as other aspects in academic life. The following are factors that could be grouped into different levels such as: pre-university academic factors, pre-university demographic factors, pre-university socio-cultural factors, pre-university socio-economic factors, university academic management factors, technological factors (Learning Analytics, B-learning), library factors, institutional factors, pedagogic factors, intellectual and affective factors. Some of the variables that the model could require are shown in

through suggestions during his university years. This integration of data and technologies cannot turn into a definitive prediction of the student since it will only inform about the probability of occurrence of any success (failing a course, year of culmination of studies, etc.) and will not give a final verdict on how the academic life of a student will be in a specific program.

**Table 1.** Variables affecting academic performance. Source: Adapted from [7], [5], [8] and [9]

Pre- University factors	Academic factors	Global score at the ICFES test (Colombia), Year of presenting the ICFES test (<2014 o ≥2014), ICFES score in Mathematics, ICFES score in Physics, ICFES score in Chemistry, ICFES score in Biology, ICFES score in Social Sciences, ICFES score in Philosophy, ICFES score in Natural Sciences, ICFES score in Critical reading, ICFES score in Citizenship, ICFES score in English.
	Demographic factors	Age, gender, civil state, type of high school where the student graduated from, work experience, time elapsed before entering university, high school degree
	Socio-cultural factors	Educational level of the parents, number of family members, occupation of the parents.
	Socio-economic factors	Place of origin, region of origin, family income, number of cars, whether the house is owned or rented, type of inscription, strata
U ni	Academic management	Scholarships obtained, syllabus, number of credits, number of credits attended, number of credits approved, final grade for each assignment,

factors	number of courses attended, number of courses failed, number of times that the student failed a course, number of repeated courses, number of cancelled courses.
Technological factors	Possibilities of internet access, number of accesses to LMS, number of accessed resources, number of downloaded documents, time per activity, x and y coordinates of the mouse, clicking time in each activity, number of clicks, time spent by the student in solving a question, participation in forums
Library-related factors	Frequency of access, types of search areas, number of books in the library
Institutional factors	Size of the classroom, organization and planning of the staff, schedule of the courses, type and size of the institution, operational process of the institution, infrastructure, policies
Pedagogic factors	Number of students per teacher, number of enrolled students, number of repeating students, didactic methods and materials used, motivation of the students, time spent by teachers in the preparation of their classes, academic formation of the teacher, assessment of the teacher, type of teacher, academic area
Intellectual and affective factors	Concentration, memory, verbal comprehension, reasoning, verbal fluidity, integration, father-son relationship, teacher-student relationship

### Regarding academic performance

Regarding academic performance, the complexity of the system can also be validated from semiotic and ontological fields since some concepts could be misinterpreted by some readers. Some authors have summarized some performance definitions, as the case studied in [10], who conceptualizes the academic performance as the sum of different and complex factors that have an impact on the person who is learning. [11] states that academic performance is a result of the learning process, stirred by the pedagogic intervention of the teacher and produced by the student; and [12] defined academic performance as a series of factors that revolve around the final results of the effort made by the student. In this case, the considered definition of academic performance is given by [7] who sees it as the main indicator of the success or failure of the student, which has led to becoming one of the pillars of analysis of the results of the learning process. In higher education, the performance of the students is one of the main indicators of educational efficiency and quality that represents the institution.

Another complex aspect of academic performance lies in how it is measured. Therefore, it can be interpreted as only the arithmetic result of approving a course [13], to which [14] points out that grades are a measure of the results of teaching, but not strictly to its quality. Another metric would involve objective tests as [15] states that they should include short and precise questions, without the subjective influence of the teacher. A third way to measure performance could be based on the number of approved courses since the number of approved courses per year is a more adequate indicator of student performance than the GPA [16]. Another indicator that represents success or failure of a student is the number of accumulated credits, since it indicates the progress made in the

program, i.e., it can compare the credits accumulated by the student over a certain time of study and the credits that he should have accumulated over a specific number of semesters according to the syllabus [17]. This models considers the appreciations of [17], [13] and [16].

### Regarding the variables involved in the model

Analyzing or predicting the academic performance of a student is influenced by multiple qualitative and quantitative variables of different nature (related to demographic, economic, academic, sociocultural, institutional, pedagogic and social network factors, etc.). Some variables of analysis can be controlled such as the ones encompassed in institutional and pedagogic settings, but they uncontrollable variables are also considered which include sociocultural, intellectual and demographic aspects. Furthermore, educational data is characterized by its nested structure. For instance, students in curricular projects, students in classrooms with curricular projects and curricular projects in the university, classrooms within curricular projects and also within faculties, etc.

### Complexity regarding analytics

The system to solve is unequivocally related to the field of analytics applied to education and its variants: data mining for education (EDM – Educational Data Mining), academic analytics and learning analytics, who turn educational data into useful information that can help take actions prior to teaching or activities to promote teaching and learning. Hence, the path to follow needs to be identified as well as choosing the Big Data tools required that can be applied to the analyzed data and the algorithms that give an approximate solution to the real issue

which can be validated.

## MODEL OF THE PERFORMANCE SYSTEM

This work represents an important challenge, which is the analysis of different techniques and technologies used for data analysis that require open architecture solutions capable of capturing structured and non-structured data and relevant information pertaining the context of administrative and academic systems. If the data can be integrated, an enlightening lifecycle of the student from the initial admission process, going through the learning experience up to graduation and even employment [18].

Proposing a methodology to solve the problem is as complex as the fact of determining which variables could affect the determination of the academic performance. This is the case of age variable; which according to [19] can have an impact on a differential academic performance; age implies, maturation differences or a diversity of values, attitudes, expectations and motivations. While there is contradicting work on the matter such as [7] that expresses the fact that a student older than the group average can negatively influence performance. Other variables which are not the exception and can present ambiguities in the research on academic performance are gender and region of origin. In terms of the former, the authors argue that there is no significant relationship between the academic performance and the gender of the students [19], [20], [21]. Other authors conclude that man and women have a distinct performance [22].

### Description of the model

It is proposed to integrate the data coming from the student, open data (government entities) and data from universities, in order to process them, analyze them and visualized them to support the decision-making process. **Error! Reference source not found.**The Figure 2 corresponds to a schematic view of the model. Initially, the aspects to be solved are identified. The data has to be analyzed to determine the variables that significantly affect the classroom (event of research: as an example, performance). Subsequently, different data analysis methodologies are used such as classification, grouping rules, regression and association that can be used to predict the future grades of the student and thus his/her performance, based on the previous record and the students who were enrolled before him/her. This can help students to find out how good will their performance be in a specific course even before enrolling in said course. The number of graduates per year can also be estimated for each program as well as their fields of action after finishing the program. Additionally, the teacher can identify students who are at risk, take the appropriate actions and adopt new strategies to improve the success rate.

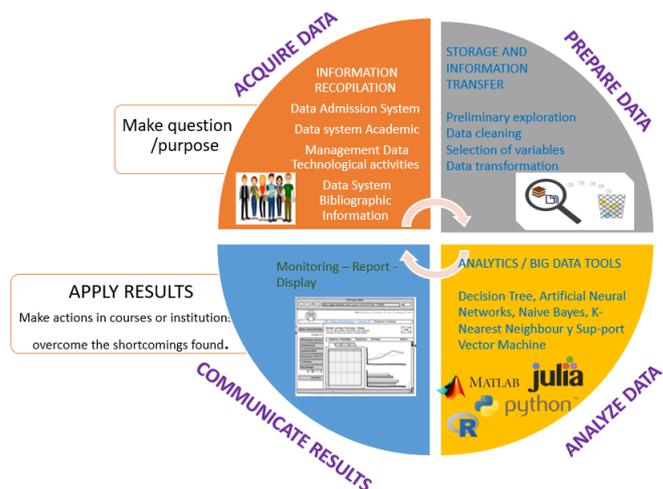


Figure 2. Academic performance model

A variety of Machine Learning algorithms (algorithms that can deliver interesting deductions based on datasets). In supervised learning algorithms, the most common task to predict the performance of students is classification (prediction of categorical-type data). There are various classification algorithms that have been applied to predict the performance of students such as Decision Trees, Artificial Neural Networks, Naïve Bayes, K-Nearest Neighbor and Support Vector Machines. Another task which is not less important to predict performance is regression (prediction of numerical data). Some of the algorithms include linear regression, multiple regression and artificial neural networks. Finally, the analytical results are presented to the end user such as online reports and recommendations that facilitate decision-making.

Nowadays, the field of software engineering has been permeated by different methodologies of application development. One of these methodologies is the Unified Modelling Language (UML) that can build software based on an object-oriented strategy. This model is initially conceived with UML developed in Enterprise Architect, where the diagrams were subdivided into groups (structured modelling diagrams, behavior diagrams and extended diagrams).

In first place, the elements are organized in packages (social data packages, simulation package and academic data package); the classes were organized within said packages to build the basic structure of the model with its attributes and methods. Figure 3 shows the class diagram for the Academic Data package. There are association, inheritance and composition relationships. Some classes are shown in a complete state defined by the UML as those that will be used further on in the development of the prototype. The *File Manager* class is in charge of reading the files and sending records for the system since the data will mostly pass in as .CSV files.

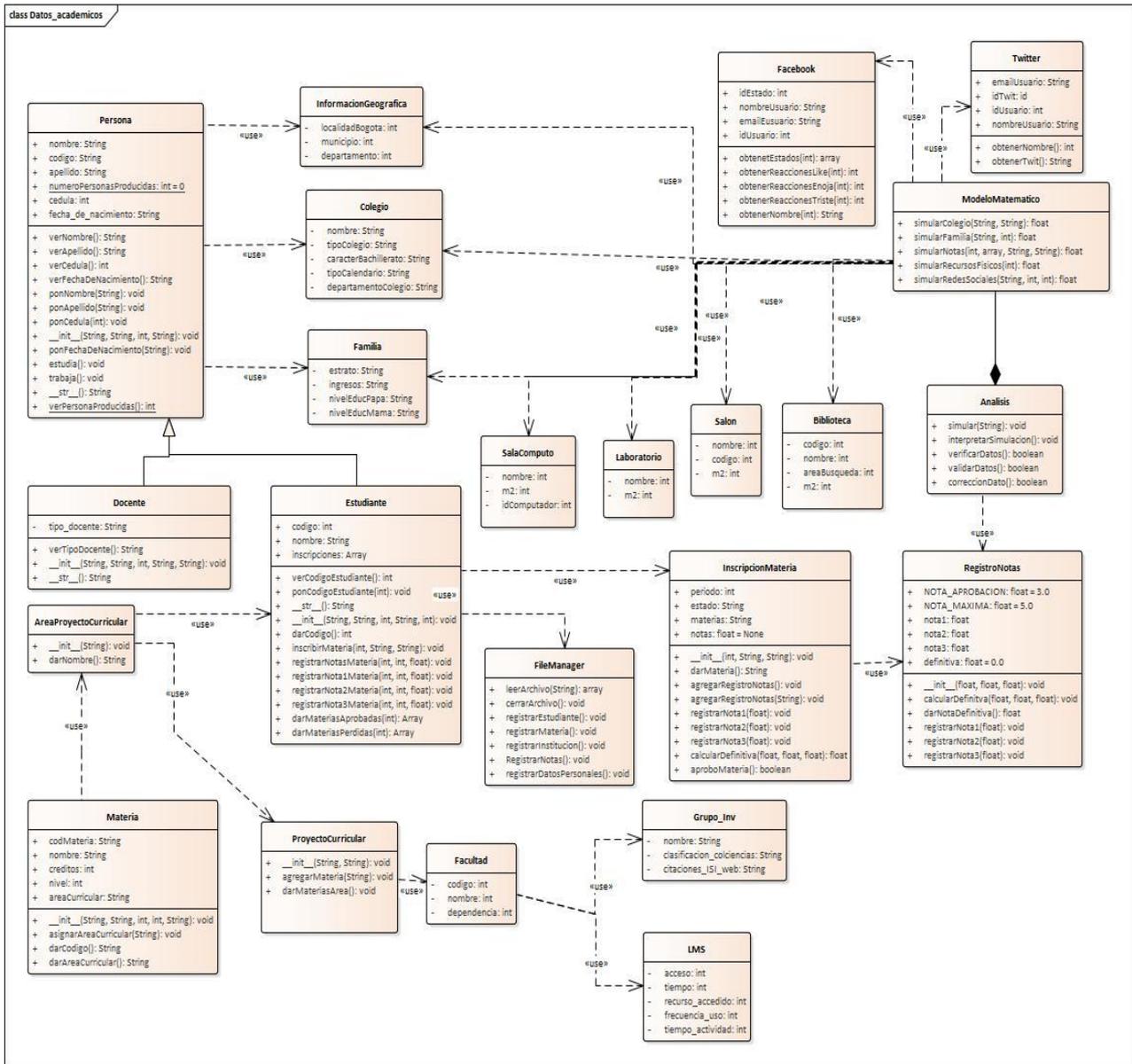


Figure 3. Class diagram for the academic data package

### DESCRIPTION OF THE PROTOTYPE

Se propone un prototipo de integración de herramientas de análisis de datos y metodologías de aprendizaje y formación, en la educación superior para la predicción de rendimiento académico. Se mostrará de una manera sencilla, pero clara como se creó el prototipo.

### Objective of the prototype

This prototype of the prediction system for academic performance focuses in two aspects. The first one consists on giving clarity to the state of the educational process inside the Distrital University (case of study). Therefore, a descriptive statistic is proposed for the gathered information. The second aspect is an attempt to develop a prediction method of the performance.

This prototype was developed in two stages. The first one involves entering the data from students and teachers by keyboard. The second one involves entering the data through .CSV files that come from the university's platforms in order to develop the academic management and prediction system. The programming language used is Python.

### Design of the academic management system

In order to progress in the code that can lead to the descriptive analysis and predictive future of academic performance based on the received input data, the *DatosAcademicos* class diagram package is used (Figure 3) which includes classes such as: *AreaProyectoCurricular* (curricular project area), *ProyectoCurricular* (curricular project), *Materia* (subject), *RegistroNotas* (record of grades), *Estudiante* (student) e *InscripcionMateria* (subject enrollment). For the

ProyectoCurricular class a constructor is designed that has the following attributes: name of the project, SNIES code (*cod\_snies*) and a dictionary called *areas* which is initialized as empty. This dictionary helps to enroll in a specific course within a certain area of the curricular project. Similar dictionaries were created for many classes (Figure 4).

```

class ProyectoCurricular:
    """Define objetos proyecto curricular"""
    def __init__(self, nombre, cod_snies):
        """Constructor de objetos proyecto curricular"""
        self.nombre = nombre
        self.cod_snies = cod_snies
        self.areas = {}
    def agregarMateria(self, materia):
        area_materia = materia.darAreaCurricular()
        if area_materia in self.areas:
            " Si el area de la materia esta en el directorio llamado areas del proyecto curricular"
            " agregar a la clave (area) del directorio el valor (materia)"
            self.areas[area_materia].append( materia )
        else:
            " como no esta creada la clave (area de la materia) en el directorio"
            nueva_materia = []
            " se crea un valor (lista) del directorio llamado areas"
            nueva_materia.append( materia )
            " A la lista creada (valor) se le agrega se crea un valor (lista) del directorio llamado areas"
            self.areas[area_materia] = nueva_materia
    def darMateriasArea(self, nombre_area ):
        return self.areas[nombre_area]
    
```

Figure 4. ProyectoCurricular class, constructor and functions in Python

**Design of the prediction process using .CSV data**

In this part of the prototype, the input data were divided into four .CSV files. For future work, it was considered that it is suitable to divide them according to the types of algorithms used. The .CSV files were named *DatosDocente.csv* (Teacher data), *DatosEstudiante.csv* (Student data), *Inscripciones.csv* (Enrollment) and *Materias.csv* (Courses). Additionally, each class previously shown was organized into different files so they could be imported into the new main file (*main.py*) as well as the Python library (*match*). When importing each .csv file, it was necessary to describe what to do with each file, i.e. tell the system to read each row detected as a register for each file.

**Testing the design of the descriptive and predictive data**

The testing data and files used in this part of the prototype correspond to the data from students, teachers, subjects, areas, grades, percentages, states, etc. from real individuals who were grouped into different .csv files that were additionally analyzed through the different classes created in each .py file as shown in Figure 5.

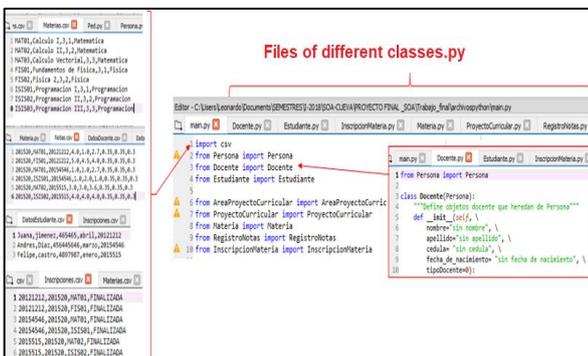


Figure 5. Testing elements for the descriptive and predictive process of the academic performance prototype

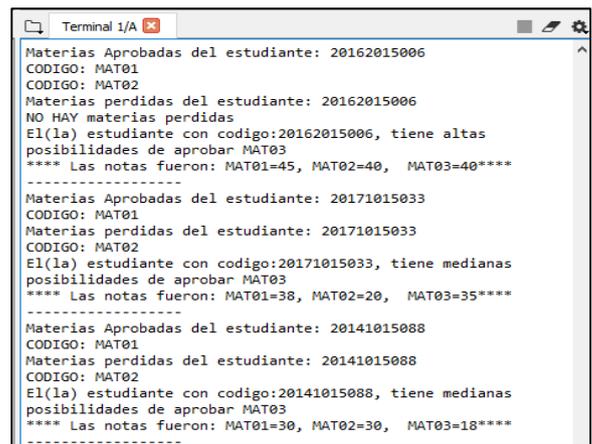
**Simulation of the descriptive and predictive data process**

In order to validate the proper operation of the prototype to guide the student who wants to know if his chances of passing a course that he has not enrolled in are high, decent or low, it was necessary to use inference rules as the ones shown here. N1 is the definitive grade of course 1, N2 is the definitive grade of course 2 and N3 is the definite grade of course 3. It is worth mentioning that this exercise is carried out for courses that belong to a specific academic area. For instance, the area of Mathematics includes the courses of Math I (N1), Math 2 (N2) and Math III (N3).

- If N1 (passed) + N2 (passed) → N3 (high possibility)
- If N1 (passed) + N2 (failed) → N3 (mid-level possibility)
- If N1 (failed) + N2 (passed) → N3 (mid-level possibility)
- If N1 (failed) + N2 (failed) → N3 (low possibility)
- If N1 [(failed) + N1 (passed)] + N2 (passed) → N3 (mid-level possibility)

**Validation**

Figure 6 shows an example of the results obtained after inputting the grades of the courses (Math I and Math II) with the purpose of receiving orientation on the possibility of passing the following course that belongs to the same area. In this case, twenty codes corresponding to real students are chosen and the system is focused on orienting. Figure 6 shows the information given by the prototype: Student 1 has high chances of passing the MAT03 course, Student 2 has average chances of passing MAT03 and Student 3 has average chances of passing MAT03. In contrast with the records of the academic history of the student, in all cases the answer is not correct. This can be possibly due to the fact that the prototype may require Machine Learning algorithms as proposed in the next phase in order to predict with more confidence the chances of passing a course (data analysis), as well as predicting the academic performance and other indicators in higher education.



Academic History			
STUDENT CODE	MAT01	MAT02	MAT03
20162015006	45	40	40
20171015033	38	20	35
20141015088	30	30	18

Figure 6. Comparison of results: Prototype in Python VS student's academic record

## CONCLUSIONS

Building a system of this nature requires a considerable amount of time since nowadays Colombian universities are limited to performing studies with descriptive statistics for short periods of time, trying to look for the reasons of low academic performance from students and their reasons for dropping out of a program. However, in search of methods and methodologies that give a more approximate solution of the problem, some machine learning tools have been identified that allow the system (data) to learn and subsequently predict results. This is a nurturing, learning and perfection-seeking process that evolves over time and is different than artisanal or simple systems.

As previously stated, another aspect to consider is that it is a complex system that considers academic performance as the main indicator of success or failure of the student. It has been widely studied both from theoretical and empirical standpoints, leading to diverse results since there is no definitive theory regarding a methodology to measure or indicate its assessment. Hence, the approach is multidimensional where the effects of numerous social, personal and school-related variables and their intertwining [17].

It is intended to design a tool or software with the main purpose of supporting the introspection of the student and improving teaching methods based on interests and personal observations. The learning analysis tool needs to offer clear and simple information that can be easily interpreted with a flexible interface for data, exploring the information as well as visualizing the results based on graphical indicators chosen individually. This process is still under construction since the next step involves analyzing the data from the Python-based machine learning algorithms in order to ease their integration into the proposed system.

## REFERENCES

- [1] L. Contreras y G. Tarazona, "Big Data: An Exploration Toward the Improve of the Academic Performance in Higher Education", *Lect. Notes Comput. Sci.*, vol. 10943, pp. 627–637, 2018.
- [2] E. Morin, *Introducción al pensamiento complejo*, Segunda ed. Barcelona, 1994.
- [3] J. Torres, *Inteligencia Integral de negocios*, Primera. Bogotá, 2017.
- [4] A. Salcedo, "Desertion in Colombian Universities", *Rev. Acad. y virtualidad*, vol. 3, n° 1, pp. 50–60, 2010.
- [5] J. Estrada, R. Quintero, y D. Vasquez, *Bajo rendimiento académico en la Universidad Distrital Francisco José de Caldas*, Primera. Bogotá: Editorial UD, 2015.
- [6] I. A. Khan y J. T. Choi, "An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction", *Int. J. Softw. Eng. Its Appl.*, vol. 8, n° 12, pp. 31–42, 2014.
- [7] D. K. García Tinisaray, "Construcción de un modelo para determinar el rendimiento académico de los estudiantes basado en learning analytics (análisis del aprendizaje), mediante el uso de técnicas multivariantes", Universidad de Sevilla, 2016.
- [8] J. Tomás, M. Expósito, y S. Sempere, "Determinantes del rendimiento académico en los estudiantes de grado. Un estudio en administración y dirección de empresas", *Rev. Investig. Educ.*, vol. 32, n° 2, pp. 379–392, 2014.
- [9] M. Oladejo, "A path-analytic study of socio-psychological variables and academic performance of distance learners in nigerian universities", Thesis, p. 197, 2010.
- [10] L. Rojas, "Revista electrónica actualidades investigativas en educación.", *Rev. electrónica Actual. Investig. en Educ.*, vol. 13, n° 1, pp. 17–25, January 2013.
- [11] J. Tourón, "La predicción del rendimiento académico: Procedimientos y resultados", *Rev. española Pedagog.*, vol. 1(25), n° 0034-9461, pp. 168–182, 1985.
- [12] J. Forteza, "Modelo instrumental de las relaciones variables motivacionales y rendimiento", *Rev. Psicol. Gen. y Apl. Rev. la Fed. Española Asoc. Psicol.* ISSN 0373-2002, Vol. 30, No. 132, 1975, págs. 75-92, vol. 30, n° 132, pp. 75–92, 1975.
- [13] M. Page, J.-L. Gaviria, y C. Gómez Bueno, *Hacia un modelo causal del rendimiento académico*. Madrid: Graficas JUMA, 1990.
- [14] T. Escudero, *Indicadores del rendimiento académico una experiencia en la Universidad de Zaragoza*. Madrid: Ministerio de Educación y Cultura, Consejo de Universidades, 1999.
- [15] O. Marrero y H. Espino, "Evaluación comparativa del poder predictor de las aptitudes sobre notas escolares y pruebas objetivas", *Rev. Educ.*, pp. 97–112, 1988.
- [16] A. Porto y L. Di Gresia, "Performance of University students and their determinants", 2005.
- [17] M. Noel, R. Ayán, M. Ángel, y R. Díaz, "Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados University students' academic achievement indicators: grades versus cumulative credits", *Rev. Educ.*, vol. 355, pp. 467–492, 2011.
- [18] T. Elias, "Learning Analytics: Definitions, Processes and Potential", 2011.
- [19] S. F. Borde, "Predictors of Student Academic Performance in the Introductory Marketing Course", *J. Educ. Bus.*, vol. 73, n° 5, pp. 302–306, May 1998.
- [20] C. Gloria, A. Pierart, y F. R. Pavés, "Learning styles, gender and academic performance", *J. Learn. Styles*, vol. 8, n° 8, pp. 71–84, 2011.
- [21] A. P. Vallejo y M. A. Risoto, "La variable Género y su relación con el autoconcepto y el rendimiento académico de alumnado universitario", *Rev. científica electrónica Educ. y Comun. en la Soc. del Conoc.*, vol. 13, n° 1, pp. 124–140, June 2013.
- [22] M. Sepúlveda, M. L. Quiero, P. T. Vergara, y J. L. Contreras, "Journal of learning styles.", *Difer. género en el Rend. académico y en el Perf. estilos y estrategias Aprendiz. en Estud. Química y Farm. la Univ. Concepción*, vol. 4, n° 7, October 2011.