MUJERES EN INGENIERÍA: EMPODERAMIENTO, LIDERAZGO Y COMPROMISO
21 - 24 DE SEPTIEMBRE 2021 CARTAGENA DE INDIAS, COLOMBIA
Encuentro Internacional de Educación en Ingeniería ACOFI
https://acofi.edu.co/eiei2021

MACHINE LEARNING TO PREDICT STUDENT ACADEMIC RISK IN ENGINEERING

Diego Patiño, David Barrera, Alejandra González, Óscar Martínez, Johan Peña, Lizeth Roldán Jiménez

Pontificia Universidad Javeriana
Bogotá, Colombia

Abstract

Identifying academic risk in an engineering program is very important since it allows to stablish policies that can be applied to prevent it and to guarantee student success. At Pontificia Universidad Javeriana in Bogotá – Colombia, we have developed some methods that can be used to predict the number of students that are in academic risk or even in dropout risk. This information can be employed to identify possible actions and interventions in our programs. We focus in the necessity to design and implement models to recognize patterns associated with academic performance and transitions of undergraduate students.

Our strategies are faced to accompany the students. Once we identify the risk, every student is evaluated by an academic adviser, who is in charge of determining the possible causes of the low performance and this way customize the support. In this sense, weaknesses are identified around CDIO competencies, for example, interpersonal competences, basic knowledge, algorithmic thinking, problem solving or personal skills. In this way, students are routed to different strategies to strengthen those abilities.

This paper starts with an introduction justifying this work. After that, a section describes the methodology to predict academic risk. Then, we present our accompaniment strategies. Some future works are also presented. In this context, this work tries to give hints about particular intervention for student in risk.

Keywords: machine learning; academic risk; drop out

Resumen

Identificar el riesgo académico en un programa de ingeniería es muy importante ya que permite establecer políticas que se pueden aplicar para prevenirlo y garantizar el éxito de los estudiantes. En la Pontificia Universidad Javeriana en Bogotá - Colombia, hemos desarrollado algunos métodos que se pueden utilizar para predecir el número de estudiantes que están en riesgo académico o incluso en riesgo de deserción. Esta información se puede utilizar para identificar posibles acciones e intervenciones en nuestros programas. Nos enfocamos en la necesidad de diseñar e implementar modelos para reconocer patrones asociados con el desempeño académico y las transiciones de los estudiantes de pregrado.

Nuestras estrategias están orientadas a acompañar a los alumnos. Una vez que identificamos el riesgo, cada alumno es evaluado por un asesor académico, quien se encarga de determinar las posibles causas del bajo rendimiento y así personalizar el apoyo. En este sentido, se identifican debilidades en torno a las competencias CDIO, por ejemplo, competencias interpersonales, conocimientos básicos, pensamiento algorítmico, resolución de problemas o habilidades personales. De esta manera, los estudiantes son encaminados a diferentes estrategias para fortalecer esas habilidades.

Este artículo comienza con una introducción que justifica este trabajo. Después de eso, una sección describe la metodología para predecir el riesgo académico. A continuación, presentamos nuestras estrategias de acompañamiento. También se presentan algunos trabajos futuros. En este contexto, este trabajo intenta dar pistas sobre la intervención particular para el estudiante en riesgo.

Palabras clave: *aprendizaje automático; riesgo académico; abandonar*

1. Introduction

The process of training students at the Pontificia Universidad Javeriana (PUJ) is described by the Student Development Model (MDE) that is inspired by the University's Educational Project. This model is the conceptual representation of the student's formative transit and describes the transitions or "critical moments" of the students from their progress in the curricula. Essential institutional interventions to facilitate this process have been identified and. It leds to the development of specific strategies to mitigate some risks. Figure 1 shows the student development model (Jaramillo, 2018).

The aim of the transition model is to identify the accompaniment routes that will allow the student to advance in his formative process. Circles are states of a student from his advance in credits of his program and the arcs that unite these states are the routes that allow to pass from one state to another. The accompaniment processes that define the transition paths are cumulative. Thus, each student must advance on the route and somehow complete all the conditions that allow him to transit between the states of formation. The processes proposed for each route coincide with strategies, projects, policies and infrastructure that the university offers, while others are in the process of development. Transition routes are based on the elements of the Integral Formation and the accompaniment that stand out as constitutive elements of the Educational Project and in the Mission of the University. Among the accompaniment processes are: Financial support, Integration and Awareness, Counseling, Academic support, Senses, well-being and community and Early warning (culture of student risk management). This model has been presented in (Gonzalez, 2018)

In particular, in order to respond to a culture of risk management, the Early Warning, Intervention and Monitoring System (SATIS – Sistema de Alertas tempranas, intervención y seguimiento) is proposed in a transversal manner, in order to promote the transit of students through each of the stages and ensure the completion of the entire cycle. The purpose of this system is to identify the situations or conditions of students who have a direct impact on their development, permanence and academic success at the University (Gonzalez, 2019).

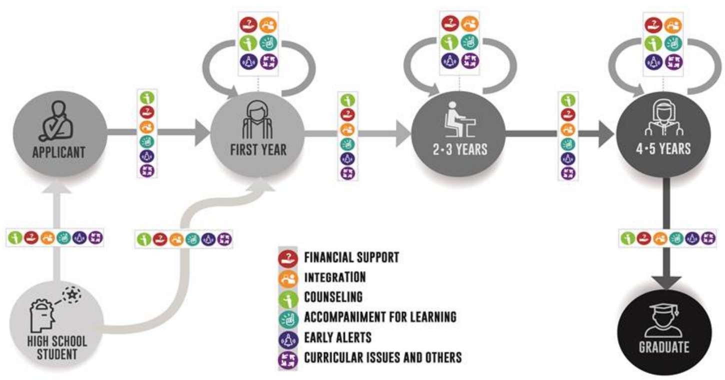

Figure 1. Student development model. (Jaramillo, 2018)

The SATIS project started in 2017 as part of the University Planning of the Academic Vice-Presidency and included the identification of individual, academic, socioeconomic and institutional risks. Within this framework, during 2018 the model for generating and prioritizing alerts and reports was established, the attention protocols were defined, reports were sent to program directors and the monitoring and intervention by directors and counselors was also implemented. SATIS has been developed from two fronts. First, we have implemented the system interface, including the parameterization of the corrective alerts mentioned. A second front is assumed by the Faculty of Engineering, which has advanced in prototypes for predictive alerts (Jayaprakash, 2014).

This article describes methods to predict academic risk or even dropout risk. This information can be employed to identify possible actions and interventions in the programs. We focus in the necessity to design models to recognize patterns associated with academic performance and transitions of undergraduate students. Thus, we intend then to make a prediction of the performance of a cohort of students and thus be able to identify intervention protocols that help to reduce dropout.

The article is divided into the following sections: Section 2 intends to give a context about the semester to semester transitions and the institutional definitions of risk are given. Section 3 presents the methodology used together with a population model based on simulation that aims to predict the number of students at risk. Section 4 presents the intervention schemes carried out. Section 5 shows the conclusions and future work.

2. Academic Risk at Javeriana University

In the University, the student regulation establishes a minimum cumulative average, which searches to be consistent with the mission of integral formation. Under this context, students are evaluated from 0 to 5 and curricula respond to a structure of academic credits. In this sense, the accumulated average is weighted according to the amount of credits and the final grade of the subject (GWA, accumulated weighted average). The minimum GWA required to consider a student in a normal academic situation is 3.4 or when the average academic period is less than 2.5. The PUJ is characterized by a culture of accompaniment for students (Torres, 2012), (Jaramillo, 2018). At this point a student at risk is identified when he is very close to this minimum GWA or below. A student can be in 5 states as shown in Figure 2, the "normal" state represents a GWA greater than or equal to 3.4. The “first academic probation” status represents the first semester in which the student with a GWA less than 3.4. The “second academic probation” status is presented when a student does not exceed the GWA condition of less than 3.4 after being in “first academic probation”. “Excluded” status is reached when the student does not exceed the GWA minimum for two consecutive semesters. “Retired” is when a student decided to not continue his studies in the program.

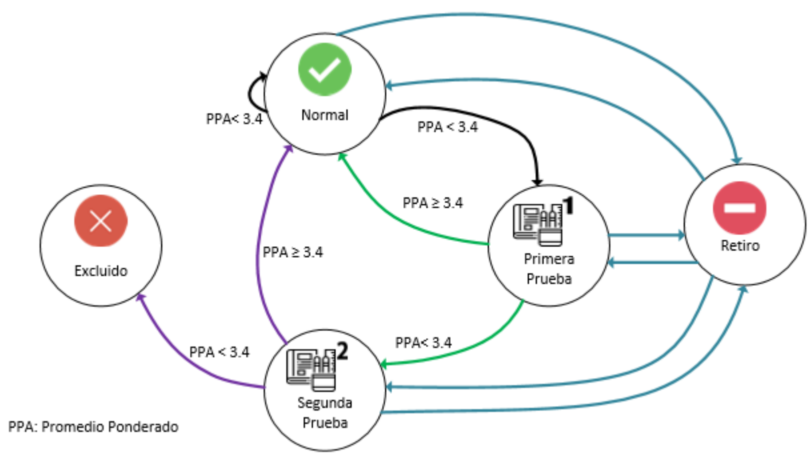

Figure 1. Academic probation model

The percentage of academic risk may not be considered significant for the university, the responsibility of training professionals for the country and accompanying the student in his life project makes this population more important. A qualitative and quantitative analysis of the its behavior allows us to identify some patterns. For example, students attending the first 4 semesters have a greater tendency to enter the test state given the conditions of adaptation (McKenzie, 2016). It has been also noticed a difference between the levels of schooling (high school-university), poor bases in mathematics and critical. Additionally, given that the amount of credits seen has a direct effect on the GWA, in the first semesters it is observed that having a poor performance in the semester significantly affects the GWA. Contrary to advanced semesters in which, due to the accumulation of credits, performance in a semester may not affect the GWA. However, in the first semesters to pass the academic test, it presents s less challenge given the inertia of the GWA and depends directly on the decision of the subjects to enroll. Finally, no significant difference was found in the risk to enter the second or exclusion test state based on the gender of the students, which means the proportion of students who enter these tests is similar between males and women. However, in the first test state, it was found that the proportion of men is usually twice that of women. The transitions of a semester to semester student can be visualized in Figure 3 following the regulations of students of the PUJ.

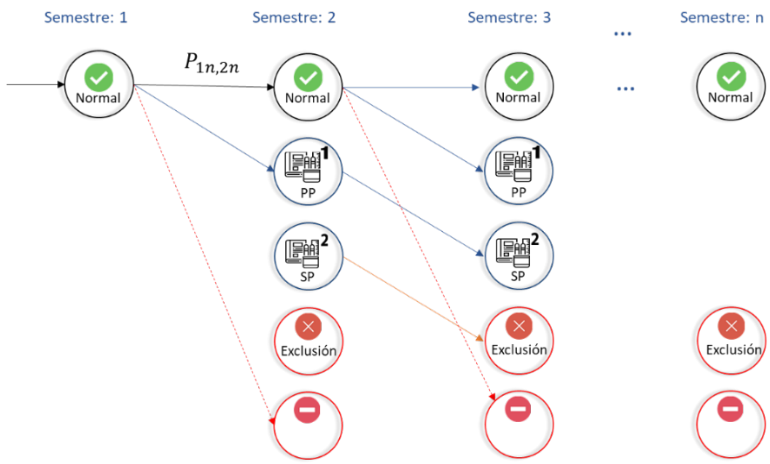

Figure 3. Transitions per semester

According to this figure, a student enters the first semester in a normal state and in the second semester can move on to a normal state, first academic probation or to retire state. In the second semester the concept of second academic probation or even dropout appears. Therefore, in order to determine the population at risk and have a predictive model, the probability of transition between semesters from one semester to another can be calculated. In the next section, we will describe the methodology and then present the results associated with the engineering programs of the PUJ. What is going to be predicted is in what status a student finishes his semester. In first semester, the model is different since we do not have any available information. In this case, we use the high school ranking and the grades that the student obtained finishing the last level. Since each phenomenon is different, the model that we use for each one changes.

3. Methodology

Methodologically, the prediction prototypes generated are developed under an exploration-intervention cycle, as shown in Figure 4. This cycle allows to generate characterizations of the phenomena studied from the dialogue with the different actors involved in the processes.
In order to follow it, the first step is to have meetings with the program Directors, counselors and professor in order to understand the behavior of the programs and the population to study. This information is used to model the different behaviors. Implementation is an essential key in this cycle. This is carried out using Matlab/Python. Results are presented also to the community and with them, some strategies of intervention are also defined. In general, the main collective construction activities with the academic community are described below:

Figure 4. Exploration – Intervention cycle

a. Exploration workshops and advance meetings: The characterization of each program began with workshops with directors and administrative staff. These spaces allowed to collect information for the classification of subjects in lines and components. Additionally, reforms need to be also identified in each program that could clarify possible behavior in the population. As examples, we can mention the elimination of subjects, changes of curriculum, definition of new lines, among others. On the other hand, meetings were held with directors and assistants to present the models and validate the results, which allowed generating alerts through risk heat maps for each program.
b. Training and ideation workshops: Training sessions were held with program assistants, in which the results were explained. It was also explained the use of the app to run the models. This app shows the prediction lists that gives information about the students that are in risk. Additionally, the counselors were trained through some intervention protocols validation workshop and were trained in the analysis of the prediction. Mentors and inductors were also trained in the intervention care channels when they detect difficulties in at-risk students.
c. Campaigns: The academic community was continuously informed of the progress of the project through presentations, advertising pieces of the student-oriented strategy were also presented. On the other hand, with the support of the Planning Office, the databases for analysis and the master databases for verification of data integrity are generated.

4. Model description

There are several models that are associated to each transition regarding the Figure 2. depending on the phenomenon, we choose different method. Models are validated with the data available in the information system of the four engineering programs (Electronics engineering, Industrial engineering, Civil engineering and Systems engineering). The models are described as follows:

a. Model for first academic probation: For the first academic probation, a neural network with a tensor model was used. The network divides two main tensors, the first one (Performance) calculates the academic performance of the cohorts, associating the historical performance of the students from 2012 to the current period. The second tensor (Individual) calculates the student's individual performance throughout his / her university life, including the academic load of the period to be predicted. The model does not predict unexpected low performance. This can be solved by including a tensor of first cut notes that allows analyzing the academic behavior in the prediction period (Géron, 2017). This model intends to make predictions from 2nd semester (See Fig. 3).

b. Model for first academic probation in first semester: Tensor model in which, in addition to the probability of the first probation, the probability of loss of the difficult courses of each program is predicted. These courses refer to those with a high morbidity history. The model includes training with the data of the whole engineering programs together. Despite the first model, there is no specific characterization of each program and this may mean the loss of information regarding the particularities. The collection of data is complex since the classification of the high school, the scores of the components of the state examination and also variables like students' gender must be taken into account. This is valid only for the transition from 1st semester to 2nd semester (See Fig. 3).
c. Model for second academic probation: This is a logistic regression that predicts the transition from first academic probation to second during two consecutive periods using the data of the programs. Given the prevalence of the phenomenon and the lack of data, the model includes training with the data of the programs together. Similar to the first model, this is valid from 2nd semester (See Fig. 3).
d. Model for drop-out prediction: It is a logistic regression that is used to predict the transition from second test to exclusion for two consecutive periods using the data of the four programs together so that the phenomenon has a significant prevalence. Similar to the first model, this is valid from 2nd semester (See Fig. 3).
e. Model for retire prediction: In this case, it was based on a population model. We care about the probability of a student to retire, but we also care about the number of the students that in a particular semester retire. Data on the academic states of all students from 2012 to the second semester of 2019 are used. The model is set as simulation, and is described in the following steps:

    i. Find the histogram of the different possible transitions between academic states by semesters as shown in Figure 3. This allows us to find the probability density function $P_{i,j}\ (e)$ where i indicates the semester origin and j the destination semester.
    ii. Initiate the model with the number of students in the first semester since 2012.
    iii. Perform Monte - Carlo simulations of at least 1000 replicas.

The model described above allows obtaining a semester-by-semester estimate of the students' academic status.

One of the fundamental characteristics of the database for all models is the gender of the students. Table 1 shows the proportion of men and women from the different engineering programs that historically entered the first academic test, based on the total number of students registered in the database.

Academic Program	Normal		First Academic Probation	
	Male	Female	Male	Female
Electronic Engineering	75%	86%	25%	14%
Civil Engineering	72%	85%	28%	15%
Industrial engineering	69%	85%	31%	15%
Systems engineering	75%	82%	25%	18%

Table 1. Proportion of men and women who historically entered to the first test.

All the models are validated with existing data. For example, first academic probation model results are compared with the real data. Table 2 shows this comparison for industrial engineering.

		Target	
		Normal	**First academic probation**
Predicted	**Normal**	82.7%	1.9%
	First academic probation	5.4%	10.1%

Table 2. Validation of a particular model...

In all cases, similar results are achieved that validate the different models.

5. Intervention

Once, alerts are predicted, some accompaniment strategies are designed. They are included in the Student Accompaniment Program (PAE, for its acronym in Spanish).
The PAE includes four lines of work that make the transition routes of the transitions model operational:

- PAE + 1: Accompaniment program for first year students. Alerts given by the Model for first academic probation in first semester are used.
- PAE + 2,3,4,5: Accompaniment program for students of year 2 and year 3. Alerts given by the for first academic probation, second academic probation and dropout are used.

Based on this, PAE designed some programs that are important. Special emphasis is given in first semester. From our population model, it has been notice that a short intervention in first semester gives a very large retention in the whole program. Some activities from PAE are:

a. Induction Program: Induction program seeks to impact the processes of integration into university life. It encourages the development of elements that allow students to assume their university role autonomously, responsibly and aware of the transcendence of the career within their life project. This is mainly for students in first semester. Special attention is put in the students with an alert from the predictive model.

b. Accompaniment of professors: The objective of this strategy is to offer professors, different tools to face the particularities of their courses taking into account the population. The aim is to ensure learnings and also to give vocational support. In this program, it is searched the link among critical courses. The idea is to give professors orientations about their teaching practices. Finally, another objective of the strategy is to generate appropriation of the transitions model. In this part, professors need to do a deeper analysis of the cases that the predictive alerts identify.
c. Mentoring program: This strategy aims to facilitate an environment of trust through peer-to-peer. The accompaniment here points out to the knowledge of the institutional processes and the understanding of the educational project. It also shows the tools and supports offered by the university for overcoming academic difficulties. Mentors facilitate the identification of risk situations associated with adaptation and integration to university life or academic performance.
d. Intelligent inscription: This is used by the program to change the courses that the students are taking according to the probability to be in second probation or excluded. Usually, it is an agreement between the student and the program director to change some courses depending on the difficulty of the student to pass them. The dropout predictive alert is the most important here.

6. Conclusion

In this article, it is shown different models that can be used to predict the academic risk of the students in the university. Based on this model, some predictive alerts can be identified. We covered since the characterization of a student in first semester until a possible dropout, and even exclusion. Tools such neural networks, logistic regression and Monte Carlo Simulation can be useful to this goal.

From the detected alerts, intervention strategies can be employed to prevent them. Professors play an important role here since they must give the necessary accompaniment. Evaluation models based on the assessment of student learning not only allow the curriculum to be regulated by identifying improvements in the course programs, CDIO competencies, curricular integration and the structure of the program (Crawley, 2014), (Al- Atabi, 2013), but also, allows to monitor the performance of students from the individual point of view and the possible generalized behavior of the population. Understanding this phenomenon intervention is done to each of the factors (curriculum-individual), responding to fulfilled facts.

By including predictive alerts, performing the intervention on unfulfilled facts allows a direct impact on the population anticipating behaviors and preventing incorrect decision making. On the other hand, the prediction of behavior patterns, allows quantifying resources in advance so that you can provide a number of counselors, mentors, tutors, classrooms and start the accompaniment processes even from the enrollment of courses.

Some future works consist in measure the real impact of the models in preventing students in academic risk. Implementation is being carries out by the engineering programs. However, it is a university decision to extend it to the whole university.

7. References Journal articles

- Al-Atabi, M. (2013). Grand Challenges for Engineering and Development of CDIO Skills. Proceedings of the 9th International CDIO Conference, Cambridge, MA, USA.
- Crawley, E., F. Malmqvist, J., Östlund, S., Brodeur, D., & Edström, K. (2014). Rethinking Engineering Education: The CDIO Approach, 2nd edition. New York: Springer-Verlag.
- Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn & TensorFlow,O'Reilly.
- González, A., Barrera, D.,León, M., Curiel,M., & Prieto,L.(2018). Student Success: On the Need for A New Standard, Proceedings of the 14th International CDIO Conference, Kanazawa Institute of Technology, Kanazawa, Japan.
- González, A., Patino, D., Roldán, L., Pena, J., Barrera, D. (2019). Toward Early Intervention: Model of Academic Performance In A CDIO Curriculum, Proceedings of the 15th International CDIO Conference, Aarhus University, Denmark.
- Jaramillo, C., Ordóñez, C., González, A., León, M., Curiel, M., Palacios, J., & Barrera, D. Formulación Del Programa De Prevención De La Deserción: Articulación De Acciones En Un Programa Transversal De Acompañamiento Para La Pontificia Universidad Javeriana. (2018). Congreso CLABES VIII, Ciudad de Panamá, Panamá.
- Jayaprakash, Sandeep & Moody, E.W. & Lauria, Eitel & J., Regan & Baron, J.D.. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. Journal of Learning Analytics. 1. 6-47.
- Torres, L. (2012). Retención estudiantil en la educación superior: revisión de la literatura y elementos de un modelo para el contexto colombiano. Bogotá: Editorial Pontificia Universidad Javeriana.

About the authors

- **Diego Patiño**, Ph D: received the Electronics Engineering degree from the National University of Colombia, Manizales-Colombia in 2002. Master degree in Automatic control and computers from the Andes University, Bogotá-Colombia, in 2005. Ph.D. degree in Automatic control and signal processing from the National Polytechnic Institute of Lorraine, Nancy, France in 2009. He is currently full-time professor at La Pontificia Universidad Javeriana at Bogotá, Colombia and head of the Electronics Department. patino-d@javeriana.edu.co
- **Alejandra González**, Ph.D is an Associate Professor in the Department of Electronics Engineering at the Pontificia Universidad Javeriana, Bogotá, Colombia. Her research interests include embedded systems, digital systems design, high-level systems design, artificial intelligence, robotic and curricular management. agonzalez@javeriana.edu.co
- **David Barrera**, MSc is an Assistant Professor at the Industrial Engineering Department at the Pontificia Universidad Javeriana, Bogotá, Colombia. Since 2015, he has been the Head of the Industrial Engineering Program. His research interests include applied operations research, and curricular management. barrera-o@javeriana.edu.co

- **Óscar Martínez**, MSc is an Electronics Engineers from the Pontificia Universidad Javeriana. He is a lecturer in the Electronics Engineering and Systems Engineering Departments at Pontificia Universidad Javeriana. His research interest includes bioengineering, neuroscience, signal processing, and artificial intelligence omartinezb@javeriana.edu.co
- **Lizeth Roldán, MSc is** an Electronics Engineer from the Pontificia Universidad Javeriana. She is a lecturer in the Electronics Engineering and Industrial Engineering Departments at Pontificia Universidad Javeriana.Her research interest includes bioengineering, Control system, signal processing and artificial intelligence roldan.l@javeriana.edu.co
- **Johan Peña**, MSc is an Electronics Engineers from the Pontificia Universidad Javeriana. He is a lecturer in the Electronics Engineering Department at Pontificia Universidad Javeriana. His research interest includes bioengineering, neuroscience, signal processing and artificial intelligence. pena.johan@javeriana.edu.co

---