

Análisis y modelización de datos para la caracterización de estudiantes de primera matrícula de la UNAD

Analysis and modelling of data for characterization of first-time students at UNAD

Juan Olegario Monroy¹

Pilar Alexandra Moreno²

Resumen

Este documento, describe las tareas iniciales llevadas a cabo en el desarrollo del proyecto "Escenario de transición de una formación presencial a una formación virtual para los estudiantes de primera matrícula de la UNAD". En dicho trabajo se realizó la identificación de los aspectos que se pueden analizar dentro de los procesos de inducción en educación superior en ambientes virtuales de aprendizaje. En tal sentido, se revisaron diferentes modelos de predicción que se han venido aplicando frente al análisis del comportamiento, motivación y deserción; los cuales pueden ayudar a establecer las variables y relaciones a considerar sobre los estudiantes de primera matrícula de la UNAD. De igual forma se examinaron diferentes estudios para el reconocimiento de las aplicaciones más usadas en el análisis y tratamiento de datos, con el fin de explorar ejercicios de modelización y análisis de información, así como de orientar la revisión sistemática que también hace parte de las etapas iniciales del proyecto, hacia la búsqueda de herramientas y estrategias usadas en procesos de inducción.

Palabras clave: análisis de datos, estudiante universitario, modelización, admisión académica, Aprendizaje en línea.

Abstract

This document describes the initial tasks carried out in the development of the project "Transition scenary from presential learning to E-learning for first-time students at UNAD". In this work, the identification of the aspects that can be analyzed within the induction processes in higher education in virtual learning environments was carried

¹ Ingeniero electrónico. Docente asistente, Universidad Nacional Abierta y a Distancia UNAD.
<https://orcid.org/0000-0003-1757-2016> / juan.monroy@unad.edu.co

² Ingeniera de sistemas. Docente asistente, Universidad Nacional Abierta y a Distancia UNAD.
<https://orcid.org/0000-0002-6990-6105> / pilar.moreno@unad.edu.co

out. In this sense, different prediction models that have been applied against the analysis of behavior, motivation and desertion were reviewed; which can help to establish the variables and relationships to consider about the first enrollment students of the UNAD. In the same way, different studies were examined for the recognition of the most used applications in the analysis and treatment of data, in order to explore modeling exercises and information analysis, as well as to guide the systematic review that is also part of the stages of the project, towards the search for tools and strategies used in induction processes.

Keywords: Data analysis, University student, Modelling, Academic admission, E-learning.

1. Introducción

La inmersión en la modalidad de educación virtual en la Universidad Nacional Abierta y a Distancia –UNAD–, es una situación de interés tanto administrativa como profesoral. En el nivel de adaptación que pueden alcanzar los estudiantes influyen diferentes factores motivacionales, económicos, de aprendizaje, de acompañamiento, recursos tecnológicos, competencias digitales, comunicativas y psicosociales, entre otros.

De ésta forma, y teniendo en cuenta lo señalado por Flechas & Moreno (2015) “la comunidad estudiantil que inicia su proceso de formación de la UNAD, está compuesta por estudiantes nativos digitales que no están muy acostumbrados a la autodisciplina, que su aprendizaje y medios de comunicación requieren de más estímulos externos, visuales e interactivos para regularse y que además provienen, en su mayoría, de una modalidad educativa presencial asistida 100 % por profesores”, surge el desarrollo del proyecto cuyo objetivo es modelar un escenario de transición de una formación presencial a una formación virtual para estudiantes de primera matrícula en la UNAD.

En este documento se presenta la explicación de las tareas iniciales de dicho proyecto, para el proceso de identificación de las variables a analizar, el modelamiento de los datos y la obtención de la caracterización de estudiantes de primera matrícula de la UNAD. Dicho proceso, se estructuró en 3 fases, así:

- Fase 1. Modelos de predicción
- Fase 2. Variables del estudio
- Fase 3. Modelamiento de datos

2. Metodología

Este análisis, realizado para la identificación de variables que orienten el estudio de estudiantes de primera matrícula de la UNAD, es importante porque permitirá brindar estrategias específicas y recomendaciones sobre los principales elementos a considerar en la caracterización de los estudiantes y el proceso de inducción involucrado, convirtiéndose en el punto de partida para el diseño del escenario de transición de una formación presencial a una formación virtual para los estudiantes de primera matrícula de la UNAD. Específicamente, la propuesta que se presenta busca reconocer diferentes técnicas para el análisis de datos que apunten al reconocimiento oportuno de las necesidades de los estudiantes de primera matrícula en su proceso de transición y adaptación hacia la formación virtual que encuentran en la UNAD, basado en el comportamiento de los estudiantes de primera matrícula, en sus procesos de inducción, en su interacción en los cursos AVA, los cuales son la base del proceso de formación.

A partir del estudio de los modelos de predicción y técnicas de análisis de datos, fue posible determinar los datos e información que se pueden considerar acerca de los estudiantes de primera matrícula de la UNAD, sin embargo, no se implementará ninguno de los modelos ya que no se requiere predecir algún comportamiento de los estudiantes sino sólo identificar las variables y relaciones que el escenario a desarrollar debe abordar. El proceso por fases para identificar las variables fue el siguiente:

2.1 Fase 1. Modelos de predicción para determinar variables de estudio

En primer lugar, se identificaron los modelos de predicción y técnicas de minería de datos, en función de las variables que en el modelo se consideran y que tienen que ver con los procesos de inducción y trabajo inicial de estudiantes en ambientes virtuales de aprendizaje. Algunos modelos identificados como más relevantes son: Numerical Analysis: Linear programming, Probability, Statistics, Logistic regression, Algorithms: 1) General Bayesian Network (GBN) y 2) Decision tree, Técnicas minería datos: 1) Clustering, 2) Árboles de decisión y 3) Reglas de asociación.

Estos modelos y técnicas, encontradas en referencias y estudios previos, permiten relacionar herramientas y estrategias usadas en procesos de inducción de estudiantes con la continuidad, motivación o deserción de su proceso de formación. Allí, también, se relacionan las técnicas de minería de datos utilizadas en estudios similares y que sirven

de base para el análisis e identificación de variables claves para el proceso de caracterización de estudiantes, que requiere el proyecto.

2.2 Fase 2. Variables del estudio

Como segunda fase, luego de evaluar los principales modelos de predicción, técnicas de minería de datos y de revisión de la aplicación de diferentes herramientas, se realiza un análisis e identificación de las variables que inciden en procesos de retención, motivación y permanencia y que se relacionan para la caracterización de los estudiantes de primera matrícula.

Para ello, se revisaron diez referentes de análisis, los cuales sugieren clasificaciones de variables desde aspectos sociales, comportamentales, académicos, personales, hasta aspectos geográficos y económicos; tanto de estudiantes como de docentes y de instituciones. De esos referentes, de manera general se pueden mencionar las siguientes variables del estudiante identificadas como ejes principales del proyecto:

Atributos demográficos: género, estado civil, etnia, raza, lugar de residencia.

Atributos de inscripción: edad de admisión, primer semestre de admisión, colegio del estudiante, carrera del estudiante.

Atributos de rendimiento: discapacidad, rendimiento en pruebas iniciales: 1) Test habilidades básicas, 2) Prueba de habilidades cognitivas.

2.3 Fase 3. Modelamiento de datos

En este ejercicio de análisis previo se definieron los siguientes pasos para el tratamiento de la información a recopilar para el proceso de caracterización.

Paso 1. Selección de los datos más relevantes y necesarios para el posterior análisis. En este punto se realizaron las siguientes operaciones:

- Identificar orígenes de datos, identificar formas de acceso, tramitar los accesos correspondientes a la información identificada requerida.

Paso 2. Depuración de los datos para establecer la estructura de datos necesaria para las posteriores etapas. Se realizó: el análisis de datos, la extracción de datos y la síntesis de datos.

Paso 3. Transformación de los datos de acuerdo a la estructura establecida para su tratamiento. Las operaciones cumplidas para

ello, son: realizar el tratamiento preliminar de datos, identificar variables actuales pertinentes para el proyecto y generar nuevas variables a partir de las ya existentes con una estructura de datos apropiada.

3. Discusión y resultados

Del proceso de análisis y modelamiento de datos, se genera una tabla de datos; con base en la información y caracterización de estudiantes de primera matrícula; brindada desde la consejería académica y recogida en los periodos 16-04 y 16-05 del año 2021.

Sobre esa tabla de datos, se establecieron los siguientes análisis, cruzando variables de acuerdo a las tres necesidades de caracterización establecidas al inicio de la investigación, así:

1. Condiciones de entrada de los estudiantes, que se tomarán de las respuestas de los estudiantes frente a los recursos y medios electrónicos con los que cuentan para acceder al campus virtual de la UNAD.

2. Las habilidades que ha desarrollado el estudiante para aprender a aprender, en su experiencia previa, que se tomarán de las respuestas dadas frente a aspectos de su proceso de aprendizaje en relación a las formas más empleadas incluyendo aspectos motivacionales.

3. Las competencias con las que cuenta el estudiante al ingresar y que le sirven para su aprendizaje, que se obtendrán del análisis de las respuestas frente a sus competencias: digitales, lectura crítica, razonamiento cuantitativo, inglés, biología, física, química.

La relación de las variables para cada uno de los ítems anteriores se dará a partir de los siguientes grupos: edad, sexo, estrato, estado civil, programa que estudia en la universidad, modalidad del bachillerato que realizó, tipo de colegio, último nivel de estudio, tiempo sin estudiar, si ha tomado cursos virtuales, situación laboral y tiempo para el desarrollo de las actividades académicas.

4. Conclusiones

Se evidencia, entonces, que al interior de la investigación es necesario realizar una caracterización en dos líneas:

- La primera caracterización, orientada a la revisión del proceso de inducción que se está llevando a cabo al interior de la Universidad. Si bien se cuenta con un proceso de inducción y lineamientos generales establecidos a nivel nacional, en cada una de las zonas se establecen

planes y acciones concretas que responden, de manera contextualizada, a la situación que reflejan los estudiantes en la encuesta que realizan al inicio del periodo académico.

- La segunda caracterización, se debe orientar al análisis y cruces de datos sobre los resultados obtenidos en las encuestas de ingreso que responden los estudiantes, las cuales abordan elementos y variables, tales como: datos personales, grupos poblacionales, datos familiares, datos académicos, recursos tecnológicos, formas de aprender, datos laborales, actividades recreativas, artísticas y deportivas, emprendimiento, vida saludable, ambiental, psicosocial, competencias y proyecto de vida.

Referencias

- Aldrich, J. & Nelson, F. (1984). *Linear probability, logit and probit model*, Sage Publications. <https://methods.sagepub.com/book/linear-probability-logit-and-probit-models>
- Amaya, Y. & Barrientos, E. (2013). *Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos aplicado al sistema de información académico de la Universidad Simón Bolívar*. (Tesis de grado). Universidad Simón Bolívar, Barranquilla.
- Castaño, E., Gallón, S., Gómez, K. & Vásquez, J. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de Economía*, 65(65), 9-36. <https://doi.org/10.17533/udea.le.n65a2639>
- Flechas, H. & Moreno, P. (2015). Gamification como estrategia didáctica para cursos en ambientes virtuales de aprendizaje. *I Encuentro Interzonal de E-Investigación, IV Cumbre Nacional de paz y II Encuentro Interzonal de Investigación "Formación del talento humano para la construcción de una cultura de paz"* (pp. 65-77). Bogotá: Universidad Nacional Abierta y a Distancia.
- Guzmán, C., Durán, D. & Franco, J. (2019). *Deserción estudiantil en la educación superior colombiana*. Bogotá: Ministerio de Educación-Viceministerio de Educación Superior.
- Martelo, R., Acevedo, D. & Martelo, P. (2018). Análisis multivariado aplicado a determinar factores clave de la deserción universitaria, *Revista Espacios*, 39(10), 13. <https://www.revistaespacios.com/a18v39n10/a18v39n10p13.pdf>

Perin, D. (2007). Can community Colleges Protect both Access and Standards? The problem of Remediation. *Teachers College Record*, 108. <https://ccrc.tc.columbia.edu/publications/access-standards-remediation.html>

Smith, J. & Naylor, R. (2001). Dropping out of University: A Statistical Analysis of the probability of Withdrawal for UK University. *Journal of the Royal Statistical Society*, 164(2), 389-405. <https://doi.org/10.1111/1467-985X.00209>