

Analíticas de aprendizaje en el análisis de datos de alumnos ingresantes a carreras de Ingeniería de la UNLP

Di Domenicantonio, Rossana M.¹; Hasperué, Waldo²

¹ Facultad de Ingeniería / Universidad Nacional de La Plata

Tel. +54 221 - 4892511 Calle 1y47 / La Plata / Buenos Aires / Argentina

² Facultad de Informática / Instituto de Investigación en Informática (III-LIDI) / Universidad Nacional de La Plata, La Plata, Argentina. Investigador Asociado, Comisión de Investigaciones Científicas

rossanadido@ing.unlp.edu.ar, whasperue@lidi.info.unlp.edu.ar

RESUMEN



En la Facultad de Ingeniería de la UNLP se dictan trece carreras y la primer materia del plan de estudios es "Matemática para Ingeniería". Una modalidad de dictado es la denominada "anticipada" para alumnos que estén finalizando el colegio secundario y si residen a más de 50km de La Plata la realizan a distancia con la plataforma Moodle.

El objetivo de este trabajo es describir el inicio de una tesis de Maestría con fin ser compartida y puesta en discusión con pares que trabajan en la misma temática. Se describe la preparación y análisis de los registros de tres cohortes de alumnos que realizaron esta modalidad con el fin de estudiar las características y patrones que permitan inferir un modelo predictivo del rendimiento académico utilizando técnicas de minería de datos y con analíticas de aprendizaje interpretar el significado de los datos. Se seleccionarán las variables de mayor impacto y según el análisis y la observación de patrones en relación con los resultados finales de los alumnos, se confeccionará un modelo predictivo que pueda inferir de manera temprana indicadores de posibles deserciones y generar estrategias de contención que contribuyan a un mejor aprovechamiento de la materia en esta modalidad

163

Palabras claves: Analítica de aprendizaje. Matemática. Moodle. Ingresantes. Ingeniería

ABSTRACT



Thirteen careers are taught at the UNLP School of Engineering and the first subject of the curriculum is "Mathematics for Engineering". One type of dictation is the so-called "anticipated" for students who are finishing high school and if they reside more than 50km from La Plata, they do it remotely with the Moodle platform.

The objective of this work is to describe the beginning of a Master's thesis in order to be shared and put into discussion with peers working on the same subject. It describes the preparation and analysis of the records of three cohorts of students who performed this modality in order to study the characteristics and patterns that allow to infer a predictive model of academic performance using data mining techniques and with learning analytics interpret the meaning of the data. The variables with the greatest impact will be selected and according to the analysis and observation of patterns in relation to the final results of the students, a predictive model will be prepared that can infer early indicators of possible dropouts and generate containment strategies that contribute to a better use of matter in this modality.

INTRODUCCIÓN

En la Facultad de Ingeniería (FI) de la Universidad Nacional de La Plata se dictan actualmente trece carreras con las diferentes especialidades. La primer materia del plan de estudios de todas las carreras es "Matemática para Ingeniería" que se ofrece en distintas modalidades de cursado durante el año. Una modalidad es para aquellos estudiantes que realizan la última etapa del colegio secundario en sus lugares de residencia y quieren realizarla de modo virtual. Esta modalidad es denominada "anticipada" por la particularidad que aún no son alumnos de la Facultad pero pueden realizarla mientras terminan el colegio secundario y luego se inscriben en la Facultad.

164

En la FI se implementa el acompañamiento virtual con la plataforma educativa Moodle desde el año 2016. Los contenidos matemáticos abordados son: conjuntos numéricos, polinomios, ecuaciones polinómicas y fraccionarias, rectas, cónicas, sistema de ecuaciones lineales y mixtos y trigonometría. Según Ander-Egg (1991), "la participación activa de todos, los docentes y alumnos, es un aspecto central del sistema de enseñanza y aprendizaje...se enseña y se aprende a través de una experiencia realizada conjuntamente en la que todos están implicados e involucrados". En el aula virtual se intenta replicar la modalidad de cursado presencial de aula taller donde los alumnos son los protagonistas y los docentes los guías de su aprendizaje. Para ello se habilitan foros de consultas donde se promueve la participación e interacción de los estudiantes en torno a la resolución de ejercicios y/o problemas y los docentes acompañan e intervienen en el momento necesario para encauzar el debate.

El objetivo del presente trabajo es describir el inicio de una tesis de Maestría en Tecnología Informática aplicada en Educación, donde se preparan y analizan los datos de los alumnos que realizaron la modalidad anticipada de Matemática para Ingeniería desde que se inició con la plataforma Moodle con el fin de poder estudiar las características, variables, patrones

y demás datos predictivos que permitan inferir un modelo que luego se utilizará para detectar falencias tempranas, evitar deserciones, mejorar el rendimiento académico y poder planificar mejoras para implementaciones futuras de esta modalidad de cursado a distancia de la materia. La propuesta presentada en este trabajo tiene como fin ser compartida y puesta en discusión con pares que trabajan en la misma temática.

Descripción de la propuesta

Partiendo de la premisa que "Lo que no se mide, no se puede mejorar" (William Thomson 1824-1907), surge la idea de analizar el rendimiento académico de alumnos que realizaron "Matemática Para Ingeniería" a distancia a través de la plataforma Moodle. Se analizarán los datos de alumnos de la cohorte 2017, 2018 y 2019 con el fin de realizar predicciones del rendimiento académico de los estudiantes y elaborar un modelo utilizando técnicas de minería de datos y mediante analítica de aprendizajes (LA) interpretar el significado de los datos obtenidos y las relaciones encontradas.

Se realizará una breve recopilación sobre "Learning Analytics" (LA) citando a diferentes autores que han realizado estudios con esta herramienta con el fin de utilizar la enorme cantidad de datos que generalmente se dispone cuando se utilizan entornos virtuales en los procesos de formación y no son aprovechados para realizar predicciones que permitan tomar decisiones oportunas.

Fundamentación

Según M. Zapata-Ros (2014) "ahora hay una nueva perspectiva: La analítica masiva de datos personalizados". Los algoritmos adecuadamente orientados por las teorías del aprendizaje personalizado, por técnicas pedagógicas y de diseño instruccional pueden, junto con los avances en minería de datos, obtener información para ajustar mejor la intervención educativa, para mejorar el rendimiento de los alumnos, y el del programa educativo. Según el mismo autor,

un desafío de naturaleza prioritaria consiste en utilizar la analítica para detectar indicadores de abandono precoz en estudios en línea.

La analítica de aprendizaje se puede definir como el proceso de determinar, evaluar e interpretar el significado de grandes volúmenes de datos educacionales; utilizando para ello algoritmos matemáticos. (Baker, 2014; Jaramillo & Paz, 2015).

Si una tecnología cobra importancia en el seno de una sociedad es porque sus miembros se apropian de ella, a la vez que la transforman y se transforman en un juego bidireccional (Burbules y Callister, 2001). En este sentido el procesamiento de gran volumen de información sobre el rendimiento de los alumnos ingresantes a carreras de ingeniería será un conjunto de datos que brindará información relevante del universo de datos estudiados dentro del contexto específico y el juego bidireccional estará dado por el uso de los resultados una vez realizado el análisis y el proceso de seguimiento y mejora de la implantación para futuros ingresantes.

La Society for Learning Analytics Research lo define como un campo de estudio y lo describe como "la medición, recolección, análisis y presentación de datos sobre los alumnos y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que se produce" (Long y Siemens, 2011, p. 34).

Romero, et al (2014) recolectaron datos de estudiantes de un foro de discusión en línea, los clasificaron en varios subconjuntos de datos y luego evaluaron la precisión predictiva de cada conjunto a través de métodos de minería de datos. Para ello utilizaron un algoritmo de clasificación y datos de interacción de los alumnos antes de un examen para predecir el rendimiento en el aprendizaje de los mismos.

Huang&Fang (2013) aplicaron cuatro modelos matemáticos para predecir el rendimiento académico de estudiantes de un curso de ingeniería utilizando para ello las calificaciones

finales de los mismos. Los resultados finales mostraron que los puntajes de los exámenes finales de los estudiantes eran predecibles con un 88% de precisión en base a ocho variables recopiladas de un sistema de gestión de aprendizaje (LMS).

Hu&Shih (2014) desarrollaron un sistema de alerta temprana basado en árboles de decisión para predecir si los estudiantes aprobarían o no. El modelo fue construido utilizando datos de 300 estudiantes y 13 variables recogidos mediante analítica en línea. Los resultados revelaron un 95% de precisión.

Salgado Reyes et al (2018), sostiene que la minería de datos educativos desarrolla modelos y métodos para explorar los datos recopilados de los entornos de aprendizajes educativos mediante analíticas de aprendizajes con el fin de detectar patrones que permitan predecir variables de interés en instituciones educativas universitarias.

Ye & Biswas, (2014) A pesar de que el rendimiento académico es una variable multifactorial, muchos de los estudios alrededor de la misma incluyen solo factores personales y socioeconómicos; sin embargo, el surgimiento y la aplicación de las nuevas tecnologías de enseñanza sobre todo el uso de las plataformas virtuales, permiten a las universidades recolectar una gran cantidad de información en tiempo real. Estos cuantiosos datos electrónicos generados proporcionan un abordaje multivariante en el estudio del rendimiento académico (citado por Salgado Reyes et al.).

Propuesta a realizar

Para el análisis de datos se estudiarán los registros de tres cohortes de alumnos ingresantes:



Ingresantes	Fecha realización	Alumnos inscriptos	Rindieron	Aprobaron	% Aprob / inscriptos
2017	Agosto 2016	109	82	32	39%
2018	Agosto 2017	147	112	49	44%
2019	Agosto 2018	193	123	49	40%

Con los datos que se obtengan de la base de datos de la plataforma de estos alumnos se confeccionará un modelo predictivo con el que se espera entender, mejorar y optimizar el aprendizaje y el rendimiento académico de futuras implementaciones. La información a ser evaluada se obtiene a través de las interacciones que los alumnos realizaron en el entorno virtual de enseñanza y aprendizaje desarrollado en Moodle (<https://www.ingreso.ing.unlp.edu.ar/>) diseñado especialmente para los estudiantes que realizan esta modalidad anticipada y que residen a un radio mayor a 50 km de la Facultad.

166

Los datos a analizar se obtuvieron de la base de datos MySQL que la plataforma Moodle ofrece. Se extrajeron los datos de aquellas tablas que tuvieran datos relevantes según el diagrama del curso y los recursos educativos que fueron utilizados de modo de poder realizar un cruce de la información por alumno con la nota final del mismo.

Los registros suministrados por la base de datos de esta plataforma serán sometidos a una limpieza primero y se definirán aquellos más relevantes que se utilicen para el modelo. El sistema utilizado para el proceso de minería de datos es Rapidminer (<https://rapidminer.com/>). Una vez realizada la definición de variables predictoras, se descartarán los irrelevantes, se analizarán, y procesarán los datos a utilizar para la confección del modelo. Conocer el estado de ciertos parámetros es clave para planificar, establecer objetivos, controlar resultados y tomar decisiones.

RapidMiner es una plataforma de análisis predictivo, aprendizaje automático, minería

de datos y análisis de negocio que permite la carga, la transformación y el modelado de grandes cantidades de datos que pueden estar distribuidas en diferentes fuentes. Presenta una interface amigable al usuario, la cual permite mediante el uso de operadores predefinidos elegir la fuente de datos a analizar, filtrarlos, limpiarlos y seleccionar las variables de interés. Luego se puede continuar con el uso de distintos operadores específicos para realizar la obtención de modelos predictivos, modelos descriptivos, y modelos temporales.

OBJETIVOS

El objetivo general de la tesis consiste en analizar los datos de tres cohortes de alumnos ingresantes a carreras de Ingeniería que realizan un curso de "Matemática para Ingeniería" con modalidad denominada "anticipada", por realizarse mientras terminan el último semestre del colegio secundario y además a distancia por residir a más de 50 km del radio de la ciudad de La Plata.

Los objetivos específicos son:

- Utilizar Learning analytics para a) analizar la cantidad de interacciones y participación en los foros ya que estos estaban diseñados para las consultas y trabajo conjunto sobre ejercicios prácticos que los alumnos debían desarrollar; b) interacciones entre los mismos alumnos en el foro de consultas y/o con los docentes; c) acceso a los simulacros de parcial realizados por los docentes en videos subidos a la plataforma; d) caminos más frecuentes; actividades menos realizadas o visitadas; e) el tiempo total en línea; f) la regularidad de accesos a la plataforma; g) los recursos didácticos más accedidos; acceso al material propuesto para cada semana; h) datos del alumno como tipo del colegio; edad u otras.

- Analizar y obtener información del procesamiento de los datos. Con ello se conformará nuevo conocimiento, que será útil tanto para los docentes como para los alumnos. A los estudiantes se les puede recomendar

actividades y recursos que favorezcan su aprendizaje, y los docentes, pueden obtener una retroalimentación objetiva para el proceso de enseñanza mediada con Moodle.

- Proponer mejoras en el diseño y materiales educativos utilizados en la plataforma. Repensar la estructura de las aulas virtuales utilizadas y su eficacia en el proceso de aprendizaje, agregando recursos educativos más visitados, consignas más orientadas y también, pensar en clasificar a los alumnos en grupos en función de sus necesidades de orientación y seguimiento.

CONCLUSIONES

Según los resultados que se desprendan del análisis de los datos y de la observación de patrones detectados en relación con los resultados finales obtenidos por los alumnos, se confeccionará un modelo predictivo que pueda inferir de manera temprana algunos indicadores de posibles deserciones o abandono, que pudiera ser atendida por los docentes a cargo y generar estrategias de contención que contribuyan a un mejor aprovechamiento de la materia con esta modalidad de cursado. En coincidencia con Maggio, M. (2016) los registros de las prácticas de enseñanza y los procesos de aprendizaje en ambientes con alta disposición tecnológica reflejan una mayor complejidad y profundidad que en otros modelos con menor acceso a la tecnología. En este sentido es un desafío la propuesta planteada donde observar, revisar, entender y construir un modelo y su validación que enriquezca el modelo actual que tiene la modalidad a distancia de la materia.

El resultado obtenido del análisis y procesamiento de los datos, conformará nuevo conocimiento, que será útil tanto para los docentes como para los alumnos. A los estudiantes se les puede recomendar actividades y recursos que favorezcan su aprendizaje, y los docentes, pueden obtener una retroalimentación objetiva para el proceso de enseñanza mediada con Moodle. Podría repensarse la estructura de las aulas virtuales y su eficacia en el proceso de

aprendizaje, agregando recursos educativos más visitados, consignas más orientadas y también, pensar en clasificar a los alumnos en grupos en función de sus necesidades de orientación y seguimiento.

La minería de datos educativos provee una gran cantidad de indicadores que pueden ser utilizados tanto en esta materia de matemática con modalidad a distancia como en otras materias de la FI que utilizan la misma plataforma con el fin de detectar patrones que motiven a docentes y alumnos a implementar estrategias de retención y mejoras en el proceso de enseñanza y aprendizaje.

BIBLIOGRAFÍA

- Ander-Egg, E. (1991). El taller, una alternativa de renovación pedagógica. Editorial Magisterio del Río de la Plata, Buenos Aires, pp 13.
- Baker, R. (2014). Educational data mining: An advance for intelligent system in education. *IEEE Intelligent Systems*, 78-82
- Burburles, N.C., y Callister, T. A. (2001). Educación: riesgos y promesas de las nuevas tecnologías de la información. Barcelona: Granica.
- Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education* , 133–145.
- Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students online learning performance. *Computers in Human Behavior*, 469–478.
- Jaramillo A. & Paz Arias H. (2015) Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje. *Revista Tecnológica ESPOL*, Vol. 28, N. 1, 64-90.

Long, P, & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education, *Educause Review*, 46(5), 31-40. Recuperado de <https://bit.ly/2UvNXuA>.

Sabulsky, G (2019) Analíticas de aprendizaje para mejorar la enseñanza y el seguimiento a través de entornos virtuales. *Revista Iberoamericana de Educación OEI*, Vol. 80. Núm. 1 - Madrid, pp 13-29

Maggio, M. (2016). Enriquecer la enseñanza. Los ambientes con alta disposición tecnológica como oportunidad". Buenos Aires: Editorial Paidós, pp 113-165.

Salgado Reyes N., Beltrán Morales J., Guaña Moya J., Escobar Teran Ch., Nicolalde Rodriguez D., Chafla Altamirano G. (2018) Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje.

168 Zapata-Ros, M. (2013). Analítica de aprendizaje y personalización. *Campus Virtuales*, 2(2), 88–118. Disponible en: <http://uajournals.com/ojs/index.php/campusvirtuales/article/view/41>

