

LAS REDES NEURONALES ARTIFICIALES Y LA DESERCIÓN EN EL NIVEL UNNIVERSITARIO

Marcelo Omar Sosa¹, Carlos Iván Chesñevar², Eugenia Cecilia Sosa Bruchmann¹

¹Departamento Computación/Facultad de Ciencias Exactas y Naturales/Universidad Nacional de Catamarca

Av. Belgrano N° 300 - Planta alta - C.P: 4700 - San Fernando del Valle de Catamarca
Teléfono: 0383- 4425610 /4420900

²Instituto de Ciencias e Ingeniería de la Computación (Conicet –U.N.S.)/Facultad de Ciencias e Ingeniería /Universidad Nacional del Sur

Av. Alem 1253 -B8000CPB Bahía Blanca, Argentina
Teléfono: 0291- 4595135 ext. 2610

sosamod1@hotmail.com, cic@cs.uns.edu.ar, sosab_ec@hotmail.com

RESUMEN

La deserción como problema en el nivel universitario representa una preocupación constante, se incrementa año a año y toma mayor importancia en el 2.020 con la situación de pandemia. Aunque se viene analizando numerosas soluciones y en la búsqueda constante de alternativas que permitan disminuirla. Con la integración generalizada de las plataformas virtuales de aprendizaje en el proceso educativo el volumen de datos recogidos creció exponencialmente. Permitió registrar cada acción realizada por los estudiantes lo que posibilitará la materialización de perfiles y el análisis de factores que influyen directamente en el rendimiento académico. Para el procesamiento de estos datos educativos se propone la utilización de redes neuronales artificiales (RNA) y el desarrollo de experimentaciones para que los resultados permitan la predicción de los posibles desempeños académicos de los estudiantes y la detección de aquellos que se encuentran en riesgo de abandono. El objetivo principal que se plantea la investigación es la disminución de la deserción basado en los resultados obtenidos de la implementación de las RNA. Se espera que esta predicción permita al docente actuar adecuadamente en temas puntuales e influir positivamente en el desempeño académico de los estudiantes.

Palabras clave: Deserción, redes neuronales artificiales.

CONTEXTO

El presente trabajo de investigación se enmarca dentro de las actividades de la tesis de maestría que se desarrolla conjuntamente con tareas dentro de los centros de investigación de informática y estadística aplicada de la Facultad de Ciencias Exactas y Naturales (FACEN) - Universidad Nacional de Catamarca (UNCA), en el marco del sistema de investigación, desarrollo e innovación. Además, los datos a procesar se obtuvieron de las actividades virtuales realizadas por los estudiantes que cursan las asignaturas seleccionadas en diferentes sedes (San Fernando del valle de Catamarca, Departamento Ancasti y San Miguel de Tucumán). Los docentes del equipo de investigación realizan sus actividades en los campus universitarios de origen y están a cargo de asignaturas de las carreras del área informática de la FACEN de la UNCA, y de la Facultad de Ciencias e Ingeniería de la Universidad Nacional del Sur (UNS).

1. INTRODUCCIÓN

Sin duda el año 2.020 afectó profundamente la vida de las personas a nivel global, produciendo cambios que parecían imposibles e impensados hasta el momento [8]. Algunos aspectos tan importantes como la educación se vieron afectados profundizando y acelerando procesos tales como la virtualización haciendo que las instituciones educativas se aferren a él como si fuera la panacea [9]. Como una consecuencia lógica de este proceso se produjo el inesperado crecimiento en la generación y almacenamiento de datos que representan las actividades virtuales (únicas posibles) en entornos de aprendizaje mientras el estudiante utiliza las diversas plataformas elegidas por cada institución educativa [10]. Además, otra consecuencia no deseada la representa la deserción, aunque si bien es un problema compartido por todos los niveles educativos, en particular en el universitario se vio incrementado de forma preocupante durante el proceso pandémico sufrido [11]. Se debe tener en cuenta que además de los factores que afectaron en años anteriores ahora se les suman otros nuevos como la imposibilidad de acceder a una conexión de internet, no contar con los dispositivos adecuados para el desarrollo de las actividades por nombrar algunos [9]. Esta situación viene siendo analizada desde años anteriores y desarrollando diferentes acciones en pos de lograr una disminución efectiva de la misma [11]. Hoy en día se presenta la oportunidad, como nunca antes, de analizar cada acción realizada por los estudiantes mientras se desenvuelven en las plataformas virtuales de aprendizaje. El volumen de datos generado es tan detallado y voluminoso que sólo puede ser procesado utilizando herramientas de Data mining y Big data [2].

Es de esperar que las instituciones educativas deseen abordar la deserción antes de que ésta ocurra concretamente. Para lo cual es necesario realizar un proceso de predicción de cómo será el desempeño de los estudiantes basado en datos iniciales e intermedios que permitan detectar a aquellos que se encuentran en riesgo de abandono. Esta predicción se

considera un aviso que permitirá al docente actuar adecuadamente apoyando a estos estudiantes en temas puntuales que le permitan influir positivamente en su desempeño.

El data mining está compuesto por numerosas técnicas que tienen como objetivo el buscar patrones e información relevante y oculta dentro de los datos almacenados en grandes bases de datos [6]. Actualmente los diferentes programas disponibles para la aplicación de las técnicas de data mining facilitan la tarea del procesamiento de grandes volúmenes de datos. En particular los datos educativos representan un campo especial para su procesamiento y la búsqueda de patrones tanto en las actividades como en el análisis de los resultados intermedios obtenidos por los estudiantes. Dentro de los datos disponibles se encuentran detalles tan precisos como la cantidad de accesos a los materiales virtualizados, tiempo de permanencia en foros, participaciones, creación de temas nuevos, aportes a las tareas en los grupos, cantidades de consultas y respuestas de preguntas realizadas por otros estudiantes, entre muchos otros. Antes del procesamiento de estos datos se debe realizar un proceso de adecuación, como por ejemplo: borrado de datos incompletos, erróneos o innecesarios y la selección de aquellos atributos que se consideren más representativos del perfil del estudiante. Del procesamiento de los datos puede extraerse información que correctamente interpretada seguramente favorecerá el proceso educativo y servirá como guía al docente en la toma de decisiones. Esta rama del data mining que procesa datos educativos se denomina Educational data mining (EDM)[1] que se aplicará a datos relacionados con el desempeño académico generados en plataformas Learning Management System (LMS).

El análisis de datos educativos puede realizarse desde distintos enfoques de acuerdo a las técnicas que se apliquen, estos son: descripción [3], predicción [4], segmentación [5].

Para este caso particular, las técnicas que se agrupan en la predicción tienen como objetivo el establecer un modelo que represente a los

datos y que permita estimar los valores que pueden tomar a futuro las variables analizadas, en el caso de su aplicación a datos educativos los atributos iniciales de los estudiantes son analizados por estas técnicas obteniéndose resultados que le predicen al docente como puede llegar a ser el desempeño académico durante las actividades a realizar. Estas predicciones se basan en el análisis de los datos históricos, iniciales y resultados intermedios que van obteniendo los estudiantes durante el desarrollo del contenido de la asignatura [4].

Unas de las técnicas más utilizadas para la predicción son las redes neuronales artificiales (RNA), constituyen algoritmos computacionales inspirados en las redes neuronales biológicas, comportándose de forma similar a las mismas en sus funciones más comunes. Las RNA son entrenadas para aprender, generalizando resultados y extrayendo patrones de un conjunto de datos [7].

Los investigadores vienen trabajando en el desarrollo de nuevos prototipos de redes neuronales basados en el tipo de problema a resolver. Para los problemas de predicción las arquitecturas más utilizadas son las denominadas perceptrón de múltiples capas, en ellas se produce el aprendizaje mediante un proceso denominado “retro propagación del error”, que se basa en obtener el mínimo de una función de error calculada a través de la salida del sistema y el modelo obtenido por el entrenamiento de la red [7].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo presenta las expectativas en cuanto a los resultados que se obtendrán de las experimentaciones previstas para su realización sobre la base de datos de los estudiantes de carreras de informática de primer año. Esta base se encuentra en la etapa de preprocesamiento y determinación de los variables y atributos más representativos de la performance desarrollada por los estudiantes en las asignaturas con contenidos básicos de programación. Representa parte de las

actividades a realizar en la tesis de maestría en Ciencias de la Computación, en donde se investiga los aportes del data mining como disciplina, al proceso educativo. Particularmente esta herramienta se orienta a prevenir la deserción utilizando su característica predictiva.

Si bien los resultados obtenidos serán presentados dentro de la tesis de maestría, posteriormente las experiencias se aplicaran en las actividades de asignaturas que aportaron los datos necesarios al proceso. En general el tema del procesamiento de los datos se viene analizando en distintos trabajos y presentaciones en congresos y reuniones científicas que van demostrando los avances y aportes encontrados mientras se desarrollan las actividades previstas en el plan de tesis.

Los objetivos principales son:

- Demostrar la capacidad predictiva de las redes neuronales artificiales en el procesamiento de datos educativos.
- Promover la utilización de esta herramienta como apoyo a la disminución de la deserción en las carreras de informática de la FACEN-UNCA.
- Comprender los procesos mentales que permiten a los estudiantes la construcción de conocimiento en programación.
- Experimentar con la selección de atributos que representen fehacientemente el perfil del estudiante de carreras de informática.

3. METODOLOGÍA

Con esta investigación se busca proponer a las redes neuronales artificiales como una herramienta que permitirá realizar predicciones sobre el rendimiento futuro de los estudiantes basados en datos históricos, iniciales e intermedios. Además de dar una base cierta para la toma de decisiones por parte de los docentes en cuanto a varios aspectos importantes del proceso educativo de modo que permitan una disminución concreta de la deserción de estudiantes en carreras de informática de la FACEN-UNCA.

El estudio de los resultados que se obtengan de la aplicación de las RNA sobre datos educativos se estima que permitirá intervenir precisamente sobre los estudiantes detectados para mejorar su desempeño y evitar la deserción de los mismos. Particularmente para en este estudio se utilizarán datos de asignaturas con contenidos básicos de programación de las diferentes carreras y subsedes de la FACEN-UNCA. Como corolario de esta investigación se estima el poder aportar una herramienta más para abordar la deserción desde una mirada científica que pueda ser aplicada por docentes para confirmar su validez.

4. FORMACIÓN DE RECURSOS HUMANOS

Algunos de los autores del trabajo se encuentran en la etapa de desarrollo de sus tesis de posgrado en carreras relacionadas con el tema de investigación, como la Maestría en Ciencias de la Computación en donde el Mgter. Sosa Marcelo Omar Diógenes desarrolla actualmente la tesis denominada: ***“Aportes de data mining a la mejora del proceso educativo con blended learning: formalización y experimentaciones”*** que se realiza con la dirección del Dr. Chesñevar Carlos Iván perteneciente a la Universidad Nacional del Sur (UNS). Además el investigador se encuentra en la etapa de planificación de su tesis doctoral en el área de minería de datos en el Doctorado en Ciencias dictado en la (FACEN) en convenio con la (UNS).

La Docente Investigadora Esp. Sosa Bruchmann Eugenia Cecilia desarrolla su tesis en la carrera Maestría en Ingeniería en Software de la Universidad Nacional de San Luis denominada ***“La experiencia del usuario desde un nuevo enfoque para el desarrollo de productos interactivos: el comportamiento emocional del usuario y la importancia de los atributos estéticos”*** dirigida por el Dr. Germán Montejano. Los docentes investigadores desarrollan actividades de dirección de tesis de la carrera de Licenciatura en Tecnología Educativa de los siguientes estudiantes: Varela

Marino del Valle cuya tesis se denomina ***“Análisis del impacto de un aula virtual en el proceso de enseñanza y aprendizaje en la escuela de educación técnica N° 7 José Alsina Alcobert”***, Tolaba Rodrigo Alejandro desarrolla la tesis denominada: ***“Aplicación del modelo Van Hiele a la enseñanza de triángulos con geogebra”***. y Santillán Mario de la carrera Especialización en docencia universitaria en disciplinas tecnológicas donde desarrolla la tesis denominada : ***“Diseño de simulador para el desarrollo de destrezas de programación”***

Además desarrollan las siguientes actividades:

- Dirección de proyectos de investigación denominado: “Digitalización de materiales didácticos para la enseñanza de la programación”.
- Investigadores pertenecientes a los centros de investigación e innovación de la FACEN-UNCA.
- Producción de artículos científicos para su presentación en congresos locales, nacionales e internacionales.
- Dictado y participación en cursos de actualización sobre los diferentes temas de investigación de área de cada docente.
- Integrantes de la revista de ciencias de la Facultad de Ciencias Agrarias de la UNCA.
- Actualización y capacitación permanente de los investigadores en talleres o workshop relacionadas con el tema del trabajo.
- Participación de los investigadores como consultores en proyecto afines que se desarrollan en la FACEN en distintas áreas.
- Examinadores de trabajos finales en las diferentes carreras que se dictan en la FACEN- UNCA.
- Dirección de tesis y tesinas finales de las carreras de computación, informática y Licenciatura en tecnología educativa.
- Planificación de seminarios para docentes en temas relacionados con la investigación.

- Participación en convenios con la Facultad de Tecnología y Ciencias Aplicadas para el desarrollo de estudios del área de datamining.

5. CONCLUSIÓN

Se espera que los resultados obtenidos de estas experimentaciones presenten a las herramientas de data mining como una alternativa al procesamiento de datos educativos. En cuanto a las RNA, se espera que los resultados que se obtengan guíen las actividades de los docentes enfocándose en los temas y estudiantes que según los resultados presentan problemas. Además que faciliten la disminución de manera concreta la deserción de los estudiantes de las carreras de informática de la FACEN-UNCA una vez implementadas en forma concreta en las asignaturas establecidas.

• 6. REFERENCIAS

1. Romero C. and Ventura S., "Educational data mining: A Survey From 1995 to 2005", *Expert System with Applications*, vol. 33, pp. 135-146, 2007.
2. Lavrac, N., Kavsec, B., Flach, P. and Todorovski, L., "Subgroup discovery with CN2-SD". *Journal of machine learning research*. 2004.
3. Jain A.K. and Dubes R.C. "Algorithm for clustering data. 1998. Englewood Cliffs. N.J. Prentice Hall.
4. Fayyad, U, Piatetsky-Shapiro, G. and Smyth, P., "The KDD process for extracting useful knowledge from volumes of data". *Communication of ADM* 1996.
5. Agrawal, R., Imielinski, T. and Swami, A.N. "Mining Association Rules between set of item in large databases". In *International conference on management of data*. 1993. Washington D.C. ACM Press.
6. Molina Félix L., "Data mining: torturando los datos hasta que confiesen". *Accedido en línea* 2 marzo de 2016.
7. Del Brío B. Martín, Sanz Molina A., "Redes neuronales y sistemas borrosos"; pról. de Lotfi A. Zadeh. January 2006. Publisher: RA-MA (Madrid, España), ISBN: 978-84-7897-743-7
8. Ordorika, I. (2020). *Pandemia y educación superior*. *Revista de la educación superior*, 49(194), 1-8.
9. Peñuelas, S. A. P., Pierra, L. I. C., González, Ó. U. R., & Nogales, O. I. G. (2020). *Enseñanza remota de emergencia ante la pandemia Covid-19 en Educación Media Superior y Educación Superior*. *Propósitos y Representaciones*, 8(SPE3), 589.
10. Morado, M. F., & Ocampo Hernández, S. (2019). *Una experiencia de acompañamiento tecno-pedagógico para la construcción de entornos virtuales de aprendizaje en educación superior*. *Revista Educación*, 43(1), 43-61.
11. Martínez, C. R. V., & Pérez, M. C. R. (2007). *La deserción estudiantil en educación superior a distancia: perspectiva teórica y factores de incidencia*. *Revista Latinoamericana de Estudios Educativos (México)*, 37(3-4), 107-122.

7. BIBLIOGRAFÍA Y TRABAJOS CONSULTADOS

- Brijesh Kumar Baradwaj and Saurabh Pal, "Mining Educational Data to Analyze Students' Performance," in *International Journal of Advanced Computer Science and Applications, India*, 2011, pp. Vol. 2, No. 6.
- Fischer Angulo Erwin Sergio. "Modelo para la automatización del proceso de determinación de riesgo de deserción en estudiantes universitarios". Santiago de Chile 2012.
- González, J. C., Ramos, S., & Hernández, S. (2017). *Modelo Difuso*

- del Rendimiento Académico Bi-explicado. Revista de Sistemas y Gestión Educativa*, 55–64.
- Grossi María Delia. "Reglas de predicción aplicables al diseño de un curso de computación". Marzo 2008.
 - 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.
 - Solarte Martinez, Guillermo Roberto, Ocampo S., Carlos Alberto. *Técnicas de clasificación y análisis de representacion del conocimiento para problemas de diagnósticos. Cientia Et Technica [En Linea] 2009, Xv (Agosto-Sin Mes) : [Fecha De Consulta: 2 de marzo de 2016]*.
 - Usman, O., & Adenubi, A. (2013). *Artificial Neural Network (ANN) Model for Predicting Students' Academic Performance. Journal of Science and Information Technology*, 23–37.
 - Ventura Cristoban Romero, Sebastian, Nykola Pechenizkiy and Rayan Beker. "Handbook of educational data mining".Chapman & Hall CRC press. 2011.