

Aplicación de Analítica de Datos en espacios de Educación Superior

Gustavo Illescas¹, Elías Todorovich¹, Claudio Aciti¹, Guillermo Rodríguez², Pia Silvestrini³

¹Instituto de Investigación en Tecnología Informática Avanzada (INTIA-Centro Asociado CIC, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA).

²Instituto Superior de Ingeniería de Software Tandil (ISISTAN-CONICET), Facultad de Ciencias Exactas, UNCPBA.

³Instituto de Investigación en Ciencias de la Salud (IICS), Facultad de Ciencias de la Salud, UNCPBA.

Paraje Arroyo Seco, Tandil (B7001BBO), Argentina.

{illescas, etodorov, caciti, grodri}@exa.unicen.edu.ar, pia.silvestrini@salud.unicen.edu.ar

RESUMEN

El objetivo de esta investigación es realizar una Analítica de datos para la toma de decisiones sobre la información disponible en bases de datos de educación superior, en particular la información vinculada a la unidad académica donde se radica este proyecto (Facultad de Cs. Exactas, UNCPBA). El foco principal de análisis estará puesto sobre los procesos de ingreso, permanencia y egreso, como así también en el estudio y aplicación de las técnicas subyacentes dentro de las aplicaciones de *Big Data* y *Machine Learning*. Las instituciones de educación superior por naturaleza son generadoras de grandes volúmenes de datos que por lo general son de uso administrativo. Sin embargo, esas bases de datos contienen una riqueza de información para la toma de decisiones que proponemos explorar y aprovechar según lo expresado en nuestra propuesta.

Palabras clave: *Analítica de Datos, Educación Superior, Indicadores, Optimización*

CONTEXTO

Esta investigación se enmarca dentro del proyecto de incentivos 03/C314 titulado “Gestión Informática del Conocimiento como soporte para la toma de decisiones Organizacionales” y se desarrolla en el Instituto de Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas

(EXA), Universidad Nacional del Centro de la provincia de Buenos Aires (UNCPBA). Colaboran con el proyecto miembros de Universidades extranjeras:

-Laboratorio en Ingeniería de Software de la Universidad Carlos III (UC3M), Madrid, España.

-Universidad de Celaya, México.

-Universidad Autónoma de Coahuila, México.

1. INTRODUCCIÓN

Las Bases de Datos de Educación Superior (BDESUP), en particular la de EXA-UNCPBA, como lo muestra la Figura 1, están conformadas por un gran volumen de datos provenientes de los diferentes sistemas de gestión como lo son, entre otros, el sistema de alumnos (Siu-Guaraní), el de evaluación docente (Kune), el de planta funcional (Majen), el de encuestas (Siu-Kolla) (Comunidad SIU), el sistema de indicadores y el Aula virtual Moodle.

Por otro lado, existen numerosas actividades que se registran en bases de datos ad-hoc o archivos independientes como consecuencia de la falta de herramientas de software adecuadas a esos fines y que luego son de difícil interpretación a una escala global (cuando se intenta generalizar), con lo que sirven como soporte para situaciones específicas o personalizadas.

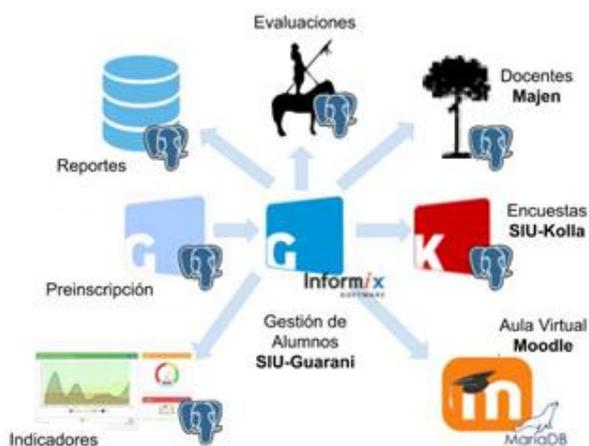


Figura 1. Esquema de Sistemas y Bases de datos disponibles en EXA. Fuente: área TICs EXA

En cuanto a los sistemas institucionales, existe la tendencia a reflejar la información a partir de la construcción de listados (también llamados reportes o informes) que son utilizados por los administrativos y equipos de gestión para realizar consultas (o filtrados) sobre el conjunto de datos disponibles y que de cierta forma son las herramientas que también se utilizan para la toma de decisiones en las unidades académicas (UA). No obstante, las BDESup contienen abundante información que no se refleja en estos listados y que generalmente no son aprovechadas para la generación de nuevo conocimiento. Por eso, la propuesta es utilizar un conjunto de técnicas y herramientas que permitan explorar todos los datos y detectar tanto información oculta como así también patrones en los datos que asistan a los tomadores de decisiones: los miembros de equipos de gestión, secretarías, directores de departamentos, coordinadores de carreras, entre otros actores de las UA, integrando así investigación con procesos de gestión.

Dadas las numerosas técnicas existentes, en una primera instancia, proponemos indagar sobre un conjunto limitado. Esto es, por un lado, las provistas por la minería de datos: patrones ocultos (Miller T. 2015; Siegel E. 2016), reglas de asociación (Concessao R. 2016), procesamiento analítico en línea (Hernández García C. et al 2016), *clustering* (Miranda M. et al 2017); y por otro, las técnicas de la optimización de la información provenientes de la rama de la estadística (Papoulis A. et al 2002) y de la investigación

de operaciones: Redes Bayesianas (Hernández Leal P. F. 2011; Oviedo B. et al 2015; Vegega C. et al 218), Cadenas de Markov (Bañuelos Saucedo A. L., 2013; Eugenia, Á. R. 2000), Procesos Jerárquicos Analíticos (Saaty, T. 2008), Árboles de decisión (Gutiérrez Pizarro C. 2014; Morrison D. 2007).

Por otro lado, a partir de la crisis sanitaria, los espacios educativos se han visto en la necesidad de utilizar plataformas de aprendizaje para interactuar activamente con sus estudiantes. Entornos educativos, tales como Moodle, se han masificado a tal punto de ser el principal medio de desarrollo académico. Esta superpoblación de uso ha expuesto una falta de evolución en el sector, el cual desaprovecha las nuevas tecnologías interactivas. En este sentido se propone realizar un asistente virtual (Carlos et al 2021; Chinedu, Ade-Ibijola 2021) que complemente activamente la utilización de la plataforma Moodle, contando con diversas funciones adecuadas al rol de quien lo utilice en la plataforma. El proyecto presentado a una beca EV-CIN se denomina Virtual Moodle Bot (VMB): Un Bot virtual en Moodle para asistir y complementar actividades académicas en espacios de educación superior.

Metodología

Todo proyecto de investigación conlleva un profundo proceso creativo para generar ideas que den solución a las preguntas de investigación que lo motivan, es por ello, que para el desarrollo de este proyecto se propone utilizar el método de trabajo conocido como *Design Thinking* (Brown T. 2009), el cual, es un método de trabajo en equipo que propicia la colaboración y la entrega frecuente de resultados a través de varias iteraciones.

Cuando se desarrolla un proyecto de carácter interdisciplinario, como el que se presenta en esta propuesta (investigadores de distintos institutos, miembros de equipos de gestión, no docentes de distintas secretarías), puede llegar a ser un reto aprovechar el talento que cada persona puede aportar desde su área de experiencia. Este reto, es otro motivo para emplear esta metodología como marco de trabajo, ya que promueve la colaboración y a

aportación de ideas a través de los siguientes principios:

- a) Definición clara de roles de trabajo.
- b) Delimitación de fases de trabajo.
- c) Definición de pautas de comportamiento.
- d) Diseño de soluciones centrado en el usuario (o la persona).
- e) Desarrollo iterativo e incremental.

Si bien el método fue diseñado en el ámbito del diseño industrial, en la actualidad, es un marco de trabajo que también se está empleando en el desarrollo de proyectos de investigación y en el desarrollo de programas académicos debido a los resultados demostrados en la generación de nuevas ideas e innovaciones.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El proyecto de incentivos que contiene esta línea de investigación comenzó a desarrollarse a mediados del 2021 luego de recibir la aprobación del organismo correspondiente. El estado de avance está relacionado con el estado actual del conocimiento sobre el tema donde los integrantes y colaboradores han realizado experiencias en la temática que han sido reportadas en diferentes comunicaciones desde 2014 a la fecha. A partir de estas experiencias es que surge la iniciativa para formular la propuesta e integrarse como equipo de trabajo. El foco principal de análisis estará puesto sobre los procesos de ingreso, permanencia y egreso siendo los dos últimos los que mayor cantidad de datos poseen dado que registran toda la actividad estudiantil. Sobre esto último, una gran ventaja es la constitución reciente de áreas que se ocupan de estas problemáticas en varias UA de la UNCPBA, lo que las convierte en las principales beneficiarias de la investigación propuesta.

Los antecedentes del equipo de trabajo están vinculados con las experiencias e investigaciones personales que los integrantes y colaboradores han desarrollado vinculadas a la temática propuesta.

-El Dr. Gustavo Illescas ha participado en actividades de EXA en el área de Coordinación de Asuntos Estudiantiles y Graduados (CAEG), en la Sec. General y en la Sec. de

Extensión. Rol: coordinador del proyecto de implantación de un sistema de indicadores (2018-actualidad).

-El Dr. Todorovich ha participado como director en el Proyecto de Mejoramiento de la Enseñanza en Carreras de Informática (PROMINF, 2013-2017), donde uno de los ejes de trabajo fue el mejoramiento de indicadores de ingreso, permanencia y promoción para la carrera de Ingeniería de Sistemas (EXA). Para ello, se puso en marcha un programa con tutores pares, que luego fue evolucionando y se relevaron los resultados para construir esos indicadores que marcaron el avance y resultado del proyecto.

-El Mag. Claudio Aciti es el actual Vicedecano y secretario general de EXA (desde 2017) donde una de las áreas que coordina es la CAEG que se encarga del seguimiento del ingreso, permanencia y egreso.

-El Dr. Guillermo Rodríguez ha participado en numerosos artículos relacionados a las técnicas de *Educational Data Mining*.

-La Médica Veterinaria María Pía Silvestrini es co-directora del proyecto de investigación "Desarrollo de un sistema de indicadores para el monitoreo del desempeño académico en las carreras de Lic. en Enfermería y Medicina. ESCS-UNCPBA".

3. RESULTADOS OBTENIDOS/ESPERADOS

A pesar de que la investigación se desarrolla en una UA en particular, la investigación constituirá un experimento trasladable al resto. Por otra parte, se espera que los resultados de las investigaciones colaboren con las problemáticas referidas a la Responsabilidad Social Universitaria (RSU) aportando indicadores como consecuencia de la aplicación de las técnicas mencionadas precedentemente.

Se listan a continuación algunos de los trabajos en desarrollo especificando la técnica utilizada:

- Análisis del rendimiento académico en la asignatura Investigación Operativa. Clasificador de Bayes.

- Deserción académica (abandono) en la carrera de Ing. de Sistemas. Análisis multivariado.
- Deserción académica (abandono) en la carrera de Ing. de Sistemas. Reglas de asociación.
- Análisis de recusantes en la asignatura Investigación Operativa. Estadística básica (dispersión, correlación, regresión lineal) y Análisis multivariado.
- Análisis de permanencia en la carrera de Ing. de Sistemas. *Educational Data Mining*.
- Análisis del desgranamiento universitario en las cohortes estudiantiles de Ing. de Sistemas. Cadenas de Markov de orden uno con técnicas de validación.
- Desarrollo de un tablero de control con Indicadores de deserción. Inclusión de métodos de pronóstico.
- Estudio de la retención/deserción en Ing. de Sistemas. Técnicas de análisis causal.

Para el caso puntual de los trabajos donde se aplican datos sobre la asignatura de grado Investigación Operativa, se pretende analizar los resultados de estos estudios preliminares a fin de evaluar si es posible generalizarlo hacia otras asignaturas de la carrera Ing. de Sistemas.

4. FORMACIÓN DE RECURSOS HUMANOS

La estructura del equipo de trabajo se muestra en la siguiente tabla:

Apellido y nombre	Título	Cargo	Funciones
Aciti, Claudio	Mg.	Prof. UNCPBA	Integrante
Illescas, Gustavo	Dr.	Prof. UNCPBA	Director
Rodríguez, Guillermo	Dr.	Ayud. UNCPBA	Integrante
Silvestrini, Pía	Med. Vet.	Prof. UNCPBA	Integrante
Todorovich, Elías	Dr.	Prof. UNCPBA	Integrante

-El Mg. Claudio Aciti es doctorando en el Doctorado en Matemática Computacional e Industrial (DCMI) de EXA bajo la dirección del Dr. Todorovich E.

-La Médica Veterinaria María Pía Silvestrini (perteneciente al IICS) ha presentado su plan de doctorado en “Interacción entre aprendizaje colaborativo, aprendizaje basado en proyectos y educación centrada en la comunidad como estrategia para la formación por competencias del estudiante de medicina”. Directores: Marín G - Illescas G.

Becarios

- Indicadores de vinculación social en la carrera de Medicina UNICEN (2021-2022) Beca Estímulo a las Vocaciones Científicas (Ev-CIN). Dirección: Silvestrini M. P. Becario: Juan Manuel Ciancio Pagano.
- Análisis de la trayectoria educativa en Salud Pública mediante minería de datos (*Educational Data Mining*). Beca Iniciación a la investigación (Fortalecimiento UNCPBA 2021-2022) Dirección: Silvestrini M. P. Becario: Gimena Dietrich. Título.
- Análisis de Redes Sociales (ARS) para la mejora de las prácticas socioeducativas. Beca Ev-CIN (presentada 2022). Dirección: Silvestrini M. P. Becario: Juan Emilio Ciancio Pagano.
- Virtual Moodle Bot (VMB): Un Bot virtual en Moodle para asistir y complementar actividades académicas en espacios de educación superior. (presentada 2022). Dirección: Illescas G., Rodríguez G. Becario: Herrera De Rosa Exequiel.

Tesis de grado

- Herramienta web para el análisis y descubrimiento de indicadores. Dirección: Illescas G. Tesista: Juan Disteffano.
- Análisis del rendimiento académico en la carrera de Ing. de Sistemas utilizando clasificadores (Random forest, K-nn vecinos más cercanos, Support vector machine y Bayes). Dirección: Illescas G., Mora-Soto A (Univ. Celaya MEX.). Tesistas: Maylen Dell Oso, Florencia Paglione.

5. BIBLIOGRAFÍA

- Bañuelos Saucedo, A. L. (2013). Análisis del avance escolar a nivel licenciatura aplicando cadenas de Markov. Universidad Nacional

- Autónoma de México (UNAM). Tesis Maestría.
- Brown, Tim (2009). *Change by Design, How Design Thinking Transforms Organizations and Inspires Innovation* Ed. Harper Collins New York ISBN 978-0-06-193774-3
 - Carlos, H.; German, S.; Dixon, S. (2021). *Tashi-Bot: A Intelligent Personal Assistant for Users in an Educational Institution*. 10.20944/preprints202108.0380.v1.
 - Chinedu O., Ade-Ibijola A. (2021). *Chatbots applications in education: A systematic review*. *Computers and Education: Artificial Intelligence*. 2. 100033. 10.1016/j.caeai.2021.100033.
 - Concessao R. (2016) *Predictive Data Analytics: for informes strategic decisions*. Editor: CreateSpace Independent Publishing Platform. Edición: 5.0. 2016.
 - Eugenia, Á. R. (2000). *Modelo markoviano para el estudio de evolución de cohortes de estudiantes de un programa académico*. Red de revistas científicas de América Latina y Caribe, España y Portugal.
 - Gutiérrez Pizarro C. (2014). *Estudio de la permanencia y deserción de los estudiantes de pregrado en una Facultad de Ingeniería, utilizando una metodología mixta de Investigación y árboles de decisión en su predicción*. Tesis doctoral Univ. de Alcalá.
 - Hernández García C., Rodríguez Rodríguez J. (2016). *Algoritmo híbrido basado en aprendizaje computacional para el manejo de datos faltantes en aplicaciones OLAP Ingeniare*. *Revista chilena de Ing.*
 - Hernández Leal P. F. (2011). *Algoritmo de Aprendizaje para Redes Bayesianas de Nodos Temporales*. Instituto Nacional de Astrofísica, Óptica y Electrónica. Ed. INAPE. México.
 - Miller T. (2015), *Marketing Data Science: Modeling Techniques in Predictive Analytics with R and Python*, Editorial Pearson Education, versión impresa y Kindle.
 - Miranda M y Guzmán J. (2017). *Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos-Form*. Univ. vol.10 no.3 La Serena. Chile
 - Morrison D. (2007). *A Naive Bayes: Decision Tree Hybrid*
 - Oviedo B., Moreno A., Puris A., Villacís A., Delgado D. (2015). *Análisis de datos educativos utilizando Redes Bayesianas*. 13th Latin American and Caribbean Conference for Engineering and Technology. Santo Domingo, República Dominicana.
 - Papoulis A., Pillai U. (2002). *Probability Random Variables and Stochastic Processes*. McGraw-Hill.
 - Saaty, T. 2008. *Decision Making with the Analytic Hierarchy Process*, Int. J. Services Sciences, Vol. 1, Nº. 1, pp.83-98
 - Siegel E. (2016), *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Ed. Wiley.
 - SIU (Sistema de Información Universitaria) (2021) <https://portal.comunidad.siu.edu.ar/>
 - Vegega C., Deroche A., Pytel P., Ramón H., Straccia L., Acosta M., Pollo-Cattaneo M (2018). *Diagnóstico del Proceso de Aprendizaje de Alumnos de Inteligencia Artificial mediante un Modelo Dinámico Bayesiano*. *Rev. Tecnología y Ciencia*. UTN.