

Evaluación de técnicas de Data Mining para la obtención de perfiles de ingresantes a la UNNOBA

Monica Sarobe, Ana Smail, Maria Rosana Piergallini, Mercedes Guasch, Marina Rodriguez, María Lanzillota, Mariana Adó, Claudia Banchoff

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44

{monica.sarobe, ana.smail, rosana.piergallini, mercedes.guasch, marina.rodriguez, maria.lanzillota, mariana.ado, cbanchoff}@itt.unnoba.edu.ar

Resumen

El ingreso a la Universidad es siempre un desafío para el aspirante como para la institución que lo albergará.

Durante la inscripción a las propuestas académicas de grado y posterior desarrollo del curso de ingreso la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA) recolecta información a través de sus sistemas de gestión. Esta información representa un importante activo para la Universidad, en la medida en que pueda ser transformada en conocimiento ya sea para la toma de decisiones que conciernen a la población estudiantil, como para el diseño de planes de acción.

La presente línea de investigación propone procesar los datos recolectados por los sistemas de gestión de la UNNOBA durante el ingreso, para obtener en principios patrones o semejanzas entre los aspirantes con idénticos resultados en el curso de ingreso. Esto permitirá entre otras cosas

determinar estrategias de aprendizajes adaptativos.

Palabras clave: KDD, Data Mining

Contexto

Esta línea de investigación forma parte del proyecto “Herramientas didácticas y nuevas formas de interacción para la inclusión de TICs en la enseñanza” aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación BIANUALES (SIB2015).

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

Introducción

La Universidad Nacional del Noroeste de la Provincia de Buenos Aires

(UNNOBA) fue creada en el año 2004, comenzando a dictar sus carreras de grado en el año 2005. Este mismo año implementó por primera vez el dictado de un curso de ingreso de modalidad presencial. Luego, a partir del año 2010 se sumó la modalidad semipresencial; y posteriormente, a partir del año 2013, se incorpora la opción de rendir una evaluación de conocimientos. Dicha evaluación, permite a los aspirantes a carreras validar sus conocimientos mediante la aprobación de un examen, quedando así exceptuados realizar el curso de ingreso.

Los aspirantes expresan la voluntad de ingresar a la Universidad realizando una preinscripción a carreras en el sistema SIU-Preinscripción. Este sistema recolecta datos demográficos, sociales, económicos y familiares en una base de datos relacional. Para aquellos alumnos que no se encuentran exceptuados del curso de ingreso, se cuenta con los resultados obtenidos en las dos materias que lo componen en cada una de las modalidades (presencial, semipresencial o evaluación de conocimientos) también en una base de datos relacional.

Los datos recolectados durante el proceso de ingreso en todos estos años conforman hoy un activo de potencial relevancia dado que posibilitan obtener patrones de los ingresantes a fin de poder establecer un perfil de los mismos.

La comunidad educativa de la universidad considera de real importancia la posibilidad de contar con una descripción o caracterización de los ingresantes a través de perfiles, dado que un mayor conocimiento de los mismos puede influir positivamente en la toma de decisiones concernientes al ingreso como también en el diseño de estrategias de enseñanza y aprendizaje.

Establecer perfiles es una actividad muy difundida en muchas áreas

disciplinarias, es análoga al proceso de determinar y clasificar patrones. El proceso de Extracción de Conocimientos a partir de datos (KDD, Knowledge Discovery from Databases) es un proceso no trivial para identificar patrones válidos, novedosos y potencialmente útiles a partir de los datos. Este busca el descubrimiento de conocimiento sin una hipótesis preconcebida. Involucra la preparación de los datos, su análisis o minería de datos e interpretación de los resultados. [6]

La Minería de Datos (Data Mining), involucra técnicas necesarias para la construcción de modelos a partir de la información disponible. Este proceso puede utilizar dos enfoques bien definidos: descriptivo o predictivo. En el primer enfoque, se intenta caracterizar las situaciones de manera tal de poder entender cuáles son las variables principales que describen el comportamiento de una situación particular. El segundo enfoque implica la utilización del modelo para establecer de antemano una situación problemática. [3]

En el caso de la determinación de perfiles de los ingresantes, un modelo descriptivo puede explicar las variables que mejor caracterizan a los aspirantes a carreras de grado con idénticos resultados en el curso de ingreso y obtener una visión global de los mismos. La simple aplicación de una técnica de MD a una muestra representativa de los datos pre-procesada (vista minable), y el conocimiento previo del problema, no garantizan patrones expresivos, novedosos y útiles. [2] Por lo que parece tener sentido la evaluación de diferentes técnicas o algoritmos para obtenerlos.

Líneas de Investigación, Desarrollo e Innovación

La presente investigación se encuadra dentro del eje “Tratamiento masivo de

datos”. En este sentido se pretende procesar los datos recolectados de los sistemas de gestión de la Universidad durante el ingreso, para obtener en principios patrones o semejanzas entre los aspirantes con idénticos resultados en el curso de ingreso.

La investigación actual sigue las siguientes líneas:

- Evaluar técnicas de selección de características.
- Evaluar técnicas de para obtener modelos descriptivos.
- Evaluar el ensamble de algoritmos para potenciar el carácter descriptivo de los modelos.
- Evaluar técnicas de visualización de los resultados.

Resultados y Objetivos

Durante la primera etapa del proyecto, como parte del pre-procesamiento de datos, hemos obtenido una vista minable de los datos extraídos del Sistema de Preinscripción a la Universidad, el cual cuenta con datos demográficos, económicos, sociales, entre otros; junto con los resultados de las tres modalidades en que se dicta el curso de ingreso y su resultado final.

Las tareas a realizar en la segunda etapa se centraran en:

- Reducción de la dimensión del problema, mediante la selección de los atributos más representativos de la información disponible.
- Selección del algoritmo que produzca los patrones más expresivos.
- Validación de la solución obtenida.
- Ensamblar con algoritmos que describan y expliquen mejor el conocimiento obtenido.

Se espera que la presente línea de investigación contribuya aportar información implícita, previamente desconocida y potencialmente útil para la toma de decisiones que conciernen a la población estudiantil y para el diseño de planes de acción tanto preventivos como correctivos.

Formación de Recursos Humanos

Se cuenta con dos alumnos avanzados de la carrera Licenciatura en Informática que poseen becas de iniciación a la investigación otorgadas por la UNNOBA y se encuentra en desarrollo dos tesis de posgrado en la UBA.

Bibliografía

- [1] WICC 2014 - Detección Temprana de la Deserción en Cursos Virtuales - Russo, Sarobe, Lencina, López Gil, Carrattini, Smail, Piergallini, Guash, Alonso, Ramón
- [2] Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios – Formia, Lanzarini - 2012
- [3] Introduction to Data Mining – Tan, Steinbach, Kumar - 2006
- [4] Selección de atributos representativos del avance académico de los alumnos universitarios usando técnicas de visualización. Un caso de estudio – Lanzarini, Charnelli, Baldino, Diaz – 2015
- [5] Educational Data Mining by Using Neural Network – Upadhyay, Lucknow - 2016
- [6] Introducción a la Minería de Datos - Orallo, Quintana, Ramirez - 2004
- [7] Minería de datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería – Ahumada, Dip, Herrera, Leguizamón, Almendra - 2015