

PATRONES ESTADÍSTICOS RELACIONADOS CON EL PERFIL DEL ALUMNO DE LA FACULTAD DE TECNOLOGÍA Y CIENCIAS APLICADAS

María A. Barrera, M. Soledad Bustos Aguiar, Nancy Lamas, Daniela Lobos Anfuso, Manuel
Baquinza y Leguizamón María Belén

Departamento Sistemas/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca
Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 03834- 435112 – int 168

mbarrera@tecno.unca.edu.ar, soledadbustosaguiar@yahoo.com.ar, nancylamas@gmail.com,
danielalobosanfuso@gmail.com, mbaquinzay@gmail.com, & mbelenleguizamon@gmail.com

Resumen

El rendimiento académico estudiantil ha sido definido como el cumplimiento de las metas, logros u objetivos establecidos en el programa o asignatura que está cursando el alumno.

Se puede analizar el rendimiento académico como una medida de las capacidades del alumno, que expresa lo que éste ha aprendido a lo largo del proceso formativo.

Existen diversos motivos o factores que pueden llevar al alumno a mostrar un pobre rendimiento académico tales como (poca motivación, desinterés, distracciones en clase, etc.)

En este trabajo se propone aplicar un proceso de descubrimiento de conocimiento o KDD (Knowledge Discovery from Data Base) a la Base de Datos de los alumnos de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCa), con el propósito de convertir una masa de datos carente de significado en información relevante, es decir, en conocimiento. Dicho conocimiento vendrá dado mediante patrones, tendencias o relaciones.

Estos resultados pueden mejorar el proceso de formación académica, estimular fundamentalmente el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un sentido automatizado para la generación de conocimiento y elevando así la calidad de la educación en la Universidad.

Palabras clave: KDD (Knowledge Discovery

from Data Base, Data Minig, Patrones, Rendimiento Académico.

Contexto

El proyecto de investigación: “Patrones estadísticos relacionados con el perfil del alumno de la Facultad de Tecnología y Ciencias Aplicadas”, se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y es evaluado por la Secretaría de Ciencia y Tecnología de la universidad, donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática.

Así mismo es prioritario, tanto para el Departamento de Informática, como para los demás departamentos de la Facultad, transferir soluciones concretas a las problemáticas que se plantean a los docentes en relación con el rendimiento académico de los alumnos.

Utilizando los conceptos del proceso de KDD y sus técnicas de Data Minig, se decidió tomar como marco de referencia a la Facultad de Tecnología y Ciencias Aplicadas. La aplicación de estas técnicas constituye una herramienta o medio para lograr el objetivo del presente proyecto. Si bien en la Facultad se han puesto en práctica técnicas de DM, no existen antecedentes de aplicación de un proceso de KDD, involucrando a todo el alumnado.

Introducción

El rendimiento académico estudiantil desde un punto de vista operativo, se ha limitado a la expresión de una nota cuantitativa o cualitativa y se encuentra que en muchos casos es

insatisfactorio lo que se ve reflejado en la deserción (abandono de los estudios universitarios) y el desgranamiento (pérdida de continuidad y/o retraso en el avance) de los alumnos de la carrera

El rendimiento supone la capacidad del alumno para responder a los estímulos educativos. En este sentido, el rendimiento académico está vinculado a la aptitud.

Existen distintos factores que inciden en el rendimiento académico. Desde la dificultad propia de algunas asignaturas, hasta la gran cantidad de exámenes que pueden coincidir en una fecha, pasando por la amplia extensión de ciertos programas educativos. Son muchos los motivos que pueden llevar a un alumno a mostrar un pobre rendimiento académico.

Por otra parte, el rendimiento académico puede estar asociado a la subjetividad del docente cuando corrige. Ciertas materias, en especial aquellas que pertenecen a las ciencias sociales, pueden generar distintas interpretaciones o explicaciones, que el profesor debe saber analizar en la corrección para determinar si el estudiante ha comprendido o no los conceptos.

Como podemos ver, hay muchos motivos por los que el rendimiento del alumno se ve afectado. Es por eso que no podemos simplificar el análisis del mismo a sólo las calificaciones obtenidas en los exámenes a lo largo del cursado de una materia a través de métodos meramente estadísticos.

En la última década, ha existido un gran crecimiento en nuestras capacidades de generar y coleccionar datos en las más diversas áreas: el comercio, la banca, astronomía, física de partículas, química, medicina, departamentos de gobierno, entre otras. Esto debido al gran poder de procesamiento de las máquinas como a su bajo costo de almacenamiento. Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información "oculta", de gran importancia estratégica, a la que no se puede acceder por las técnicas clásicas de recuperación de la misma.

A partir de la aplicación de un grupo de técnicas de Data Mining (DM) como el clustering, árboles de decisión, algoritmos de aprendizaje inductivo, sumarización, modelización de dependencias, regresión,

entre otros, se pretende clasificar a los estudiantes encontrando patrones ocultos y reglas que los caractericen. Estos resultados pueden mejorar el proceso de formación académica, estimular fundamentalmente el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un sentido automatizado para la generación de conocimiento y elevando de esta manera la calidad de la educación en la Universidad.

Líneas de investigación y desarrollo

El proyecto aborda la aplicación de un proceso de descubrimiento de conocimiento o KDD a la Base de Datos de los alumnos de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

El proceso de KDD entre otras sofisticadas técnicas aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad.

Este proceso conlleva diversas actividades entre ellas interpretación de los datos sobre los que se va a trabajar, depuración de los mismos, aplicación de algoritmos que ofrece la técnica de DM y análisis de resultados. Hay que destacar que todo este proceso es interactivo e iterativo.

Resultados y Objetivos

El grupo de trabajo ha realizado algunos análisis planteados, pero no posee aún resultados finales, ya que el mismo ha comenzado su ejecución en enero del pasado año. Se tiene como objetivo proveer los conocimientos para la formación científica, tecnológica y complementaria, sobre las tecnologías de Gestión de Datos en Base de Datos (BD), buscando manipular estructuras de almacenamiento de grandes volúmenes de información para lograr acceso rápido y seguro a los datos. Por lo que se pretende investigar y experimentar nuevas técnicas y metodologías para lograr la agilidad en las distintas operaciones relacionadas a las BD. Esto permitirá obtener información, que con simples accesos o consultas a los datos no sería posible y se podrá, de acuerdo a cada

situación, elegir el algoritmo correcto para obtener resultados.

Objetivo General:

Determinar patrones estadísticos relacionados con el perfil del alumno de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

Objetivos Específicos:

- Determinar el rendimiento académico de los alumnos por año y carrera de la Facultad de Tecnología y Ciencias Aplicadas.
- Descubrir la incidencia en el rendimiento académico, teniendo en cuenta los entornos socio-económico, socio-familiar, socio-sanitario y habitacional de los alumnos.

Formación de Recursos Humanos

El proyecto cuenta con integrantes que se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto y con una becaria de Beca de Estimulo a las Vocaciones Científicas (CIN) 2012 otorgada por la Secretaria de Ciencia y Tecnología de UNCa. Además los mismos son docentes de las cátedras Gestión de Datos y Estructura de Datos de la carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas; estando ambas materias estrechamente ligadas al tema en cuestión y donde tienen por objetivo proveer los conocimientos para la formación científica, tecnológica y complementaria sobre las tecnologías de Gestión de Datos en Base de Datos, buscando manipular estructuras de almacenamiento de grandes volúmenes de información para lograr acceso rápido y seguro a los datos, por lo consiguiente siempre están abocados a investigar y experimentar nuevas técnicas y metodologías para lograr la agilidad en las distintas operaciones relacionadas a las BD.

Específicamente el proyecto de investigación prevé el programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología

y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.

- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.
- Participación en talleres o workshops de herramientas informáticas relacionadas con las BD y KDD.
- Celebración de convenios con otras universidades, para capacitación de los recursos humanos en KDD.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

Bibliografía

- CHAPMAN & HALL, *The Top Ten Algorithms in Data Mining*
- DOÑA J.M., QUINTANA O.P., VALESANI M.E., VALLEJOS O.A. *Analysis of Agregation Methods in Incomplete Database Systems. Information Processing and Management of Uncertainty in Knowledge-Based System (IPMU 2008)*. ISBN: 9978-84-612-3061-7
- YANG C. YUAN. *Multiple Imputation for Missing Data: Concepts and New Development.*, SAS Institute Inc., Rockville, MD. P267-25.
- PIATETSKY-SHAPIRO G. FRAWLEY WJ. (1991) - *Knowledge Discovery in Databases Ed. AAAI/MIT Press.*
- THURASINGHAM, B. (1999) – *Data Mining. Technologies, Techniques Tools and Trends – Ed. CRC Press LLC.*
- WITTEN, IAN H. AND FRANK EIBE (2005) – *Data Mining: Practical Machine Learning Tools and Techniques – Morgan Kaufmann 2° Edición.*
- JEAN - MICHEL – FRANCO (1997) – *El Data Warehouse. El Data Mining – Ed. Gestión 2000*