

## WICC 2014 XVI Workshop de Investigadores en Ciencias de la Computación

---

# ESTUDIO DE TECNICAS DE DATA MINING APLICADAS AL ANALISIS DE DATOS GENERADOS CON LA METODOLOGIA BLENDED LEARNING

Marcelo Omar Sosa, Sosa Bruchmann Eugenia Cecilia

Departamento 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

[sosamod1@hotmail.com](mailto:sosamod1@hotmail.com), [sosab\\_ec@hotmail.com](mailto:sosab_ec@hotmail.com)

## Resumen

La utilización de las nuevas tecnologías de la información y la comunicación (*Ntic*) en todos los niveles educativos y particularmente en el nivel universitario, permite el desarrollo de contenidos con la aplicación de la metodología de blended learning. Ésta representa el dictado combinado, con actividades desarrolladas en forma presencial y virtual sustentadas por la tecnología y la utilización de internet como medio de comunicación. Las ventajas comparativas de este tipo de metodología, se vienen estudiando desde diversos aspectos, concluyéndose que representa una metodología abarcativa y completa que utiliza de forma racional la tecnología en el aula.

Las actividades que permite esta metodología generan un gran volumen de datos relacionados con las acciones que desarrollan los alumnos en esta plataforma. Van desde el tiempo de permanencia en cada página hasta las secuencia de acciones relacionadas con las actividades propuestas. De este tipo de datos pueden extraerse conocimiento potencialmente útil que posteriormente puede ser utilizado para mejorar el diseño, tanto de materiales, la organización de actividades y evaluaciones entre otros aspectos. Para la obtención de este conocimiento es necesario la utilización de técnicas de data mining que permiten el procesamiento de grandes volúmenes de datos.

**Palabras clave:** Técnicas de data mining; blended learning.

## Contexto

El presente trabajo se encuentra dentro de las líneas de investigación en que se desarrollan las actividades de investigación del equipo. Actualmente los docentes están a cargo de las asignaturas Sistemas de información y comunicación, Simulación y Análisis comparativo de lenguajes (*Mgter. Sosa Marcelo Omar*) e ingeniería de software y Técnicas digitales (*Lic. Sosa Bruchmann Eugenia*) de las carreras de Profesorado en Computación y Tecnicatura en Informática de la Facultad de Ciencias Exactas y Naturales (*Fa.C.E.N.*) de la Universidad Nacional de Catamarca (*U.N.Ca.*). Estas carreras además son dictadas en las subsedes de la Provincia de Tucumán en las localidades de Concepción y en San Miguel de Tucumán donde los docentes están a cargo de las mismas asignaturas.

## Introducción

La aplicación en la educación superior de la metodología de blended learning genera una dinámica diferente a la hora desarrollar las actividades previstas dentro de los contenidos de las diferentes asignaturas [6]. Estas actividades generan un gran volumen de datos que se almacenan en distintos espacios destinados a ello. Entre ellos podemos mencionar los archivos log pertenecientes al servidor en el que se almacenan las interacciones entre el docente y los alumnos, además de los datos que se generan durante el desarrollo de las actividades propuestas por el

docente. Para ello se utilizan diferentes herramientas como el chat, foro de discusión, la descarga de materiales de consulta, el desarrollo de los trayectos de aprendizaje que incluye lectura de materiales, evaluaciones entre otros [4].

Basados en el número de alumnos y actividades que se proponen en las asignaturas analizadas (asignaturas con contenido de programación de las carreras de tecnicatura y profesorado en computación de la *Fa.C.E.N.*), nos encontramos con un volumen de datos muy elevado que no puede ser analizado por métodos tradicionales ya sean automatizados o manuales. Es por ello que las técnicas de la minería de datos (*data mining*) representadas por el modelo lógico que analiza los datos para la extracción del conocimiento, permiten obtener desde el punto de vista del usuario la visualización clara de los diferentes resultados de su aplicación. Este conocimiento representa un aporte importante a la toma de decisiones para el docente, permitiéndoles un mejoramiento en el proceso de diseño y evaluación tanto de contenidos, métodos de enseñanza y formas de presentación de contenidos de las asignaturas a los alumnos. Desde el punto de vista de las técnicas de data mining podemos encontrar en numerosas bibliografías clasificaciones como la que se muestra a continuación:

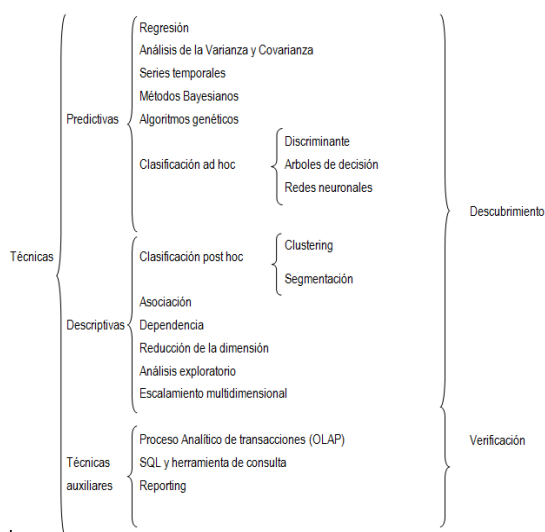


Figura Nª1: Clasificación de técnicas de datamining.  
Fuente: Minería de datos conceptos, técnicas y sistemas  
Pag. 9.

Si bien todas ellas pueden Utilizarse en el análisis de los datos en cuestión, algunas de ellas ya fueron probadas con resultados satisfactorios por investigadores que trabajan en el área de data minig educativo [5].

El procesamiento de los datos educativos en los que se aplica las técnicas de data mining proporcionan resultados sobre el descubrimiento de patrones de comportamiento de los alumnos frente a los contenidos virtuales, clasificación de tipos de alumnos de acuerdo al rendimiento, cumplimiento de las actividades propuestas por el docente, presentación de trabajos y comportamiento en grupo. Con respecto a los contenidos permite el análisis de acuerdo a la utilización que hacen los alumnos de los mismos, evaluando la estructura, la secuenciación, la profundidad de los concepto y el método más adecuado para la evaluación de la adquisición de las competencias que requiere cada asignatura.

Las técnicas más utilizadas por los investigadores en esta área para el procesamiento de datos educativos son: El análisis de secuencias, el descubrimiento de patrones de comportamiento y reglas de clasificación y agrupamiento. Si bien estas técnicas son independientes para mejorar los resultados obtenidos en muchos casos son aplicadas en forma combinada.

Las técnicas de clasificación y agrupamiento o Clustering [9] permiten la creación de grupos o conjunto de datos, sobre todo cuando el universo de análisis es demasiado grande. Para el caso de los datos que se obtienen con la metodología de blended learning permite comprenderlos más fácilmente al agrupar el comportamiento de los alumnos frente a los contenidos, la permanencia en determinadas páginas y la secuencia en las que se maneja el contenido de las mismas. Por otra parte es utilizado en combinación con otras técnicas de predicción sobre todo en los niveles universitarios [2].

Las técnicas de asociación [7] o descubrimiento de patrones [3], son aplicadas en diversos conjuntos de datos, para el caso de

los datos en estudio se utilizan para descubrir la asociación entre las páginas visitadas por los alumnos para hacer determinar que estructura sería más adecuada para facilitar la comprensión de los contenidos. Además para establecer el grado de asimilación de los alumnos en cada tema cuando se los evalúa virtualmente.

Por lo general se desarrolla un modelo para representar el comportamiento de los alumnos en las actividades virtuales, y se propone un modelo de navegación basado en la secuencia de consulta y en las asociaciones que realizan con sus comportamientos los alumnos.

Al igual que las otras técnicas, se puede encontrar trabajos en donde se combina con otras como los métodos heurísticos.

La técnica de análisis de secuencia analiza y encuentra [8] secuencias en los datos procesados. Esta técnica es aplicada en diferentes campos, en el caso de los datos generados con la metodología de blended learning analizan repeticiones de caminos utilizados por los alumnos para llegar a los resultados que presentan. Por lo general estos caminos pueden tomarse en forma individual o promediarlos para un estudio de comportamiento en grupo. En algunos casos puede ser conveniente la representación en forma de árbol, que muestre más claramente las relaciones entre cada nodo que representa las actividades. Esto permite al docente plasmar gráficamente las estructuras que subyacen en el comportamiento de los alumnos y ajustar el diseño de los materiales para facilitar el aprendizaje.

### Líneas de investigación y desarrollo

El trabajo se enmarca en la investigación de aplicación de técnicas y algoritmos de minería de datos en bases de datos educativas. Con el principal objetivo de cuantificar la efectividad del proceso de enseñanza y aprendizaje, organizar adecuadamente el contenido de la asignatura, aplicar el agrupamiento de alumnos de acuerdo a los perfiles encontrados, entre otros.

El presente trabajo está íntimamente relacionado con trabajos desarrollados en el área y próximos a ser presentados en encuentros de investigadores en el tema.

A partir de estos trabajos se analiza la posibilidad de desarrollar tesis de posgrado.

### Resultados y Objetivos

Los resultados esperados para este trabajo son los de estudiar las diferentes técnicas de data mining para seleccionar las más adecuadas para analizar el gran volumen de información generada con la utilización de la metodología de blended learning. El volumen de datos que se genera con esta metodología es demasiado grande para ser analizado por medio de otras técnicas y mucho menos manualmente.

Tiene como objetivos principales:

- Comprender el funcionamiento de las diferentes técnicas de data mining.
- Analizar los resultados que proporcionan cada técnica.
- Valorar el conocimiento extraído con la aplicación de cada técnica.
- Buscar un método de selección de variables que sean más representativas de las actividades de los alumnos.
- Identificar las variables que aportan más conocimiento al análisis de datos para ser extraídas de la información que proporciona el alumno.
- Utilizar herramientas de datamining disponibles en el mercado con el objetivo de seleccionar la de más fácil aplicación.
- Diseñar líneas de investigación que relacione el área de datamining con los datos generados con el blended learning.

## Formación de Recursos Humanos

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 D. desarrolla actualmente la tesis “Estudio y selección de técnicas y herramientas de data mining para ser aplicados a bases de datos utilizados en el blended learning” que se realiza con la dirección del Dr. Cheznívar Iván de la Universidad Nacional del Sur (U.N.S). 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 Facultad de Ciencias Exactas y Naturales (Fa.C.E.N.) en convenio con la Universidad Nacional del Sur (U.N.S). La Docente Investigadora Lic. Eugenia Cecilia Sosa Bruchmann desarrolla su tesis en la carrera Maestría en ingeniería de Software de la Universidad Nacional de San Luis.

Durante el año 2013 se realizó la presentación y defensa de la tesis de Maestría en Docencia Universitaria por el Mgter. Sosa Marcelo Omar D. realizada en la Facultad de Ciencias Agrarias de la Universidad Nacional de Catamarca (U.N.Ca.)

Además los docentes dirigen tesis de la carrera de Licenciatura en Tecnología Educativa de los alumnos Albornoz, Marcelo y Romero, Claudio.

Dentro de las actividades que desarrollan podemos destacar:

- La creación de un grupo de alumnos colaboradores, que desarrollan investigación con acompañamiento de docentes.
- La selección de temas posibles para tesis en el área relacionadas con la investigación.
- La producción científica para su presentación en congresos locales, nacionales e internacionales.

- Participación de los integrantes en cursos de actualización y posgrado en el área de estudio.
- La actualización y capacitación permanente de los investigadores en talleres o workshop relacionadas con el tema del trabajo.
- La participación de los investigadores como consultores en proyecto afines que se desarrollan en la Facultad de Ciencias exactas y Naturales en distintas áreas.
- La planificación de seminarios para docentes en temas relacionados con la investigación y resultados obtenidos en la investigación.

## REFERENCIAS

- 1- Horton, W. *Designing Web-Based Training*. John Wiley&Sons. 2000.
- 2- Luan, J. *Data Mining, Knowledge Management in Higher Education, Potential Applications*. Workshop Associate of Institutional Research International Conference. Toronto. 2002.
- 3- Agrawal, R.; Srikant, R. *Mining sequential patterns*. Int. Conf. Data Engineering, pp. 3-14. Taiwan. 1995.
- 4- Zaïane, O.Z. *Web Usage Mining for a Better Web-Based Learning Environment*. Conference on Advanced Technology for Education. pp 60-64. Alberta. 2001.
- 5- Zytkow J.; Klosgen W., *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press. 2001.
- 6- Minaei-Bidgoli, B.; Punch, W.F. *Predicting student performance: an application of data mining methods with the educational webbased system LON-CAPA*. IEEE Frontiers in Education. Pp 1-6. 2003.
- 7- Agrawal, R.; Imielinski, T.; Swami, A. *Mining association rules between sets of items in large databases*. ACM SIGMOD Conference on Management of Data. pp. 207-216. 1993.

- 8- Agrawal, R.; Srikant, R. *Mining sequential patterns*. *Int. Conf. Data Engineering*, pp. 3-14. Taiwan. 1995.
- 9- Arabie, P.; Hubert, J.; De Soete, G. *Clustering and Classification*. World Scientific Publishers. 1996.

Escuela Politécnica Superior.  
Universidad de Córdoba. España-  
2011.

### BIBLIOGRAFÍA CONSULTADA

- C. ROMERO & S. VENTURA. *Data Mining E-Learning*. Editorial WIT Press, 2006.
- C. ROMERO, S. VENTURA & E. GARCÍA. *Data Mining course management systems: Moodle case study and tutorial*. *Computers and Education*, 2007.
- SCHEUER, O., & MCLAREN, B. M. (2012). *Educational data mining*. *Encyclopedia of the sciences of learning*, Springer, New York, 1075-1079.
- TALAVERA, L., AND GAUDIOSO, E. 2004. *La minería de datos estudiantiles para caracterizar los grupos de comportamiento similares en espacios de colaboración*. Trabajo presentado en el Workshop de inteligencia artificial CSCL, 16<sup>th</sup> Conferencia Europea sobre Inteligencia Artificial, Valencia España.
- LAROSE, D. T. (2005). *Discovering Knowledge in Data, an introduction to Data Mining*. John Wiley & Sons. Estados Unidos.
- ESTADO ACTUAL DE LA APLICACIÓN DE LA MINERÍA DE DATOS A LOS SISTEMAS DE ENSEÑANZA BASADA EN WEB  
Cristóbal Romero Morales, Sebastián Ventura Soto, Cesar Hervás Martínez  
Departamento de Informática y Análisis Numérico.