

## INTEGRANDO TECNOLOGÍA Y EDUCACIÓN MEDIANTE MINERÍA DE DATOS, UNA APLICACIÓN PRÁCTICA

Mag. Graciela Beguerí, Mag. Alejandra Malberti, Mag. Raúl O. Klenzi  
Instituto de Informática – Departamento de Informática  
Facultad de Ciencias Exactas, Físicas y Naturales – Universidad  
Nacional de San Juan  
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", San Juan  
{grabada, amalberti, rauloscarklenzi}@gmail.com

### Resumen

Este artículo muestra como a través de proyectos de investigación se pueden incorporar, integrar y explotar simultáneamente distintas herramientas libres, provenientes de los aportes de las Tecnologías de la Información y la Comunicación –TIC’s-, en Educación en general y en particular en las carreras de grado del Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales – Universidad Nacional de San Juan. FCFN-UNSJ.

Uno de los trabajos insertos en el proyecto “Minería de Datos en Determinación de Patrones de Uso y Perfiles de Usuarios” busca investigar, implementar y evaluar herramientas de minería de datos y de estadística de código libre y abierto.

En este sentido se trabaja con los datos de una encuesta tomada a docentes-investigadores de la biblioteca universitaria “Dr. Emiliano Pedro Aparicio” perteneciente a la facultad citada anteriormente; pretendiendo encontrar atributos inductores sobre la satisfacción o percepción de estos usuarios con respecto al mencionado establecimiento.

**Palabras clave:** Estadística - Minería de Datos – RapidMiner - R

### Contexto

Este trabajo está inmerso en el proyecto “Minería de Datos en Determinación de Patrones de Uso y Perfiles de Usuarios”- código 21/E889-, que se desarrolla en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan. El contexto de aplicación se enmarca en las carreras del Departamento de Informática - FCFN-UNSJ. En las asignaturas de las carreras se trabaja con las herramientas RapidMiner, R y R-Commander, desarrolladas en código abierto.

El proyecto, se encuentra Aprobado y Subsidiado por CICITCA - Consejo de Investigación Científica, Técnica y de Creación Artística de la Secretaría de Ciencia y Técnica de la UNSJ.

### Introducción

El avance de la tecnología de Minería de Datos-MD ha permitido salvar algunos factores que desalentaban su aplicación, como ser: los productos comerciales eran muy costosos y en general los análisis estadísticos y de MD eran abordados mediante el uso inconexo de diferentes productos.

Hoy en día existe la posibilidad de contar con una enorme diversidad de software de código libre y abierto que caracteriza al trabajo en repositorios de datos y que puede ser utilizado tanto

académicamente como comercialmente, abriendo así oportunidades educativas y laborales [15] [16].

En MD se puede trabajar con un mayor volumen de datos, así como con datos mucho más complejos, diferencias fundamentales con la estadística.

Aunque los algoritmos con que se vale MD están basados o se nutren de modelos estadísticos, aún persiste una desvinculación académica y práctica entre ambos tópicos.

Diversos trabajos expresan que:

*La minería de datos (en inglés, data mining) se define como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de datos. En la actual sociedad de la información, donde día a día se multiplica la cantidad de datos almacenados casi de forma exponencial, la minería de datos es una herramienta fundamental para analizarlos y explotarlos de forma eficaz para los objetivos de cualquier organización. La minería de datos se define también como el análisis y descubrimiento de conocimiento a partir de datos.*

*La minería de datos hace uso de todas las técnicas que puedan aportar información útil, desde un sencillo análisis gráfico, pasando por métodos estadísticos más o menos complejos, complementados con métodos y algoritmos del campo de la inteligencia artificial y el aprendizaje automático que resuelven problemas típicos de agrupamiento automático, clasificación, predicción de valores, detección de patrones, asociación de atributos, etc. Es, por tanto, un campo multidisciplinar que cubre*

*numerosas áreas y se aborda desde múltiples puntos de vista, como la estadística, la informática (cálculo automático) o la ingeniería [4] [13].*

Numerosos estudios se han centrado en analizar la popularidad de los software en MD. Así se puede encontrar que, en una consulta respondida por 1103 votantes, se indica a RapidMiner como una de las herramientas de uso más frecuente en MD [2]. Ver figura 1.

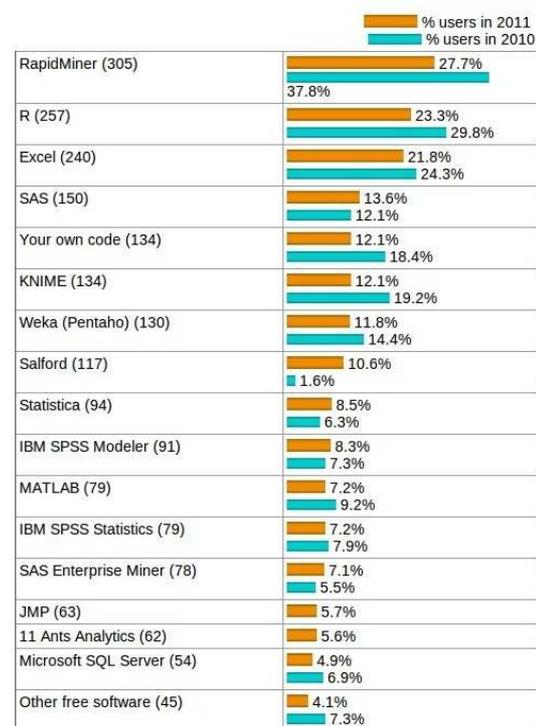


Figura 1. Algunas de las respuestas de los 1103 votantes a la pregunta: ¿Qué herramientas de minería/análisis de datos que utilizó en los últimos 12 meses para un proyecto real (no sólo para evaluación)?

También R es señalado en primer lugar, siendo R un programa básicamente estadístico [3]. Ver figura 2.

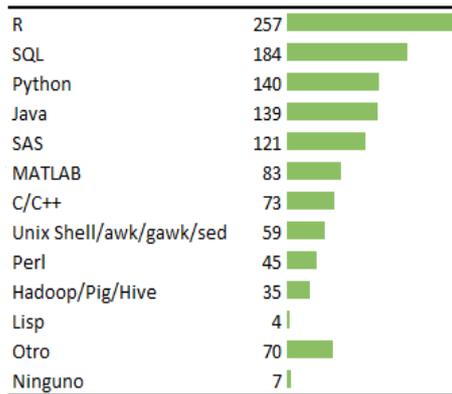


Figura 2. Respuesta de 570 personas consultadas sobre ¿Qué herramientas de minería/análisis de datos utilizó en los últimos 12 meses?

Dado que R también se puede tratar como una extensión de RapidMiner, lo adecuado es trabajarlos de manera integrada. Esto constituye un aporte tendiente a eliminar brechas que en diferentes oportunidades se producen entre áreas disciplinares y/ o asignaturas que hacen uso de TIC's en el proceso de enseñanza aprendizaje.

Asimismo, desde lo conceptual se puede observar que numerosa bibliografía de Minería de Datos contiene explícitamente capítulos de Estadística, presentándose la misma situación de forma inversa [1] [4] [5] [6] [7] [8] [9] [10] [13] [14].

En resumen, para lograr la interdisciplinariedad en el desarrollo curricular de las carreras universitarias es necesario abordar tanto la disyunción entre teoría y práctica de cada asignatura como con las restantes. Ello podría contribuir a superar la desvinculación entre la formación inicial y la práctica profesional y de esta manera construir puentes que las acerquen.

## Líneas de investigación y desarrollo

Las tareas de reconocimiento de perfiles de docentes-investigadores se realizarán,

a partir de la aplicación de herramientas estadísticas y de minería de datos, a los datos obtenidos en encuestas sobre satisfacción o percepción de estos usuarios con respecto a la biblioteca universitaria. Este proceso permitirá aplicar, evaluar, comparar y complementar la información extraída a través de la integración de las herramientas.

El Tipo de Diseño de la investigación propuesto es de Campo. Para ello, la población a estudiar corresponde a:

- Datos de docentes-investigadores de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan.
- Datos obtenidos a partir de instrumentos de recolección de información, desde un constructo validado.

## Resultados y Objetivos

Con esta propuesta se pretende:

- Aplicar técnicas de Minería de Datos y Estadística que permitan la reducción de atributos, sin pérdida de capacidad descriptiva, de los datos originales.
- Realizar un análisis comparativo de las técnicas consideradas.
- Correlacionar el conocimiento adquirido desde las encuestas a docentes con el ya obtenido con los alumnos y extraer conclusiones que favorezcan la toma de decisiones.
- Aplicar y evaluar las potencialidades de las herramientas integradas RapidMiner, R y R-Commander.
- Coadyuvar a lograr una vinculación real entre asignaturas y áreas disciplinares de la carrera.

## Formación de Recursos Humanos

En el proyecto de investigación “Minería de Datos en Determinación de Patrones de Uso y Perfiles de Usuarios” participan docentes del Departamento de Informática, los que a su vez son responsables de las asignaturas Algoritmos y Estructuras de Datos, Probabilidad y Estadística, Sistemas de Datos, Inteligencia Artificial y Teoría de la Información.

Hace tiempo que las tareas proyectadas inciden en la vinculación entre las asignaturas nombradas y beneficia de manera directa a las carreras del Departamento de Informática.

Desde años anteriores se vienen generando Trabajos Finales de Grado para las carreras Licenciatura en Sistemas de Información y Licenciatura en Ciencias de la Computación así como Tesis de Posgrado.

Durante el ciclo lectivo 2011, en el marco del proyecto citado, se iniciaron tres tesis de posgrado y dos trabajos finales de grado, habiéndose concluido uno de estos y encontrándose el otro en instancia de evaluación final.

## Referencias

- [1] Berry, Michael J.A.; Linoff, Gordon S. *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management - Second Edition*- Wiley Publishing. 2004
- [2] Figura 1. <http://www.kdnuggets.com/polls/2011/tools-analytics-data-mining.html>
- [3] Figura 2. <http://www.r-bloggers.com/kdnuggest-r-most-commonly-used-software-for-data-mining-analytics/>
- [4] Han, J; Kamber, M. *Data Mining: Concepts y Techniques. Second Editions*. Morgan Kaufmann Publisher. 2006.
- [5] Hand, David; Mannila, Heikki; Smyth, Padhraic *Principles of Data Mining*. The MIT Press. 2001
- [6] Hastie ,T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer-Verlag, 2001.
- [7] Hernández Orallo J., Ramirez Quintana, J, Ferri Ramirez, C. *Introducción a la Minería de Datos*. Pearson-Prentice Hall. 2008
- [8] Larose, D. *Data Mining. Methods and Models*. Department of Mathematical Sciences Central Connecticut State University Wiley. A John Wiley & Sons, Inc Publication. 2006.
- [9] Larose, D. *Discovering Knowledge In Data - An Introduction to Data Mining*. John Wiley & Sons, Inc., Publication. 2005
- [10] Myatt, G.J. *Making Sense of Data. A Practical Guide to Exploratory Data Analysis and Data Mining*. Wiley-Interscience a John Wiley & Sons, Inc., Publication. 2007
- [11] R <http://www.r-project.org/>
- [12] Rapid-I. <http://rapid-i.com/api/rapidminer-5.1/com/rapidminer/tools>. 2011.
- [13] Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin *Introduction to Data Mining* Addison-Wesley. 2006
- [14] Unwin, A.; Theus, M.; Hofmann, H. *Graphics of Large Datasets: Visualizing a Million (Statistics and Computing)*. Springer, 2006.
- [15] Web 1- <http://www-users.cs.umn.edu/~kumar/dmbook/resources.htm>
- [16] Web 2 - <http://www.junauza.com/2010/11/free-data-mining-software.html>