

# Aplicación de la Técnica Regresión Logística de la Minería de Datos en el proceso de Descubrimiento de Conocimiento (KDD) en Bases de Datos Operativas o Transaccionales

Juan Camilo Giraldo Mejía<sup>1</sup>, Fabio Alberto Vargas Agudelo<sup>1</sup>  
Tecnológico de Antioquia, TdeA. Institución Universitaria  
jgiraldo1@tdea.edu.co, fvargas@tdea.edu.co

## Resumen

El artículo presenta la caracterización de la técnica Regresión Logística de la Minería de Datos (Data Mining). Igualmente, se muestra la funcionalidad y aplicación de la técnica para apoyar al proceso de obtención de conocimiento (Knowledge Discovery in Databases o KDD), a encontrar información interesante a partir de Datos "ocultos". La funcionalidad de la técnica se ejemplifica con los resultados obtenidos en un trabajo de investigación que se realizó buscando encontrar el nivel de innovación y desarrollo tecnológico en algunas empresas de Colombia.

La finalidad es mostrar el proceso de obtención de conocimiento de un sistema de bases de datos Transaccionales u operativos para empresas de bienes y servicios. En él, se desarrollan los antecedentes conceptuales e investigativos y la caracterización de los conceptos fundamentales relacionados con el proceso de descubrir conocimiento, Minería de Datos y la Técnica de Regresión Logística.

## Palabras clave:

SGBD, KDD, Minería de Datos.

## Abstract

The article presents the characterization of the logistic regression technique of data mining (Data Mining). Also shows the functionality and application of technology to support the process of obtaining knowledge (Knowledge Discovery in Databases or KDD), to find interesting information from data "hidden". The functionality of the technique is exemplified by the results of a research which was carried out seeking to find the level of innovation and technological development in some companies in Colombia.

The aim is to show the process of obtaining knowledge of a database system Transactional or operational goods and services companies. In the develop conceptual and research background and characterization of the fundamental concepts related to the process of discovering knowledge, data mining and logistic regression.

## Keywords:

DBMS, KDD, Data Mining

## Introducción

En los últimos años, ha existido un gran crecimiento en nuestras capacidades de generar y coleccionar datos debido básicamente, al gran poder de procesamiento de las máquinas como a su bajo coste 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 información. El descubrimiento de esta información oculta es posible gracias a la Minería de Datos (DataMining), que 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, pero es el descubrimiento del conocimiento (KDD) el que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados (Vallejos, 2006).

“La posesión real de toda la información obtenida depende de nuestra capacidad para hacer ciertas operaciones con la información textual, por ejemplo: buscar información interesante” Guzmán (2005).

Posteriormente, con el apoyo de la tecnología de información, justificada en las bases de datos relacionales, los ordenadores personales y la computación gráfica, el acceso a la información y su representación por parte de los usuarios finales, comenzaron a ser cada día más frecuentes. Todo esto enmarcado en el concepto de Descubrimiento u obtención de Conocimiento (KDD).

Existen diversos métodos durante el proceso de KDD para el análisis de la información. Cada una de ellos ofrece herramientas conceptuales para extraer cierta inteligencia o conocimiento de los datos acumulados en la empresa. Estas tecnologías son las técnicas de Minería de Datos. La Minería de Datos utiliza algoritmos, como Regresión Logística, para explorar los repositorios de datos y extraer conocimiento a partir de los datos.

“La técnica de la regresión logística se originó en la década de los 60 con el trabajo de Cornfield, Gordon y Smith. En 1967 Walter y Duncan la utilizan ya en la forma que la conocemos actualmente, o sea para estimar la probabilidad de ocurrencia de un proceso en función de ciertas variables. Su uso se incrementa desde principios de los 80 como consecuencia de los adelantos ocurridos en el campo de la computación.” Domínguez (2001).

La regresión logística es un procedimiento cuantitativo de gran utilidad para problemas donde la variable dependiente toma valores en un conjunto finito.

El objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico). (Moliner, 2001)

## Conclusiones

La regresión logística es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación de diferente contexto, de ahí su amplia utilización.

El momento en que la mayoría de los profesionales tenga acceso a herramientas de Minería de Datos para obtener conocimiento de las Bases de Datos transaccionales u operativas trae consigo un mejoramiento en el tipo de organizaciones que se encuentran compitiendo en el mercado debido a que se elevan los niveles de calidad desde las mejores tomas de decisiones por parte de sus analistas.

## Referencias

- [1] [Molinero, 2001] Molinero, L. (2001). "La Regresión Logística", Asociación de la sociedad Española de Hipertension, España.
- [2] [Molinero, 2006] Vallejos, S. (2006). "Minería de Datos", Facultad de Ciencias Exactas, Naturales, y Agrimensura, Universidad Nacional del Nordeste.
- [3] [Domínguez, 2001] Domínguez, E. (2001). "Regresión logística, un ejemplo de su uso en endocrinología", Instituto Nacional de Endocrinología, Revista Cubana Endocrinol, Cuba.
- [4] [Valcárcel, 2004] Valcárcel, V. (2004). "Data Mining y el descubrimiento de conocimiento". Revista de la Facultad de Ingeniería Industrial. Vol. (7) 2: pp. 83-86 (2004) UNMSM ISSN: 1560-9146. Diciembre de 2004.
- [5] [Lezcano, 2002] lezcano, R. (2002). "Minería de Datos". Trabajo de Investigación. Universidad Nacional del Nordeste. 2002.
- [6] [Guzmán, 2005] Guzmán, G. (2005). "Búsqueda de colocaciones en la Web para sinónimos de Wordnet". Universidad de Guanajuato. 2005.
- [7] [Mures, 2005] Mures, J. (2005). "Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de Resultados", Revista pecvnia, Universidad de León.
- [8] [Puga, 2011] Puga, J. "Eventos por Variable en Regresión Logística y Redes Bayesianas para Predecir Actitudes Emprendedoras", Revista Electrónica de Metodología Aplicada, Universidad de Almería.
- [9] [Aguayo, 2007] Aguayo, M. (2007). "Como hacer una Regresión Logística con SPSS paso a paso", Servicio de Medicina Interna, Hospital Universitario Virgen Macarena, Sevilla España.