

**Anteproyecto del Trabajo de Fin de Carrera**

**Minería de datos aplicada al análisis de  
sentimiento en opiniones en internet**

**Escuela Universitaria de Informática**

**Universidad Politécnica de Madrid**

**Madrid, Abril de 2011**

Autor: Javier Pelado García

Tutor: Ángel Arroyo Castillo

Departamento: Sistemas Inteligentes Aplicados

# 1. INTRODUCCIÓN

Antes de la web social, preguntábamos a nuestros amigos cuando necesitábamos una recomendación en cualquier ámbito: películas, libros, servicios, hoteles... Hoy en día, Internet hace posible encontrar opiniones en casi cualquier ámbito.

Desde el punto de vista de una compañía de productos y servicios, el análisis de las opiniones sobre la misma y su oferta que circulan por Internet se hace esencial hoy en día. Podríamos decir que hay una relación directa entre la inversión en la optimización del 'clima social' respecto a la marca y los rendimientos comerciales de dichos productos/servicios.

El análisis del sentimiento en opiniones, se realiza aplicando técnicas de minería de datos sobre los textos que las expresan.

La **minería de datos (DM, Data Mining)** consiste en la extracción no trivial de información que reside de manera implícita en los datos. Dicha información era previamente desconocida y podrá resultar útil para algún proceso. En otras palabras, la minería de datos **prepara, sondea y explora** los datos para sacar la información oculta en ellos.

Bajo el nombre de minería de datos se engloba todo un conjunto de técnicas encaminadas a la extracción de conocimiento procesable, implícito en las bases de datos. Está fuertemente ligado con la supervisión de procesos industriales ya que resulta muy útil para aprovechar los datos almacenados en las bases de datos.

Las bases de la minería de datos se encuentran en la **Inteligencia Artificial** y en el análisis estadístico. Mediante los modelos extraídos utilizando técnicas de minería de datos se aborda la solución a problemas de predicción, clasificación y segmentación.

La minería de datos utiliza muy diversas técnicas para analizar los datos disponibles y poder clasificarlos. Las técnicas o algoritmos más utilizados para dicho cometido, pertenecientes al campo de la Inteligencia Artificial, son los algoritmos **ID3** y **C4.5**, que son *Árboles de decisión*.

## 2. OBJETIVOS

Crear una herramienta web que facilite la extracción de opiniones de las distintas redes sociales más influyentes del momento, para su posterior polarización o clasificación.

La herramienta deberá ser capaz de ofrecer una interfaz amigable para el entrenamiento del sistema experto en clasificación de opiniones y posterior testeo de dicho sistema con opiniones extraídas de internet mostrando los resultados de una forma visual más cómoda de interpretar que las tablas de datos.

Como objetivo de estudio, observando los resultados ofrecidos por el sistema, podremos demostrar el peso de las opiniones de los internautas en las redes sociales y su influencia en la popularidad e integración en el mercado, o por el contrario en el fracaso, de los nuevos productos o servicios.

## 3. ELECCIÓN DEL ENTORNO DE TRABAJO

Al tratarse de un portal web de entrenamiento, serán necesarios dos entornos distintos que trabajen conjuntamente:

- Un entorno web que contenga el portal que hará de interfaz de dicho entrenador. Éste estará formado por un servidor web **apache** con intérprete de **php** y **javascript** y un gestor de bases de datos **mysql**.
- El otro entorno será el cerebro del sistema, dado que se encargará de construir los clasificadores que buscarán el sentimiento en los textos u opiniones. Para el desarrollo de este módulo, usaré un entorno **java**, dado que usaré los algoritmos que contiene la librería **WEKA** para los entrenamientos y está escrita en este lenguaje.

Para comunicar los dos módulos usaré sockets.

## 4. BIBLIOGRAFÍA

- “WEKA”.

Machine Learning Group at University of Waikato.

<http://www.cs.waikato.ac.nz/ml/weka/>

- “Introducción a la Minería de Datos”.

José Hernández Orallo, M. José Ramírez Quintana, Cèsar Ferri Ramírez.

<http://users.dsic.upv.es/~flip/LibroMD/>

- “Minería de opiniones: Realidades y ficciones”.

Oscar Marín Miró.

<http://www.paradigmatecnologico.com/mineria-de-opiniones-realidades-y-ficciones-i/>

- “Minería de Datos”.

Wikipedia.

[http://es.wikipedia.org/wiki/Miner%C3%ADa\\_de\\_datos](http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos)