

UNIVERSIDAD NACIONAL DE LA MATANZA INTELIGENCIA DE NEGOCIOS

Trabajo Práctico de Explotación de Información

El siguiente caso de Inteligencia de Negocios se enfoca en la resolución de un problema de Explotación de Información y Minería de Datos.

Descripción del Problema

La multinacional eBay Inc. (propietaria de la conocida web eBay.com para subasta de productos a través de internet) desea conocer mejor el comportamiento de las transacciones producidas en su web para, a la vista de los resultados, diseñar nuevos servicios que mejoren la experiencia de los usuarios vendedores para así incrementar las ventas y, por tanto, mejorar los ingresos de la compañía.

Así, se propone aplicar inteligencia de negocios de cara a extraer conocimiento útil para la toma de decisiones a partir de algunos datos disponibles. Concretamente, el objetivo es construir un modelo que clasifique las subastas entre competitivas y no competitivas. Una subasta se considera competitiva si recibe al menos dos ofertas sobre el objeto subastado.

Los datos disponibles corresponden a 1972 subastas a través de eBay.com realizadas durante dos meses. Se incluyen variables que describen el objeto (categoría), el vendedor (mediante su valoración o rating) y los términos de la subasta fijados por el vendedor (duración de la subasta, precio de inicio, moneda y día de la semana en el que finaliza la subasta). Además, se dispone del precio al cual se cerró la subasta. Estos datos están contenidos dentro del archivo *subastas.xls*.

Se pretende identificar los criterios para que una subasta sea o no competitiva así como comprender qué relaciones provocan dicho factor de cara a diseñar estrategias de negocio para que los clientes vendedores aumenten la tasa de subastas competitivas en sus productos.

En base a la descripción dada y aplicando las fases y tareas de la metodología CRISP-DM se pide:

- Realizar un análisis previo para entender cómo son los datos del archivo subastas.xls a fin de conocer sus características y la necesidad o no de realizar un Preprocesamiento previo de los mismos para su uso.
- Determinar el o los procesos de explotación de información y algoritmos de minería de datos a utilizar durante la fase de modelado.
- Utilizando la herramienta Tanagra, desarrollar el/los modelo/s necesario/s con los procesos de explotación identificados. En esta tarea, deberá estudiar el comportamiento de los distintos algoritmos de clasificación (vistos en la materia) mediante el diseño experimental apropiado y el análisis comparado de resultados. Además, deberá extraer conclusiones a partir del conocimiento aprendido mediante estos algoritmos para comprender las relaciones entre las variables (también llamadas características o predictores) que favorecen las subastas competitivas.
- Aplicar las métricas de evaluación de modelos (vistas en la materia) para seleccionar los resultados más adecuados.
- Escribir un informe de los resultados obtenidos y el conocimiento descubierto lo suficientemente profundo para que resulte de utilidad a la compañía.

Material de Referencia

- Britos, P., García-Martínez, R. (2009). *Propuesta de Procesos de Explotación de Información*. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- Metodología CRISP-DM
- Apuntes de clase

Fecha de Entrega: 4 de Diciembre de 2017