

Propuesta de Proyecto de Electiva III

Mariana Paucar Castillo

Minería de datos con Weka para la predicción del precio de automóviles de segunda mano

Carrera de Ingeniería en Sistemas Computacionales
Universidad de Guayaquil

1 Definición del problema de investigación

Profundizar en un concepto muy interesante relacionado con la obtención y extracción de información relevantes que podemos encontrar en una colección de datos. Estamos hablando de la Minería de Datos. Bajo este nombre se agrupan todas aquellas técnicas que nos ayudan a extraer conocimientos e información relevantes que se encuentran implícitos en las bases de datos.

2 Datos

Primero nos centraremos en el estudio de la minería de datos, analizaremos y marcaremos en detalle los conceptos más importantes de esta metodología de tratamiento de datos. Hablaremos de sus orígenes, de su presente y del futuro que tiene esta “tecnología” a largo plazo. La segunda parte del proyecto consistirá en crear un caso de estudio real, del cual podamos extraer datos, analizarlos y aplicar diferentes métodos y técnicas de minería de datos.

3 Metodología

Procederemos ahora a relatar el proceso de tratamiento de los datos obtenidos y exportados al fichero csv mediante Weka. Estos datos, habrá que procesarlos detenidamente para que los resultados que obtengamos sean lo más precisos posibles, es decir, con la menor media de error que seamos capaces de conseguir. Para ello, tendremos que eliminar atributos que no nos proporcionen información relevante para predecir el precio del vehículo. Registros con campos vacíos, que sólo introducen e inducen a errores, valores mal escritos y no asociables, etc.

4 Evaluación de la metodología

Pasaremos a analizar y exponer los resultados de tres algoritmos de los múltiples algoritmos implementados en Weka. La selección de estos métodos ha estado basada en los resultados que deseamos obtener, Weka permite aplicar unos métodos u otros en concordancia con el propósito del estudio. En este caso,

el propósito era el de predecir el precio de un vehículo, por lo que, bajo la recomendación del director de este proyecto y varias pruebas con otros algoritmos, expondré los resultados obtenidos con estos tres en concreto únicamente.

5 Resultados esperados

La minería de datos es una herramienta con un potencial increíble y aplicable en un sinfín de proyectos, circunstancias y finalidades. A través de este proyecto, hemos podido introducir un poco en este mundo de la obtención de información relevante de masas de datos compactas. Hemos introducido el concepto de minería de datos, así como analizado varios de sus numerosos tipos y familias de métodos de cálculo y obtención de resultados. Evidentemente, se puede profundizar muchísimo más en la minería de datos. No obstante, hemos podido aprender a utilizar una herramienta muy potente y con licencia OpenGPL para la minería de datos llamada Weka, hemos analizado sus características y hemos surcado por encima de sus numerosas posibilidades y funcionalidades.