

Factores ambientales en la determinación del sueldo

Práctica Final de la Asignatura de
Tratamiento Inteligente de la Información y Aplicaciones
Convocatoria de Febrero, 2009/2010

Juan A. Botía Blaya
juanbot@um.es

1 Introducción

En esta práctica, vamos a comprobar si es posible determinar el sueldo de una persona a partir de determinadas variables que definen su situación. Para ello, vamos a utilizar el conjunto `adult`¹ en el que se adjuntan descripciones sobre adultos, en forma de ejemplares. De entre los atributos que podemos encontrar en este conjunto, tenemos los siguientes:

- Edad
- Tipo de trabajo
- Ocupación
- Sexo
- Horas a la semana
- País de origen

En la práctica se trata de responder a la pregunta ¿en qué medida es determinante nuestro contexto para determinar nuestra nómina? ¿es algún factor más relevante que el resto? ¿qué conjunto mínimo de variables podría servir para determinar nuestra *solvenia*? ¿En qué medida podría aprovechar alguna entidad información fácil de obtener para generar otra información no suministrada directamente por nosotros?

Se pide responder a estas preguntas, realizando un análisis exploratorio de los datos. Esto nos permitirá tener una vista clara y detallada del conjunto de datos y de cada uno de los atributos relevantes que ahí aparecen. Tras eso, deberíamos ser capaces de realizar minería de datos (i.e. la tarea es clara, clasificación, aunque no deberíamos restringir nuestra actividad si es que de otra forma se genera conocimiento nuevo sobre este dominio).

Se valorará en la práctica

¹<http://archive.ics.uci.edu/ml/machine-learning-databases/adult/>

- la creatividad a la hora de realizar el análisis exploratorio de datos,
- la potencia de la técnica elegida, siempre que se haya razonado correctamente por qué se ha escogido cada técnica usada,
- la claridad y validez de las conclusiones generadas en el trabajo.

La entrega de dicha práctica deberá incluir:

- Una memoria en la que se expliquen las respuestas a las preguntas realizadas y cómo se ha llegado hasta ellas mediante el análisis exploratorio de datos y el data mining.
- La presentación a realizar por el alumno en la defensa del trabajo.

Los trabajos se han de entregar al profesor el 4 de febrero, hasta las 12:00 AM, mediante el correo electrónico a juanbot@um.es. El trabajo no debe considerarse entregado hasta no recibir por parte del profesor un mensaje de respuesta al envío. Si dicho mensaje no se recibe, deberá repetirse dicho envío. Se ruega usar el prefijo [TIIA0910] en el subject. La defensa de trabajos tendrá lugar el 11 de febrero, (día del examen) y se publicará con antelación el correspondiente horario de defensa.