

Introduccion a Data Mining – 2020

Caso IDM-I

Considere el siguiente caso:

Una cadena de hoteles que gestiona numerosos establecimientos en varios países registra información acerca de sus huéspedes en los diferentes hoteles. La gerencia desea implementar un **proyecto de Data Mining** a efectos de extraer el máximo provecho de esos datos. Más precisamente, la cadena desea conocer mejor a sus clientes de manera de poder informarlos sobre eventos especiales, promociones especiales, etc., durante su estancia como así también después de esta.

A la llegada de un huésped, se registran sus datos demográficos y se le pide completar un cuestionario cuando se le hace entrega de una credencial que le permite ingresar a distintos lugares recreativos tales como piscinas, bares, etc. Esos lugares son gratuitos, pero vía la credencial es posible registrar cuando un huésped hace uso de alguno de esos servicios. Además, la credencial sirve como tarjeta de crédito para pagar ciertas bebidas como así también para pagar productos en las pequeñas tiendas que la cadena posee. Todas esas transacciones son registradas en una base de datos.

Por medio de su canal de TV privado, el hotel informa a sus huéspedes sobre los próximos eventos, actividades y promociones especiales, etc. Sin embargo, puesto que los huéspedes pasan la mayor parte de su tiempo afuera y no mirando TV, el hotel necesita un sistema para enviar anuncios altamente personalizados de manera de garantizar que sólo se envíen anuncios en los que el huésped esté posiblemente interesado. Además, el sistema de data mining también tiene que dar una sugerencia razonable sobre qué anuncios enviar a un huésped particular ya durante los primeros días de su estadía, cuando el sistema tiene pocas posibilidades de aprender los hábitos de ese huésped particular. Después de la estadía, se le solicita al huésped que complete un formulario de evaluación. Este formulario contiene una lista de preguntas en las que se debe consignar un puntaje de 1 a 5. Para cada respuesta se pueden explicar los motivos o agregar comentarios adicionales breves.

Durante el invierno, la cadena envía publicaciones de una selección de sus hoteles a antiguos huéspedes, para obtener nuevas reservas en uno de sus hoteles. La selección para este mailing personalizado tiene que ser hecha de modo tal que sólo las publicaciones de los hoteles más interesantes para un huésped particular son enviadas a ese huésped. Por lo expresado, es importante considerar que en la mayoría de los casos un huésped de hotel no elige muchas veces exactamente el mismo hotel.

En conclusión, el sistema de data mining debe proporcionar información para:

- 1. Decidir cuáles anuncios de TV van a ser enviados a los huéspedes durante su estadía; un subproblema de este primer problema es decidir cuales anuncios van a ser enviados durante los primeros días de la estadía de un huésped.**
- 2. Decidir sobre la selección de hoteles para el mailing privado durante el verano.**

Recorrer los diferentes pasos del proceso KDD, explicando cómo se aplican en este caso. Indicar para cada paso cuáles técnicas usaría y justificar su elección.