

# **Maestría en Explotación de Datos y Gestión del Conocimiento**

## **Introducción a Data Mining**

### **Estudio de Caso 1**



:

- ***Díaz, Darío Ezequiel***
- ***Figuerola, Martín***
- ***Mazzola, Gabriel Agustín***
- ***Oña, Ana***

# Caso 1 - Intro Data Mining

**La cadena desea conocer mejor a sus clientes de manera de poder informarles sobre eventos especiales, promociones especiales, etc., durante su estancia como así también después de esta.**

**Indicar para cada paso cuáles técnicas usaría y justificar su elección.**

**Consignas:**

**1. Decidir cuáles anuncios de TV van a ser enviados a los huéspedes durante su estadía; un subproblema de este primer problema es decidir cuáles anuncios van a ser enviados durante los primeros días de la estancia de un huésped.**

Los datos con los que contaremos para el proceso de data mining en nuestro ejemplo vienen de diferentes fuentes:

- Reserva online : datos sobre fechas de estadía, cantidad de pasajeros y motivo del viaje (booking, tripadvisor, despegar y página de reservas propia de la cadena )
- Check-in : datos a rellenar por el recepcionista en el momento del check in como Nombre, nro. DNI/ Pasaporte, dirección de email, dirección de los pasajeros y datos del vehículo si corresponde.
- Encuesta: se le ofrece un formulario de encuesta voluntario, a cambio de un voucher por una consumición en el bar, con el objetivo de motivar al pasajero a llenar dicho formulario, donde se pregunta sobre el motivo del viaje, si ha estado o no antes en algún hotel de la cadena, preferencias sobre lugares de vacaciones, deportes que practica etc..
- Cuestionario para la credencial
- Encuesta de satisfacción (hotel): al retirarse se le invita a llenar una encuesta de satisfacción sobre las instalaciones y el servicio.
- Review/reseña de páginas de reserva: se recolectarán datos de las reseñas de los clientes en las diferentes plataformas de reserva (booking, tripadvisor, etc). donde se puntúa la satisfacción del cliente, la calidad de los servicios y se deja un comentario acerca de su experiencia durante la estadía.
- Encuesta de ocupación hotelera: los establecimientos hoteleros y parahoteleros deben responder una encuesta elaborada y aplicada por el INDEC y las direcciones estadísticas provinciales. Las variables son: categoría del establecimiento, disponibilidad de habitaciones, unidades y plazas, ocupación mensual, unidades y plazas, entrada de viajeros y pernoctaciones según lugar de residencia habitual,

tarifa promedio, participación porcentual del ingreso en los diferentes servicios ofrecidos, consumo y facturación de los servicios de gas, electricidad y agua.

A continuación describimos las etapas del proceso KKD que aplicarían en este caso:

### **1a) Etapa de selección de datos**

Datos solicitados a los pasajeros:

- Nombre y apellido
- DNI/Pasaporte
- Datos del acompañante
- Modelo de auto, patente
- Dirección donde vive y localidad
- Teléfono de contacto
- Mail
- Grupo familiar
- Motivo del viaje (Turismo; Negocios – Industria)
- Número de socio/pasajero frecuente

Además de los datos mencionados anteriormente, se tendrán en cuenta otros datos provenientes del uso y consumo de la credencial:

- Asistencia a spa/gym/bar/reservas de canchas/uso de amenities
- Tiempo de permanencia en la habitación
- Uso de la cochera
- Nombre del producto
- Nombre del servicio
- Categoría del producto (alimentos, bebidas, ropa, etc.)
- Categoría del servicio (lavandería, transporte, excursiones)
- Actualidad (Recency) tiempo entre compras/consumos
- Frecuencia: el número de veces que un cliente compra
- Valor monetario de la compra

## **1b) Etapa de pre-procesamiento/limpieza**

La primera etapa consiste en recolectar los datos obtenidos a partir de distintas fuentes (check in manual, encuestas en papel, formularios web, webAPIs de sitios de reserva, etc). Luego se ejecutaría el proceso de limpieza de los datos que consiste en:

- Identificar valores atípicos o outliers
- Identificar columnas que tienen el mismo valor o ninguna variación y eliminarlas
- Eliminar registros duplicados
- Reconocer valores vacíos como faltantes
- Imputación de valores faltantes

## **1c) Etapa de transformación/reducción**

En la etapa de transformación de datos se utilizan técnicas para cambiar el tipo o la distribución de las variables de datos. Las transformaciones van a depender del tipo de dato según sea categórico o numérico.

Cuando la característica categórica es ordinal, podemos usar la codificación ordinal (por ejemplo, la calificación del servicio). Cuando las características categóricas no tienen ningún orden (nominal), usamos One Hot Encoding (por ejemplo género) para codificar una variable categórica en variables binarias.

Con respecto a las variables numéricas, se le puede aplicar transformaciones para escalar los datos mediante normalización (escala una variable al rango 0 y 1) o estandarización (escala una variable a una Gaussiana estándar).

Luego los datos se pasan de un formato relacional de la base de producción a un formato de datawarehouse que permite trabajar con los datos de una manera más eficiente. En el caso del Hotel este proceso es automático y ocurre durante la madrugada cuando la base de producción está menos requerida. Esto involucra los datos de todos los hoteles de la cadena.

## **1d) Etapa de minería de datos**

Una vez concluida la limpieza de los datos; el preprocesamiento y el armado de la Data Warehouse, comenzamos con la etapa de la Minería de Datos. Las técnicas que podemos aplicar son:

### **Clasificación**

Dividimos los clientes en segmentos pre-definidos; como por ejemplo:

- geográfico
  - país
  - provincia
  - localidad
- demográfico
  - edad
  - género
  - ocupación (Industria)
- perfil del pasajero
  - tipo de pasajero: potencial, primera vez, pasajero frecuente, corporativos, grupos
  - tipo de viaje: mochileros, jubilados, familia, viajeros solos, parejas, vacacionistas, aventureros, turistas médicos, viajeros de negocios, etc.
  - ocasión de uso: turismo, trabajo

### **Clustering**

La agrupación en clústers agrupa a los clientes en función del conocimiento del dominio y la base de datos, pero no se basa en definiciones de grupo predeterminadas. Esta función es beneficiosa porque ayuda a los hoteleros a comprender quiénes son los clientes. Por ejemplo, se podría construir clústers utilizando como información el comportamiento de compra de los huéspedes, los datos recabados a través de las encuestas de satisfacción, el consentimiento para recibir promociones o descuentos y los reviews que los usuarios dejan en las plataformas externas al hotel (booking, tripadvisor) una vez finalizada su estadía.

### **Detección de desviaciones o anomalías de datos**

La detección de desviaciones descubre anomalías en los datos, como un aumento repentino de las compras por parte de un cliente. La información de este tipo puede resultar útil si una cadena hotelera quiere agradecer a un huésped por su reciente aumento en el gasto u ofrecer una promoción.

Los gerentes de marketing también pueden intentar establecer correlaciones entre los aumentos inesperados en las desviaciones con variables incontrolables del entorno empresarial que no están representados en la base de datos (por ejemplo, un aumento de los precios de los combustibles, una devaluación brusca del tipo de cambio, anuncios de congelamiento de precios, incrementos significativos de las tasas de interés).

### **Predicción**

Estimar el valor futuro de las variables continuas basándose en patrones y tendencias dentro de los datos. Por ejemplo, la función de pronóstico se puede utilizar para predecir la demanda en diferentes épocas del año y para eventos o fechas especiales en base a los datos recolectados. Con el pronóstico, también se pueden usar las tendencias de los datos para proyectar qué servicios hoteleros son de creciente importancia para los consumidores y serán impulsores clave del futuro del consumidor.

### **Reglas de Asociación**

En una primera instancia, el modelo podría recomendar una lista de eventos y promociones. Para ello es necesario capturar las preferencias de los pasajeros a través de los comportamientos históricos y así generar una lista de recomendaciones basadas en usuarios de perfil similar.

Luego, cuando el modelo aprende los hábitos de ese huésped particular, inserta anuncios personalizados en la lista de recomendación basados en los comportamientos de dicho usuario. Para ello es necesario determinar cuáles son las compras o actividades que los huéspedes realizan durante su estadía. El modelo podría tener en cuenta qué tan recientemente un cliente realizó una compra, la frecuencia o el número de veces que un cliente compra y el valor monetario de la compra.

### **1d) Etapa de interpretación/evaluación de datos**

En esta etapa buscamos identificar los patrones que son interesantes para el caso, los cuales deben cumplir ciertas condiciones: que sean de fácil comprensión, válidos para nuevos datos, potencialmente útiles y originales o inéditos. Por ejemplo, patrones de consumo de los huéspedes de los distintos eventos y promociones que ofrece el hotel (descuentos de happy hour, compras en tiendas, consumo de gimnasio o spa, asistencia a eventos).

## **1e) Presentación del Conocimiento**

Utilizar herramientas de visualización (histogramas, diagramas de tendencia, gráficos de torta, tablas de frecuencia, tablas de contingencia) y un informe de conclusiones para condensar grandes cantidades de datos en un formato conciso y comprensible.

### **Plan de Acción: Anuncios de TV van a ser enviados a los huéspedes durante su estadía**

En base al análisis de los datos, los anuncios de TV que serán enviados a los huéspedes serán los siguientes:

- Durante las primeras 48 horas los anuncios estarán basados en los datos que obtuvimos del check-in, del sistema de reserva y datos históricos de clientes con perfiles similares (segmentos, clusters). Por ejemplo, en este caso; al registrarse un cliente dentro de determinada edad, que viaje por turismo, con un grupo familiar y en vehículo propio se le ofrecerán servicios como excursiones, descuentos en restaurantes, actividades a realizar en familia, etc.
- Conforme la estadía del pasajero continúa, a partir de la obtención y el procesamiento de los nuevos datos de consumo y comportamientos de los huéspedes, se personalizan los anuncios en función de sus intereses: descuentos en tiendas y restaurantes de la cadena, souvenirs, restaurantes/happy hours, sightseeing / circuitos turísticos.

## **2. Decidir sobre la selección de hoteles para el mailing privado durante el invierno.**

En función de los clusters armados durante la etapa de data mining y una vez definidos quiénes son, qué están buscando, cómo se comportan, y cuál es la percepción y satisfacción de los servicios brindados, podremos predecir la intención de revisita de los huéspedes a algún hotel de la cadena y con esto proceder al armado de las campañas de marketing para la próxima temporada de verano. Estas campañas estarán dirigidas según los siguientes perfiles de clientes:

- Clientes Fieles (aquellos con estadía en varios hoteles de la cadena)
- Clientes Fieles Temporada (aquellos con estadía en varios hoteles de la cadena durante la misma temporada)

- Clientes Única compra (aquellos que no han regresado en un determinado periodo de tiempo)
- Clientes Nuevos (aquellos con estadía reciente)
- Clientes Disconformes (aquellos que dieron malas reseñas en encuestas internas y/o en plataformas de internet)

Una vez definido nuestro mercado objetivo y teniendo en cuenta que en la mayoría de los casos un huésped de hotel no elige muchas veces exactamente el mismo hotel, el contenido estará enfocado a la promoción de distintos destinos turísticos donde se encuentre un hotel de la cadena y que la persona todavía no haya visitado. Para esto también se deberá contar con un listado de hoteles con sus características, para poder ofrecer destinos afines a los intereses del usuario; por ejemplo, playa, cordillera, canchas de golf, parques temáticos, turismo rural, turismo cultural, etc.

Adicionalmente, se personalizará la oferta en base a incentivos como bonificaciones, descuentos, transporte; de acuerdo al perfil del pasajero. Por ejemplo, el correo dirigido a clientes disconformes incluiría ofertas que ofrezcan desayuno gratis o descuentos en la próxima estadía con el objetivo de mejorar su experiencia y así la opinión sobre la cadena.

## **Conclusiones :**

Después de haber planteado diferentes fuentes heterogéneas de datos y habiéndolo sometido a éstos al proceso completo de data mining, pudimos generar suficiente conocimiento acerca de los huéspedes, vinculándolos con el concepto de PRO-SUMIDOR, el consumidor que adicionalmente produce datos que permiten revelar sus preferencias, intereses, aspiraciones, deseos, etc.. Con internet se ha creado un escenario complejo para el prosumidor, que ha pasado de ser una figura ecléctica que combina las funciones relacionadas con el consumidor y productor a una figura que trasciende el ámbito económico, donde el prosumidor no solo se manifiesta en la creación de contenidos, sino que se encuentra ahora a ambos lados de un proceso productivo de una economía colaborativa.