

Ejercicio de reclutamiento para Personalización en Despegar.com

1) Usando el dataset provisto, construya y exponga la probabilidad de compra de un hotel para un usuario dependiendo de las variables existentes. Valoraremos el uso de una técnica distinta a logit o probit. Si se le presenta algún problema para computarlo, exponga de qué manera lo haría, qué significado tendría y cómo se interpretarían los resultados.

Algunas preguntas guías son:

¿En qué día de navegación es más probable que compre?

¿Cómo cambia la probabilidad de compra cuando cambian la cantidad de acciones?

¿Si tuviera que quedarse con 2 variables, cuáles son las más importantes?

¿Cómo mediría el ajuste del modelo?

¿Qué momento elegiría para hacer una oferta de precio? (con qué cantidad de días-acciones-etc.).

2) Piense en cómo aplicaría el conocimiento obtenido generando reglas para dar un cupón de descuento a un usuario que incentive la compra.

Describa paso a paso qué variables tendría en cuenta, y cómo lo implementaría (en qué instancia, forma de mostrarlo, comunicación, etc).

3) Agregue todo lo que le parezca que contribuya a los puntos 1) y 2), como descriptivas, gráficos, o código de análisis.

Flujo de compra de hoteles

El flujo típico por el cual se llega a comprar un hotel en despegar.com es el siguiente.

Al realizar un usuario una búsqueda por un destino en particular, este es presentado con una página de resultados de hoteles (search). Al “clickear” uno de estos hoteles, se presenta una página de detalle del mismo con la descripción del hotel, donde al seleccionar una habitación y presionar reservar se redirige al usuario a un formulario de checkout. Luego de completar todos los datos correspondientes y presionar comprar, el usuario es presentado con una página donde se agradece por haber realizado la compra. (Este flujo puede ser verificado fácilmente navegando el sitio)

Además de estos eventos típicos (search, detail, checkout y thanks), existen otras acciones que puede realizar el usuario y que fueron agrupadas en el dataset.

Descripción de las variables

“userid”: Identificador único del browser de navegación.

“Date”: Fecha (YYYY-MM-DD) de la sesión actual

“cant_actions”: Cantidad de eventos del usuario en la sesión previa

“search”: Cantidad de eventos de búsqueda de la sesión previa

“detail”: Cantidad de eventos de detalle de la sesión previa

“checkout”: Cantidad de eventos de checkout de la sesión previa

“thanks”: Cantidad de eventos de compra de la sesión previa

“comprador”: Es la variable que queremos predecir, si compró o no algún hotel.

“comprador_sess”: Si compró un hotel en la sesión previa

“used_tablet”: Si el dispositivo usado es una tablet

“used_smartphone”: Si el dispositivo usado es un celular

“timeprox”: Distancia en milisegundos entre la sesión actual y la previa

“longsess”: Duración en milisegundos de la sesión previa

“cantsesiones”: Cantidad de sesiones previas a la actual en los últimos 15 días

“dur_media”: Duración media (en segundos) de la estadía buscada en la sesión previa

“ant_media”: Anticipación media (en segundos) entre la fecha de la sesión previa y las fechas de check-in buscadas