

Proyecto final

Problema

Usted es el científico de datos de una empresa de telefonía celular y tiene que proporcionar soluciones para hacer frente a las problemáticas de un sector que ha llegado a saturación del mercado. Tanto su empresa como sus competidores directos tienen que disputarse por una base de clientes limitada, de tal forma que usted tiene que responder a un objetivo estratégico definido por la dirección así:

“Mantener y fidelizar a nuestros clientes por medio de un servicio de calidad que se adapte a sus necesidades particulares.”

Su compañía dispone de una base de datos histórica de personas que hace un año eran clientes propios. Algunos de esos clientes siguen siéndolo hoy en día, otros ya no lo son.

La idea es crear un modelo predictivo de deserción que permita identificar los clientes propensos a irse a las empresas en directa competencia con la suya en el próximo año.

Puntos a desarrollar

1. **Limpieza y EDA:** Verifiquen si hay problemas de calidad de datos.
Se espera una primera sección de evaluación de la calidad de los datos y de entendimiento de la relación entre las variables predictivas y la variable objetivo (**OJO!** Solo poner gráficos y análisis de las relaciones importantes, **menos es más!**)
2. **Modelos predictivos:** Entrenen modelos predictivos (al menos 3 familias de modelos) que permitan estimar si el cliente abandonará la empresa o no. Escoja el mejor modelo, buscando sus parámetros óptimos.
Se espera una sección donde se establezca el protocolo de evaluación y los procesos de entrenamiento y evaluación de los modelos.
3. **Cambio de representación del dataset:** Extraiga los componentes principales, analice sus niveles de varianza explicada, e interprete los 3 más importantes, en función de las variables originales.
4. **Caracterización de los clientes:** Analice los clientes que se han ido, creando un modelo de segmentación de los clientes que desertan la compañía, teniendo en cuenta sus datos socio-demográficos y comportamientos de consumo del servicio de telefonía. Interpretar el perfil de clientes asignado a cada segmento, caracterizándolos de tal manera que le permita sugerir 3 a 5 campañas de fidelización.

Diccionario de datos:

- - ESTADO: Describe si el usuario sigue con la compañía (VINCULADO) o no (RETIRADO)
- - INGRESOS: Promedio de ingresos del cliente en pesos
- - CASA: Precio de la casa en la que vive el cliente en pesos
- - PRECIO_DISPOSITIVO: Precio del celular del cliente en pesos
- - MESES: Antigüedad del usuario en meses
- - DURACION: Promedio de duración de las llamadas hechas por el cliente en minutos
- - SOBRECARGO: Promedio de minutos que se sobrepasa el usuario en un mes
- - SALDO_RESTANTE: Promedio de minutos de su plan que le quedan al usuario sin utilizar cada mes
- - SATISFACCION: nivel de satisfacción del usuario de 0 a 10 (muy satisfecho), obtenido a partir de una encuesta.

Rúbricas de puntuación

1ª entrega: Sesión 9

Calidad de datos	Visualización de datos	Extracción de intuiciones de los datos	TOTAL PROYECTO PRIMERA ENTREGA
1.0	2.0	2.0	5.0

2ª entrega: Sesión 12

Entendimiento de los datos y limpieza	Protocolo de entrenamiento y evaluación de modelos	Entrenamiento de los 3 modelos	TOTAL PROYECTO SEGUNDA ENTREGA
1.0	1.0	3.0	5.0

3ª entrega: Sesión 15

Transformación de los datos por PCA	Caracterización de perfiles de clientes	TOTAL PROYECTO TERCERA ENTREGA
2.0	3.0	5.0