

Proyecto final

Problema

Usted es el encargado de analítica 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.

Se han identificado dos proyectos de analítica de datos que permitirán alcanzar tal objetivo, que tendrá que desarrollar en las dos partes siguientes.

Parte 1.

Cree 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.

1. ¿Encuentra alguna anomalía en los datos? (0.3)
2. La empresa considera que los valores de variables que estén a más de 4 desviaciones estándar del promedio deberían ser consideradas excepcionales, y por lo tanto no se deben considerar en los análisis. Identifíquelas y apártelas del dataset (0.3)
3. Analice la correlación entre las variables y explique lo que puede implicar desde el punto de vista de PCA. (0.2)
4. Debe entrenar 3 tipos de modelos predictivos de diferentes familias:
 - a. Defina el **protocolo** de evaluación que va a utilizar para calibrar los modelos y estimar la calidad del modelo final. (0.3)
 - b. Establezca las **métricas** que va a utilizar, justificando su escogencia (0.2)
 - c. Calibre **3** tipos de modelos diferentes: K-NN, árbol de decisión y algún otro que propongan, utilizando las métricas y protocolo definido (1.0)
 - d. Evalúe los 3 modelos encontrados, escoja el mejor, **explicando y concluyendo lo que encontró**, utilizando las métricas y protocolo definido (0.5)

Parte 2.

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.

1. Definición del número de campañas a realizar (0.6)

2. Extraiga los componentes principales, analice sus niveles de varianza explicada, e interprete los 3 más importantes en función de las variables originales. (0.6)
3. Compare de los clusters obtenidos utilizando K-Means y Clustering jerárquico, seleccionando los resultados de uno de los dos métodos. **Justifique.** (0.4)
4. Caracterice los clusters, **etiquetando** el segmento; proponga una estrategia de fidelización para cada uno de los clusters. **Justifique** (0.6)

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.

Notas:

- Entregar un archivo **PDF o HTML únicamente**
- El proyecto es por grupos de máximo **2 personas**.
- **Fecha de entrega: 19 de junio**
- No olvidar responder a los puntos formulados. En la primera parte hay 3 puntos sencillos y un cuarto con 4 sub puntos, cada uno de ellos muestra su peso en la nota final.
- Por favor limitarse en las explicaciones y justificaciones con unas cuantas pocas líneas por pregunta es suficiente.
- Por favor limitarse en los gráficos utilizados, escogiendo únicamente los que tengan una importancia argumentativa.
- Tengan cuidado al utilizar markdown, puede que los clusters / colores aleatorios que ustedes ven en RStudio no correspondan con los del documento generado. Primero génenlo y luego hagan la interpretación sobre los resultados que da el markdown.