

## Ejercicio: Abandono de clientes en empresa de Telecomunicaciones

Una empresa de telecomunicaciones tiene una tasa de abandono de clientes elevada y nos ha proporcionado sus datos para ver si podemos encontrar algo de valor que ayude a mejorarla.

Utilizando el Excel adjunto (“Abandono\_clientes\_Telco.xlsx”), contesta a las siguientes preguntas:

1. (1 punto) Describe la problemática de negocio. ¿Qué necesita mejorar la empresa? ¿Por qué necesita mejorarlo?

El negocio tiene un problema de *tipo de clasificación*. Necesita mejorar su modelo de estimación de abandono para poder emplear tácticas que ayuden a reducir la tasa de abandono de clientes.

2. (1 punto) ¿Cómo ayudaría un modelo de predicción de abandono a resolver el problema de negocio?

Con un modelo de predicción de abandono la empresa puede estimar cuáles son los clientes que cumplen con características previamente observadas en aquellos que terminaron por abandonar la empresa y, con esta información, podrían establecer mejores tácticas para evitarlo.

3. (1 punto) Identifica cuál sería la variable objetivo y su tipo de variable. ¿De qué clase de modelo se trataría?

La variable objetivo es la variable “Abandono” y es una variable categórica. El modelo es supervisado de clasificación.

4. (2 puntos) Identifica qué variables de entrada son categóricas, cuales son continuas y si alguna podría considerarse ordinal.

- Variables categóricas:

Genero, Jubilado, Casado, Hijos, ServicioTelefono,  
MultiplesLineas, ServicioInternet, SeguridadOnline, OnlineBackup,

ProteccionDispositivo, SoporteTecnico, StreamingTV,  
StreamingMovies, FacturaElectronica, MetodoPago, Abandono.

- Variables Continuas:  
MesesServicio, CuotaMensual, CuotaTotal
- Variables Ordinales:  
Contrato.

**5. (1 punto) ¿Cuál es la calidad general de los datos? ¿Eliminarías alguna variable o registro por falta de calidad?**

En los datos provistos, tenemos un total de 7043 registros, 20 variables de los cuales 0% es el porcentaje de valores nulos. No eliminaría ninguna variable o registro por falta de calidad solo convertiría “Jubilados” a una variable categórica.

**6. (1.5 puntos) Describe al menos 3 gráficas (ej.: histograma de la cuota mensual) que dibujarías para entender estos datos, y que información esperarías obtener de ellas.**

1. Histograma para ver la relación entre meses de servicio con la cuota mensual de aquellos que han abandonado.
2. Diagrama de barras para ver la relación entre el abandono de los clientes con el servicio contratado.
3. Diagrama de dispersión para ver si el tipo de contrato(mensual, anual, ...) es directamente proporcional al abandono por parte del usuario.

**7. (1 punto) ¿Qué variables esperarías que fueran las más relevantes? ¿Eliminarías alguna variable que no aporte información al modelo?**

Las variables que esperaba sean relevantes son el tipo de contrato, los meses que tiene y los tipos de servicios contratados (internet, teléfono, streaming). Sobre eliminar, creo que el método de pago no me aportaría información, tampoco el estado marital o el estado de jubilación.

**8. (0.5 puntos) Menciona 3 algoritmos de Machine Learning que podrías emplear en la modelización.**

1. K primeros vecinos
2. Árboles de Decisión
3. Redes Neuronales

9. (1 punto) Suponiendo que se han entrenado tres modelos y se obtienen las siguientes matrices de confusión. ¿Qué modelo escogerías? ¿Por qué?

Escogería el segundo modelo por las siguientes razones:

1. Lo observado(realidad) contra lo estimado(modelo dice):
  - a. Cuando vemos el porcentaje de verdaderos positivos con respecto a los verdaderos totales dicho por el modelo, para el primer modelo representaría un 53%, en el caso del segundo modelo, **representa un 96%**; para el tercer modelo representaría un 16%.
  - b. Con respecto a la proporción de verdaderos negativos con respecto a los negativos totales dichos por el modelo, en el caso del modelo 1 un 80%, **el modelo 2 tenemos un 65%**, mientras que el modelo 3 un 94%.
2. Con respecto a la **precisión** con que los modelos escogen correctamente, los verdaderos positivos es de un 50% mientras que, escogiendo los negativos, el primer modelo lo hace con un 75% de exactitud, **el segundo con un 98%** y el tercero con un 82%.

Modelo 2 (acierto 73%)		Abandono	
		Si	No
Modelo dice:	Si	1800	1800
	No	69	3374

Modelo 3 (acierto 73%)		Abandono	
		Si	No
Modelo dice:	Si	300	300
	No	1569	4874

Modelo 1 (acierto 73%)	Abandono
---------------------------	----------

		Si	No
Modelo dice:	Si	1000	1000
	No	869	4174