

Ejercicio sobre Minería de Datos y Aprendizaje Automático

Situación:

Una empresa de telecomunicaciones desea reducir la tasa de abandono de clientes. Para ello, decide aplicar técnicas de minería de datos y aprendizaje automático para predecir qué clientes tienen una alta probabilidad de cancelar su servicio en los próximos 3 meses.

Parte 1: Preparación de los Datos

1. ¿Cuáles son las principales fases del preprocesado de los datos? Menciona y explica al menos 3 técnicas utilizadas en esta etapa.
2. ¿Por qué es importante el muestreo en el preprocesamiento de datos? Explica una situación en la que sería recomendable usar muestreo estratificado en este caso.

Parte 2: Modelización

3. ¿Qué diferencia hay entre un modelo supervisado y uno no supervisado? ¿Cuál crees que es más adecuado para este problema? Justifica tu respuesta.
4. Considerando que queremos clasificar a los clientes en dos grupos (probable abandono o retención), ¿qué tipo de modelo supervisado recomendarías? Explica por qué.
5. En un árbol de decisión, ¿qué criterio se utiliza para dividir los datos en diferentes nodos? Explica brevemente el concepto de "pureza" en este contexto.

Parte 3: Evaluación y Producción

6. Se ha entrenado un modelo de clasificación y se obtiene una matriz de confusión. ¿Cuáles son las métricas clave para evaluar su desempeño? Define **precisión**, **recall** y **exactitud**.
7. Una vez entrenado el modelo, se quiere poner en producción. ¿Qué método recomendarías para aplicar la inferencia en tiempo real en una plataforma web? Explica brevemente.