**1. Clasificación de Cada Problema**

El primer paso consiste en clasificar cada problema de negocio según el tipo de tarea de aprendizaje de máquina que representa:

* **Problema A: Predecir el monto de ventas semanales.**
  + **Clasificación:** **Regresión**. El objetivo es predecir un valor numérico continuo (el monto de las ventas).
* **Problema B: Detectar si un cliente abandonará un servicio.**
  + **Clasificación:** **Clasificación Binaria**. El objetivo es predecir una categoría discreta con dos posibles resultados (el cliente "abandonará" o "no abandonará").
* **Problema C: Agrupar a los clientes por comportamiento de gasto.**
  + **Clasificación:** **Aprendizaje No Supervisado (Clustering)**. El objetivo es encontrar grupos o segmentos naturales en los datos sin tener etiquetas predefinidas.

**2. Modelos Propuestos y Justificación**

Para cada problema, se selecciona un algoritmo adecuado, justificando la elección con sus ventajas y limitaciones.

**Problema A: Regresión (Predicción de Ventas)**

* **Modelo Propuesto:** **Gradient Boosting**
* **Justificación:** Los modelos de Gradient Boosting son el estándar de la industria para datos tabulares.
  + **Ventajas:** Suelen alcanzar la **máxima precisión** al poder capturar relaciones complejas y no lineales entre las variables (ej. estacionalidad, efecto de promociones). Son computacionalmente eficientes.

**Problema B: Clasificación**

* **Modelo Propuesto:** **Random Forest (Bosque Aleatorio)**.
* **Justificación:** Es un modelo de ensamble robusto y de alto rendimiento.
  + **Ventajas:** Maneja bien diferentes tipos de datos, es **menos propenso al sobreajuste** que un único árbol de decisión y ofrece **importancia de variables**, lo que permite entender qué factores influyen más en el abandono de un cliente (ej. "días de inactividad").

**Problema C: Clustering (Segmentación de Clientes)**

* **Modelo Propuesto:** **K-Means**.
* **Justificación:** Es el algoritmo de clustering más conocido y utilizado.
  + **Ventajas:** Es **simple de entender, rápido y escalable** a grandes conjuntos de datos. Es un excelente punto de partida para la segmentación.

**3. Análisis de Riesgos y Desafíos**

La implementación de estos modelos conlleva varios desafíos comunes que deben ser gestionados:

|  |  |  |
| --- | --- | --- |
| Riesgo / Desafío | Descripción | Problemas Afectados |
| **Calidad de los Datos** | Los modelos son sensibles a datos incorrectos, faltantes o atípicos. "Basura entra, basura sale". | A, B, C |
| **Overfitting (Sobreajuste)** | El modelo aprende el ruido de los datos de entrenamiento y no generaliza bien a datos nuevos. Es un riesgo alto en modelos complejos. | A, B |
| **Datos Desbalanceados** | En el problema de abandono, es probable que haya muchos más clientes que "no abandonan" que los que sí. El modelo podría ignorar a la clase minoritaria, que es la más importante. | B |
| **Selección de K** | En K-Means, elegir un número incorrecto de clústeres (K) puede generar segmentos de clientes que no tienen sentido para el negocio. | C |

4. Esquema

1. Preparación de datos ⟶ 2. Selección de modelo ⟶ 3. Entrenamiento ⟶ 4. Evaluación ⟶ 5. Despliegue.