# Informe de Aprendizaje Automático

## 1. Clasificación del problema

Problema A: Predecir el monto de ventas semanales de una cadena de supermercados.  
→ Tipo de tarea: Regresión  
  
Problema B: Detectar si un cliente abandonará un servicio de streaming.  
→ Tipo de tarea: Clasificación  
  
Problema C: Agrupar a los clientes bancarios por comportamiento de gasto sin tener etiquetas.  
→ Tipo de tarea: No supervisado

## 2. Selección de modelo y justificación

Problema A: Regresión

Modelo: Regresión lineal o Árbol de decisión regresivo  
Justificación: Permite predecir valores continuos como el monto de ventas.  
Ventajas: Fácil de interpretar, rápido de entrenar.  
Limitaciones: Sensible a outliers, puede no capturar relaciones no lineales.

Problema B: Clasificación

Modelo: Árbol de decisión o Random Forest  
Justificación: Clasifica clientes en 'abandona' o 'no abandona'.  
Ventajas: Maneja variables categóricas, interpretable.  
Limitaciones: Puede sobreajustarse si no se poda correctamente.

Problema C: No supervisado

Modelo: K-means  
Justificación: Agrupa clientes según patrones de gasto sin etiquetas.  
Ventajas: Simple, eficiente en grandes volúmenes.  
Limitaciones: Sensible a la escala de datos y al número de clusters elegido.

## 3. Identificación de riesgos

• Maldición de la dimensionalidad: Muchos atributos pueden dificultar el aprendizaje, especialmente en clustering.  
• Overfitting: Modelos complejos pueden ajustarse demasiado a los datos de entrenamiento.  
• Datos desbalanceados: En clasificación, si hay pocos casos de abandono, el modelo puede ignorarlos.  
• Datos faltantes o ruidosos: Afectan la calidad del modelo y su capacidad predictiva.

## 4. Reflexión personal sobre la aplicación práctica de los conceptos

Integrar estos conceptos en proyectos reales permite tomar decisiones basadas en datos, anticipar comportamientos y optimizar recursos. La clave está en entender el problema, elegir el modelo adecuado y evaluar continuamente su desempeño. Además, la colaboración entre perfiles técnicos y de negocio es esencial para lograr soluciones efectivas y sostenibles.

## 5. Métricas de evaluación (Opcional)

• Regresión: MAE, RMSE, R²  
• Clasificación: Accuracy, Precision, Recall, F1-score  
• Clustering: Silhouette Score, Within-cluster SSE

## 6. Esquema gráfico del flujo de trabajo para el Problema A

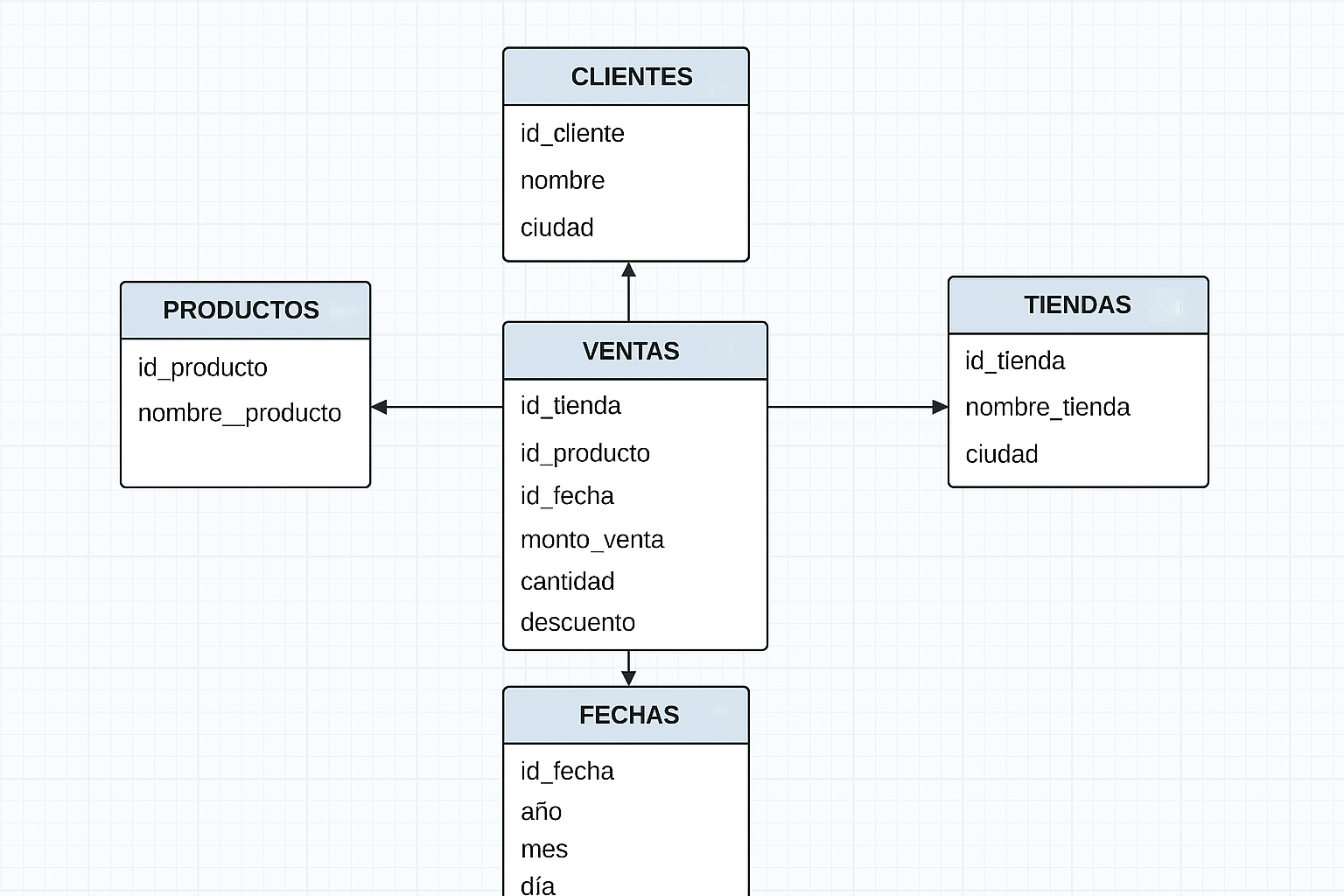


Figura: Flujo de trabajo de aprendizaje automático para predecir ventas semanales.