MODULO 06 - EJERCICIO 02-B

ALEXIS YURI M.

Simule las etapas de un proyecto de aprendizaje automático aplicable a una necesidad real desde la definición del problema hasta la validación del modelo.

1) Selección del problema

Estimación de demanda semanal para una cadena de retail.

El objetivo es predecir la cantidad de ventas que se realizarán por sucursal y por producto cada semana, con el fin de optimizar el inventario y evitar quiebres de stock o exceso de inventario.

2) Esquema breve:

Tipo de aprendizaje: Supervisado.

Tipo de tarea: Regresión (predicción continua de cantidad vendida)

Datos necesarios y posibles fuentes:

Historial de ventas: ventas semanales por tienda y por producto.

Calendario de ventas: días de la semana, feriados, promociones, campañas de marketing, eventos especiales.

Inventario: niveles de stock actuales, quiebres de stock previos (para manejar “ventas no realizadas”).

Precios: histórico de precios, descuentos, promociones.

Características de productos y sucursales: tamaño de la sucursal, ubicación, categorías de productos, etc.

3) Etapas a seguir

Exploratory Data Analysis (EDA)

Análisis de tendencias:

Ver si hay una tendencia estacional en las ventas (por semana, mes).

Ver si hay alzas de demanda debido a campañas, eventos o feriados.

Revisión de distribuciones:

Verificar las distribuciones de las ventas para cada sucursal y producto.

Identificar valores atípicos (outliers) por eventos especiales o errores en los datos.

Revisión de desbalanceo:

Por ejemplo si las ventas suelen ser de tipo long tail (muchos productos con pocas ventas y pocos productos con muchas ventas).

Valores faltantes:

Revisar los valores faltantes y determinar cómo tratarlos (por ejemplo, relleno con la media o usar interpolación).

División de datos:

Entrenamiento, validación y test:

Entrenamiento: Datos históricos hasta el mes actual.

Validación: Datos del mes pasado (para validar el modelo).

Test: Datos de la última semana del trimestre (para evaluar rendimiento final).

Características de entrada (Variables explicativas):

Ventas pasadas: ventas de semanas anteriores (p. ej., ventas\_{t-1},

ventas\_{t-2}, ventas\_{t-4}, etc.).

Media móvil: medias de ventas de las últimas 4, 8 y 12 semanas.

Indicadores de campaña/promoción: variables binarias para indicar campañas activas o promociones.

Precios: precio del producto en la semana y descuentos aplicados.

Escalado y codificación:

Si se utilizan modelos sensibles a escala (como SVR), aplicar StandardScaler.

Para variables categóricas (por ejemplo, sucursal, categoría de producto), usar One-Hot Encoding.

4) Algoritmos candidatos:

- Predicción simple: ventas\_t = ventas\_{t-1} o ventas\_t = media\_{t-52} (modelo de referencia para comparar).

- Regresión Lineal: para predecir la demanda a partir de características de entrada (ventas pasadas, precios, indicadores de campaña).

- Modelos basados en árboles de decisión, muy eficaces para capturar interacciones no lineales y manejar desbalanceos en los datos.

5) Métricas de evaluación:

- sMAPE (Error Porcentual Absoluto Medio Simétrico): Mide el error relativo de forma simétrica.

- RMSE (Error Cuadrático medio): Mide la diferencia promedio entre las predicciones y los valores reales.

- WAPE (Error Porcentual Absoluto Ponderado): Es útil cuando se quiere penalizar más los errores en productos de mayor volumen.

- Precisión y Recall por producto: Se usan estas métricas de clasificación binaria si el negocio se enfoca en no tener quiebres de stock, puede ser útil monitorear cómo el modelo está priorizando las predicciones en productos críticos.