

Ejercicio propuesto

Predicción y Análisis de Ventas de una Tienda Online

Contexto

Una tienda online llamada **TecnoShop** vende productos electrónicos (audífonos, teclados, cámaras, laptops, etc.) y desea analizar su información de ventas para **identificar patrones, predecir ventas futuras y visualizar sus resultados estratégicos en Power BI**.

La empresa te ha contratado como analista de datos para realizar un estudio completo de minería de datos y business intelligence utilizando Python y Power BI.

Objetivo general

Aplicar técnicas de análisis de datos, modelado predictivo y visualización de información usando las herramientas **Pandas, NumPy, scikit-learn y Power BI**, integrando los resultados de modelos como **Regresión Lineal, K-Veinos más Cercanos y Árboles de Decisión**.

Instrucciones generales

1. Dataset base:

Genera o descarga un dataset simulado de ventas con las siguientes columnas (mínimo 200 registros):

- Id
- Producto
- Categoría
- Precio
- Unidades vendidas
- Mes
- Región
- Publicidad
- Satisfacción Cliente

Parte 1: Preparación de los datos (Pandas y NumPy)

1. Cargar los datos en un DataFrame de Pandas.
2. Explorar el dataset: conteo, valores nulos, medias, etc.
3. Crear nuevas columnas calculadas, por ejemplo:
 - $\text{IngresoTotal} = \text{Precio} * \text{UnidadesVendidas}$
 - $\text{Rentabilidad} = \text{IngresoTotal} * 0.2$
4. Aplicar algunas operaciones con NumPy, como normalización o cálculo de estadísticas (promedios, desviación estándar, etc.).
5. Guardar el DataFrame limpio en un nuevo archivo `ventas_procesadas.csv`.

Parte 2: Modelos de aprendizaje supervisado (scikit-learn)

Usando el dataset procesado: Modelo 1 –

Regresión lineal

- Objetivo: Predecir las Unidades Vendidas en función del Precio y la Publicidad.
- Graficar la línea de regresión con un scatter plot (matplotlib).
- Evaluar con R^2 y error cuadrático medio. Modelo 2 – K-Vecinos más Cercanos (KNN)

- Objetivo: Clasificar el nivel de Satisfacción del Cliente (por ejemplo, *Alta / Media / Baja*) según Precio y Publicidad.
- Usar KNeighborsClassifier.
- Mostrar matriz de confusión. Modelo 3 – Árbol de Decisión
- Objetivo: Predecir la Categoría del Producto según Precio, Publicidad y Unidades Vendidas.
- Visualizar el árbol (usando plot_tree o graphviz).
- Interpretar qué variables son más relevantes en la decisión.

Al final, guarda los resultados (predicciones y métricas de cada modelo) en un nuevo CSV: resultados_modelos.csv

Parte 3: Integración con Power BI

Exporta los siguientes archivos desde tu notebook:

1. ventas_procesadas.csv
2. resultados_modelos.csv

creación de dashboards en Power bi:

Importa ambos CSV y crea **4 dashboards**:

Dashboard	Contenido	Visualizaciones sugeridas
1. Análisis de Ventas	Comparación de ingresos por categoría, mes y región	Barras, mapa, tarjeta de total de ingresos
2. Rendimiento de Productos	Precio vs Unidades Vendidas, Rentabilidad promedio	Scatter plot, barra agrupada, tarjetas
3. Predicción de Ventas (Regresión)	Línea de tendencia real vs predicha	Scatter plot + línea de tendencia
4. Modelos de Clasificación	Resultados de KNN y Árbol de Decisión	Gráfico de barras con precisión por modelo, árbol visualizado, matriz de confusión importada como imagen o tabla