

Universidad Yachay Tech

Maestría en Ciencia de Datos - Taller II

Título: Aprendizaje de Máquina para el Análisis del Rendimiento de Ventas en Comercio Electrónico

Objetivos:

1. **Algoritmos de aprendizaje automático:** Utilizar algoritmos de aprendizaje automático para el análisis de datos de comercio
2. **Clasificación de clientes:** Clasificación de clientes en varias categorías.
3. **Segmentación de Clientes:** Utilizar técnicas de agrupamiento para segmentar a los clientes basándose en su comportamiento de compra.
4. **Predicción de Ventas:** Desarrollar un modelo predictivo para pronosticar ventas futuras basándose en datos históricos.

Instrucciones:

Conjunto de Datos: Utilizar el conjunto de datos Online Retail II del Repositorio de Aprendizaje Automático de UCI Irvine Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>). Este conjunto de datos contiene todas las transacciones que ocurren para un comercio electrónico con sede en el Reino Unido entre el 01/12/2009 y el 09/12/2011.

Todo el trabajo debe ser realizado por medio de Sckit-learn.

Parte 0: Preparación

- Exploración y limpieza de Datos
 - Descargar el conjunto de datos Online Retail II.
 - Cargar el conjunto de datos utilizando pandas y realizar una exploración inicial para entender la estructura y contenido de los datos.
 - Limpiar los datos manejando valores faltantes, eliminando registros duplicados y filtrando entradas irrelevantes (por ejemplo, transacciones sin ID de clientes o pedidos cancelados).
- Análisis Exploratorio de Datos
 - Analizar la distribución de variables clave (por ejemplo, cantidad de productos comprados, precio de los productos y fecha de compra).
 - Identificar los productos más vendidos y visualizar sus tendencias de ventas a lo largo del tiempo.
 - Explorar los patrones estacionales en los datos de ventas e identificar los períodos de máximas ventas.

Parte 1: Clasificación de clientes:

- Definir una nueva columna en los datos para trabajar con dos clases: “Clientes Normales” y “Clientes Premium”. ¿Cómo podríamos definir estas categorías, por volumen de compras, cantidades, etc? Definir este campo.
- Entrenar un modelo de clasificación para la clasificación de los clientes en estas dos categorías.

Parte 2: Segmentación de Clientes

- Usar el agrupamiento K-means para categorizar a los clientes en grupos distintos y analizar las características de cada segmento.
- Usar el agrupamiento Mean Shift para categorizar a los clientes en grupos distintos y analizar las características de cada segmento.
- Comparar los resultados de las dos técnicas anteriores

Parte 3: Predicción de Ventas

- Crear nuevas características, si es necesario, para ayudar en el modelado predictivo (por ejemplo, época del año, categorías de productos).
- Dividir los datos en conjuntos de entrenamiento y prueba.
- Elegir un algoritmo de aprendizaje automático adecuado (por ejemplo, regresión lineal, árbol de decisión o bosque aleatorio) para predecir ventas futuras.
- Evaluar el modelo utilizando métricas de rendimiento apropiadas.