

Prueba Técnica científico De Datos

Contexto de negocio:

Estás trabajando con un conjunto de datos de ventas de una cadena minorista. El conjunto de datos contiene información sobre las ventas diarias de varias tiendas, incluyendo datos como la fecha de la venta, el número de unidades vendidas, el precio por unidad, y la categoría del producto.

Tu tarea es construir un modelo de predicción para estimar el número de unidades que se venderán en el futuro basado en las ventas pasadas y otras características del conjunto de datos.

Tienes un archivo CSV llamado `sales_data.csv` con las siguientes columnas:

- Date: La fecha de la venta.
- Store: El identificador único de la tienda.
- Category: La categoría del producto.
- Units_Sold: El número de unidades vendidas.
- Unit_Price: El precio por unidad.

Objetivos de la prueba:

1. Realiza un análisis exploratorio de los datos para identificar patrones, tendencias, valores atípicos, y posibles relaciones entre variables.
2. Crea visualizaciones que ayuden a comprender mejor los datos.
3. Limpia los datos eliminando valores nulos y manejando datos faltantes o duplicados.
4. Normaliza o estandariza las características numéricas si es necesario.
5. Codifica las características categóricas utilizando técnicas adecuadas.
6. Crea un modelo de regresión para predecir las ventas futuras basado en los datos históricos.
7. Evalúa el modelo utilizando métricas adecuadas (por ejemplo, RMSE, MAE).
8. Interpreta los resultados del modelo y proporciona recomendaciones sobre cómo mejorar el modelo o ajustar las estrategias de ventas basadas en los resultados.

Entregables

1. Informe de Análisis Exploratorio de Datos (EDA) - PDF

- a. Un informe detallado que describa los patrones, tendencias y relaciones identificadas en el conjunto de datos. Este informe debe incluir:
 - i. Estadísticas descriptivas (media, mediana, desviación estándar, etc.) para cada variable.

- ii. Análisis de correlaciones entre las variables.
- iii. Identificación de valores atípicos y su posible impacto en el análisis.

2. Informe de Preprocesamiento de Datos - .py

- a. Un informe que detalle los pasos tomados para limpiar y preprocesar los datos. Este informe debe incluir:
 - i. Métodos utilizados para manejar valores nulos, faltantes y duplicados, y justificación para cada método.
 - ii. Proceso de normalización o estandarización de las características numéricas, si fue necesario.
 - iii. Técnica(s) utilizada(s) para codificar características categóricas (por ejemplo, one-hot encoding, label encoding) y justificación de la elección.

3. Modelo Predictivo y Evaluación - .py / PDF

- a. Un archivo que contenga el modelo predictivo entrenado para predecir las ventas futuras basadas en los datos históricos.
- b. Descripción detallada del proceso de construcción del modelo, incluyendo:
 - i. Algoritmo(s) de aprendizaje utilizados (por ejemplo, regresión lineal, regresión polinómica, etc.).
 - ii. Procesos de selección de características y de tuning de hiperparámetros.
 - iii. Evaluación del modelo utilizando métricas apropiadas (por ejemplo, RMSE, MAE), y comparación entre diferentes modelos (si se construyeron varios)
 - iv. Justificación de la elección del modelo final.

4. Informe de Interpretación de Resultados y Recomendaciones - PDF

- a. Un informe final que interprete los resultados del modelo predictivo y proporcione recomendaciones estratégicas. Este informe debe incluir:
 - i. Interpretación de los coeficientes del modelo y cómo las diferentes características influyen en las ventas.

- ii. Identificación de posibles mejoras en la precisión del modelo (por ejemplo, más datos, diferentes características, técnicas de modelado avanzadas).
 - iii. Recomendaciones prácticas para el equipo de ventas basadas en los resultados del modelo (por ejemplo, enfoque en productos específicos, estrategias de precios).
- b. Reflexión sobre las limitaciones del análisis y sugerencias para futuros trabajos.