

PROPUESTA DE MONOGRAFÍA (Entregable 1)

Título del proyecto	Análisis comparativo de algoritmos de regresión para predecir el monto transaccionado por ticket en el sector retail.	
Estudiante 1		
Nombres completos	Lina María Beltrán Durango	e-mail:
		GitHub: https://github.com/lmbd92/DataScienceMonograph
Estudiante 2		
Nombres completos	Mario Earles Otero Andrade	e-mail:
		GitHub: https://github.com/lmbd92/DataScienceMonograph

1. Descripción del problema

El sector retail se enfrenta al desafío de **predecir los montos transaccionados por ticket** en sus proyecciones de venta, lo cual es fundamental para optimizar sus estrategias de negocio y tomar decisiones informadas. En este contexto, se plantea la necesidad de realizar un análisis comparativo de algoritmos de regresión con el objetivo de determinar cuál de ellos ofrece el mejor rendimiento en términos de precisión y eficiencia. Los resultados de este análisis comparativo serán de utilidad para las áreas analíticas anexas al sector retail, ya que les proporcionarán una base para seleccionar el algoritmo más apropiado para sus necesidades específicas.

Este análisis aportaría a la búsqueda de soluciones para las siguientes problemáticas de negocio:

- **Predicción precisa de ventas:** La capacidad de predecir con precisión el monto transaccionado por ticket permitirá a las empresas del sector retail tener una visión más clara de sus ingresos futuros. Esto les ayudará a planificar de manera más efectiva la gestión de inventario, la asignación de recursos y la elaboración de presupuestos.
- **Optimización de estrategias de precios:** Al comprender mejor el monto transaccionado por ticket, las empresas podrán ajustar sus estrategias de precios para maximizar sus ganancias. Identificar patrones y tendencias en los datos de transacciones les permitirá establecer precios óptimos, promociones adecuadas y descuentos atractivos para los clientes.
- **Mejora de la planificación de inventario:** Conociendo la predicción del monto transaccionado por ticket, las empresas podrán realizar una planificación de inventario más precisa y eficiente. Esto evitará la escasez de productos populares y reducirá los costos asociados al exceso de inventario.
- **Toma de decisiones informadas:** Al contar con modelos predictivos robustos, basados en el análisis comparativo de algoritmos de regresión, las empresas del sector retail podrán tomar decisiones más informadas en áreas como la expansión a nuevos mercados o la selección de productos y proveedores.

En general, este análisis ayudaría a las empresas del sector retail a comprender mejor su rendimiento financiero, optimizar sus operaciones y estrategias comerciales, y mejorar la satisfacción del cliente al ofrecer experiencias más personalizadas y adecuadas a sus necesidades.

2. Descripción del dataset

Este dataset se llama “Online Retail II UCI” contiene todas las transacciones que se produjeron en una tienda minorista en línea registrada y con sede en el Reino Unido entre el 12/01/2009 y el 12/09/2011. La empresa vende principalmente artículos de regalo únicos para toda ocasión. Muchos clientes de la empresa son mayoristas. Cuenta con 8 características entre numéricas, categóricas y temporales, con un total de 1.067.371 registros.

Para más detalles, el link donde se puede encontrar el dataset es el siguiente:

[Online Retail II UCI | Kaggle](#)

3. Métricas de desempeño

Para evaluar el desempeño de los modelos de regresión, se utilizarán varias métricas, como el error cuadrático medio (MSE), el coeficiente de determinación (R^2), el error absoluto medio (MAE), y el error absoluto porcentual medio (MAPE). Además, se utilizarán técnicas de validación cruzada de manera que se pueda evaluar la capacidad del modelo para generalizar a datos nuevos que no han sido utilizados en el entrenamiento.

4. Criterio de desempeño

Los criterios de desempeño para los modelos de regresión serán la precisión, la simplicidad, la estabilidad y la capacidad de generalización. La precisión se evaluará mediante las métricas de desempeño mencionadas anteriormente. La simplicidad se evaluará en función de la complejidad del modelo y la facilidad de interpretación. La estabilidad se evaluará mediante la sensibilidad del modelo a cambios en los datos de entrada. La capacidad de generalización se evaluará utilizando técnicas de validación cruzada.