



Predicción de la demanda en tiendas retail con series de tiempo

Gabriel Fernando Franco Calvo

E-mail: gfernando.franco@udea.edu.co

GitHub: <https://github.com/fdofrancoc>

Jessica Torres Franco

E-mail: jessica.torresf@udea.edu.co

GitHub: <https://github.com/JessieFranco>

Nombre del Orientador

Yohany Pemberthy Salas

E-mail: pemberthy777@hotmail.com

Especialización en Analítica y Ciencia de Datos

Universidad de Antioquia

Facultad de Ingeniería

Departamento de Ingeniería de Sistemas

Medellín, Colombia

2023

1. INTRODUCCIÓN

En el mundo de las tiendas de abarrotes tradicionales, los pronósticos son una herramienta crucial para mantener el equilibrio entre la oferta y la demanda. Los dueños de estas tiendas tienen que bailar delicadamente con la cantidad de inventario que se debe comprar para evitar quedarse con productos perecederos con exceso de existencias o con una falta de los artículos populares que generan una pérdida de ingresos.

Ante este desafío, es importante contar con herramientas efectivas que permitan hacer pronósticos precisos. Una de las posibles soluciones es el uso de modelos de recomendación de productos y modelos de tiempo para predecir el consumo en los meses posteriores. Los modelos de recomendación de productos son una forma efectiva de determinar qué productos se venden juntos con más frecuencia, lo que permite hacer pronósticos más precisos sobre la demanda de los mismos. Estos modelos se basan en el análisis de datos de ventas y la identificación de patrones que permiten sugerir productos complementarios que pueden ser atractivos para los clientes.

Por otro lado, los modelos de serie de tiempo son una herramienta clave para prever la demanda futura de los productos. Estos modelos se basan en el análisis de datos históricos de ventas y en la identificación de tendencias y patrones estacionales que permiten hacer pronósticos precisos sobre el consumo en los meses posteriores. De esta forma, los tenderos pueden ajustar su inventario y asegurarse de tener suficiente stock para satisfacer la demanda en el momento adecuado.

En resumen, los pronósticos son cruciales para las tiendas de abarrotes tradicionales, y contar con herramientas efectivas como modelos de recomendación de productos y modelos de tiempo pueden ayudar a evitar la pérdida de ingresos y clientes molestos por falta o exceso de inventario. Los tenderos deben estar siempre al tanto de las últimas tendencias y tecnologías en el ámbito de los pronósticos para asegurarse de estar un paso adelante en este competitivo mercado.

2. DESCRIPCIÓN DEL DATASET

Los datos a utilizar provienen de una competencia de Kaggle que contiene información a nivel de tienda y producto con las ventas diarias realizadas en Ecuador (<https://www.kaggle.com/competitions/store-sales-time-series-forecasting/overview/description>). El problema contiene 7 archivos:

- **Vacaciones:** Contiene información acerca de los días festivos y/o eventos contiene 5 columnas fecha, tipo de día (festivo, adicional, puente, etc), local (Nacional, regional, etc), lugar, nombre local y si el día es transferido o no.
- **Oil:** Contiene información del precio del petroleo en Ecuador y contienen 2 columnas **Fecha** y **Precio** .
- **Stores:** Contiene información de las tiendas y contiene 5 columnas Ciudad, Estado, tipo y clúster (Es una etiqueta que permite identificar tiendas similares)
- **Train:** Contiene información para entrenar el modelo y posee 6 columnas:
 - **Id:** Contiene información única de compra
 - **store_nbr:** Identifica la tienda donde los productos son vendidos
 - **family:** Identifica el tipo de producto vendido
 - **Sales:** Venta total para el producto en un tienda en particular
 - **Onpromotion:** Contiene información del número de items que se encuentran en promoción en ciertas fechas
- **Test:** Contiene la misma información que la base de Train a excepción de las ventas totales

3. MÉTRICAS DE DESEMPEÑO

La métrica utilizada para la evaluación del desempeño del modelo es el **RMSLE** (Root Mean Squared Logarithmic Error) que es calculada como:

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(1 + \hat{y}_i) - \log(1 - y_i))^2} \quad (1)$$

donde

- n es el número total de instancias
- \hat{y}_i es el valor predicho del objetivo para la instancia i
- y_i es el valor real del objetivo para la instancia i

La métrica a utilizar debe generar valores menores al 5 %, en el contexto del problema, predecir una venta de artículos mayor al posible resultaría en costos para la empresa. Además, estos costos podrían no ser recuperables debido al tipo de mercancía que se esté vendiendo.

Referencias

- [1] Adhikari, R. and R. Agrawal (2013). An Introductory Study on Time Series Modeling and Forecasting. Lap Lambert Academic Publishing GmbH KG.
- [2] Aggawal, C. C. (2018). *Neural Network and Deep Learning: A Textbook* (2018 ed.). Springer
- [3] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
- [4] Pinedo Chapa, Joely Mireilli. Propuesta de un modelo de pronósticos de demanda y gestión de inventarios para la planeación de demanda en prendas de vestir juvenil. Edu.pe. Recuperado el 21 de abril de 2023, de https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/623528/Pinedo_CJ.pdf?sequence=5