

Sales Time Series Forecasting

Juan Felipe Vásquez Uribe

Jhon David Ballesteros Vargas

Natalia de Jesús Polo Peña

Introducción a la Inteligencia Artificial para ciencias e ingeniería.



Facultad de Ingeniería

Medellín

2023

Descripción del Problema

Se construirá un modelo que predice con mayor precisión las ventas unitarias de miles de artículos vendidos en diferentes tiendas Favorita. Se pretende pronosticar las ventas de los miles de familias de productos que se venden en las tiendas Favorita ubicadas en Ecuador.

Los pronósticos son especialmente relevantes para las tiendas de abarrotes tradicionales, que deben operar delicadamente con la cantidad de inventario que se debe comprar. Si la predicción se hace por un poco más, los tenderos se quedan con productos perecederos con exceso de existencias. Por el contrario, pronosticar un poco por debajo, y los artículos populares se agotan rápidamente, lo que genera una pérdida de ingresos y clientes molestos. Un pronóstico más preciso, gracias a técnicas de Machine Learning, podría ayudar a garantizar que los minoristas complazcan a los clientes al tener suficientes productos correctos en el momento correcto.

Data Set

El Data set que se usara para este modelo es de una competencia de Kaggle que proporciona datos desde el 31 de diciembre de 2012 hasta el 14 de agosto de 2017. En este data set se encuentra información de las ventas de miles de familias de productos que se venden en diferentes tiendas en Ecuador, además los datos de entrenamiento incluyen fechas e información del producto y la tienda, así como si el artículo tenía promoción y sus cifras de ventas, los demás archivos proveen información complementaria que será útil para el modelo.

Train.csv

Test.csv

Stores.csv

Oil.csv

holidays_events.csv

Métricas de desempeño

La métrica de desempeño usada en este modelo será el error logarítmico cuadrático medio (RMSLE) el cual se calcula mediante la formula:

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

Donde **n** es el número total de observaciones, **pi** es la predicción de la variable objetivo y **ai** es el valor real de la variable objetivo, **log** es el logaritmo natural.

Criterio de desempeño

Este modelo ayudará a predecir las ventas unitarias de miles de artículos vendidos en diferentes tiendas, por lo que debe tener una tasa de acierto de más del 70%, ya que, una previsión más precisa puede reducir el desperdicio de alimentos relacionado con el exceso de existencias y mejorar la satisfacción del cliente.

Bibliografía

- Store sales time series forecasting - <https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data?select=train.csv>