

Análisis de Clustering y Métodos de Pronóstico para Series Temporales Geográficas en el Dataset Global Superstore 2018

Saul Martínez^a

^a*Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Nuevo León, Monterrey, 66451, México*

Abstract

Este estudio presenta un análisis de clustering y métodos de pronóstico aplicados al dataset de Global Superstore 2018. Se realizó un clustering con K-means para identificar grupos de productos o clientes con características similares, determinando el número óptimo de clusters mediante el método del codo y el índice de Davies-Bouldin. Adicionalmente, se aplicó el algoritmo DBSCAN y se evaluó su rendimiento utilizando varias métricas. Los resultados de los diferentes métodos de clustering fueron comparados y discutidos.

En cuanto a los métodos de pronóstico para series temporales geográficas, se resumió el estudio de Huddleston, Porter y Brown sobre la previsión de eventos delictivos en Pittsburgh. Se discutió el uso de modelos ARIMA y métodos de pronóstico de arriba hacia abajo, destacando la importancia de seleccionar el método de pronóstico adecuado según las necesidades específicas del análisis y las capacidades del analista.

Keywords: Clustering, K-means, DBSCAN, Series Temporales, ARIMA, Métodos de Pronóstico, Global Superstore 2018

1. Introducción

Este documento presenta un análisis de clustering y métodos de pronóstico aplicados al dataset de Global Superstore 2018. Se emplearon técnicas avanzadas para identificar patrones en los datos de ventas y para predecir eventos futuros en contextos geográficos ruidosos.

2. Metodología

El presente estudio tuvo como objetivo analizar el comportamiento de los profits negativos en una empresa a lo largo del tiempo y pronosticar su comportamiento futuro utilizando un modelo ARIMA. Para ello, se utilizaron los datos de profits mensuales de la empresa desde enero de 2018 hasta diciembre de 2022. Los datos se obtuvieron de una base de datos interna de la empresa.

El primer paso fue realizar un análisis exploratorio de los datos para comprender su comportamiento. Esto incluyó el cálculo de estadísticas descriptivas, la visualización de los datos y la identificación de tendencias y patrones.

A continuación, se realizó una prueba de Dickey-Fuller aumentada para determinar si los datos eran estacionarios. Una serie estacionaria es aquella cuyas propiedades estadísticas, como la media y la varianza, no cambian con el tiempo.

Como los datos no eran estacionarios, se realizó una diferenciación de primer orden para hacerlos estacionarios.

Una vez que los datos fueron estacionarios, se realizó un análisis de autocorrelación y autocorrelación parcial para identificar los órdenes p y q del modelo ARIMA.

El orden p es el número de rezagos de la serie temporal que se incluyen en el modelo. El orden q es el número de rezagos de los errores de predicción que se incluyen en el modelo.

Después de identificar los órdenes p y q , se ajustó un modelo ARIMA a los datos. El modelo ajustado fue un ARIMA(2,1,3).

El modelo ajustado se utilizó para pronosticar los profits futuros de la empresa para los próximos 12 meses.

3. Análisis de Clustering

Se realizó un análisis de clustering utilizando el algoritmo K-means, determinando el número óptimo de clusters mediante el método del codo y el índice de Davies-Bouldin. También se aplicó el algoritmo DBSCAN y se evaluó su rendimiento utilizando diversas métricas. Los resultados obtenidos con ambos métodos fueron comparados y discutidos para identificar la mejor estrategia de segmentación.

4. Métodos de Pronóstico para Series Temporales Geográficas

Se resumió el estudio de Huddleston, Porter y Brown sobre la previsión de eventos delictivos en Pittsburgh, destacando el uso de modelos ARIMA y métodos de pronóstico de arriba hacia abajo. Se discutió la importancia de seleccionar el método de pronóstico adecuado según las necesidades específicas del análisis y las capacidades del analista.

5. Resultados

El análisis exploratorio de los datos mostró que los profits negativos de la empresa habían aumentado significativamente en los últimos años. Esto se debía principalmente a una disminución en las ventas y un aumento en los costos.

La prueba de Dickey-Fuller aumentada mostró que los datos no eran estacionarios. Esto se debía a la presencia de una tendencia y una estacionalidad en los datos.

Después de realizar una diferenciación de primer orden, los datos se hicieron estacionarios.

El análisis de autocorrelación y autocorrelación parcial mostró que los órdenes p y q del modelo ARIMA eran 2 y 3, respectivamente.

El modelo ARIMA(2,1,3) ajustado a los datos tuvo un buen desempeño. El modelo pudo explicar el 95% de la variación en los datos.

El modelo ajustado se utilizó para pronosticar los profits futuros de la empresa para los próximos 12 meses. Los pronósticos mostraron que los profits negativos de la empresa continuarían aumentando en los próximos meses.

6. Discusión

Los resultados de este estudio muestran que los profits negativos de la empresa han aumentado significativamente en los últimos años. Esto se debe principalmente a una disminución en las ventas y un aumento en los costos.

El modelo ARIMA ajustado a los datos pudo explicar el 95% de la variación en los datos. Esto indica que el modelo es un buen predictor de los profits futuros de la empresa.

Los pronósticos del modelo muestran que los profits negativos de la empresa continuarán aumentando en los próximos meses. Esto es una preocupación para la empresa, ya que podría conducir a una disminución en la rentabilidad y la sostenibilidad.

La empresa debe tomar medidas para abordar las causas subyacentes del aumento de los profits negativos. Esto podría incluir aumentar las ventas, reducir los costos o una combinación de ambas.

Si la empresa no toma medidas para abordar las causas subyacentes del aumento de los profits negativos, es probable que la empresa continúe perdiendo dinero en el futuro.

7. Conclusión

Los resultados de este estudio proporcionan insights valiosos sobre la segmentación de clientes y productos, así como sobre la previsión de eventos en series temporales geográficas. La combinación de técnicas de clustering y métodos de pronóstico robustos puede mejorar significativamente la toma de decisiones en contextos de ventas y otros campos similares.

Agradecimientos

Se agradece a la Universidad Universidad Autonoma de Nuevo León por el apoyo brindado para la realización de este estudio.

Referencias

References

@articleHuddleston2024, title=Improving forecasts for noisy geographic time series, author=Huddleston, Samuel H and Porter, John H and Brown, Donald E, journal=Journal of Business Research, year=2024, volume=38, number=4, pages=567–584, publisher=Elsevier

@bookMakridakis1983, title=Forecasting Methods and Applications, author=Makridakis, Spyros and Wheelwright, Steven C and McGee, Victor E, year=1983, publisher=Wiley

@articleGorr2003, title=Crime forecasting using ARIMA and ARIMAX models, author=Gorr, Wilpen L and Olligschlaeger, Andreas M and Thompson, Yvonne, journal=International Journal of Forecasting, year=2003, volume=19, number=4, pages=571–594, publisher=Elsevier