資料探勘專題作業二

訓練模型預測數值

指導教授:

許中川 教授

成員:

M11123047 劉穎謙

日期:

2023年1月5號

摘要

在這競爭激烈的消費市場中，如何掌握消費者的消費動機成為非常重要的課題。本研究使用交易資料集，藉由Apriori演算法及FP-Growth演算法分析出適合推薦給消費者的產品。

一、緒論

1.1動機

在這市場競爭非常熱烈的時代中，如何更有效且準確地預測消費者的喜好，成為了許多商家及電商所面臨的問題。本研究想有效地分析消費者的消費紀錄，來創造更多的商機。

1.2目的

為了創造商機，本研究最終希望利用消費紀錄來推薦相關可能購買的產品，找到潛在的商品組合來增加消費動機。

二、方法

圖1.實驗方法

三、實驗

3.1資料集

3.1 交易資料集 筆數:157396

|  |  |
| --- | --- |
| 屬性名稱 | 型態 |
| ITEM\_ID | Int |
| ITEM\_NO | String |
| PRODUCT\_TYPE | String |
| CUST\_ID | Int |
| TRX\_DATE | Date |
| INVOICE\_NO | String |
| QUANTITY | Int |

3.2前置處理

QUANTITY為零或負值的交易代表退貨或註銷，因此將其刪除。

3.3實驗設計

選擇屬性ITEM\_ID及INVOICE\_NO兩個欄位，利用INVOICE\_NO做資料分組。再使用Apriori與FP-Growth演算法分析，比較兩種演算法所需花費的時間，最後提供輸入產品代號，產生推薦產品的功能。

3.4實驗結果

Support = 0.001

|  |  |
| --- | --- |
| 演算法 | 花費時間 |
| Apriori | 17.139(秒) |
| FP-Growth | 1.325(秒) |

Support = 0.0007

|  |  |
| --- | --- |
| 演算法 | 花費時間 |
| Apriori | 60.367(秒) |
| FP-Growth | 1.296(秒) |

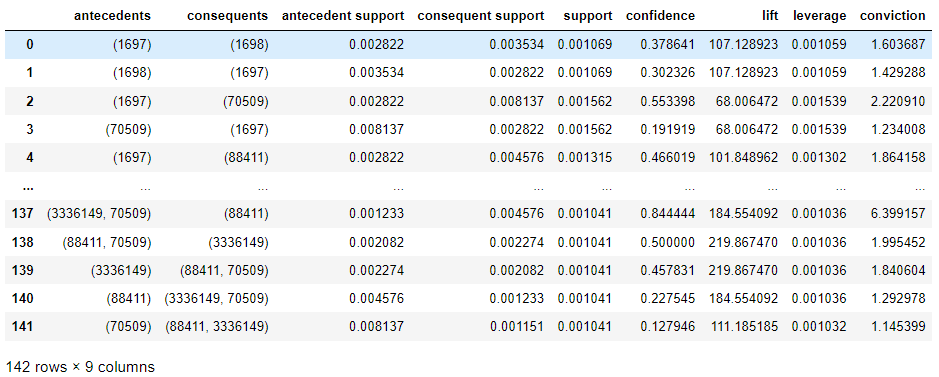


圖2.Apriori support=0.001之關聯結果

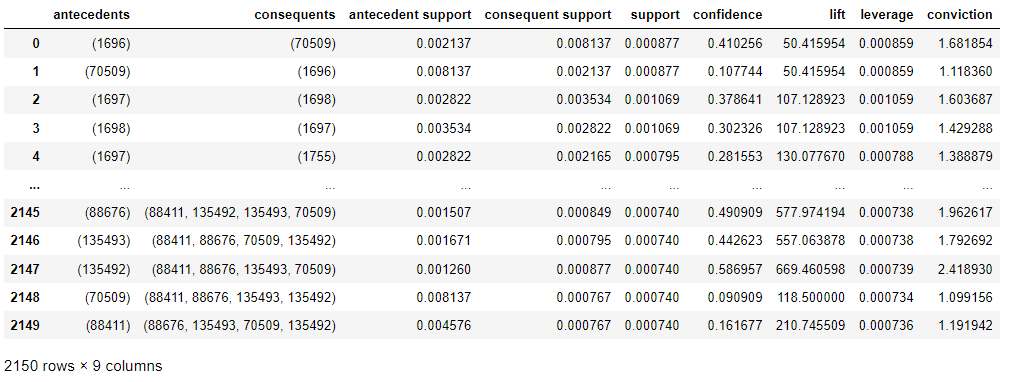
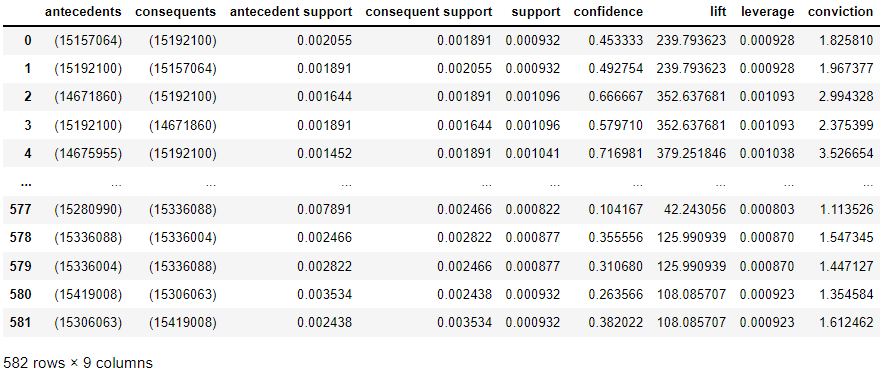
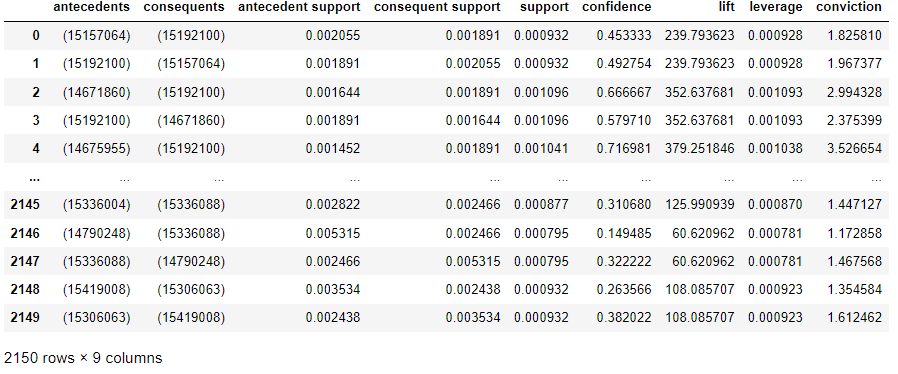


圖3.Apriori support=0.0007之關聯結果

圖4. FP-Growth support=0.001之關聯結果

圖5. FP-Growth support=0.0007之關聯結果

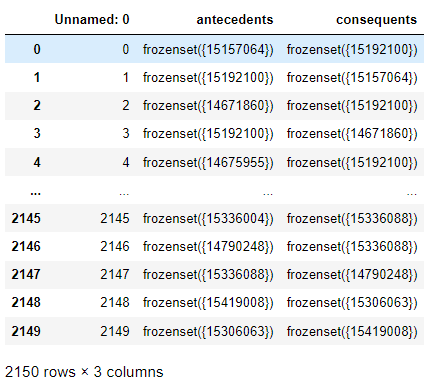


圖6.存檔&讀檔



圖7.推薦的商品代號

四、結論

當支持度設定0.001時，表示1000次購買紀錄中需出現過1次組合才可形成關聯規則，因此關聯規則的數量較少；支持度設定0.0007時為10,000次購買紀錄須出現過7次組合會形成關聯規則，因此支持度0.001與0.0007相比，支持度0.0007會出現較多的關聯規則。最終透過時間花費的計算結果可以發現FP-Growth演算法所需要的分析時間比Apriori演算法來的少。

五、參考文獻

Mike Ku (2020, November 1). [Pandas教學]資料分析必懂的Pandas DataFrame處理雙維度資料方法. LEARN CODE WITH MIKE.

<https://www.learncodewithmike.com/2020/11/python-pandas-dataframe-tutorial.html>

Chwang (October 15, 2020).Machine Learning -關聯分析-Apriori演算法-詳細解說啤酒與尿布的背後原理 Python實作-Scikit Learn一步一步教學

<https://chwang12341.medium.com/machine-learning-%E9%97%9C%E8%81%AF%E5%88%86%E6%9E%90-apriori%E6%BC%94%E7%AE%97%E6%B3%95-%E8%A9%B3%E7%B4%B0%E8%A7%A3%E8%AA%AA%E5%95%A4%E9%85%92%E8%88%87%E5%B0%BF%E5%B8%83%E7%9A%84%E8%83%8C%E5%BE%8C%E5%8E%9F%E7%90%86-python%E5%AF%A6%E4%BD%9C-scikit-learn%E4%B8%80%E6%AD%A5%E4%B8%80%E6%AD%A5%E6%95%99%E5%AD%B8-76b7778f8f34>

plusone (2017, December 29). [Day10]Pandas Groupby使用！

<https://ithelp.ithome.com.tw/articles/10194027>

Sebastian Raschka (2022, December 24). apriori: Frequent itemsets via the Apriori algorithm. mlxtend.

<http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/#apriori-frequent-itemsets-via-the-apriori-algorithm>