

# **Blog Chủ đề 2: Product Hub – Tìm sản phẩm trung tâm để tối ưu chiến lược cross-selling trong siêu thị**

---

**Nhóm tác giả : Nhóm 3**

**Lớp : CNTT17-13**

**Thành viên : Nguyễn Văn Vinh**

**Đỗ Văn Vinh**

**Bạch Ngọc Lương**

**Lại Thành Đoàn**

**Ngày : 13/12/2025**

---

## **1. THUẬT TOÁN APRIORI**

### **1.1. Khái niệm**

Thuật toán Apriori được công bố bởi R. Agrawal và R. Srikant vào năm 1994 vì để tìm các tập phổ biến trong một bộ dữ liệu lớn. Tên của thuật toán là Apriori vì nó sử dụng kiến thức đã có từ trước (prior) về các thuộc tính, vật phẩm thường xuyên xuất hiện trong cơ sở dữ liệu. Để cải thiện hiệu quả của việc lọc các mục thường xuyên theo cấp độ, một thuộc tính quan trọng được sử dụng gọi là thuộc tính Apriori giúp giảm phạm vi tìm kiếm của thuật toán.

Thuật toán sử dụng hai chiến lược chính:

- Xây dựng từ dưới lên (bottom-up)

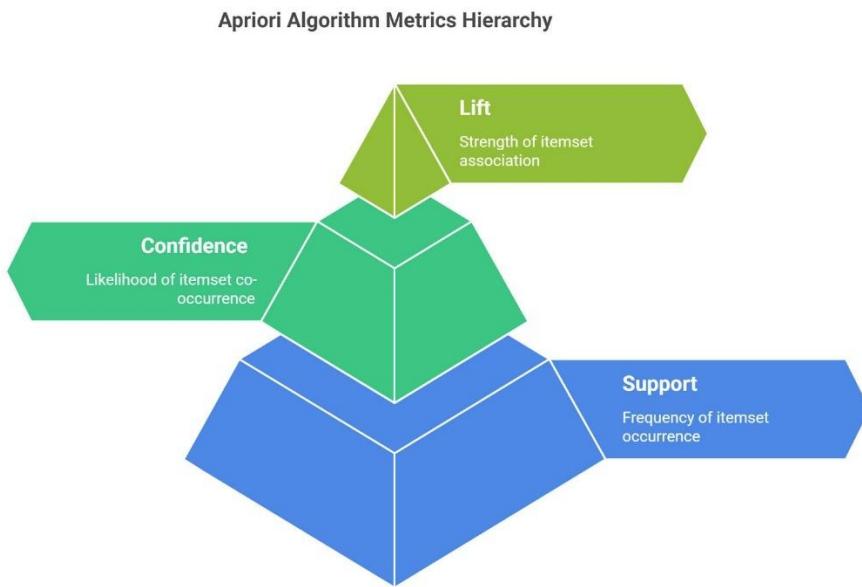
Apriori không xét tất cả tổ hợp sản phẩm ngay từ đầu (vì sẽ rất tốn tài nguyên), mà xây dựng dần theo từng cấp độ:

1. Bắt đầu với **1 sản phẩm** (1-itemset)

2. Chỉ giữ lại các sản phẩm có **support** đủ lớn
3. Kết hợp chúng để tạo **2-itemset**
4. Tiếp tục mở rộng lên 3-itemset, 4-itemset...

Nguyên lý quan trọng:

=> **Nếu một tập sản phẩm không phỗ biến, thì mọi tập con mở rộng từ nó cũng không phỗ biến.**



- Tìm kiếm theo chiều rộng (breadth-first search)

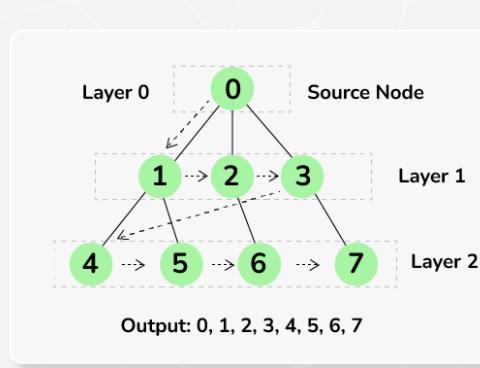
Apriori sử dụng tìm kiếm theo chiều rộng (BFS) để duyệt các tập sản phẩm:

- Xét tất cả các itemsets cùng kích thước trước
- Sau đó mới chuyển sang mức tiếp theo

Cách này giúp:

- So sánh support dễ dàng
- Cắt tỉa nhanh các nhánh không tiềm năng

# Breadth First Search



Ý tưởng:

- Từ các tập phô biến có kích thước  $k - 1$ , thuật toán sẽ sinh ra các tập ứng viên có kích thước  $k$ .
- Các tập ứng viên này là những tập có khả năng trở thành tập phô biến nếu thoả mãn ngưỡng độ hộ trợ tối thiểu. Sau đó, thuật toán kiểm tra từng tập ứng viên để xác định tập nào thực sự là tập phô biến kích thước  $k$ .
- Quá trình lặp lại cho đến khi không thể tìm thêm bất kỳ tập phô biến nào nữa.

## 1.2. Mục tiêu

Mục tiêu của Apriori là:

Tìm ra các mối quan hệ mua kèm giữa các sản phẩm dựa trên dữ liệu giao dịch trong quá khứ.

Nói một cách dễ hiểu, Apriori giúp trả lời câu hỏi:

*“Khách hàng thường mua những sản phẩm nào cùng nhau?”*

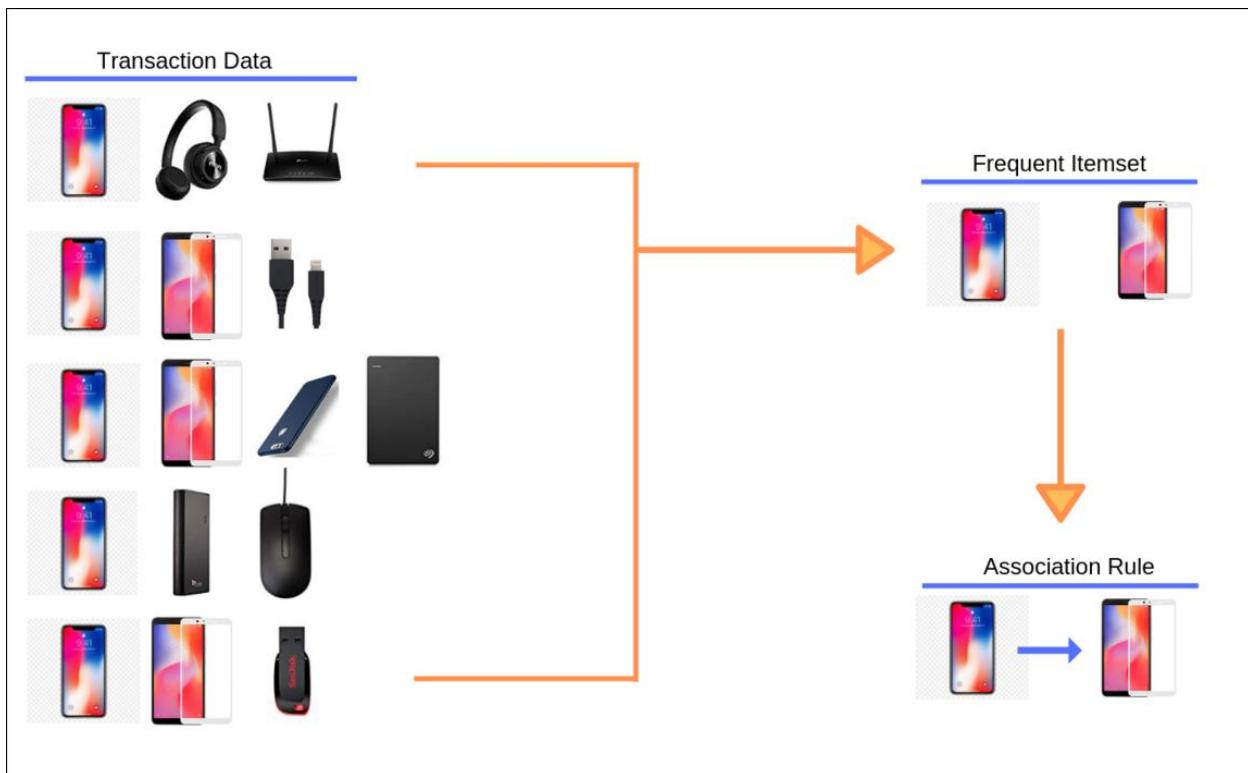
Thuật toán này tìm ra các **luật kết hợp (Association Rules)** dựa trên dữ liệu giao dịch mua sắm, từ đó giúp doanh nghiệp:

- Hiểu hành vi mua hàng

- Tối ưu bán chéo (cross-selling)
- Sắp xếp hàng hóa hiệu quả hơn

Tìm ra các **mẫu mua sắm thường xuyên** và các **luật kết hợp có ý nghĩa** từ dữ liệu giao dịch lớn.

Ví dụ:



Nếu nhiều khách mua điện thoại kèm với kính cường lực  
→ ta có luật kết hợp: điện thoại kèm với kính cường lực

Điều này có nghĩa khi khách hàng mua điện thoại thì khả năng cao họ sẽ mua thêm kính cường lực đây cũng chính là nền tảng cho **cross-selling** và **Product Hub**.

### 1.3. Bài toán mà Apriori giải quyết

Trong thực tế bán lẻ:

- Một siêu thị có hàng nghìn sản phẩm

- Có hàng trăm nghìn hóa đơn
- Không thể kiểm tra thủ công mọi tổ hợp sản phẩm

Thì Apriori được sinh ra để giải quyết vấn đề:

- Tìm các nhóm sản phẩm thường xuyên xuất hiện cùng nhau
- Loại bỏ các tổ hợp không có ý nghĩa
- Giảm đáng kể chi phí tính toán

#### **1.4. Các thành phần cốt lõi của thuật toán Apriori**

Thuật toán Apriori xoay quanh 3 chỉ số quan trọng, là:

##### 1. Support (Độ hỗ trợ)

- Đo tần suất xuất hiện của một sản phẩm hoặc một tập sản phẩm
- Công thức (diễn giải):

$$\text{Support} = \frac{\text{Số giao dịch chứa sản phẩm}}{\text{Tổng số giao dịch}}$$

Ý nghĩa:

- Cho biết sản phẩm có phổ biến hay không
- Các sản phẩm quá ít xuất hiện sẽ bị loại sớm

##### 2. Confidence (Độ tin cậy)

- Đo xác suất mua B khi đã mua A
- Luật dạng:  $A \rightarrow B$

Ý nghĩa:

- Confidence cao  $\rightarrow$  khả năng mua kèm cao
- Dùng để đánh giá độ mạnh của luật

### 3. Lift (Độ nâng)

- Đo mức độ mua kèm vượt trội so với ngẫu nhiên

Cách hiểu đơn giản:

- Lift = 1 → mua kèm ngẫu nhiên
- Lift > 1 → có liên quan tích cực
- Lift < 1 → liên quan tiêu cực

Trong bài toán Product Hub, Lift là chỉ số quan trọng nhất để xác định mối quan hệ thực sự có ý nghĩa.

Đánh giá ảnh hưởng của các độ đo

Thí nghiệm	Tham số thay đổi	Số luật sinh ra	Tỷ lệ giảm	Hiện tượng quan sát được	Ý nghĩa quản trị
Gốc (Baseline)	(Mặc định)	1,794	-	Tìm thấy cả sản phẩm đại trà ( <i>Jumbo Bag</i> ) và hàng ngách ( <i>Herb Marker</i> ).	Bức tranh toàn cảnh về hành vi mua sắm.
TN 1: Tăng Support	min_support = 0.03	21	~99%	Các nhóm hàng ngách biến mất hoàn toàn. Chỉ còn lại hàng bán chạy nhất (Best-sellers).	Lọc tìm Sản phẩm chủ lực (Key Drivers) để đảm bảo doanh thu nền tảng.
TN 2: Tăng Confidence	min_conf = 0.6	599	~66%	Loại bỏ các luật "lò mò". Giữ lại các cặp có độ kết dính cao ( <i>Herb Marker</i>	Lọc tìm Combo chắc thắng. Chỉ đóng gói chung những

				quay lại do Conf ~95%).	món khách chắc chắn sẽ mua kèm.
<b>TN 3: Tăng Lift</b>	<b>min_lift = 3.0</b>	<b>1,773</b>	<b>~1%</b>	Số lượng luật gần như không đổi. Hầu hết luật tìm được đều có mối quan hệ rất mạnh.	Xác thực Chất lượng mối quan hệ. Chứng tỏ khách hàng mua sắm theo nhiều nhu cầu thực sự, không phải ngẫu nhiên.

#### Nhận xét : **Kết luận: Ảnh hưởng của từng độ đo đến bài toán**

Từ các thí nghiệm trên, nhóm rút ra kết luận về vai trò của từng chỉ số trong bài toán phân tích giỏ hàng:

#### **1. Support (Độ hỗ trợ) - "Bộ lọc độ phổ biến"**

- Tác động:** Là tham số nhạy cảm nhất. Chỉ cần tăng nhẹ Support, số lượng luật sẽ giảm theo cấp số nhân.
- Vai trò:** Dùng để loại bỏ nhiễu (noise) từ các giao dịch hiếm gặp.
- Bài học:**
  - Đặt Support cao:** Để tìm sản phẩm xu hướng, bán chạy (Trending/Best-seller).
  - Đặt Support thấp:** Để khai thác thị trường ngách (Long-tail) - nơi thường chứa các cơ hội lợi nhuận cao nhưng ít người biết (như ví dụ *Herb Marker*).

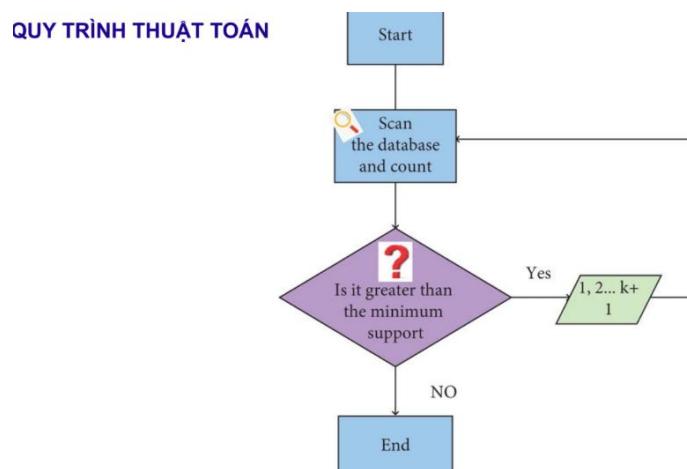
#### **2. Confidence (Độ tin cậy) - "Độ chắc chắn của dự báo"**

- **Tác động:** Tác động vừa phải đến số lượng luật. Confidence càng cao, độ chính xác của luật càng lớn.
- **Vai trò:** Đánh giá khả năng thành công của lời gợi ý.
- **Bài học:** Dùng Confidence để quyết định hình thức khuyến mãi:
  - **Confidence thấp (30-50%):** Chỉ nên gợi ý nhẹ ("Khách hàng thường xem thêm...").
  - **Confidence cao (>60%):** Mạnh dạn đóng gói Combo cứng (Bundle) hoặc giảm giá khi mua kèm.

### 3. Lift (Độ nâng) - "Thước đo chất lượng mối quan hệ"

- **Tác động:** Dùng để kiểm chứng giá trị thực của luật, loại bỏ sự ngẫu nhiên.
- **Vai trò:** Trả lời câu hỏi "Hai món này đi chung với nhau là do chúng hợp nhau, hay chỉ do chúng đều là hàng bán chạy?".
- **Bài học:** Trong dự án này, Lift đóng vai trò then chốt để phát hiện ra các nhóm sản phẩm "**trời sinh một cặp**".  
 ◦ Luật có **Lift cao (như Herb Marker ~74)** là mục tiêu vàng cho chiến lược Bán chéo (Cross-selling), bát ké Support của nó thấp hay cao.

#### 1.5. Quy trình hoạt động



## **1. Start**

## **2. Scan the database and count**

- Quét toàn bộ dữ liệu
- Đếm số lần xuất hiện của từng itemset

## **3. So sánh với min\_support**

- Nếu  $\geq \text{min\_support}$  → giữ lại
- Nếu  $< \text{min\_support}$  → loại bỏ

## **4. Tăng kích thước tập ( $k \rightarrow k+1$ )**

- Ghép các itemset hợp lệ
- Quay lại bước scan

## **5. Kết thúc**

- Khi không còn itemset nào thỏa mãn

## **2. Thuật toán Apriori với chủ đề “Product Hub (Tìm kiếm sản phẩm trọng tâm)”**

### **2.1. Giới thiệu chủ đề**

Trong môi trường bán lẻ hiện đại, việc tăng doanh thu không chỉ dựa vào một sản phẩm đơn lẻ mà dựa vào mối quan hệ mua kèm giữa các sản phẩm. Các sản phẩm được mua cùng nhau thường xuyên đóng vai trò là “cầu nối” trong hành vi tiêu dùng — gọi là Product Hub. Việc xác định các sản phẩm trung tâm này không chỉ giúp tối ưu cross-selling (bán chéo) mà còn ảnh hưởng trực tiếp đến bố cục sắp xếp hàng hóa trong siêu thị.

Trong bài viết này, chúng ta sẽ:

1. Xác định sản phẩm xuất hiện nhiều nhất trong các luật kết hợp.
2. Đánh giá vai trò của Product Hub trong chiến lược kinh doanh.
3. Đề xuất cách bố trí hàng hóa tối ưu trong siêu thị theo phát hiện này.

## 2.2. Tại sao cần tìm Product Hub?

Khi khách hàng vào siêu thị, họ thường mua không chỉ một mặt hàng mà **nhiều mặt hàng liên quan**. Ví dụ:

- Mua cà phê → dễ mua bánh quy, sữa
- Mua giấy in → dễ mua mực in, sổ tay

Những sản phẩm lặp lại nhiều lần trong các giao dịch mua kèm là điểm kết nối hành vi người dùng, và nếu biết cách khai thác, các sản phẩm này sẽ trở thành trung tâm cho chiến lược bán hàng hiệu quả.

## 2.3. Product Hub là gì?

Product Hub không chỉ đơn thuần là *sản phẩm bán chạy nhất* — mà là sản phẩm xuất hiện nhiều nhất trong các luật kết hợp mạnh mẽ được sinh ra từ thuật toán Apriori.

Một sản phẩm được coi là Product Hub nếu:

- Xuất hiện thường xuyên trong các *association rules*, ở cả vế antecedent (về trái → sản phẩm dẫn tới sản phẩm khác), và vế consequent (về phải → sản phẩm được gợi ý mua tiếp).
- Các luật đó đều có confidence và lift cao, nghĩa là mối quan hệ mua kèm là đáng tin cậy và có ý nghĩa hơn so với mua ngẫu nhiên.

## 2.4. Mục tiêu

Mục tiêu của nhóm là áp dụng luật kết hợp Apriori để phân tích hành vi mua sắm của khách hàng, từ đó:

- Xác định các sản phẩm trung tâm (Product Hub) – những sản phẩm xuất hiện thường xuyên trong nhiều luật kết hợp.
- Đánh giá vai trò của các sản phẩm này trong chiến lược bán chéo (Cross-selling).

- Để xuất cách bố trí hàng hóa trong siêu thị nhằm tối ưu doanh thu và trải nghiệm khách hàng.

## 2.5. Ý tưởng

Apriori dùng để làm gì?

Apriori được dùng để tìm ra các sản phẩm thường được mua cùng nhau dựa trên lịch sử giao dịch. Thay vì nhìn từng hóa đơn riêng lẻ, Apriori giúp phát hiện quy luật mua sắm lặp lại của khách hàng.

Tại sao phù hợp cho bài toán gio hàng?

Dữ liệu gio hàng có đặc trưng:

- Mô hóa đơn chứa nhiều sản phẩm
- Số lượng sản phẩm rất lớn
- Hành vi mua không ngẫu nhiên

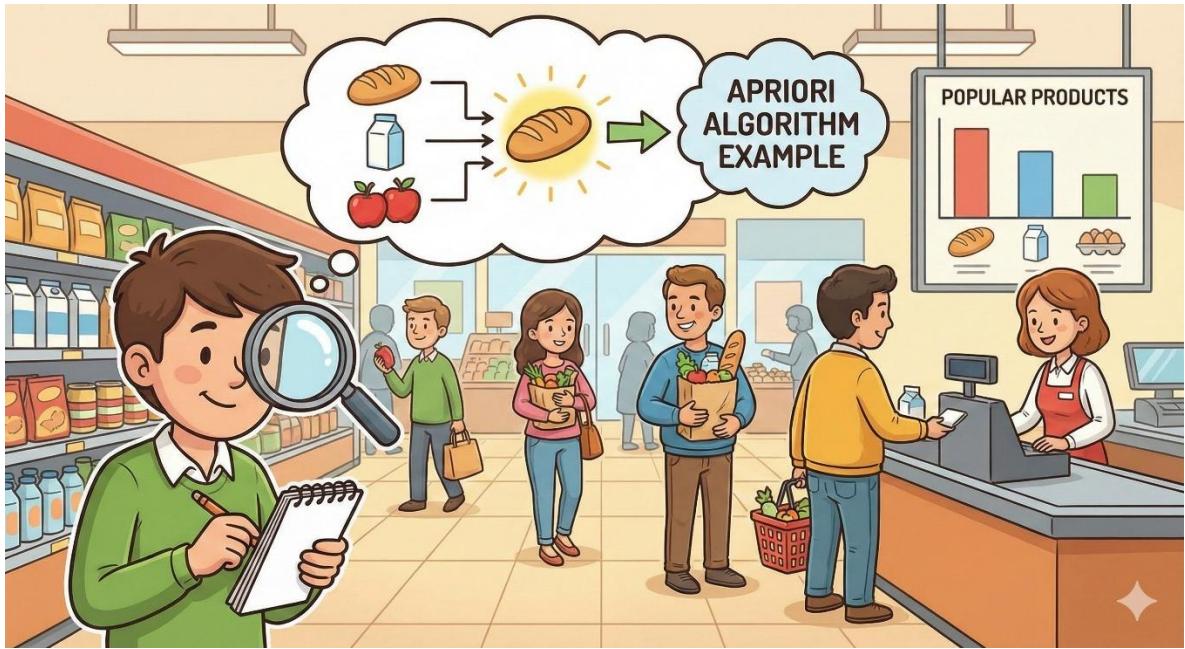
Apriori đặc biệt phù hợp vì nó loại bỏ sớm các tổ hợp ít xuất hiện, giúp giảm đáng kể chi phí tính toán.

Ý tưởng :

Hãy tưởng tượng Apriori giống như một người quan sát tinh ý trong siêu thị. Người này không quan tâm khách hàng mua gì một lần, mà quan sát hàng trăm nghìn hóa đơn để phát hiện ra quy luật:

“Nếu khách hàng mua sản phẩm A, rất có khả năng họ sẽ mua thêm sản phẩm B.”

Thay vì thử mọi khả năng kết hợp (rất tốn tài nguyên), Apriori loại bỏ sớm các sản phẩm ít được mua. Nếu một sản phẩm hiếm khi xuất hiện, thì mọi tập hợp chứa nó cũng không đáng để phân tích tiếp.



## 2.6. Quy trình Thực hiện

Pipeline của nhóm được xây dựng theo hướng chuẩn hóa và tự động hóa, gồm các bước:

1. Load & làm sạch dữ liệu
2. Tạo ma trận basket (One-hot / Boolean)
3. Áp dụng Apriori để tìm frequent itemsets
4. Trích xuất luật kết hợp (association rules)
5. Trực quan hóa (đặc biệt là Network Graph)
6. Phân tích insight & đề xuất kinh doanh

Toàn bộ pipeline được tự động chạy bằng Papermill, đảm bảo tính tái lập và nhất quán.

## 2.7. Tiền xử lý dữ liệu

### Các bước làm sạch dữ liệu

Dữ liệu bán lẻ thực tế chứa nhiều nhiễu, nhóm đã thực hiện:

- Loại bỏ các dòng thiếu mô tả sản phẩm (Description rỗng)

- Loại bỏ các giao dịch bị hủy  
→ InvoiceNo bắt đầu bằng ký tự "C"
- Loại bỏ các giao dịch có:
  - Quantity  $\leq 0$
  - UnitPrice  $\leq 0$
- Chỉ giữ dữ liệu tại United Kingdom để đồng nhất hành vi mua sắm

Thống kê nhanh sau khi lọc

- Số giao dịch sau lọc: 485,123 giao dịch
- Số sản phẩm duy nhất: 4,007 sản phẩm

Việc làm sạch giúp đảm bảo các luật kết hợp phản ánh hành vi mua sắm thực tế của khách hàng.

## 2.8. Áp dụng Apriori

### Chuẩn bị dữ liệu

Dữ liệu được chuyển sang **ma trận giỏ hàng dạng Boolean**:

- Dòng: hóa đơn (InvoiceNo)
- Cột: sản phẩm (Description)
- Giá trị:
  - True nếu sản phẩm xuất hiện  $\geq 1$  lần trong hóa đơn
  - False nếu không xuất hiện

### Tham số sử dụng (đúng theo code nhóm)

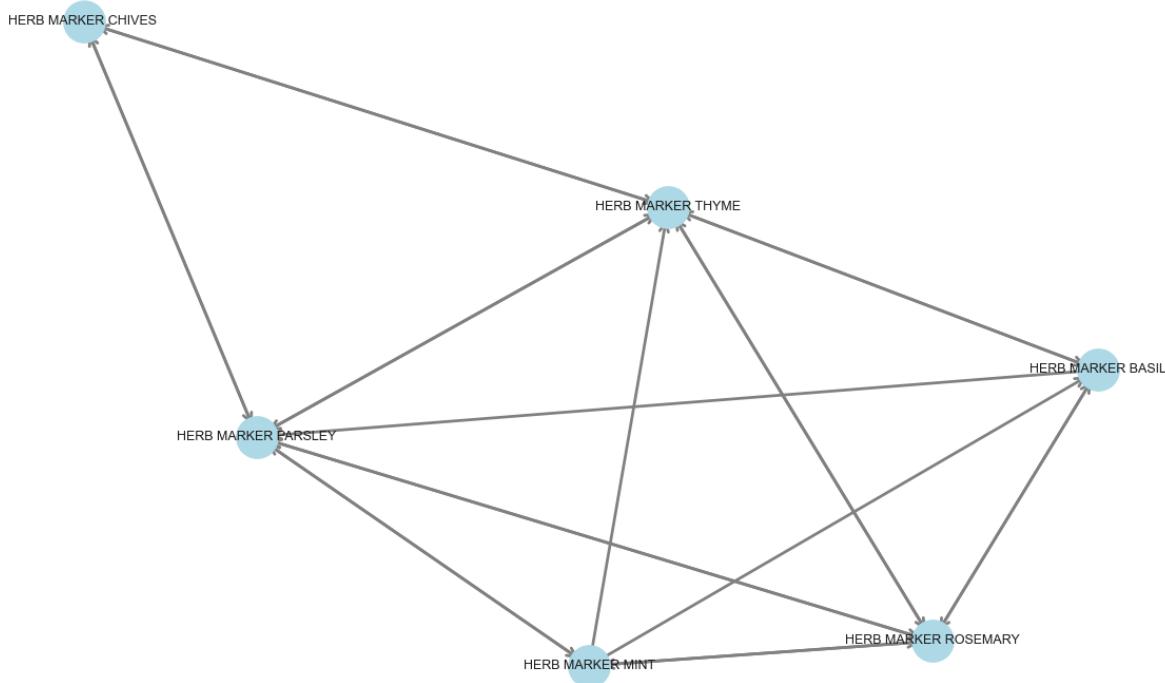
- min\_support = 0.01
- metric = "lift"

- min\_threshold = 1.0
  - **Bộ lọc luật sau khi sinh:**
    - FILTER\_MIN\_SUPPORT = 0.01
    - FILTER\_MIN\_CONF = 0.3
    - FILTER\_MIN\_LIFT = 1.2
  - **Số luật phân tích:** TOP\_N\_RULES = 50
- ...

```
from mlxtend.frequent_patterns import apriori, association_rules
frequent_itemsets = apriori(
    basket_bool,
    min_support=0.01,
    use_colnames=True
)
rules = association_rules(
    frequent_itemsets,
    metric="lift",
    min_threshold=1.0
)
rules = rules.sort_values(["lift", "confidence"], ascending=False)
...
```

## 2.9.Trực quan hóa (Visualization)

Mạng lưới các luật kết hợp (Arrow: antecedent → consequent)



Hình 1 – Network Graph các luật kết hợp

Mỗi nút (node) đại diện cho một sản phẩm

Mỗi mũi tên (edge) biểu diễn một luật kết hợp có hướng  
(từ *antecedent* → *consequent*)

Độ dày của cạnh tỉ lệ với chỉ số Lift, thể hiện mức độ liên kết mạnh giữa hai sản phẩm

### 1. Ý nghĩa cấu trúc mạng

Trong hình có thể quan sát thấy một cụm sản phẩm thảo mộc (HERB MARKER) gồm:

- HERB MARKER THYME
- HERB MARKER BASIL
- HERB MARKER PARSLEY

- HERB MARKER MINT
- HERB MARKER ROSEMARY
- HERB MARKER CHIVES

Các sản phẩm này được liên kết chặt chẽ với nhau bằng nhiều mũi tên hai chiều, cho thấy chúng thường xuyên được mua cùng nhau trong các giao dịch.

## 2. Nhận diện Product Hub

Trong mạng lưới, HERB MARKER THYME và HERB MARKER PARSLEY nổi bật vì:

- Có nhiều mũi tên đi ra và đi vào
- Xuất hiện trong nhiều luật kết hợp mạnh
- Đóng vai trò trung tâm kết nối (Product Hub) giữa các sản phẩm khác

Điều này cho thấy:

Khi khách hàng mua một trong hai sản phẩm này, xác suất họ mua thêm các loại thảo mộc khác là rất cao.

## 3. Ý nghĩa chỉ số Lift trong hình

Các cạnh trong mạng đều tương đối đậm, phản ánh:

- Lift > 1.2 (đã được lọc trước khi vẽ)
- Mỗi quan hệ mua kèm không phải ngẫu nhiên
- Việc mua sản phẩm A làm tăng đáng kể xác suất mua sản phẩm

## 4. Ý nghĩa kinh doanh

Từ Hình 1 có thể rút ra các kết luận thực tiễn:

- Nhóm sản phẩm thảo mộc nên được bố trí gần nhau trên cùng một khu vực kệ

- Có thể chọn HERB MARKER THYME làm sản phẩm mồi (anchor product) cho các chiến dịch:
  - Cross-selling
  - Combo mua kèm
  - Giảm giá nhóm sản phẩm

Kết luận : Hình 1 cho thấy rõ một cụm Product Hub với các mối liên kết mạnh giữa các sản phẩm thảo mộc. Việc trực quan hóa luật kết hợp dưới dạng mạng lưới giúp dễ dàng nhận diện các sản phẩm trung tâm, từ đó hỗ trợ hiệu quả cho chiến lược cross-selling và bố trí hàng hóa trong siêu thị.

### **Hình 2 : Top 50 luật kết hợp theo Lift**

1. Hình minh họa 50 luật kết hợp mạnh nhất được sinh ra từ thuật toán Apriori, được sắp xếp giảm dần theo chỉ số Lift.

- Trục dọc (Y-axis):  
Các luật kết hợp có dạng:
- Sản phẩm A, Sản phẩm B → Sản phẩm C

Trong đó:

- Vết trái (antecedent): các sản phẩm mua trước
- Vết phải (consequent): sản phẩm thường được mua kèm
- Trục ngang (X-axis):  
Giá trị Lift của từng luật.

2. Lift là gì và vì sao quan trọng?

Lift đo lường mức độ liên kết thực sự giữa các sản phẩm:

- Lift = 1  
→ mua kèm ngẫu nhiên, không có ý nghĩa
- Lift > 1  
→ mua kèm có liên quan
- Lift càng cao  
→ mối quan hệ càng mạnh và đáng tin

Trong hình, nhiều luật có Lift rất cao ( $\approx 60-70$ ) → cho thấy đây là nhóm sản phẩm mua cùng nhau gần như “mặc định”.

### 3. Insight quan trọng từ biểu đồ

Quan sát biểu đồ cho thấy:

#### 3.1 Một nhóm sản phẩm nổi bật rõ ràng

Hầu hết các luật đều xoay quanh nhóm:

- HERB MARKER PARSLEY
- HERB MARKER THYME
- HERB MARKER ROSEMARY
- HERB MARKER BASIL
- HERB MARKER MINT
- HERB MARKER CHIVES

Đây là một cụm sản phẩm (product cluster) rất chật chẽ.

#### 3.2 Sản phẩm xuất hiện lặp lại nhiều lần

Một số sản phẩm xuất hiện liên tục trong cả:

- Antecedent

- Consequent

Đặc biệt là:

- HERB MARKER THYME
- HERB MARKER PARSLEY
- HERB MARKER ROSEMARY

=> Những sản phẩm này đóng vai trò trung tâm (Product Hub).

#### 4. Liên hệ với Product Hub

Từ biểu đồ này, ta có thể kết luận:

Product Hub không phải là sản phẩm bán chạy nhất, mà là sản phẩm xuất hiện nhiều nhất trong các luật có Lift cao.

Trong trường hợp này:

- Các sản phẩm thuộc nhóm HERB MARKER là hub
- Chúng đóng vai trò:
  - Kết nối các sản phẩm khác
  - Dẫn dắt hành vi mua kèm

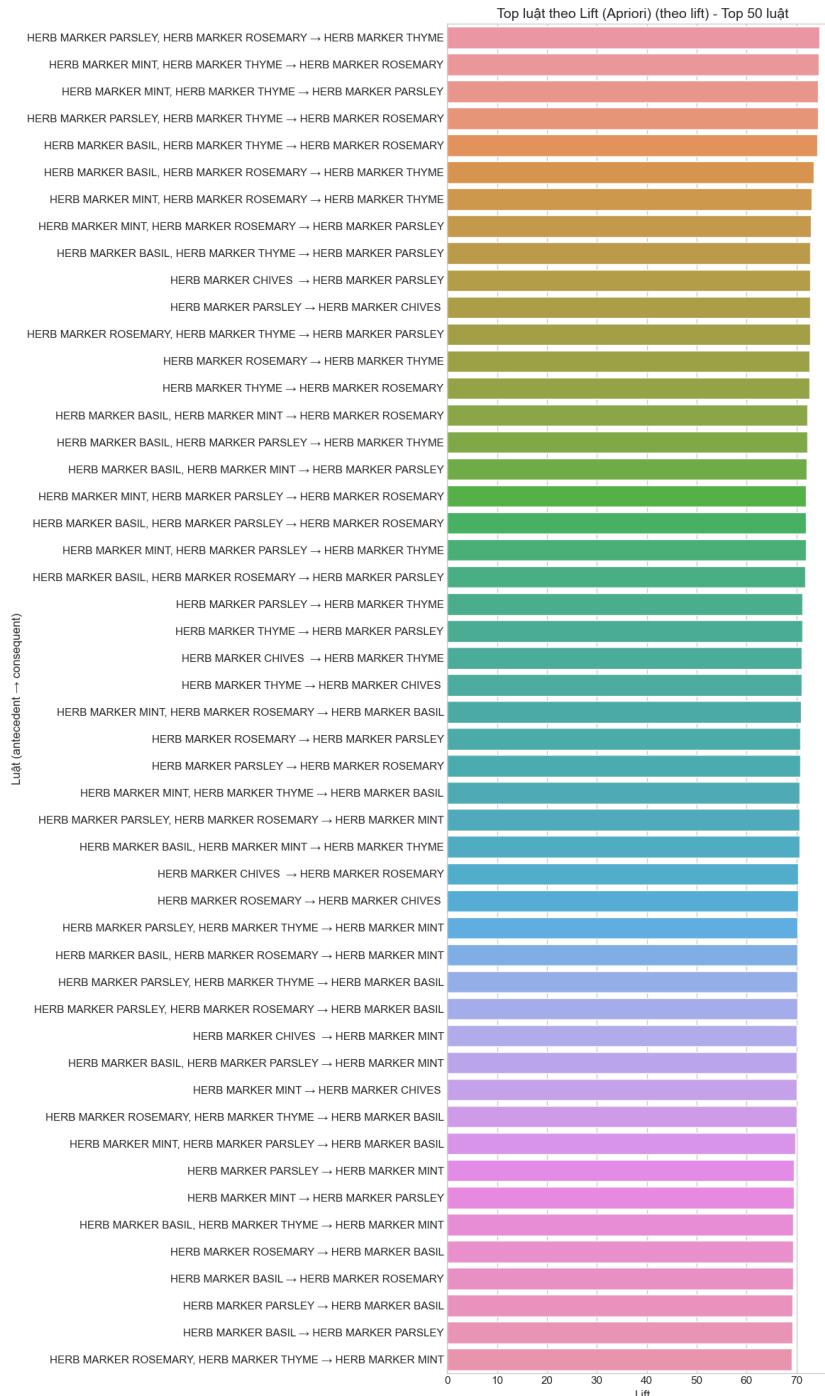
#### 5. Ý nghĩa kinh doanh

Từ biểu đồ này, doanh nghiệp có thể:

- Tạo combo sản phẩm  
Ví dụ: Parsley + Thyme + Rosemary
- Bố trí kệ hàng theo cụm  
Các sản phẩm HERB đặt gần nhau để kích thích mua thêm

- Cross-selling thông minh

Khi khách mua 1 loại herb → gợi ý các loại còn lại



Hình 2 : Top 50 luật theo Lift

Hình 2 thể hiện top 50 luật kết hợp có giá trị Lift cao nhất. Kết quả cho thấy một nhóm sản phẩm HERB MARKER xuất hiện dày đặc trong các luật mạnh, đặc biệt là Parsley, Thyme và Rosemary. Việc lặp lại liên tục của các sản phẩm này trong cả về trái và về phải cho thấy chúng đóng vai trò Product Hub, có khả năng dẫn dắt hành vi mua kèm và rất phù hợp cho chiến lược cross-selling.

## 2.10. Insight

**Insight #1:** Nhóm sản phẩm “HERB MARKER” tạo thành một cộng đồng mua sắm rất chật chẽ

Quan sát network graph, ta thấy các sản phẩm trong nhóm HERB MARKER (Parsley, Thyme, Rosemary, Basil, Mint, Chives) tạo thành một cụm liên kết dày đặc, với nhiều cạnh hai chiều giữa các nút. Điều này cho thấy các sản phẩm này không được mua riêng lẻ, mà thường xuyên xuất hiện cùng nhau trong cùng một giỏ hàng.

Insight này khẳng định đây là một **product family** rõ ràng, phù hợp để phân tích theo hướng *Product Hub* thay vì từng sản phẩm đơn lẻ.

Nếu là cửa hàng trưởng tôi sẽ không bán bán riêng lẻ từng sản phẩm mà sẽ xây dựng 1 hệ sinh thái gồm 6 sản phẩm

### **Hành động cụ thể:**

- Trong quản lý kệ:
  - Không tách Parsley, Thyme, Basil... ra các khu khác nhau
  - Gộp chúng thành 1 khu “Herb & Spices for Cooking”
- Trong báo cáo:
  - Theo dõi doanh thu theo nhóm herb
- Trong khuyến mãi:
  - Khuyến mãi theo combo

## **Insight #2: HERB MARKER THYME và HERB MARKER PARSLEY đóng vai trò Product Hub trung tâm**

Trong network graph, các nút HERB MARKER THYME và HERB MARKER PARSLEY có:

- Số lượng cạnh kết nối lớn
- Vừa đóng vai trò antecedent (dẫn tới sản phẩm khác)
- Vừa xuất hiện thường xuyên ở consequent

Điều này cho thấy hai sản phẩm này đóng vai trò hub trung tâm, là “điểm giao” trong hành vi mua sắm. Khi khách hàng mua Thyme hoặc Parsley, xác suất họ mua thêm các loại herb khác tăng lên đáng kể.

Đây chính là Product Hub đúng nghĩa, không chỉ bán chạy mà còn “kéo” các sản phẩm khác đi kèm.

Nếu là cửa hàng trưởng tôi chọn THYME và PARSLEY làm “mồi chiến lược” để chèo kéo khách hàng đặt HERB MARKER THYME & PARSLEY là Product Hub trung tâm xắp xếp ở nơi dễ thấy nhất.

## **Insight #3: Giá trị Lift rất cao chứng minh hành vi mua kèm là có chủ đích, không ngẫu nhiên**

Biểu đồ Top luật theo Lift cho thấy nhiều luật có Lift rất cao (lên tới ~70), đặc biệt giữa các cặp sản phẩm herb. Lift cao đồng nghĩa với việc:

Khách mua sản phẩm A có khả năng mua B cao hơn rất nhiều lần so với mua ngẫu nhiên.

Ví dụ các luật như:

- Parsley, Rosemary → Thyme
- Mint, Thyme → Rosemary

đều nằm trong nhóm top lift, chứng minh mối liên hệ mua kèm rất mạnh và ổn định.

Điều này xác nhận rằng các luật phát hiện được có giá trị kinh doanh thực tế, không phải nhiều dữ liệu.

Nếu là cửa hàng trưởng tôi sẽ thiết kế các combo và cross selling vì qua quan sát có thể thấy lift rất cao nên việc mua thường là có chủ đích ví dụ như khi khách chọn Thyme → gợi ý ngay Rosemary

**Insight #4:** Mỗi quan hệ mua sắm mang tính “hai chiều”, không phải một chiều

Quan sát cả network graph và danh sách luật lift cao cho thấy:

- Nhiều cặp sản phẩm có luật  $A \rightarrow B$  và  $B \rightarrow A$
- Ví dụ: Thyme  $\leftrightarrow$  Rosemary, Parsley  $\leftrightarrow$  Basil

Điều này cho thấy không tồn tại một sản phẩm “phụ” rõ ràng, mà các sản phẩm herb được mua theo tập hợp thói quen nấu ăn, nơi khách hàng linh hoạt lựa chọn.

Insight này rất quan trọng khi thiết kế combo hoặc bố trí kệ, vì không nên ép một thứ làm “sản phẩm phụ”.

Nếu là cửa hàng trưởng tôi sẽ :

1.Combo linh hoạt:

- Cho phép khách chọn 3 trong 6 loại herb

2.Trưng bày:

- Các hộp herb **đặt ngang hàng**
- Không làm 1 loại to – loại khác nhỏ

3.Nội dung quảng bá:

- Tránh câu kiểu: “*Mua Thyme tặng Basil*”
- Thay bằng: “*Bộ herb dùng cùng nhau*”

**Insight #5:** Nhóm HERB MARKER phù hợp để tối ưu cross-selling và bố trí theo cụm

Từ cả hai hình, có thể kết luận rằng:

- Các sản phẩm HERB MARKER nên được đặt gần nhau
- Cross-selling hiệu quả nhất khi:
  - Chọn 1–2 Product Hub (Thyme, Parsley) làm trung tâm
  - Bao quanh bằng các sản phẩm còn lại (Mint, Basil, Rosemary, Chives)

Thay vì quảng bá từng sản phẩm riêng lẻ, chiến lược hiệu quả hơn là:

- Bundle theo “herb cooking set”
- Khuyến mãi xoay quanh Product Hub

Đây là nhóm sản phẩm lý tưởng để tối ưu không gian kệ và doanh thu trên mỗi giờ hàng.

Nếu là cửa hàng trưởng tôi sẽ

Thiết kế lại layout kệ theo cụm sản phẩm

### Hành động cụ thể:

- Layout kệ:

[Parsley] [Thyme]

↓      ↓

[Basil] [Rosemary]

↓

[Mint] [Chives]

- Bảng gợi ý:

- “Bạn thường mua kèm các loại này khi nấu món Âu”

## **2.11. Kết luận & Đề xuất Kinh doanh**

**Kết luận :** Thông qua việc áp dụng thuật toán Apriori trên dữ liệu giờ hàng bán lẻ, nhóm đã xác định rõ ràng sự tồn tại của Product Hub trong nhóm sản phẩm HERB MARKER. Các sản phẩm như *Thyme, Parsley, Rosemary, Basil, Mint* và *Chives* không chỉ xuất hiện thường xuyên trong cùng một giao dịch mà còn hình thành một mạng lưới mua sắm chặt chẽ, thể hiện rõ qua cả network graph và các luật có Lift rất cao.

Kết quả phân tích cho thấy:

- Hành vi mua kèm trong nhóm HERB MARKER không mang tính ngẫu nhiên, mà phản ánh thói quen nấu ăn và nhu cầu sử dụng đồng thời nhiều loại gia vị thảo mộc.
- Một số sản phẩm như HERB MARKER THYME và HERB MARKER PARSLEY đóng vai trò Product Hub trung tâm, có khả năng “kéo theo” các sản phẩm khác trong giờ hàng.
- Mỗi quan hệ mua sắm mang tính hai chiều, cho thấy khách hàng linh hoạt trong việc lựa chọn sản phẩm, thay vì phụ thuộc vào một sản phẩm duy nhất.

Như vậy, Apriori không chỉ giúp phát hiện các luật kết hợp về mặt kỹ thuật, mà còn cung cấp cơ sở dữ liệu vững chắc để ra quyết định kinh doanh, đặc biệt trong các chiến lược cross-selling và tối ưu bố trí hàng hóa.

### **Đề xuất kinh doanh cụ thể cho cửa hàng**

Dựa trên các insight thu được, nếu đứng trên vai trò cửa hàng trưởng / quản lý bán lẻ, có thể triển khai các chiến lược sau:

#### **1. Tối ưu bố trí kệ hàng theo cụm (Cluster-based Layout)**

- Gộp toàn bộ nhóm HERB MARKER thành một khu vực riêng trên kệ thay vì phân tán ở nhiều vị trí.

- Đặt HERB MARKER THYME và HERB MARKER PARSLEY ở vị trí trung tâm hoặc ngang tầm mắt, vì đây là các Product Hub có khả năng kích hoạt mua kèm cao.
- Các sản phẩm còn lại (Mint, Basil, Rosemary, Chives) được bố trí bao quanh theo dạng “vòng tròn” hoặc “cụm”.

Cách bố trí này giúp:

- Tăng khả năng khách hàng nhìn thấy nhiều sản phẩm liên quan cùng lúc
- Gia tăng giá trị trung bình trên mỗi giỏ hàng (basket value)

## 2. Thiết kế combo và gói sản phẩm (Bundle Strategy)

- Tạo các combo nấu ăn như:
  - “Combo gia vị Ý”
  - “Combo herb cơ bản cho người mới nấu ăn”
- Mỗi combo nên lấy Product Hub (Thyme hoặc Parsley) làm sản phẩm chính, đi kèm 2–3 sản phẩm herb khác.

Chiến lược này:

- Khai thác trực tiếp mối quan hệ mua kèm đã được chứng minh bằng Lift cao
- Giúp khách hàng ra quyết định nhanh hơn, đặc biệt với người ít kinh nghiệm nấu ăn

## 3. Chiến lược khuyến mãi xoay quanh Product Hub

- Áp dụng giảm giá nhẹ hoặc ưu đãi cho Product Hub thay vì giảm đồng loạt tất cả sản phẩm.
- Ví dụ:

- Mua HERB MARKER THYME → giảm 10% khi mua thêm Rosemary hoặc Basil
- Tránh giảm giá mạnh các sản phẩm “vệ tinh” vì có thể làm mất vai trò trung tâm của hub.

Điều này giúp:

- Kích thích mua thêm sản phẩm liên quan
- Bảo toàn biên lợi nhuận tổng thể

#### 4. Ứng dụng vào gợi ý bán hàng (Online & POS)

- Với cửa hàng online hoặc hệ thống POS:
  - Khi khách thêm Thyme hoặc Parsley vào giỏ → gợi ý ngay các sản phẩm herb liên quan
- Nội dung gợi ý có thể dựa trên:
  - “Khách hàng thường mua kèm...”
  - “Phù hợp cho món ăn bạn đang chuẩn bị”

Đây là ứng dụng trực tiếp của Apriori vào hệ thống recommendation, giúp cá nhân hóa trải nghiệm mua sắm.

#### 5. Điều chỉnh chiến lược theo mùa và sự kiện

- Nhóm HERB MARKER rất phù hợp cho các chiến dịch theo mùa:
  - Mùa lễ hội
  - Mùa nấu ăn tại nhà
  - Các dịp cuối năm
- Có thể:

- Thay đổi vị trí kệ theo mùa
- Đẩy mạnh combo herb trong các dịp cao điểm

Việc này giúp cửa hàng linh hoạt theo hành vi khách hàng, thay vì giữ bố trí cố định quanh năm.

## 2.12. Link Code & Notebook

- **Notebook:** preprocessing\_and\_eda.ipynb, basket\_preparation.ipynb, apriori\_modelling.ipynb
- **Repo:** Xem repo dự án ở [đây](#)

## 2.13. Slide Trình bày

- Link slide của nhóm xem tại [đây](#)