

PHÂN TÍCH GIỎ HÀNG (MARKET BASKET ANALYSIS) VỚI THUẬT TOÁN APRIORI

1. Giới thiệu bài toán (Business Understanding)

Trong lĩnh vực bán lẻ, việc hiểu rõ hành vi mua sắm đồng thời của khách hàng là chìa khóa để tối ưu doanh thu. Dự án này áp dụng kỹ thuật **Phân tích giỏ hàng (Market Basket Analysis)** sử dụng thuật toán **Apriori** trên bộ dữ liệu bán lẻ trực tuyến (Online Retail).

Mục tiêu:

- Tìm ra các mối quan hệ ẩn giữa các sản phẩm (Ví dụ: Khách mua A thường mua kèm B).
- Đưa ra các quyết định kinh doanh (Bố trí hàng hóa, gợi ý combo, khuyến mãi chéo) dựa trên số liệu thực tế thay vì cảm tính.

2. Quy trình xử lý dữ liệu (Data Pipeline)

Để đảm bảo kết quả chính xác, nhóm đã xây dựng một pipeline xử lý dữ liệu gồm các bước:

- Thu thập dữ liệu:** Sử dụng bộ dữ liệu Online Retail II từ UCI Machine Learning Repository.
- Làm sạch dữ liệu (Data Cleaning):**
 - Loại bỏ các đơn hàng bị hủy (Invoice chứa 'C').
 - Lọc dữ liệu chỉ lấy thị trường Anh (United Kingdom) để đồng nhất hành vi.
 - Loại bỏ các giao dịch thiếu ID khách hàng hoặc có giá trị lỗi (âm).
- Chuẩn bị giỏ hàng (Basket Preparation):** Chuyển đổi dữ liệu từ dạng dòng sang dạng ma trận nhị phân (One-hot encoding), nơi mỗi hàng là một hóa đơn và các cột là sản phẩm (1: có mua, 0: không mua).

3. Kết quả chạy mô hình (Modeling Results)

Nhóm đã thiết lập thuật toán Apriori với các tham số ngưỡng như sau:

- Min Support:** 1% (Sản phẩm phải xuất hiện trong ít nhất 1% tổng số đơn hàng).
- Min Confidence:** 30% (Độ tin cậy của luật phải đạt ít nhất 30%).
- Min Lift:** > 1.2 (Chỉ lấy các quy luật có tương quan dương mạnh).

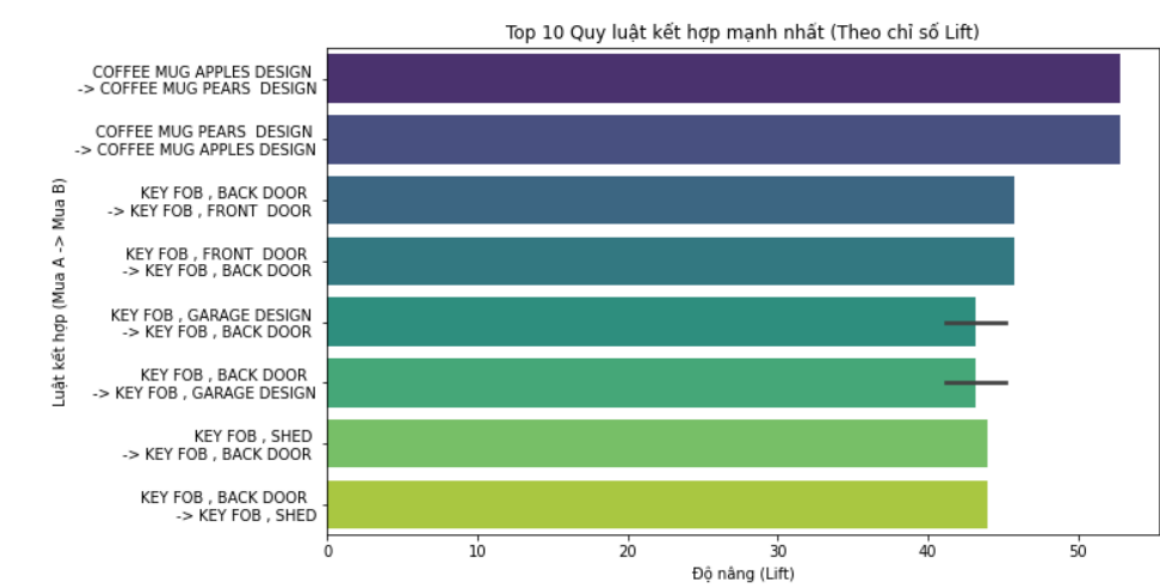
Kết quả thu được: Mô hình đã tìm ra các luật kết hợp có độ mạnh (Lift) rất cao, đặc biệt nổi bật ở hai nhóm sản phẩm chính.

Bảng số liệu Top 10 luật kết hợp mạnh nhất:

=== KẾT QUẢ TOP 10 LUẬT KẾT HỢP MẠNH NHẤT ===			
	antecedents \		
132	(COFFEE MUG APPLES DESIGN)		
133	(COFFEE MUG PEARS DESIGN)		
356	(KEY FOB , BACK DOOR)		
357	(KEY FOB , FRONT DOOR)		
774	(KEY FOB , GARAGE DESIGN)		
771	(KEY FOB , BACK DOOR , KEY FOB , SHED)		
775	(KEY FOB , SHED)		
770	(KEY FOB , BACK DOOR , KEY FOB , GARAGE DESIGN)		
773	(KEY FOB , BACK DOOR)		
772	(KEY FOB , GARAGE DESIGN, KEY FOB , SHED)		
	consequents	support	confidence \
132	(COFFEE MUG PEARS DESIGN)	0.010685	0.718310
133	(COFFEE MUG APPLES DESIGN)	0.010685	0.784615
356	(KEY FOB , FRONT DOOR)	0.011523	0.632184
357	(KEY FOB , BACK DOOR)	0.011523	0.833333
774	(KEY FOB , BACK DOOR , KEY FOB , SHED)	0.010685	0.614458
771	(KEY FOB , GARAGE DESIGN)	0.010685	0.784615
775	(KEY FOB , BACK DOOR , KEY FOB , GARAGE DESIGN)	0.010685	0.515152
770	(KEY FOB , SHED)	0.010685	0.910714
773	(KEY FOB , GARAGE DESIGN, KEY FOB , SHED)	0.010685	0.586207
772	(KEY FOB , BACK DOOR)	0.010685	0.750000

	lift
132	52.746046
133	52.746046
356	45.718391
357	45.718391
774	45.120111
771	45.120111
775	43.907468
770	43.907468
773	41.146552
772	41.146552

Biểu đồ trực quan hóa độ mạnh (Lift) của các luật:



4. Phân tích Insight & Đề xuất kinh doanh (Business Insights)

Dựa trên bảng kết quả và biểu đồ, nhóm rút ra 5 Insight cốt lõi sau đây:

Insight 1: Hiệu ứng "Sưu tập trọn bộ" (The Collection Effect)

- **Quan sát:** Cặp sản phẩm COFFEE MUG APPLES DESIGN và COFFEE MUG PEARS DESIGN có chỉ số **Lift cực khủng (~52.7)**.
- **Ý nghĩa:** Chỉ số Lift > 50 là con số rất hiếm gặp, cho thấy hai sản phẩm này gần như "sinh ra là dành cho nhau". Khách hàng mua cốc hình Táo gần như chắc chắn sẽ mua cốc hình Lê để đủ bộ sưu tập.
- **Hành động đề xuất:** Không nên bán lẻ tẻ. Hãy đóng gói thành **"Combo Gift Set"** (Bộ cốc trái cây) với giá ưu đãi nhẹ. Điều này vừa tăng giá trị đơn hàng trung bình (AOV) vừa giúp khách hàng ra quyết định nhanh hơn.

Insight 2: Hành vi "Đồng bộ hóa an ninh" (Security Synchronization)

- **Quan sát:** Nhóm sản phẩm Móc khóa (Key Fob) gồm: BACK DOOR, FRONT DOOR, GARAGE, SHED có sự liên kết chặt chẽ với nhau (Lift dao động từ 41 - 45).
- **Ý nghĩa:** Khách hàng không mua móc khóa ngẫu nhiên. Họ có xu hướng mua một lần cho toàn bộ các cửa trong nhà (Cửa trước, cửa sau, nhà kho, gara) để dễ phân biệt chìa khóa.
- **Hành động đề xuất:** Tại trang thanh toán, nếu giỏ hàng khách có "Key Fob Front Door", hệ thống cần tự động gợi ý: *"Bạn có muốn mua thêm móc khóa cho Cửa sau và Gara để đồng bộ không?"*.

Insight 3: Sức mạnh của chỉ số Lift

- **Quan sát:** Các luật tìm được đều có Lift > 40, cao hơn rất nhiều so với ngưỡng tối thiểu (1.2).
- **Ý nghĩa:** Đây là các mối quan hệ thực chất, không phải ngẫu nhiên. Chiến lược bán chéo (Cross-selling) áp dụng cho các cặp sản phẩm này sẽ có tỷ lệ thành công (Conversion rate) rất cao.

Insight 4: Tối ưu hóa hiển thị (Visual Merchandising)

- **Đề xuất:** Trên giao diện website hoặc app, hình ảnh của Cốc Táo và Cốc Lê phải luôn được đặt cạnh nhau. Nếu khách hàng xem chi tiết sản phẩm "Móc khóa Gara", mục "Sản phẩm thường mua cùng" phải hiển thị ngay "Móc khóa Cửa sau".

Insight 5: Quản lý tồn kho thông minh

- **Đề xuất:** Vì các sản phẩm này bán theo cặp/nhóm (Cốc đi theo cặp, Móc khóa đi theo bộ), việc nhập hàng cần được đồng bộ. Nếu kho hết "Cốc Táo", doanh số của "Cốc Lê" có thể sẽ bị sụt giảm nghiêm trọng theo. Cần thiết lập cảnh báo tồn kho chung cho cả nhóm sản phẩm này.

5. Kết luận

Thông qua việc áp dụng thuật toán Apriori, chúng ta đã chuyển đổi dữ liệu giao dịch thô thành các tri thức kinh doanh có giá trị. Các quy luật tìm thấy (Cốc đôi, Bộ móc khóa) là cơ sở vững

chắc để xây dựng các chiến dịch Marketing cá nhân hóa và tối ưu hóa doanh thu cho cửa hàng.