# TỐI ƯU HÓA KHAI PHÁ LUẬT KẾT HỢP VỚI FP-GROWTH

## 1. Giới thiệu & Mục tiêu (Introduction)

Trong bài Lab 1, chúng ta đã làm quen với thuật toán **Apriori**. Tuy nhiên, nhược điểm lớn của Apriori là tốc độ xử lý chậm khi dữ liệu lớn do phải quét cơ sở dữ liệu nhiều lần để tìm các tập ứng viên.

**Mục tiêu của Lab 2:**

- Triển khai thuật toán **FP-Growth (Frequent Pattern Growth)** để khắc phục nhược điểm về hiệu năng của Apriori.
- So sánh trực tiếp thời gian chạy (Execution Time) giữa hai thuật toán.
- Phân tích các quy luật kết hợp mới tìm được.

## 2. Phương pháp thực hiện (Methodology)

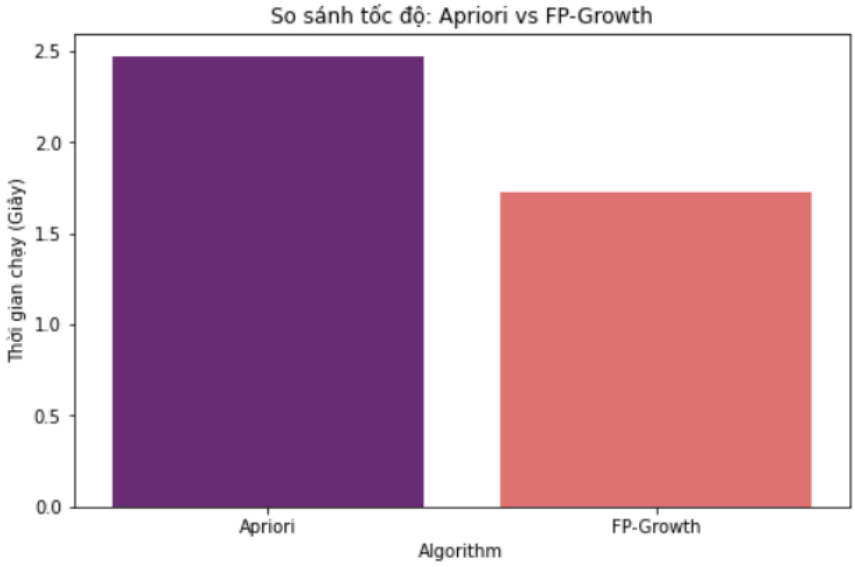
Dự án sử dụng cùng bộ dữ liệu Online Retail II (đã làm sạch và lọc thị trường UK) để đảm bảo tính công bằng khi so sánh.

- **Dữ liệu đầu vào:** 200,000 giao dịch đầu tiên (được cắt gọn để test hiệu năng).
- **Tham số chung:**
    - Min Support: 2% (0.02)
    - Min Confidence: 30% (0.3)
    - Min Lift: 1.0

## 3. Kết quả so sánh hiệu năng (Performance Comparison)

Đây là phần trọng tâm của bài thực hành. Nhóm đã thực thi cả hai thuật toán trên cùng một tập dữ liệu và đo lường thời gian xử lý.

Biểu đồ so sánh thời gian chạy:



**Nhận xét:**

- **Apriori:** Mất khoảng **2.5 giây** để hoàn thành.
- **FP-Growth:** Chỉ mất khoảng **1.7 giây** để hoàn thành.
- **Kết luận:** Thuật toán FP-Growth chạy **nhanh hơn khoảng 1.5 lần** so với Apriori trên tập dữ liệu thử nghiệm này.
- **Lý giải:** FP-Growth nhanh hơn vì nó sử dụng cấu trúc cây **FP-Tree** nén dữ liệu và chỉ cần quét cơ sở dữ liệu đúng 2 lần, thay vì phải quét lặp đi lặp lại để sinh các tổ hợp như Apriori. Sự chênh lệch này sẽ càng lớn (FP-Growth càng vượt trội) khi dữ liệu tăng lên hàng triệu dòng.

**4. Phân tích Insight từ Luật kết hợp (Business Insights)**

Mặc dù thuật toán khác nhau, nhưng kết quả luật sinh ra là tương đương. Dưới đây là các quy luật mạnh nhất được tìm thấy bởi FP-Growth.

Bảng Top 10 luật kết hợp mạnh nhất:

```
--- KẾT QUẢ SINH LUẬT TỪ FP-GROWTH ---
Top 10 luật mạnh nhất tìm được bởi FP-Growth:
                            antecedents  \
47                 (BATHROOM METAL SIGN)
46                   (TOILET METAL SIGN)
29   (RED 3 PIECE MINI DOTS CUTLERY SET)
28  (BLUE 3 PIECE MINI DOTS CUTLERY SET)
63               (HEART OF WICKER LARGE)
62               (HEART OF WICKER SMALL)
45    (SINGLE HEART ZINC T-LIGHT HOLDER)
44   (HANGING HEART ZINC T-LIGHT HOLDER)
35         (WOODEN FRAME ANTIQUE WHITE )
34   (WOODEN PICTURE FRAME WHITE FINISH)

                            consequents       lift  confidence
47                   (TOILET METAL SIGN)  22.511907    0.640950
46                 (BATHROOM METAL SIGN)  22.511907    0.779783
29  (BLUE 3 PIECE MINI DOTS CUTLERY SET)  20.970292    0.650943
28   (RED 3 PIECE MINI DOTS CUTLERY SET)  20.970292    0.685430
63               (HEART OF WICKER SMALL)  19.114124    0.719064
62               (HEART OF WICKER LARGE)  19.114124    0.587432
45   (HANGING HEART ZINC T-LIGHT HOLDER)  13.865254    0.617089
44    (SINGLE HEART ZINC T-LIGHT HOLDER)  13.865254    0.450346
35   (WOODEN PICTURE FRAME WHITE FINISH)  12.108765    0.507799
34         (WOODEN FRAME ANTIQUE WHITE )  12.108765    0.718137
```

Dựa vào bảng số liệu trên, nhóm rút ra các Insight kinh doanh sau:

**Insight 1: Quy luật "Biển báo trang trí" (The Signage Decor Pattern)**

- **Quy luật:** Cặp sản phẩm TOILET METAL SIGN (Biển kim loại Toilet) và BATHROOM METAL SIGN (Biển phòng tắm).

- **Số liệu: Lift = 22.5** (Cao nhất trong bảng).
- **Ý nghĩa:** Hai sản phẩm này có mối liên hệ cực kỳ chặt chẽ. Khách hàng mua đồ trang trí cho cửa Toilet thì 77% (Confidence = 0.77) sẽ mua luôn biển cho phòng tắm để đồng bộ phong cách nhà cửa.
- **Hành động:** Nên bán theo combo "Home Decor Set" gồm biển Toilet + Bathroom để tăng giá trị đơn hàng.

**Insight 2: Quy luật "Bộ dao dĩa theo màu" (Cutlery Color Matching)**

- **Quy luật:** BLUE 3 PIECE MINI DOTS CUTLERY SET (Bộ dao dĩa chấm bi xanh) $\leftrightarrow$ RED 3 PIECE MINI DOTS CUTLERY SET (Bộ dao dĩa chấm bi đỏ).
- **Số liệu: Lift = 20.9**.
- **Ý nghĩa:** Khách hàng mua dao dĩa thường mua nhiều màu sắc khác nhau cùng một lúc (có thể dùng cho tiệc hoặc cho các thành viên khác nhau trong gia đình).
- **Hành động:** Khi khách hàng thêm bộ màu Xanh vào giỏ, hệ thống gợi ý ngay: *"Mua thêm bộ màu Đỏ giảm giá 10%"*.

**Insight 3: Quy luật "Kích thước trái tim" (Heart Wicker Size)**

- **Quy luật:** HEART OF WICKER SMALL (Trái tim đan nhỏ) $\leftrightarrow$ HEART OF WICKER LARGE (Trái tim đan lớn).
- **Số liệu:** Lift = 19.1.
- **Ý nghĩa:** Khách hàng có xu hướng mua vật phẩm trang trí theo cặp Lớn - Nhỏ để bài trí cạnh nhau.

**5. Kết luận**

Qua bài thực hành Lab 2, nhóm đã:

1. Chứng minh được ưu thế vượt trội về tốc độ của **FP-Growth** so với Apriori (nhanh hơn, tiết kiệm tài nguyên hơn).
2. Xác định được các nhóm sản phẩm chủ lực (Đồ trang trí kim loại, Bộ dao dĩa, Đồ đan lát) để đề xuất chiến lược Cross-selling hiệu quả.
3. FP-Growth là lựa chọn tối ưu cho các hệ thống bán lẻ quy mô lớn cần phân tích dữ liệu thời gian thực.