**✅ TOÀN BỘ CÁC GIAI ĐOẠN CỦA THUẬT TOÁN APRIORI**

**🔹 Giai đoạn 1: Khởi tạo – Tìm tập mục đơn thường xuyên**

* Duyệt toàn bộ cơ sở dữ liệu để **đếm số lần xuất hiện** của từng mục (item).
* Giữ lại các mục có **support ≥ min\_support**.
* Gọi là tập L₁ (Frequent 1-itemsets).

**🔹 Giai đoạn 2: Sinh các tập mục ứng viên**

* Từ tập Lₖ (các tập mục thường xuyên có k phần tử), sinh ra tập các **ứng viên độ dài k+1** (gọi là Cₖ₊₁) bằng phép **nối (join)** các tập con.
* Sau đó, **lọc** các ứng viên không hợp lệ bằng cách **loại bỏ các tập con không thường xuyên** (theo nguyên lý Apriori: *"mọi tập con của tập thường xuyên đều phải thường xuyên"*).

**🔹 Giai đoạn 3: Đếm tần suất**

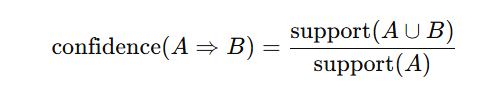
* Với mỗi tập ứng viên trong Cₖ₊₁, duyệt toàn bộ cơ sở dữ liệu để đếm số lần xuất hiện.
* Giữ lại các tập có **support ≥ min\_support**, gọi là Lₖ₊₁.

**🔹 Giai đoạn 4: Lặp lại quá trình**

* Quay lại bước 2 với Lₖ₊₁. Tiếp tục lặp đến khi **không còn tập ứng viên nào đủ điều kiện** (tức là Lₖ = ∅).

**🔹 Giai đoạn 5: Phát sinh luật kết hợp**

* Với mỗi **tập mục thường xuyên L có độ dài ≥ 2**, thực hiện:
  1. Lấy tất cả các tập con không rỗng A của L.
  2. Tính B = L \ A, tạo luật A ⇒ B.
  3. Tính **confidence** của luật:



* 1. Nếu **confidence ≥ min\_confidence**, giữ lại luật.

**✅ KẾT LUẬN**

* **Apriori** là một thuật toán hiệu quả dùng để **khai phá luật kết hợp** trong tập dữ liệu lớn.
* Ưu điểm:
  + Dễ cài đặt, dễ hiểu.
  + Áp dụng nguyên lý Apriori để **giảm không gian tìm kiếm**.
* Nhược điểm:
  + Phải quét cơ sở dữ liệu nhiều lần ⇒ **tốn thời gian với tập dữ liệu lớn**.
  + Sinh ra rất nhiều **ứng viên trung gian**.
* Phù hợp cho các ứng dụng như:
  + Phân tích giỏ hàng (Market Basket Analysis)
  + Gợi ý sản phẩm
  + Phân tích hành vi người dùng

📌 **Ghi nhớ:**

* **Support** đo lường độ phổ biến.
* **Confidence** đo lường độ tin cậy của luật.
* Cần đặt ngưỡng min\_support và min\_confidence hợp lý để khai thác tri thức hiệu quả mà không bị nhiễu.