**BÁO CÁO KHAI THÁC DỮ LIỆU - TUẦN 4**

*Họ tên : Nguyễn Tiến Phong*

*MSSV : 20280071*

**Các bước tự triển khai thuật toán Apriori :**

* Text

  Description automatically generatedĐọc và xử lý dữ liệu
* A screenshot of a computer

  Description automatically generated with low confidenceTìm các ứng viên 1-itemsets, từ đó suy ra các ứng viên k-itemsets

Sau đó tìm các ứng viên frequent\_itemsets có support lớn hơn hoặc bằng ngưỡng min\_support

**min\_support = 0.2**

A black screen with white text

Description automatically generated with low confidence

* Tạo ra các luật kết hợp từ các frequent\_itemsets và tìm các frequent\_itemsets có confidence phù hợp với ngưỡng min\_confidence

**min\_confidence = 0.5**

* A picture containing graphical user interface

  Description automatically generatedTạo ra các luật kết hợp từ các frequent\_itemsets và tìm các frequent\_itemsets có lift phù hợp với ngưỡng min\_lift

**min\_lift = 1.0**

**So sánh với thuật toán sử dụng thư viện Apyori :**

* Kết quả sau khi được lọc với một số điều kiện thì gần giống với kết quả của thuật toán apriori khi tự triển khai thủ công.
* Một số khác biệt nhỏ như thứ tự sắp xếp các frequent\_itemsets, cách xử lý các luật kết hợp, thời gian chạy và bộ nhớ cần thiết. Các frequent\_itemsets được tạo từ thư viện Apyori bao gồm cả các k-itemsets và cả thành phần của chúng nên chúng ta cần phải lọc lại để đạt được kết quả mong muốn.
* Tuy nhiên sau khi triển khai thuật toán thủ công qua nhiều phương thức thì thuật toán dễ tiếp cận hơn. Còn thư viện Apyori được xây dựng để tối ưu hóa thuật toán, có thể thực hiện nhanh và hiệu quả hơn.