

# Báo cáo Lab 02 FIM

Giáo viên hướng dẫn :

- Lê Hoài Bắc
- Nguyễn Ngọc Đức
- Nguyễn Duy Khánh
- Nguyễn Khánh Toàn

Sinh viên thực hiện :

- 18120363 - Đặng Văn Hiến
- 18120647 - Lê Thanh Viễn

**I. Giới thiệu chung:**

MSSV	Họ Tên	Công việc	% đóng góp (tối đa 100%)
18120647	Lê Thanh Viễn	Viết báo cáo	50%
18120363	Đặng Văn Hiến	Cài đặt thuật toán apriori	50%

## II. Thực hiện:

### 1. Ý tưởng:

- Tạo tất cả tập item theo tổ hợp của tất cả item
- Lặp theo kích thước tập item tăng dần :
  - o Chọn tập item con có độ hỗ trợ lớn hơn minsup vào kết quả rồi ngược lại xóa tập item nhỏ hơn và các tập cha của nó

### 2. Thực hiện:

#### Hàm Apriori

```
function apriori(tran_arr, minsup, item_list)
# Chứa tất cả tập item có thể theo kích thước
item_set_arr = [Set(collect(combinations(item_list, k))) for k = 1: length(item_list)]

# Mảng tập phổ biến
result = []

# Tăng kích thước cho lần duyệt tập item
for k = 1: length(item_list)
    print("\n=> Liệt kê tập item kích thước $k : $(item_set_arr[k])\n")

    # Duyệt tập item
    for item_set in item_set_arr[k]
        # Tính độ hỗ trợ của tập item này
        sup = sum([issubset(item_set, tran) for tran in tran_arr]) / length(tran_arr)

        # Loại tập item độ hỗ trợ nhỏ hơn minsup
        if sup < minsup
            print("$item_set có độ hỗ trợ nhỏ hơn minsup nên loại và bỏ tất cả tập item chứa nó\n")

            # bỏ tất cả tập item cha chứa tập item này
            for j = k + 1: length(item_list)
                if length(item_set_arr[j]) == 0
                    break
                end
                filter!(x -> !issubset(item_set, x), item_set_arr[j])
            end
        else
            print( "$item_set có độ hỗ trợ lớn hơn minsup nên thêm vào tập phổ biến\n")
        end
    end
end
```

### III. Điểm mạnh và yếu của 2 thuật toán:

#### a. Apriori:

- Ưu điểm : Đơn giản và dễ hiểu.
- Nhược điểm : Phải quét toàn bộ dữ liệu mỗi lần đếm tần suất tập item, cần nhiều bộ nhớ.

#### b. Tree Projection:

- Ưu điểm : Cần ít bộ nhớ hơn và ít lần quét qua dữ liệu hơn.
- Nhược điểm: Đơn giản và dễ hiểu.

### IV. Ứng dụng:

- Apriori làm việc tốt với lượng dữ liệu lớn.
- Tree Projection làm việc tốt với lượng dữ liệu nhỏ và các tập phổ biến không quá dài.