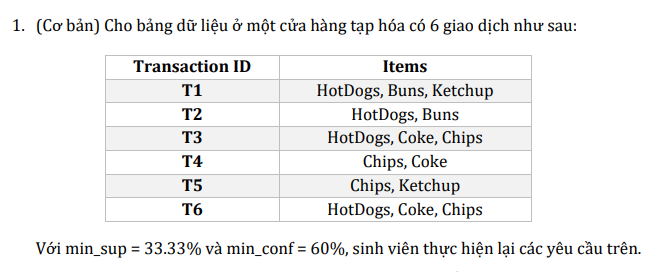
Tên: Nguyễn Cẩm Tú

MSSV : 20520837

**BÀI TẬP THỰC HÀNH LAB 02:**

**Câu 1:**



A.

* Với min\_sup = 33.33% và tổng số giao dịch là 6 => Tần số xuất hiện tối thiểu của phần tử để thỏa min\_sup là 2.
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {HotDogs: 4, Buns: 2, Ketchup: 2, Coke: 3, Chips: 4}
  + - * Các tập phổ biến thỏa min\_sup 2 là: HotDogs, Buns, Ketchup, Coke, Chips
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{HotDogs}, {Buns}, {Ketchup}, {Coke}, {Chips}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{HotDogs, Buns}: 2, {HotDogs, Ketchup}: 1, {HotDogs, Coke}: 2, {HotDogs, Chips}: 2, {Buns, Ketchup}: 1, {Buns, Coke}: 0, {Buns, Chips}: 0, {Ketchup, Coke}: 0, {Ketchup, Chips}: 1, {Coke,Chips}: 3}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {HotDogs, Buns}, {HotDogs, Coke}, {HotDogs, Chips}, {Coke, Chips}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{HotDogs, Buns}, {HotDogs, Coke}, {HotDogs, Chips}, {Coke, Chips}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{HotDogs, Buns, Coke}:0, {HotDogs, Buns, Chips}:0, {HotDogs, Coke, Chips}:2}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {HotDogs, Coke, Chips}
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 = {{HotDogs, Coke, Chips}}
* Vì L3 chỉ còn 1 tập hợp nên không tiến hành tìm tập ứng cử viên 4 phần tử.

b. Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng giải thuật FP-growth

Xây dựng cây FP:

+ Tập phổ biến 1 phần tử và tần suất xuất hiện

|  |  |
| --- | --- |
| Item | Supp |
| HotDogs | 4 |
| Buns | 2 |
| KetChup | 2 |
| Coke | 3 |
| Chips | 4 |

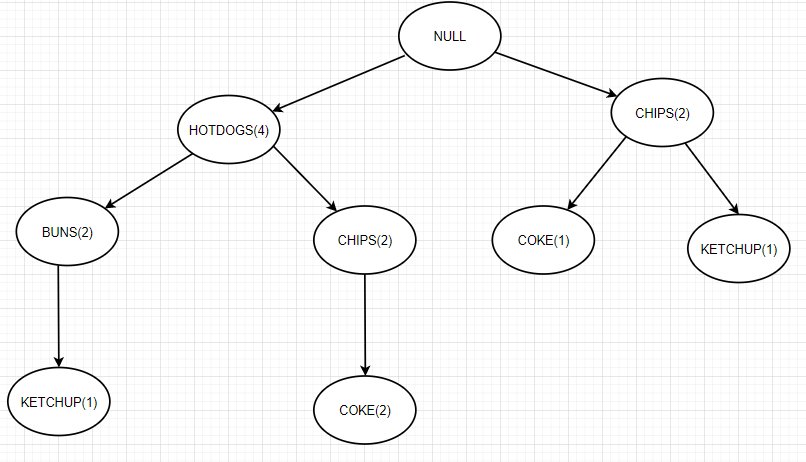
Sau khi sắp xếp giảm dần :

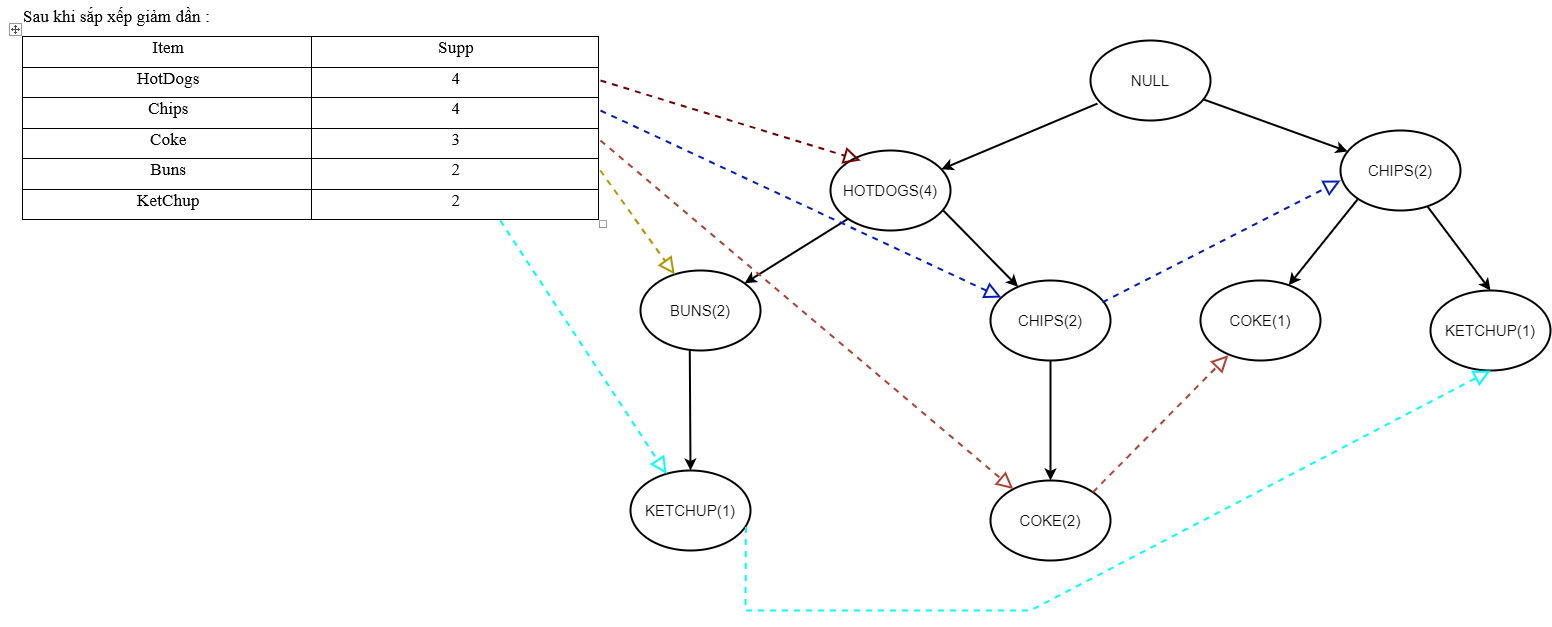
|  |  |
| --- | --- |
| Item | Supp |
| HotDogs | 4 |
| Chips | 4 |
| Coke | 3 |
| Buns | 2 |
| KetChup | 2 |

Sau khi điền cột Ordered frequent items vào

|  |  |  |
| --- | --- | --- |
| Transaction ID | Items | Ordered frequent items |
| T1 | HotDogs, Buns, Ketchup | HotDogs, Buns, Ketchup |
| T2 | HotDogs, Buns | HotDogs, Buns |
| T3 | HotDog,Coke,Chips | HotDog,Coke,Chips |
| T4 | Chips,Coke | Chips,Coke |
| T5 | Chips,Ketchup | Chips,Ketchup |
| T6 | HotDogs,Coke,Chips | HotDogs,Coke,Chips |

Cây FP:



Bảng Header:

Cơ sở mẫu điều kiện:

|  |  |  |
| --- | --- | --- |
| **Item** | **Conditional pattern base** | **Support count** |
| Ketchup | HotDogs,Buns | 1 |
| Chips | 1 |
| Buns | HotDogs | 2 |
| Coke | HotDogs,Chips | 2 |
| Chips | 1 |
| Chips | HotDogs | 2 |
| HotDogs | NULL | |

Xây dựng cơ sở điều kiện cho mỗi mục phổ biến:

|  |  |  |
| --- | --- | --- |
| Item | **Conditional Pattern Base** | **Conditional FP-Tree (supp >= 2)** |
| Chips | **HotDogs : 2** | **HotDogs:2 | Chips** |
| Coke | **HotDogs,Chip :2**  **Chips : 1** | **Chips:3**  **Hotdogs:2**  **Chips,HotDogs :2 |Coke** |
| Buns | **HotDogs : 2** | **HotDogs: 2 |Buns** |
| Ketchup | **HotDogs , Buns :1**  **Chips :1** | **{}** |
| Hotdogs | **{}** | **{}** |

Tìm tất cả các luật kết hợp dựa trên các luật phổ biến ở câu a

* Xét tập phổ biến {HotDogs, Buns}: 2 có min\_conf = 60%, có 2 tập con không rỗng như sau :
* {HotDogs}: 4, {Buns}: 2 {HotDogs, Buns}: 2
* Ta có

|  |  |
| --- | --- |
| HotDogs 🡪 Buns | 2/4 = 50% |
| Buns 🡪 HotDogs | 2/2 = 100% |

* Với min\_conf = 60% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Buns => HotDogs**
* Xét tập phổ biến {HotDogs, Chips}: 2 có min\_conf = 60% và có các tập con không rỗng như sau:
* {HotDogs}: 4, {Chips}: 4, {HotDogs, Chips}: 2
* Ta có

|  |  |
| --- | --- |
| HotDogs 🡪 Chips | 2/4 = 50% |
| Chips 🡪 HotDogs | 2/4 = 50% |

* Với min\_conf = 60% dựa vào bảng trên ta có luật kết hợp ta không tìm được luật thỏa yêu cầu min\_conf = 60%
* Xét tập phổ biến {HotDogs, Coke}: 2 có min\_conf = 60% có các tập con không rỗng như sau
* {HotDogs}: 4, {Coke}: 3, {HotDogs, Coke}: 2
* Ta có

|  |  |
| --- | --- |
| HotDogs 🡪 Coke | 2/4 = 50% |
| Coke 🡪 HotDogs | 2/3 = 6.67% |

* Với min\_conf = 60% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là:

**Coke => HotDogs**

* Xét tập phổ biến {Chips, Coke}: 3 có min\_conf = 60% có các tập con không rỗng như sau
* {Chips}:4, {Coke}: 3, {Chips, Coke}:3
* Ta có

|  |  |
| --- | --- |
| Chips 🡪 Coke | 3/4 =75% |
| Coke 🡪 Chips | 3/3 = 100% |

* Với min\_conf = 60% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là:

**Coke => Chips, Chips => Coke**

* Xét tập phổ biến {HotDogs, Chips, Coke}: 2 có min\_conf = 60%, có các tập con khác rỗng như sau
* {HotDogs}: 4, {Coke}: 3, {Chips}: 4, {HotDogs, Coke}: 2, {HotDogs, Chips}: 2, {Chips, Coke}: 3, {HotDogs, Chips, Coke}: 2
* Ta có

|  |  |
| --- | --- |
| HotDogs 🡪{Chips, Coke} | 2/4 = 50% |
| Coke 🡪 {HotDogs, Chips} | 2/3 = 66,67% |
| Chips 🡪 {HotDogs, Coke} | 2/4 = 50% |
| {Chips, Coke} 🡪 HotDogs | 2/3 =66,67% |
| {HotDogs, Chips} 🡪 Coke | 2/2 = 100% |
| {HotDogs, Coke} 🡪 Chips | 2/2 = 100% |

* Với min\_conf = 60% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là:

**Coke 🡪 {HotDogs, Chips}, {Chips, Coke} 🡪 HotDogs, {HotDogs, Chips} 🡪 Coke, {HotDogs, Coke} 🡪 Chips**

**Câu 2 :**



Với min\_sup = 30% và min\_conf = 80%, sinh viên thực hiện lại các yêu cầu trên.

+ Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán Apriori

Với min\_sup = 30% và tổng số giao dịch là 10 => Tần số xuất hiện tối thiểu của phần tử để thỏa min\_sup là 3.

Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:

C1 = {Kéo: 2, Compa: 3, Thước: 2, Tập trắng: 5, Bút bi: 5, Bút màu: 2, Tẩy:2}

Các tập phổ biến thỏa min\_sup ≥ 3 là: Compa, Tập trắng, Bút bi

Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):

L1 = {{Compa}, {Tập trắng}, {Bút bi}}

Tập các ứng viên 2 phần tử và tần số xuất hiện của nó

C2 = L1⋈L1 = {{Compa; Tập trắng}: 3, {Compa; Bút bi}:3, {Bút bi, Tập trắng}:4}

Các tập phổ biến thỏa min\_sup ≥ 3 là: {Compa; Tập trắng}, {Compa; Bút bi}, {Bút bi, Tập trắng}

Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):

L2 = {{Compa; Tập trắng}, {Compa; Bút bi}, {Bút bi, Tập trắng}}

Tập các ứng viên 3 phần tử và tần số xuất hiện của nó

C3 = L2⋈L2 = {{Compa; Tập trắng; Bút bi}:3}

Các tập phổ biến thỏa min\_sup ≥ 3 là: {Compa; Tập trắng; Bút bi}

Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):

L3 = {{Compa; Tập trắng; Bút bi}}

Vì L3 chỉ còn 1 tập hợp nên không tiến hành tìm tập ứng cử viên 4 phần tử.

+ Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán FP-Growth

Tập phổ biến 1 phần tử và tần suất xuất hiện của nó

|  |  |
| --- | --- |
| **Item** | **Supp** |
| Kéo | 2 |
| Compa | 3 |
| Thước | 2 |
| Tập trung | 5 |
| Bút bi | 5 |
| Bút màu | 2 |
| Tẩy | 2 |

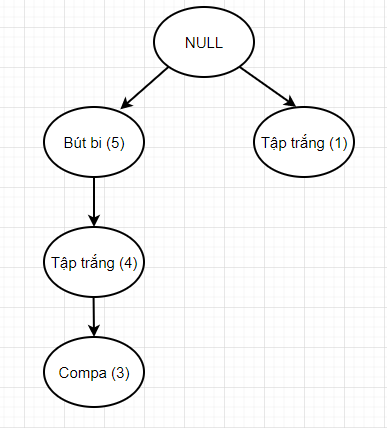
Sắp xếp giảm dần và tần xuất thỏa điều kiện >= min\_supp

|  |  |
| --- | --- |
| **Item** | **Supp** |
| Bút bi | 5 |
| Tập trắng | 5 |
| Compa | 3 |

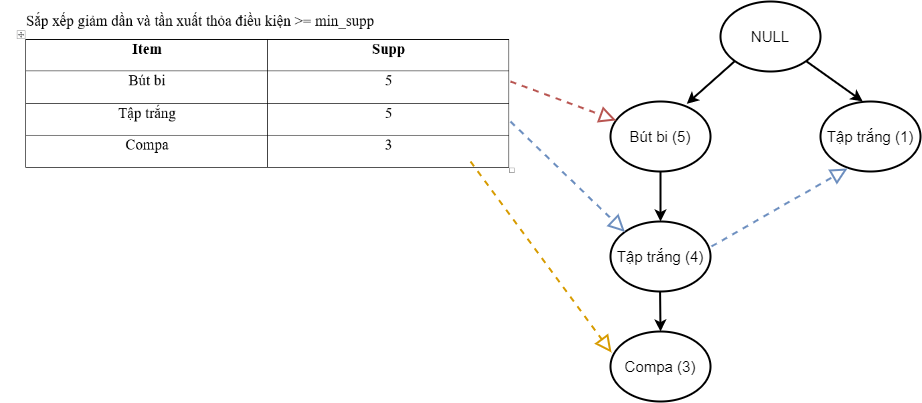
Điền cột Order Frequent Items vào:

|  |  |  |
| --- | --- | --- |
| **Transaction ID** | **Item** | **Ordered frequent items** |
| T1 | Compa,Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T2 | Kéo, Thước, Tập trắng, Bút bi | Bút bi, Tập trắng |
| T3 | Compa, Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T4 | Kéo, Compa, Tập trắng, Bút bi | Bút bi, Tập trắng, Compa |
| T5 | Thước |  |
| T6 | Bút bi | Bút bi |
| T7 | Tập trắng | Tập trắng |
| T8 | Tẩy |  |
| T9 | Bút màu, Tẩy |  |
| T10 | Bút màu |  |

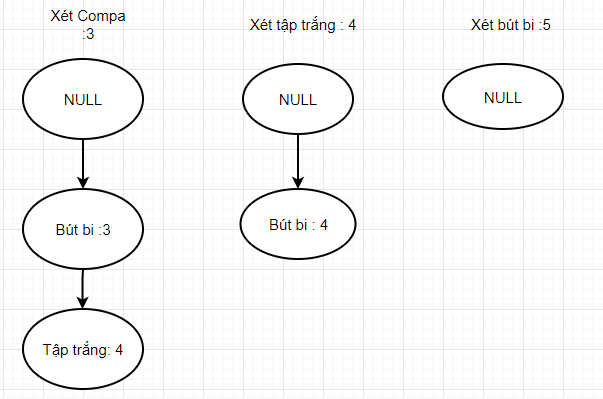
* Vẽ cây FP:



Bảng Header :



Cây FP điều kiện :



Kết quả các mẫu thường xuyên :

• Lấy những phần tử có 2 phần tử trở lên và thỏa min\_supp ta được các tập hợp sau đây:

o L2 = {{Compa; Bút bi}:3, {Compa; Tập trắng}: 3, {Bút bi; Tập trắng}:4}

o L3 = {{Compa; Bút bi; Tập trắng}:3}

Tìm luật kết hợp dựa trên các tập phổ biến

* Xét tập phổ biến {Compa; Bút bi}: 3 có min\_conf = 80%, có các tập con khác rỗng
* {Compa}:3, {Bút bi}: 5, {Compa; Bút bi}:3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 Bút bi | 3/3 = 100% |
| Bút bi 🡪 Compa | 3/5 = 60% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **Bút bi**
* Xét tập phổ biến {Compa; Tập trắng}: 3 có min\_conf = 80%, có các tập con khác rỗng
* {Compa}:3, {Tập trắng}: 5, {Compa; Tập trắng}:3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 Tập trắng | 3/3 = 100% |
| Tập trắng 🡪 Compa | 3/5 = 60% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **Tập trắng**
* Xét tập phổ biến {Bút bi; Tập trắng}:4 có min\_conf = 80%, có các tập con khác rỗng
* {Bút bi}: 5, {Tập trắng}: 5, {Bút bi; Tập trắng}: 4
* Ta có

|  |  |
| --- | --- |
| Bút bi🡪 Tập trắng | 4/5 = 80% |
| Tập trắng 🡪 Bút bi | 4/5 = 80% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Bút bi** 🡪 **Tập trắng, Tập trắng** 🡪 **Bút bi**
* Xét tập phổ biến {Compa; Bút bi; Tập trắng}: 3 có min\_conf = 80%, có các tập con khác rỗng
* {Compa}: 3, {Bút bi}: 5, {Tập trắng}: 5, {Compa; Bút bi}: 3, {Compa; Tập trắng}: 3; {Bút bi; Tập trắng}: 4, {Compa; Bút bi; Tập trắng}: 3
* Ta có

|  |  |
| --- | --- |
| Compa 🡪 {Bút bi; Tập trắng} | 3/3 = 100% |
| Bút bi 🡪 {Tập trắng; Compa} | 3/5 = 60% |
| Tập trắng 🡪 {Bút bi; Compa} | 3/5 = 60% |
| {Bút bi; Tập trắng} 🡪 Compa | 3/4 =75% |
| {Tập trắng; Compa} 🡪 Bút bi | 3/3 = 100% |
| {Bút bi; Compa} 🡪 Tập trắng | 3/3 = 100% |

* Với min\_conf = 80% dựa vào bảng trên ta có luật kết hợp thỏa yêu cầu là: **Compa** 🡪 **{Bút bi; Tập trắng}, {Tập trắng; Compa}** 🡪 **Bút bi, {Bút bi; Compa}** 🡪 **Tập trắng**

**Câu 3:**



Vì B = { Tuổi, Năng lực làm việc, Thăng chức } nên ta có bảng dữ liệu:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tuổi (T)** | **Năng lực làm việc ( NL)** | **Thăng chức**  **( TC)** |
| 1 | 20..25 | Giỏi | Có |
| 2 | 20..25 | Khá | Không |
| 3 | 26..30 | Giỏi | Có |
| 4 | 31..40 | Trung bình | Có |
| 5 | 26..30 | Trung bình | Không |
| 6 | 31..40 | Khá | Không |
| 7 | 26..30 | Khá | Không |
| 8 | 31..40 | Khá | Có |
| 9 | >40 | Giỏi | Không |
| 10 | 26..30 | Giỏi | Có |

Đưa dữ liệu về dạng như câu 2:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **Tuổi (T)** | | | | **Năng lực làm việc ( NL)** | | | **Thăng chức**  **( TC)** | |
| **20..25** | **26..30** | **31..40** | **>40** | **Giỏi** | **Khá** | **TB** | **Có** | **Không** |
| **1** | **x** |  |  |  | **x** |  |  | **x** |  |
| **2** | **x** |  |  |  |  | **x** |  |  | **x** |
| **3** |  | **x** |  |  | **x** |  |  | **x** |  |
| **4** |  |  | **x** |  |  |  | **x** | **x** |  |
| **5** |  | **x** |  |  |  |  | **x** |  | **x** |
| **6** |  | **x** |  |  |  | **x** |  |  | **x** |
| **7** |  |  | **x** |  |  | **x** |  |  | **x** |
| **8** |  | **x** |  |  |  | **x** |  | **x** |  |
| **9** |  |  |  | **x** | **x** |  |  |  | **x** |
| **10** |  | **x** |  |  | **x** |  |  | **x** |  |

Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán Apriori

* Với min\_sup = 30% và tổng số giao dịch là 10 => Tần số xuất hiện tối thiểu của phần tử để thỏa min\_sup là 3.
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {{20..25}:2, {26..30}:5, {31..40}:2, {>40}: 1, Giỏi: 4, Khá: 4, Trung bình: 2, Có: 5, Không: 5}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {26..30}, Giỏi, Khá, Có, Không
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{26..30}, {Giỏi}, {Khá}, {Có}, {Không}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{26..30, Giỏi}: 2, {26..30, Khá}: 2, {26..30, Có}:3, {26..30, Không}: 2, {Giỏi, Có}:3, {Giỏi, Không}:1, {Khá, Có}:1, {Khá, Không}:3}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {26..30, Có}, {Giỏi, Có}, {Khá, Không}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{26..30, Có}, {Giỏi, Có}, {Khá, Không} }
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{26..30, Giỏi, Có}: 2}
  + - * Các tập phổ biến thỏa min\_sup 3 là: không có
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 =
* Vì L3 không có tập hợp nào nên không tiến hành tìm tập ứng cử viên 4 phần tử.

Tìm luật có về phải là thăng chức

* Xét tập phổ biến {26..30, Có} có min\_conf = 80%
* {26..30}: 5, {Có}: 5, {26..30, Có}: 3
* Ta có

|  |  |
| --- | --- |
| 26..30🡪Có | 3/5= 60% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {26..30, Có} không có luật kết hợp có về phải là thuộc tính Thăng chức.
* Xét tập phổ biến {Giỏi, Có} có min\_conf = 80%
* {Giỏi}: 4, {Có}: 5, {Giỏi, Có}: 3
* Ta có

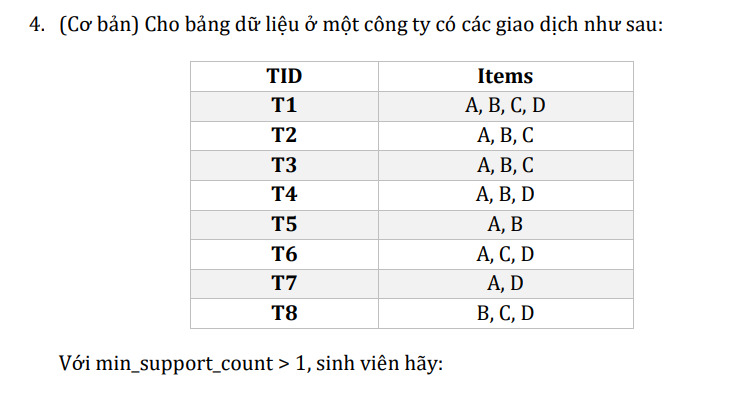
|  |  |
| --- | --- |
| Giỏi🡪Có | 3/4= 75% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {Giỏi, Có} không có luật kết hợp có về phải là thuộc tính Thăng chức.
* Xét tập phổ biến {Khá, Không} Không min\_conf = 80%
* {Khá}: 4, {Không}: 5, {Khá, Không}: 3
* Ta có:

|  |  |
| --- | --- |
| Khá🡪Không | 3/4= 75% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {Khá, Không} không có luật kết hợp có về phải là thuộc tính Thăng chức.
  + - * Không có luật kết hợp cần tìm với min\_conf = 80%.

**Câu 4:**



+ Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán Apriori

* Với min\_sup \_count > 1
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {A:7, B:6, C:5, D:5}
  + - * Các tập phổ biến thỏa min\_sup 2 là: A, B, C, D
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{A}, {B}, {C}, {D}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{A, B}: 5, {A, C}: 4, {A, D}: 3, {B, C}: 4, {B, D}: 3, {C, D}: 3}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{A, B, C}: 3, {A, B, D}: 2, {A, C, D}: 2, {B, C, D}: 2}
  + - * Các tập phổ biến thỏa min\_sup 2 là: {A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 = {{A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}}
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C4 = L3⋈L3 = {{A, B, C, D}: 1}
  + - * Không có tập phổ biến 4 phần tử thỏa min\_sup 2 là
      * Các tập phổ biến 4 phần tử (tập các ứng viên thoải min\_support\_count):
* L4 =

Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng giải thuật FP-growth:

* Tập phổ biến 1 phần tử và tần suất xuất hiện:

|  |  |
| --- | --- |
| **Item** | **Supp** |
| A | 7 |
| B | 6 |
| C | 5 |
| D | 5 |

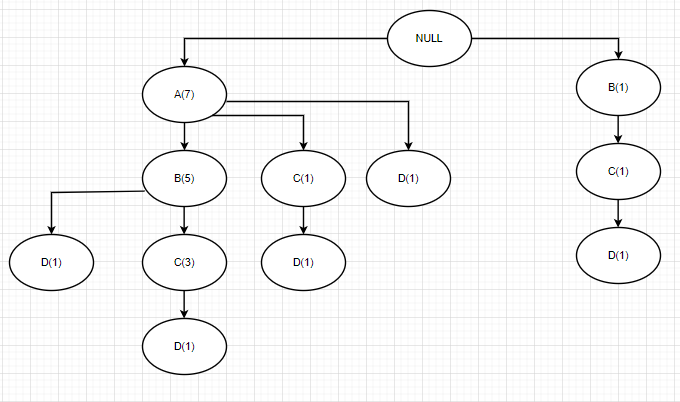
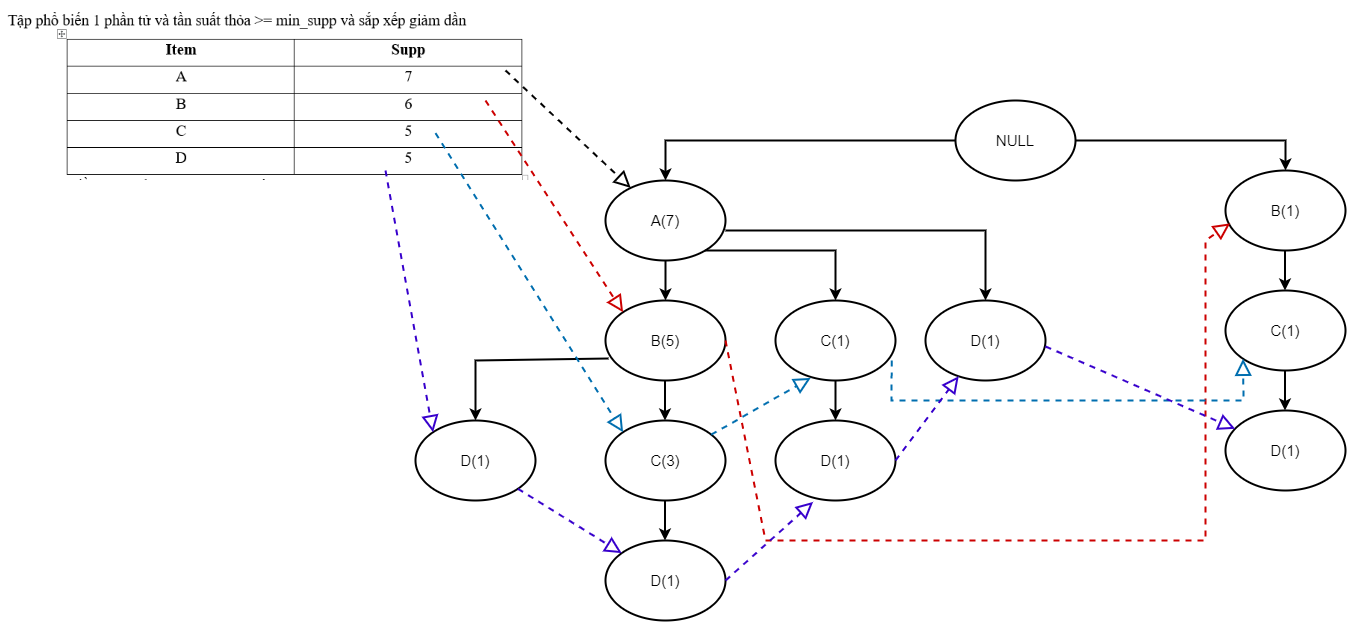
Tập phổ biến 1 phần tử và tần suất thỏa >= min\_supp và sắp xếp giảm dần

|  |  |
| --- | --- |
| **Item** | **Supp** |
| A | 7 |
| B | 6 |
| C | 5 |
| D | 5 |

* Điền cột Order Frquent Items vào:

|  |  |
| --- | --- |
| **Transaction ID** | **Items( Order)** |
| T1 | A,B,C,D |
| T2 | A,B,C |
| T3 | A,B,C |
| T4 | A,B,D |
| T5 | A,B |
| T6 | A,C,D |
| T7 | A,D |
| T8 | B,C,D |

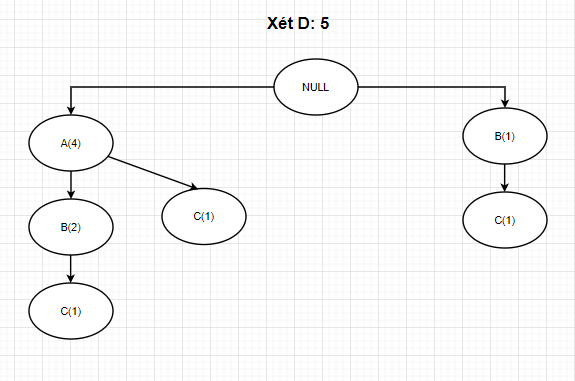
Cây FP:

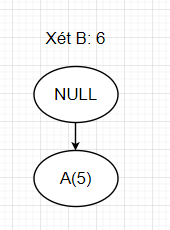
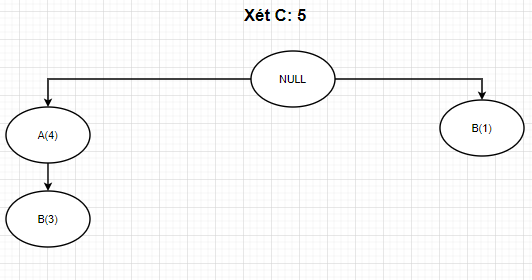


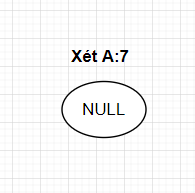
Cơ sở mẫu điều kiện :

|  |  |  |
| --- | --- | --- |
| Item | Conditional pattern base | Support Count |
| D | A,B,C | 1 |
| A,C | 1 |
| A,B | 1 |
| A | 1 |
| B,C | 1 |
| C | A,B | 3 |
| A | 1 |
| B | 1 |
| B | A | 5 |
| A | NULL | |

Xây dựng cây FP điều kiện:





* Lấy những phần tử có 2 phần tử trở lên và thỏa min\_supp ta được các tập hợp sau đây:
* L2 = {{A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}}
* L3 = {{A, B, C}, {A, B, D}, {A, C, D}, {B, C, D}}