BÀI TẬP THỰC HÀNH \_ CÂU 1

(Cơ bản) Cho bảng dữ liệu ở một cửa hàng tạp hóa có 6 giao dịch như sau:

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

Với min\_sup = 33.33% và min\_conf = 60%

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

* 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.

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

* Xây dựng cây FP
* Tập phổ biến 1 phần tử và tần suất xuất hiện của nó

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

* Tập phổ biến 1 phần tử và tần xuất thỏa điều kiện min\_supp và đã được sắp xếp giảm dần

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

* Điền cột Order Frequent Items vào

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

* Vẽ cây FP

**Diagram

Description automatically generated**

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

|  |  |
| --- | --- |
| Item | Conditional Pattern Base |
| Chips | {HotDogs}: 2 |
| Coke | {HotDogs, Chips,}: 2, {Chips}: 1 |
| Buns | {HotDogs}: 2 |
| Ketchup | {HotDogs, Buns}: 1, {Chips}: 1 |

* Xây dựng cây FP điều kiện

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

* Kết quả các mẫu thường xuyên

|  |  |  |
| --- | --- | --- |
| Item | Conditional FP-Tree | Frequent Pattern |
| Buns | {{HotDogs}:2} |Buns | {{Buns}:2, {HotDogs, Buns}:2} |
| Chips | {{HotDogs}:2} |Chips | {{Chips}:2, {HotDogs, Chips}:2} |
| Coke | {{Chips}:3, {Hotdogs}: 2, {Hotdogs, Chips}: 2} |Coke | {{Coke}: 3, {Coke, Chip}:3, {Hotdogs, Coke}: 2, {Hotdogs, Chips, Coke}: 2} |
| Ketchup | {} | {{Ketchup}:2} |
| HotDogs | {} | {{HotDogs}:4} |

* Lấy những phần tử có 2 phần tử trở lên và thỏa min\_supp ta được các tập hợp sau đây:
* L2 = {{HotDogs, Buns}: 2, {HotDogs, Chips}: 2, {HotDogs, Coke}:2, {Chips, Coke}:3}
* L3 = {{HotDogs, Chips, Coke}:2}

1. **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%
* {HotDogs}: 4, {Buns}: 2 {HotDogs, Buns}: 2
* Ta có

|  |  |
| --- | --- |
| HotDogs 🡪 Buns | 2/4 = 50% |
| Buns 🡪 HotDogs | 2/2 = 1 |

* 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%
* {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%
* {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%
* {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%
* {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**