**Bài 1**: Tìm luật sử dụng giải thuật Apriori với bộ dữ liệu sau, cho biết Support threshold=50%. Hãy tính xác suất Confidence của tập kết quả.

| **Transaction** | **List of items** |
| --- | --- |
| T1 | I1, I2, I3 |
| T2 | I2, I3, I4 |
| T3 | I2, I4, I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I2, I3, I5 |
| T6 | I1, I2, I3, I4 |

Support threshold=50% => 0.5\*6= 3 => min\_sup=3

* Bước 1: K = 1
* C1:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1 | 4 |
| I2 | 6 |
| I3 | 4 |
| I4 | 4 |
| I5 | 2 |

* Vì I5 < min\_sup (2 < 3) nên xóa I5. L1:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1 | 4 |
| I2 | 6 |
| I3 | 4 |
| I4 | 4 |

* **Bước 2:** K = 2
* C2:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1, I2 | 4 |
| I1, I3 | 3 |
| I1, I4 | 2 |
| I2, I3 | 4 |
| I2, I4 | 4 |
| I3, I4 | 2 |

* Vì {I1, I4} và {I3, I4} < min\_sup (2 < 3) nên xóa {I1, I4} và {I3, I4}. L2:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1, I2 | 4 |
| I1, I3 | 3 |
| I2, I3 | 4 |
| I2, I4 | 4 |

* **Bước 3:** K = 3
* **{I1, I2, I3} là phổ biến.** C3:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1, I2, I3 | 3 |

* L3:

| **Itemset** | **sup\_count** |
| --- | --- |
| I1, I2, Ỉ3 | 3 |

* **Bước 4:** Confidence(A🡪B) = sup\_count(A∪B)/sup\_count(A)
* {I1, I2} => {I3}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I1, I2} = (3/4) \* 100 = 75%
* {I1, I3} => {I2}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I1, I3} = (3/3) \* 100 = 100%
* {I2, I3} => {I1}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I2, I3} = (3/4) \* 100 = 75%
* {I1} => {I2, I3}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I1} = (3/4) \* 100 = 75%
* {I2} => {I1, I3}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I2} = (3/6) \* 100 = 50%
* {I3} => {I1, I2}, Confidence = sup\_count {I1, I2, I3} / sup\_count {I3} = (3/4) \* 100 = 75%

🡪 Nếu độ tin cậy tối thiểu là 50% thì tất cả các quy tắc trên có thể được coi là quy tắc kết hợp mạnh mẽ.