

Bước 1: Thiết lập thông số

- Tổng số dòng (N): 16
- MinSup = 0.25 (25%) ⇒ Số lần xuất hiện tối thiểu = 16 × 0.25 = 4. (Cái nào xuất hiện dưới 4 lần là loại ngay).
- MinConf = 0.75 (75%): ⇒ Chỉ nhìn các luật có độ chính xác từ 75% trở lên.

Bước 2: Tìm tập phổ biến (Frequent Itemsets)

1. Đếm từng món (1-Itemset)

Nhìn vào bảng và đếm số lượng số 1 của từng cột:

- Kem: 10 lần (Dòng 1,3,5,6,7,8,11,13,14,16) → Giữ
- Bơ: 11 lần → Giữ
- Bánh mì: 8 lần → Giữ
- Chuối: 9 lần → Giữ

2. Đếm cặp đôi (2-Itemset)

Đếm số dòng mà cả 2 cột đều là số 1:

- (Kem, Bơ): 5 lần (Dòng 1, 5, 7, 11, 14) → Giữ
- (Kem, Bánh mì): 4 lần (Dòng 5, 6, 14, 16) → Giữ (Vừa đủ mập)
- (Kem, Chuối): 5 lần (Dòng 3, 6, 9, 13, 16) → Giữ
- (Bơ, Bánh mì): 6 lần (Dòng 2, 5, 8, 10, 12, 14) → Giữ
- (Bơ, Chuối): 4 lần (Dòng 2, 8, 10, 12) → Giữ
- (Bánh mì, Chuối): 6 lần (Dòng 2, 6, 8, 10, 12, 16) → Giữ

(Tất cả các cặp đều đạt chuẩn ≥ 4)

3. Đếm bộ ba (3-Itemset)

Ghép các cặp lại và kiểm tra xem dòng nào cả 3 cột đều là 1:

- (Kem, Bơ, Bánh mì): Dòng 5, 14 → Tổng 2 lần. (LOẠI vì < 4)
- (Kem, Bơ, Chuối): Không có dòng nào cả 3 cùng là 1. (LOẠI)
- (Kem, Bánh mì, Chuối): Dòng 6, 16 → Tổng 2 lần. (LOẠI)
- (Bơ, Bánh mì, Chuối): Dòng 2, 8, 10, 12 → Tổng 4 lần. (GIỮ)

⇒ Kết luận bước 2: Chỉ có duy nhất một tập 3 phần tử phổ biến là (Bơ, Bánh mì, Chuối) với số lần xuất hiện là 4.

Bước 3: Sinh luật và Kiểm tra (Tinh Confidence)

Chúng ta sẽ ưu tiên tìm luật từ tập lớn nhất tìm được là (Bơ, Bánh mì, Chuối) (Support = 4).

Công thức: $Confidence(A \rightarrow B) = \frac{Support(A \text{ và } B)}{Support(A)}$

Xét các luật có thể xảy ra:

- Luật: Nếu mua (Bơ, Chuối) → Mua (Bánh mì)
 - Support((Bơ, Chuối, Bánh mì)) = 4
 - Support((Bơ, Chuối)) = 4 (Lấy từ bước 2)
 - Confidence = 4/4 = 1.0 (100%)
 - Kết luận: 100% > 75% → LUẬT HOP LỆ (Rất mạnh)
- Luật: Nếu mua (Bơ, Bánh mì) → Mua (Chuối)
 - Support((Bơ, Bánh mì)) = 6
 - Confidence = 4/6 ≈ 0.67 (67%)
 - Kết luận: 67% < 75% → LOẠI
- Luật: Nếu mua (Bánh mì, Chuối) → Mua (Bơ)
 - Support((Bánh mì, Chuối)) = 6
 - Confidence = 4/6 ≈ 0.67 (67%)
 - Kết luận: 67% < 75% → LOẠI

Bài 2.

1. Thiết lập tham số (Config)

- Tổng số dòng (N): 20.
- Chọn MinSup = 50% (0.5): ⇒ Môn học hoặc combo nào phải xuất hiện ít nhất 10 lần mới được xét.
- Chọn MinConf = 70% (0.7): ⇒ Chỉ lấy luật nào có độ chính xác trên 70%.

2. Bước 1: Tìm tập phổ biến 1 phần tử (1-Itemset)

- CSLT (Cơ sở tập trình): 15 lần → Giữ (15 > 10)
- KTLT (Kỹ thuật lập trình): 13 lần → Giữ
- CTDL (Cơ sở dữ liệu): 12 lần → Giữ
- OOP (Lập trình hướng đối tượng): 15 lần → Giữ
- LT CSDL (Lý thuyết CSDL): 15 lần → Giữ

3. Bước 2: Tìm tập phổ biến 2 phần tử (Cặp đôi)

- {CSLT, OOP}:
 - Đếm được: Dòng 1, 4, 5, 6, 7, 11, 12, 13, 14, 15, 19, 20.
 - Tổng: 12 lần.
 - Support: 12/20 = 60% (Đạt chuẩn > 50%).
- {CSLT, KTLT}:
 - Đếm được: Dòng 1, 2, 5, 6, 10, 11, 12, 14, 15, 18, 19.
 - Tổng: 11 lần.
 - Support: 11/20 = 55% (Đạt).
- {OOP, LT CSDL}:
 - Đếm được: 11 lần.
 - Confidence = 4/6 ≈ 0.67 (67%)
 - Support: 55% (Đạt).
- {CSLT, LT CSDL}:
 - Đếm được: 10 lần.
 - Support: 50% (Vừa đủ mập Đạt).

4. Bước 3: Sinh luật và Tính Confidence

Luật A: Nếu học CSLT → Sẽ học OOP

- Dữ liệu:
 - Số sinh viên học cả 2: 12
 - Số sinh viên học CSLT: 15
- Tính toán:

$$Confidence = \frac{12}{15} = 0.8 \quad (80\%)$$

- Kết luận: 80% > 70% → Luật này HOP LỆ.
- Ý nghĩa: Một sinh viên đã đăng ký "Cơ sở tập trình" thì 80% khả năng họ cũng học "OOP".

Luật B: Nếu học OOP → Sẽ học CSLT

- Dữ liệu:
 - Số sinh viên học cả 2: 12
 - Số sinh viên học OOP: 15
- Tính toán:

$$Confidence = \frac{12}{15} = 0.8 \quad (80\%)$$

- Kết luận: HOP LỆ.

Mở rộng: Thủ tính cho cặp {OOP, LT CSDL} (Support chung = 11)

Luật C: Nếu học LT CSDL → Sẽ học OOP

- Support(LT CSDL) = 15.

$$Conf = \frac{11}{15} \approx 0.733 \quad (73.3\%)$$

- Kết luận: 73.3% > 70% → HOP LỆ.

Luật D: Nếu học OOP → Sẽ học LT CSDL

- Support(OOP) = 15.

$$Conf = \frac{11}{15} \approx 0.733 \quad (73.3\%)$$

- Kết luận: HOP LỆ.

Bài 3.

Bước 1: Thiết lập thông số đầu vào

- Tổng số giao dịch (N): 16 dòng.
- MinSup = 0.4 (40%).
 - Ngưỡng đếm = 16 × 0.4 = 6.4.
 - ⇒ Một tập hợp phải xuất hiện từ 7 lần trở lên mới được giữ lại (làm tròn lên vì không có 0.4 lần xuất hiện).
- MinConf = 0.8 (80%).
 - Luật tìm ra phải có độ tin cậy từ 80% trở lên.

Bước 2: Tìm tập phổ biến (Frequent Itemsets)

1. Đếm 1 phần tử (L1)

Ta đếm số lần xuất hiện (số 1) của từng cột:

- Kem: 10 lần (Dòng 1,3,5,6,7,8,11,13,14,16) → Support = 10/16 (Pass)
- Bơ: 11 lần → Support = 11/16 (Pass)
- Bánh mì (BM): 8 lần → Support = 8/16 (Pass)
- Chocolate: 9 lần → Support = 9/16 (Pass)
- Chuối: 12 lần → Support = 12/16 (Pass)

2. Đếm cặp đôi (L2) - Chỉ giữ cặp ≥ 7 lần

Sau khi ghép và đếm (bước này tốn công nhất), ta lọc ra được các cặp "sống sót":

- (Kem, Chuối): 9 lần (Dòng 1,3,5,6,8,11,13,14,16).
- (Kem, Chocolate): 7 lần (Dòng 1,5,7,8,11,14,16).
- (Bơ, Chocolate): 7 lần.
- (Bơ, Chuối): 7 lần.
- (Bánh mì, Chuối): 7 lần (Dòng 2,5,6,8,12,14,16).
- (Chocolate, Chuối): 7 lần.

(Các cặp khác đều dưới 7 nên bị loại).

3. Đếm bộ ba (L3)

{Kem, Chocolate, Chuối}: Chỉ xuất hiện 6 lần (Dòng 1,5,8,11,14,16). (LOẠI vì 6 < 7).

- Các bộ ba khác kiểm tra cũng đều dưới 7.

⇒ Kết luận: Thuật toán dừng lại ở tập 2 phần tử. Chúng ta chỉ tìm luật từ các cặp đôi.

Bước 3: Sinh luật và Tính toán (Lift & Leverage)

Điều kiện: MinConf ≥ 0.8. Chúng ta sẽ kiểm tra các cặp tiềm năng ở trên.

1. Cặp {Kem, Chuối} (Support chung = 9)

- Luật: Kem → Chuối
 - Conf = $\frac{Support(Kem, Chuối)}{Support(Kem)} = \frac{9}{10} = 0.9$ (90% > 80% → NHẬN)
- Luật: Chuối → Kem
 - Conf = $\frac{9}{12} = 0.75$ (75% < 80% → LOẠI)

2. Cặp {Bánh mì, Chuối} (Support chung = 7)

- Luật: Bánh mì → Chuối
 - Conf = $\frac{Support(BM, Chuối)}{Support(BM)} = \frac{7}{8} = 0.875$ (87.5% > 80% → NHẬN)
- Luật: Chuối → Bánh mì
 - Conf = $\frac{7}{12} \approx 0.58$ (LOẠI)

(Các cặp còn lại như {Kem, Chocolate}, {Bơ, Chuối}... khi tính thì Confidence đều dưới 0.8.)

Bước 4: Tính Lift và Leverage chi tiết

Bây giờ ta tính các chỉ số nâng cao cho 2 luật vừa tìm được.

A. Luật: Nếu mua KEM thì sẽ mua CHUỐI ((Kem) → {Chuối})

- Sup(L ∪ R) = 9/16 = 0.5625
- Sup(Kem) = 10/16 = 0.625
- Sup(Chuối) = 12/16 = 0.75

1. Lift:

$$Lift = \frac{0.5625}{0.625 \times 0.75} = \frac{0.5625}{0.46875} = 1.2$$

(Lift > 1: Mỗi liên hệ tích cực)

2. Leverage:

$$Leverage = 0.5625 - (0.625 \times 0.75) = 0.5625 - 0.46875 = 0.09375$$

B. Luật: Nếu mua BÁNH MÌ thì sẽ mua CHUỐI ({Bánh mì} → {Chuối})

- Sup(L ∪ R) = 7/16 = 0.4375
- Sup(Bánh mì) = 8/16 = 0.5
- Sup(Chuối) = 12/16 = 0.75

1. Lift:

$$Lift = \frac{0.4375}{0.5 \times 0.75} = \frac{0.4375}{0.375} \approx 1.167$$

2. Leverage:

$$Leverage = 0.4375 - (0.5 \times 0.75) = 0.4375 - 0.375 = 0.0625$$

KẾT QUẢ CUỐI CÙNG (Bảng trình bày)

| STT | Luật Kết Hợp (Rule) | Support(A&B) | Confidence | Lift | Leverage |
|-----|---------------------|--------------|------------|-------|----------|
| 1 | Kem → Chuối | 0.5625 | 90% | 1.2 | 0.09375 |
| 2 | Bánh mì → Chuối | 0.4375 | 87.5% | 1.167 | 0.0625 |

Bài 4.

Bước 1: Thiết lập thông số đầu vào

- Tổng số giao dịch (N): 16 dòng.
- MinSup = 0.3 (30%).
 - Ngưỡng đếm = 16 × 0.3 = 4.8.
 - ⇒ Một tập hợp phải xuất hiện từ 5 lần trở lên mới được giữ lại.
- MinConf = 0.6 (60%).
 - Luật tìm ra phải có độ tin cậy từ 60% trở lên.

Bước 2: Tìm tập phổ biến (Frequent Itemsets)

1. Đếm 1 phần tử (L1)

Ta đếm số lần xuất hiện (đầu X) của từng người:

- Tuấn Trăn (TT): 12 lần → Support = 12/16 = 0.75 (Pass)
- Phùng Anh Đào (PAD): 12 lần → Support = 12/16 = 0.75 (Pass)
- Kiều Minh Tuấn (KMT): 13 lần → Support = 13/16 = 0.8125 (Pass)
- Thủy Ngân (TN): 13 lần → Support = 13/16 = 0.8125 (Pass)

2. Đếm cặp đôi (L2) - Chỉ giữ cặp ≥ 5 lần

Các cặp có tần suất xuất hiện cao nhất:

- (Tuấn Trăn, Thủy Ngân): 10 lần (Dòng 1,2,4,5,6,7,11,12,14,16).
- (P.Anh Đào, Thủy Ngân): 10 lần.
- (Kiều Minh Tuấn, Thủy Ngân): 10 lần.
- (Tuấn Trăn, Kiều Minh Tuấn): 9 lần.
- (P.Anh Đào, Kiều Minh Tuấn): 9 lần.
- (Tuấn Trăn, P.Anh Đào): 8 lần.

(Tất cả các cặp đều thỏa mãn ngưỡng MinSup ≥ 5. Ta chọn 2 cặp có tần suất cao nhất (10 lần) để diện để tính toán chi tiết).

Bước 3: Sinh luật và Tính toán (Lift & Leverage)

Điều kiện: MinConf ≥ 0.6. Xét các cặp tiềm năng nhất có Support chung = 10.

1. Cặp {Tuấn Trăn, Thủy Ngân} (Support chung = 10)

- Luật: Tuấn Trăn → Thủy Ngân
 - Conf = $\frac{Support(TT, TN)}{Support(TT)} = \frac{10}{12} \approx 0.83$ (83% > 60% → NHẬN)
- Luật: Thủy Ngân → Tuấn Trăn
 - Conf = $\frac{10}{13} \approx 0.77$ (77% > 60% → NHẬN)

2. Cặp {Phùng Anh Đào, Thủy Ngân} (Support chung = 10)

- Luật: P.Anh Đào → Thủy Ngân
 - Conf = $\frac{Support(PAD, TN)}{Support(PAD)} = \frac{10}{12} \approx 0.83$ (83% > 60% → NHẬN)
- Luật: Thủy Ngân → P.Anh Đào
 - Conf = $\frac{10}{13} \approx 0.77$ (77% > 60% → NHẬN)

Bước 4: Tính Lift và Leverage chi tiết

Chọn 2 luật xuất có độ tin cậy cao nhất (83%) để tính chỉ số nâng cao.

A. Luật: Nếu có TUẤN TRẦN thì có THỦY NGÂN ((TT) → {TN})

- Sup(L ∪ R) = 10/16 = 0.625
- Sup(Tuấn Trăn) = 12/16 = 0.75
- Sup(Thủy Ngân) = 13/16 = 0.8125

1. Lift:

$$Lift = \frac{0.625}{0.75 \times 0.8125} = \frac{0.625}{0.609375} \approx 1.026$$

(Lift > 1: Có mối liên hệ tích cực nhẹ)

2. Leverage:

$$Leverage = 0.625 - (0.75 \times 0.8125) = 0.625 - 0.609375 = 0.0156$$

B. Luật: Nếu có PANH ĐÀO thì có THỦY NGÂN ((PAD) → {TN})

- Sup(L ∪ R) = 10/16 = 0.625
- Sup(P.Anh Đào) = 12/16 = 0.75
- Sup(Thủy Ngân) = 13/16 = 0.8125

1. Lift: (Thông số Support giống hết trường hợp A)

$$Lift = \frac{0.625}{0.75 \times 0.8125} \approx 1.026$$

2. Leverage:

$$Leverage = 0.0156$$

KẾT QUẢ CUỐI CÙNG (Bảng tóm tắt các luật mạnh nhất)

| STT | Luật Kết Hợp (Rule) | Support(A&B) | Confidence | Lift | Leverage |
|-----|-----------------------|--------------|------------|-------|----------|
| 1 | Tuấn Trăn → Thủy Ngân | 0.625 | 83.3% | 1.026 | 0.0156 |
| 2 | P.Anh Đào → Thủy Ngân | 0.625 | 83.3% | 1.026 | 0.0156 |

| | | | | | |
|---|-----------------------|-------|-------|-------|--------|
| 3 | Thủy Ngân → Tuấn Tiến | 0.625 | 76.9% | 1.026 | 0.0156 |
| 4 | Thủy Ngân → P.Anh Đào | 0.625 | 76.9% | 1.026 | 0.0156 |