

Bước 1: Thiết lập thông số

- Tổng số dòng (N): 16
- MinSup = 0.25 (25%) \Rightarrow Số lần xuất hiện tối thiểu = $16 \times 0.25 = 4$. (Cái nào xuất hiện dưới 4 lần là loại ngay)
- MinConf = 0.75 (75%) \Rightarrow Chỉ nhận 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, 11, 13, 14, 16) \rightarrow Giữ
 - Bơ: 11 lần \rightarrow Giữ
 - Bánh mì: 8 lần \rightarrow Giữ
 - Chuối: 9 lần \rightarrow 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) \rightarrow Giữ
 - (Kem, Bánh mì): 4 lần (Dòng 5, 6, 14, 16) \rightarrow Giữ (Vì đã đếm)
 - (Kem, Chuối): 3 lần (Dòng 3, 6, 9, 13, 16) \rightarrow Giữ
 - (Bơ, Bánh mì): 6 lần (Dòng 2, 5, 8, 10, 12, 14) \rightarrow Giữ
 - (Bơ, Chuối): 4 lần (Dòng 2, 8, 10, 12) \rightarrow Giữ
 - (Bánh mì, Chuối): 3 lần (Dòng 2, 6, 8, 10, 12, 16) \rightarrow Giữ

(Tất cả các cặp đều đạt chuẩn ≥ 1)

3. Đếm bộ ba (3-Itemset)

Giữ các cặp lại và kiểm tra xem dòng nào có 3 cột đều là 1:

- (Kem, Bơ, Bánh mì): Dòng 5, 14 \rightarrow Tổng 2 lần. (Loại vì < 4)
- (Kem, Bơ, Chuối): Không có dòng nào có 3 cột là 1. (Loại)
- (Kem, Bánh mì, Chuối): Dòng 6, 16 \rightarrow Tổng 2 lần. (Loại)
- (Bơ, Bánh mì, Chuối): Dòng 2, 8, 10, 12 \rightarrow Tổng 4 lần. (Giữ)

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

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

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

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

Xét các luật có thể xảy ra:

1. Luật: Nếu mua (Bơ, Chuối) \rightarrow Mua (Bánh mì)

- Support((Bơ, Chuối), Bánh mì) = 4
 - Support((Bơ, Chuối)) = 4 (Lý từ bước 2)
 - Confidence = $4/4 = 1.0 (100\%)$
 - Kết luận: $100\% > 75\% \rightarrow$ LUẬT HỢP LÊ (Rất mạnh)
2. Luật: Nếu mua (Bơ, Bánh mì) \rightarrow Mua (Chuối)
- Support((Bơ, Bánh mì)) = 6
 - Confidence = $6/4 = 0.67 (67\%)$
 - Kết luận: $67\% < 75\% \rightarrow$ LOẠI
3. Luật: Nếu mua (Bánh mì, Chuối) \rightarrow Mua (Bơ)
- Support((Bánh mì, Chuối)) = 6
 - Confidence = $6/4 = 0.67 (67\%)$
 - Kết luận: $67\% < 75\% \rightarrow$ LOẠI

Bài 2.

1. Thiết lập tham số (Config)

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

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

- CSLT (Các số lặp trình): 15 lần \rightarrow Giữ (15 > 10)
- KTLT (Kỹ thuật lặp trình): 13 lần \rightarrow Giữ
- CTDL (Cấu trúc dữ liệu): 12 lần \rightarrow Giữ
- OOP (Lập trình hướng đối tượng): 15 lần \rightarrow Giữ
- LT CSDL (Lý thuyết CSDL): 15 lần \rightarrow Giữ

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

1. (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%).

2. (CSLT, KTLT):

- Đếm được: Dòng 1, 2, 5, 6, 10, 11, 12, 14, 15, 18, 19.
- Tổng: 11 lần.
- Support: $11/20 = 55\%$ (Đạt).

3. (OOP, LT CSDL):

- Đếm được: 10 lần.
- Support: 50% (Vì đã mape Đạt).

4. Bước 3: Sinh luật và Tính Confidence

Luật A: Nếu học CSLT \rightarrow Sẽ học OOP

* Đầu tiên:

- 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: $80\% > 70\% \rightarrow$ Luật này HỢP LÊ.

* Ý nghĩa: Một sinh viên đã đăng ký "Có số lặp trình" thì 80% khả năng họ cũng học "OOP".

Luật B: Nếu học OOP \rightarrow Sẽ học LT CSDL

* Đầu tiên:

- Số sinh viên học cả 2: 12
- Số sinh viên học OOP: 15

* Tính toán:

$$Confidence = \frac{12}{15} \approx 0.733 \quad (73.3\%)$$

* Kết luận: HỢP LÊ.

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

Luật C: Nếu học LT CSDL \rightarrow Sẽ học OOP

* Đầu tiên:

- Support(LT CSDL) = 15.

* Tính toán:

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

* Kết luận: HỢP LÊ.

Bài 3.

Bước 1: Thiết lập thông số đầu vào

* Tổng số giao dịch (N): 16 dòng.

* MinSup = 0.4 (40%):

- Ngưỡng đếm = $16 \times 0.4 = 6.4$.
- \Rightarrow Một cặp hợp phải xuất hiện từ 5 lần trở lên mới được giữ lại (làm tròn lên vì không có 0.4 lần xuất hiện).

* MinConf = 0.8 (80%):

- Lưu ý: \Rightarrow 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)

Tìm số lần xuất hiện (số 1) của từng cột:

- Kem: 10 lần (Dòng 1, 3, 5, 6, 7, 11, 13, 14, 16) \rightarrow Support = 10/16 (Pass)
- Bơ: 11 lần \rightarrow Support = 11/16 (Pass)
- Bánh mì (BM): 8 lần \rightarrow Support = 8/16 (Pass)
- Chocolate: 7 lần \rightarrow Support = 7/16 (Pass)
- Chuối: 12 lần \rightarrow 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, 5, 6, 7, 11, 13, 14, 16).
- (Kem, Chocolate): 7 lần (Dòng 1, 5, 7, 9, 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, 6, 7, 11, 14, 16). (Loại vì $6 < 7$).

* Các bộ ba khác kiểm tra đúng đều dưới 7.

= Kết luận: Thuật toán dừng lại ở step 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 \rightarrow Chuối

* $Conf = \frac{Support(Kem, Chuối)}{Support(Kem)} = \frac{9}{10} = 0.9$ (90% > 80% \rightarrow NHÂN)

* Luật: Chuối \rightarrow Kem

* $Conf = \frac{Support(Kem, Chuối)}{Support(Chuối)} = \frac{9}{12} = 0.75$ (75% < 80% \rightarrow LOẠI)

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

* Luật: Bánh mì \rightarrow Chuối

* $Conf = \frac{Support(BanhMi, Chuoi)}{Support(BanhMi)} = \frac{7}{8} = 0.875$ (87.5% > 80% \rightarrow NHÂN)

* Luật: Chuối \rightarrow Bánh mì

* $Conf = \frac{Support(BanhMi, Chuoi)}{Support(Chuoi)} = \frac{7}{12} \approx 0.58$ (LOẠI)

(Các cặp còn lại như (Kem, Chocolate), (Bơ, Chocolate)... khi tính thi \rightarrow 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.

1. Luật: Nếu mua KEM thì sẽ mua CHUỐI ((Kem) \rightarrow (Chuối))

* Lift:

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

($Lift > 1$: Mỗi lần hệ tích cJC).

2. Leverage:

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

KẾT QUẢ CUỐI CÙNG (Bảng trình bày)

STT	Luật	Kết Hợp (Rule)	Support(A&B)	Confidence	Lift	Leverage
1	Kem \rightarrow Chuối		0.5625	90%	1.2	0.09375
2	Bánh mì \rightarrow 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 = $16 \times 0.3 = 4.8$.
- \Rightarrow Một cặp hợp phải xuất hiện từ 5 lần trở lên mới được giữ lại (làm tròn lên vì không có 0.4 lần xuất hiện).

* MinConf = 0.6 (60%):

- Lưu ý: \Rightarrow 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)

Tìm số lần xuất hiện (số 1) của từng cột:

- Tuấn Trần \rightarrow Thúy Ngân: 10 lần (Dòng 1, 2, 4, 5, 6, 7, 11, 12, 14, 16).
- (Panh Dao, Thuy Ngan): 10 lần.
- (Kieu Minh Tuoi, Thuy Ngan): 10 lần.
- Tuấn Trần, Kieu Minh Tuoi: 9 lần.
- (Panh Dao, Kieu Minh Tuoi): 9 lần.
- Tuấn Trần, Panh Dao: 8 lần.

(Tất cả các cặp đều thỏa mãn $MinSup \geq 5$. Ta chọn 2 cặp có tần suất cao nhất (10 lần) để diễn đế tính toán chi tiết).

3	Thúy Ngân → Tuấn Trần	0.625	76.9%	1.026	0.0156
4	Thúy Ngân → P.Anh Đào	0.625	76.9%	1.026	0.0156