|  |  |
| --- | --- |
| **TID** | **Items** |
| 1 | A,B,C,D,E,F |
| 2 | B,C,D,E,F,G |
| 3 | A,D,E,H |
| 4 | A,D,F,I,J |
| 5 | B,D,E,K |

APRIORI ALGORITHM

Assume Support = 60%

Generate Frequent item sets using Apriori algorithm. Also generate the association rules with confidence = 80%

Solution:

Support = (60/100)\*5 =3

**Step 1: C1 candidate generation (1-item set)**

|  |  |
| --- | --- |
| **Item** | **count** |
| A | 3 |
| B | 3 |
| C | 2 |
| D | 5 |
| E | 4 |
| F | 3 |
| G | 1 |
| H | 1 |
| I | 1 |
| J | 1 |
| K | 1 |
|  |  |

**Step 2: L1 generation**

|  |  |
| --- | --- |
| **Item** | **count** |
| A | 3 |
| B | 3 |
| D | 5 |
| E | 4 |
| F | 3 |

**Step 3: C2 (2 –item set)**

|  |  |
| --- | --- |
| Item | Count |
| A,B | 1 |
| A,D | 3 |
| A,E | 2 |
| A,F | 2 |
| B,D | 3 |
| B,E | 3 |
| B,F | 2 |
| D,E | 4 |
| D,F | 3 |
| E,F | 2 |

**Step 4: L2 (2-item frequent set)**

|  |  |
| --- | --- |
| A,D | 3 |
| B,D | 3 |
| B,E | 3 |
| D,E | 4 |
| D,F | 3 |

**Step 5: C3 (3-item set)**

|  |  |
| --- | --- |
| A,B,D | 1 |
| A,D,E | 2 |
| A,D,F | 2 |
| B,D,E | 3 |
| B,D,F | 2 |
| D,E,F | 2 |

**Step 6: L3 Generation**

|  |  |
| --- | --- |
| B,D,E | 3 |

**Step 7: Frequent item sets: L1 U L2 U L3**

|  |  |
| --- | --- |
| **Frequent Itemset** | **Frequent Patterns (Apriori)** |
| L1  (1-item frequent set) | A:3 |
| B:3 |
| D:5 |
| E:4 |
| F:3 |
| L2 (2-item frequent set) | A,D:3 |
| B,D:3 |
| B,E:3 |
| D,E:4 |
| D,F:3 |
| L3 (3-item frequent set) | B,D,E:3 |

Rules generation and checking its validity when confidence = 80%

A->D confidence= support (AUD)/support (A) D->A = confidence= support (AUD)/support (D)

= 3/3 =1 =100% = 3/5=60%

B->D=100% D->B =60%

B->E = 100% E->B =75%

D->E= 80% E->D =100%

D->F = 60% F->D = 100%

B->DE= confidence=support(BUDUE)/support(B) DE->B confidence= support(BDE)/support(DE)

=3/3=100% =3/4 =75%

D->BE = 3/5=60% BE->D = 3/3 =100%

E->BD= 3/4= 75% BD->E = 3/3 = 100%

BD->E = 3/3 =100% E->BD = ¾ =75%

BE->D =100% D->BE = 3/3 =100%

DE->B = 100% B->DE = 3/3 =100%

FP-Tree (Frequent Pattern) : Divide and Conquer

|  |  |
| --- | --- |
| TID | Items |
| 1 | A,B,C,D,E,F |
| 2 | B,C,D,E,F,G |
| 3 | A,D,E,H |
| 4 | A,D,F,I,J |
| 5 | B,D,E,K |

Support = 60%

Support\_count = 60/100\*5 = 3

Generate Frequent item sets using FP-Tree algorithm. Generate rule with confidence = 80%

1-item set

|  |  |
| --- | --- |
| Item | count |
| A | 3 |
| B | 3 |
| C | 2 |
| D | 5 |
| E | 4 |
| F | 3 |
| G | 1 |
| H | 1 |
| I | 1 |
| J | 1 |
| K | 1 |

1-item frequent set

|  |  |
| --- | --- |
| Item | count |
| A | 3 |
| B | 3 |
| D | 5 |
| E | 4 |
| F | 3 |

Reordering 1-item frequent set in descending order w.r.t count value

|  |  |
| --- | --- |
| Item | count |
| D | 5 |
| E | 4 |
| A | 3 |
| B | 3 |
| F | 3 |

Reordering original item set (Transaction Database) in descending order w.r.t count value

|  |  |
| --- | --- |
| TID | Items |
| 1 | D,E,A,B,F |
| 2 | D,E,B,F |
| 3 | D,E,A |
| 4 | D,A,F |
| 5 | D,E,B |
|  |  |

|  |  |
| --- | --- |
| TID | Items |
| 1 | D,E,A,B,F |
| 2 | D,E,B,F |
| 3 | D,E,A |
| 4 | D,A,F |
| 5 | D,E,B |
|  |  |

A:1

NULL

D:5

B:1

F:1

A:2

E:4 1

|  |  |
| --- | --- |
| Item | count |
| D | 5 |
| E | 4 |
| A | 3 |
| B | 3 |
| F | 3 |

B:2

F:1

F:1

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Conditional Pattern Base | Conditional FP-Tree | Frequent Patterns |
| F | {{D,E,A,B:1},{D,E,B:1},{D,A:1}} | {D:3, E:2, A:2, B:2} | {F:3, FD:3} |
| B | {{D,E,A:1},{D,E:2}} | {D:3, E:3, A:1} | {B:3, BD:3, BE:3, BDE:3} |
| A | {{D,E:2},{D:1}} | {D:3, E:2} | {A:3, AD:3} |
| E | {D:4} | {D:4} | {E:4, ED:4} |
| D | - | - | {D:5} |

Rules generation and checking its validity when confidence = 80%

A->D confidence= support (AUD)/support (A) D->A = confidence= support (AUD)/support (D)

= 3/3 =1 =100% = 3/5=60%

B->D=100% D->B =60%

B->E = 100% E->B =75%

D->E= 80% E->D =100%

D->F = 60% F->D = 100%

B->DE= confidence=support(BUDUE)/support(B) DE->B confidence= support(BDE)/support(DE)

=3/3=100% =3/4 =75%

D->BE = 3/5=60% BE->D = 3/3 =100%

E->BD= 3/4= 75% BD->E = 3/3 = 100%

BD->E = 3/3 =100% E->BD = ¾ =75%

BE->D =100% D->BE = 3/3 =100%

DE->B = 100% B->DE = 3/3 =100%