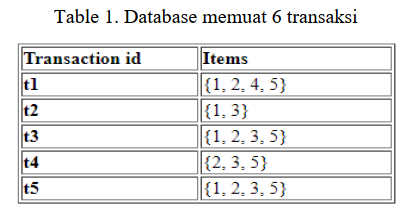
**Tugas DMBI 4**

**FI, AR, dan SP Mining**



1. *Minimum support threshold* = 50%

Kandidat *frequent itemset* dengan panjang 1:

C1 = {{1}, {2}, {3}, {4}, {5}}

*Support* untuk *itemset* dengan panjang 1:

s(1) = 4/5

s(2) = 4/5

s(3) = 4/5

s(4) = 1/5

s(5) = 4/5

*Frequent itemset* dengan panjang 1:

F1 = {{1}, {2}, {3}, {5}}

Kandidat *frequent itemset* dengan panjang 2:

C2 = {{1,2}, {1,3}, {1,5}, {2,3}, {2,5}, {3,5}}

*Support* untuk *itemset* dengan panjang 2:

s({1,2}) = 3/5

s({1,3}) = 3/5

s({1,5}) = 3/5

s({2,3}) = 3/5

s({2,5}) = 4/5

s({3,5}) = 3/5

*Frequent itemset* dengan panjang 2:

F2 = {{1,2}, {1,3}, {1,5}, {2,3}, {2,5}, {3,5}}

1. *Join*:

C3 = {{1,2,3}, {1,2,5}, {1,3,5}, {2,3,5}}

*Pruning*:

Tidak ada kandidat yang dihapus karena tidak ada kandidat yang memiliki subset yang tidak *frequent*. Dengan begitu, kandidat *frequent itemset* dengan panjang 3 tidak berubah setelah proses *join*.

C3 = {{1,2,3}, {1,2,5}, {1,3,5}, {2,3,5}}

1. *Support* untuk *itemset* dengan panjang 3:

s({1,2,3}) = 2/5

s({1,2,5}) = 3/5

s({1,3,5}) = 2/5

s({2,3,5}) = 3/5

*Minimum support threshold* = 50%

*Frequent itemset* dengan panjang 3:

F3 = {{1,2,5}, {2,3,5}}

1. *Frequent 3-itemset* yang dipilih = {1,2,5}

*Minimum confidence* = 80%

*Confidence* dari setiap *rule* yang dihasilkan:

c({1,2} → {5}) = = = 1 (*rule* valid)

c({1,5} → {2}) = = = 1 (*rule* valid)

c({2,5} → {1}) = = = 0,75 (*rule* tidak valid)

Karena *rule* {2,5} → {1} tidak valid:

{2} → {1,5} (*rule* tidak valid)

{5} → {1,2} (*rule* tidak valid)

c({1} → {2,5}) = = = 0,75 (*rule* tidak valid)

*Association rule* yang dihasilkan dari *frequent 3-itemset* {1,2,5}:

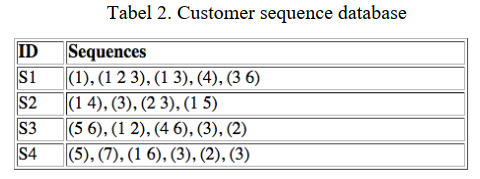
{1,2} → {5}

{1,5} → {2}

|  |  |
| --- | --- |
| *Frequent itemset* | *Support count* |
| {1} | 4 |
| {2} | 4 |
| {3} | 4 |
| {5} | 4 |
| {1,2} | 3 |
| {1,3} | 3 |
| {1,5} | 3 |
| {2,3} | 3 |
| {2,5} | 4 |
| {3,5} | 3 |
| {1,2,5} | 3 |
| {2,3,5} | 3 |

*Frequent closed itemset* = {{1}, {3}, {2,5}}

*Frequent maximal itemset* = {{1,3}, {1,2,5}, {2,3,5}}



1. *Minimum support threshold* = 60%

Kandidat *frequent 1-subsequence*:

C1 = {<(1)>, <(2)>, <(3)>, <(4)>, <(5)>, <(6)>, <(7)>}

*Support* untuk *1-subsequence*:

s(<(1)>) = 4/4

s(<(2)>) = 4/4

s(<(3)>) = 4/4

s(<(4)>) = 3/4

s(<(5)>) = 3/4

s(<(6)>) = 3/4

s(<(7)>) = 1/4

*Frequent 1-subsequence*:

F1 = {<(1)>, <(2)>, <(3)>, <(4)>, <(5)>, <(6)>}

Kandidat *frequent 2-subsequence*:

C2 = {<(1)(2)>, <(2)(1)>, <(1,2)>, <(1)(3)>, <(3)(1)>, <(1,3)>, <(1)(4)>, <(4)(1)>, <(1,4)>, <(1)(5)>, <(5)(1)>, <(1,5)>, <(1)(6)>, <(6)(1)>, <(1,6)>, <(2)(3)>, <(3)(2)>, <(2,3)>, <(2)(4)>, <(4)(2)>, <(2,4)>, <(2)(5)>, <(5)(2)>, <(2,5)>, <(2)(6)>, <(6)(2)>, <(2,6)>, <(3)(4)>, <(4)(3)>, <(3,4)>, <(3)(5)>, <(5)(3)>, <(3,5)>, <(3)(6)>, <(6)(3)>, <(3,6)>, <(4)(5)>, <(5)(4)>, <(4,5)>, <(4)(6)>, <(6)(4)>, <(4,6)>, <(5)(6)>, <(6)(5)>, <(5,6)>}

*Support* untuk *2-subsequence*:

… banyak banget :(

*Frequent 2-subsequence*:

F2 = {<(1)(2)>, <(1)(3)>, <(3)(2)>, <(4)(3)>)

1. *Join*:

C2 = {<(1)(3)(2)>, <(4)(3)(2)>}

*Pruning*:

Kandidat <(4)(3)(2)> terhapus karena memiliki *subsequence* yang *infrequent*, yaitu <(4)(2)>. Dengan begitu, kandidat *frequent 2-subsequence* menjadi:

C2 = {<(1)(3)(2)>}

1. *Support* untuk *3-subsequence*:

s(<(1)(3)(2)>) = 3/4

*Frequent 3-subsequence*:

F3 = {<(1)(3)(2)>}

|  |  |
| --- | --- |
| Frequent subsequence | Support count |
| <(1)> | 4 |
| <(2)> | 4 |
| <(3)> | 4 |
| <(4)> | 3 |
| <(5)> | 3 |
| <(6)> | 3 |
| <(1)(2)> | 4 |
| <(1)(3)> | 4 |
| <(3)(2)> | 3 |
| <(4)(3)> | 3 |
| <(1)(3)(2)> | 3 |

*Frequent closed subsequence* = {<(5)>, <(6)>, <(1)(2)>, <(1)(3)>, <(4)(3)>}

*Frequent maximal subsequence* = {<(5)>, <(6)>, <(4)(3)>}