

Istraživanje podataka - zadaci za vežbe 9

1. Dat je skup podataka

Transaction ID	Items Bought
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, e}
7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, d}

- Nacrtati rešetku skupova stavki koja odgovara datom skupu podataka. Označiti čvorove u rešetki sledećim slovima:
 - N : ako se skup stavki ne smatra kandidatom po Apriori algoritmu. Tj. ako 1) skup stavki nije generisan u koraku generisanja kandidata, ili 2) je generisan u koraku generisanja kandidata ali je kasnije uklonjen u koraku čišćenja kandidata jer neki njegov podskup nije čest.
 - F : kandidat skupa stavki je čest po Apriori algoritmu.
 - I : Ako se skup stavki smatra retkim posle određivanja podrške.
- Koliki je procenat čestih skupova stavki?
- Koliki je odnos čišćenja Apriori algoritma za ovaj skup podataka? Odnos čišćenja je definisan kao procenat skupova stavki koji nisu generisani za vreme generisanja kandidata ili su eliminisani u koraku čišćenja kandidata.
- Koliki je odnos *lažnog alarma* (procenat kandidatskih skupova stavki koji su obeleženi kao retki posle prebrojavanja podrške)?

Praktični zadaci

1. Primenom pravila pridruživanja proveriti da li postoji zavisnosti u tabeli *Adult*. Diskutovati dobijene rezultate. (*IBM IM*)
2. Primenom pravila pridruživanja proveriti da li postoji zavisnost medju upisanim izbornim predmetima. Diskutovati dobijene rezultate. (*IBM IM*)
3. Primenom pravila pridruživanja proveriti da li postoji zavisnost medju podacima u skupu *transactions*. Diskutovati dobijene rezultate. (*KNIME*)