

Istraživanje podataka 1 - Vežbe 13

1 Zadaci

1. Primenom algoritma *Apriori* u alatu IBM SPSS Modeler proveriti da li postoji zavisnost među podacima u skupu u datoteci *transactions.csv*. Skup *transactions* sadrži podatke u transakcionom obliku. Atributi skupa su:

- *transaction* - identifikator transakcije
- *product* - stavka transakcije

Zadatak rešiti korišćenjem podataka u

- transakcionom obliku
- tabelarnom obliku.

Koja pravila pridruživanja su zanimljiva?

2. Primenom algoritma *Apriori* u alatu IBM SPSS Modeler proveriti da li postoji zavisnost među upisanim izbornim predmetima.

Napraviti i model sa pravilima pridruživanja samo za izborne predmete studenata smeru Informatika i Računarstvo i informatika.

Skup u datoteci *izborni_predmeti.xlsx* sadrži podatke o upisanim izbornim predmetima studenata. U jednom redu je indeks studenta, naziv jednog od njegovih izbornih predmeta koje je upisao i naziv smeru koji studira.

Koja pravila pridruživanja su zanimljiva?

3. Primenom pravila pridruživanja u alatu IBM SPSS Modeler proveriti da li postoji zavisnost među upisanim izbornim predmetima studenata smeru Informatika i Računarstvo i informatika. Koristiti čvor *Association Rules*.

Skup u datoteci *izborni_predmeti.xlsx* sadrži podatke o upisanim izbornim predmetima studenata. U jednom redu je indeks studenta, naziv jednog od njegovih izbornih predmeta koje je upisao i naziv smeru koji studira.

Koja pravila pridruživanja su zanimljiva?

4. **Zadatak za vežbu:** Korišćenjem čvora *Association Rules* izdvojiti pravila pridruživanja iz skupa u datoteci *smer_predmet_rok.csv* koji sadrži podatke o uspešnosti polaganja ispita studenata određenog smeru u određenom roku. Atributi su:

- *NAZIV_SMERA* - naziv smeru studenata koji su polagali ispit

- *NAZIV_PREDMETA* - naziv predmeta koji je polagan
- *GODINA_ROKA* - godina u kojoj je održan ispit
- *OZNAKA_ROKA* - oznaka ispitnog roka
- *PROSEK* - prosečna ocena na položenim ispitima
- *BROJ_POLAGANJA* - koliko puta su u proseku studenti koji su izašli na ispit polagali taj predmet

U radnom toku postaviti da važe sledeći uslovi:

- koriste se svi atributi osim *GODINA_ROKA*
- numerički atributi se dele u 4 grupe
- izdvaja se najviše 500 najboljih pravila pridruživanja prema Lift meri
- minimalna podrška pravila je 5%
- minimalna pouzdanost pravila je 60%
- izdvajaju se pravila koja su zanimljiva prema Lift meri

Odgovoriti na pitanja:

- Među izdvojenim pravilima pridruživanja, koje pravilo je najzanimljivije?
- Pronaći najbolje pravilo prema Lift meri za 8. transakciju u skupu kada se stavke koje su u glavi pravila javljaju u transakciji.
- Koja stavka je najčešća u skupu i kolika joj je podrška?