

## Istraživanje podataka - zadatak za vežbe 5

1. Skup podataka *smer\_predmet\_rok.csv* sadrži sledeće podatke: naziv smer, naziv predmeta, godinu ispitnog roka, oznaku ispitnog roka, prosečnu ocenu položenih ispita studenata izdvojenog smer, naziv predmeta u izdvojenom ispitnom roku i prosečan broj polaganja ispita studenata izdvojenog smer, naziv predmeta u izdvojenom ispitnom roku. Koristeći alat KNIME izdvojiti pravila pridruživanja kojima se proverava da li postoji zavisnost između: naziva smer, naziva predmeta, oznake ispitnog roka, prosečne ocene i prosečnog broja polaganja. Prilikom pravljenja modela zadati sledeće uslove:

- najmanja podrška je 5%;
- najmanja pouzdanost je 30%.

Pomoć: Za pravljenje kolekcije stavki (vrednosti datih kolona) možete koristiti *Create Collection Column*.

Odgovoriti na pitanja:

- Koliko je čestih skupova stavki u modelu?
- Kolika je najmanja, a kolika najveća podrška čestih skupova stavki?
- Koji je najveći čest skup stavki?
- Koliko je pravila pridruživanja u modelu?
- Kolika je najmanja, a kolika najveća pouzdanost pravila?
- Koje pravilo je najzanimljivije po Lift meri koje sadrži stavku *Informatika*? Objasniti zašto.

U okviru modela uraditi:

- Napraviti tabelu koja sadrži pravila koja su nabolja prema podršci ili Lift meri. Za svako pravilo izdvojiti: podršku, pouzdanost, Lift meru, telo pravila i glavu pravila.
- Napraviti tabelu koja sadrži pravila koja sadrže stavku *Informatika*.
- U datoteku *pravila.csv* izdvojiti sva dobijena pravila.