

Istraživanje podataka - primer prvog testa, april 2018. g.

Broj indeksa	Ime i prezime

Zadaci se rade ... minuta. PISATI ČITKO - NEČITKI ZADACI NEĆE BITI PREGLEDANI! Maksimalan broj poena je X. Broj poena se izračunava tako što se saberu osvojeni poeni po zadacima, dobijeni zbir podeli sa dva i zaokruži. Broj poena po zadacima je:

Zadatak	1	2	3	4	5	6	7	8	9	10	Zbir	Ukupno
maks	Zbir/2
<i>Osvojeno</i>												

- Za sledeće attribute, navesti koje su vrste: masa tela izražena u kg, veličina garderobe (u oznakama XS, S, M, ...) i naziv pesme. Obrazložiti odgovor.
- Dati su vektori $X = \{0, 1, 1, 0, 0, 1, 1, 1, 0, 1\}$ i $Y = \{1, 1, 0, 0, 1, 1, 1, 0, 0, 0\}$. Primeniti dve mere sličnosti za binarne podatke.
- Kako se statistički računaju elementi van granica za numeričke attribute pri njihovom predstavljanju pomoću kućica (*box plot*). Navesti formulu i dati primer kućice sa elementima van granica.
- Dat je skup podataka o izbornim predmetima. Atributi su: indeks studenta, naziv izbornog predmeta i naziv smera koji student studira. Svaki red u skupu podataka sadrži podatak o jednom izbornom predmetu studenta i naziv smera koji studira. Student može izabrati najviše 5 izbornih predmeta. Ukupan broj različitih izbornih predmeta je 100. Pomoću pravila pridruživanja traže se veze između izbornih predmeta koje su studenti izabrali i smera koji studiraju.
 - Kako biste dati skup transformisali u tabelarni oblik za primenu pravila pridruživanja?
 - Šta je jedna transakcija u ovom skupu?
 - Kog tipa je dobijeni skup podataka nakon transformacije.

INDEKS	NAZIV PREDMETA	NAZIV SMERA
20060009	Taja A	Racunarstvo i informatika
20060009	Taja B	Racunarstvo i informatika
20060009	Programske paradigme	Racunarstvo i informatika
...
20060011	Teorija izracunljivosti	Informatika
20060011	Programske paradigme	Informatika

- Za dati skup transakcija izdvojiti maksimalno česte skupove stavki. Za zahtevanu podršku uzeti vrednost 1/3.
 - Izračunati pouzdanost i Lift za sledeća pravila : $\{A, D\} \rightarrow \{C\}$ ili $\{C, D\} \rightarrow \{A\}$. Koje pravilo je zanimljivo/zanimljivije prema Lift meri? Obrazložiti odgovor.

1	$\{A, B\}$
2	$\{A, B, C, D\}$
3	$\{A\}$
4	$\{B\}$
5	$\{A, C, D\}$
6	$\{A, B, C, D\}$
7	$\{A, B\}$
8	$\{C, D\}$
9	$\{A, C, D\}$