Pitanje 1

Točno

Broj bodova: 1,50 od 1,50

Pretpostavite da se koristi algoritam sažimanja osjetljivog na bliskost (engl. *Locality Sensitive Hashing*). Ukupna duljina sažetaka jest 256 elemenata te se pritom koristi 16 pojaseva svaki duljine 16 elemenata. Što će se dogoditi ako se broj pojaseva smanji na 8, a duljina jednog pojasa smanji na 32 elemenata.

Odaberite jedan odgovor:

- a. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost neće se mijenjati.
- b. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se povećati.
- c. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se smanjiti.
- d. Broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost povećati.
- e. Broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost povećati. ✓
- f. Odznači odgovor (ne želim odgovoriti)

Your answer is correct.

Ispravan odgovor je: Broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost povećati.

Pitanje 2

Točno

Broj bodova: 2,00 od 2,00

U tablici ispod zadana je matrica ocjena korisnika za pojedine filmove (engl. *user-item matrix, utility matrix*). Prazna polja u matrici predstavljaju ocjene koje nedostaju. Korištenjem algoritama suradničkog filtriranja (engl. *Collaborative Filtering*) potrebno je izračunati ocjenu za korisnika $\mathbf{U_4}$ i film $\mathbf{M_2}$ ako se koristi *Item-Item* pristup suradničkog filtriranja.

Bitne napomene:

- Za računanje sličnosti među filmovima koristi se *Pearson Correlation Coefficient.*
- Sustav koristi najviše **k = 2** najsličnijih filmova za izračun ocjene.
- Prilikom izračuna, filmovi čija sličnost je manja od 0 se ne uzimaju u obzir.

	$\mathbf{U_1}$	$\mathbf{U_2}$	$\mathbf{U_3}$	$\mathbf{U_4}$	$\mathbf{U_5}$
$\mathbf{M_1}$	2	3	4	1	3
$\mathbf{M_2}$	-	3	3	-	4
$\mathbf{M_3}$	-	2	3	1	5
$\mathbf{M_4}$	1	2	5	2	3
$\mathbf{M_5}$	2	5	4	-	5

Vaš odgovor zaokružite na 2 decimale!

Odaberite jedan odgovor:

a.
3.63

b. Odznači odgovor (ne želim odgovoriti)

c.
1.00 ✓

d.
2.50

e.
1.43

f.
3.00

Your answer is correct.

Ispravan odgovor je:
1.00

Pitanje 3

Točno

Broj bodova: 1,50 od 1,50

Neka je zadana sljedeća "*Utiliity*" matrica u kojoj su znakom "-" označene ocjene koje nedostaju:

	I_1	I_2	I_3	I_4
U_1	-	-	5	1
U_2	-	4	4	1

Izračunajte sličnost između korisnika U_1 i U_2 korištenjem *Pearson Correlation Coefficient* (PCC) mjere sličnosti.

Napomena: *Odgovor zaokružite na 2 decimale*!

Odaberite jedan odgovor:

- a. -0.44
- b. 0.5
- c. 0.12
- d. 0.72
- e. 0.87 ✓
- f. Odznači odgovor (ne želim odgovoriti)

Your answer is correct.

Ispravan odgovor je: 0.87

Pitanje 4

Točno

Broj bodova: 1,50 od 1,50

U postupku generiranja pravila asocijacije $I \rightarrow j$ nad skupom predmeta I parametar pouzdanost (eng. confidence) definira se kao (potpora - eng. support):

Odaberite jedan odgovor:

- a. conf($I \rightarrow j$) = support(j) / support($I \cup j$)
- b. conf($I \rightarrow j$) = support(j) / support(I)
- c. Odznači odgovor (ne želim odgovoriti)
- d. conf($I \rightarrow j$) = support(I) / support(j)
- e. conf($I \rightarrow j$) = support($I \cup j$) / support(I) ✓
- f. conf($I \rightarrow j$) = support(I) / support($I \cup j$)

Your answer is correct.

Ispravan odgovor je: conf($I \rightarrow j$) = support($I \cup j$) / support(I)

Pitanje 5

Netočno

Broj bodova: -0,30 od 1,50

Promatramo računanje Rabinovog sažetka nad n-gram prozorom proizvoljnog teksta koristeći kodiranje polinomima. Ako je o_i broj operacija potrebnih za računanje sažetka i-tog n-gram prozora (i je zadan u [1, n]). Onda vrijedi:

Odaberite jedan odgovor:

- a. $o_i > o_{i+1}$, za svaki i >= 1
- b. Odznači odgovor (ne želim odgovoriti)
- c. $o_1 < o_i$, za svaki i > 1
- d. $o_i = o_{i+1}$, za svaki i >= 1 ✗
- e. $o_1 > o_i$, za svaki i > 1
- f. $o_i < o_{i+1}$, za svaki i >= 1

Your answer is incorrect.

Ispravan odgovor je: $o_1 > o_i$, za svaki i > 1

Pitanje 6

Točno

Broj bodova: 2,00 od 2,00

Neka se u skupu podataka nalaze predmeti A, B, C, D i E.

Pritom, neka su sljedeći podskupovi podataka česti na odabranom uzorku:

{B}, **{C}**, **{D}**, **{B, C}**, **{C, D}**

Negativna granica Toivonenog algoritma za zadani primjer je:

Odaberite jedan odgovor:

- a. Odznači odgovor (ne želim odgovoriti)
- b. {E}, {A, B}, {A, C}, {B, D}
- c. {A}, {E}, {B, D} ✓
- d. {A}, {B, D}
- e. {E}, {B, D}
- f. {A}, {B, E}, {B, C, D}

Your answer is correct.

Ispravan odgovor je: {A}, {E}, {B, D}

Pitanje 7

Točno

Broj bodova: 1,50 od 1,50

Prilikom stvaranja skupovnih reprezentacija tekstualnih dokumenata za potrebe algoritma MinHash, označite tvrdnju koja vrijedi kada se razmatraju dulji tekstualni dokumenti.

Odaberite jedan odgovor:

- a. Uputno je koristiti kraće *shingleove*
- b. Ništa od navedenoga ne stoji.
- c. Uputno je koristiti duže *shingleove* ✓
- d. Veličina *shingleova* nije bitan parametar za algoritam MinHash.
- e. Odznači odgovor (ne želim odgovoriti)

Your answer is correct.

Ispravan odgovor je: Uputno je koristiti duže *shingleove*

Pitanje 8

Točno

Broj bodova: 2,00 od 2,00

Pretpostavite da su provođenjem algoritma MinHash nad izvornim skupovnim reprezentacijama dobiveni sljedeći sažeci:

	s0	**s1**	**s2**	**s3**	**s4**
$\mathbf{b_0}$	2	5	4	7	3
$\mathbf{b_1}$	7	6	2	8	5
$\mathbf{b_2}$	4	3	3	5	6
$\mathbf{b_3}$	4	3	2	4	2

Nadalje, pretpostavite da se koristi algoritam sažimanja osjetljivog na bliskost (engl. Locality Sensitive Hashing) s veličinom pojasa b=2.

Unatar prvog pojasa koristi se sljedeća funkcija sažimanja:

$f_1 = (b_0 * 10 + b_1) \bmod 6$,

Unutar drugog pojasa koristi se sljedeća funkcija sažimanja:

$f_2 = (b_2 * 10 + b_3) \bmod 6$.

Npr. sažetak $\mathbf{s_0}$ u prvom pojasu se raspršti u pretinac p=2:

$p = (2 * 10 + 7) \% 6 = 2$.

Koliko će biti parova kandidata za sličnost kada se završi algoritam za oba pojasa?

Odaberite jedan odgovor:

- a. 3
- b. 4 ✓
- c. Odznači odgovor (ne želim odgovoriti)
- d. 5
- e. 2
- f. 1

Your answer is correct.

Ispravan odgovor je: 4

Pitanje 9

Točno

Broj bodova: 2,00 od 2,00

U točki $C_{x,y}$ = (1, 2) zadan je centroid oko kojeg su podaci normalno distribuirani. Odredite Mahalanobisovu udaljenost točke $T_{x,y}$= (3, 4) od centroida, ako varijance po dimenzijama iznose: v_x = 0.01 i v_y = 0.02.

Odaberite jedan odgovor:

- a. 24.49 ✓
- b. 223.6
- c. Odznači odgovor (ne želim odgovoriti)
- d. 14.14
- e. 42.42
- f. 8.26

Your answer is correct.

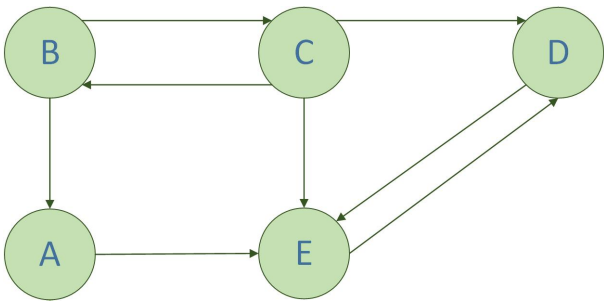
Ispravan odgovor je: 24.49

Pitanje 10

Točno

Broj bodova: 1,50 od 1,50

Zadan je graf na slici za koji je potrebno izračunati vektor utjecaja (engl. *rank vector*).



Primjeni li se metoda uzastopnog potenciranja s ciljem izračunavanja vektora ranga *r* na zadani graf, hoće li dobiveni rezultati biti vjerodostojni?

Odaberite jedan odgovor:

- a. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*). ✓
- b. Odznači odgovor (ne želim odgovoriti)
- c. Da, rezultati će biti vjerodostojni.
- d. Ne, u grafu postoje ciklusi.
- e. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*), ali i "mrtvi čvorovi" (engl. *dead end nodes*).
- f. Ne, u grafu postoje "mrtvi čvorovi" (engl. *dead end nodes*).

Your answer is correct.

Ispravan odgovor je: Ne, u grafu postoji "paukova zamka" (engl. *spider trap*).

Pitanje 11

Točno

Broj bodova: 1,50 od 1,50

Neka je zadana sljedeća "*Utiliity*" matrica u kojoj su znakom "-" označene ocjene koje nedostaju:

	I_1	I_2	I_3	I_4
U_1	-	-	5	1
U_2	-	4	4	1

Izračunajte sličnost između korisnika U_1 i U_2 korištenjem *Cosine similarity* mjere sličnosti.

Napomena: *Odgovor zaokružite na 2 decimale*!

Odaberite jedan odgovor:

- a. 0.12
- b. 0.87
- c. 0.72 ✓
- d. 0.5
- e. Odznači odgovor (ne želim odgovoriti)
- f. -0.44

Your answer is correct.

Ispravan odgovor je: 0.72

Pitanje 12

Točno

Broj bodova: 1,50 od 1,50

Kod računanja skupa sličnih sažetaka koristeći permutacijske tablice, sustav koristi 64-bitne sažetke i parametar p u svim tablicama je jednak 16. U sustavu se nalazi ukupno 2^32 sažetaka. Prosječni broj sažetaka za koji sustav računa Hammingovu udaljenost, za svaki upit, jest:

Odaberite jedan odgovor:

- a. 2^{64}
- b. 2^{32}
- c. 2^{16} ✓
- d. Odznači odgovor (ne želim odgovoriti)
- e. 2^{48}
- f. 2^{8}

Your answer is correct.

Ispravan odgovor je: 2^{16}

Pitanje 13

Netočno

Broj bodova: -0,40 od 2,00

Algoritam simhash u svojoj kanonskoj inačici koristi 3-bitne sažetke, a jedinke su riječi (odvojene razmacima). Interna funkcija sažimanja je definirana s **h(x) = duljina(x) % 8**, gdje je x riječ, a duljina(x) broj znakova riječi. Decimalni simhash sažetak teksta "**with or without you**" jest:

Odaberite jedan odgovor:

- a. 1
- b. 6
- c. Odznači odgovor (ne želim odgovoriti)
- d. 5
- e. 7 ✗
- f. 2
- g. 3
- h. 4

Your answer is incorrect.

Ispravan odgovor je: 6

Pitanje 14

Točno

Broj bodova: 1,50 od 1,50

Korištenje mape (engl. hash table) učinkovitije je za pohranu brojača u algoritmu A-Priori ako se u skupu podataka pojavljuje manje od *X* od ukupnog mogućeg broja različitih podskupova podataka.

Odaberite jedan odgovor:

- a. X = 1/2
- b. X = 3/5
- c. X = 2/3
- d. X = 3/4
- e. Odznači odgovor (ne želim odgovoriti)
- f. X = 1/3 ✓

Your answer is correct.

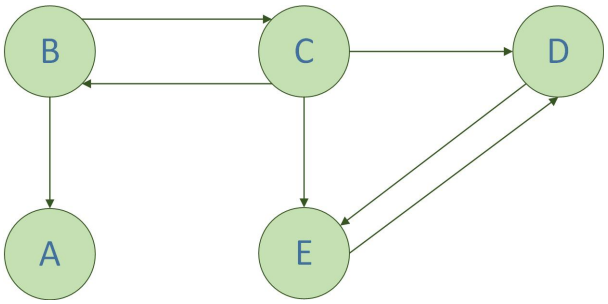
Ispravan odgovor je: X = 1/3

Pitanje 15

Točno

Broj bodova: 1,50 od 1,50

Zadan je graf na slici za koji je potrebno izračunati vektor utjecaja (engl. *rank vector*).



Primjeni li se metoda uzastopnog potenciranja s ciljem izračunavanja vektora ranga *r* na zadani graf, hoće li dobiveni rezultati biti vjerodostojni?

Odaberite jedan odgovor:

- a. Odznači odgovor (ne želim odgovoriti)
- b. Ne, u grafu postoje ciklusi.
- c. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*), ali i "mrtvi čvorovi" (engl. *dead end nodes*). ✓
- d. Da, rezultati će biti vjerodostojni.
- e. Ne, u grafu postoje "mrtvi čvorovi" (engl. *dead end nodes*).
- f. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*).

Your answer is correct.

Ispravan odgovor je: Ne, u grafu postoji "paukova zamka" (engl. *spider trap*), ali i "mrtvi čvorovi" (engl. *dead end nodes*).

Pitanje 16

Točno

Broj bodova: 2,00 od 2,00

Neka su zadana dva dokumenta D_1="ABECEDA" i D_2="CEDAR". Izračunajte sažetke dokumenata koristeći *MinHash* algoritam uz duljinu shingleova L=3, koristeći dvije funkcije sažimanja f_1 i f_2 umjesto permutacija prema zadanoj tablici.

r	shingle	$f_1 = (r + 1) \bmod 6$	$f_2 = (r + 2) \bmod 6$
0	ABE	1	2
1	BEC	2	3
2	ECE	3	4
3	CED	4	5
4	EDA	5	0
5	DAR	0	1

Izračunajte sličnost dobivenih sažetaka kao omjer**:** *broj redaka u kojima su sažeci jednaki* i *ukupni broj redaka u sažecima!*

Odaberite jedan odgovor:

- a. 1.0
- b. 0.0
- c. 0.5 ✓
- d. Odznači odgovor (ne želim odgovoriti)
- e. 0.25
- f. 0.75

Your answer is correct.

Ispravan odgovor je:
0.5

Pitanje 17

Točno

Broj bodova: 1,50 od 1,50

Algoritam grupiranja CURE rješenje nalazi u N prolaza po skupu podataka pri čemu je:

Odaberite jedan odgovor:

- a. N = 5
- b. Odznači odgovor (ne želim odgovoriti)
- c. N = 4
- d. N = 2 ✓
- e. N = 3
- f. N = 1

Your answer is correct.

Ispravan odgovor je: N = 2

Pitanje 18

Točno

Broj bodova: 1,50 od 1,50

Kako bi se mogao primijeniti algoritam grupiranja BFR podaci moraju biti:

Odaberite jedan odgovor:

- a. Normalno distribuirani oko centroida, a dimenzije prostora moraju biti zavisne
- b. Normalno distribuirani oko centroida, a dimenzije prostora nezavisne ✓
- c. Odznači odgovor (ne želim odgovoriti)
- d. Najviše udaljeni za $\sqrt{d}$
- e. Normalno distribuirani oko centroida, bez ograničenja na dimenzije prostora
- f. Distribucija podataka nije važna.

Your answer is correct.

Ispravan odgovor je: Normalno distribuirani oko centroida, a dimenzije prostora nezavisne

Pitanje 1

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Neka je zadana sljedeća *"Utility"* matrica u kojoj su znakom "-" označene ocjene koje nedostaju:

	I_1	I_2	I_3	I_4
U_1	-	-	5	1
U_2	-	4	4	1

Izračunajte sličnost između korisnika U_1 i U_2 korištenjem *Cosine similarity* mjere sličnosti.

Napomena: *Odgovor zaokružite na 2 decimale!*

Odaberite jedan odgovor:

- a. 0.5
- b. 0.12
- c. -0.44
- d. 0.72 ✓
- e. Odznači odgovor (ne želim odgovoriti)
- f. 0.87

Your answer is correct.

Ispravan odgovor je: 0.72

Pitanje 2

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Ako su podskupovi predmeta {a, b}, {b, d} i {c, d} česti, onda je sigurno čest i skup:

Odaberite jedan odgovor:

- a. {a, c}
- b. {a, b, c}
- c. Odznači odgovor (ne želim odgovoriti)
- d. {a} ✓
- e. {a, b, c, d}
- f. {b, c}

Your answer is correct.

Ispravan odgovor je: {a}

Pitanje 3

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Složenost algoritma hijerarhijskog grupiranja iznosi:

Odaberite jedan odgovor:

- a. O(n)
- b. O(n^2 log(n)) ✓
- c. Odznači odgovor (ne želim odgovoriti)
- d. O(n^2)
- e. O(n log(n))
- f. O(log(n))

Your answer is correct.

Ispravan odgovor je: O(n^2 log(n))

Pitanje 4

Točno

Broj bodova: 2,00 od 2,00

Označi pitanje

Pretpostavite da su provođenjem algoritma MinHash nad izvornim skupovnim reprezentacijama dobiveni sljedeći sažeci:

	s0	s1	s2	s3	s4
$\mathbf{b}_0$	2	5	4	7	3
$\mathbf{b}_1$	7	6	2	8	5
$\mathbf{b}_2$	4	3	3	5	6
$\mathbf{b}_3$	4	3	2	4	2

Nadalje, pretpostavite da se koristi algoritam sažimanja osjetljivog na bliskost (engl. Locality Sensitive Hashing) s veličinom pojasa b=2.

Unatar prvog pojasa koristi se sljedeća funkcija sažimanja:

$f_1 = (b_0 * 10 + b_1) \bmod 6$,

Unutar drugog pojasa koristi se sljedeća funkcija sažimanja:

$f_2 = (b_2 * 10 + b_3) \bmod 6$.

Npr. sažetak $\mathbf{s}_0$ u prvom pojasu se raspršuje u pretinac p=2:

$p = (2 * 10 + 7) \% 6 = 2$.

Koliko će biti parova kandidata za sličnost kada se završi algoritam za oba pojasa?

Odaberite jedan odgovor:

- a. 3
- b. 4 ✓
- c. 5
- d. 2
- e. Odznači odgovor (ne želim odgovoriti)
- f. 1

Your answer is correct.

Ispravan odgovor je: 4

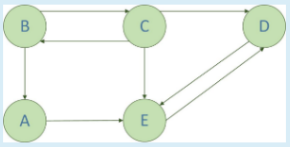
Pitanje 5

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Zadan je graf na slici za koji je potrebno izračunati vektor utjecaja (engl. *rank vector*).



Primjeni li se metoda uzastopnog potenciranja s ciljem izračunavanja vektora ranga *r* na zadani graf, hoće li dobiveni rezultati biti vjerodostojni?

Odaberite jedan odgovor:

- a. Odznači odgovor (ne želim odgovoriti)
- b. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*). ✓
- c. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*), ali i "mrtvi čvorovi" (engl. *dead end nodes*).
- d. Ne, u grafu postoje "mrtvi čvorovi" (engl. *dead end nodes*).
- e. Da, rezultati će biti vjerodostojni.
- f. Ne, u grafu postoje ciklusi.

Your answer is correct.

Ispravan odgovor je: Ne, u grafu postoji "paukova zamka" (engl. *spider trap*).

Pitanje 6

Točno

Broj bodova: 2,00 od 2,00

Označi pitanje

U točki $C_{x,y} = (1, 2)$ zadan je centroid oko kojeg su podaci normalno distribuirani. Odredite Mahalanobisovu udaljenost točke $T_{x,y} = (3, 4)$ od centroida, ako varijance po dimenzijama iznose: $v_x = 0.01$ i $v_y = 0.02$.

Odaberite jedan odgovor:

- a. Odznači odgovor (ne želim odgovoriti)
- b. 24.49 ✓
- c. 14.14
- d. 8.26
- e. 223.6
- f. 42.42

Your answer is correct.

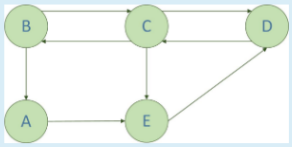
Pitanje 7

Točno

Broj bodova: 2,00 od 2,00

Označi pitanje

Za zadani graf na slici metodom uzastopnog potenciranja izračunati vrijednost rang vektora *r* (engl. *rank vector*) za prve tri iteracije algoritma. Prilikom računanja potrebno uključiti teleportirajuće poveznice (engl. *teleports*) u graf te pritom vjerojatnost da će slučajni šetač (engl. *random walker*) slijediti poveznicu iz originalnog grafa iznosi β = 0.8. Rang vektor u početku (u 0-toj iteraciji) treba inicijalizirati tako da svi čvorovi imaju jednak rang/utjecaj.



Napišite vrijednost utjecaja/ranga čvora **A** nakon **3. iteracije.**

Napomena: *Vrijednost dobivenu nakon 3. iteracije zaokružite na 3. decimale.*

Odaberite jedan odgovor:

- a. 0.340
- b. 0.283
- c. 0.086 ✓
- d. 0.177
- e. Odznači odgovor (ne želim odgovoriti)
- f. 0.114

Pitanje 8

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Kod računanja skupa sličnih sažetaka koristeći permutacijske tablice, sustav koristi ***m*** jedinica memorije te odgovara na upit za ***s*** sekundi. Broj permutacijskih tablica ***t*** je (znak ~ označava proporcionalnost):

Odaberite jedan odgovor:

- a. ~m i ~s
- b. Odznači odgovor (ne želim odgovoriti)
- c. ~m i ~1/s ✓
- d. ne utječe na m ili s
- e. ~1/m i ne utječe na s
- f. ~m i ne utječe na s

Your answer is correct.

Ispravan odgovor je: ~m i ~1/s

Pitanje 9

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Algoritam grupiranja CURE rješenje nalazi u N prolaza po skupu podataka pri čemu je:

Odaberite jedan odgovor:

- a. N = 3
- b. N = 1
- c. N = 2 ✓
- d. N = 5
- e. N = 4
- f. Odznači odgovor (ne želim odgovoriti)

Your answer is correct.

Ispravan odgovor je: N = 2

Pitanje 10

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Neka je zadana sljedeća "*Utility*" matrica u kojoj su znakom "-" označene ocjene koje nedostaju:

	I_1	I_2	I_3	I_4
U_1	-	-	5	1
U_2	-	4	4	1

Izračunajte sličnost između korisnika U_1 i U_2 korištenjem *Pearson Correlation Coefficient* (PCC) mjere sličnosti.

Napomena: *Odgovor zaokružite na 2 decimale!*

Odaberite jedan odgovor:

- a. -0.44
- b. 0.87 ✓
- c. 0.5
- d. Odznači odgovor (ne želim odgovoriti)
- e. 0.72
- f. 0.12

Pitanje 11

Točno

Broj bodova: 2,00 od 2,00

Označi pitanje

Neka je zadan skup košara:

$B_1 = \{1, 2, 3\}$ $B_4 = \{1, 3, 4\}$
$B_2 = \{1, 4\}$ $B_5 = \{4\}$
$B_3 = \{1, 3, 4\}$ $B_6 = \{2, 3, 4\}$

Pod pretpostavkom da prag potpore (engl. support threshold) iznosi 3, koliko podskupova podataka je često?

Odaberite jedan odgovor:

- a. 6 ✓
- b. 3
- c. 2
- d. 4
- e. 5
- f. Odznači odgovor (ne želim odgovoriti)

Pitanje 12

Točno

Broj bodova: 2,00 od 2,00

Označi pitanje

U tablici ispod zadana je matrica ocjena korisnika za pojedine filmove (engl. *user-item matrix, utility matrix*). Prazna polja u matrici predstavljaju ocjene koje nedostaju. Korištenjem algoritama suradničkog filtriranja (engl. *Collaborative Filtering*) potrebno je izračunati ocjenu za korisnika $\mathbf{U_4}$ i film $\mathbf{M_2}$ ako se koristi *Item-Item* pristup suradničkog filtriranja.

Bitne napomene:

- Za računanje sličnosti među filmovima koristi se *Pearson Correlation Coefficient*.
- Sustav koristi najviše **k = 2** najsličnijih filmova za izračun ocjene.
- Prilikom izračuna, filmovi čija sličnost je manja od 0 se ne uzimaju u obzir.

	U_1	U_2	U_3	U_4	U_5
M_1	2	3	4	1	3
M_2	-	3	3	-	4
M_3	-	2	3	1	5
M_4	1	2	5	2	3
M_5	2	5	4	-	5

Vaš odgovor zaokružite na 2 decimale!

Odaberite jedan odgovor:

- a. 1.43
- b. 3.63
- c. 3.00
- d. Odznači odgovor (ne želim odgovoriti)
- e. 1.00 ✓
- f. 2.50

Pitanje 13

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

U postupku generiranja pravila asocijacije $I \rightarrow j$ nad skupom predmeta I parametar pouzdanost (eng. confidence) definira se kao (potpora - eng. support):

Odaberite jedan odgovor:

- a. $\text{conf}(I \rightarrow j) = \text{support}(I) / \text{support}(I \cup j)$
- b. $\text{conf}(I \rightarrow j) = \text{support}(j) / \text{support}(I)$
- c. Odznači odgovor (ne želim odgovoriti)
- d. $\text{conf}(I \rightarrow j) = \text{support}(I) / \text{support}(j)$
- e. $\text{conf}(I \rightarrow j) = \text{support}(j) / \text{support}(I \cup j)$
- f. $\text{conf}(I \rightarrow j) = \text{support}(I \cup j) / \text{support}(I)$ ✓

Pitanje 14

Netočno

Broj bodova: -0,30 od 1,50

Označi pitanje

Pretpostavite da se koristi algoritam sažimanja osjetljivog na bliskost (engl. *Locality Sensitive Hashing*). Ukupna duljina sažetaka jest 256 elemenata te se pritom koristi 16 pojaseva svaki duljine 16 elemenata. Što će se dogoditi ako se broj pojaseva poveća na 32, a duljina jednog pojasa smanji na 8 elemenata.

Odaberite jedan odgovor:

- a. Broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost povećati.
- b. Broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost povećati. ✗
- c. Odznači odgovor (ne želim odgovoriti)
- d. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se smanjiti.
- e. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost će se povećati.
- f. Broj lažno negativnih (engl. *false negative*) i broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost neće se mijenjati.

Your answer is incorrect.

Ispravan odgovor je: Broj lažno negativnih (engl. *false negative*) parova kandidata za sličnost će se smanjiti, dok će se broj lažno pozitivnih (engl. *false positive*) parova kandidata za sličnost povećati.

Pitanje 15

Netočno

Broj bodova: -0,30 od 1,50

Označi pitanje

Promatramo računanje Rabinovog sažetka nad n-gram prozorom proizvoljnog teksta koristeći kodiranje polinomima. Ako je o_i broj operacija potrebnih za računanje sažetka i-tog n-gram prozora (i je zadan u [1, n]). Onda vrijedi:

Odaberite jedan odgovor:

- a. $o_i = o_{i+1}$, za svaki i >= 1 ✗
- b. Odznači odgovor (ne želim odgovoriti)
- c. $o_1 > o_i$, za svaki i > 1
- d. $o_1 < o_i$, za svaki i > 1
- e. $o_i > o_{i+1}$, za svaki i >= 1
- f. $o_i < o_{i+1}$, za svaki i >= 1

Your answer is incorrect.

Ispravan odgovor je: $o_1 > o_i$, za svaki i > 1

Pitanje 16

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Prilikom stvaranja skupovnih reprezentacija tekstualnih dokumenata za potrebe algoritma MinHash, označite tvrdnju koja vrijedi kada se razmatraju dulji tekstualni dokumenti.

Odaberite jedan odgovor:

- a. Uputno je koristiti kraće *shingleove*
- b. Odznači odgovor (ne želim odgovoriti)
- c. Uputno je koristiti duže *shingleove* ✓
- d. Ništa od navedenoga ne stoji.
- e. Veličina *shingleova* nije bitan parametar za algoritam MinHash.

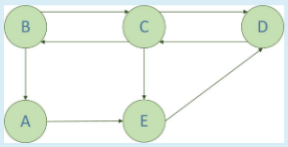
Pitanje 17

Točno

Broj bodova: 1,50 od 1,50

Označi pitanje

Zadan je graf na slici za koji je potrebno izračunati vektor utjecaja (engl. *rank vector*).



Primjeni li se metoda uzastopnog potenciranja s ciljem izračunavanja vektora ranga *r* na zadani graf, hoće li dobiveni rezultati biti vjerodostojni?

Odaberite jedan odgovor:

- a. Odznači odgovor (ne želim odgovoriti)
- b. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*).
- c. Ne, u grafu postoje ciklusi.
- d. Da, rezultati će biti vjerodostojni. ✓
- e. Ne, u grafu postoje "mrtvi čvorovi" (engl. *dead end nodes*).
- f. Ne, u grafu postoji "paukova zamka" (engl. *spider trap*), ali i "mrtvi čvorovi" (engl. *dead end nodes*).

Pitanje **18**

Netočno

Broj bodova: 0,00 od 2,00

Označi pitanje

Ulaz u 3-bitni algoritam Simhash je niz od *jedne jedinice, dvije dvojke, ..., pet petica*; dakle, 122 ... 55555. Prije računanja Simhash sažetka, nad ulaznim nizom se provede TF-IDF pred-procesiranje te je ulaz u algoritam Simhash zapravo matrica "pojavljivanja" jedinki ulaznog niza (budući da imamo samo jedan ulazni niz, matrica ima jedan redak). Dodatno, koristi se inačica algoritma Simhash s težinskim inkrementiranjem koristeći upravo izračunate TF-IDF koeficijente za povećanje pozicije internog vektora (napomena: kanonska inačica algoritma obavlja pred-procesiranje *implicitno* i uvijek koristi +/-1 koeficijente). Ako se koristi sljedeća formula za izračun težinskih faktora:

$\text{TF-IDF} = 1 + \log_{10}(tf)$

gdje je *tf* broj pojavljivanja promatrane jedinke u ulaznom nizu, onda je (dekadski) Simhash sažetak ulaznog niza jednak:

Odaberite jedan odgovor:

- a. 5
- b. 2
- c. Odznači odgovor (ne želim odgovoriti) ✗
- d. 4
- e. 3
- f. 1

Your answer is incorrect.

Ispravan odgovor je: 1