

# Razvoj softvera II

## Sistemi preporuke



# Sadržaj



- Kupci i proizvodi
- Kreiranje ponude
- Sistem preporuke
- Content-based filtering
- Collaborative filtering
- Association rules learning



# Uvod

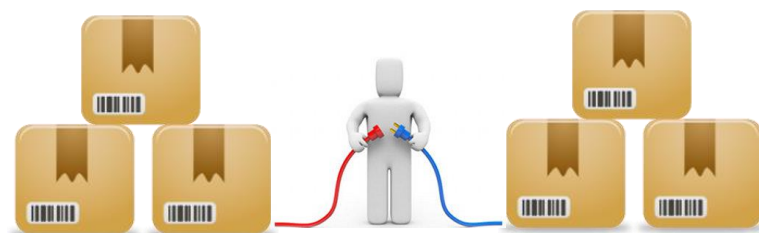
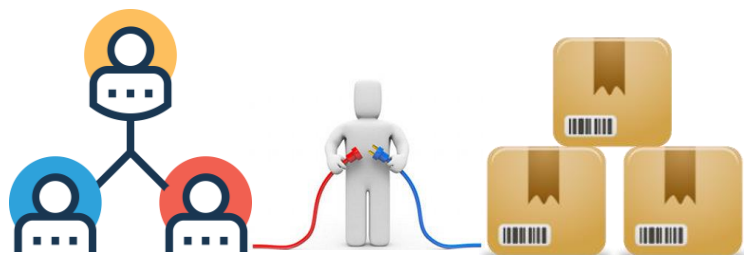
- Online prodaja je proizvode učinila dostupnim milionima kupaca
- Svi kupci su unikatni, te posjeduju jasne preference i očekivanja od proizvoda koje pretražuju
- Pogled kupca – prevelik broj proizvoda, te nedostatak vremena i strpljenja za njihovu pretragu i analizu
- Pogled kompanija koji nude proizvode – ukoliko se kupcu ponude adekvatni proizvodi to bi zasigurno trebalo dovesti do povećanja prodaje





# Kupci i proizvodi

- Prepoznavanje relacija (konekcija) između pojedinca i proizvoda je od izuzetne važnosti za prodaju (neki kupci preferiraju sportsku opremu, neki tehničke uređaje, neki unikatne proizvode i sl.)
- Postoje relacije između proizvoda (vozila istih karakteristika, knjige istog žanra, uređaji istih funkcionalnosti i sl.)
- Postoje sličnosti između kupaca (vole iste knjige ili vozila, imaju zajedničke prijatelje i sl.)





# Kreiranje ponude

- Servisi kao što su Amazon, eBay i Netflix prikupljaju na milione zapisa o korisnicima, a njihova adekvatna analiza omogućava da se dobiju odgovori na neke od sljedećih pitanja:
  - koje knjige kupac voli? (kupac  $\leftrightarrow$  proizvod)
  - ukoliko je kupio mobilni uređaj koje još dodatke bi kupio uz njega? (proizvod  $\leftrightarrow$  proizvod)
  - ako kupac Jasmin poznaje kupca Adela, a kupac Adel poznaje kupca Zanina, da li postoji vjerovatnoća da i kupac Jasmin poznaje kupca Zanina? (kupac  $\leftrightarrow$  kupac)
- Problem: *kako kupcima ponuditi (preporučiti) ono što ih zaista interesuje?*
- Rješenje: *sistemi preporuke*

# Sistem preporuke



- Sistemi preporuke imaju za cilj osigurati personalizaciju i kupcima ponuditi ono što odgovara njihovim potrebama. Postoji nekoliko vrsta podataka koji su potrebni za rad ovakvih sistema:
  - Podaci o ponašanju kupca (ocjene proizvoda, broj pregleda, količina i učestalost kupovine i sl.)
  - Demografski podaci (lokacija, starost i sl.)
  - Atributi proizvoda (marka i snaga vozila, žanr knjige i sl.)
- Na osnovu pomenuti podataka sistemi preporuke, koristeći odgovarajuće algoritme, mogu u značajnoj mjeri personalizirati okruženje i interakciju sa kupcima





# Sistem preporuke

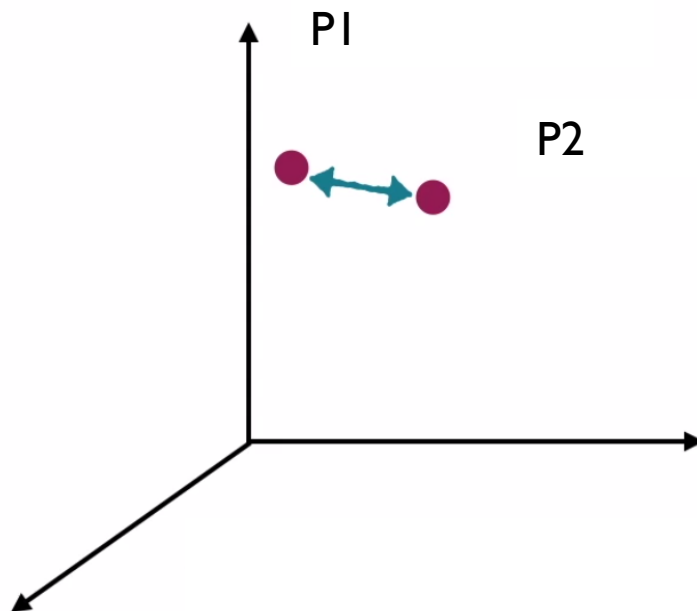
## Algoritmi

- Content-based filtering (**CBF**)
  - Omogućava preporuku proizvoda sličnih onima koje je kupac prethodno pregledao, ocijenio, komentarisao, kupovao (knjige istog žanra ili pisca, vozila iste snage ili proizvođača i sl.)
- Collaborative filtering (**CF**)
  - Omogućava preporuku proizvoda koje su drugi kupci, sa sličnim preferencijama, već ocijenili. Ukoliko su kupci slični, onda će vjerovatno imati slične zahtjeve i potrebe (ako se kupcu Zaninu sviđa marka vozila Hyundai, onda bi se ta marka vozila mogla sviđati i kupcu Jasminu sa kojim posjeduje veliku sličnost).
  - Može se govoriti o dvije vrste kolaborativnog filtriranja: **User-based** i **Item-based**
- Association rules learning (**ARL**)
  - Omogućava preporuku proizvoda na osnovu učestalosti zajedničke kupovine (ako se prilikom kupovine mobilnog uređaja gotovo uvijek kupuje i zaštita za ekran, onda se može prepoznati neka asocijacija između njih).



# Content-based filtering

- Posmatraju se vrijednosti atributa proizvoda, ali prvenstveno one koje imaju najveći značaj za korisnika (rijetko nam je bitno koji hemijski elementi ulaze u sastav šasije nekog vozila i sl.)
- Svaka vrijednost atributa se posmatra kao koordinata u prostoru
- Sličnost između proizvoda se može posmatrati kroz njihovu udaljenost u prostoru



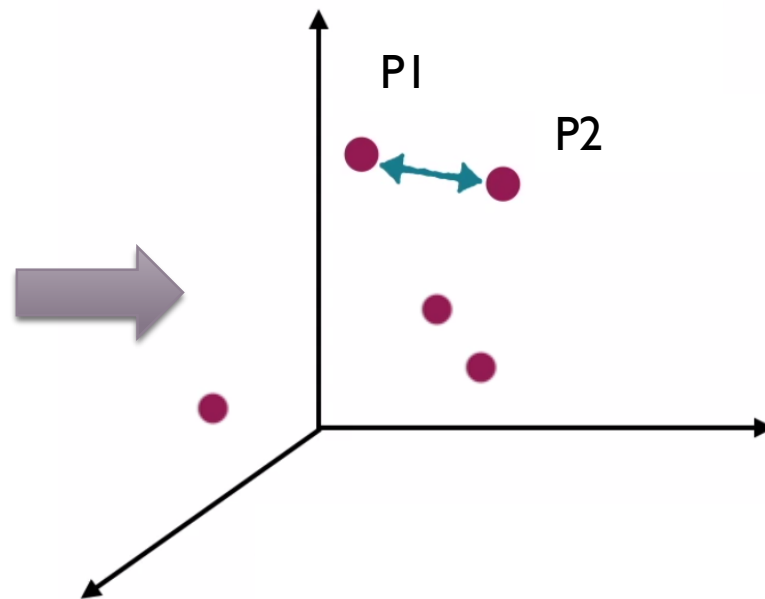




# Content-based filtering

- Za određivanje sličnosti između proizvoda se mogu koristiti različite metrike, a neke od njih su: Euclidean-ova, Hamming-ova ili Correlation (korelacijska) distanca
- U narednoj tabeli su prikazane ocjene osobina ( $O_i$ ) za 5 različitih proizvoda, te su na osnovu njih pomenuti proizvodi predstavljeni u prostoru

	O1	O2	O3	O4	O5
P1	5	3	4	4	4
P2	3	1	2	3	3
P3	4	3	4	3	5
P4	3	3	1	5	4
P5	1	5	5	2	1



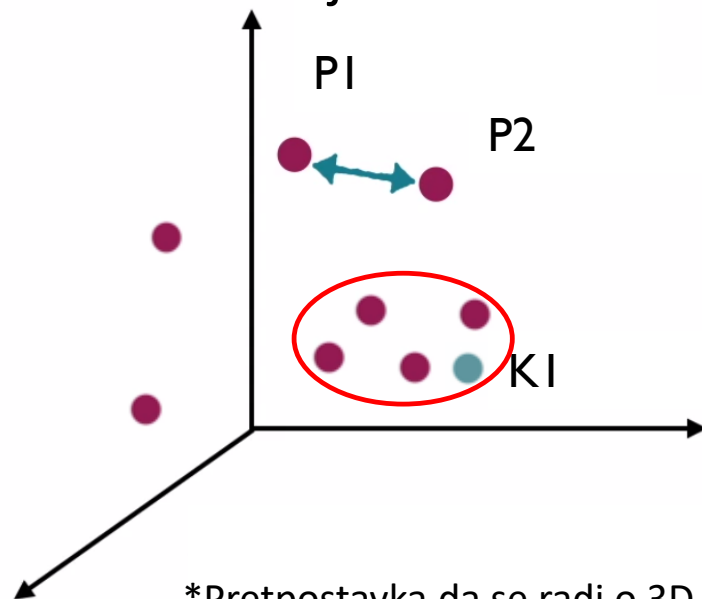
\*Pretpostavka da se radi o 3D prostoru



# Content-based filtering

- Da bismo znali koje proizvode preporučiti kupcu, potrebno je i kupca predstaviti u prostoru u kome se nalaze proizvodi. Vrijednosti, odnosno vektor, kojim će kupac biti predstavljen u prostoru mogu, ali ne moraju, biti prosječne vrijednosti ocjena kojim je kupac ocijenio osobine posmatrane kategorije proizvoda
- Proizvodi koji budu najbliži kupcu su oni koji bi trebali biti preporučeni

	O1	O2	O3	O4	O5
K1	5	3	4	4	4



\*Pretpostavka da se radi o 3D prostoru



# User-based CF

- Sličnost između dva korisnika se mjeri pomoću Pearson korelacije.

	p1	p2	p3	p4	p5
a	5	3	4	4	?
b	3	1	2	3	3
c	4	3	4	3	5
d	3	3	1	5	4
e	1	5	5	2	1

$$\text{slicnost}(a, b) = \frac{\sum_{p \in P} (o_{a,p} - \bar{o}_a)(o_{b,p} - \bar{o}_b)}{\sqrt{\sum_{p \in P} (o_{a,p} - \bar{o}_a)^2} \sqrt{\sum_{p \in P} (o_{b,p} - \bar{o}_b)^2}}$$

$a, b$  – korisnici sistema;  $a, b \in K$

$P$  – skup proizvoda koje su ocijenili korisnici;  $p \in P$

$o_{a,p}$  – ocjena kojom je korisnik **a** ocijenio proizvod **p**

$\bar{o}_a$  – prosječna ocjena korisnika **a**



# User-based CF

$$\text{slicnost}(a, b) = \frac{\sum_{p \in P} (o_{a,p} - \bar{o}_a)(o_{b,p} - \bar{o}_b)}{\sqrt{\sum_{p \in P} (o_{a,p} - \bar{o}_a)^2} \sqrt{\sum_{p \in P} (o_{b,p} - \bar{o}_b)^2}}$$

	BMW	Audi	VW	Mercedes	Hyundai	Prosjek	Sličnost
Zanin	5	3	4	4	?	4	
Jasmin	3	1	2	3	3	2,25	0,634441
Goran	4	3	4	3	5	3,5	0,516398
Adel	3	3	1	5	4	3	0
Denis	1	5	5	2	1	3,25	-0,54719

$$\begin{aligned} & \text{slicnost}(\text{Zanin}, \text{Jasmin}) \\ &= \frac{(5 - 4)(3 - 2,25) + (3 - 4)(1 - 2,25) + (4 - 4)(2 - 2,25) + (4 - 4)(3 - 2,25)}{\sqrt{(5 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (4 - 4)^2} \sqrt{(3 - 2,25)^2 + (1 - 2,25)^2 + (2 - 2,25)^2 + (3 - 2,25)^2}} = 0,63 \end{aligned}$$

# User-based CF



	BMW	Audi	VW	Mercedes	Hyundai	Prosjek	Sličnost
Zanin	5	3	4	4	?	4	
Jasmin	3	1	2	3	3	2,25	0,634441
Goran	4	3	4	3	5	3,5	0,516398
Adel	3	3	1	5	4	3	0
Denis	1	5	5	2	1	3,25	-0,54719

$$\text{pretpostavka}(a, p) = \overline{o_a} + \frac{\sum_{n \in NK} \text{slicnost}(a, n)(o_{n,p} - \overline{o_n})}{\sum_{n \in NK} \text{slicnost}(a, n)}$$

$$\text{pretpostavka}(\text{Zanin}, \text{Hyundai}) = 4 + \frac{0,63 * (3 - 2,25) + 0,51 * ..}{0,63 + 0,51 + ..}$$



# Item-based CF

- Item-based CF računa preporuku na osnovu sličnosti između proizvoda.

$$\text{slicnost}(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| |\vec{B}|}$$

$$A = [3, 5, 4, 1]$$

$$B = [1, 3, 3, 5]$$

$$A \cdot B = \sum_{i=1}^n A_i B_i = A_1 B_1 + A_2 B_2 + \dots + A_n B_n$$

$$A \cdot B = (3 * 1) + (5 * 3) + (4 * 3) + (5 * 1)$$

Hyundai
?
3
5
4
1

Audi
3
1
3
3
5

$$|A| = \sqrt{A_1^2 + A_2^2 + \dots + A_n^2} = \sqrt{3^2 + 5^2 + 4^2 + 1^2}$$



# Collaborative filtering

- Jedna od najjednostavnijih načina za procjenu sličnosti je Hamming-ova udaljenost kojom se izračunava pojam neslaganja
- Hamming-ova udaljenost je pogodna prvenstveno u slučajevima kada se za vrijednosti atributa proizvoda koriste diskretne vrijednosti
- U narednom primjeru, neslaganje između kupca K1 i K2 je 60% ili 0.6

	O1	O2	O3	O4	O5
K1	5	3	4	4	4
K2	5	2	4	3	1

neslaganje = 0.6



# Association rules learning

- Ukoliko se neki proizvodi često kupuju zajedno, onda će se na osnovu te relacije i bazirati preporuka
  - Support – vjerovatnoća kupovine posmatranog proizvoda (mobitel)
  - Confidance – vjerovatnoća kupovine drugog proizvoda nakon kupovine prvog (mobitel i zaštita za ekran)
  - Lift – promjena u vjerovatnoći da će drugi proizvod biti kupljen nakon kupovine prvog
- Tokom određivanja asocijacija se može koristiti Apriori algoritam koji u svakoj fazi eliminiše proizvode koji imaju vrijednost asocijacije ispod određenog praga

$$P(mobitel) = 3\%$$

$$P(zastita\_ekran \mid mobitel) = 7\%$$

$$Lift(mobitel \Rightarrow zastita\_ekran) = \frac{7}{3} = 2,33$$



---

KRAJ PREZENTACIJE

