

# Sustavi za preporuke na internetu

---

**Bikić, Mila**

**Undergraduate thesis / Završni rad**

**2022**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:131:104044>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-09-02**



Sveučilište u Zagrebu  
Filozofski fakultet  
University of Zagreb  
Faculty of Humanities  
and Social Sciences

*Repository / Repozitorij:*

[ODRAZ - open repository of the University of Zagreb  
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU  
FILOZOFSKI FAKULTET  
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI  
Ak. god. 2021./2022.

Mila Bikić

## **Sustavi za preporuke na internetu**

Završni rad

Mentor: prof.dr.sc. Sanja Seljan

Zagreb, srpanj 2022.

## Izjava o akademskoj čestitosti

Izjavljujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

*Mila Bikić*

---

Mila Bikić



# Sadržaj

1. Uvod.....	1
2. Sustavi za preporuku.....	3
2.1. Kontekstno utemeljeni sustavi .....	5
2.1.1. TF-IDF mjera .....	6
2.1.2. Računanje sličnosti .....	8
2.2. Kolaborativno filtriranje.....	9
2.2.1. Kolaborativno filtriranje utemeljeno na memoriji .....	11
2.2.2. Kolaborativno filtriranje utemeljeno na modelu.....	11
2.2.3. Računanje preporuka .....	12
2.3. Druge vrste sustava za preporuku .....	13
3. Evaluacija sustava za preporuku .....	15
3.1. Metode evaluacije .....	15
3.2. Metrike evaluacije .....	16
4. Primjene sustava za preporuku na internetu .....	19
4.1. Netflix.....	19
4.2. Deezer.....	22
5. Stavovi mladih o sustavima za preporuke .....	26
5.1. Cilj i pretpostavke istraživanja.....	26
5.2. Metodologija istraživanja.....	26
5.3. Rezultati istraživanja .....	28
5.4. Rasprava .....	38
6. ISTRAŽIVANJE.....	40
6.1. Preporuka u kontekstno utemeljenom sustavu .....	40
6.2. Preporuka u kolaborativnom filtriranju.....	42
7. Zaključak.....	45
8. Literatura.....	47

Sažetak .....	51
Summary .....	52

# 1. Uvod

Prema Hrvatskom jezičnom portalu (n.d.)<sup>1</sup> smatra se da je preporuka „povoljna ocjena o svojstvima koga/čega“. Preporuke se nalaze u pozadini mnogih web-stranica svih tipova – internetske trgovine, usluge digitalnog strujanja, novinskih portala, i tako dalje. Mogu biti ručno preporučene od urednika, jednostavne liste, npr. top 10, najpopularnije stavke na web stranici, najnovije stavke, i slično, ili mogu biti prilagođene pojedinačnim korisnicima uz pomoć sustava za preporuku (Vyas, 2018)<sup>2</sup>.

Sustavi za preporuku (eng. recommender systems – RS) su softverski alati i tehnike koje se koriste za generiranje preporuka za stavke koje će određenom korisniku biti zanimljive. Koriste se kod donošenja raznih odluka kao npr. koju glazbu slušati, koje članke čitati ili koje proizvode kupiti. Određeni sustav za preporuku je obično fokusiran na određenu vrstu stavke (npr. knjiga, glazba ili CD) te je njegovo korisničko sučelje i tehnika koja se koristi za preporuku prilagođena je za davanje korisnih i efektivnih preporuka za tu vrstu stavke.

Sustavi za preporuku prvobitno su usmjereni osobama koji nisu dovoljno iskusne ili kompetentne za evaluaciju potencijalno prevelikog broja sličnih stavki koje određena web stranica nudi (Ricci, Rokach & Shapira, 2015, str. 1)<sup>3</sup>. Koriste se u svrhu navođenja korisnika na neku akciju kao što je kupovanje određenog proizvoda ili gledanje određenog filma ili kao alat protiv prezasićenosti informacija. Naime, oni smanjuju veliki skup stavki na manji skup korisniku najzanimljivijih stavki. Utemeljeni su u područjima pretraživanja informacija i filtriranja informacija (Jannach, Zanker & Felfernig, 2012, str. 3)<sup>4</sup>. Važan su dio platformi za digitalno strujanje sadržaja kao što su *Netflix*, *Spotify* i *Pandora*, i web-stranica za internetskih trgovina poput *Amazon*, ali i društvenih mreža kao što su *Facebook*, *Instagram* i *TikTok*. Primjerice, kad korisnik pristupi nekoj internetskoj trgovini, npr. svojoj omiljenoj online knjižari, na početnoj stranici će uz popis trenutačno najprodavanijih knjiga biti i popis personaliziranih preporuka. Taj popis uključuje stavke poput knjiga koje su po nekom kriteriju

---

<sup>1</sup> *Preporuka*. (n.d.). Hrvatski Jezični Portal. <https://hjp.znanje.hr/index.php?show=search> (Pristupljeno: 19. svibnja 2022.)

<sup>2</sup> Vyas, M. (2018). *Recommendation Systems*. Medium. <https://meenavyas.medium.com/recommendation-systems-1d20d159634d> (Pristupljeno: 19. svibnja 2022.)

<sup>3</sup> F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.

<sup>4</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

slične onima koje je korisnik u prošlosti kupovao ili pregledavao (Burke, Felfernig & Göker, 2011)<sup>5</sup>.

U prvom teorijskom dijelu ovoga rada predstavljaju se različite vrste sustava za preporuku te se teorijski objašnjavaju važni koncepti poput TF-IDF mjere, računanja sličnosti te računanja preporuka. Predstavljen je tijek evaluacije sustava za preporuku, metode izvan mrežne analitike, korisničke studije i testiranja na mreži. Opisane su sljedeće metrike evaluacije: srednja apsolutna greška, korjenovana prosječna kvadrirana greška, preciznost, odziv, R-rezultat metrika i mjera klikanja. Dani su primjeri korištenja sustava za preporuku na platformama *Netflix* i *Deezer*. U drugom praktičnom dijelu rada predstavljenu su rezultati istraživanja kojemu je cilj ustanoviti upoznatost mladih ljudi u dobi od 16 do 30 s ovom temom te njihove stavove o sustavima za preporuku. Nadalje su izrađeni modeli sustava za preporuku nad dva različita skupa podataka u programskom jeziku *Python*. Na kraju slijedi zaključak, popis literature te prilog.

---

<sup>5</sup> Burke, R., Felfernig, A., & Göker, M. H. (2011). Recommender Systems: An Overview. *AI Magazine*, 32(3), 13–18. <https://doi.org/10.1609/aimag.v32i3.2361>



## 2. Sustavi za preporuku

Prema Jannachu i sur. (2012)<sup>6</sup>, ideja za sustav koji bi prikupljao mišljenja ljudi u svrhu olakšavanja pronalaska korisnog i zanimljivog sadržaja nastala je početkom devedesetih godina dvadesetog stoljeća. Kad su se trgovine, knjižare i kina prebacivale na internet, ukazala se potreba za automatiziranim sustavima koji bi pomogli korisnicima pretraživati stavke i olakšati im pregledavanje interneta ili online kupnju (str. ix). Računalni program za knjižnicu *Grundy* iz 1979. godine se smatra začetnikom automatiziranih sustava za preporuke, a nadalje se izdvaja *Tapestry* iz 1992. godine koji je korisnicima omogućio dodavanje oznaka tj. anotacija mailovima i pretraživanje istih prema tim oznakama (Marković & Tomljanović, 2018)<sup>7</sup> te prvi uvodi terminologiju za kolaborativno filtriranje. Sustav *GroupLens* je uveo pristup kolaborativnom filtriranju kojeg je moguće automatizirati i distribuirati putem mreže, a koristio se za filtriranje novosti. Tehnološki institut u Massachusettsu je razvio sustav *Ringo* koji se koristio za filtriranje glazbenih albuma i izvođača, a sustav *Bellcore Video Recommender* je razvijen za filtriranje filmova (Jannach i sur., 2012, str. ix)<sup>8</sup>. Na početku su se takvi sustavi koristili da korisniku preporuča vijesti za pročitati i web stranice za posjetiti (Felfernig, Friedrich & Schmidt-Thieme, 2007)<sup>9</sup>.

Slika 1 prikazuje dvije najraširenije vrste sustava za preporuku: kolaborativno filtriranje i kontekstno utemeljeni sustavi. Razlike između njih su sljedeće:

- Kolaborativno filtriranje, koji se najčešće koristi u internetskim trgovinama, polazi od stajališta da ako su korisnici u prošlosti slično ponašali, primjerice čitali iste članke, u budućnosti će nastaviti s istim ponašanjem. Dakle, sustav će im preporučiti članke koje oni nisu čitali, ali drugi slični korisnici jesu.
- Kontekstno utemeljeno filtriranje će preporučiti korisniku slične članke koje je u prošlosti čitao (Dasca, 2019)<sup>10</sup>.

---

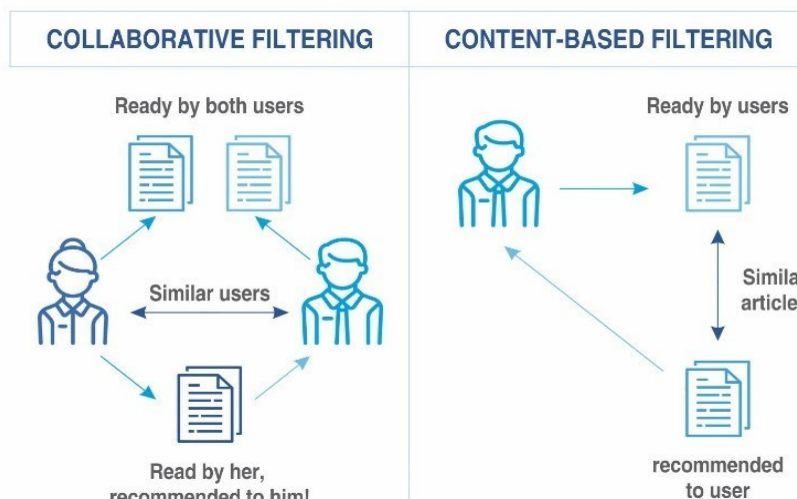
<sup>6</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

<sup>7</sup> Marković, M., & Tomljanović, Z. (2018). "Kolaborativno filtriranje." *Math.e*, 34(1), 1-22. <https://hrcak.srce.hr/222763>

<sup>8</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

<sup>9</sup> Felfernig, Alexander, Gerhard Friedrich, and Lars Schmidt-Thieme. 2007. "Guest Editors' Introduction: Recommender Systems". *IEEE Intelligent Systems* 22 (3): 18-21. doi:10.1109/mis.2007.52.

<sup>10</sup> *Machine Learning for Recommender Systems - A Primer*. (2019). Dasca. <https://www.dasca.org/world-of-big-data/article/machine-learning-for-recommender-systems-a-primer> (Pristupljeno: 21. svibnja 2022.)



Slika 1. Najraširenije vrste sustava za preporuku (Dasca, 2019)<sup>11</sup>

Dasca (2019)<sup>12</sup> izdvaja tri faze preporuke: prikupljanje informacija, učenje i preporuka. Tijekom faze prikupljanja informacija prikupljaju se relevantni podaci o korisniku, atributima stavki, ponašanju korisnika, povratne informacije, i slično. Korisnikove povratne informacije mogu se prikupljati eksplicitno ili implicitno. Pri eksplicitnom prikupljanju, korisnika se pita da ocjenjuje stavke. Primjerice, web stranica ili aplikacija može korisnika pitati da ocjeni stavku na skali od 1-5, pitati ga da pretraži stavku, da rangira stavke od najbolji do najgorih, pitati ga da stvori listu najdražih stvari, i slično. Ovom metodom dobivaju se vjerodostojne informacije, ali većina ljudi ne ocjenjuje stavke tako da većina stavki ostaje neocijenjena.

Implicitna metoda prikupljanja povratnih informacija uključuje promatranje i učenje povratnih informacija korisnika. Najčešće se prati korisnikova povijest pregledavanja stavki, vrijeme pregledavanja, ako se radi o internetskoj trgovini prati se povijest kupovine, i slično (Vyas, 2018)<sup>13</sup>. U fazi učenja uvodi se sustav za preporuku te se provjerava može li sustav koristiti prikupljene informacije za izvođenje preporuka. U zadnjoj fazi, sustav predviđa odabir korisnika ili stavke koje bi se korisniku najviše svidjele (Dasca, 2019)<sup>14</sup>.

<sup>11</sup> *Machine Learning for Recommender Systems - A Primer*. (2019). Dasca. <https://www.dasca.org/world-of-big-data/article/machine-learning-for-recommender-systems-a-primer> (Pristupljeno: 21. svibnja 2022.)

<sup>12</sup> Ibid.

<sup>13</sup> Vyas, M. (2018). *Recommendation Systems*. Medium. <https://meenavyas.medium.com/recommendation-systems-1d20d159634d> (Pristupljeno: 19. svibnja 2022.)

<sup>14</sup> *Machine Learning for Recommender Systems - A Primer*. (2019). Dasca. <https://www.dasca.org/world-of-big-data/article/machine-learning-for-recommender-systems-a-primer> (Pristupljeno: 21. svibnja 2022.)

## 2.1. Kontekstno utemeljeni sustavi

U kontekstno utemeljenim sustavima su sustavi za preporuke koje korisniku preporučuju sadržaj sličan onome kojeg je u prošlosti konzumirao. Koristi se sadržaj korisnika i njegovi interesi za preporuku drugih stavki. Svrha kontekstno utemeljenih sustava jest klasificirati stavke prema ključnim riječima, naučiti što korisnik voli, pretražiti bazu podataka prema tim ključnim riječima i koristiti prikupljene informacije za preporuku sličnih stavki. Primjerice, ključne riječi u sustavu za preporuku filmova bi bile informacije o trajanju filma, žanr, glavni glumci i slično, a za korisnika dob, spol, posao i druge osobne informacije (Mishra, 2022)<sup>15</sup>.

Na web stranicama Google (2022)<sup>16</sup> ističe kao prednosti kontekstno utemeljenog sustava sljedeće:

- Sustavu nisu potrebne informacije o drugim korisnicima jer su preporuke jedinstvene za svakog korisnika.
- Model može sakupiti specifične interese korisnika te mu preporučiti stavku koja je interesantna malom broju korisnika.

S druge strane, kao nedostatke ističe da:

- Značajke stavki su do određene mjere ručno izrađene tako da model može biti samo jednako dobar kao i njegove ručno izrađene značajke. Ovo također zahtjeva veliku količinu znanja o domeni.
- Sustav može samo preporučiti stavke utemeljene na već postojećim interesima korisnika te je njegova mogućnost proširivanja postojećih interesa korisnika ograničena.

---

<sup>15</sup> Mishra, U. (2022). *What is a Content-based Recommendation System in Machine Learning?* Analytics Steps. <https://www.analyticssteps.com/blogs/what-content-based-recommendation-system-machine-learning> (Pristupljeno: 28. travnja 2022.)

<sup>16</sup> 15. Google. (n.d.). *Content-based Filtering Advantages & Disadvantages*. Google Developers. <https://developers.google.com/machine-learning/recommendation/content-based/summary> (Pristupljeno: 27. travnja 2022.)

### 2.1.1. TF-IDF mjera

Prema Kumar, Pavan, Potluri i Mohanty (2021)<sup>17</sup>, TF-IDF mjera (eng. Term Frequency-Inverse Document Frequency) je tehnika vektorizacije teksta koja se koristi u obradi prirodnog jezika. Prednost TF-IDF mjere jest da omogućuje računanje sličnosti između različitih dokumenata, ali nedostatak je da je ne uzima semantiku riječi u obzir te supojavljivanje druge riječi u dokumentu. Kod TF-IDF mjere riječi koje se pojavljuju češće u jednom dokumentu, ali rijetko u ostalim dokumentima će biti relevantnije za temu dokumenta (Ricci i sur., 2015, str. 123)<sup>18</sup>.

Jannach i sur. (2012)<sup>19</sup> navode da se TF temelji na pretpostavki da važnije riječi se češće pojavljuju u dokumentu i mjera je učestalosti pojavljivanja određene riječi u dokumentu. Potrebna je normalizacija duljine dokumenta kako bi se uzela u obzir veličina dokumenta. U protivnom će dulji dokumenti biti relevantniji od kraćih dokumenata. Formula kojom se računa normalizirana frekvencija termina  $TF(i, j)$  glasi (str. 55):

$$TF(i, j) = \frac{freq(i, j)}{maxOthers(i, j)}$$

U kojoj je  $freq(i, j)$  apsolutni broj pojave ključne riječi  $i$  u dokumentu  $j$ ,  $OtherKeywords(i, j)$  označava skup ostalih ključnih riječi koje se pojavljuju u  $j$  te se maksimalna frekvencija  $maxOthers(i, j)$  računa kao  $max(freq(z, j))$ ,  $z \in OtherKeywords(i, j)$ .

IDF je mjera koja se koristi za smanjivanje težine tokena koji se jako često pojavljuju u svim dokumentima jer takve riječi nisu korisne i riječi koje se pojavljuju u samo nekim dokumentima su važnije. IDF se računa pomoću formule (Jannach i sur., 2012, str. 56):

$$IDF(i) = \log \frac{N}{n(i)}$$

---

<sup>17</sup> Kumar, P. P., Vairachilai, S., Potluri, S., & Mohanty, S. N. (2021). Recommender Systems: Algorithms and Applications. CRC Press. <https://doi.org/10.1201/9780367631888>

<sup>18</sup> Francesco, Ricci, Rokach Lior, Bracha Shapira, and Paul B. Kantor. 2015. *Recommender Systems Handbook*. 2. izdanje. New York: Springer.

<sup>19</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

U kojoj je  $N$  broj svih preporučljivih dokumenata,  $n(i)$  broj dokumenata iz  $N$  u kojima se token  $i$  pojavljuje. Formula kojom se računa TF-IDF za ključnu riječ  $i$  u dokumentu  $j$  se računa kao umnožak tih dvaju veličina (Jannach i sur., 2012, str. 56):

$$TF - IDF(i, j) = TF(i, j) * IDF(i)$$

U svrhu uklanjanja nerelevantnih informacija iz TF-IDF vektora i smanjenja veličine vektora, Jannach i sur. (2012)<sup>20</sup> izdvajaju sljedeće tehnike (str. 56-57):

- Stop riječi i korjenovanje – stop riječi (npr. članovi *the*, *a/an*, prijedlozi *on*, *under*) se mogu ukloniti jer se pojavljuju u gotovo svim dokumentima. Korjenovanjem se zamjenjuje varijante određene riječi sa korijenom te riječi (npr. riječ *dokumentima* bi se zamijenio sa riječi *dokument*).
- Smanjivanje veličine – ovom metodom se koristi samo  $n$  broj najinformativnijih riječi. Važno je ne koristiti premali ili preveliki broj riječi. U slučaju premalog broja riječi neće biti uključeni svi važni elementi dokumenta, a u slučaju prevelikog broja riječi, stvara se prevelika buka zbog velikog broja nerelevantnih riječi.
- Izrazi – korištenjem fraza umjesto pojedinačnih termina moguće je poboljšati točnost vektora jer su izrazi informativniji od pojedinačnih riječi. Izraze se može generirati iz ručno napisanih popisa ili korištenjem tehnika statističke analize.

Primjer računanje TF-IDF mjere je prikazan na Slici 2. U stupcima TF upisani su brojevi pojave svake riječi u dokumentima D1, D2 i D3. U stupcima TF (normalizirano) prikazane su normalizirane vrijednosti pojave svake riječi. U stupcu IDF je IDF vrijednost za svaki broj, a u stupcima TF-IDF je izračun TF-IDF mjere za svaku riječ u svakom dokumentu.

---

<sup>20</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

RIJEČ	TF			TF (normalizirano)			IDF	TF-IDF		
	D1	D2	D3	D1	D2	D3		D1	D2	D3
stablo		2	1	0	1	0,333333	0,176091	0	0,176091	0,058697
cvijet		2	1	0	1	0,333333	0,176091	0	0,176091	0,058697
zemlja		1		0	0,5	0	0,477121	0	0,238561	0
saditi	2		1	1	0	0,333333	0,176091	0,176091	0	0,058697
mokro			3	0	0	1	0,477121	0	0	0,477121
lonca	1			0,5	0	0	0,477121	0,238561	0	0

Slika 2. Primjer računanja TF-IDF mjere. (Izvor: vlastita izrada)

### 2.1.2. Računanje sličnosti

Metode računanja sličnosti ili težine između korisnika ili stavki su brojne, ali najčešće korištena mjera sličnosti u sustavima za preporuku je kosinusna mjera sličnosti. Vrijednosti ove sličnosti iznose od -1 do 1, s tim da 1 znači da su stavke iste, a -1 da stavke nemaju nikakve sličnosti. Sličnost između stavki je veća što je manji kut između vektora na kojem se nalaze (Subramanian, 2020)<sup>21</sup>. Kosinusna mjera sličnosti u računalnom učenju za vektore  $A$  i  $B$  se računa pomoću sljedeće formule:

$$\text{Kosinusna mjera sličnost}(A, B) = \cos \theta = \frac{A * B}{\|A\| \|B\|}$$

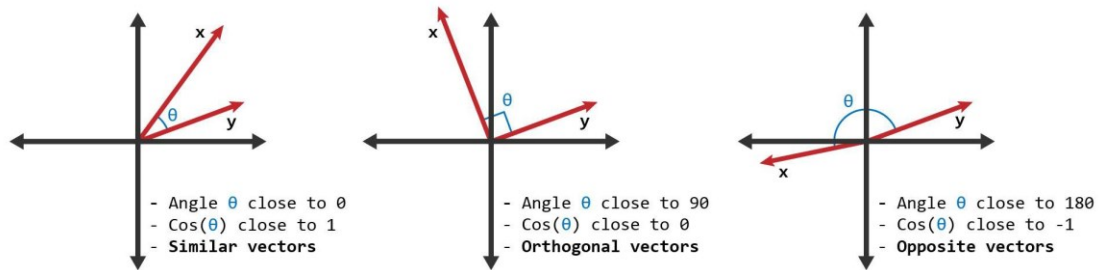
U kojoj je  $\theta$  kut između vektora,  $A*B$  je skalarni produkt, a  $\|A\|$  i  $\|B\|$  predstavljaju L2 normu vektora (Karabiber, 2022)<sup>22</sup>. L2 norma se ujedno zove i euklidska mjera jer se koristi za računanje euklidske udaljenosti od baze vektorskog prostora. Računa se kao korijen zbroja kvadriranih vrijednosti vektora (Brownlee, 2018)<sup>23</sup>. Slika 2 prikazuje da su dva vektora slična ako je kut  $\theta$  između njih bliži  $0^\circ$  te je tada kosinusna sličnost blizu 1. U slučaju da kut  $\theta$  iznosi blizu  $90^\circ$ , kosinusna sličnost je blizu 0 te takve vektore nazivamo ortogonalnim vektorima. Ako je kut  $\theta$  blizu  $180^\circ$ , kosinusna sličnost je blizu -1, a takvi vektori se nazivaju suprotnim vektorima. Kosinusna sličnost se koristi za računanje sličnosti kad je prisutan veliki broj nepotpunih podataka. U slučaju da se brojni nepotpuni podaci uključe u računanje sličnosti, oni

<sup>21</sup> Subramanian, D. (2020). *Cosine Similarity, the metric behind recommendation systems*. Indiaai.Gov.In. <https://indiaai.gov.in/article/cosine-similarity-the-metric-behind-recommendation-systems> (Pristupljeno: 27. travnja 2022.)

<sup>22</sup> Karabiber, F. (2022). *Cosine Similarity*. Learn Data Science. <https://www.learndatasci.com/glossary/cosine-similarity/> (Pristupljeno: 22. lipnja 2022.)

<sup>23</sup> Brownlee, J. (2018). *Gentle Introduction to Vector Norms in Machine Learning*. Machine Learning Mastery. <https://machinelearningmastery.com/vector-norms-machine-learning/> (Pristupljeno: 22. lipnja 2022.)

se podudaraju s obzirom da svi imaju vrijednost 0 te bi povećali vrijednost sličnosti (Karabiber, 2022)<sup>24</sup>.



Slika 3. Prikaz vektora u vektorskom prostoru (Karabiber, 2022)<sup>25</sup>.

Slika 4 predstavlja dva skupa podataka, odnosno vektore, te rezultat koji se dobije primjenom kosinusne mjere sličnosti nad tim skupovima. Rezultat izračuna iznosi 0,892 te iz toga proizlazi da su ova dva vektora slična.

Skup podataka A	Skup podataka B
23	17
25	23
34	22
32	54
67	67
12	32
11	42
21	55
<b>Kosinusna sličnost:</b>	<b>0,892492028</b>

Slika 4. Primjer izračuna kosinusne mjere sličnosti. (Izvor: vlastita izrada)

## 2.2. Kolaborativno filtriranje

Goldberg, Nichols, Oki i Terry (1992)<sup>26</sup> su definirali kolaborativno filtriranje (eng. collaborative filtering – CF) 1992. godine i osmislili eksperimentalni sustav za upravljanje

<sup>24</sup> Karabiber, F. (2022). *Cosine Similarity*. Learn Data Science. <https://www.learndatasci.com/glossary/cosine-similarity/> (Pristupljeno: 22. lipnja 2022.)

<sup>25</sup> Ibid.

<sup>26</sup> Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70. <https://doi.org/10.1145/138859.138867>

elektroničkim dokumentima *Tapestry*. *Tapestry* je, slično kao i drugi sustavi u to vrijeme, koristio filtriranje, ali u sustavu *Tapestry*, ljudi su mogli anotirati dokumente, te filtrirati ih prema anotacijama. Kolaborativno filtriranje je, u suštini, sustav u kojem ljudi surađuju (tj. kolaboriraju) da bi filtrirali po njihovim reakcijama na neki dokument. U preporuku su uključena dva faktora: dokument i njegove anotacije. Kolaborativno filtriranje polazi sa stajališta da ako su korisnici u prošlosti imali zajedničke interese (primjerice, ako su gledali ili kupili istu knjigu ili film), u budućnosti će imati sličan ukus (Jannach i sur., 2012, str. 2-3)<sup>27</sup>.

Isinkaye, Folajimi i Ojokoh (2015)<sup>28</sup> izdvajaju kao prednosti kolaborativnog filtriranja sljedeće:

- Moguće ga je koristiti u domenama gdje nema mnogo stavki i gdje je otežano računalno analiziranje sadržaja (npr. mišljenja i ideali).
- Prisutan je faktor iznenađenja (eng. serendipity) što omogućuje relevantne preporuke bez prisutnosti takvog sadržaja na profilu korisnika.

Nedostaci ovog pristupa uključuju:

- Problem hladnog starta (eng. cold start problem) – situacija u kojem sustav preporuke nema adekvatnu količinu informacija o određenom korisniku ili stavki te nije u mogućnosti preporučiti relevantan sadržaj. Zbog nedostatka bilo kakvih anotacija korisnika, jer ovakvi profili pripadaju novim korisnicima, sustavu nije poznat ukus korisnika.
- Problem oskudnosti podataka – posljedica nedovoljnog broja informacija u slučajevima kad su samo nekoliko od ukupnog broja stavki korisnici ocijenili, što uzrokuje oskudnu bazu podataka, nemogućnost lociranja bliskog susjeda te davanje loših preporuka.
- Skalabilnost – sustav preporuke koji je uspješan na ograničenoj bazi podataka ne mora biti dostatno uspješan kad se poveća veličina baze podataka.
- Sinonimnost – odnosi se na različite nazive ili unose izrazito sličnih stavki. Većina sustava za preporuku otežano razlikuje bliske stavke kao što su npr. *baby clothes* i *baby cloth*. Za rješavanje ovog problema koriste se metode obrade prirodnog jezika: automatsko proširenje termina, tezaursus. Međutim, primjena ovih metoda može utjecati na učinkovitost sustava za preporuku.

---

<sup>27</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

<sup>28</sup> Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>



Kao dva glavna pristupa kolaborativnom filtriranju, Ricci i sur. (2015)<sup>29</sup> navode:

- Kolaborativno filtriranje utemeljeno na memoriji;
- Kolaborativno filtriranje utemeljeno na modelu (str. 38).

### **2.2.1. Kolaborativno filtriranje utemeljeno na memoriji**

Kolaborativno filtriranje bazirano na memoriji izravno koristi korisnikove ocjene stavki (eng. user-item ratings) za predviđanje ocjena novih stavki. Ovaj pristup može biti orijentiran prema korisniku – koristeći ocjene osoba koje slično ocjenjuju kao željeni korisnik (takozvani *susjedi*) dolazi se do predviđene ocjene željenog korisnika – ili prema stavki – korisnikova ocjena se predviđa koristeći korisnikove ocjene sličnih stavki. U ovakvom sustavu prisutan je faktor iznenađenja. Naime, sustav može korisniku preporučiti stavku koja ne odgovara nužno korisnikovom ukusu ili nije široko poznata stavka jer je netko od korisnikovih susjeda tu stavku visoko ocijenio (Francesco i sur., 2015, 38-39)<sup>30</sup>..

### **2.2.2. Kolaborativno filtriranje utemeljeno na modelu**

Kolaborativno filtriranje utemeljeno na modelu koristi ocjene korisnika za učenje modela za predviđanje. Ovaj sustav ističe se prepoznavanjem karakteristika stavki koje se korisniku se sviđa, ali nije u mogućnosti preporučiti stavku koja nema te karakteristika. Drugim riječima, nije prisutan faktor iznenađenja (Ricci i sur., 2015, 39-40)<sup>31</sup>.

---

<sup>29</sup> Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.

<sup>30</sup> Ibid.

<sup>31</sup> Ibid.

### 2.2.3. Računanje preporuka

Prvi korak kod kolaborativnog filtriranja je normalizacija ocjena koje su korisnici dali stavkama. Zbog tendencija korisnika da stavkama daju samo visoke ili niske ocjene, vrijednosti ocjena je potrebno normalizirati kako bi imale manju težinu i bile uravnotežene sa ostalim ocjenama. Normalizacija se izvodi računanjem prosjeka svih ocjena te se individualne ocjene oduzimaju od prosjeka (Raman, 2022)<sup>32</sup>.

Sljedeći korak je računanje sličnosti. Za izračun sličnosti između stavke  $i$  i stavke  $j$  za kolaborativno filtriranje orijentirano prema stavkama, prvo se uzimaju korisnici koji su ocijenili stavku  $i$  i stavku  $j$  te se njih primjenjuje mjera sličnosti, a za izračun sličnosti za kolaborativne sustave orijentirane prema korisniku prvo se računa mjera sličnosti između korisnika koji su ocijenili iste stavke. Za kraj, koriste se ocjene najbližijih korisnika za računanje predviđene ocjene stavke koristeći formulu:

$$p_{a,i} = \frac{\sum_{j \in k} r_{aj} W_{ij}}{\sum_{j \in k} W_{ij}}$$

U kojoj je  $k$  skup najbližijih stavki koje je korisnik  $a$  ocijenio, a  $w(i,j)$  je sličnost između stavki  $i$  i  $j$  (Raman, 2022)<sup>33</sup>.

	<b>Film A</b>	<b>Film B</b>	<b>Film C</b>	<b>Film D</b>
<b>Korisnik A</b>	3	4	2	?
<b>Korisnik B</b>	4	4	3	5
<b>Korisnik C</b>	3	4	?	4
<b>Korisnik D</b>	?	4	5	1

Tablica 1. Tablica korisnosti.

Tablica 1 prikazuje tablicu korisnosti u kojoj su zapisani podaci o ocjenama koje su korisnici A, B, C i D dali filmovima A, B, C i D. Upitnik označava filmove koje korisnici nisu ocijenili.

<sup>32</sup> Raman, K. (2020). *Matrix Factorization Explained | What is Matrix Factorization?* Great Learning. <https://www.mygreatlearning.com/blog/matrix-factorization-explained/> (Pristupljeno: 27. svibnja 2022.)

<sup>33</sup> Ibid.

Prvi korak kod računanja preporuka u kolaborativnom filtriranju je normalizacija ocjena. Prosjek svih ocjena za Korisnika A iznosi 3, za Korisnika B iznosi 4, za Korisnika C iznosi 3,66, a za Korisnika D iznosi 3,33. Za dobivanje normaliziranih ocjena, prosjek je potrebno oduzeti od svake ocjene tog korisnika. Na primjer, normalizirana ocjena Korisnika A za Film A se računa kao  $3-3$ , odnosno normalizirana ocjena iznosi 0. Za Film B Korisnika B normalizirana ocjena iznosi  $-1$ , odnosno računa se kao  $3-4$ . Tablica 2 prikazuje normalizirane ocjene korisnika A, B, C i D.

	<b>Film A</b>	<b>Film B</b>	<b>Film C</b>	<b>Film D</b>
<b>Korisnik A</b>	0	1	-1	?
<b>Korisnik B</b>	1	1	0	2
<b>Korisnik C</b>	-0,67	0,33	?	0,33
<b>Korisnik D</b>	?	0,67	1,67	-2,33

Tablica 2. Tablica korisnosti s normaliziranim ocjenama.

Za izračun predviđene ocjene Korisnika A za Film D, potrebno je usporediti vektor Korisnika A sa vektorima ostalih korisnika primjenom formule za kosinusnu mjeru sličnost, pri čemu prazne vrijednosti zamjenjujemo s nulom. Sličnost između Korisnika A i Korisnika B iznosi 0,7071678, između Korisnika A i C 0,22627417, a između Korisnika A i D -0,38735309. Ubacivanjem najbližijih korisnika, odnosno korisnika B i C, u formulu za računanje predviđene ocjene dobiva se ocjena 4.

### 2.3. Druge vrste sustava za preporuku

Ricci i sur. (2015)<sup>34</sup> kao druge manje korištene vrste sustava za preporuku navode demografske sustave za preporuku (eng. demographic recommender system), filtriranje temeljeno na znanju (eng. knowledge-based recommender system) te hibridne sustave za preporuku.

Demografski sustavi za preporuku preporuke filtriraju prema demografskim značajkama o korisniku. Polazi se od stajališta da različite demografske skupine će htjeti različite preporuke.

<sup>34</sup> Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.

Filtriranje temeljeno na znanju koristi znanje iz domene o tome kako određena stavka odgovara korisnicima i kako je stavka korisna korisniku. Pri računanju sličnosti dobivena vrijednost izražava korisnost stavke za korisnika. Sakupljaju se informacije o korisničkim zahtjevima te su preporuke objašnjene korisnicima. Prema Ricci i sur. (2015)<sup>35</sup>, podvrste filtriranja temeljenog na znanju su:

- Sustavi na temelju ograničenja (eng. constraint-based systems) koji su u suštini slični filtriranju temeljenom na znanju, ali umjesto da se preporuke računaju na sličnosti, one se računaju pomoću postojećih baza podataka sa pravilima o svojstvima stavki;
- Sustavi temeljeni na zajednici (eng. community-based systems) koji preporučuju stavke na osnovi korisnikovih prijatelja (str. 14).

Hibridni sustavi za preporuku su sustavi koji kombiniraju neke od prethodno navedenih vrsta filtriranja koristeći prednosti jedne vrste sustava za uklanjanje nedostatka druge vrste sustava.

---

<sup>35</sup> Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.

### 3. Evaluacija sustava za preporuku

Široko korištenje sustava za preporuku na raznovrsnim domenama na internetu je prouzročilo potrebu za razvitkom realističnih i točnih metoda za evaluaciju sustava za preporuku. Za učinkovitu evaluaciju potrebno je odabrati najprikladniju tehniku za evaluaciju sustava, uzeti u obzir domenu u kojoj se sustav koristi, te u čemu je određena tehnika evaluacije bila uspješna. Evaluacija sustava za preporuku se tradicionalno odvija izvan mrežno koristeći postojeće baze podataka aktivnosti korisnika (Jannach i sur., 166)<sup>36</sup>.

#### 3.1. Metode evaluacije

Chen i Liu (2017)<sup>37</sup> izdvajaju tri klasične metode evaluacije sustava za preporuku kroz koje novi sustav za preporuku mora biti evaluiran prije nego je službeno stavljen u upotrebu:

- Izvan mrežna analitika (eng. offline analytics);
- Korisnička studija (eng. user study);
- Testiranje na mreži (eng. online experiment) (str. 1247-1248).

Izvan mrežna analitika je najjeftinija i najlakša metoda za implementaciju. Podaci za bazu podataka koja se koristi u ovoj metodi se unaprijed prikupljaju. Ti podaci uključuju korisnikove odabire i ocjene određenih stavki te se koriste za simulaciju interakcija između korisnika i sustava za preporuku. Mogu se prikupljati iz nasumično iz zapisa aktivnosti korisnika, ili se uzima cjelokupni zapis aktivnosti korisnika u određenom vremenskom razdoblju. U sljedećem koraku ti se skupovi podataka dijele na skupove podataka za treniranje i skupove podataka za testiranje sustava za preporuku. Prednost ovog pristupa je da ne zahtjeva interakciju sa stvarnim korisnicima, što čini njegovu implementaciju povoljnom, te je moguće evaluirati razne vrste sustava za preporuku. S druge strane, evaluacija je ograničena na evaluaciju točnosti predviđanja ili preciznosti rangiranja stavki.

---

<sup>36</sup> Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.

<sup>37</sup> Chen, M. (2017). Performance Evaluation of Recommender Systems. *International Journal of Performability Engineering*, 13(8), 1246–1256. <https://doi.org/10.23940/ijpe.17.08.p7.12461256>

Korisnička studija zahtjeva angažiranje ispitivača koji će aktivno koristiti sustav za preporuku, ispunjavati unaprijed postavljene zadatke, promatrati i zapisivati svoje ponašanje, primjerice, koje su zadatke ispunili i koliko im je vremena trebalo. Ovom metodom testira se interakcija korisnika sa sustavom za preporuku. Za razliku od izvan mrežne analitike, moguće je dobiti povratnu informaciju o kompleksnosti zadataka, korisničkom sučelju, i slično. Nedostatak ovog pristupa je da je izrazito skup, s obzirom da zahtjeva angažiranje velikog broja ispitivača koji moraju ispuniti veliki broj zadataka. Također zahtjeva da su ispitivači imaju slične karakteristike onima u stvarnom sustavu, tj. imaju slične interese, hobije, slične dobi, sličnog udjela ženskih i muških osoba, i tako dalje.

U testiranju na mreži provodi se opsežan eksperiment uz nasumično odabrane korisnike na sustavu za preporuke koji se već koristi. Ova metoda omogućuje evaluaciju ili usporedbu sustava za preporuke u stvarnim zadacima koje ispunjavaju stvarni korisnicu sustava, a dobivenu rezultati su najrealniji. Prednost testiranja na mreži je mogućnost evaluacije učinkovitosti svih aspekta sustava za preporuku, poput dugoročne profitabilnosti sustava i zadržavanja korisnika, umjesto evaluacije pojedinih metrika. S druge strane, zbog mogućnosti lošeg radnog učinka sustava za preporuku koji se testiraju, moguć je gubitak povjerenja korisnika u sustav za preporuku.

### 3.2. Metrike evaluacije

Za računanje točnosti preporuka Silvera, Zhang, Lin, Liue i Ma (2017)<sup>38</sup> navode da se koriste se metrike greške (eng. error metrics), primjerice srednja apsolutna greška (eng. Mean Absolute Error – MAE) čija formula glasi:

$$MAE = \frac{\sum_{i \in R_u} p(i) - r(i)}{|R_u|}$$

Za koju je  $i$  stavka,  $R_u$  popis svih preporuka za korisnika  $u$ ,  $p$  je stvarna ocjena korisnika, a  $r$  ocjena koju je sustav za preporuke predvidio. Za računanje velikih pogrešaka u predviđanju

---

<sup>38</sup> Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2017). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5), 813–831. <https://doi.org/10.1007/s13042-017-0762-9>

ocjena koristi se korjenovana prosječna kvadrirana greška (eng. Root Mean Squared Error – RMSE).

$$RMSE = \sqrt{\frac{\sum_{i \in R_u} (p(i) - r(i))^2}{|R_u|}}$$

Preciznost (eng. precision) mjeri stopu preporučenih stavki koje su se korisniku svidjele, odnosno koje je konzumirao. Računa se pomoću sljedeće formule:

$$preciznost = \frac{|C_u \cap R_u|}{R_u}$$

U kojoj je  $C_u$  broj svih preporučenih stavki koje je korisnik  $u$  konzumirao, a  $R_u$  popis svih preporuka za korisnika  $u$ .

Odzivom (eng. recall) se računa mjera konzumacije preporučenih stavki  $R_u$  u odnosu sa stavkama  $C_u$  koje je korisnik  $u$  konzumirao:

$$odziv = \frac{|C_u \cap R_u|}{C_u}$$

Metrika R-rezultat (eng. R-score) se koristi u slučajevima kad je potrebno u obzir uzeti rangiranje preporuka te se veća vrijednost pridaje stavkama na vrhu ranga. R-rezultat se računa formulom:

$$rang(R_u) = \sum_{j=1}^{|R_u|} \frac{\max(r(i_j) - d, 0)}{2 \frac{j-1}{\alpha-1}}$$

U kojoj je  $R_u$  popis svih preporuka za korisnika  $u$ ,  $r(i,j)$  je ocjena stavke  $i$  u rangu,  $d$  je medijan ocjene, a  $\alpha$  je vrijednost opadanja rangiranja.

Mjera klikanja (eng. click-through-rate – CTR) se koristi za računanje udjela preporučenih stavki na koje je korisnik  $u$  kliknuo  $C_u$  u odnosu sa svim preporučenim stavkama  $R_u$  te se računa pomoću formule (str. 815-816):

$$CTR = \frac{C_u}{R_u}$$

Osim prethodno navedenih metrika koje se koriste za mjerenje korisnosti (eng. utility), tj. preciznosti sustava, Silvera i sur. (2017) izdvaja metrike za mjerenje noviteta (eng. novelty), raznolikosti (eng. diversity), neočekivanosti (eng. unexpectedness), faktora iznenađenja (eng.

serendipity) i pokrivenosti (eng. coverage) sustava za preporuku (str. 815). Za odabir odgovarajuće metrike za evaluaciju Schröder, Thiele i Lehner (2011)<sup>39</sup> navode da je potrebno uzeti u obzir sljedeće:

- U slučaju da su stavke implicitno ocijenjene, odnosno da ne postoji razlika između ocijenjenih i neocijenjenih stavki, metrike točnosti predviđanja se ne mogu koristiti, s obzirom da ne postoje neocijenjene stavke za koje je moguće predvidjeti ocjenu i izmjeriti točnost predviđanja.
- Ako su stavke ocijenjene binarno, preporučeno je koristiti klasifikacijsku metriku ili metriku rangiranja.
- U slučaju da su korisnicima bitne samo najviše ili najniže rangirane stavke, umjesto pojedinačnih ocjena za svaku stavku, ovdje je također preporučeno koristiti klasifikacijsku metriku ili metriku rangiranja.
- Kod ograničenog popisa najviše rangiranih stavki, nije prikladno koristiti metrike koje mjere sveukupnu točnost predviđanja i mjere koje mjere sveukupnu točnost rangiranja.
- Pri prikazivanju velikog broja preporuka odjednom, bitan je redoslijed u kojem će se prikazivati preporuke. U ovom slučaju, nije dostatno koristiti osnovne metrike poput preciznosti i odaziva te je potrebno koristiti metriku koja uzima u obzir redoslijed preporučenih stavki.
- Potrebno je uzeti u obzir brzinu kojom korisnik gubi interes u niže rangirane stavke. Ako korisnici rijetko gledaju rangiranje nisko rangiranih stavki, klasifikacijska greška ili pogreška u rangiranju kod niže rangiranih stavki nije bitna.

---

<sup>39</sup> Schröder, G., Thiele, M. & Lehner, W. (2011). *Setting Goals and Choosing Metrics for Recommender System Evaluations*.



## 4. Primjene sustava za preporuku na internetu

U ovom poglavlju pobliže će se razmotriti sustavi za preporuku koji su prisutni na platformi za digitalno strujanje video sadržaja *Netflix* te na usluzi za strujanje glazbe *Deezer*.

### 4.1. Netflix

*Netflix* je američka tvrtka osnovana 1997. godine koja nudi prikaz filmova u stvarnom vremenu (eng. streaming filmova) i TV emisija uz plaćanje mjesečne pretplate. U početku, *Netflix* je nudio usluge posudbe DVD diskova putem pošte, a *Netflixova* platforma za online streaming je uvedena 2007. godine. S obzirom da *Netflix* ne sadrži reklame, njegov profit proizlazi iz plaćanja mjesečne pretplate. Prema Rešetar i Seljan (2021)<sup>40</sup>, *Netflix* uvodi 2000. godine sustave za preporuku koji ih značajno razlikuje od konkurencije. Procjenjuje se da je njihov sustav za preporuku odgovoran za 80% sati gledanja sadržaja na platformi te se tako ističe kao važan dio modela zarade platforme *Netflix* (Spandana, 2020)<sup>41</sup>. Krysik (2021)<sup>42</sup> navodi da *Netflix* koristi razne modele za filtriranje podataka, uključujući kontekstno utemeljene sustave za uspoređivanje individualnih stavki kao što su opis, komentari i recenzije, i kolaborativno filtriranje za usporedbu ponašanja i sklonosti korisnika. Naknadno sortira filtrirane podatke i prikazuje ih na početnoj *stranici* *Netflix*a. Razlikujemo nekoliko vrsta poredaka koje *Netflix* koristi:

- Personalizirano rangiranje videa (eng. personalised video ranking – PVR) – algoritam razdvaja sadržaj po svojstvu (npr. znanstveno fantastični filmovi, akcijski filmovi, komedije, itd.) i spaja ih s odgovarajućim korisnikom prema njegovom ukusu.

---

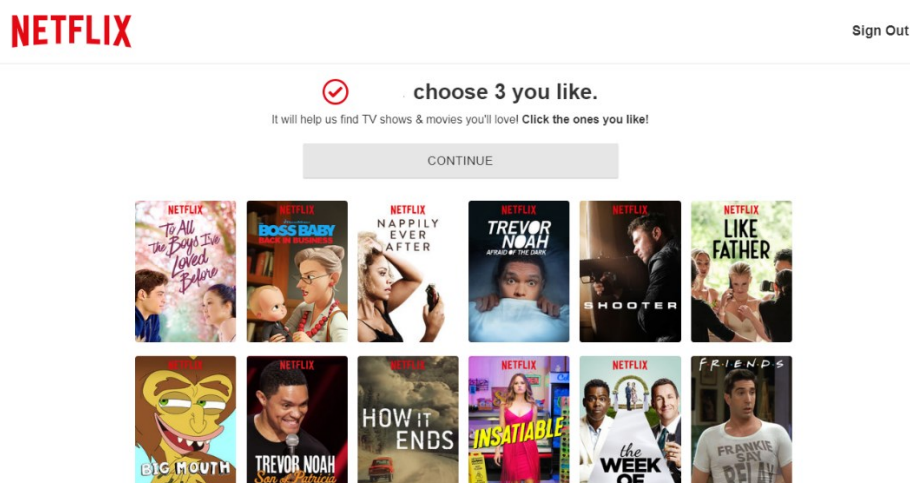
<sup>40</sup> Rešetar, J., Seljan, M. *The Case study of Netflix. Company analysis and risk management strategies in the global business environment – a case study collection.* (urednik: Danijela Miloš Sprčić), 2021, 279-318.

<sup>41</sup> Spandana, S. (2020). *Why Am I Seeing This? New America.* <https://www.newamerica.org/oti/reports/why-am-i-seeing-this/case-study-netflix/> (Pristupljeno: 1. svibnja 2022.)

<sup>42</sup> Krysik, A. (2021). *Netflix Algorithm: Everything You Need to Know About the Recommendation System of the Most Popular Streaming Portal.* Recostream. <https://recostream.com/blog/recommendation-system-netflix> (Pristupljeno: 1. svibnja 2022.)

- Rangiranje najboljeg sadržaja (eng. top-N ranking) – određuje najpopularniji sadržaj u određenoj kategoriji uzimajući u obzir samo najviše rangirani sadržaj individualnog korisnika.
- Rangiranje zanimljivog sadržaja za nastavak gledanja – provjerava sadržaj koji je korisnik nije završio gledati i računa vjerojatnost nastavka gledanja tog sadržaja te ga rangira. Algoritam uzima u obzir parametre kao što su uređaj kojim se sadržaj gledao, prije koliko vremena je korisnik gledao sadržaj, gdje je stao i slično.
- Rangiranje popularnih filmova – algoritam analizira koja vrsta sadržaja je popularna u vrijeme gledanja. U obzir se uzimaju razni vremenski trendovi kao što su važni svjetski događaji (npr. u vrijeme pandemije koronavirusa, algoritam će preporučiti filmove povezane sa pandemijama) i praznici i slične prigode (primjerice, u prosincu algoritam će preporučiti božićne filmove).

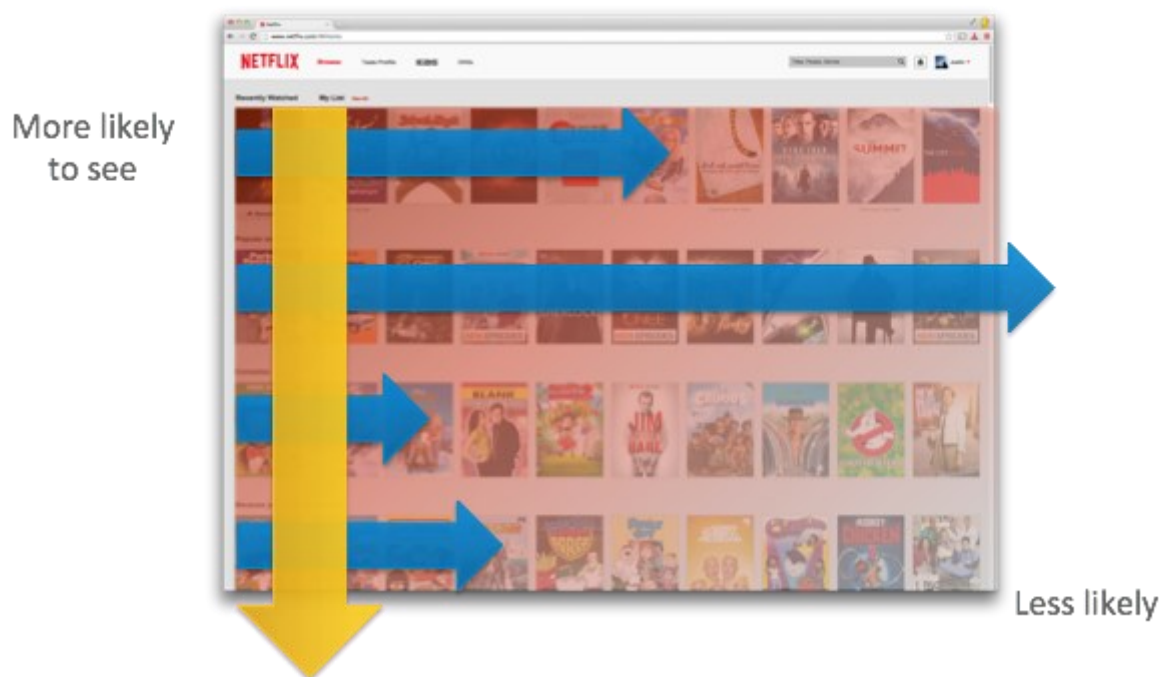
Kako bi se izbjegnulo problem hladnog starta, pri stvaranju novog korisničkog računa, *Netflix* traži korisnika da odabere nekoliko naslova koji se korisniku sviđaju (Slika 5). Odabir naslova nije obvezan te u slučaju da korisnik ne odabere nijedan naslov, korisnikova početna stranica prikazuje raznoliki set popularnih naslova. Nakon što korisnik započne koristiti uslugu, sustav započinje preporuku temeljenu prema povijesti gledanja umjesto prema prethodno odabranim naslovima. S vremenom, sadržaj u nedavnoj povijesti gledanja postaje važniji od starijeg sadržaja te će sustav za preporuku se temeljiti na novijem sadržaju (Netflix, 2022)<sup>43</sup>.



Slika 5. Rješenje za problem hladnog starta implementiran na platformi Netflix. (Preuzeto sa: <https://thebusylady.com/how-to-sign-up-for-netflix/>)

<sup>43</sup> Netflix. (2022). *How Netflix's Recommendations System Works*. Netflix Help Center. <https://help.netflix.com/en/node/100639> (Pristupljeno: 7. svibnja 2022.)

Primarna funkcija početne stranice na platformi *Netflix* je omogućiti korisniku jednostavno pronalaženje sadržaja koji će mu se svidjeti. Stoga, je početna stranica je primarni način interakcije korisnika sa preporukama. S obzirom da *Netflix* u svojem katalogu ima znatno veći broj sadržaja nego što je moguće prikazati, potrebno je prilagoditi početnu stranicu svakom korisniku kako bi bila relevantna, u skladu sa korisnikovim interesima i namjerama, a istovremeno da omogućuje istraživanje kataloga. Na ovaj problem nailaze sve web stranice na kojima je potrebno koherentno prikazati manji broj stavki iz većeg skupa. Početna stranica je organizirana u redove, svaki od kojih je smisljeno nazvan, kako bi korisnici brzo odlučili postoji li mogućnost da se njima zanimljiv sadržaj nalazi u određenom setu. U protivnom, preskaču red te promatraju sljedeći. Sadržaj može biti grupiran po metapodacima (npr. žanr, godina prikazivanja) ili kao rezultat sustava za preporuku. Kako bi svi redovi bili relevantni za korisnika, razvijen je algoritam za personalizaciju početne stranice prema korisniku (Alvino i Basilico, 2015)<sup>44</sup>.



Slika 6. Navigacija korisnika početnom stranicom (Alvino i Basilico 2015)<sup>45</sup>.

<sup>44</sup> Alvino, C., & Basilico, J. (2015). *Learning a Personalized Homepage - Netflix TechBlog*. Medium. <https://netflixtechblog.com/learning-a-personalized-homepage-aa8ec670359a> (Pristupljeno: 7. svibnja 2022.)

<sup>45</sup> Ibid.

Pri stvaranju početne stranice, potrebno je uzeti u obzir pozicije na kojima se najviše provodi interakcija. To je sadržaj koji se nalazi u gornjem lijevom kutu te ako je on relevantan, smanjuje se vrijeme traženja sadržaja za gledanje (Alvino i Basilio, 2015)<sup>46</sup>. *Netflix* je svjestan da je potrebno brzo pridobiti korisnikovu pažnju kako osoba ne bi izgubila interes te stoga su razvili poseban algoritam za odabir slika koje se prikazuju na početnoj stranici. Algoritam u obzir uzima sadržaj koji će se korisniku koji primjerice gleda filmove sa glumicom Umom Thurman, za film *Pulp Fiction*, prikazati će se slika na kojoj je glumica Uma Thurman, a korisniku koji gleda filmove sa glumcem John Travoltom, za isti film prikazati će se slika na kojoj je John Travolta (Blattmann, 2018)<sup>47</sup>.



Slika 7. Primjer korištenja algoritma za odabir slika (Blattmann 2018)<sup>48</sup>.

## 4.2. Deezer

*Deezer* je francuska platforma za zvučnu reprodukciju u realnom vremenu (eng. streaming platforma audio sadržaja) pokrenuta 2007. godine koja se koristi u više od 180 zemalja diljem svijeta. Njihov katalog se sastoji od više od 90 milijuna pjesama, podcasta, audio knjiga i radijskih kanala ("About Deezer", 2022)<sup>49</sup>. Slično kao i druge platforme za streaming glazbe, *Deezerova* početna stranica se sastoji takozvanih vrtuljaka (eng. swipeable carousels), koji

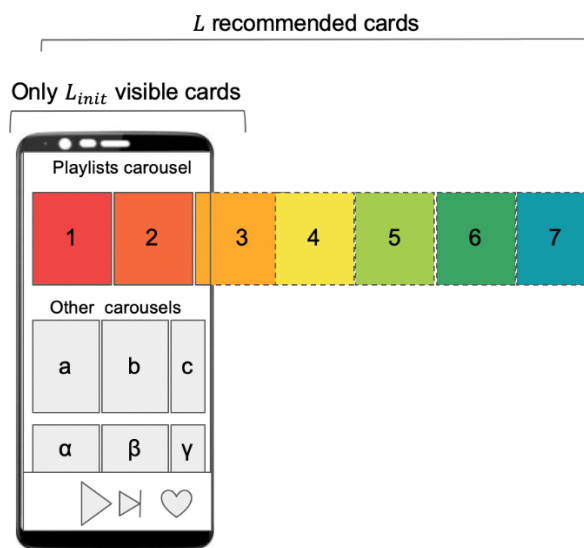
<sup>46</sup> I Alvino, C., & Basilio, J. (2015). *Learning a Personalized Homepage - Netflix* TechBlog. Medium. <https://netflixtechblog.com/learning-a-personalized-homepage-aa8ec670359a> (Pristupljeno: 7. svibnja 2022.)

<sup>47</sup> Blattmann, J. (2018). *Netflix: Binging On The Algorithm*. Medium. <https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59> (Pristupljeno: 7. svibnja 2022)

<sup>48</sup> Ibid.

<sup>49</sup> *About Deezer*. (n.d.). Deezer. <https://www.deezer.com/en/company/about> (Pristupljeno: 19. svibnja 2022.)

prikazuju nekoliko stavki koje se moguće pregledavati povlačenjem preko ekrana. Te stavke mogu biti albumi, popisi za reprodukciju, izvođači i slično. Ti vrtuljci se koriste za prikazivanje preporučenog sadržaja na korisnikovoj početnoj stranici (Bendada, Salha i Bontempelli, 2020)<sup>50</sup>. *Deezerove* beskonačne personalizirane liste za reprodukciju se nazivaju *Protok* (eng. Flow) i omogućuju biranje raspoloženja (npr. zabava, motivacija, opuštanje, itd.). *Deezerove* pametne liste za reprodukciju se sastoje od 40 pjesama koje je sustav odabrao. Sve liste za reprodukciju osim Dnevni i Obiteljskih popisa za reprodukciju ažuriraju se jednom tjedno (Deezer Music, 2022)<sup>51</sup>.



Slika 8. Vrtuljci reprodukcijjskih lista (Bendada i sur., 2020)<sup>52</sup>.

Afchar i sur. (2022)<sup>53</sup> navodi ključne razlike koje je potrebno uzeti u obzir za preporuku glazbe:

- Trajanje stavke je često znatno manje nego od drugih stavki;
- Glazbeni zapisi su složeni što znači da se osim zvučnog zapisa sastoje od MIDI i tekstualnih metapodataka, te sustav za preporuku glazbe (eng. music recommender

<sup>50</sup> Bendada, W., Salha, G., & Bontempelli, T. (2020). Carousel Personalization in Music Streaming Apps with Contextual Bandits. *Fourteenth ACM Conference on Recommender Systems*, 25(420). <https://doi.org/10.1145/3383313.3412217>

<sup>51</sup> *Deezer Music*. (2022). Deezer Support. [https://support.deezer.com/hc/en-gb/articles/201331091-Deezer-Music-\(Pristupljeno: 18. svibnja 2022.\)](https://support.deezer.com/hc/en-gb/articles/201331091-Deezer-Music-(Pristupljeno: 18. svibnja 2022.))

<sup>52</sup> Bendada, W., Salha, G., & Bontempelli, T. (2020). Carousel Personalization in Music Streaming Apps with Contextual Bandits. *Fourteenth ACM Conference on Recommender Systems*, 25(420). <https://doi.org/10.1145/3383313.3412217>

<sup>53</sup> Afchar, D., Melchiorre, A.B., Schedl, M., Hennequin, R., Epure, E.V., & Moussallam, M. (2022). *Explainability in Music Recommender Systems*. <https://doi.org/10.48550/arXiv.2201.10528>.

system – MRS) mora u obzir uzeti i naslovnice albuma, glazbeni spot i skupljene povratne informacije o korisniku;

- Slušatelji reproduciraju drugačije pjesme ovisno o raspoloženju, lokaciji i socijalnoj situaciji u kojoj se nalaze;
- Glazba se sluša u sesijama tako da je potrebno pri preporuci se fokusirati na kreiranje popisa za reprodukciju ili automatski nastavak trenutnog popisa za reprodukciju.

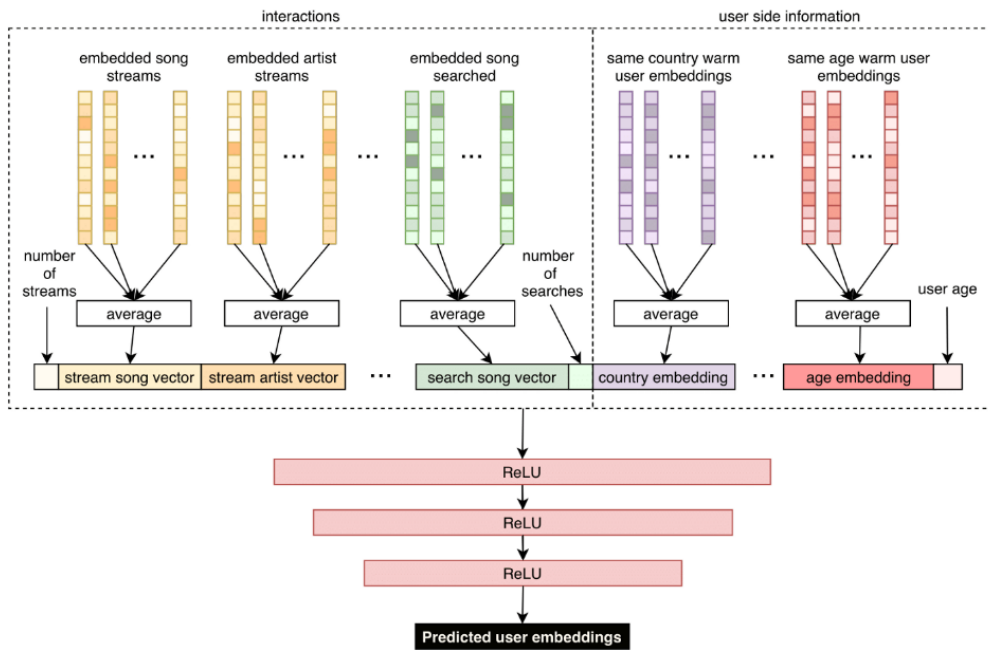
Sustav za preporuku usluge *Deezer* je kombinacija kontekstno utemeljenog i kolaborativnog filtriranja. Prati koje pjesme se korisniku sviđaju, koje je ponovno poslušao te koje je preskočio. Za utvrđivanje raspoloženja pjesme koristi velike podatke sa platforme *Last.fm* koja se koristi za praćenje konzumacije i preporuku glazbu (Antal, 2021)<sup>54</sup>. Pri kreiranju sustava za preporuku koriste se razne metode koje uključuju latentne faktorske modele i grafičke tehnike rudarenja, i tehnike utemeljene na dubokom učenju. Koriste se i tehnike iz obrade prirodnog jezika i procesiranja zvučnog signala da bi se glazba anotirala ili prikazala u vektorskom prostoru (Afchar i sur., 2022)<sup>55</sup>. *Deezer* je posvetio pozornost problemu hladnog starta i razvio sustav kojim novi korisnici dobivaju djelomično personalizirane preporuke prije kraja prvog dana korištenja platforme. Latentni modeli koje *Deezer* koristi za svoje vrtuljke se treniraju samo jednom tjedno, a novi korisnici ne smiju čekati nekoliko dana do personaliziranog sadržaja. Novim korisnicima ne smiju se preporučiti loše preporuke ni samo ljestvice popularnosti. Za ovaj problem, *Deezer* je razvio neuronsku mrežu s tri skrivena sloja koja koristi podatke prikupljene na dan registracije starijih korisnika Deezer-a koji su već prikazani u vektorskom prostoru (Slika 9). Koriste se informacije o zemlji i godini rođenja, interakcije sa glazbom npr. koja im se glazba sviđala i koja im se nije sviđjela, što su pretraživali, i slično. Koristeći ove podatke te poziciju starijih korisnika na vektoru, ova neuronska mreža prikazuje nove korisnike u vektorskom prostoru te preporuča pjesme koje slušaju susjedi na vektoru (Briand, 2021)<sup>56</sup>.

---

<sup>54</sup> Antal, D. (2021). *Listen Local: Why We Need Alternative Recommendation Systems*. Daniel Antal, CFA. <https://danielantal.eu/post/2020-12-15-alternative-recommendations/> (Pristupljeno: 18. svibnja 2022.)

<sup>55</sup> Afchar, D., Melchiorre, A.B., Schedl, M., Hennequin, R., Epure, E.V., & Moussallam, M. (2022). *Explainability in Music Recommender Systems*. <https://doi.org/10.48550/arXiv.2201.10528>.

<sup>56</sup> Briand, L. (2021). *Recommending music to new users - Deezer I/O*. Medium. <https://deezer.io/recommending-music-to-new-users-182263648b82> (Pristupljeno: 18. svibnja 2022.)



Slika 9. Deezeova neuronska mreža (Briand, 2021)<sup>57</sup>.

<sup>57</sup> Briand, L. (2021). *Recommending music to new users - Deezer I/O*. Medium. <https://deezer.io/recommending-music-to-new-users-182263648b82> (Pristupljeno: 18. svibnja 2022.)

## **5. Stavovi mladih o sustavima za preporuke**

U praktičnom dijelu rada provodit će se istraživanje o stavovima mladih prema sustavima za preporuku na internetu. Postavljena pitanja biti će o sustavima za preporuke općenito, interakciji s preporučenim sadržajem te će se dio pitanja odnositi na preporučene sadržaje na uslugama digitalnog strujanja, s obzirom da je u takvim uslugama sustav za preporuke u prvom planu i interakcija s preporučenim sadržajem je omogućena na početnoj stranici.

### **5.1. Cilj i pretpostavke istraživanja**

Cilj istraživanja je ustanoviti mišljenja mladih u dobi od 16 do 30 godina starosti o sustavima za preporuku na internetu, utvrditi koliko su upoznati sa sustavima za preporuke, postoje li određene internet usluge koje koriste isključivo zbog preporuka, te istražiti do koje mjere je preporučeni sadržaj relevantan za njih.

Istraživanje polazi od sljedećih pretpostavka:

1. Gotovo svi mladi ljudi su svjesni postojanja sustava za preporuku;
2. Mladi ljudi smatraju preporuke koje dobivaju na internetu su u skladu s njihovim interesima;
3. Sustavi za preporuke pomažu mladima otkriti nove interese;
4. Mladi nisu svjesni svih podataka koje sustavi za preporuku skupljaju;
5. Mladi ljudi koriste usluge za strujanje zbog njihovih preporuka.

### **5.2. Metodologija istraživanja**

Istraživanje je provedeno u razdoblju od 2. do 14. lipnja 2022. Obuhvaćen je slučajni uzorak studenata koji je dobiven distribucijom ankete putem društvene mreže *Facebook*. Anketni upitnik je distribuiran u nekoliko *Facebook* grupa koje koriste većinom mladi (grupe



studentskih domova u Zagrebu, grupe za distribuciju anketa, grupe za ponudu i potražnju poslova).

Anketa se sastoji od 34 različitih tipova: pitanja jednostrukog i mnogostrukog odgovora, pitanja oblikovana Likertovom ljestvicom te pitanja otvorenog tipa. Na pitanja, tj. tvrdnje oblikovane Likertovom ljestvicom ispitanik je mogao procijeniti slaganje s određenom tvrdnjom na skali od 1 do 5, pri čemu 1 označava "uopće se ne slažem", 2 "donekle se ne slažem", 3 "niti se slažem, niti se ne slažem", 4 "donekle se slažem", a 5 "u potpunosti se slažem". Prva skupina pitanja (1-4) odnosi se na osnovne podatke o ispitaniku: spol, dob, status na tržištu rada te ispitanikova povezanost sa IT industrijom. Pitanje o dobi ispitanika je ujedno i eliminacijsko pitanje – u slučaju da je ispitanik odabrao opcije *Mlađi od 16* ili *Stariji od 30*, anketa bi za korisnika bila završena.

Druga grupa pitanja propitkuje ispitanikove (5-7) navike pri korištenju interneta: količina vremena provedena na internetu dnevno, koje društvene mreže i koje digitalne usluge strujanja ispitanik koristi.

Treća grupa pitanja (8-19) sadrži 9 pitanja oblikovanih Likertovom skalom, dva pitanja jednostrukog odgovora te jedno pitanje otvorenog tipa. Ova grupa pitanja propitkuje svjesnost o sustavima za preporuku, ispitanikovu interakciju sa preporučenim sadržajem, pomaže li preporučeni sadržaj pri otkrivanju novih interesa te uzrokuju li sustavi za preporuku osjećaj praćenja. Ovdje se također propitkuje ispitanika jeli ikada upravljao interesima da bi poboljšao preporuke, koristi li određene internet usluge isključivo zbog njihovih preporuka te, u slučaju da ispitanik koristi ovakve usluge, moguće ih je navesti.

Sljedeća, odnosno četvrta grupa pitanja (20-21) odnosi se na glazbene usluge strujanja i propitkuje broj ispitanika koji koristi glazbene usluge strujanja te koji su to usluge. U slučaju da ispitanik odgovori da ne koristi usluge strujanja glazbe, ispitanik ne može pristupiti petoj grupi pitanja nego pristupa šestoj grupi pitanja.

Peta grupa pitanja (22-27) se također odnosi na glazbene usluge strujanja. Sastoji se od 6 tvrdnji oblikovanih Likertovom ljestvicom te propitkuju interakciju s preporučenom glazbom na glazbenim uslugama strujanja, svjesnost o podacima koji se prikupljaju te korištenje drugih usluga u svrhu dobivanja dodatnih preporuka.

Šesta grupa pitanja (28-29) odnosi se na usluge strujanja video sadržaja. Propitkuje se broj ispitanika koji koristi usluge strujanja video sadržaja te koje su to usluge. Ako ispitanik

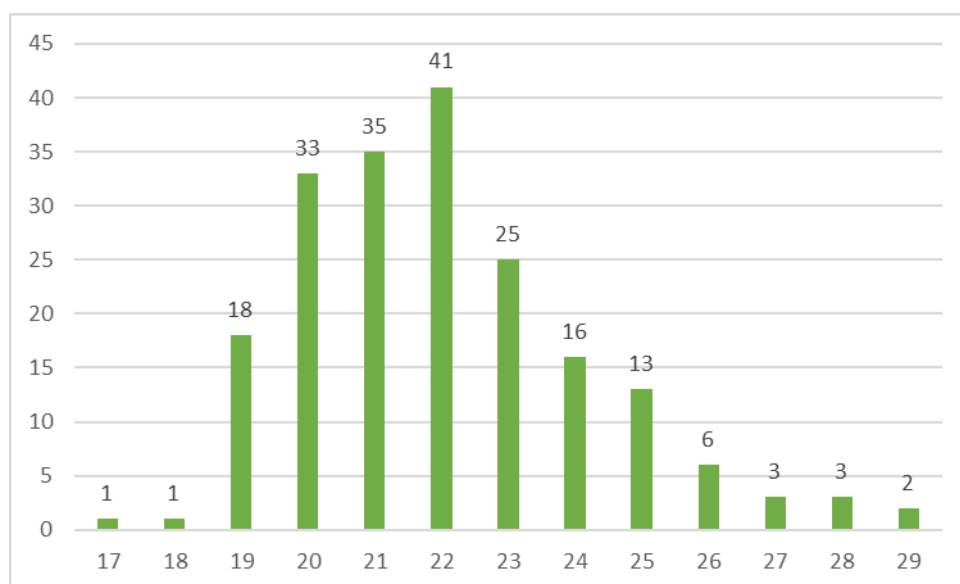
odgovori da ne koristi takve usluge, ispitanik završava anketu. U protivnom, nastavlja na sedmu grupu pitanja.

U posljednjoj, odnosno sedmoj grupi (30-34) pitanja postavljeno je 6 tvrdnji oblikovanih Likertovom ljestvicom. Tvrdnje se odnose na interakciju s preporučenim video sadržajem, svjesnost o podacima koji se prikupljaju te korištenje drugih usluga u svrhu pribavljanja dodatnih preporuka. Ovdje anketa završava. Na kraju je moguće dati povratnu informaciju na ispunjeni anketni upitnik.

### 5.3. Rezultati istraživanja

Anketnom upitniku pristupilo je sveukupno 201 ispitanika, od kojih je 4 (2%) bilo eliminirano zbog neodgovarajuće dobi. Konačan broj ispitanika koji su sudjelovali u anketi nakon eliminacijskog pitanja iznosi 197.

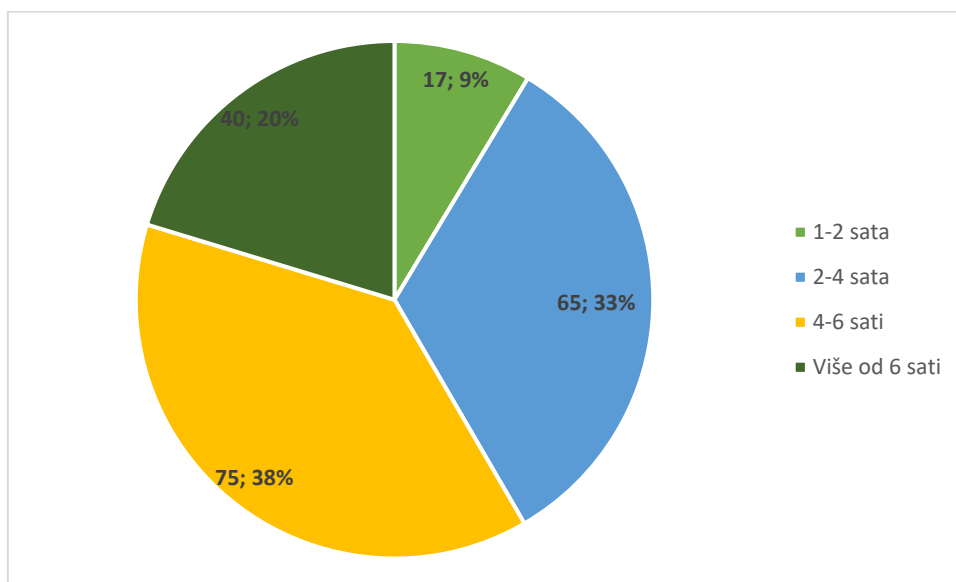
U istraživanju je sudjelovalo 148 ženskih ispitanika (75,1%) i 49 muških ispitanika (24,9%). Slika 10 prikazuje distribuciju dobi ispitanika. Vidljivo je da je najveći broj ispitanika između 19 i 25 godina (91,9%), a najmanje ispitanika 17 i 18 godina (1%). Najveći udio ispitanika čine studenti, kojih ima 184 (94,8%).



Slika 10. Distribucija dobi ispitanika.

Upitniku je pristupilo 10 zaposlenih osoba (5%), 2 učenika (1%) i 1 nezaposlena osoba (0,5%). Njih 68 (34,5%) je izjavilo da studiraju, trenutačno studiraju i/ili su zaposleni u području informatike, računarstva ili informacijskih i komunikacijskih znanosti, dok njih 129 (65,5%) je odgovorilo niječnim odgovorom.

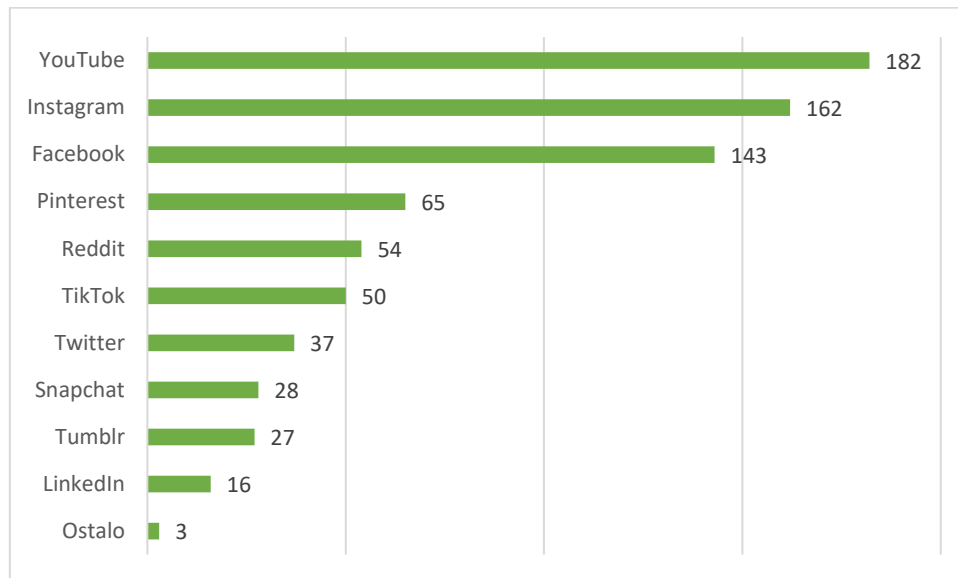
U svrhu pobližnjeg upoznavanja sa navikama korištenja interneta ispitanika, postavljena su 3 pitanja. Na pitanje o količini vremena provedenog na internetu tijekom jednog dana, 75 ispitanika (38,1%) je izjavilo da provodi od 4 do 6 sati na internetu. Njih 65 (33%) provodi od 2 do 4 sata na internetu, dok više od 6 sati na internetu provodi 40 ispitanika (20,3%). Najmanje ljudi, njih 17 (8,6%) je odgovorilo da provodi od 1 do 2 sata na internetu, a niti jedan ispitanik ne provodi manje od 1 sata dnevno.



Slika 11. Količina vremena provedena na internetu dnevno.

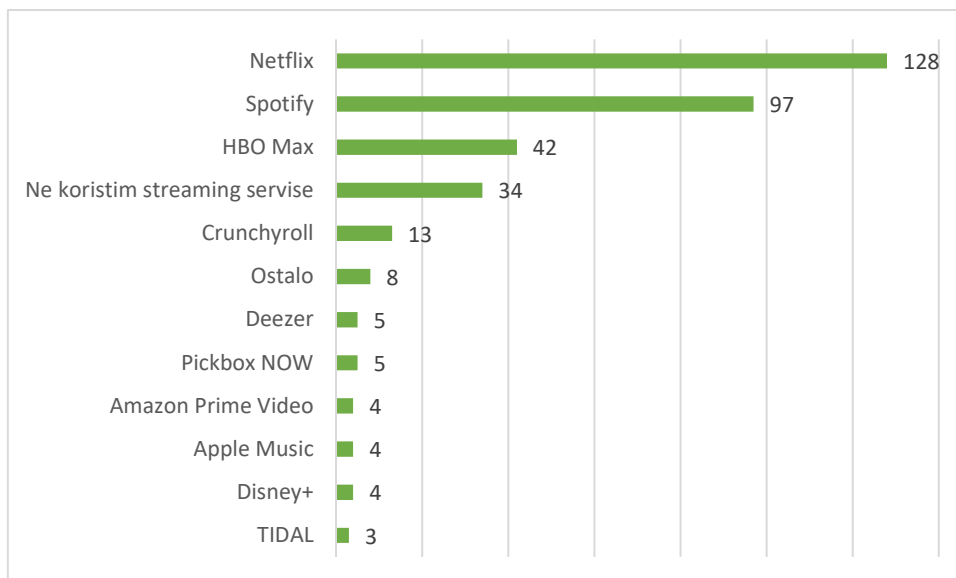
Od društvenih mreža, najveći broj ispitanika koristi *YouTube*, njih 182. Zatim sljedeći je *Instagram* kojeg koristi 162 ispitanika. Sljedeći je *Facebook* kojega koristi 143 ispitanika. Društvene mreže *Pinterest*, *Reddit*, *TikTok*, *Twitter*, *Snapchat*, *Tumblr* i *LinkedIn* se koriste u manjoj mjeri od prethodnih tri. Ispitanici su imali mogućnost navesti i druge nenavede društvene mreže te su

dva korisnika nadopisali da koriste platformu za strujanje prijenosa uživo *Twitch*, a jedan korisnik je naveo društvenu mrežu *4chan*.



Slika 12. Društvene mreže koje koriste ispitanici.

Na upit o korištenju streaming platformi, 128 ispitanika je navelo da koristi *Netflix*, 97 ispitanika koristi *Spotify*, 42 njih *HBO Max*, a 34 ispitanika se izjasnilo da ne koriste niti jednu uslugu strujanja. Ostale ponuđene opcije uključuju *Crunchyroll*, *Deezer*, *Pickbox NOW*, *Amazon Prime Video*, *Apple Music* i *Disney+* te njih koristi manji broj ispitanika. Osmam ispitanika su odabrali opciju ostalo. Dva ispitanika su navela da koriste *YouTube Music*, a ostali odgovori su *YouTube*, *HRTi*, *RTL Play* i *VRV*, te besplatne stranice za strujanje video sadržaja *Gogoanime* i *Movieorca*.



Slika 13. Usluge strujanja koje koriste ispitanici.

Sljedeća grupa pitanja odnosi se na sustave za preporuke te sarži 12 pitanja, od kojih su 9 tvrdnje za koje ispitanici moraju odabrati stupanj slaganja. Tablica 3 donosi prikaz broja odgovora koji odražava različite stupnje slaganja s navedenim tvrdnjama.

	1 – Uopće se ne slažem.	2 – Donekle se ne slažem.	3 – Niti se slažem, niti se ne slažem.	4 – Donekle se slažem.	5 – U potpunosti se slažem.
Svjestan sam sustava za preporuku na internetu.	6	10	39	44	98
Smatram da su preporuke koje dobivam na internetu u skladu sa mojim interesima.	6	31	85	65	10
Smatram da preporuke koje dobivam na internetu mi pomažu otkriti nove interese.	21	59	63	46	8
Reklame na internetu u skladu su sa mojim interesima i potrebama (npr. kad želim kupiti mobitel, razne reklame mi preporučuju mobitele).	21	30	40	72	34

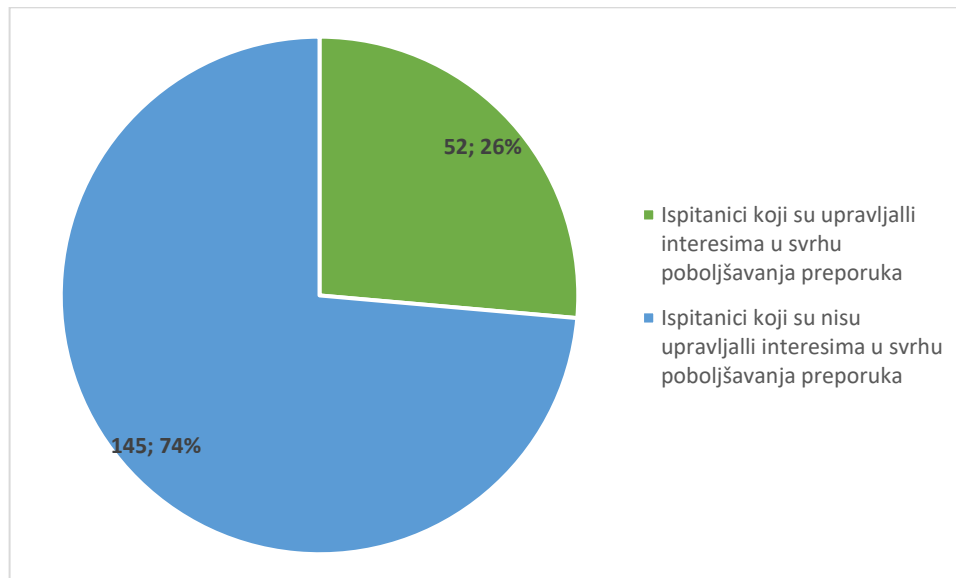
Prilikom online kupovine, kupujem stavke koje mi internet trgovina preporuča.	83	60	37	11	6
Čitam članke koje mi sustav preporuča.	46	60	48	28	15
Zbog sustava za preporuku, provodim više vremena na internetu nego što bih htio.	60	54	29	37	17
Zbog sustava za preporuku određenom sadržaju pristupam u anonimnom načinu.	70	35	29	31	32
Ponekad su preporuke toliko specifične da mislim da me netko prati.	13	20	33	43	88

Tablica 3. Broj odgovora koji odražava različiti stupanj slaganja s tvrdnjama.

Za prvu tvrdnju, 72,1% ispitanika smatra da su svjesni sustava za preporuku na internetu, a prosječni stupanj slaganja na ovu tvrdnju iznosi 4,11. Druga tvrdnja preispituje relevantnost preporučenog sadržaja za ispitanike. Najveći broj ispitanika, njih 85 (43,1%) niti se slaže, niti ne slaže s navedenom tvrdnjom, a prosječni stupanj slaganja iznosi 3,21. Iz svega proizlazi da većina ispitanika je neutralna prema sadržaju koji im je preporučen, ali broj ljudi koji smatra da je preporučeni sadržaj relevantan je veći od broja ljudi koji smatra da je nerelevantan. Na tvrdnju o utjecaju preporučenog sadržaja na otkrivanje novog sadržaja, 40,6% ispitanika je izrazilo da preporučeni sadržaj im ne pomaže otkriti novi interese, a prosječni stupanj slaganja je 2,8. Ovo tvrdnja se odnosi na prisutnost faktora iznenađenja u sustavima za preporuku. Zaključno je da u većini slučajeva sustavi za preporuku ne pomažu mladim ljudima otkriti nove interese te faktor iznenađenja nije dostatno prisutan. Njih 53,8% smatra da su preporučene reklame relevantne za njih, a prosječni stupanj slaganja je 3,35. Stavke koje im internet trgovina preporuča ne kupuje 83,3% ispitanika, a prosječni stupanj slaganja je 1,97. Preporučene članke ne čita 63,8% ispitanika, a prosječni stupanj slaganja iznosi 2,52. Njih 60,3% ne smatra da zbog sustava provode više vremena na internetu nego što bi htjeli, a prosječno su odgovorili s 2,48. S tvrdnjom o pregledavanju sadržaja u anonimnom načinu 57,9% ispitanika se ne slažu, a

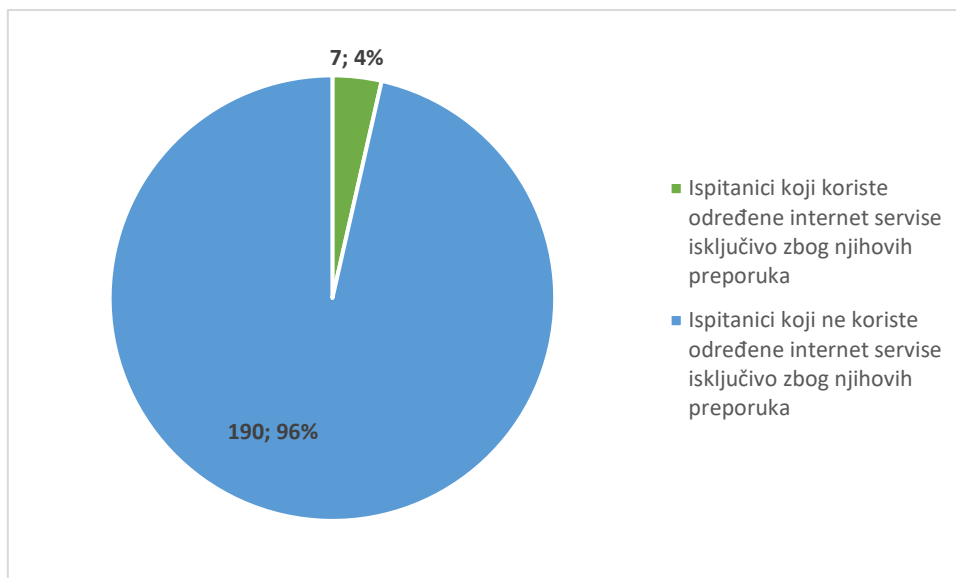
prosječni stupanj slaganja je bio 2,59. Na posljednju tvrdnju o specifičnim preporukama koje izazivaju osjećaj praćenja složilo se 51,3% ispitanika te je prosjek slaganja 3,88.

U svrhu poboljšavanja preporuka, 145 ispitanika (73,6%) nije nikad upravljalo interesima određene web stranice. Mogućnost upravljanja interesima nude usluge poput pretraživača *Google* te se tako može upravljati sa preporukama koje korisniku sustav generira.



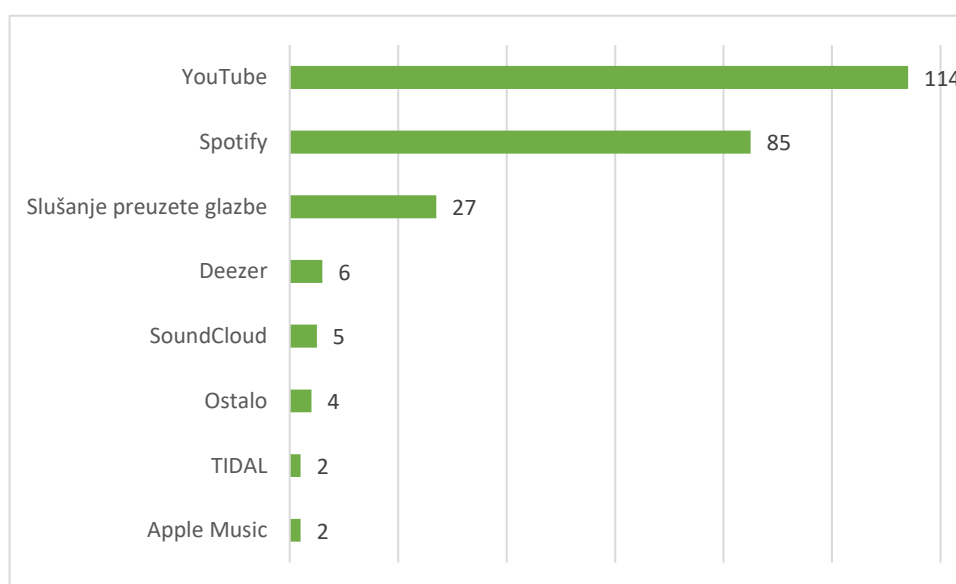
Slika 14. Broj i postotak ispitanika koji su upravljali interesima određene web stranice u svrhu poboljšavanja preporuka.

Samo 3,6% ispitanika je odgovorilo da određene internet usluge koristi isključivo zbog njihovih preporuka. Nastavno na to pitanje, ispitanici koji su odgovorili potvrdno su zamoljeni da nabroje takve usluge te su naveli sljedeće: *AliExpress*, *Google*, *Last.fm*, *TV Time*, *IMDb*, *Letterboxd* i *Pinterest*. Jedan ispitanik izdvojio je Facebook i njegove preporuke događaja, a drugi ispitanik izjavio je sljedeće: „*TikTok* ima odličan algoritam koji pruža specifičan sadržaj koji mi odgovara, a istovremeno ubacuju sadržaj kojeg konzumiraju osobe sa sličnim interesima. Pomoću opcije Ne zanima me (eng. Not interested in this) moguće je smanjiti ili u potpunosti ukloniti sadržaj koji me ne interesira.“



Slika 15. Broj i postotak ispitanika koji koriste određene internet servise isključivo zbog njihovih preporuka.

U nadolazećim pitanjima, naglasak je na preporukama koje korisnici dobivaju na uslugama digitalnog strujanja. U ovom uzorku 154 ispitanika je izjavilo da koristi usluge strujanja glazbe. Kako bi se pobliže ustanovila važnost usluga strujanja pri konzumaciji glazbe, sljedeće pitanje se odnosi na glavne načine slušanja glazbe. *YouTube* je glavni način slušanja glazbe za 114 ispitanika, pa slijedi *Spotify* kojeg koristi 85 ispitanika. Glazbu preuzima 27 ispitanika, a ostali ponuđene usluge (*Deezer*, *SoundCloud*, *TIDAL*, *AppleMusic*) koristi mali broj ispitanika. Ostali načini slušanja koje su ispitanici naveli da koriste su radio, *YouTube Music* i *NewPipe*.



Slika 16. Glazbene usluge strujanja koje koriste ispitanici.



Pristup sljedećoj grupi pitanja bio je omogućen isključivo ispitanicima koji su u prethodnoj grupi pitanja odgovorili da koriste usluge strujanja glazbe. Tablica 4 prikazuje broj odgovora koji odražavaju stupnje slaganja s određenom tvrdnjom.

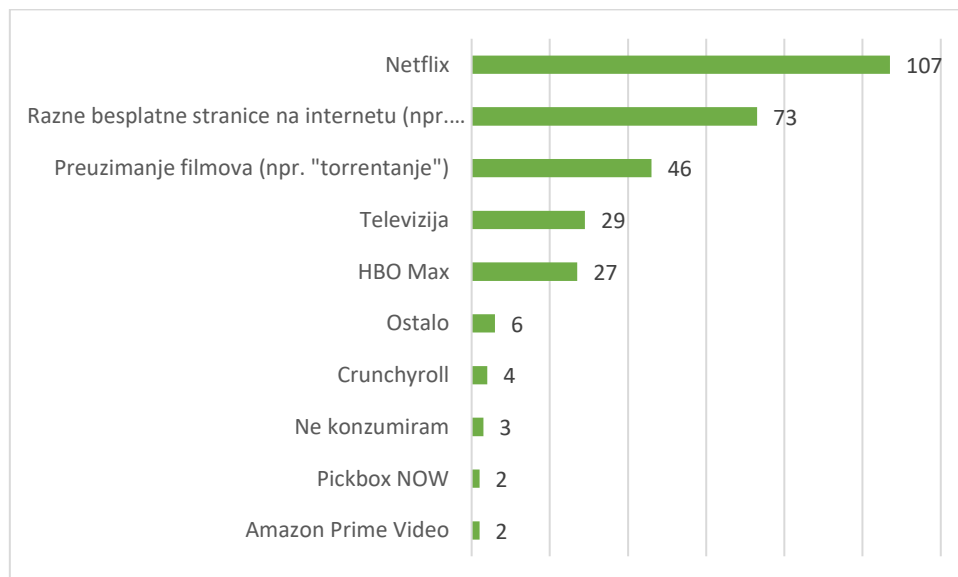
	1 – Uopće se ne slažem	2 – Donekle se ne slažem	3 – Niti se slažem, niti se ne slažem	4 – Donekle se slažem	5 – U potpunosti se slažem
Slušam glazbu koje mi streaming servis preporuča.	14	24	37	53	26
Smatram da je preporučena glazba u skladu s mojim ukusom.	6	13	57	62	16
Često dodajem preporučenu glazbu na vlastite reproduksijske liste.	25	25	26	48	30
Svjestan sam koje podatke sustav za preporuku prikuplja pri konzumaciji glazbe.	26	35	28	25	30
Streamanje glazbe olakšava mi pronalazak nove glazbe koja mi se sviđa.	9	18	38	44	45
Koristim servise koje prate moju konzumaciju glazbe kako bih dobio dodatne preporuke (npr. Last.fm).	77	22	27	14	14

Tablica 4. Broj odgovora koji odražava različiti stupanj slaganja s tvrdnjama..

Tablica 4 ukazuje na sljedeće: 51,2% ispitanika sluša glazbu koje im sustav za preporuku glazbe preporuča, a prosječni stupanj slaganja iznosi 3,34; 50,6% ispitanika smatra da je preporučena glazba u skladu s njihovim ukusom, a prosječna ocjena slaganja je 3,45; 50,6% ispitanika redovito dodaje preporučenu glazbu na vlastite reproduksijske liste, a prosječni stupanj slaganja je 3,21; 39,6% ispitanika nije svjesno koji podaci se prikupljaju u svrhu poboljšanja preporuka te su u prosjeku slaganje ocijenili s 2,99; 57,8% ispitanika usluge strujanja glazbe pomažu pronaći novu glazbu koja im se sviđa, a u prosjeku stupanj slaganja je 3,64; 64,3% ispitanika

koristi usluge koje prate njihovu konzumaciju glazbe u cilju dobivanja dodatnih preporuka, a u ovoj tvrdnji je prosječni stupanj slaganja iznosio 2,13.

Iduća grupa pitanja odnosi se na usluge strujanja video sadržaja poput televizijskih emisija, serija i filmova. Takve usluge koristi 155 ispitanika, a kod upita o glavnim načinima gledanja ovakvog sadržaja njih 107 odabralo je *Netflix*. Besplatne stranice na internetu poput *Filmovizije* koristi 73 ispitanika, dok njih 46 filmove preuzima sa interneta. Televiziju gleda 29 ispitanika, a *HBO Max* njih 27. Usluge strujanja *Crunchyroll*, *Pickbox NOW* i *Amazon Prime Video* koristi mali broj ispitanika, a njih troje ne konzumiraju ovakav sadržaj. Pod opciju ostalo ispitanici su istaknuli *YouTube*, *Google*, snimanje TV sadržaja pomoću *Max TV-a*, *Twitch* i aplikaciju *Stremio*.



Slika 17. Usluge strujanja video sadržaja koje koriste ispitanici.

Posljednju grupu pitanja ispunjavali su isključivo ispitanici koji su u prethodnoj grupi pitanja odgovorili da koriste usluge strujanja video sadržaja. Broj odgovora koji odražavaju stupnje slaganja s određenom tvrdnjom prikazan je u Tablici 5.

	1 – Uopće se ne slažem	2 – Donekle se ne slažem	3 – Niti se slažem, niti se ne slažem	4 – Donekle se slažem	5 – U potpunosti se slažem
Gledam video sadržaj koje mi streaming servis preporuča.	34	32	38	34	17
Smatram da je preporučeni video sadržaj u skladu s mojim ukusom.	23	29	50	44	9
Svjestan sam koje podatke sustav za preporuku prikuplja prilikom konzumacije video sadržaja.	28	32	36	35	24
Streamanje video sadržaja olakšava mi pronalazak novog video sadržaja koji mi se sviđa.	25	30	36	35	24
Koristim servise poput IMDb ili Letterboxd kako bih pratio svoju konzumaciju video sadržaja i dobio dodatne preporuke.	64	24	16	17	34

Tablica 5. Broj odgovora koji odražava različiti stupanj slaganja s tvrdnjama.

Podaci izražavaju sljedeće: preporučeni sadržaj ne gleda 42,6% ispitanika, a prosječni stupanj zadovoljstva iznosi 2,79; preporučeni sadržaj je u skladu s ukusom 34,2% ispitanika, a prosjek slaganja iznosi 2,92; 38,7% ispitanika ne znaju koje podatke sustav za preporuku prikuplja u svrhu poboljšanja tehnologije, a zadovoljstvo iznosi 2,97 u prosjeku; strujanje olakšava pronalazak novog video sadržaja 38% ispitanika, a slaganje je prosječno ocijenjeno s 3,05; usluge za praćenje konzumacije video sadržaja u svrhu pribavljanja dodatnih preporuka ne koristi 56,77% ispitanika.

Na kraju anketnog upitnika bilo je moguće dati povratnu informaciju na ispunjenu anketu te je jedan ispitanik napisao sljedeće: „Nisam primijetio nikakve preporuke u uslugama strujanja koje koristim, osim *YouTube*, koji ovisi o mojoj povijesti gledanja i pretraživanja, odnosno 'pretplatama.'“

## 5.4. Rasprava

Cilj ovog istraživanja je bio utvrditi stavove mladih ljudi prema sustavima za preporuku na internetu te utvrditi njihovu upoznatost s ovakvim sustavima. Također se ispitalo koje usluge koriste isključivo zbog preporuka, ako takve postoje te istražiti relevantnost sadržaja koji im je preporučen. Tri četvrtine ispitanika je bilo ženskog spola, a samo jednu četvrtinu uzorka čine muške osobe. Približno trećina uzorka je izjavila da je da studiraju, trenutačno studiraju i/ili su zaposleni u području informatike, računarstva ili informacijskih i komunikacijskih znanosti. Kod ovih osoba veća je vjerojatnost da su upoznati sa sustavima za preporuke te da imaju više tehničkog znanja o temi istraživanja.

Prema rezultatima istraživanja o navikama ispitanika pri korištenju interneta možemo zaključiti da se uzorak sastoji od aktivnih korisnika interneta. Niti jedan ispitanik nije naveo da koristi internet manje od jednog sata dnevno. Nadalje, svi ispitanici su korisnici društvenih mreža, ali gotovo jedna petina ispitanika ne koristi usluge digitalnog strujanja.

Sa izjavom da su svjesni sustava za preporuku na internetu se donekle slažu ili u potpunosti slažu 142 ispitanika, što čini 72,1% ukupnog uzorka. Ovime je potvrđena prva pretpostavka istraživanja da gotovo svi mladi ljudi su svjesni postojanja sustava za preporuku.

Dva postavljena pitanja su preispitivala znanje ispitanika o podacima koje sustavi za preporuku prikupljaju, jedan od kojih se odnosio na usluge za strujanje glazbe, a drugi na usluge strujanja video sadržaja. Od 144 ispitanika koji su ocijenili svoj stupanj slaganja sa tvrdnjom da su svjesni podataka koje sustav za preporuku prikuplja pri konzumaciji glazbe, 55 ispitanika se donekle složilo ili u potpunosti složilo. Na tvrdnju da su svjesni koje podatke sustav prikuplja prilikom konzumacije video sadržaja 51 ispitanika od 155 koji su pristupili pitanju je odgovorilo da se donekle slaže ili u potpunosti slaže sa tvrdnjom. S ovim tvrdnjama se u prosjeku složilo 35,5% ispitanika te je time potvrđena četvrta pretpostavka da mladi ljudi nisu svjesni svih podataka koje sustavi za preporuku skupljaju na internetu.

Sa tvrdnjom da su preporuke u skladu s njihovim interesima slaže se 75 ispitanika, a najveći broj ispitanika, njih 85, se izjasnilo da se niti slaže niti ne slaže sa tvrdnjom, odnosno da su neutralni. Ukupno 131 ispitanika smatraju da su preporuke ponekad specifične do mjere da izazivaju osjećaj praćenja. Njih 78 od 154 ispitanika koji koriste usluge strujanja glazbe se slaže da je preporučena glazba u skladu s njihovim ukusom, a njih 89 smatra da im strujanje glazbe olakšava otkriti novu glazbu koja im se sviđa. S druge strane, kod strujanja video sadržaja manji

broj ljudi složilo s istim tvrdnjama. Naime, od 155 ispitanika koji su pristupili ovom dijelu upitnika njih 53 se slaže da je preporučeni video sadržaj u skladu s njihovim ukusom, dok njih 59 smatra da im strujanje video sadržaja olakšava pronalazak novog zanimljivog sadržaja. Ovi rezultati djelomično potvrđuju drugu pretpostavku. Drugim riječima, mladi ljudi smatraju da je dio preporuka sa interneta u skladu s njihovim interesima, ali neke platforme će im ponuditi manje relevantne preporuke od drugih platformi.

U istraživanju je djelomično potvrđena i peta pretpostavka. Dobiveni rezultati ukazuju da preporuke u uslugama strujanja glazbe pomažu korisnicima otkriti nove stavke – novu glazbu koja im se sviđa, u kojoj mogu uživati te ju dodati na vlastite reproduksijske liste. No rezultati ne ukazuju na isto za usluge strujanja video sadržaja, s obzirom da 38,1% ispitanika koji koriste usluge strujanja video sadržaja se slažu sa tvrdnjom da im je pronalazak novog sadržaja olakšan korištenjem usluga strujanja video sadržaja. Zaključno je da ispitanici koriste samo određene usluge digitalnog strujanja zbog njihovih preporuka, dok im druge usluge daju relevantne preporuke te na korištenje takvih usluga digitalnog strujanja utječu drugi faktori.

Jedna od prednosti sustava za preporuke poput kolaborativnog filtriranja je takozvani faktor iznenađenja, čija prisutnost omogućuje preporuku relevantnih sadržaja bez postojanja sličnog sadržaja na profilu. Ipak, 80 ispitanika se uopće nije složilo ili donekle nije složilo sa tom tvrdnjom, a njih 63 je neutralno, dok je 54 ispitanika se donekle složilo ili u potpunosti složilo sa ovom tvrdnjom. Ovi rezultati sugeriraju da sustavi za preporuke ne pomažu mladima otkriti nove interese, s obzirom da se samo 27,4% ispitanika složilo s navedenom tvrdnjom, te time treća pretpostavka istraživanja nije potvrđena.

Utvrđeno je da 7 ispitanika, to jest 3.6% uzorka, koristi određene internet usluge isključivo zbog preporuka. Takve usluge uključuju: internetske trgovine poput *AliExpress*, web pretraživač *Google*, platforma za praćenje konzumacije i preporuku glazbe *Last.fm*, filmsku bazu podataka *IMDb*, društvene mreže poput *Pinterest* i *TikTok*, i tako dalje. Takve usluge su raznih domena što ukazuje i na rasprostranjenost sustava za preporuku na internetu.

## 6. ISTRAŽIVANJE

U ovom poglavlju prikazan je praktičan rad u izradi modela sustava za preporuke. Kao alat koristiti će se programski jezik *Python* u distribuciji *Anaconda*. Izrađen su dva modela sustava za preporuke: prvi je kontekstno utemeljeni model, a drugi model utemeljen na kolaborativnom filtriranju.

### 6.1. Preporuka u kontekstno utemeljenom sustavu

Za izradu preporuka u kontekstno utemeljenom sustavu prema Kharwal (2020)<sup>58</sup> kao izvor baze podataka korištena je internetska filmska baza podataka *TMDb*. Prva datoteka pod nazivom *tmdb\_5000\_movies.csv* sastoji se od 4806 različitih filmskih naslova i sadrži podatke o budžetu, žanrovima, web stranici filma, ID broju filma, originalnom jeziku, originalnom naslovu, kratki opis radnje, produkcijski studio, zemlje produkcije, datumu izlaska, zarada, trajanju filma, govornim jezicima, statusu, sloganu, naslovu, prosječnom broju ocjena, te broju ocjena. Druga datoteka pod nazivom *tmdb\_5000\_credits.csv* sadrži sljedeće informacije o tim filmovima: ID-filma, naslov, glumačka postava, filmska ekipa.

Za obradu navedenih datoteka potrebno je koristiti biblioteku za analizu podataka `pandas` koju je moguće uvesti u Python uz pomoć funkcije `import`. Sljedeći korak je učitavanje skupa podataka što se izvodi koristeći funkciju `read_csv` koja kao argument prima putanju datoteke koja se učitava. Ovom funkcijom se datoteka pretvara u *DataFrame* oblik podataka, tj. postaje tablica. Ovaj kod izgleda kao što slijedi:

```
import pandas as pd
informacije = pd.read_csv("podaci/tmdb_5000_credits.csv")
filmovi = pd.read_csv("podaci/tmdb_5000_movies.csv")
```

---

<sup>58</sup> Kharwal, A. (2020). *Movie Recommendation System with Machine Learning*. Thecleverprogrammer. <https://thecleverprogrammer.com/2020/05/20/data-science-project-movie-recommendation-system/> (Pristupljeno: 23. kolovoza 2022.)

U svrhu obrađivanja podataka u ovim datotekama potrebno ih je spojiti u jednu tablicu. Prvo je potrebno preimenovati naziv stupca *movie\_id* u *id* u tablici učitanoj u varijabli *informacije* koristeći metodu `rename` kako bi naziv stupca bio konzistentan s nazivom stupca *id* u tablici učitanoj pod varijablom *filmovi*. Metoda `merge` se koristi u svrhu spajanja ovih dviju tablica te se takva tablica pohranjuje u varijablu *filmovi\_spojeno*. U novu varijablu se sprema tablica sa spojenim podacima bez nepotrebnih stupaca. Ovo se izvodi uz pomoć metode `drop`.

Na očišćenu tablicu se primjenjuje TF-IDF mjera koju je moguće unijeti u Python iz biblioteke `sklearn`. Funkcija `TfidfVectorizer` se koristi za primjenu TF-IDF mjere nad određenim dokumentima – u ovom slučaju nad filmova. U varijablu *tfidf\_matrica* se sprema matrica na koju je primijenjena TF-IDF mjera. Kod je prikazan u nadolazećem paragrafu:

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_mjera = TfidfVectorizer(min_df = 3, max_features = None,
                              strip_accents = "unicode", analyzer = "word",
                              token_pattern = r"\w{1,}",
                              ngram_range=(1, 3),
                              stop_words = "english")
tfidf_matrica =
tfidf_mjera.fit_transform(filmovi_ocisceno["overview"].values.astype('U'))
```

Za učenje sličnosti između stavki u varijabli *tfidf\_matrica* potrebno je uvesti `sigmoid_kernel` iz biblioteke `sklearn`. U službenoj dokumentaciji Scikit-Learn<sup>59</sup> navodi da funkcija `sigmoid_kernel` računa sigmoidnu funkciju (eng. sigmoid kernel) između dvaju matrica.

Sigmoidna funkcija je stroj potpornih vektora (eng. support vector machine – SVM) koji se koristi kao metoda klasifikacije (Kovač i sur., 2022<sup>60</sup>). Prema Ricci i sur. (2015)<sup>61</sup> u kontekstu sustava za preporuku strojevi potpornih vektora uzimaju niz stavki koje se određenom korisniku sviđaju i niz stavki koje mu se ne sviđaju i pretvaraju ih u dva vektora u hiperravnini, pri čemu se maksimalno povećava prostor između tih vektora. Takva hiperravnina predstavlja klasifikator za preporuku stavki. Stavke se preporučuju samo ako se nalaze na strani hiperravnine gdje su ostale stavke koje se korisniku sviđaju. Nakon uvođenja sigmoidne

---

<sup>59</sup> *sklearn.metrics.pairwise.sigmoid\_kernel*. (n.d.). Scikit-Learn. Retrieved August 28, 2022, from [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.sigmoid\\_kernel.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.sigmoid_kernel.html) (Pristupljeno: 28. Kolovoza 2022.)

<sup>60</sup> Kovač, A., Dunder, I., Seljan, S. (2022). An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services. *Proceedings of MIPRO 2022*.

<sup>61</sup> Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.

funkcije stvara se funkcija `preporuka` koja kao argumente prima naziv filma te varijablu `sig` u kojoj je pohranjen rezultat sigmoidne funkcije. U funkciji `preporuka` u varijabli `idx` pohranjuje se indeks broj filmskog naslova film. Rezultat primjene sigmoidne funkcije nad filmom pod indeksom `idx` se pohranjuje u varijablu koja se potom sortira silazno. Funkcija vraća stavke pod indeksom 1 do 11. Stavka pod indeksom 0 nije uzeta u obzir jer se radi o filmu za kojeg se računaju preporuke s obzirom da ima najveću sličnost. Rezultat korištenja funkcije `preporuka` nad filmskim naslovom „The Shining“ prikazan je na Slici 18.

```
In [6]: preporuka("The Shining")
Out[6]:
2397          Hotel Rwanda
403           Last Action Hero
3325          Albert Nobbs
2063          Arlington Road
3652          Decoys
697           The Truman Show
835           Mr. Popper's Penguins
618           Mystery Men
4772          Down Terrace
1804          Snow White: A Tale of Terror
Name: original_title, dtype: object
```

Slika 18. Preporuke za film *Isijavanje*. (Izvor: vlastita izrada)

## 6.2. Preporuka u kolaborativnom filtriranju

Ovaj model prema Garodia (2020)<sup>62</sup> koristi skup podataka kojeg čini 610 korisnika s ukupno 100.836 ocjena i 3684 oznaka na 9742 različitih filmova. Sastoji se od 4 datoteke programa Excel s vrijednostima razdvojenima zarezom. Datoteka *links.csv* sadrži ID filma, ID filma na internetsku bazu podataka filmova *IMDb*, te ID filma *TMDb*. Datoteka *movies.csv* sadrži ID filma, naslov te žanrove. U datoteci *ratings.csv* se pohranjeni su podaci o ID-ju korisnika, ID-ju filma, ocjeni, te podaci o vremenskoj oznaci. U datoteci *tags.csv* pohranjen je ID korisnika, ID filma, oznaka te vremenska oznaka.

---

<sup>62</sup> 35. Garodia, S. (2020). *Recommendation System Using Collaborative Filtering in Python*. Medium. <https://medium.com/analytics-vidhya/recommendation-system-using-collaborative-filtering-cc310e641fde> (Pristupljeno: 26. kolovoza 2022.)



Kao i kod kontekstno utemeljenog modela, prvi korak je učitavanje datoteka. Ovaj model koristiti će samo dvije od četiri datoteke – *movies.csv* i *ratings.csv*. Nakon učitavanja `pandas` biblioteke u Python, datoteke se učitavaju u program. Datoteke pohranjujemo u varijablama `filmovi` i `ocjene`. Također se učitava biblioteka `numpy` u program. Navedene linije koda su:

```
import numpy as nd
import pandas as pd
```

U svrhu spajanja datoteka u jednu datoteku koristi se metoda `merge`, te se takva tablica pohranjuje u varijablu `filmovi`. Filmovi se grupiraju koristeći operaciju `groupby` iz biblioteke `pandas`, a prosječne ocjene se računaju uz pomoć metode `mean`. U sljedećem koraku se dodaje stupac *Total Rating* u tablicu koja sadrži broj ocjena za svaki film pojedinačno. Koristi se metoda `count`, zajedno sa operacijom `groupby`. Koristi se metoda `pivot_table` iz biblioteke `pandas` da bi se stvorila matrica koja prikazuje zasebne ocjene svakog korisnika za svaki film pojedinačno i pohranjuje se u novu varijablu

Za filmske naslove koje korisnici nisu ocijenili prikazana je vrijednost `NaN`. Za primjenu kosinusne vrijednosti nad ovom matricom potrebno je `NaN` vrijednosti pretvoriti u ocjene 0. U varijabli `ocjene_nula` pohranjuje se takva matrica. Koristeći metodu `copy` stvara se kopija matrice `film_korisnik`, a primjenom `fillna(0)` `NaN` vrijednosti se mijenjaju brojem 0. Funkcijom `import` se učitava mjera kosinusne sličnosti iz biblioteke `sklearn`. S obzirom da je `matrica_slicnosti` takozvani `numpy.ndarray` tip podataka, potrebno je koristiti funkciju `DataFrame` iz `pandas` biblioteke u svrhu pretvorbe u tablicu.

Funkcija `def` se koristi za izradu funkcije koja prima argumente `naziv` i `id_korisnik` te računa predviđene ocjene za sve filmove u tablici. Uvjetna petlja `if` provjerava nalazi li se određeni filmski naslov u tablici. U slučaju da uvjet nije ostvaren, odnosno da je `False`, funkcija vraća 2.5 te javlja da navedeni naslov se ne nalazi u matrici. U `if` bloku se pohranjuje tablica sa ID brojem svakog korisnika te sličnost pojedinog korisnika sa korisnikom `id_korisnik`. Mjera sličnosti je izražena decimalnim brojem od 0 do 1, u kojem 0 označava najmanju, a 1 najvišu razinu sličnosti. U zasebnu varijablu se pohranjuje se ocjena svakog korisnika za film `naziv`. U idućoj varijabli pohranjuju se korisnici koji nisu ocijenili film `naziv`. Koristeći funkciju `dropna` uklanjaju se `NaN` vrijednosti iz `vrijednost_ocjena`. U varijabli `kosinusna_vrijednost` pohranjuje se tablica `kosinusna_vrijednost` bez tablice `indeks_neocjenjen`. Ovime se za računanje preporuka uzimaju u obzir isključivo korisnici

koji su ocijenili film naziv. Na kraju se računa predviđena ocjena za film naziv. Funkcija vraća ocjenu u obliku decimalnog broja.

U svrhu dobivanja popisa filmova izrađena je funkcija `top_10_preporuka` koja kao argument uzima ID korisnika te vraća sortiranu listu 10 filmova koje korisnik `id_korisnik` nije ocijenio. Filmovi su sortirani od najveće do najmanje predviđene ocjene. Rezultat primjene ove funkcije nad korisnikom pod ID brojem 10 je prikazan na Slici 19.

```
In [29]: top_10_preporuka(10)
C:\Users\Mila Bikić\OneDrive - Sveučilište u Zagrebu,
Filozofski fakultet\FFZG\III\Završni rad\Praktični rad
\kolaborativno_filtriranje.py:57: RuntimeWarning: invalid
value encountered in double_scalars
  ocjene_filma = np.dot(vrijednost_ocjena,
kosinusna_vrijednost)/kosinusna_vrijednost.sum()
Out[29]:
[("Salem's Lot (2004)", 5.0),
 ('20 Million Miles to Earth (1957)', 5.0),
 ('61* (2001)', 5.0),
 ('7 Faces of Dr. Lao (1964)', 5.0),
 ('9/11 (2002)', 5.0),
 ('A Detective Story (2003)', 5.0),
 ('A Flintstones Christmas Carol (1994)', 5.0),
 ('A Perfect Day (2015)', 5.0),
 ('A Plasticine Crow (1981)', 5.0),
 ('Act of Killing, The (2012)', 5.0)]
```

Slika 19. Top 10 preporuka za korisnika pod ID brojem 10. (Izvor: vlastita izrada)

## 7. Zaključak

Glavni cilj ovog rada je bio predstaviti sustave za preporuku u kontekstu u kojem se koriste na internetu, njihove vrste i primjene te utvrditi stavove mladih ljudi u dobi od 16-30 godina o sustavima za preporuke, ali i saznati koliko su upoznati s ovim sustavima, koriste li određene internetske usluge isključivo zbog preporuka te smatraju li da im je preporučeni sadržaj relevantan.

U teorijskom dijelu rada dana je definicija, prikazan povijesni razvoj sustava za preporuku, te su predstavljene glavne vrste sustava za preporuku, tj. kontekstno utemeljeni sustavi i kolaborativno filtriranje. Prikazani su temeljni principi mjerenja poput TF-IDF mjere koja se koristi za računanje sličnosti između različitih dokumenata te kosinusne mjere sličnosti. Sustavi za preporuke su softverski alati i tehnike koje se koriste za generiranje preporuka za stavke koje će određenom korisniku biti zanimljive, a koriste se kod donošenja raznih odluka kao npr. koju glazbu slušati, koje članke čitati ili koje proizvode kupiti. Prisutni su na raznim domenama poput društvenih mreža kao što su *Facebook*, *Instagram* i *TikTok*, internetskim trgovinama kao što je *Amazon* i *AliExpress*, uslugama digitalnog strujanja kao *Spotify*, *Pandora*, *Netflix*, web pretraživačima poput *Google*.

U radu su detaljnije prikazani sustavi za preporuke koje koriste internetske usluge *Netflix* i *Deezer*. Istaknuta je važnost početne stranice za interakciju s preporučenim sadržajima te nabrojane razne vrste popisa koje takvi usluge koriste. U praktičnom okruženju su objašnjeni načini postupanja sa problemom hladnog starta. *Netflix* pri stvaranju novog korisničkog računa navodi korisnika da odabere nekoliko naslova iz kataloga koje se novom korisniku sviđa, dok je *Deezer* razvio sustav koji pruža djelomično personalizirane preporuke do kraja prvog dana korištenja sustava.

U istraživačkom dijelu rada su prikazani rezultati anketnog upitnika o sustavima za preporuke u kojoj je sudjelovalo 201 ispitanika, od kojih 197 je bilo u ciljanoj dobnoj skupini od 16 do 30 godina. Utvrđeno je da 72,1% ispitanika smatraju da su upoznati sa sustavima za preporuku. Ukupno 85 od 197 sudionika, odnosno 43, 1% uzorka je neutralno prema preporukama koje im sustavi predlažu, ali vidljivo je kod digitalnog strujanja glazbe veći broj ljudi, njih 78 od 154 koji su pristupili tom dijelu upitnika, je zadovoljan sa preporukama. Ipak, kod usluga strujanja video sadržaja samo 53 od 155 ispitanika koji koriste ovakve usluge su zadovoljni sa svojim preporukama, ali ne na svim uslugama. Manji broj ispitanika koristi određene internetske usluge

isključivo zbog preporuka. Samo 3,6% uzorka se izjasnilo da koristi određene web-usluge isključivo zbog preporuka. Kao takve usluge su izdvojili internetske trgovine *AliExpress*, web pretraživač *Google*, platformu za praćenje konzumacije i preporuku glazbe *Last.fm*, filmsku bazu podataka *IMDb*, i društvene mreže kao što su *Pinterest* i *TikTok*. Potvrđene su dvije od pet pretpostavki, a jedna je djelomično potvrđena. Potvrđeno je da su gotovo svi mladi ljudi svjesni postojanja sustava za preporuku te da nisu svjesni podataka koje sustavi za preporuku skupljaju. Djelomično je potvrđeno da mladi ljudi koriste usluge strujanja zbog njihovih preporuka i da su preporuke relevantne za njih. Ovim istraživanjem nije potvrđeno da preporuke pomažu mladima otkriti nove interese.

Na kraju su izrađena dva različita modela sustava za preporuku u programskom jeziku *Python*. Prvi je kontekstno utemeljen model koji koristi TF-IDF mjeru za računanje sličnosti zadanog filmskog naslova sa ostalima. Kao rezultat vraća popis deset najbližijih filmova. Drugi model koristi kolaborativno filtriranje kako bi zadanom korisniku preporučio deset filmova za koje predviđa najviše ocjene. Za izračun sličnosti koristila se kosinusna mjera sličnosti.

## 8. Literatura

1. *About Deezer*. (n.d.). Deezer. <https://www.deezer.com/en/company/about> (Pristupljeno: 19. svibnja 2022.)
2. Afchar, D., Melchiorre, A.B., Schedl, M., Hennequin, R., Epure, E.V., & Moussallam, M. (2022). *Explainability in Music Recommender Systems*. <https://doi.org/10.48550/arXiv.2201.10528>.
3. Alvino, C., & Basilico, J. (2015). *Learning a Personalized Homepage - Netflix TechBlog*. Medium. <https://netflixtechblog.com/learning-a-personalized-homepage-aa8ec670359a> (Pristupljeno: 7. svibnja 2022.)
4. Antal, D. (2021). *Listen Local: Why We Need Alternative Recommendation Systems*. Daniel Antal, CFA. <https://danielantal.eu/post/2020-12-15-alternative-recommendations/> (Pristupljeno: 18. svibnja 2022.)
5. Bendada, W., Salha, G., & Bontempelli, T. (2020). Carousel Personalization in Music Streaming Apps with Contextual Bandits. *Fourteenth ACM Conference on Recommender Systems*, 25(420). <https://doi.org/10.1145/3383313.3412217>
6. Blattmann, J. (2018). *Netflix: Binging On The Algorithm*. Medium. <https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59> (Pristupljeno: 7. svibnja 2022.)
7. Briand, L. (2021). *Recommending music to new users - Deezer I/O*. Medium. <https://deezer.io/recommending-music-to-new-users-182263648b82> (Pristupljeno: 18. svibnja 2022.)
8. Brownlee, J. (2018). *Gentle Introduction to Vector Norms in Machine Learning. Machine Learning Mastery*. <https://machinelearningmastery.com/vector-norms-machine-learning/> (Pristupljeno: 22. lipnja 2022.)
9. Burke, R., Felfernig, A., & Göker, M. H. (2011). Recommender Systems: An Overview. *AI Magazine*, 32(3), 13–18. <https://doi.org/10.1609/aimag.v32i3.2361>
10. Chen, M. (2017). Performance Evaluation of Recommender Systems. *International Journal of Performability Engineering*, 13(8), 1246–1256. <https://doi.org/10.23940/ijpe.17.08.p7.12461256>

11. *Deezer Music*. (2022). *Deezer Support*. <https://support.deezer.com/hc/en-gb/articles/201331091-Deezer-Music-> (Pristupljeno: 18. svibnja 2022.)
12. Felfernig, A., Friedrich, G., & Schmidt-Thieme, L. (2007). Guest Editors' Introduction: Recommender Systems. *IEEE Intelligent Systems*, 22(3), 18–21. <https://doi.org/10.1109/mis.2007.52>
13. Garodia, S. (2020). *Recommendation System Using Collaborative Filtering in Python*. Medium. <https://medium.com/analytics-vidhya/recommendation-system-using-collaborative-filtering-cc310e641fde> (Pristupljeno: 26. kolovoza 2022.)
14. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70. <https://doi.org/10.1145/138859.138867>
15. Google. (n.d.). *Content-based Filtering Advantages & Disadvantages*. Google Developers. <https://developers.google.com/machine-learning/recommendation/content-based/summary> (Pristupljeno: 27. travnja 2022.)
16. Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>
17. Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). *Recommender Systems: An Introduction* (Illustrated ed.). Cambridge University Press.
18. Karabiber, F. (2022). *Cosine Similarity*. Learn Data Science. <https://www.learndatasci.com/glossary/cosine-similarity/> (Pristupljeno: 22. lipnja 2022.)
19. Kharwal, A. (2020). *Movie Recommendation System with Machine Learning*. Thecleverprogrammer. [https://thecleverprogrammer.com/2020/05/20/data-science-project-movie-recommendation-system/#google\\_vignette](https://thecleverprogrammer.com/2020/05/20/data-science-project-movie-recommendation-system/#google_vignette) (Pristupljeno: 23. kolovoza 2022.)
20. Kovač, A., Dunder, I., Seljan, S. (2022). An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services. *Proceedings of MIPRO 2022*.
21. Krysik, A. (2021). *Netflix Algorithm: Everything You Need to Know About the Recommendation System of the Most Popular Streaming Portal*. Recostream. <https://recostream.com/blog/recommendation-system-netflix> (Pristupljeno: 1. svibnja 2022.)

22. Kumar, P. P., Vairachilai, S., Potluri, S., & Mohanty, S. N. (2021). *Recommender Systems: Algorithms and Applications*. CRC Press. <https://doi.org/10.1201/9780367631888>
23. *Machine Learning for Recommender Systems - A Primer*. (n.d.). Dasca.Org <https://www.dasca.org/world-of-big-data/article/machine-learning-for-recommender-systems-a-primer> (. Pristupljeno: 21. svibnja 2022.)
24. Marković, M., & Tomljanović, Z. (2018). "Kolaborativno filtriranje." *Math.e*, 34(1), 1-22. <https://hrcak.srce.hr/222763>.
25. Mishra, U. (2022). *What is a Content-based Recommendation System in Machine Learning?*. Analytics Steps. <https://www.analyticssteps.com/blogs/what-content-based-recommendation-system-machine-learning> (Pristupljeno: 28. travnja 2022.)
26. Netflix. (2022). *How Netflix's Recommendations System Works*. Netflix Help Center. <https://help.netflix.com/en/node/100639> (Pristupljeno: 7. svibnja 2022.)
27. *Preporuka*. (n.d.). Hrvatski Jezični Portal. <https://hjp.znanje.hr/index.php?show=search> (Pristupljeno: 19. svibnja 2022.)
28. Raman, K. (2020). *Matrix Factorization Explained | What is Matrix Factorization?* Great Learning. <https://www.mygreatlearning.com/blog/matrix-factorization-explained/> (Pristupljeno: 27. svibnja 2022.)
29. Rešetar, Janko, Seljan, Marko. *The Case study of Netflix. Company analysis and risk management strategies in the global business environment – a case study collection*. (urednik: Danijela Miloš Sprčić), 2021, 279-318.
30. Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2. izd.). Springer.
31. Schröder, G., Thiele, M. & Lehner, W. (2011). *Setting Goals and Choosing Metrics for Recommender System Evaluations*.
32. Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2017). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5), 813–831. <https://doi.org/10.1007/s13042-017-0762-9>
33. *sklearn.metrics.pairwise.sigmoid\_kernel*. (n.d.). Scikit-Learn. [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.sigmoid\\_kernel.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.sigmoid_kernel.html) (Pristupljeno: 28. kolovoza 2022.)

34. Spandana, S. (2020). *Why Am I Seeing This?* New America. <https://www.newamerica.org/oti/reports/why-am-i-seeing-this/case-study-netflix/> (Pristupljeno: 1. svibnja 2022.)
35. Subramanian, D. (2020). *Cosine Similarity, the metric behind recommendation systems.* Indiaai.Gov.In. <https://indiaai.gov.in/article/cosine-similarity-the-metric-behind-recommendation-systems> (Pristupljeno: 27. travnja 2022.)
36. Vyas, M. (2018). *Recommendation Systems.* Medium. <https://meenavyas.medium.com/recommendation-systems-1d20d159634d> (Pristupljeno: 19. svibnja 2022.)



# Sustavi za preporuku na internetu

## Sažetak

Mnoge internetske usluge kao što su *Amazon*, *Deezer* i *Netflix* koriste sustave za preporuku (eng. recommender systems) kako bi korisnicima pružili preporučene personalizirane sadržaje. Ovakvi sustavi mogu postati i sastavni dio aplikacija. Cilj ovog rada je bio predstaviti sustave za preporuku u kontekstu u kojem se koriste na internetu, njihove vrste i primjene te utvrditi stavove mladih ljudi u dobi od 16-30 godina o sustavima za preporuke, ali i saznati koliko su upoznati s ovim sustavima, koriste li određene internetske usluge isključivo zbog preporuka te smatraju li da im je preporučeni sadržaj relevantan.

U teorijskom dijelu rada prikazani su osnovni principi rada i algoritama sustava za preporuku, predstavljena je povijest razvitka sustava za preporuke, objašnjene su razlike između primijenjenih tipova filtriranja (kontekstno utemeljeni sustavi, kolaborativno filtriranje, filtriranje temeljeno na znanju, demografski i hibridni sustavi) i algoritmi, te predstavljeni načini i metrike evaluacije sustava za preporuku. Navedeni su primjeri korištenja sustava za preporuku na popularnim internetskim uslugama *Netflix* i *Deezer*. U praktičnom dijelu prikazani su rezultati provedenog istraživanja kojim su ispitani stavovi i upoznatost mladih sa sustavima za preporuku te njihov utjecaj na korištenje odabranih internetskih usluga. Na kraju su izrađeni modeli kontekstno utemeljenog sustava i kolaborativnog filtriranja u programskom jeziku *Python*.

**Ključne riječi:** sustavi za preporuku, preporuke na internetu, streaming servisi, Netflix, Deezer

# Recommender Systems on the Internet

## Summary

Numerous internet services such as *Amazon, Deezer and Netflix* use recommender systems to provide users with recommended personalized content. These systems can become an integral part of application software. The aim of this thesis is to present recommender systems within the framework of the Internet, enumerate recommender types and applications, and to identify opinions about this topic of young people aged 16-30, as well as to find out the extend of familiarity with recommenders, whether they use any Internet service exclusively because of its recommendations and whether they find the recommended content relevant.

In the theoretical segment of this paper the basic principles and algorithms of recommender systems are presented, an outline of the historical development of recommender systems is shown, and types of recommender systems and algorithms are differentiated (content-based recommendation, collaborative filtering, knowledge-based, demographic, hybrid systems), and methods and metrics of recommender system evaluation are given. Then, examples of recommender systems used by popular internet services *Netflix* and *Deezer* are given. In the research segment of the paper the results of a conducted survey on the attitudes and awareness of young people about recommender systems are presented, and on their impact on the usage of chosen Internet services. Lastly, a content-based and a collaborative filtering-based recommender system were created in the **Python** programming language.

**Key words:** recommender systems, online recommendations, streaming services, Netflix, Deezer