

LATVIJAS UNIVERSITĀTE

MAGISTRA DARBS

RĪGA 2022

LATVIJAS UNIVERSITĀTE
DATORIKAS FAKULTĀTE

Televīzijas satura personalizētu ieteikumu sistēmas izstrāde

Autors: Roberts Kiršteins

Stud. apl. Nr. rk10076

Darba vadītājs: prof. Kārlis Podnieks

RĪGA 2022

ANOTĀCIJA, ATSLĒGVĀRDI

Mūsdienu cilvēkam ir pieejams ļoti plašs televīzijas satura piedāvājums. Līdz ar to daudz laika tiek pavadīts nevis skatoties interesējošu filmu vai raidījumu, bet meklējot piemērotu saturu. Ieteikumu sistēmas mērķis ir ieteikt cilvēkam potenciāli saistošu produktu vai pakalpojumu, sašaurinot pieejamo pakalpojumu un produktu klāstu, atvieglojot izvēli un piedāvājot cilvēkam to, ko cilvēks visticamāk vēlētos redzēt vai pirkt. Darba mērķis ir izveidot interaktīvās televīzijas satura ieteikumu sistēmu, ar kuras palīdzību katram pakalpojuma lietotājam tiktu aprēķināti individuāli satura ieteikumi. Aprēķini balstītos uz SIA Tet televīzijas klientu skatīšanās paradumiem, kur tiktu izmantoti pieejamie televīzijas programmu skatīšanās dati, pieejamie elektroniskās satura programmas (EPG) metadati, kuri tiktu papildināti ar tīmeklī pieejamajiem datiem. Izaicinājums televīzijas satura rekomendācijās ir mazais datu apjoms, kas pieejams par TV saturu, nekvalitatīvie programmu dati, kā arī limitētais laiks, kurā saturs ir televīzijā pieejams.

Atslēgvārdi: TV satura ieteikumi, ieteikumu sistēmas, datizrace, mašīnmācīšanās.

ABSTRACT

Constructing a personalized television content recommender system.

An increasingly large volume of TV content is available to a modern-day man. Consequently, a lot of time is spent looking for appropriate content and not watching a movie or broadcast of interest. The recommendation system aims to propose to a person potentially interesting product or service, narrowing down the range of services and products available, making it easier to choose and offering person something he would most likely want to see or buy. The goal is to create a recommender system for interactive TV content that would calculate individual content recommendations for each user. The calculations would be based on the viewing habits of Tet TV customers, where available TV program viewing data and electronic program guide (EPG) metadata supplemented with the data available on the web would be used. The challenge in TV content recommendations is the small amount of data available on TV content, poor EPG data, and the limited time span at which the content is available on television.

Keywords: TV content recommendations, recommendation system, data mining, machine learning.

AUTOREFERĀTS

Darba rezultātā izveidota EPG satura personalizētu ieteikumu veidošanas sistēma. Darbs sastāv no datu tīrīšanas, analīzes, papildu datu ieguves, personalizētu ieteikumu izveides un pielāgošanas Latvijas TV skatīšanās specifikai, kā arī risinājuma iegūto rezultātu izvērtēšanas un salīdzināšanas ar citiem līdzīgiem pētījumiem.

Darba autors izpētījis dažādas ieteikumu sistēmu pieejas, kuras izmantotas gan veidojot pieprasījumuvideo, gan EPG ieteikumu sistēmas, izvērtējis labākās pieejas, kuras varētu praktiski pielietot SIA Tet ieteikumu sistēmas veidošanā. Apskatīti 27 zinātniskās literatūras avoti – konferenču raksti un žurnālu publikācijas. Darbā kopumā izmantoti 47 literatūras avoti.

Esošo risinājumu apskatā izklāstītas bieži pielietotas ieteikumu sistēmu veidošanas paradigmas un izmantotie modeļi, kā arī identificētas tipiskākās problēmas un izaicinājumi, ar kādiem nākas saskarties, veidojot TV satura ieteikumu sistēmas. Ieteikumu sistēmas veidošanas sadaļā autors aprakstījis savu pieredzi, veidojot šādu sistēmu, un aprakstījis saskatītos izaicinājumus, kā arī aprakstījis izvēlētos risinājumus un tos pamatojis.

Darba praktiskajā daļā autors ir ieguvis datus no SIA Tet datubāzēm, izvērtējis datu nozīmīgumu ieteikumu veidošanā, novērtējis datu kvalitāti un kvantitāti. Datu apstrādes, tīrīšanas, analīzes kods ir autora paša rakstīts, izmantojot dažādas *python* bibliotēkas. Autors veicis datu tīrīšanu, kā arī identificējis potenciālo datu trūkumu un izpētījis iespējas iegūt papildu satura metadatus. Darbā iegūti papildu dati no atvērto datu portāla *Wikidata* un izmantotas dažādas ieteikumu veidošanas pieejas, lai veidotu ieteikumu sistēmu.

Ieteikumu sistēmas kvalitatīvo un nesaistē veikto testu rezultāti liecina, ka ieteikumu sistēma ir funkcionējoša un ģenerētie ieteikumi ir kvalitatīvi. Izvērtējot pielietoto metožu sarežģītību un reāli patērēto laiku, nonākts pie secinājuma, ka risinājums būtu spējīgs darboties tiešsaistē, un ir lietojams, lai ģenerētu personalizētus ieteikumus SIA Tet EPG sistēmā.

Darba teksts ir pārlasīts un darbā lietotie termini ir pārbaudīti Latvijas Nacionālajā terminoloģijas vai tezaurs portālā. Darbs rakstīts ievērojot metodiskos norādījumus.

Visa informācija, kas tikusi iegūta no citiem autoriem, ir atzīmēta ar atsaucēm.

SATURS

SATURS.....	6
APZĪMĒJUMU SARAKSTS.....	8
IEVADS	9
1. MOTIVĀCIJA	12
2. ESOŠO RISINĀJUMU APSKATS	13
2.1. Ieteikumu sistēmu paradigmas.....	13
2.1.1 Content-based (CB) pieeja.....	13
2.1.2 Collaborative filtering pieeja	14
2.1.3 Hibrīda pieeja	16
2.2. RS konteksts	16
2.3. Ieteikumu sistēmu risinājumi.....	18
2.3.1 Pētījumi VOD RS.....	18
2.3.2 Pētījumi EPG RS	20
2.4. Ieteikumu skaidrojuma nozīme.....	24
2.5. Metožu precizitātes novērtēšana	25
2.6. EPG ieteikumu sistēmu problēmas apgabals un izmērs	26
3. AUTORA IEGULDĪJUMS, RISINĀJUMS, PIEEJA	28
4. SISTĒMAS ARHITEKTŪRA, VISPĀRĪGAIS RISINĀJUMS.....	29
5. IETEIKUMU SISTĒMAS VEIDOŠANA	30
5.1. Datu pieejamība un apraksts	30
5.2. Datu kvalitātes novērtēšana un datu tīrīšana.....	33
5.2.1 Datu kvalitāte filmām	34
5.2.2 Datu kvalitāte seriāliem un šoviem	35
5.3. Lietotāju IPTV skatīšanās paradumu analīze.....	36
5.3.1 Interesējošā satura noteikšana	36

5.3.2	Valodas izvēle.....	40
5.3.3	Interaktīvo funkciju izmantošana	42
5.3.4	Laika konteksts	42
5.4.	Papildus datu ieguve	45
5.4.1	Datu iegūšana no Wikidata.....	45
5.4.2	Datu vākšana no tīmekļa satura datubāzēm.....	46
5.4.3	Atslēgas vārdu ģenerēšana, izmantojot anotācijas	47
5.5.	Algoritmi un metodes ieteikumu veidošanai	47
5.5.1	Kosinusa līdzības algoritms – CB pieeja.....	48
5.5.2	Matricu faktORIZĀCIJA, izmantojot ALS metodi – CF pieeja	50
5.5.3	Populārākais saturs	54
5.6.	Analītiskais apraksts	55
5.7.	Ieteikumu sistēmas kvalitātes novērtējums.....	56
5.7.1	Ieteikumu izvērtēšana filmām nesaistē.....	57
5.7.2	Ieteikumu izvērtēšana seriāliem nesaistē.....	59
5.7.3	Ieteiktais satura apjoms no kataloga.....	61
5.7.4	Ieteikumu sistēmas kvalitatīvie testi.....	61
EKSPERIMENTU REZULTĀTI, NOVĒRTĒJUMS		63
SECINĀJUMI		64
PIELIKUMI.....		72
1.	pielikums. <i>Github</i> repozitorija apraksts	73
2.	pielikums. Iegūto datu piemērs filmu metadatiem no <i>Wikidata</i>	74
3.	pielikums. Ieteikumu paraugs filmām.	75
4.	pielikums. Ieteikumu paraugs seriāliem.	76

APZĪMĒJUMU SARAKSTS

RS – ieteikumu sistēma.

EPG – elektroniskā satura programma.

ML – mašīnmācīšanās.

CS – “cold start” jeb nepietiekama lietotāja un pakalpojuma mijiedarbība, lai izsecinātu lietotāja preferences. Parasti jauniem pakalpojuma lietotājiem.

CB – “content based filtering” ir viena no ieteikumu sistēmu veidošanas paradigmām, kas balstās uz lietotāju un produktu raksturojošajiem datiem.

CF – “collaborative filtering” ir viena no ieteikumu sistēmu veidošanas paradigmām, kas balstās uz mijiedarbībām starp lietotāju un produktu.

VOD – “video on demand” jeb video pēc pieprasījuma ir pakalpojuma veids, kurš darbojas pēc video nomas principa, kur pieejamās satura vienības ir skatāmas lietotājam ērtā laikā.

SVOD - “subscription video on demand” ir viens no VOD iedalījumiem - abonēšanas pakalpojums filmu un seriālu datubāzei.

TVOD – “transactional video on demand” ir viens no VOD iedalījumiem - individuālu filmu noma uz noteiktu laika periodu.

iTV – interaktīvā televīzija.

IPTV – interneta protokola televīzija.

KNN – tuvāko kaimiņu metode klasteru ģenerēšanai.

MF – matricu faktORIZĀCIJA.

ALS – mainīgu mazāko kvadrātisko kļūdu algoritms matricu faktoru aprēķināšanai.

SPARQL - SPARQL protokola un RDF vaicājumu valoda (SPARQL Protocol and RDF Query Language).

IEVADS

Pēdējo desmitgažu laikā kopā ar Google, Amazon, Youtube, Netflix un citu līdzīgu portālu ienākšanu cilvēku ikdienā, kā neizbēgama ikdienas sastāvdaļa kļuvusi arī ieteikumu sistēmas (RS). Tās ir sastopamas gan e-komercijas, gan mārketinga, gan straumēšanas un televīzijas pakalpojumu jomās. Vai tā ir Google interneta meklētājprogramma, kura palīdz cilvēkam atrast aktuālāko rakstu, vai tas ir Rimi interneta veikals, kurā tiek ieteikti līdzīgie un saistītie produkti, RS ir kļuvušas par gandrīz katra cilvēka ikdienas sastāvdaļu.

RS ir algoritmu un nosacījumu kopums, kura mērķis ir lietotājam ieteikt viņu potenciāli interesējošus produktus un pakalpojumus (piemēram, TV raidījumus, pārtikas produktus, zinātniskos rakstus, utt.). RS ir nozīmīga daudzu nozaru uzņēmumu sastāvdaļa, it sevišķi to, kuru pakalpojumi ir nodrošināti digitālajā vidē. RS var būtiski atvieglot pakalpojuma lietošanas ērtumu, palīdzēt uzņēmumam atrādīt lielāku pieejamo produktu klāstu, pielāgojoties klienta vēlmēm un interesēm, bieži palielinot pakalpojuma vērtību klienta acīs, kā arī ģenerējot uzņēmumam papildus ienākumus, atvieglot konkurenci un ļaujot digitālajā vidē mijiedarboties ar klientu. Tas ļauj klientam piedāvāt alternatīvus, iespējams, labākus un dārgākus produktus vai pakalpojumus, kā arī piedāvā iespēju klientam ieteikt papildus produktus un pakalpojumus, kas varbūt arī nebija plānoti primāri pirkt.

Kā piemērs tam, ka labu RS izveide ir nopietns datizraces un mašīnmācīšanās izaicinājums, var minēt Netflix izsludināto konkursu 2009. gadā. "Netflix prize" bija atklāts konkurss par labāko RS algoritmu, ar mērķi uzlabot kompānijas esošo risinājumu precizitāti. Galvenā balva bija 1 miljons dolāru [1].

RS izstrāde un pētīšana ir aktuāla datizraces problēma jau vairākus gadu desmitus [2, 3, 4, 5], sākot ar vienkāršu kanālu filtrēšanu [6], beidzot ar dziļās mašīnmācīšanās RS [7]. Atkarībā no pielietojuma sfēras, šo risinājumu veidošana var stipri atšķirties. Tāpēc, veidojot RS, būtiski saprast attiecīgā biznesa specifiku un ieteikumu mērķi.

Televizors vēl joprojām ir populārākā multimediju patēriņa iekārta Latvijā un lielā daļā pasaules [8, 9, 10], līdz ar to televizorā patērētā video satura apmērs ir ievērojams un, lai gan pēdējos gados video patēriņa daļa tā sauktajos lineārajos (klasiskajos) TV kanālos samazinās, tā joprojām ir lielākā video patēriņa daļa gan Latvijā [11, 12], gan citur pasaulē [10]. Abonējot TV

pakalpojumu Latvijā, piedāvāto kanālu skaits parasti svārstās no dažiem desmitiem līdz pat 150 vietējo un starptautisko TV kanālu. Lietotājiem ir ļoti daudz izvēles iespēju, un šīs izvēles neatvieglo tas, ka mūsdienās plaši pieejama ir interneta protokola televīzija (IPTV), kura, pretēji kabeļtelevīzijai, var piedāvāt arī interaktīvo TV, kas ļauj TV pārraides ierakstīt, apstādināt, un skatīties vēlreiz gan sākot skatīties konkrēto vienību no jauna, gan arī skatīties televīziju t.s. arhīvā, kur daļa pārraižu pieejamas noskatīties vēl pāris dienas pēc tiešraides. Līdz ar to, RS var būtiski uzlabot TV patērētāja skatīšanās pieredzi, atvieglojot interesējošā satura meklēšanas procesu un iesakot raidījumus, filmas un kanālus, kas patērētājam varētu būt interesanti.

Klasiskajā digitālajā vai lineārajā televīzijā video saturs ir pieejams tikai tiešraides formātā, līdz ar to šādai televīzijai veidot rekomendācijas ir ļoti izaicinoši. Interaktīvajai televīzijai (ITV) ir papildus iespējas veikt raidījumu ierakstu, kā arī skatīties lielu daļu satura arhīvā, kas pieejams aptuveni 10 dienas. Papildus sarežģītumus rekomendāciju aprēķinam rada tas, ka bieži jādarbojas ar nepilnīgu informāciju apjomu. T.i., pieejamie raidījumu metadati no EPG ir sliktā kvalitātē un ar limitētu aprakstošo lauku skaitu. EPG televīzijas skatītājiem piedāvā nepārtraukti atjaunotas izvēlnes, kurās atspoguļotas dažādos kanālos šobrīd raidītās un plānotās pārraides. EPG palīdz lietotājam izvēlēties kanālus, taču nesniedz personalizētus satura ieteikumus. Klasiskos rekomendāciju algoritmu aprakstos parasti pieņemts, ka ir zināmi tādi raidījumu parametri kā raidījuma režisors, aktieri, raidījuma atslēgas vārdi un pat vērtējums no starptautiski atzītām televīzijas satura datubāzēm, taču EPG pieejamie lauki aprobežojas ar satura vienības tipu, žanru, nosaukumu un, iespējams, anotāciju. EPG dati pamatā tiek iegūti no TV kanālu satura piegādātājiem un tiem ir stipri atšķirīga kvalitāte un aizpildījums. Rekomendāciju veidošanu padara sarežģītāku arī tas, ka Latvijas ģeopolitiskajā telpā TV saturs tiek patērēts vairākās valodās - latviešu, krievu, angļu. Jāņem vērā arī TV skatītāja valodu preferences un zināšanas. Latvijas un citu mazo valstu veidotajam oriģinālajam saturam ir problēmas ar aprakstošo datu kvalitāti un kvantitāti, sevišķi sarunu un realitātes raidījumiem, jo to mērķauditorija ir maza, un starptautiskajās filmu datubāzēs informācija par tiem nav pieejama. Tāpēc rekomendācijām nav iespējams pilnvērtīgi izmantot uz saturu aprakstošajiem datiem balstītus algoritmus (*content-based*). Savukārt, uz mijiedarbībām balstītiem (*collaborative filtering*) algoritmiem prasība ir zināmi konkrētu programmu skatījumi un patērētāja preferences. Lineārās TV gadījumā šāda pieeja var tikt lietota, lai veidotu ieteikumus raidījumu atkārtojumiem vai raidījumiem arhīvā, bet ne tiešraides raidījumiem. Rezultātā CF pieeja IPTV rekomendācijās izmantota reti.

Maģistra darbā veidota rekomendāciju sistēma, balstoties uz SIA Tet interaktīvas televīzijas skatījuma datiem, EPG metadatiem, kā arī izmantota papildu informācija par satura vienībām tīmeklī. RS veidota kā hibrīds, izmantojot dažādas datizraces metodes - gan *collaborative filtering* (CF), gan *content-based* (CB) un šīs metodes papildinātas ar nosacījumiem, kas veidojas, analizējot klienta paradumus. Analīzē ir izmantoti TV patēriņa dati un EPG metadati un noskaidroti lietotāja skatīšanās laika parametri, kanālu un valodu preferences, iecienītākie žanri un raidījumu tipi. Risinājumā veidoti ieteikumi gan nākotnes tiešraides TV programmām, gan arhīva raidījumiem, izmantojot CF un CB pieejas. CB algoritmā izmantoti EPG pieejamie un papildus iegūtie metadati, iesakot līdzīgās filmas pēc to raksturojošajiem parametriem. CF algoritmā izmantoti lietotāju mijiedarbības dati ar raidījumiem, veidojot satura ieteikumus no citu, līdzīgu klientu, interesējoša satura. Papildus tam tiek aprēķināts personalizēts populārākais saturs. Šie risinājumi apvienoti un, ņemot vērā raidījumu pieejamību, veidoti personalizēti satura ieteikumi.

1. MOTIVĀCIJA

Kā jau iepriekš minēts, TV tiešraides pakalpojumu aktualitāte tuvākajos gados saglabāsies augstā līmenī, par spīti jaunāko paaudžu tendencei biežāk izvēlēties straumēšanas un pieprasījumvideo (VOD) pakalpojumus lineārās TV vietā [11, 12, 13]. Ziņu raidījumi un sporta pārraides primāri tiek skatītas tiešraidē un pārsvarā lineārajā TV, lai gan popularitāti iegūst arī straumēšanas pakalpojumi [13].

Baltijas tirgū pagaidām nav neviena lineārā TV pakalpojuma sniedzēja, kas nodrošinātu personalizētus šī pakalpojuma satura ieteikumus, un, ņemot vērā to, ka Latvijā TV saturu patērē dažādās valodās [12], šādam nolūkam neder arī specializēti RS izstrādātāju piedāvātie produkti, kuri IPTV piedāvā tikai uz CB balstītus risinājumus [14].

Patērētāja ieguvums no TV RS ir laika ekonomija – personalizēta ieteikumu sistēma lietotājam ļauj ātrāk sameklēt izklaidējo vai izzinošo satura programmu, kas patērētājam varētu interesēt konkrētā brīdī, kā arī RS ļautu ieskicēt potenciāli interesantu saturu TV kanālos, kurus lietotājs ikdienā nepatērē, taču kur tiek raidīti lietotājam potenciāli interesējoši raidījumi.

Skatoties no uzņēmuma puses, personalizēta RS var palīdzēt celt pakalpojuma vērtību, tādejādi uzlabojot pakalpojuma veicinātāju un noliedzēju attiecību (Net promoter score – NPS) un, palielinot pakalpojuma abonēšanas laiku, samazinot mainības koeficientu (churn rate), kā arī uzņēmumu nostādīt labākā pozīcijā attiecībā pret citiem līdzīgu pakalpojumu sniedzējiem šajā piesātinātajā biznesa jomā. Lietotāju pāreja uz VOD zināmā mērā ir skaidrojama tieši ar iespēju ērti meklēt saturu un šajā vidē personalizēti ieteikumi jau ir norma.

2. ESOŠO RISINĀJUMU APSKATS

Šajā nodaļā sīkāk aplūkosim RS veidošanas pieejas un prakses, kuras var tikt izmantotas video ieteikumu veidošanai.

Ieteikumu sistēmas var veidot no vienkāršiem, cilvēku definētiem likumiem, taču personalizētas un kvalitatīvas ieteikumu sistēmas izveidei nepieciešams lietot kādu no mašīnmācīšanās (ML) metodēm. ML izmanto datorus, lai simulētu cilvēka mācīšanos un ļautu datoriem iegūt zināšanas no reālās pasaules procesiem un datiem, un, balstoties uz šīm zināšanām, efektīvi uzlabotu dažādu uzdevumu sniegumu.

Mācīšanās ir zināšanu ieguves process. Cilvēki dabīgi mācās no pieredzes, jo viņiem piemīt spriestspēja. Pretēji tam datoriem nav spriestspējas un tie mācās ar algoritmiem. Mūsdienu literatūrā aprakstīts plašs klāsts ML algoritmu un tos var klasificēt pēc to mācīšanās procesa. ML iedalās četrās galvenajās klasēs: pārraudzītā, nepārraudzītā, daļēji pārraudzītā un stimulētā mācīšanās [15]. Ieteikumu sistēmu veidošanā var tikt izmantota gan pārraudzītā, gan nepārraudzītā ML.

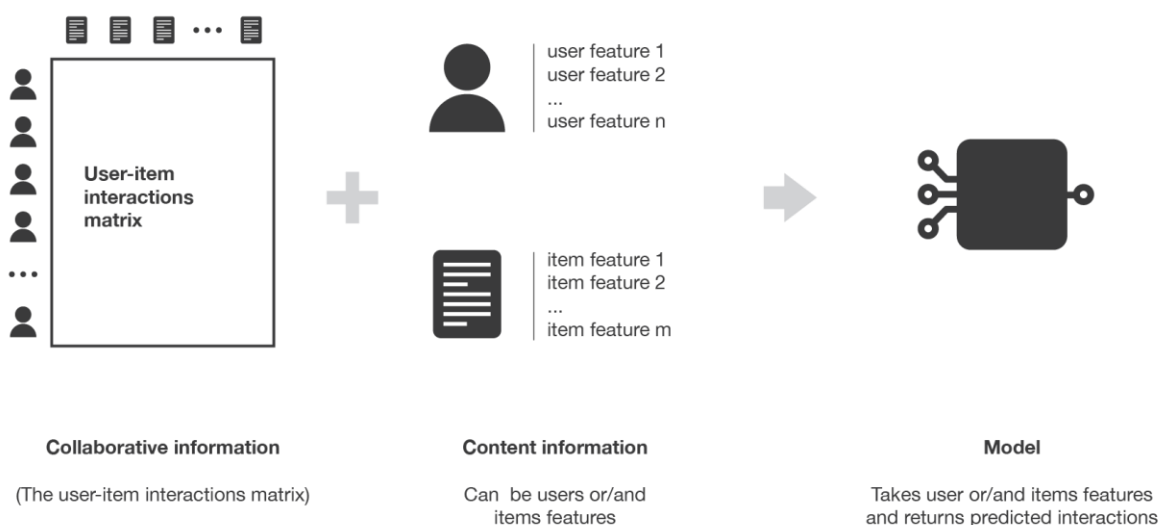
2.1. Ieteikumu sistēmu paradigmas

RS problēma risināta daudz un dažādos veidos jau pēdējos 20 gadus, un ir izstrādāti dažādi algoritmi, kurus pēc darbības principa var iedalīt sekojošās paradigmās – CB pieeja un CF pieeja. Pirmā pieeja darbojas kopā ar aprakstošajiem datiem vai pat izmantojot tikai aprakstošos datus. Otrā pieeja balstās uz mijiedarbībām, tāpēc nav izmantojama gadījumos, kad nav datu par mijiedarbībām ar produktu. Filmas, seriālus un raidījumus šajā sadaļā sauksim par vienumiem.

2.1.1 *Content-based (CB) pieeja*

CB pamatā ir patērētāja mijiedarbība ar vienumiem, kas papildināta ar patērētāju vai vienumu raksturojošu informāciju vai gan patērētāju, gan vienumu raksturojoša informācija. Video

rekomendāciju kontekstā tie būtu, piemēram, patērētāja demogrāfiskie parametri un filmu un seriālu metadati [16].



2.1. att. CB paradigmas metožu pārskats [16].

Viena no vienkāršākām pieejām – aprēķināt katrai vienībai līdzīgās vienības, izvēloties kādu no attāluma mēriem, kā piemēram, kosinusa līdzības koeficientu, Manhetenas, Minkovska vai Eiklīda attālumu, un līdzīgās vienības izmantot ieteikumu veidošanā, iesakot tās, ar kurām mijiedarbība vēl nav bijusi.

Cita pieeja CB ieteikumu veidošanā ir regresijas modeļu izmantošana, lai prognozētu, kā lietotājs mijiedarbosies ar vienumiem, piemēram, kādu balli liktu. Ja vēsturiski dati par konkrētu vērtējumu nav pieejami, var izmantot bināras klasifikācijas metodes, piemēram, lēmumu kokus, loģisko regresiju vai kādu citu modeli. Šādi modeļi rēķinātu varbūtību, ka lietotājam konkrētā satura vienība interesē. Šīs ir pārraudzītās ML pieejas, jo treniņa kopā tiek precīzi nodefinēts, kāda ir lietotāja mijiedarbība ar vienumu.

2.1.2 Collaborative filtering pieeja

CF ieteikumu sistēmu pamatā ir tikai iepriekšējā mijiedarbība, kas reģistrēta starp patērētājiem un produktiem vai pakalpojumiem. Šīs mijiedarbības tipiski tiek glabātas lietotāja-

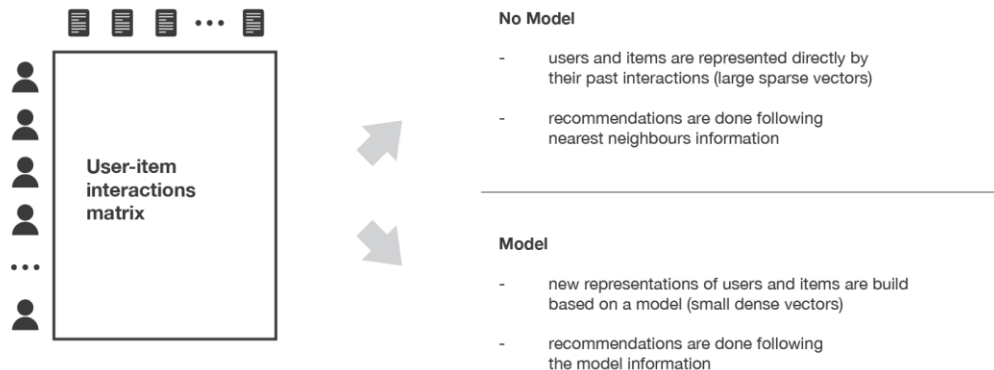
produkta mijiedarbības matricās [16]. Pēc būtības CF ideja vienkāršākajā variantā ir līdzīga asociāciju likumu meklēšanai, tikai vienas sesijas vietā tiek ņemtas vērā arī vēsturiskas patērētāja intereses, līdz ar to ieteikumi veidojas personalizētāki. Metodes var iedalīt divās grupās – atmiņā bāzētās un modelī bāzētās metodes (2.2. att.):

- Atmiņā bāzētās metodes var iedalīt lietotājs-lietotājs tipa metodēs un viens-vienums tipa metodēs. Lietotājs - lietotājs metodē principā tiek meklēti lietotājam līdzīgākie lietotāji, izmantojot, piemēram, KNN un ieteiktas vēl neredzētas filmas, kas patika līdzīgajiem lietotājiem. Viens - viens metodes pieeja ir meklēt vienumus, ar kuriem lietotāji kopumā mijiedarbojas līdzīgi – piemēram slikti vērtē maza budžeta filmas. Ja kādam lietotājam tieši šādas filmas patīk, šis modelis ieteiks filmas, uz kurām vairums lietotāju reaģēja līdzīgi. Atmiņā bāzētu metožu lielākais mīnuss ir tāds, ka tiek glabātas lielas un maz aizpildītas datu matricas un šādi risinājumi slikti mērogojas. Algoritmu sarežģītība ir $O(ndk)$, kur n ir lietotāju skaits, d ir vienumu skaits un k – tuvāko kaimiņu skaits.

- Modelī bāzēto metožu galvenā ideja ir matricu faktorizācija, kur visa lietotājs-vienums matrica tiek saspiesta lietotājs-faktors matricas P un faktors-vienums matricas Q reizinājuma matricā, kuras ir daudz mazākas un blīvākas par sākotnējo lietotājs-vienums matricu – formula (2.1). Ideja balstās tajā, ka eksistē tādi latentie faktori (skaitā daudz mazāki par reālo lietotāja mijiedarbību skaitu), ar kuriem var tuvināti aprakstīt gan lietotājus, gan vienumus. Šie lietotāju un vienumu aprakstošie vektori latentu pazīmju telpā ir blīvi, un lietotāja-vienuma mijiedarbība tiek uzdots kā šo divu blīvo vektoru reizinājums. Modelī bāzētā pieeja neapskata katru mijiedarbību starp lietotāju un vienumu, bet izmanto latentu faktoru vektorus, lai pēc iespējas labāk atjaunotu konkrētas matricas vērtības un varētu aizpildīt arī tās, kur mijiedarbība nav bijusi.

$$R \approx PQ^T; \quad (2.1)$$

CF metodes pluss ir tāds, ka tā kļūst arvien labāka, parādoties vairāk mijiedarbībām starp lietotāju un produktu, kā arī parādoties vairāk lietotājiem, taču mīnuss – šī metode cieš no CS problēmas. Gan atmiņā, gan modelī bāzētās metodes var ieteikt lietotājiem vienības, aprēķināt vienībai līdzīgās vienības, lietotājam līdzīgos lietotājus un katrai vienībai pielasīt lietotājus, kas visticamāk varētu ar vienumu mijiedarboties. CF metodes pēc mācīšanās procesa ir nepārraudzītas, modelim netiek dotas pazīmes ar lietotāja preferencēm, modelis tās iteratīvi iemācās pats.



2.2. att. CF paradigmas metožu pārskats [16].

2.1.3 Hibrīda pieeja

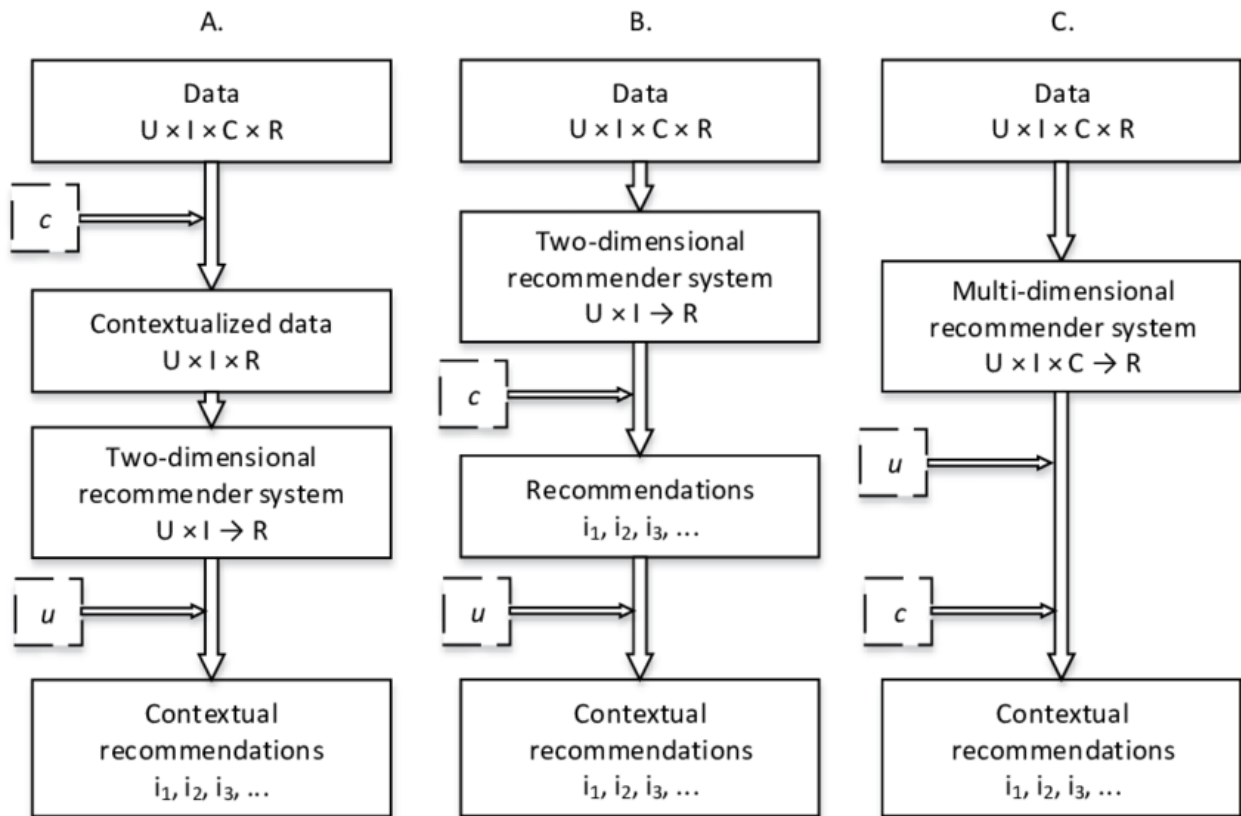
Dažādu pieeju apvienojumu dēvē par hibrīda pieeju. Hibrīda RS var realizēt dažādos veidos: veidojot atsevišķu CF un CB daļu, kuru rezultātus apvieno; pievienojot CB algoritmam CF īpašības vai otrādi; apvienojot abas pieejas vienā modelī, bieži izmantojot neirona tīklus. Hibrīda metodes parasti atrisina arī problemātisko ieteikumu veidošanu klientiem, kuriem nav vai nav pietiekami daudz mijiedarbības datu ar pakalpojumu. Mūsdienās lielai daļai RS pamatā ir tieši šīs jaukto modeļu pieejas. Kā piemērs var kalpot *Netflix*, kur ieteikumi tiek veidoti, salīdzinot skatīšanās un meklēšanas paradumus starp lietotājiem (CF) un iesakot pakalpojuma lietotājam tās filmas, kuru metadati ir līdzīgi filmām, ko lietotājs augstu vērtējis [17].

2.2. RS konteksts

Tāpat kā ieteikumi var mainīties atkarībā no lietotājiem, tie var mainīties arī atkarībā no konteksta. Konteksts var būt skatīšanās laiks, diena, kanālu grupa, skatīšanās vieta vai iekārta (piemēram, skatoties saturu mobilajā telefonā, specefektiem bagātas filmas būs mazāk aktuālas,

taču ziņas un sporta raidījumi varētu interesēt vairāk). Ir trīs veidi, kā konteksts var papildināt ieteikumu sistēmu [18]:

- Konteksta pirms-filtrēšana jeb filtrs pirms datu izmantošanas modelī. Šajā gadījumā modelī tiek izmantoti tikai konkrētā konteksta dati, piemēram, tiek izmantotas tikai filmas, kas vērtētas brīvdienās. Tad dati tiek izmantoti rekomendāciju algoritmā, lai prognozētu interesējošo saturu.
- Konteksta pēc-filtrēšana. Šajā gadījumā modelis tiek trenēts ar visiem datiem un konteksta filtrs uzlikts rezultātiem.
- Kontekstuālā modelēšana – šajā gadījumā konteksts tiek pievienots modelī un rekomendācijas tiek rēķinātas, ņemot vērā kontekstu. Protams, pats modelis ir sarežģītāks.



2.2. att. RS konteksta pievienošana [18].

2.3. Ieteikumu sistēmu risinājumi

2.3.1 *Pētījumi VOD RS*

Vispirms tiks apskatīti risinājumi, kas veido VOD ieteikumus, un šādos scenārijos parasti pieejams lielāks datu apjoms un nav limitēta satura pieejamība laikā. Lielais vairums RS pētījumu ir saistīti ar VOD ieteikumiem, kuros plaši lietota gan dziļā mašīnmācīšanās [19], gan vienkāršāki mašīnmācīšanās modeļi [15]. VOD ir gan pateicīgāka platforma ieteikumu veidošanai, gan arī prasīgāka vide tieši ieteikumu dēļ – ja lineārajā TV bez ieteikumiem var nosacīti iztikt, jo TV kanālos ir satura plānotāji un saturs principā tiek rādīts nepārtraukti, tad VOD gadījumā, noskatoties vienu vienumu, nav loģiska turpinājuma – atkal saturs ir jāmeklē.

2.3.1.1 *Vispārējs VOD RS metožu novērtējums*

Bambini et al [20] ģenerēja VOD ieteikumus, izmantojot atmiņā bāzētu CF metodi ar vienums-vienums pieeju, izmantojot CF ar modeļa pieeju (matricu faktorizācija), kā arī CB metodes algoritmus. Tika novērots, ka CF metodes filmu ieteikumiem darbojās labāk kā CB metodes.

2.3.1.2 *Laika konteksta izmantošana VOD ieteikumu veidošanā*

Ullah et al 2012. gadā apskatījuši VOD ieteikumu sistēmas ar individuālām lietotāja laika preferencēm [21]. Tradicionālie modeļi tiek papildināti pēc sekojošas metodes:

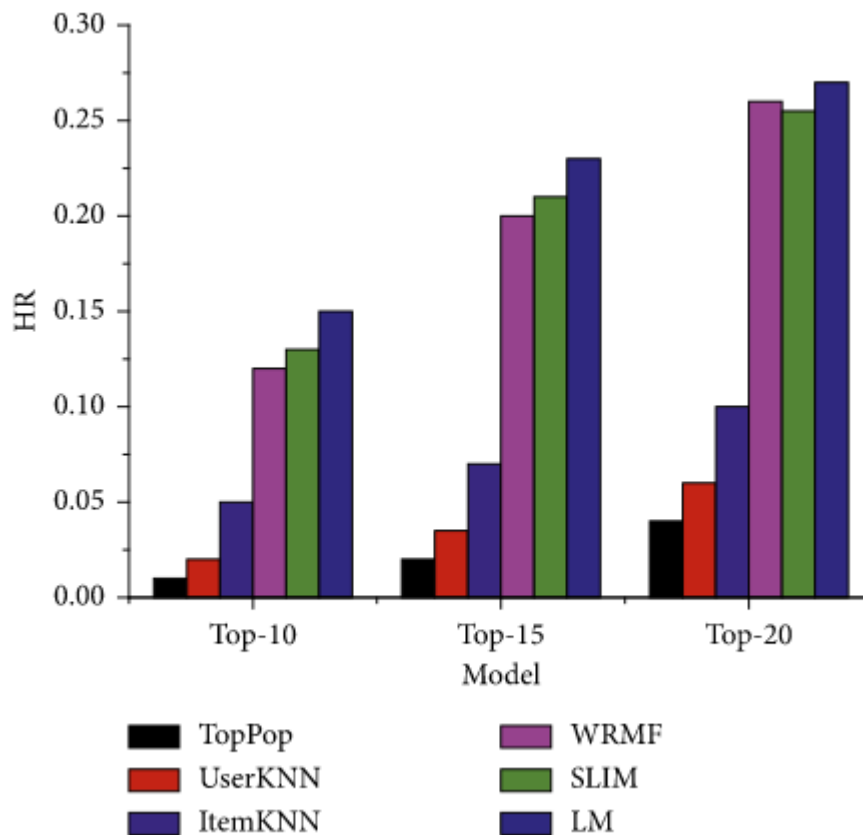
- Modelis tiek veidots nesaistē, lai RS veidotie ieteikumi tiktu nodoti klientam pēc iespējas ātrāk.
- Modelis katram lietotājam ieteikumus rēķina konkrētā laika periodā, lai samazinātu aprēķinu apjomu.
- Modelis ir daudz mazāks kā sākotnējā datu matrica, kas tiek izmantota citos CF modeļos, un līdzīgie vienumi tiek aprēķināti ātrāk.
- Šis modelis atrisina mērogojamības problēmu, jo tiek daudzkārtēji samazināts apstrādājamo datu apjoms, kā arī tiek ņemts vērā skatīšanās laika faktors un ieteikumi veidojas precīzāki.

2.3.1.3 RS kvalitātes izmaiņas laikā

Cremonesi et al fokusējušies uz VOD RS izpēti, izvērtējot gan CB, gan CF metodes (gan tuvāko kaimiņu pieeju, gan latentu faktoru pieeju), kā arī dažādas hibrīdu metodes. Viņi izvērtē arī RS evolūciju laikā, pieaugot lietotāju sniegtajām atsauksmēm par skatītajiem vienumiem. Tika noskaidrots, ka tuvāko kaimiņu metodes strādā labāk pie mazākiem datu apjomiem, taču pieaugot vēsturisko datu apjomam, latentu faktoru modeļi var sniegt labākus rezultātus, ja vien datos nav t.s. “garās astes uzvedība”, kur populārākie vienumi sastāda lielāko daļu skatījumu, savukārt, lielākā daļa vienumu praktiski nav skatīti. Viņi arī parādīja, ka lietotāja viedoklis par RS nekorelē ar vienu modeļa precizitātes parametru, drīzāk ar saliktu precizitātes, novitātes un nejausības salikumu. Kā arī tika parādīts, ka uz populārākajiem vienumiem balstīti ieteikumi strādā pārsteidzoši labi kā pirmie ieteikumi lietotājam, kurš nesen sācis lietot pakalpojumu un kura preferences nav skaidri zināmas [22].

2.3.1.4 Hibrīdu ieteikumu sistēmas un to salīdzinājums ar vienkāršākām metodēm

Yuan et al [23] 2021. gada rakstā salīdzina klasisko metožu ieteikumu sistēmu darbību ar autoru izstrādātu hibrīda risinājumu metodi, kur mijiedarbojas CF algoritms ar asociāciju meklēšanas algoritmu. Princips ir līdzīgs klasiskajiem CF modeļi bāzētajiem risinājumiem, izmantojot matricu faktorizāciju un latentu faktoru meklēšanu, un vērtējumu prognozēšana tiek veikta, izmantojot faktoru-vienumu un lietotāju-faktoru reizinājumu. Rezultāti tiek apvienoti ar asociāciju meklēšanas algoritmu. Respektīvi, netiek ieteikts saturs, kur asociācijas atbalsts ir zem noteikta līmeņa. Modeļa performance tiek salīdzināta ar vienkāršu topu ieteikumiem, vienkāršiem CF lietotājs-lietotājs un viens-vienums risinājumiem, kā arī ar divu hibrīda modeļu performanci - svērtā regulētā matricas faktorizācija modeļa (WRMF) un retā lineārā modeļa (SLIM) (attēls 2.4).



2.4. att. Dažādu modeļu performances salīdzinājums pēc HR indeksa [23].

HR indekss tiek definēts formulā (2.2) [24]:

$$HR = \frac{|U_{hit}^L|}{|U_{all}|}, \quad (2.2)$$

kur $|U_{hit}^L|$ ir lietotāju skaits, kuriem pareizā atbilde ir iekļauta top L ieteikumu sarakstā un $|U_{all}|$ ir kopējais lietotāju skaits datu kopā.

2.3.2 Pētījumi EPG RS

Šajā nodaļā tiks apskatīti EPG satura ieteikumu sistēmu risinājumi un pieejas. EPG ieteikumi ir mazāk populāri gan jau 2.3.1 nodaļā minēto iemeslu dēļ, gan arī tāpēc, ka EPG metadati ir sliktākas kvalitātes – satura iepircēji pērk kanālus, kuri tālāk nodrošina metadatu kvalitāti, bet VOD tās bieži ir individuālas filmas, kuras bez kvalitatīviem metadatiem nevarētu atrast TV

platformā. EPG var nedefinēt bāzes algoritmus: populārākie kanāli, populārākie lietotāja kanāli un populārākie lietotāja kanāli dotajā laika period [25], ar kuriem lietotāju var primāri uzrunāt, neveidojot sarežģītas ieteikumu sistēmas.

2.3.2.1 Hibrīds *iALS1* modelis EPG ieteikumu sistēmām

Zibriczky et al 2013. gada rakstā [26] salīdzina dažādus risinājumus EPG rekomendācijām, izmantojot CF, CB metodes un populārāko vienumu ieteikumus. CB metodē kā salīdzināmie vienumi tiek izmantoti raidījumu metadatu vektori. T.i., tiek izmantoti dati ar lielāko piesesumu modelim: raidījumu nosaukumi, žanri, kanāli, kategorijas, vecuma ierobežojumi. Lietotāju pazīmju vektori tiek kalkulēti kā vidējā vērtība no vēsturiski skatītajiem raidījumiem, un ieteiktās filmas tiek noteiktas ar kosinusa līdzības algoritmu, aprēķinot lietotāju vektoru un filmu vektoru savstarpējo līdzību un tādā veidā veidojot ieteikumus. Kā CF metodes algoritms izmantots - *iALS1* [27] (mainīgu mazāko kvadrātu metode netiešām atsauksmēm). *iALS1* modelē lietotāja – vienuma preferenču matricu ar zema ranga tuvinājumu – formula (2.3), kur matrica R ir tuvinājums, ko apraksta divu zemāka rangu matricu reizinājums. Uzdevums ir minimizēt prognozes kļūdu lietotāja preferenču noteikšanā. Te var tikt prognozēts gan vērtējums, gan netieša interese par raidījumu un optimizācijas uzdevumu veic ar gradienta nolaišanas metodi un mainīgu mazāko kvadrātu metodi. Savukārt, autoru populārāko vienumu algoritmu loģika ir rekomendēt globāli populārākos seriālus un filmas, kā arī jau iesākto seriālu nākamās sērijas [26].

$$R \approx PQ^T, \quad (2.3)$$

kur R lietotāja – vienuma mijiedarbības matrica,

P - lietotāja latentu pazīmju matrica,

Q – vienumu latentu pazīmju matrica.

Papildus tam Zibriczky et al eksperimentēja ar hibrīdu risinājumu CF un CB metodēm, integrējot vienumu metadatus faktorizācijas modelī (hibrīda *iALS1*), kura galvenā doma ir uzlabot ieteikumus lietotājiem, kuriem ir maz raidījumu skatījumu (netiešie satura vērtējumi, kurus var izsecināt, ņemot vērā to, cik daudz lietotājs konkrēto vienību ir noskatījies). Hibrīda risinājuma ideja ir veidot matricu $R^{(U+M) \times I}$ kā divu matricu apvienojumu – viena ir sākotnējā matrica R ar netiešajiem vērtējumiem, otra – matrica S , kas ir metadati - vienums matrica, kura pēc būtības ir līdzīga lietotājs – vienums matricai, kur elements $s_{ij} = 1$, ja metadati atbilst satura vienībai un 0, ja neatbilst. Metadatus var uzskatīt kā pseido - lietotājus un ar tiem saistītos notikumus kā pseido –

notikumus un šo matricu var faktorizēt ar iALS1. Līdz ar to, ja ir pieejami tikai metadati, var ģenerēt jēgpilnus ieteikumus lietotājam un gadījumā, ja parādās reāli skatījumi, šī daļa kļūst par dominējošo un tiek nodrošināta gluda pāreja no CB metodes uz CF metodi.

2.3.2.2 *Laika konteksta izmantošana EPG ieteikumu sistēmās*

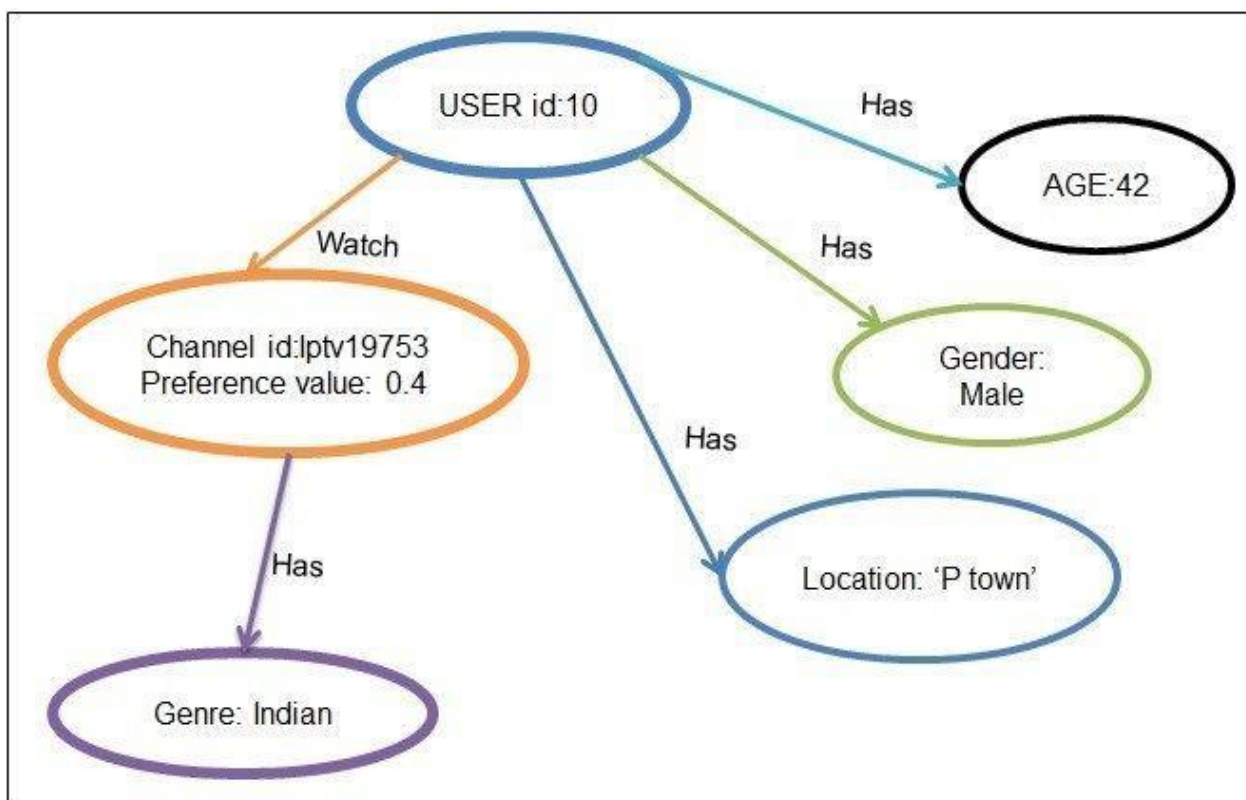
Kim et al savā 2018. gada rakstā [28] apskatītajā EPG RS risinājumā balstās uz programmu žanriem un kanāliem kā galvenajiem RS parametriem, papildinot tos ar laika kontekstu. T.i., par cik lietotāju iecienītās pārraides lineārajā TV tiek raidītas periodiski, lietotāju skatīšanās paradumi arī bieži novērojami periodiski. Līdz ar to TV ieteikumiem jāņem vērā arī laika konteksts, lai izvērtētu lietotāja skatīšanās paradumus [25]. Kā arī lietotāja skatījumus var raksturot vairāki periodiskas skatīšanās apvienojumi, piemēram, ja lietotājs skatās ziņas 21.00 katru dienu, tā būs viena periodiska uzvedība, ko var raksturot ar stundu periodiskumu, kur notikums atkārtojas vienā laikā nedēļas ietvaros, taču, ja filmas vai sporta pārraides pārsvarā tiek raidītas, piemēram, sestdienā, vai lietotājs TV skatās tikai konkrētās dienās, to var raksturot dienu periodiskums. Vairākas šādas periodiskas uzvedības tiek apvienotas, lai aprēķinātu lietotāja TV skatīšanās periodiskumu līdz ar to RS jāmodelē katra lietotāja vairākas laika kontekstuālās preferences [28]. Līdz ar to Kim et al piedāvā metodi, kura RS veido, apvienojot dažādus laika kontekstus bez papildu informācijas par TV programmām (neskaitot jau minēto žanru un kanālu), un šis risinājums ir izmantojams arī, piemēram, digitālajai TV, kur informācija par skatījumiem netiek padota pakalpojuma sniedzējam un rekomendācijas jāērķina pašā TV satura dekodētājā. Lai efektīvi modelētu sarežģītus laika skatīšanās paradumus, tiek piedāvāta metode, kas veic vairākas laika konteksta profilēšanas un veido dažādas laika preferenču matricas ar kanāliem un žanriem. Piedāvātā metode apvieno dažādās laika preferenču matricas ar lietotāja iecienītajiem kanāliem un žanriem. Tā kā tiek veidoti vairāki preferenču modeļi, tie var savstarpēji konfliktēt. Lai to risinātu, tiek izstrādāts kvadrātiskās programmēšanas modelis, kur optimizēšanas problēma tiek formulēta ar pēc iespējas mazāku ierobežojumu skaitu, lai optimizācijas process būtu ātrs un mērogojams. Kanālu un žanru preferences tiek apvienotas, lai veidotu tiešraides rekomendācijas pašā EPG platformā [28].

2.3.2.3 *Ontoloģiju izmantošana IPTV RS*

Khan et al 2018. piedāvā kontekstuālu ontoloģiju hibrīda RS IPTV (COHRS) [4]. Atsaucoties uz to, ka esošie risinājumi cieš no CS problēmas, modeļu pārāk lielas pielāgošanas datiem un datu nepietiekamības, autori piedāvā lietot ontoloģijas pieeju. Kā zināms, ontoloģija

datorzinātnes kontekstā apzīmē zināšanas noteiktā jomā, kas noformētas gan cilvēkam saprotamā, gan mašīnlasāmā formā. Dažādas vienību pazīmes varētu formulēt zem noteiktas klases. Klases instances ir specifiskas, un tām ir atšķirīgas vērtības. Attiecība starp klasēm parasti ir binārā semantiskā attiecība. Piemēram, klasi “Lietotājs” un klasi “Dzimums” varētu semantiski saistīt ar rekvizītu “ir”. Pārveidojot semantisko zināšanu kontekstā, tas nozīmē, ka “Lietotājam” ir “Dzimums”. Ontoloģiju galvenā priekšrocība pār modeli, būvētu uz klasiskiem atslēgas vārdiem, ir bagātākas semantiskās zināšanas un izpratne par noteiktu domēnu [29].

Lietotāju raksturojošie dati un lietotāja preferences no neapstrādātiem datiem tiek pārveidotas, un tiek būvēti ontoloģiskie lietotāju profili (2.4.att). Ontoloģiskā pieeja papildina lietotāju profilēšanas procesu, tādējādi uzlabojot ieteikumu sistēmas precizitāti. Eksperimentālie novērtējumi liecina, ka COHRS spēj novērst iepriekš apskatītos RS trūkumus, piemēram, pārmērīgu piemērošanu un procesu padarīt efektīvāku [4].



2.4. att. Piemērs ontoloģiju lietotāja profilam [4].

2.3.2.4 CF izmantošana IPTV rekomendācijām

Kim et al [30] 2011. gadā piedāvāja shēmu, kā veidot personalizētus TV ieteikumus lineārās TV skatītājiem ar CF metodi – iztiekot bez vienumu vērtējumiem, bet izsecinot vai saturs ir

interesējies lietotāju vai nē, balstoties uz to, cik daudz no attiecīgā vienuma raidlaika lietotājs noskatījies. Lietotāji vispirms ir sagrupēti līdzīgu lietotāju klasteros, balstoties uz skatīšanās paradumiem. Lai veidotu ieteikumus, tiek izmantota CF metode, kas ļauj konkrētajam lietotāju klasterim piemeklēt potenciāli interesantākos raidījums pēc to metadatiem, un šiem raidījumiem tiek piešķirts rangs ar autoru piedāvātu metodi, kuras pamatā ir “*Best Match*” modelis, kurš tiek izmantots meklētājos, lai atgrieztu vaicājumam atbilstošākos rezultātus. Šajā gadījumā lietotāja intereses var saukt par vaicājumu un vienumu metadatus – par rezultātu. Kad noskaidroti top-N ranga raidījumi, lietotājam tiek veidoti personalizēti ieteikumi. Šāda metode vidēji dod 77% precizitāti.

2.3.2.5 IPTV ieteikumu veidošana, izmantojot raidījumu subtitrus

Rautiainen et al 2013. gadā piedāvā sistēmu, kas palīdz lietotājam izvēlēties saturu, ko skatīties nelineāri – arhīvā vai ierakstā (saturu ar nobīdi laikā) [31]. Risinājuma pamatā tiek indeksētas TV tiešraides un, izmantojot mašīnmācīšanās un datizraces metodes, iegūti vienumu raksturojošu vārdu kopsavilkumi, kā arī notverti svarīgāko momentu attēli. Šo iegūto informāciju tālāk izmanto satura meklēšanai – gan meklējot saturu pēc atslēgas vārdiem, gan brīvā tekstā. Papildus tam no indeksētajiem TV raidījumiem tiek veidota CB RS nelineārajam saturam, kur lietotājs saturu var izvēlēties pēc atbilstošajiem attēliem. Eksperimentu rezultāti liecina, ka piedāvātie CB risinājumi lietotāju vidū bija populārāki kā klasiskā satura meklēšana pēc nosaukumiem un metadatiem. Īpaši izteikti tas bija nelineārajā TV.

2.4. Ieteikumu skaidrojuma nozīme

Knijnenburg et al [32] 2012. gadā apskatīja ieteikumu paskaidrojumu ietekmi uz lietotāju apmierinātību ar RS un tās lietotāju pieredzi. Lai izvērtētu ieteikumu kvalitāti, autori izvēlējās lietotāju centrētu pieeju. Metode testēta ar A/B testiem un aptaujām. Visos autoru pētījumos tiek novērots, ka uztvere par RS kvalitāti un daudzveidību ir svarīgi starpnieki, lai noteiktu objektīvu sistēmas aspektu ietekmi uz trim lietotāju pieredzes sastāvdaļām: procesu (piemēram, cik tas ir sarežģīts), sistēmu (piemēram, sistēmas efektivitāte) un iznākumu (piemēram, apmierinātība). Testējot RS ar reāliem lietotājiem, algoritmus nevar testēt izolēti; ir jāņem vērā dažādi sistēmas aspekti un personiskās un situāciju īpašības jāapvieno, lai izvērtētu lietotāju pieredzi. Piedāvātās

pieejas nostādnes ļauj uzņēmumam būt soli tuvāk lietotājam, kuram algoritma precizitāte var būt sekundāra.

Ieteikumu skaidrojumi spēlē lielu lomu RS ieviešanā. Ja lietotājam tiek paskaidrots, kā personalizētais ieteikums veidots, tas var uzlabot ticību ieteikumu kvalitātei un ieteikumus tādejādi ir vērts papildināt ar īsu paskaidrojumu kā “citi lietotāji, kas skatījās šo, skatījās arī šo” vai “šis viens ir līdzīgs šim, ko esat augstu vērtējis”. Līdzīgu praksi varam novērot Amazon, Google, Netflix un citos lielo kompāniju ieteikumos.

2.5. Metožu precizitātes novērtēšana

Lai novērtētu RS efektivitāti, tipiski tiek mērīta un novērtēta kļūdas metrika (t.i., vidējā kvadrātiskā kļūda un vidējā absolūtā kļūda) vai klasifikācijas uzdevumu precizitātes metrikas, ko var aprēķināt no klasifikācijas kļūdu matricas (t.i., precizitātes (*precision*) formula (2.3), atdeves (*recall*) formula (2.4), kā arī nekorekti ieteikto vienumu īpatsvara (*fallout*) formula (2.5)).

$$P = \frac{TP}{TP+FP}, \quad (2.3)$$

$$R = \frac{TP}{TP+FN}, \quad (2.4)$$

$$F = \frac{FP}{FP+TN}, \quad (2.5)$$

kur TP ir testa rezultāts, kur korekti prognozēta interese, FP – testa rezultāts, kur nekorekti prognozēta interese, TN – testa rezultāts, kur korekti prognozēts, ka viens neinteresē, FN – testa viens, kur nekorekti prognozēts, ka viens neinteresē. Ideālas klasifikācijas gadījumā FP un FN tiecas uz nulli, un precizitāte, atdeve tiecas uz viens, savukārt, nekorekti ieteikto vienumu daļa – uz nulli [33].

RS atšķirībā no standarta klasifikācijas uzdevumiem nav vienas pareizās atbildes.

Modeļa precizitāti var izvērtēt nesaistē – analizējot vēsturiskos skatījumus un izmantojot *krosvalidācijas* metodes vai arī prognozējot lietotāju vērtējumus izlaistajiem vienumiem (izmantojot, piemēram, *leave-one-out* metodi). Otrs variants ir risinājuma izvērtēšana tiešsaistē jeb atgriezeniskās saites analizēšana no produkcijā esošas RS. Šajā gadījumā atdevi varētu definēt kā

ieteiktā vienuma noskatīšanos konkrētā laika periodā. *ContentWise* kompānijas 2010. gada rakstā par ieteikumu sistēmām ieviests empīriskās atdeves jēdziens, kurš tiek aprēķināts kā ieteikto vienumu skatījumu skaits pret visu skatījumu skaitu lietotāja līmenī noteiktā laika nogrieznī [14]. Šis laiks var būt definēts no pāris stundām līdz pāris dienām atkarībā no vides, kur ieteikums tiek nodots. Lineārās TV kontekstā, protams, galvenais ierobežojums ir laiks, kad saturs ir pieejams. Rakstā arī minēts interesants fenomens – rekomendāciju atdeve ir labāka mazāk populāru vienību ieteikumiem. Protams, risinājumu laižot produkcijā, ir jābūt pārliecībai, ka ieteikumi strādā labi. Lin et al parādīja [34], ka modeļi, kuri kā modeļa vērtēšanas mērus ņem vērā gan precizitāti, gan dažādību, gan novitāti, pārspēj tradicionālos modeļus, kur modeļus izvērtē tikai pēc precizitātes mēra.

Papildus jau minētajiem RS kritērijiem ir būtiski, lai ieteikumi nekoncentrējas uz mazu daļu vienumu. Īpaši, ja runa ir par VOD ieteikumiem. T.i., lai, lietotājam noskatoties “Harijs Poters un filozofu akmens”, visas rekomendācijas nesastāvētu tikai no nākamajām franšīzes daļām. Nav jēgas no ieteikumiem, ja tie neliek lietotājam izpētīt kaut ko ārpus jau ierastā. Līdzīgi kā ieteikumi skatīties “Rīta Panorāma” lietotājam, kas to jau regulāri dara, ir bezjēdzīgi. Līdz ar to varam arī ieviest mēru, kas mēra cik liela daļa no satura klāsta tiek ieteikta.

Lai pārbaudītu RS darbību produkcijā, var izmantot “A/B testu” metodi, kur līdzīgiem lietotājiem iesaka vienumus pēc dažādām metodēm, vai arī atstāt t.s. kontroles grupu, kurai ieteikumus vispār neatrāda, un salīdzināt rezultātus starp šīm grupām. T.i., vai lietotāji, kuriem satura ieteikumus nosūtīja, ieteikto saturu skatījās biežāk kā tie, kuriem ieteikumi netika sūtīti un kuri to noskatījās bez ieteikumu palīdzības (saturu atrodot bez ieteikumiem).

Papildus tam var tikt veidotas aptaujas un organizētas intervijas ar lietotājiem, lai saprastu vai ieteikumi atbilst interesēm vai nē, kā arī saprastu, kāpēc.

2.6. EPG ieteikumu sistēmu problēmas apgabals un izmērs

Apskatot esošos risinājumus EPG ieteikumu veidošanā, skaidrs, ka nav vienotas pieejas un viens labākais risinājums, kā arī tie stipri atšķiras no VOD ieteikumu veidošanas pieejām. Ieteikumu sistēmu lauks ir ļoti plašs, pieejas ir dažādas, taču literatūrā nav piedāvāts pilnīgs

risinājums, un, veidojot jaunu ieteikumu sistēmu, jāsaskaras ar datu kvalitātes problēmām, jāanalizē dati un skatīšanās paradumi, lai noteiktu lietotāja interesi par konkrētiem satura vienumiem, lai varētu veidot korektu lietotāju-vienumu mijiedarbību matricu. Papildus tam nav apskatīta Latvijai raksturīgā daudzu valodu problēma, kas atspoguļojas gan metadatu apstrādē, gan ieteikumu veidošanā, kā arī nav apskatītas metodes, kā metadatu katalogu papildināt vai, piemēram, izmantot tēmu modelēšanas metodes, izvilkt vienumu anotāciju tēmas un izmantot tās ieteikumu veidošanai. Ieteikumu sistēmas izveide ir komplicēts datu ieguves, analīzes, tīrīšanas process, kurā nepieciešams izmantot dažādas statistikas un datizraces metodes, lai nonāktu līdz vēlamajam rezultātam.

3. AUTORA IEGULDĪJUMS, RISINĀJUMS, PIEEJA

Darba autors maģistra darba ietvaros izstrādājis EPG ieteikumu sistēmu, kura katram televīzijas pakalpojuma abonētājam var aprēķināt personalizētus satura ieteikumus. Risinājuma būtība ir - ņemot vērā IPTV skatīšanās specifiku un katram lietotājam raksturīgos skatīšanās paradumus, veidot satura ieteikumus, apvienojot rezultātu no vairākiem ieteikumiem ģenerējošajiem algoritmiem. Darbā izstrādāto sistēmu iespējams mērogot un integrēt TV EPG katalogā, uzlabojot lietotāja skatīšanās pieredzi. Apskatīti gan uz satura metadatiem, gan lietotāju-vienumu mijiedarbību matricām balstīti algoritmi, kā arī ņemts vienkāršs populārākā satura aprēķins, lai veidotu personalizētus ieteikumus. Ieteikumi tiek aprēķināti un var tikt filtrēti gan tiešraides un nākotnes saturam, gan arhīvā pieejamam saturam.

Maģistra darba ietvaros darba autors ir iepazīties ar ieteikumu veidošanas paradigmām un tām atbilstošajiem algoritmiem gan VOD ieteikumu, gan EPG ieteikumu aprēķināšanai un izvēlēties literatūrā efektīvus un datiem piemērotus algoritmus un pieejas ieteikumu veidošanai. Ieteikumu sistēmu veidošana literatūrā pamatā aprobežojas ar to, kurš modelis dos labāku rezultātu, kā uzlabot ātrdarbību un kā apstrādāt lielus datu apjomus, taču nav pieejama informācija par to, kā neapstrādātiem skatījuma datiem ģenerēt reālus ieteikumus – maģistra darba ietvaros, soli pa solim tiek nonākts līdz rezultātam, sākot ar datu ieguvī no SIA Tet un beidzot ar ieteikumu sistēmas rezultātu pārbaudi.

Autors maģistra darbā koncentrējas uz ieteikumu sistēmas koncepta izveidi un verificēšanu, kuru potenciāli varētu tālāk attīstīt, pielāgot SIA Tet vajadzībām un izmantot Tet+ un iTV platformās. Darbs sevī ietver datu ielasīšanu, skatījumu datu un metadatu, lietotāju paradumu analīzi, datu tīrīšanu un papildu datu ieguvī, satura intereses noteikšanu, piemērotu modeļu meklēšanu, trenēšanu un testēšanu uz vēsturiskiem skatījumu datiem. Darba mērķis ir apzināt, ar kādām problēmām nākas saskarties, veidojot ieteikumu sistēmas, un saprast, vai plānotā pieeja strādās labu ieteikumu veidošanai.

5. IETEIKUMU SISTĒMAS VEIDOŠANA

Šajā nodaļā tiks apskatīta datu ielasīšana, analīze, apstrāde un apsvērumi filtru un kontekstu ieviešanai, kā arī interešu definēšana un ieteikumu aprēķina loģika.

5.1. Datu pieejamība un apraksts

Darba veikšanai tiek izmantoti sekojoši datu avoti no SIA Tet televīzijas analītikas datiem par 2020. gada TV sezonu (2020. gada septembris – 2021. gada augusts):

- f_tv_traffic tabula (5.1 tabula):

fakta tabula par TV skatīšanās notikumiem, kura sastāv no 53 kolonām, no kurām ieteikumu sistēmai var izmantot 5.1. tabulā attēlotos laukus. Šajā tabulā uzkrātas lietotāja – satura vienības mijiedarbības lietotāja un servisa līmenī, kā arī atspoguļots, vai notikums skatīts tiešraidē, arhīvā, ierakstā. Skatīšanās ilguma aprēķinā ņemts maksimāli 3h intervāls pēc pēdējās lietotāja mijiedarbības ar TV, lai izslēgtu garo notikumu uzskaiti, kur TV saturs netiek skatīts. Kopā par apskatīto periodu pieejami 2.3 miljardi ierakstu.

5.1. tabula

f_tv_trafiks kolonu tips un apraksts

Kolonnas nosaukums tabulā	Datu tips	Datu skaidrojums
idd_wh_service	BIGINT	unikāls abonēšanas servisa identifikators
broadcast_id	BIGINT	pārraidē un unikālais ID (atslēga, pēc kuras apvienot ar TV skatījumu tabulu)
event_type	STRING	notikuma tips – vai skatīta tiešraide, arhīvs, lietota pauze, utt.

Kolonnas nosaukums tabulā	Datu tips	Datu skaidrojums
len_broadcast_s_calc	BIGINT	notikuma skatīšanās ilgums sekundēs
d_epg_start	DATE	pārtraides skatīšanās sākuma datums
d_epg_end	DATE	pārtraides skatīšanās beigu datums
dt_epg_start	TIMESTAMP	pārtraides skatīšanās sākuma datums un laiks
dt_epg_end	TIMESTAMP	pārtraides skatīšanās beigu datums un laiks

- *d_epg_metadata* tabula (5.2 tabula):

dimensiju tabula, kas sastāv no 53 kolonām, no kurām ieteikumu sistēmai var izmantot laukus, kas aprakstīti 7.2. tabulā. *d_epg_metadata* ir tabula, kurā uzkrāti TV kanālu pārtraīžu metadati, tādi kā pārtraides sākuma un beigu laiks, pieejamība arhīvā, nosaukums, žanrs, cenzūra, epizodes numurs, anotācija u.c. Kopā par apskatīto periodu pieejami 1.6 miljoni ierakstu.

5.2. tabula

d_epg_metadata kolonu tips un apraksts

Kolonnas nosaukums tabulā	Datu tips	Datu skaidrojums
broadcastid	BIGINT	pārtraides unikālais ID (atslēga, pēc kuras apvienot ar TV skatījumu tabulu)
d_epg_start	DATE	pārtraides sākuma datums
d_epg_end	DATE	pārtraides beigu datums
dt_epg_start_timestamp	DATETIME	pārtraides sākuma datums un laiks
dt_epg_end_timestamp	DATETIME	pārtraides beigu datums un laiks
episode_id	INT	pārtraides epizodes numurs
series_id	BIGINT	unikālais pārtraides ID (TV šoviem un seriāliem)

Kolonnas nosaukums tabulā	Datu tips	Datu skaidrojums
season_nr	INT	pārraides sezonas numurs
censure	INT	pārraides vecuma ierobežojums
is_adult	BOOLEAN	vai saturs ir paredzēts pieaugušajiem
series_title_original	STRING	seriāla, šova oriģinālais nosaukums
title_full_original	STRING	vienības oriģinālais nosaukums
title_full_lv	STRING	vienības nosaukums angļiski
title_full_lv	STRING	vienības nosaukums latviski
title_full_ru	STRING	vienības nosaukums krieviski
series_title_en	STRING	seriāla, šova nosaukums angļiski
series_title_lv	STRING	seriāla, šova nosaukums latviski
series_title_ru	STRING	seriāla, šova nosaukums krieviski
summary_long_en	STRING	pārraides anotācija
series_summary_long_en	STRING	seriāla, šova anotācija
genre_en	STRING	saturs vienības žanrs angļiski
recordable	BOOLEAN	vai pārraidi var ierakstīt
restartable	BOOLEAN	vai pārraide pieejama arhīvā
genre_l1	STRING	žanra tipa pirmais līmenis
channel_name	STRING	Kanāla nosaukums (atslēga, pēc kuras apvienot ar TV kanālu tabulu)

- d_tv_channel_metadata tabula (5.3 tabula):

dimensiju tabula, kurā apkopoti dati par kanālu pieejamajām audio valodām un kanālu tematiku, piemēram, filmu kanāli, populārzinātniskie kanāli, bērnu kanāli. Tabulā par attiecīgo laika periodu pieejami 219 ieraksti.

x_tv_channel_metadata kolonu tips un apraksts

Kolonnas nosaukums tabulā	Datu tips	Datu skaidrojums
title_main	STRING	Kanāla nosaukums (atslēga, pēc kuras apvienot ar EPG metadatu tabulu)
genre_name	STRING	Kanālu grupas žanrs
languages	STRING	Kanālā pieejamās audio valodas

Datu iegūšanai un analīzei tiek izmantota brīvpieejas datorprogramma Knime, pieslēdzoties SIA Tet Apache Hadoop datu ezera platformai ar lielo datu dzini Impala. Visai datu atlasei no *d_epg_metadata* tabulas tiek izmantots filtrs `is_adult=False`. Modeļu veidošanai un pārbaudei faktiski izmantota tikai puse no laika perioda – respektīvi, 2020. gada 1. novembris līdz 2021. gada 10. aprīlis, lai izvairītos lielu sporta notikumu un no vasaras sezonālītātes datus.

5.2. Datu kvalitātes novērtēšana un datu tīrīšana

Šajā sadaļā tiks aprakstīta datu izpēte, sīkāk izvērtēti pielietojamie lauki gan no datu aizpildījuma un kvalitātes viedokļa, gan no lietojamības TV satura ieteikumu veidošanā.

Ieteikumu aprēķinā atsevišķi tiks izdalīti filmu, seriālu un šovu dati un skatījumi to konceptuālo atšķirību un aprakstošo datu dēļ. *d_epg_metadata* katrai satura vienībai ir pieejams unikāls TV pārraides ID, kā arī satura vienību nosaukums, taču filmu gadījumā nav pieejami tādi aprakstošie lauki kā seriāla nosaukums, epizodes numurs, sezonas numurs un seriāla apraksts, līdz ar ko, veidojot filmu ieteikumus, šos laukus nav jāņem vērā. Seriāliem vienības nosaukums ir epizodes nosaukums. Veidojot seriālu ieteikumus, nepieciešams darboties visa seriāla nevis epizodes līmenī un izmantot datus, kas saistīti ar seriāla nosaukumu un seriālu aprakstošajiem datiem.

Abām iepriekš minētajām grupām tiek veikta atbilstošo datu analīze, tīrīšana un uzlabošana. Datu tīrīšana ir būtiska vairāku iemeslu dēļ – pirmkārt, lai varētu viennozīmīgi noteikt satura vienības, kuras lietotājs noskatījies neatkarīgi no kanāla un valodas, otrkārt, lai varētu pēc iespējas precīzāk pielasīt papildu metadatus no citām datubāzēm.

5.2.1 Datu kvalitāte filmām

Veidojot datu pieprasījumu no *d_epg_metadata* tabulas, lietojam filtru `show_type='movie'`, atlasot datus par filmām. Izvērtējot datu aizpildījumu (5.4 tabula), novērtēts, ka nepilnīgi aizpildīti dati par dažādiem filmu raksturojošajiem parametriem: filmu nosaukumiem dažādās valodās, kā arī cenzūru un anotāciju, papildu tam novērtēts, ka 9% un 10% gadījumu aizpildīti arī tādi lauki, kuri uz filmām neattiecas (*season_nr*, *episode_id*).

5.4. tabula

Nepilnīgi aizpildītie datu lauki vienību tipam “movie”

Kolonas nosaukums	Kolonas aizpildījums (% no visiem ierakstiem)
<i>episode_id</i>	9%
<i>season_nr</i>	10%
<i>censure</i>	26%
<i>summary_long_en</i>	45%
<i>title_full_lv</i>	52%
<i>title_full_en</i>	58%
<i>title_full_ru</i>	75%

Manuāli izvērtējot šos gadījumus, nonākts pie secinājuma, ka daļa šo ierakstu ir filmas, daļa – seriāli, taču ieraksti, kas atbilst seriāliem, nosaukuma laukā pārsvarā satur attiecīgi informāciju par epizodes numuru, sērijas numuru vai satur tādus vārdus kā ”epizode”, ”seriāls”, ”series” (piemēram, “Dr. K's Exotic Animal ER 4 (Season 4 Extension) - Hide N' Snake”), līdz ar to pieņemts lēmums šos ierakstus filtrēt pēc nosaukumiem. Papildu tam novērots, ka filmu nosaukumi mēdz atšķirties gan viena kanāla, gan dažādu kanālu ietvaros, un tāpat novērotas valodu

neatbilstības dažādos vienības nosaukumu laukos. Nosaukumu tīrīšanai uzrakstīts *python* kods, datus apstrādājot kā *pandas* datu tabulas, kur pārsvarā izmantotas regulārās izteiksmes (regex), lai meklētu, notīrītu un izvilktu teksta fragmentus no teksta. Koda paraugs 1. pielikumā minētajā repozitorijā.

Filmu datu un nosaukumu tīrīšanas rezultātā 5% no unikālajām filmu vienībām tiek notīrītas, jo satur kādu no seriālu atslēgas vārdiem. No atlikušajām filmām 38% pēc datu tīrīšanas tiek notīrīti kā dublikāti. Kopumā filmu datubāzē tiek notīrīti 41% ierakstu.

5.2.2 *Datu kvalitāte seriāliem un šoviem*

Turpmāk ar seriāliem jāsaprot seriāli un šovi, jo ar tiem maģistra darba ietvaros tiks strādāts līdzīgi. Seriālus no EPG kataloga atlasām līdzīgi kā filmas – tikai šajā gadījumā atlasām seriālus un šovus. Papildus liekam filtru pēc *genre_11 not in ('Sporta raidījumi', 'Mūzikas raidījumi', 'Ziņas')* - izmetam raidījumus, kas pārsvarā tiek skatīti tiešraidē vai nav seriāli (nav vairākas epizodes, sezonas). Sporta raidījumu gadījumā būtu jāreķina lietotāja sporta raidījumu intereses un jānosaka interesējošās komandas, kas nav šī maģistra darba mērķis. Tabulā 5.5. attēlots nepilnīgi aizpildīto datu lauku saraksts seriāliem.

Seriālu nosaukums pieejams tikai 33% no visiem ierakstiem, apvienojot *series_title_full_original*, *series_title_en* laukus, *title_full_original* pieejams pilnīgi visiem ierakstiem. To arī izmantosim, lai pielasītu seriāla nosaukumu trūkstošajiem ierakstiem. *title_full_lv*, *title_full_eng* lauki nav izmantojami, jo tajos bieži ir individuālu epizožu nosaukumi, savukārt, seriālu nosaukumi konkrētās valodās ir reti aizpildīti, līdz ar to seriālu nosaukumi pārsvarā pieejami vienā valodā.

Nosaukumu tīrīšanā seriāliem pārsvarā uzmanība tiek pievērsta kanālu pievienoto lieko vārdu notīrīšanai (tādu kā “tiešraide”, “jauna sezona”), atslēgas vārdu un daļu nosaukumu notīrīšanai, lai saglabātu tikai oriģinālo seriāla nosaukumu. Daļai raidījumu, kur nosaukumus neizdodas iztīrīt ar vispārīgām regulārajām izteiksmēm, tiek izveidots katalogs, kas šos nosaukumus izvelk no EPG nosaukuma lauka. Kopumā no unikālajām *series_title_original*, *series_title_en*, *series_title_lv*, *series_title_ru* pēc datu tīrīšanas saglabājas 39% no sākotnējā apjoma.

Nepilnīgi aizpildītie datu lauki vienību tipam “series” un “tv show”

Kolonnas nosaukums	Kolonas aizpildījums (% no visiem ierakstiem)
series_title_lv	16%
series_title_ru	24%
series_summary_long_en	26%
title_full_lv	27%
series_title_original	28%
series_title_en	30%
censure	31%
series_id	33%
season_nr	37%
summary_long_en	43%

5.3. Lietotāju IPTV skatīšanās paradumu analīze

Šajā sadaļā tiks analizēti SIA Tet IPTV lietotāju skatīšanās paradumi, izvērtēts, vai SIA Tet lietotāji var tikt iedalīti kādās grupās pēc to skatīšanās paradumiem un vai šīm lietotāju grupām pieejas ieteikumu veidošanā var atšķirties.

5.3.1 *Interesējošā satura noteikšana*

Viens no sarežģītājiem un neviennozīmīgajiem jautājumiem, apskatot un analizējot lineārās TV skatījumus, ir intereses noteikšana. Ieteikumu sistēmu aprakstos un paraugos bieži tiek apskatīti dati un risinājumi [35, 20], kur skatītās vienības, pārsvarā filmas, tiek vērtētas pilnībā vai daļēji kādā konkrētā skalā, dodot ieteikumu sistēmas izstrādātājiem pilnīgu informāciju par lietotāja preferencēm. Taču, lai šādu informāciju saņemtu, nepieciešamas papildus izstrādes pie produkta, papildus informācijas glabāšana datubāzēs, kā arī aktuāla ir datu kvalitāte un kvantitāte – vai

lietotājs, noskatoties filmu, pasūtot precī internetā vai izlasot žurnāla rakstu, to reāli arī novērtēs? Un – vai novērtējums tiešām vienmēr atspoguļo lietotāja domas par produktu?

Lineārās TV gadījumā šādu novērtējumu ievākšana ir nepraktiska un visticamāk būtu neefektīva, līdz ar to nākas strādāt ar skatījumu datiem jeb t.s. “netiešo interesi” (*implicit interest*). Literatūrā pieejami dažādi varianti, kā netiešo interesi interpretēt un lietot ieteikumu aprēķinos [27, 36]:

- A. Pārvērst interesi par bināru vērtību – attiecīgi, patika vai nepatika.
- B. Lietot skatījumu apjomu vai skaitu.
- C. Skatījumu skaitu vai apjomu normalizēt vai standartizēt.

Taču tāpat nepieciešams saprast vai jebkas, kas skatīts, var tikt pieņemts kā lietotājam interesējošs? Vai katru mijiedarbību nepieciešams interpretēt un ievietot modelī? Vai noskatītas 20 min no filmas, kas ir 90 min gara, liecina, ka lietotājam saturs interesēja vai gluži pretēji – ka neinteresēja? Vai varbūt nebija laika noskatīties visu? Vai ir kaut kāda robeža, kuru definēt, lai noteiktu, ka lietotājam konkrētais saturs interesē? Jautājums ir sarežģīts, ņemot vērā lietotāju dažādos skatīšanās paradumus un TV lomu cilvēka ikdienā. Viens lietotājs TV var ieslēgt fonā, veicot citus darbus, kāds var neizslēgt vai aizmirst TV izslēgt, aizejot uz veikalu, kāds var TV skatīties pirms iemigšanas, un realitātē saturs nemaz netiek skatīts.

Papildus tam satura vienības metadatos konkrētai satura vienības pārraidei tiek noteikts tikai sākuma un beigu laiks, taču nav zināms reālais brīdis, kad tiek raidītas reklāmas un cik tās ir garas. To nevar noteikt arī no skatīšanās fakta tabulas – skatījumiem tiek noteikts tikai sākuma un beigu laiks, līdz ar to no datiem var iegūt kopējo skatīšanās laiku uz konkrētu satura vienību, ieskaitot reklāmas. Latvijas Republikas Elektronisko plašsaziņas līdzekļu likuma [37] 42.panta (Reklāmas un televīzijas veikala apjoms) pirmais punkts skan sekojoši: “*Reklāmai un televīzijas veikalam (izņemot televīzijas veikala skatlogu) atvēlētais laiks elektronisko plašsaziņas līdzekļu televīzijas programmās no pulksten 6.00 līdz 18.00 un no pulksten 18.00 līdz 24.00 nedrīkst pārsniegt 20 procentus no attiecīgajos laika intervālos iekļautā satura.*”. Pēc būtības tas nozīmē, ka aptuveni likumiski 20% no raidlaika var sastādīt reklāmas, taču šis apjoms var variēt attiecīgo intervālu robežās.

Līdz ar ko darbā intereses noteikšanai pieņemam, ka satura vienība skaitās noskatīta, ja noskatīts 70% vai vairāk no raidlaika, pieņemot, ka reklāmas laiks var sasniegt 30% un reklāmas

var tikt pārtītas (skatoties ierakstā vai arhīvā) vai kanāls uz reklāmas laiku pārslēgts, vai pieņemot, ka kādu laiku no satura vienības lietotājs nenoskatās, piemēram, nokavē sākumu vai pārtrauc skatīties neilgi pirms beigām.

Ņemot vērā lineārās TV skatīšanās specifiku, ieteikumu sistēmas izstrādes laikā tiek ieviests arī kritērijs, kad par lietotāja interesi par konkrētu satura vienību vēl nevar spriest – t.i., pārslēdzot kanālu ar pulti, neizmantojot konkrētu kanāla numuru vai EPG interaktīvo kanāla izvēles iespēju, tiek reģistrēts īss skatīšanās notikums katrā pārslēgtajā kanālā. Otrs variants – beidzoties iepriekšējai satura vienībai, sāk raidīt nākamo – pēc iepriekšējās satura vienības noskatīšanās lietotājs var TV izslēgt, vai mainīt kanālu uz kādu tajā brīdī vairāk interesējošu, līdz ar to šis īsais notikums var tikt reģistrēts pat saturu neskatoties. Šo minimālo kritēriju definējam 5 minūšu apmērā.

5.3.1.1 Filmu intereses noteikšana

Filmu intereses noteikšanai, veidojot ieteikumu sistēmu, tiek pieņemts, ka noskatīta filma klientam interesē. Savukārt filmas, kas noskatītas daļēji (t.i., mazāk par 70%, bet vairāk par 5 min), neinteresē.

5.3.1.2 Seriālu un šovu intereses noteikšana

Seriālu un filmu gadījumā intereses noteikšana ir sarežģītāka, jo seriāli un šovi sastāv no vairākām satura vienībām – sērijām vai epizodēm. Lai noteiktu lietotāja interesi par šovu vai seriālu, veicam reālu skatīšanās analīzi noskatītajām epizodēm (kritēriji tiek pieņemti tādi pat kā filmu gadījumā). Analīzē izmantojam 2021. gada marta un aprīļa TV skatījumu datus servisa līmenī (viena TV pakalpojuma līmenī, vienam klientam tie var būt vairāki – var pieņemt, ka katrs ir citā fiziskā adresē). Analīzē izmantojam tikai tos seriālus vai šovus, kuri tiek raidīti (ir pieejami tiešraidē) abos mēnešos, un definējam interesi, ņemot vērā marta skatījumus, un izvērtējam vai interese saglabājas arī aprīlī. Tiek pieņemts, ka saturs interesē, ja ir noskatīta kaut viena epizode, taču tālāk interesi iedalām divās grupās – liela interese un maza interese. Interese tiek saukta par lielu, ja izpildās vismaz viens no nosacījumiem:

- A. Lietotājs noskatījies 3 vai vairāk epizodes.
- B. Lietotājs noskatījies 35% vai vairāk no pieejamajām epizodēm.

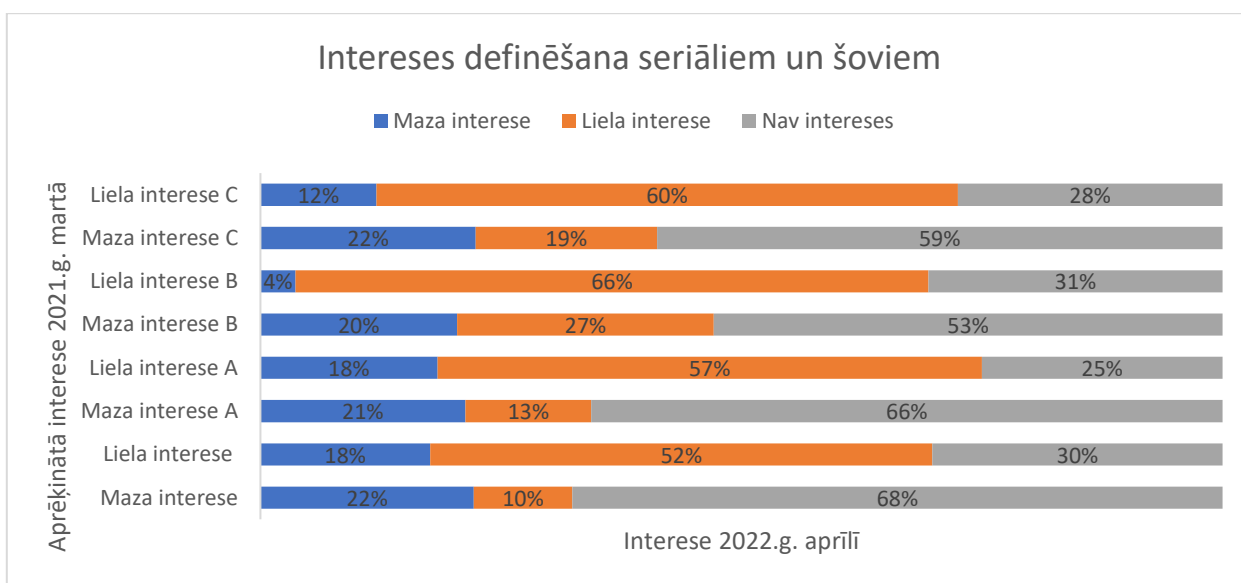
C. Lietotāja raidījuma skatījums sastāda 20% vai vairāk no visa seriālu un raidījumu skatīšanās laika.

Ja neviens no nosacījumiem neizpildās, tiek pieņemts, ka interese ir maza. Faktiskā situācija ar individuālu kritēriju noteikto interesi un tās saglabāšanos mēnesī redzama 7.1. attēlā. Kopā marta mēnesī liela interese noteikta 40% raidījumu un servisu kombināciju, maza – pārējiem 60%, kuri neatbilst kritērijiem.

Galvenais kritērijs intereses noteikšanai ir A kritērijs – liela interese tiek noteikta 32% raidījumu un servisu kombināciju. Kā redzams 7.1. attēlā “Liela interese A” – kritērijam 57% gadījumu saglabājas liela interese arī nākamajā mēnesī, 18% gadījumu samazinās, 25% gadījumu pazūd. Savukārt, maza interese marta mēnesī aprīļa mēnesī 21% gadījumu saglabājas maza, 13% gadījumu palielinās un 66% gadījumu pazūd. Kritērijā A, epizožu skaitu samazinot līdz divām, liela interese aprīļa mēnesī saglabājas tikai 44% lietotāju.

Kritērijs B ieviests, jo mūsdienās populāri kļuvuši seriāli ar mazu epizožu skaitu, kā arī tāpēc, lai interesi noteiktu arī seriāliem, kuri nesen sākuši rādīt TV ekrānos un, piemēram, fiziski bijis iespējams noskatīties tikai divas sērijas. Šis kritērijs atbilst 19% raidījumu un servisu kombinācijām.

Kritērijs C, savukārt, pielasa seriālu interesi tiem lietotājiem, kuri vispār TV skatās maz – t.i., ja 20% no kopējā mēneša skatījumu skaita tiek veltīts konkrētam seriālam, varam pieņemt, ka šis saturs lietotājam interesē. C kritērijs pielasās ~1% servisu – raidījumu kombināciju.



5.1. att. Intereses definēšana seriāliem un šoviem

5.3.2 Valodas izvēle

Latviešu un krievu valodas jautājums Latvijā ir aktuāls sarunu temats gan iedzīvotāju, gan politiķu, gan TV satura veidotāju, gan izplatītāju vidū. SIA Tet klienti ir dažādu nacionalitāšu, un arī skatīšanās paradumi valodu ziņā nav vienkāršs jautājums. Galvenais ieteikumu sistēmas mērķis ir veicināt klientu apmierinātību ar pakalpojumu un palielināt skatīšanās apjomu, līdz ar to jāizvairās lietotājam ieteikt saturu valodā, kuru lietotājs nesaprot, nelieto vai nevēlas izmantot.

Viens veids, kā atlasīt klientam aktuālāko valodu, būtu izmantot klienta paša norādītu vēlamu saziņas valodu. Šie dati ir klienta līmenī, taču skatījumi ir servisa līmenī, pie kam no datiem nevar noteikt vai reālais pakalpojuma lietotājs ir tas pats, kas pakalpojuma abonētājs. Var gadīties, ka īpašums, kur pieslēgts pakalpojums, tiek izīrēts, vai pakalpojumu lieto cits ģimenes loceklis. Jāpatur prātā arī tas, ka lietotājs var skatīties saturu vairākās valodās.

Reālākais veids, kā pielasīt lietotājam interesējošu valodu, ir analizējot kanālus, kuri tiek skatīti. Kanālu metadatos ir pieejama informācija gan par kanālu audio valodas iespējām, gan par subtitru valodām. Pēc šiem datiem varam aptuveni noteikt, kurš lietotājs skatās saturu latviski, kurš – krievu valodā un kurš skatās arī angļu valodas saturu, u.t.t., un no šiem pieņēmumiem veidot papildus filtrus ieteikumu sistēmas aprēķinātajiem rezultātiem.

Izmantojot datizraces klasteru veidošanas algoritmu KNN un t.s. “elkoņa metodi”, var noteikt klasterus, kādos iedalās lietotāji pēc viņu skatīšanās valodu preferencēm. Tiek izmantoti dati laika periodā no 2020. gada septembra līdz 2021. gada aprīlim, un dati tiek ņemti kā īpatsvars no visa skatīšanās apjoma servisa līmenī. Tiek ņemtas tikai primārās kanālu valodas (pirmās, kas kanālā minētas pie pieejamajām audio valodām), kā arī tiek atstātas tikai trīs skatītākās valodas – latviešu, krievu un angļu, pārējo valodu īpatsvars sastāda <0.5% no visiem skatījumiem. Datu piemērs 5.6. tabulā.

5.6. tabula

Datu piemērs lietotāju valodu preferenču klasteru izveidei

idd_wh_service	en	lv	ru
DTVS0000019	0.009	0.98	0.011
DTVS0000032	0.08	0.01	0.91
DTVS0000052	0.0	0.914	0.086
DTVS0000055	0.128	0.738	0.135
DTVS0000059	0.091	0.013	0.897

Pēc elkoņa metodes noteikti 5. klasteri, kuru vidējās vērtības redzamas tabulā 5.7. Klasteriem arī noteiktas potenciālās valodu preferences, kuras ņemt vērā, piedāvājot konkrēta kanāla saturu, balstoties uz satura pieejamo valodu. Kā redzams, ir lietotāji, kuri izteikti dod priekšroku latviešu kanāliem, ir tādi, kas izteikti dod priekšroku kanāliem krievu valodā. Šie klasteri, respektīvi, 1. un 3. un 4., relatīvi maz skatās kanālus ar angļu valodu kā kanāla pamatvalodu, un 3. klastera lietotājiem principā interesē tikai saturs latviešu kanālos.

Lietotāju valodu preferenču klasteri pēc KNN metodes un tos raksturojošās vidējās vērtības

Klastera numurs	en	lv	ru	Noteiktās valodu preferences
0	0.13	0.67	0.2	Visas
1	0.07	0.06	0.87	RU
2	0.58	0.22	0.2	Visas
3	0.04	0.92	0.04	LV
4	0.1	0.4	0.51	LV+RU

5.3.3 Interaktīvo funkciju izmantošana

Kā jau minēts ievada daļā – SIA Tet iTV piedāvā iespēju saturu skatīties gan tiešraidē, gan raidījumus pauzēt, ierakstīt un skatīties arhīvā. Taču liela daļa lietotāju šīs funkcijas neizmanto un pieturās pie klasiskās lineārās TV funkcionalitātes. No ieteikumu viedokļa, daudz vienkāršāk un pārlicinošāk ieteikumus var aprēķināt saturam, kuram jau ir mijiedarbība ar lietotājiem, līdz ar to veidot ieteikumus saturam, kas pieejams arhīvā, var gan izmantojot CB, gan CF metodes, gan abu metožu apvienojumu, kā arī aprēķinot populārāko saturu – to var aprēķināt tikai saturam, kuram bijusi mijiedarbība ar lietotāju. Ieteikumu sistēmas mērķis ir lietotājam atvieglot satura meklēšanu un arī celt pakalpojuma vērtību. No uzņēmuma viedokļa ir būtiski uzsvērt un parādīt abonētā pakalpojuma funkcionalitāti, ieteikumus veidojot un nododot lietotājam gan tiešraides, gan arhīva saturam, tas, kādā secībā atrādīt tiešraides, arhīva vienumus jau ir uzņēmuma ziņā, ieteikumi maģistra darba ietvaros tiks rēķināti gan tiešraidei, gan arhīvam visiem lietotājiem.

5.3.4 Laika konteksts

Izvērtējot lietotāju skatīšanās paradumus, jāņem vērā arī laiks un diena, kurā lietotājs patērē saturu. Lai gan iTV piedāvā iespēju katram lietotājam izveidot savu kontu, šī iespēja tomēr tiek izmantota reti un analizēt TV skatījumus kontu līmenī vismaz pagaidām nav vērts. Vienu

abonēšanas pakalpojumu var izmantot vairāki mājsaimniecības locekļi, kā arī viena un tā paša lietotāja patērētais saturs dažādās nedēļas dienās un dažādos diennakts laikos var atšķirties. Daļēji tā varētu būt lietotāja paša izvēle, taču jāņem vērā arī tas, ko TV kanāli attiecīgajā brīdī piedāvā. Kā minēts 5.3.3 nodaļā, ne visi lietotāji aktīvi izmanto interaktīvās TV opcijas, līdz ar to lietotājs var pielāgoties tam, kas konkrētajā brīdī tiek raidīts tiešraidē. Protams, arī TV kanālu satura plānotāji no savas puses ir izvērtējuši, kādas lietotāju grupas konkrētajā brīdī TV saturu patērē un kādas ir to intereses gan ar aptauju palīdzību, gan analizējot TV trafika datus.

Veidojot ieteikumu sistēmu, pašus ieteikumus var veidot, ņemot vērā laika kontekstu, kā jau aprakstīts nodaļās 2.2, 2.3.2.2 un 2.3.1.2. Vispirms tiek apskatīts, kādas ir lietotāju laika un skatīšanās dienu preferences. Izmantojot lietotāju skatīšanās notikumu datus, tiek aprēķināts, kurā stundā un dienā konkrētais notikums ir sācies. Izmantojot līdzīgu pieeju, kā valodu klasteru noteikšanā, tiek izdalītas 3 skatītāju grupas – tabula 5.7. Nosakot klasterus, vispirms aprēķins veidots ņemot stundu gradāciju, tad pielāgots rezultātiem un pārveidots uz tabulā redzamajiem laika periodiem.

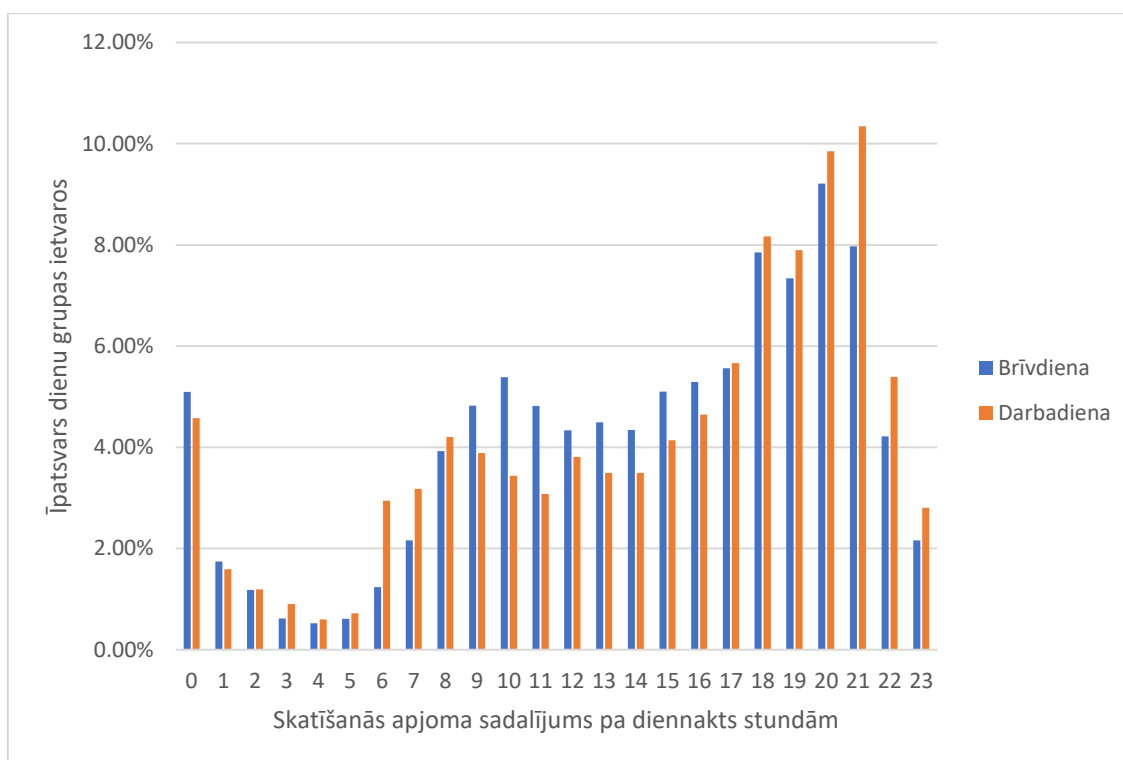
5.7. tabula

Lietotāju skatīšanās laika klasteri pēc KNN metodes un tos raksturojošās vidējās vērtības

Klastera nr.	Laika periods 1-6	Laika periods 6-9	Laika periods 9-12	Laika periods 12-18	Laika periods 18-23
cluster_0	2%	8%	10%	27%	47%
cluster_1	1%	3%	4%	12%	76%
cluster_2	8%	14%	15%	29%	26%

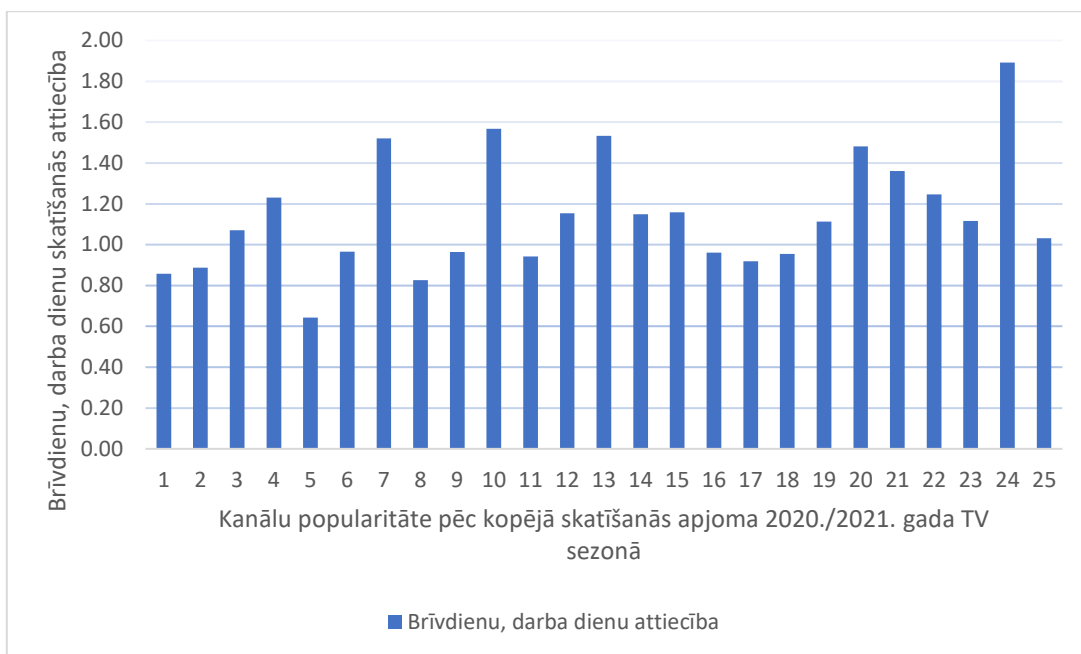
Šīs skatīšanās paradumu grupas var ņemt vērā, izvērtējot, kurš populārākais saturs lietotājam būs aktuāls, kā arī, kārtojot algoritmu ģenerētos ieteikumus, ņemot vērā laika kontekstu.

Tāpat būtiski ir ņemt vērā vai ir brīvdiena vai darba diena. Attēlā 5.2. redzams skatīšanās stundu īpatsvars darba dienās un brīvdienās. Redzams, ka darba dienās populārāks ir agrs rīts un vakara stundas, brīvdienās – dienas vidus un vakars.



5.2. att. Brīvdienu, darba dienu vidējais skatīšanās apjoma sadalījums pa stundām.

Savukārt, attēlā 5.3. redzams gada laikā skatītāko kanālu brīvdienu un darbadienu skatīšanās stundu attiecība. Skatīšanās laiks ir normalizēts, ņemot vērā, cik daudz TV satura kopumā tiek patērēts darba dienās un cik – brīvdienās. Vērtība 1 nozīmē, ka kanāls ir vienlīdz populārs darba dienās un brīvdienās, attiecīgi <1 nozīmē, ka kanāls populārāks brīvdienās, >1 - ka darba dienās. Tas nozīmē, ka ir kanāli, piemēram, 5. kanāls, kuru brīvdienās patērē par ~40% mazāk kā ikdienā, savukārt, 24. kanālu brīvdienās patērē gandrīz divreiz vairāk.



5.3. att. Brīvdienų, darba dienu normalizēta skatīšanās apjoma attiecība top 25 kanāliem.

5.4. Papildus datu ieguve

Izvērtējot EPG metadatu tabulas datus, ir skaidrs, ka ļoti kvalitatīvus CB ieteikumus no šiem datiem veidot nevar aprakstošo lauku skaita dēļ. Ja datu kvalitāti var daļēji uzlabot, metadatus tīrot un apstrādājot, tad papildus aprakstošos laukus, tādus kā aktieri, režisori, vērtējumi, atslēgas vārdi, ir jāmeklē citur vai arī šie dati jāpērk. Šajā nodaļā tiks apskatītas bezmaksas iespējas iegūt papildus datus par satura vienībām.

5.4.1 Datu iegūšana no Wikidata

Wikidata ir atvērto datu portāls, kurš apkopo informāciju par visdažādākajām tēmām, tai skaitā arī TV saturu, seriāliem, filmām, šoviem, aktieriem, utt. Wikidata ir brīvi pieejams datu avots, un informāciju var papildināt gan cilvēks, gan dators. Datus ērti iegūt, izmantojot kādu no SPARQL vaicājumu ģenerētājiem, lai piekļūtu resursu aprakstīšanas standarta (RDF) grafu datiem.

Viens no tādiem vaicājumu ģenerētājiem ir *Wikidata Query Service* [38], kurā var vai nu ģenerēt nepieciešamos vaicājumus interaktīvi vai arī manuāli rakstīt SPARQL kodu.

Maģistra darba ietvaros tiek ievākti metadati par *Wikidata* portālā pieejamajām filmām, pārraidēm un seriāliem, atsevišķi atlasot saturu, kurš pieejams angļu valodā, tad atlasot to, kas pieejams krievu valodā, bet nav pieejams angļu valodā un, visbeidzot, saturu, kas pieejams tikai latviešu valodā. Par cik vaicājumi ir apjomīgi, datu ielasīšana notiek iteratīvi, pa izlaišanas gadiem.

No *Wikidata* var iegūt tādu informāciju, kā nosaukumu vairākās valodās, satura žanru, aktierus, direktoru, producentus, izlaišanas gadu, valsti, kurā notiek darbība, oriģinālvalodu, kā arī unikālos identifikatorus un saites uz dažādām ar saturu saistītajām mājaslapām, piemēram, uz *IMDb*, *Facebook*, *RottenTomatoes* vai *Youtube* lapu konkrētajai satura vienībai. Šīs saites vai identifikatorus var tālāk izmantot, lai konkrētās vienības atlasītu no interesējošajām mājaslapām un ļautu veikt robotizētu datu ievākšanu no tām. Datu ieguves kods no *Wikidata* atrodams 1. pielikumā minētajā repozitorijā.

Kopumā no *Wikidata* iegūti 110 tūkstoši filmu ierakstu, datu piemērs pielikumā nr. 2. Papildu informāciju izdevies iegūt par 54% unikālajām filmu vienībām, kas raidītas 2020./2021. gada TV sezonā SIA Tet translētajos kanālos.

Seriāliem datu ieguves pieeja un kods ir līdzīgs, taču informācija par seriāliem ir sliktāk aizpildīta – tikai 10 tūkstoši ierakstu. Seriālu gadījumā aktieru, režisoru lauki faktiski ieteikumu veidošanai nav derīgi, jo seriālu formāts neļauj vienam aktierim spēlēt gana daudz seriālos, lai uz to varētu balstīt ieteikumus. Kopumā no apskatītajos EPG datos piefiksētajiem seriāliem tikai 8% izdodas pielasīt papildu metadatus. Ieteikumos vērts pievienot tikai pirmo izlaišanas gadu un ražotājvalsti.

5.4.2 *Datu vākšana no tīmekļa satura datubāzēm*

IMDb un *RottenTomatoes* ir divas no plašāk pazīstamajām video satura kritiķu un skatītāju atsauksmju un vērtējumu apkopojošajām mājaslapām. Papildus atsauksmēm, video un bilžu ieskatiem, dažādiem topiem, kā arī jaunumiem, kas saistīti ar konkrēto satura vienību, mājaslapās

pieejami dažādi satura vienību metadati, piemēram, pilns aktieru saraksts, filmas izmaksas un ieņēmumi, un cita informācija.

Izmantojot no *Wikidata* iegūtos vienību identifikatorus konkrētos portālos, datus no šīm mājaslapām iespējams efektīvi un robotizēti iegūt, jo nav jāveic pilna tīmekļa lapu pārļaušana vai jāsimulē meklēšana, bet tikai jāuzģenerē konkrēta saite un jāievāc dati. *Python* šo ērti izdarīt, izmantojot *bs4 (BeautifulSoup)* un *requests* bibliotēkas, taču maģistra darba ietvaros šie dati netiek ievākti konkrēto tīmekļa portālu lietošanas noteikumu dēļ.

5.4.3 Atslēgas vārdu ģenerēšana, izmantojot anotācijas

Satura vienību anotācijas satur tekstuālu darba raksturojumu, iekļaujot informāciju par satura vienības stāsta norises vietu, kontekstu, galvenajiem notikumiem un personāžiem. Cilvēkam, lasot šādu informāciju, bieži ir skaidrs vai satura vienība interesēs vai nē, taču tiešā veidā ieteikumu sistēmā anotācijas iekļaut nevar. Taču ir dažādi datizraces paņēmieni, kas ļauj no teksta izvilkt svarīgāko informāciju jeb tā saucamos atslēgas vārdus, ko jau var izmantot ieteikumu veidošanā. Ja anotācijas pieejamas vairākās valodās, vislabāk izvēlēties angļu valodas anotācijas, jo *python* ir pieejamas gatavas bibliotēkas, kas palīdz notīrīt tā sauktos “stopwords” jeb vārdus, kas tekstā varbūt bieži sastopami, taču nesatur nekādu informāciju par teksta nozīmi un jēgu. Tādi vārdi ir saikļi, iespraudumi, sasveicināšanās un pieklājības frāzes un citi.

No EPG datu analīzes izriet, ka unikālajām filmām anotāciju aizpildījums ir 45%, seriāliem tie ir 17%. Ir dažādas datizraces metodes, kas ļauj no teksta izvilkt atslēgas vārdus, taču maģistra darba ietvaros tas netiks veikts.

5.5. Algoritmi un metodes ieteikumu veidošanai

Ieteikumu sistēma maģistra darba ietvaros tiek būvēta, izmantojot gan CB, gan CF pieeju, kā arī izmantojot populārāko saturu.

CB plusi ir šīs metodes spēja ieteikt vienumus “uz priekšu”, taču ieteikumu kvalitāte ir ļoti atkarīga gan no tā, kas ieteikumu sistēmā tiek ievadīts kā interesējošs saturs un, protams, no datu kvalitātes. TV filmu ieteikumu ģenerēšanai CB ir visbūtiskākā metode, jo lielākā daļa filmu satura tiek skatīta tiešraidē, atkārtojumi ir pieejami tikai dažām filmām, un arhīvu lieto tikai daļa lietotāju. Tiesa, labi satura ieteikumi varētu lietotājus mudināt vairāk lietot arhīva un ieraksta funkcijas. CF metode filmu gadījumā spēj ieteikt tikai atkārtojumus, saturu, kas bijis pieejams kādā citā kanālā un arhīva saturu.

Seriālu gadījumā CB arī ir būtiska, taču seriāli pēc būtības ir labāk piemēroti CF ieteikumiem – jo var sameklēt līdzīgos lietotājus un ieteikt saturu, kas aktuāls viņiem. CF lielākā priekšrocība ir tāda, ka šī metode nav tik ļoti atkarīga no datu kvalitātes un var iztikt bez papildu metadatiem, kā arī spēj pamanīt sakarības datus, ko ar citām metodēm nesanāktu atrast.

Populārākais saturs, savukārt, noder veidojot ieteikumus lietotājiem, kas nesen abonējuši pakalpojumu vai arī tiem, kas skatījušies mazu satura apjomu. Lai gan vispārīgi, tomēr šie ieteikumi lietotājiem norāda uz citu lietotāju iecienītāko saturu.

5.5.1 Kosinusa līdzības algoritms – CB pieeja

Kosinusa līdzības algoritma mērķis ir noteikt divu vienumu savstarpējo līdzību. Aprēķinot kosinusa līdzību, tiek noteikts leņķa kosinuss starp diviem vienumus aprakstošiem vektoriem jeb skaitļu virknēm daudzdimensiju telpā - abu vektoru skalārais reizinājums tiek dalīts ar vektoru garumu reizinājumu – formula (5.1) [39].

$$\text{Kosinusa līdzība } (A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}. \quad (5.1)$$

Aprēķinātā kosinusa līdzība būs intervālā no [-1;1], kur “1” nozīmē, ka pazīmju vektori sakrīt, bet “-1” – ka pazīmju vektori ir pretēji, savukārt, “0”, ka tie ir ortogonāli, ko var interpretēt tā, ka šiem vektoriem nav nekā kopīga. Veidojot ieteikumu sistēmu, visas vienumu raksturojošās vērtības ir pozitīvas, un līdz ar to vektoru kosinusa līdzība ir intervālā [0;1]. Aprēķinot visu vienumu savstarpējo līdzību, iegūstam kosinusu līdzības matricu, kurā tabulas indeksi un kolonnas ir vienību identifikatori, bet šūnu vērtības – abu vienību līdzības koeficients. Šo līdzību matricu var izmantot, katram lietotājam iesakot vienības, kas vistuvāk atbilst jau skatītajām vienībām.

Ieteikumu veidošanai kosinusu līdzība starp vienumiem tiek rēķināta, izmantojot python *sklearn.metrics.pairwise.cosine_similarity* bibliotēku [40]. Kosinusa līdzības algoritmam nav parametru, ar kuriem algoritmu optimizēt un nav mācīšanās procesa, līdzība tiek sarēķināta viennozīmīgi un vienumu līdzības korektumu matemātiski izvērtēt nevar.

Šo metodi literatūrā visbiežāk izmanto, lai aprēķinātu dokumentu vai tekstu savstarpējo līdzību. Šajā gadījumā tiktu skaitīts vārdu skaits tekstā, un kā pazīmes tiktu izmantots katrs vārds ar tam piemēroto biežumu tekstā. Katru satura vienību raksturo noteiktas pazīmes un parametri, piemēram, konkrēts žanrs vai aktieris. Ievades dati pārsvarā ir reti aizpildītas tabulas formā (*scipy.sparse*), taču var izmantot arī *python* populārās *pandas* tabulas – katra rinda ir vienība, kurai rēķināsim līdzību ar citām vienībām, katra kolona ir vienību raksturojoša pazīme. Vērtības var būt dažādas, taču pārsvarā tiks lietota bināra reprezentācija – 0, ja pazīme raksturo vienumu, 1, ja neraksturo, piemēram, aktieris ir vai nav filmā. Ja pazīme saturiski ir nozīmīgāka kā citas (piemēram, nepieciešams uzsvērt žanru), vērtības var palielināt, tādējādi palielinot leņķi starp vektoriem, kuriem šīs pazīmes vērtība būs 0.

Izmantojot šo metodi, ieteikumu sistēma ģenerē CB ieteikumus, kas balstīti uz saturu raksturojošām pazīmēm. Satura eksperts nosaka robežu aprēķinātajai satura līdzībai, un lietotājam līdzīgās vienības var pielasīt, katrai noskatītajai satura vienībai pielasot līdzīgās, tad lietotāja līmenī šīs līdzības summējot. Tādējādi lietotājam primāri tiek ieteikts saturs, kas ir līdzīgs vairākām skatītajām satura vienībām. Otrs variants ir ieteikt satura vienības pēc lielākā koeficienta vērtības.

CB algoritma izveidē filmām tiek izmantoti sekojoši metadatu lauki – raidlaiks (iepriekš definētajos intervālos), darbadiena vai brīvdiena, kanāla nosaukums, cenzūra, žanrs, kanāla valodas un kanāla žanrs, ražotāja valsts, izdošanas gads, režisors un aktieri. Katrai pazīmei, ko pārvērs binārā formā, tiek aprēķināts atbilstošo ierakstu skaits un izmantoti tikai tie lauki, kas parādās noteiktu skaitu reižu, piemēram, tiek ņemti vērā tikai aktieri, kuri ir vismaz 50 filmās. Tādējādi tiek ģenerētas 212 vienumu aprakstošas pazīmes. Pārsvarā pazīmju vērtības ir bināras, taču valoda un žanrs tiek uzsvērts, vērtību palielinot līdz 2. No pieejamajām satura vienībām tiek izmestas tās, kuras raksturo mazāk kā sešas pazīmes (pretējā gadījumā šādas satura vienības ir ļoti līdzīgas un bojā algoritma rezultātu). Aprēķina paraugs līdzīgo filmu noteikšanai 1. pielikumā minētajā repozitorijā.

Seriālu ieteikumu aprēķinam ir līdzīga loģika kā filmām, tikai nav pieejams aktieru lauks.

5.5.2 Matricu faktorizācija, izmantojot ALS metodi – CF pieeja

CF modelī bāzēta pieeja aprakstīta 2.1.2 nodaļā. ALS (*alternating least squares*) ir matricu faktorizācijas pieeja, kas optimizāciju veic iteratīvi, vienu no matricām paturot konstantu, pielāgojot otru, minimizējot rekonstrukcijas kļūdas matricas vērtību pēc mazāko kvadrātu metodes. Minimālo kļūdu aprēķina pēc formulas (5.2) [41].

$$\min \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u n_{x_u} \|x_u\|^2 + \sum_i m_{y_i} \|y_i\|^2), \quad (5.2)$$

kur p_{ui} , ir lietotāja preference, definēta nodaļā 7.3,

c_{ui} ir pārliecības mērs, novērojot p_{ui} , fiksēts kā $2(1 + \alpha_{ui})$,

α pārliecības koeficients,

r_{ui} ir mijiedarbības matricas elements,

λ ir regulēšanas parametrs,

n_x, m_y ir vērtējumu jeb mijiedarbību skaits lietotājam u un vienumam i .

CF ieteikumu ģenerēšanai tiek izmantota *python* bibliotēka *implicit* [42], kur nodefinētas funkcijas gan modeļa trenēšanai, testa un treniņa kopas nodalīšanai, izvērtēšanai un arī ieteikumu aprēķinam. Modeļa trenēšanas process un izvērtēšana ņemta no Ethan Rosenthal bloga [43].

Modeli var pielāgot, izmantojot sekojošus parametrus:

- Faktoru skaits – latento faktoru skaits, ar kuru tiks aprakstīti lietotāji un satura vienumi.
- Regulācijas parametrs λ – parametrs, kas regulē modeļa trenēšanas procesu, palīdzot izvairīties no modeļa pārākās pielāgošanas datiem.
- Iterāciju skaits – ALS optimizācijas reižu skaits.

Modeļa labākās versijas atrašanai ir vairāki varianti, un testēšanas rezultāti var atšķirties atkarībā no tā, uz cik ieteiktajām vienībām rezultātu izvērtējam. Vienību skaita parametrs K – top ieteikumu skaits, kuros tiek meklēts *leave_one_out* vienums.

Modeļa pirmreizējā trenēšanā izmantojam datus par vienu mēnesi – 2021. gada martu, izmantojot visas lietotāju skatītās filmu vienības. Testa kopā tiek atstāta viena filma visiem tiem lietotājiem, kuri skatījušies vismaz 3 filmas, tādi no visas kopas ir 80%. Pārējās lietotāja – satura mijiedarbības izmantotas modeļa trenēšanā. Šāds laika posms izvēlēts pirmajai parametru pielāgošanai, jo parametru pārmeklēšana ir laikietilpīga. Mijiedarbības matricas aizpildījums ir 0.6% - t.i., katalogā ir ļoti daudz vienību, ar kurām lietotāji nav mijiedarbojušies.

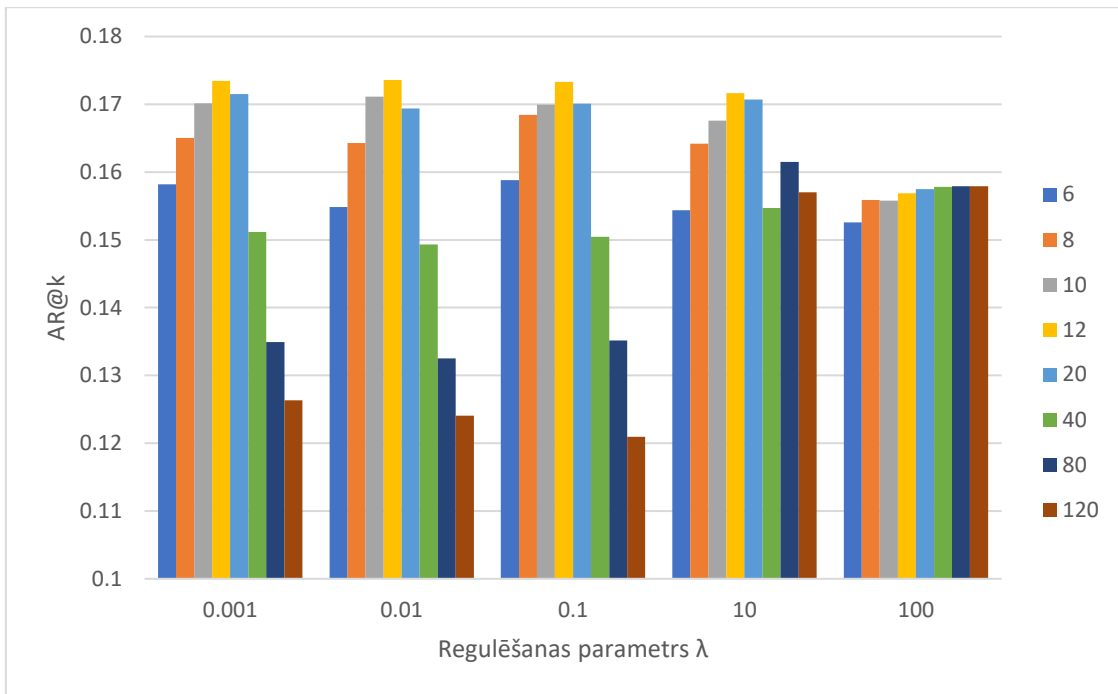
Vispirms tiks apskatīts modeļa pielāgošanas process filmu ieteikumu veidošanai. Veicam labāko parametru meklēšanu, piefiksējot rezultātu ik pa divām iterācijām, sākot no 2 un beidzot ar 38. Regulācijas parametra λ vērtības tiek mainītas pa kārtai, sākot no $1e-3$ līdz $1e2$. Faktoru skaits tiek izvēlēts sekojošs: 6, 8, 10, 12, 20, 40, 80, 120. Katra iterācija tiek pārbaudīta uz ieteiktajām top 3, 5, 10 un 20 vienībām. Modelis tiek izvērtēts, pārbaudot, vai top k ieteikumi satur testa kopā iekļauto satura vienību, formula (5.3) [44]. Pie $k=1$ formula aprēķina Faktiski *implicit* bibliotēkā šis mērs ir saukts par precizitāti, bet aprēķins atbilst atdevei [45]. Kopā parametru pārmeklēšanā notestētas 3040 dažādas parametru kombinācijas.

$$R@k = \frac{1}{N} \sum_i^N R_i(\text{rel}), \begin{cases} R_i(\text{rel}) = 1, \text{ ja vienība ir top } k \text{ ieteikumos} \\ R_i(\text{rel}) = 0, \text{ ja vienība nav ieteiktajos vienumos} \end{cases} \quad (5.3)$$

N – interesējošo vienību skaits lietotājam,

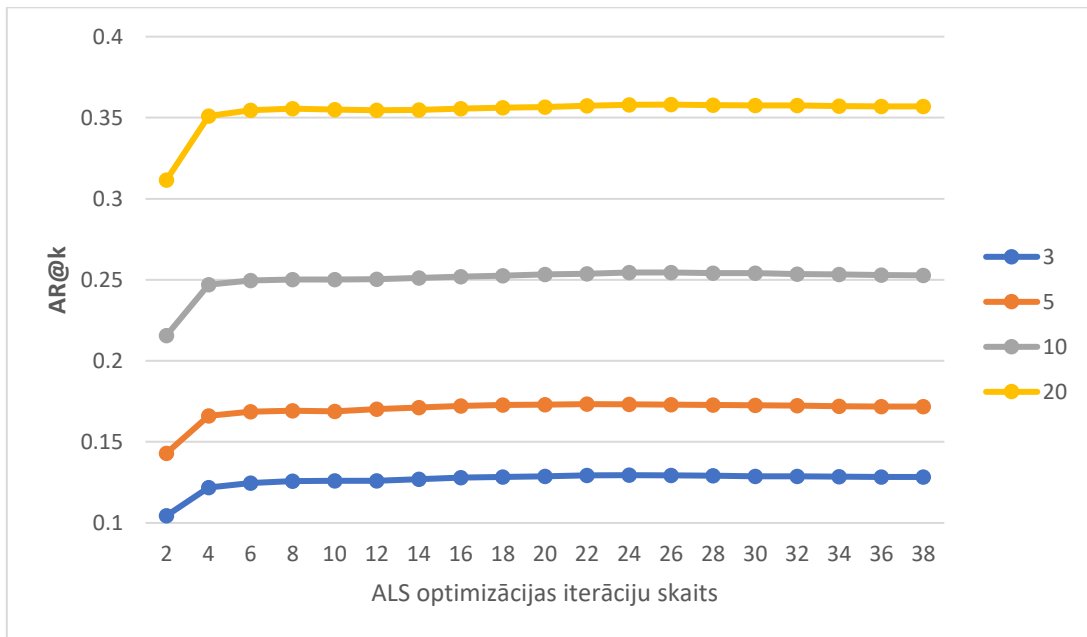
$R_i(\text{rel})$ - indikators, vai interesējošā vienība ir TOP k ieteikumos.

$AR@k$ ir vidējā $R@k$ vērtība visiem lietotājiem. Tālāk analizēti rezultāti pie $k=5$. Labākie $AR@5$ rezultāti tiek iegūti pie mazas λ vērtības un pie faktoru skaita 12, taču var novērot (attēls 5.4), ka $AR@5$ dod labus rezultātus arī pie ļoti lielām λ un liela latentu faktoru skaita. Taču 120 latentie faktori nozīmē, ka modelis ir sarežģīts, kā arī lielās λ dēļ modelim krītas sarēķināto ieteikumu pārlicība, tāpēc tiek izvēlēts latentu faktoru skaits 12. Faktiski ir redzēti 27%-37% ieteikumu pie attiecīgi top 50 un top 100 ieteikumiem katram lietotājam, kas arī ir likumsakarīgi, ņemot vērā, ka 30% no mēneša laikā pieejamajām filmām ir skatītas <100 reižu (specifiskas tematiskās pakas, nakts stundas, darba laiks, utt). CF modeļi balstās uz lietotāju un vienumu mijiedarbību un līdz ar to, vienumi ar mazu mijiedarbību skaitu tiks ieteikti retāk kā CB gadījumā [46].



5.4. att. AR@k (k=5) atkarībā no latentu faktoru skaita un λ .

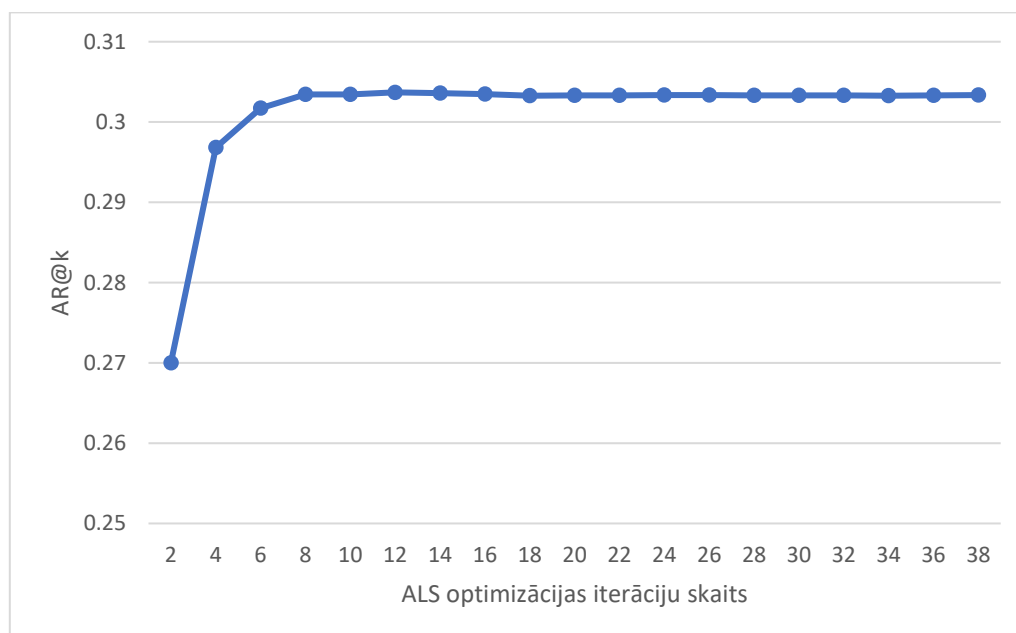
Modeļa precizitāte ($\lambda=0.1$ un latentu faktoru skaits = 12) pie dažādām k vērtībām un itērāciju skaita redzams attēlā 5.5. Var novērot, ka modeļa precizitāte būtiski izmainās tikai pie pirmajām optimizācijas procesa itērācijām. Labākais rezultāts sasniegts pie itērāciju skaita 24.



5.5. att. AR@k ($\lambda=0.1$ un faktoru skaits = 12) filmām atkarībā no trenēšanas itērāciju skaita pie dažādiem k.

Modeli trenējot uz datiem par lielāku laika periodu – pusgadu, AR@k pie tiem pašiem nosacījumiem krītas (pie k=10 tas sasniedz 11%, pie k=3, 5%), taču arī unikālo filmu skaits šajā periodā ir trīs reizes lielāks kā mēneša laikā. Šajā gadījumā mijiedarbības matricas aizpildījums – 1.4%, t.i., lai gan pusgada laikā ir vairāk satura vienību, procentuāli lietotājiem ir bijušas gandrīz trīs reizes vairāk mijiedarbību ar unikālajām satura vienībām kā mēneša laikā, to, galvenokārt, var izskaidrot ar to, ka dažādi kanāli saturu var rādīt atkārtoti, kā arī saturs var pārklāties vairākos kanālos.

Seriālu modeļa trenēšana notiek līdzīgi, sākumā izvēloties viena mēneša periodu, lai atrastu labākos parametrus. Šajā gadījumā maksimālā precizitāte pie k=5 ir 5%, ko varētu skaidrot ar seriālu skatīšanās un interešu aprēķina specifiku - mēnesis ir pārāk mazs periods, lai korekti noteiktu latentos faktorus, kas raksturo lietotāja preferences. Šī iemesla dēļ ir nolemts modeļa trenēšanai izmantot garāku laika periodu – 180 dienas, izmantojot tikai k=5. Mijiedarbības matricas aizpildījums 1.9%. 93% no visiem lietotājiem ir vismaz divi interesējoši seriāli, kā arī, tikai 20% seriālu interesentu skaits ir mazāks par 100. Matricas Šajā gadījumā precizitāte ir ievērojami labāka, modeļa trenēšanas grafiks attēlā 5.6 (k=5, $\lambda=0.1$ un latentu faktoru skaits = 20). Līdzīgi kā filmām, modeļa trenēšanai nav nepieciešams liels iterāciju skaits, labākais rezultāts 30.4% sasniegts pie 12. iterācijas. Unikālais ieteikto seriālu skaits sastāda 25-35% no visiem seriāliem, kas ir zemāks rādītājs kā filmām.



5.6. att. AR@k ($\lambda=0.1$, k=5 un faktoru skaits = 20) seriāliem atkarībā no trenēšanas iterāciju skaita.

CF pieejas ieteikumu aprēķina paraugs līdzīgo filmu noteikšanai un optimālo modeļa parametru piemeklēšanai atrodas 1. pielikumā minētajā repozitorijā.

5.5.3 *Populārākais saturs*

Popularitāti pirms vienuma publicēšanas ar pieejamajiem datiem tiešā veidā aprēķināt nevar. Var izdarīt pieņēmumus, ka kanāli paši konkrētos laikos mēdz raidīt līdzīgas popularitātes un pat žanru saturu. Šo informāciju var izmantot populārākā satura ieteikumu veidošanai.

Populārākā satura ieteikumus varētu savā ziņā varētu uztvert kā CF algoritmu – tiek veidoti ieteikumi, balstoties uz paša lietotāja un citu lietotāju mijiedarbību ar satura vienībām. Runājot par TV satura ieteikumiem, varam izdalīt divas populārākā satura grupas:

1. Vēsturiski populārākais saturs.
2. Prognozēts nākotnes populārākais saturs, balstoties uz vēsturiskajām kanālu skatījumu stundām.

Pirmais gadījums faktiski ir vēsturisko datu grupēšana. Otrā gadījumā tiek aprēķināts, kuri kanāli skatīti visvairāk iepriekšējā laika periodā, piemēram, mēnesī, laika intervālā no 18.00-21.00, un tiek pieņemts, ka pārraides, kas tiks raidītas šajā pašā laikā nākamajā nedēļā, būs tikpat populāras arī nākošajā nedēļā. Populārāko saturu var papildināt ar lietotāja preferencēm, piemēram, lietotāja iecienītākajiem kanāliem un arī lietotāja iecienītākajām skatīšanās stundām un žanriem.

Maģistra darba ietvaros, populārākais saturs aprēķināts abos veidos. Pirmajā gadījumā tiek aprēķinātas populārākās k vienības, un tās pieņemtas kā katra lietotāja ieteikumi. Otrā gadījumā aprēķins ir sarežģītāks:

- a) Katram lietotājam tiek aprēķinātas satura preferences, ņemot vērā žanru, laiku un to, vai lietotājs skatās saturu darba dienā vai brīvdienā.
- b) Katram lietotājam tiek ņemtas pirmās 10 preferences.
- c) Lai aprēķinātu populārākos skatīšanās laikus un noskaidrotu, kurā laikā populārākais saturs būs pieejams nākamajā nedēļā, tiek ņemti pēdējā mēneša dati, un aprēķināts populārākā satura raidlaiks pēc līdzīgiem kritērijiem kā lietotājam.

- d) Katram kanālam atlasītas 10 populārākās parametru kombinācijas.
- e) Parametru kombinācijām tiek pielasītas atbilstošas satura vienības testa laika periodā (par kuru tiek veidoti ieteikumi) jeb tiek atlasīts pieejamais saturs.
- f) Ieteikumu pielasīšana katram lietotājam.
- g) Ieteikumi tiek sarindoti pēc preferencēm, primāri ņemot lietotāja iepriekš skatītos kanālus

Piemērs d punkta parametru kombinācijām tabulā 5.8. Populārākā satura aprēķina kods atrodas 1. pielikumā minētajā repozitorijā.

5.8. tabula

TV3 kanāla top 10 skatītās parametru kombinācijas

hour	is_weekend	channel_name	genre_en
18-23	Nē	TV3	News
18-23	Nē	TV3	RealityShow
18-23	Nē	TV3	GameShow
1-6	Nē	TV3	Drama
12-18	Nē	TV3	Drama
9-12	Jā	TV3	Leisure

5.6. Analītiskais apraksts

Šajā nodaļā izvērtēsim risinājuma un algoritmu sarežģītību (*time complexity*), kā arī mērogošanas iespējamību.

Maģistra darba ietvaros datu apstrāde un ieteikumu aprēķins pamatā veikts lietotnē *Knime* un *Python jupyter notebook* vidē un datu apstrāde *Python* tiek realizēta ar bibliotēkas *Pandas* datu tabulām. Ieteikumu aprēķiniem izmantots dators ar 16GB operatīvo atmiņu un procesoru *AMD Ryzen 7 PRO 4750U with Radeon Graphics 1.70 GHz*. Aprēķiniem izmantots CPU.

Kosinusa līdzības algoritma laika un atmiņas sarežģītība ir $O(MN)$, kur M ir vektoru (vienību) skaits un N ir pazīmju skaits. Savukārt ALS matricu faktORIZācijas laika sarežģītība vienai iterācijai ir $O((M + N)K^3 + MNK^2)$ [47], kur K ir latentu faktoru skaits, N – lietotāju skaits un M – vienību skaits. Pilna parametru pārmeklēšana uz lokālā datora aizņēma 14h, taču paša modeļa atkārtota trenēšana ar zināmiem parametriem un maza iterāciju skaita aizņem vien pāris sekundes.

Abi ieteikumu sistēmā izmantotie algoritmi ir pieejami *Apache Spark* lielo datu apstrādes sistēmā, un aprēķini ir realizējami ar lietojumprogrammu saskarni *pyspark*, tādējādi ir iespējams veikt aprēķinu paralelizāciju un ieteikumu sistēmas rezultātus atjaunot pēc nepieciešamības bieži, piemēram, reizi stundā. Papildus tam procesu var optimizēt, latentos faktorus pārrēķināt un atjaunot ieteikumus tikai lietotājiem ar TV trafiku noteiktajā periodā.

5.7. Ieteikumu sistēmas kvalitātes novērtējums

Dažādo ieteikumu kvalitāti vislabāk izvērtēt praksē – redzot lietotāja mijiedarbību ar ieteiktajām vienībām. T.i., vai lietotājs redz ieteikto vienību, vai tas ir e-pasta ziņojums ar ieteiktajām filmām, kurš ir atvērts, vai tie ir slīdošie ieteikumu logi platformā, vai ieteikumi aplikācijā. Nākamajā solī tiek izvērtēts vai tie, kas ir redzējuši ieteikto satura vienību, to reāli skatās. Un, ja skatās – cik daudz skatās. Kvalitatīvas ieteikumu sistēmas izveide ir iteratīvs process, testējot dažādus ieteikumu variantus, pārbaudot hipotēzes par filtru darbību, ņemot vērā sezonalitāti un pat laika apstākļu ietekmi. Pie kam pašu ieteikumu kvalitāte ir tikai viens no faktoriem no lietotāja skatu punkta. Jāņem vērā arī tas, kur ieteikumi lietotājam tiek nogādāti, vai tie ir redzami un vai lietotājam ir ērti ar tiem mijiedarboties. Vai iespējams ieteikto vienību uzreiz skatīties, vai tā pašam jāatceras un jāmeklē? Vai ieteikumi ir kvalitatīvi noformēti, vai ir pievienota bilde un apraksts?

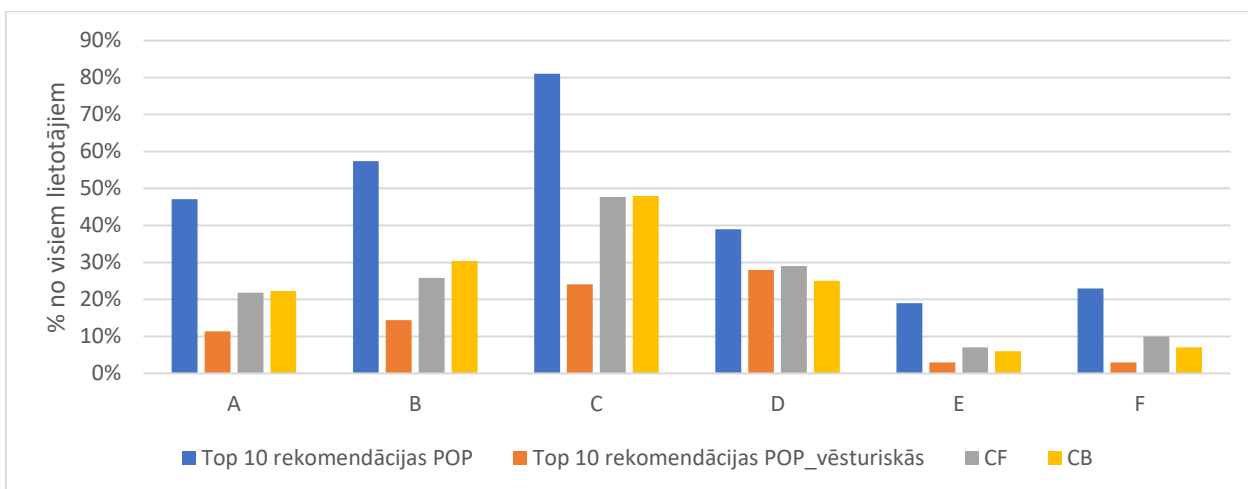
Maģistra darba ietvaros personalizētas izsūtnes lietotājam kvantitatīvai sistēmas izvērtēšanai netiek sūtītas, taču ieteikumu kvalitāti var izvērtēt arī citādi – veicot ieteikumu sistēmas testus vai aprēķinot ieteikumus uz konkrētu datumu, un izvērtējot lietotāja reālos skatījumus tuvākajā nedēļā (nesaistes pārbaude). Šajā gadījumā gan netiks pievienots laika konteksts ar konkrētajos laikos un dienās ieteiktajām filmām un seriāliem. Jāņem vērā, ka bez tā, ieteikumi ir vispārīgāki.

Ieteikumu sistēmu izvērtējot tiešsaistē, ir būtiska arī ieteikumu secība, kur precizitāte un atdeve tiktu vērtēta atkarībā no tā, kurā pozīcijā ieteikums atradās, taču nesaistē šāda metrika ir lieka, jo netiek novērota lietotāja reakcija uz ieteikumu.

5.7.1 Ieteikumu izvērtēšana filmām nesaistē

Filmu un seriālu ieteikumi tiks izvērtēti atsevišķi, filmām tiks apskatīts skatīto vienību skaits pret 10 ieteiktajām filmām no katras no metodēm – kosinusa līdzības, matricu faktorizācijas, kā arī populārākā satura pagājušajā nedēļā un provizoriska populārākā satura nākamajā nedēļā, ņemot vērā lietotāja iecienītākos kanālus, žanrus un skatīšanās laikus. Ieteikumi sarēķināti uz 2021. gada 1. aprīli, faktiskie skatījumi pielasīti tekošajās 9 dienās, t.i., 2.-10. aprīlī. Apskatīti tikai lietotāji, kuriem šajā periodā ir vismaz viena noskatīta filma. Pieeju rezultāti atspoguļoti 5.7. attēlā. Metodes salīdzinātas sekojošos griezumos:

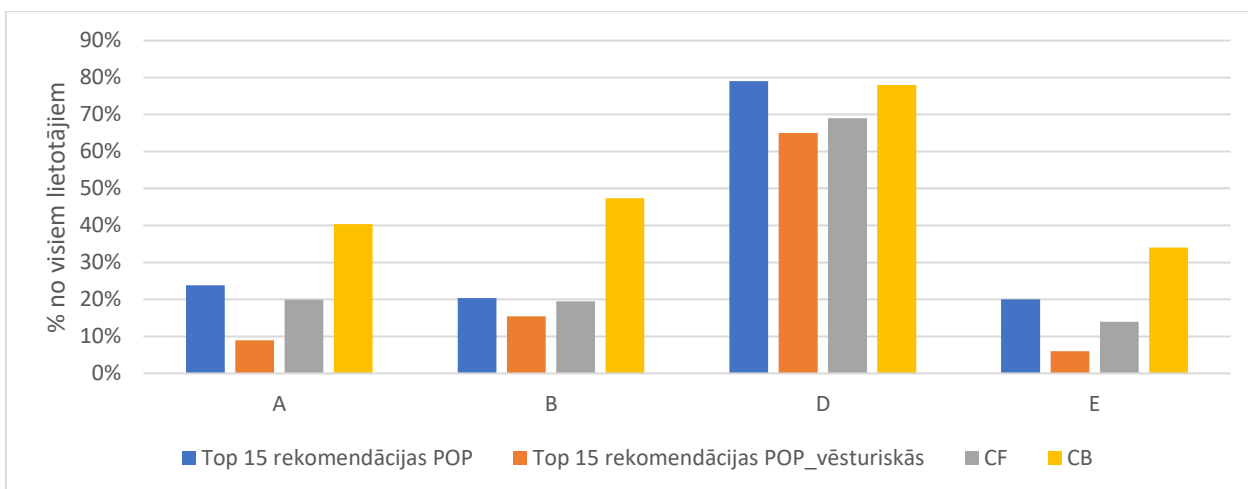
- A. Procenti no visiem lietotājiem, kuri noskatījušies vismaz vienu rekomendēto filmu.
- B. Procenti no lietotājiem, kas izmanto arhīvu un noskatījušies vismaz vienu rekomendēto vienību.
- C. Procenti no lietotājiem, kas noskatījušies vismaz 10 filmas un noskatījušies vismaz vienu rekomendēto.
- D. Vidējā ieteikumu atdeve lietotājiem, kas noskatījušies vismaz vienu rekomendēto vienību.
- E. Vidējā ieteikumu atdeve visiem lietotājiem.
- F. Vidējā ieteikumu atdeve lietotājiem, kas noskatījušies vismaz 10 filmas.



5.7. att. Filmu ieteikumu rezultāti, testa periods 9 dienas, k=10

A, B, C griezumā definē HR - formula (2.2), D, E, F – vidējo atdevi $AR@k$ - formula (5.3). Faktiski redzams, ka vislabākā parametru vērtība visos gadījumos ir populārākā satura ieteikumiem, taču arī CB un CF pieejas strādā labi. Visu lietotāju vidū atdeves (E griezums) ir 6% CB algoritmam, 7% CF algoritmam. Paturot prātā 2.5. nodaļā minētos ieteikumu sistēmas parametrus, ieteikumu sistēma, kas precīzi paredz visas lietotāja mijiedarbības, nedos vēlamus rezultātus. Jāņem arī vērā, ka tie ir 10 ieteikumi uz nedēļu un ir liela iespēja, ka konkrētajā laikā TV nemaz netiek skatīts. Var novērot, ka lietotājiem, kuri izmanto arhīvu (B griezums), ieteikumi ir mazliet precīzāki, kā arī – lietotājiem, kuri skatās lielu satura apjomu, ieteikumi būs atbilstošāki (C griezums).

Līdzīgu pieeju nesaistes EPG ieteikumu izvērtēšanai izmantojuši Zibrincký et al 2013. gadā [26]. Testa periods ir ņemts vienas dienas ietvaros, kā arī ņemtas 15 ieteikumu vienības vidējās atdeves aprēķinam. Līdzīgi pielāgojot testa periodu un rekomendējamo vienību skaitu, filmu rezultāti redzami attēlā 5.5. Lietotāju skaits, kas noskatījies 10 vai vairāk filmas vienas dienas laikā ir pārāk mazs, lai šos kritērijus apskatītu. Papildu esošajiem filtriem CB un CF tiek pievienoti lietotāja top 10 kanāli, un ieteikts tikai saturs, kas raidīts tajos.

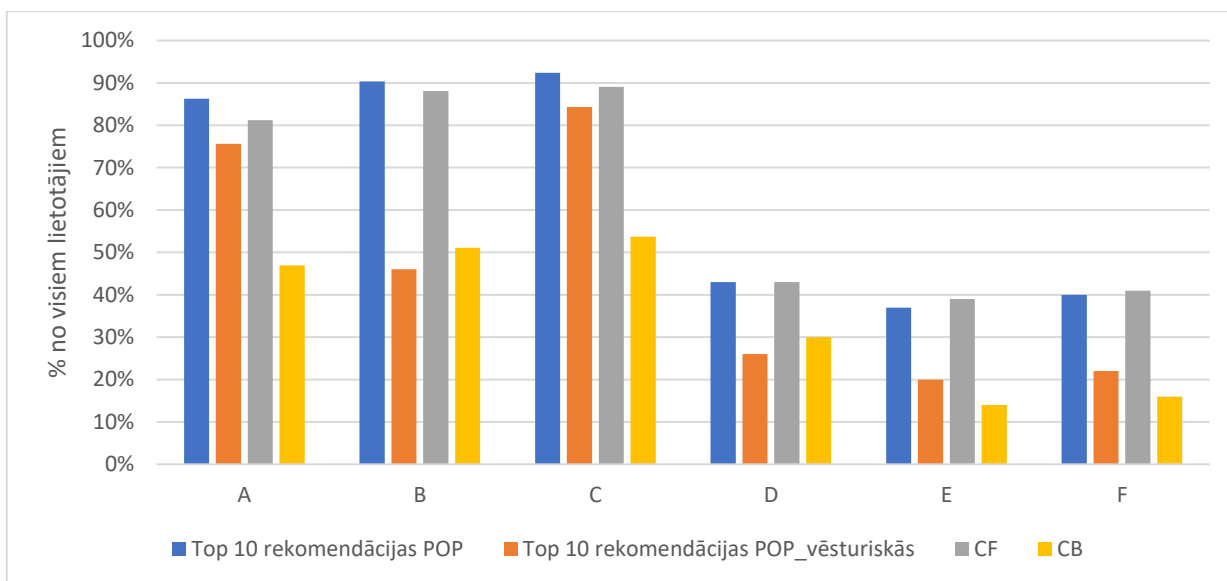


5.8. att. Filmu ieteikumu rezultāti, testa periods 1 diena, k=15

Zibrincký et al pētījumā AR@k filmām attiecīgi ir 31% CB risinājumam, 14% CF pieejai un populārākajam saturam labākais rezultāts – 12%. Salīdzinājumam, maģistra darba ieteikumu sistēmas rezultāti 5.8. attēlā, respektīvi 34% CB un 14% CF algoritmam un 20% populārākajam saturam. Lai gan izmantoti citi TV skatījumu dati, kopumā var secināt, ka ieteikumu kvalitāte ir līdzīga un līdz ar to, ka maģistra darba ietvaros izveidotā ieteikumu sistēma strādā labi.

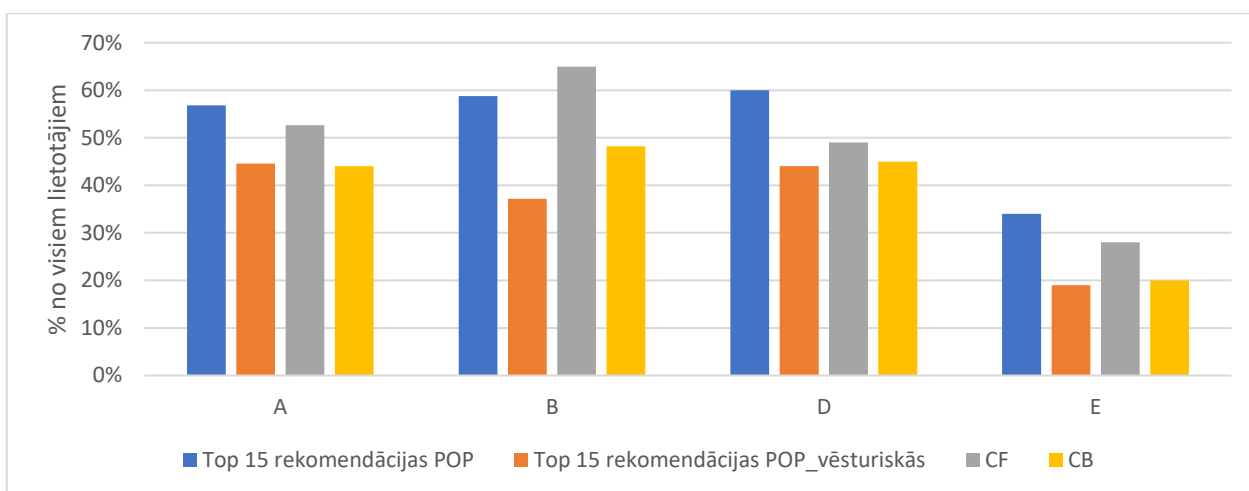
5.7.2 Ieteikumu izvērtēšana seriāliem nesaistē

Seriālu izvērtēšanu primāri veicam līdzīgi kā filmu gadījumā – apskatām 10 ieteiktās vienības un to skatījumus. Rezultāti attēlā 5.9. Kā var redzēt, visi ieteikumi strādā daudz labāk kā filmu gadījumā, bet tas arī ir sagaidāms – seriāli tika skatīti iepriekšējā periodā un tos turpina skatīties arī testa periodā. Šajā gadījumā daļa no ieteikumiem nevis iesaka jaunu saturu, bet iesaka “turpini skatīties”. Šādi ieteikumi var būt noderīgi un noteikti var atvieglot lietotāja skatīšanās pieredzi – piemēram, EPG varētu rādīt sadaļu “turpini skatīties”, kur ātri varētu piekļūt saturam, kas lietotājam garantēti interesē. Var novērot arī to, ka pie šādiem nosacījumiem CF pieeja strādā labāk kā CB un pat labāk kā populārākais saturs lietotāja līmenī (pēc kritērija E – vidējās atdeves visiem lietotājiem).



5.9. att. Seriālu ieteikumu rezultāti, testa periods 9 dienas, k=10

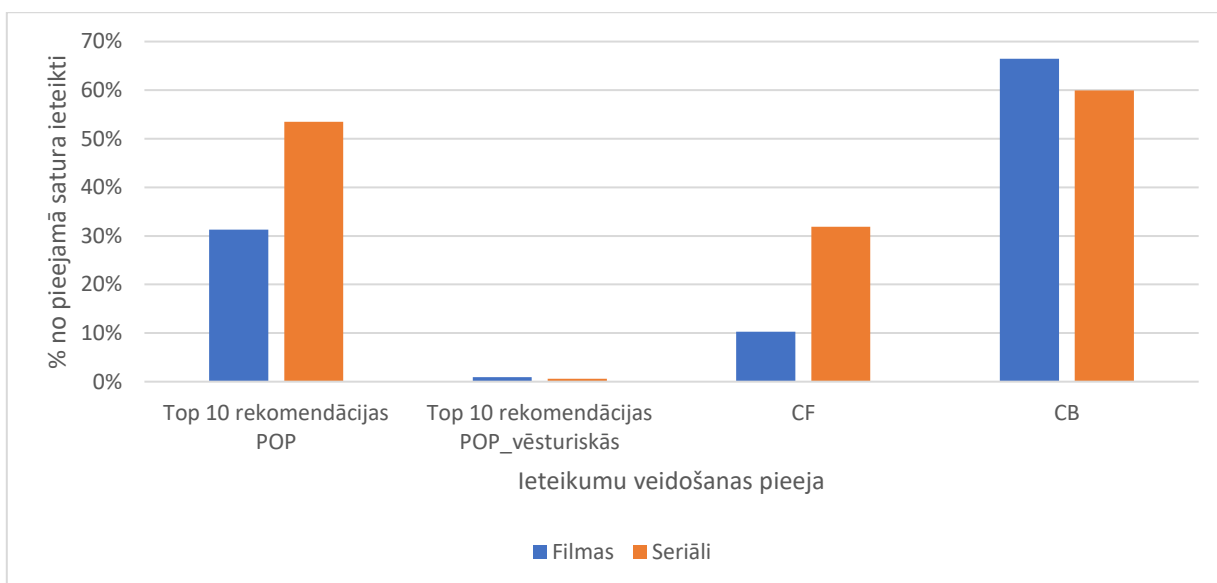
Lai risinājumu varētu salīdzināt ar Zibrincký et al pētījumu, nepieciešams sarēķināt 15 ieteikumus katram lietotājam uz vienu dienu – t.i., otro aprīli. Rezultāti attēlā 5.10. Vislabākais Zibrincký et al pētījuma rezultāts CB risinājumam – 54%, CF – 50% un populārākajam saturam – 35%. Maģistra darba risinājuma precizitāte - CB risinājumam – 20%, CF – 28% un populārākajam saturam – 34%. Populārākajam saturam precizitāte līdzīga, taču gan CF, gan CB precizitātes maģistra darbā ir daudz zemākas. Atšķirībā no Zibrincký et al pētījuma, maģistra darbā nav apskatīti ziņu raidījumi un laika prognožu raidījumi, kas varētu precizitāti uzlabot. CB gadījumā atšķirību papildus var skaidrot ar mazo metadatu apjomu, ko izdevies iegūt no *Wikidata*.



5.10. att. Seriālu ieteikumu rezultāti, testa periods 1 diena, k=15

5.7.3 Ieteiktais saturs apjoms no kataloga

Ieteikumu sistēmas viens no raksturojošajiem parametriem ir arī ieteiktā saturs daudzveidība. T.i., cik liela daļa no pieejamā saturs tiek ieteikta un cik personalizēti ir ieteikumi. Attēlā 5.11 redzams katras pieejas ģenerēto unikālo ieteikto vienību skaits no visa vienību skaita gan seriāliem, gan filmām. Redzams, ka CB pieeja iesaka vairāk kā pusi kataloga pie 10 ieteiktajām vienībām, arī top 10 personalizētās saturs vienības nosedz lielu daļu kataloga. CF pieeja veido ieteikumus par mazāku saturs daļu, sevišķi mazs skaits ir filmu ieteikumiem, kas arī ir likumsakarīgi, jo CF nevar ieteikt saturs, ko lietotāji nav jau skatījušies.



5.11. att. Unikālo ieteikto vienību skaits no konkrētajā nedēļā pieejamā saturs atkarībā no ieteikumu ģenerēšanas pieejas

5.7.4 Ieteikumu sistēmas kvalitatīvie testi

Ieteikumu sistēmas kvalitāte notestēta arī uz 8 SIA Tet darbiniekiem – datu un biznesa analītiķiem, kuri abonē un lieto interaktīvās TV pakalpojumu. Katram testa dalībniekam tika sagatavots skatītā saturs saraksts par laiku periodu 2021. gada 15. novembris līdz 2022. gada 16. maijs, noteikts pēc 5.3.1.1 un 5.3.1.2 nodaļās aprakstītās loģikas. Šie skatījumi tika izmantoti kā pamats personalizētu ieteikumu aprēķinam, izmantojot 5.5 nodaļā aprakstītās pieejas. Ieteikumi

apkopotī individuālos *excel* failos un nosūtīti katram lietotājam. Katram lietotājam sagatavoti ieteikumi atsevišķi filmām un seriāliem, izmantojot sekojošas, lietotājam nezināmas, pieejas:

- A – ar CB algoritmu rēķināti ieteikumi.
- B – populārākās satura vienības iepriekšējā nedēļā (kuras pieejamas arī tekošajā nedēļā).
- C - personalizēts populārākais saturs.
- D – ar CF algoritmu rēķināti ieteikumi.

Katrā no pieejām atlasīti 10 vienumi, D ieteikumi seriāliem rēķināti divos veidos: vienā - iekļaujot skatīto saturu, otrā – neiekļaujot. Pārējās pieejās skatītais saturs pēdējā pusgada laikā nav iekļauts. Testu piemēri 3. un 4. pielikumā. Testa dalībniekiem lūgts katru vienību izvērtēt kā atbilstošu, neitrālu vai neatbilstošu, kā arī brīvā veidā sniegt komentārus.

Apkopojot visu testa dalībnieku atbildes, secināts, ka labāka atbilstība kopumā ir D un C risinājumiem seriālu gadījumā (vidējais atbilstošu ieteikumu skaits 7 un 6.17) un A, C, D ieteikumiem filmu gadījumā (vidējais atbilstošu ieteikumu skaits attiecīgi, 4.5; 5.83 un 5.33). Viennozīmīgi sliktākie ieteikumi ir B pieejai abos gadījumos (3.17 un 1.5). No kvalitatīvajiem testiem secināms, ka ieteikumi ir lietotājiem atbilstoši.

Kvalitatīvo testu rezultātā piefiksēti daži ieteikumu sistēmas trūkumi un iezīmēti potenciāli ieteikumu sistēmas uzlabojumi. Pirmais – lietotājiem, kuri skatās animācijas filmas ar vecuma ierobežojumu 12+, tiek piedāvātas arī filmas mazākiem bērniem, kas lietotājam varētu nebūt aktuālas. Šo varētu risināt, uzsverot vecuma ierobežojumu CB ieteikumos. Otrais – lietotāji, kuru bērni skatās animācijas filmas, ieteikumos dominē animācijas filmas. Šo varētu risināt dažādi – ieteikumus rēķinot lietotāja profila līmenī, ieteikumiem pievienojot laika kontekstu, animācijas ieteikumus rēķinot atsevišķi un limitējot animāciju vienību skaitu. Trešais – daļa filmu un seriālu, kas tiek ieteiktas, jau ir redzētas, taču agrāk. Būtu nepieciešams veikt analīzi, lai noskaidrotu, kurš saturs ir tāds, ko lietotāji labprāt skatās atkārtoti un pēc cik ilga laika, kā arī – kuru saturu nepieciešams atkārtoti neieteikt ilgāku laiku. Ceturtais – B pieejā ieteikts saturs krievu valodā, ko daļa lietotāju atzīst par nepiemērotu. Citās pieejās gan šis nav novērots.

EKSPERIMENTU REZULTĀTI, NOVĒRTĒJUMS

Izvērtējot eksperimentu rezultātus nesaistē, secināts, ka maģistra darba ietvaros veidotā ieteikumu sistēma strādā labi un lietotājiem bieži tiek ieteikts tiem aktuāls saturs. Pārbaudot ieteikumu sistēmas darbību ar līdzīgiem parametriem kādi izmantoti Zibrincký et al 2013. gada [26] pētījumā, var secināt, ka sistēmas darbojas līdzīgi. Testos vislabākie rezultāti ir populārākā satura ieteikumiem, kurš nesaistē ir vienīgais risinājums, kurā izmantoti lietotāja kanālu filtri ieteikumu kārtošā un filtrēšanā. Novērots, ka CF (*collaborative filtering*) ieteikumi labāk strādā seriālu satura ieteikumu veidošanai.

Izvērtējot ieteikumu sistēmas personalizācijas līmeni – t.i., cik dažādu saturu RS lietotājiem iesaka, novērots, ka vislabākie rezultāti ir CB (*content based filtering*) pieejai (pēc šīs metodes aprēķinātie ieteikumi aptver vislielāko pieejamā satura daļu), kas atbilst literatūrā aprakstītajiem metodes plusiem.

Izvērtējot RS kvalitatīvos testus, novērots, ka rezultāti ir labāki kā nesaistes testos, kas, galvenokārt, ir skaidrojams ar ieteikumu atrādīšanu gala lietotājam un kopumā liecina, ka ieteikumu atrādīšana lietotājam varētu sniegt pozitīvu efektu, ievērojami uzlabojot nesaistes testu rezultātus. Zibrincký et al tā paša pētījuma [26] ietvaros veikuši arī tiešsaistes ieteikumu testēšanu EPG sistēmā, kur novērots, ka lietotājiem, kuri izmanto RS piedāvātos ieteikumus, TV trafika apjoms pret tiem pakalpojuma lietotājiem, kuri to neizmanto, palielinās. Raksta autori arī novērojuši, ka lietotāju RS lietošanas regularitāte laika gaitā pieaug, kas nozīmē, ka lietotāji novērtē RS. Ņemot vērā maģistra darbā izstrādātās RS kvalitatīvo testu rezultātu, var pieņemt, ka šīs RS ieviešana EPG sniegtu līdzīgu rezultātu.

SECINĀJUMI

Video satura ieteikumu sistēmu pamatā ir divas paradigmas, CF (*collaborative filtering*) un CB (*content based filtering*). Abas metodes var lietot kopā, veidojot hibrīda metodes, vai atsevišķi. Papildus tam ieteikumus var veidot arī ar triviāliem populārāko kanālu, personalizētu kanālu un personalizētu laiku ieteikumiem.

CF metodes pamatā ir lietotāja un video satura mijiedarbība. Ja lietotājs reti skatās TV vai ir nesen uzsācis pakalpojuma lietošanu, šie dati ir pieejami mazā apjomā un CF cieš no t.s. CS problēmas – rekomendāciju kvalitāte šādiem klientiem būs nepietiekama. Taču kā novērots, veidojot ieteikumu sistēmu, saturs kanālos mēdz dublēties un atkārtoties, kas ļauj CF izmantot arī iesakot saturu, kas vēl nav raidīts. CF metode labi strādā seriālu ieteikumu veidošanai.

CB metodes pamatā ir kvalitatīvi un kvantitatīvi metadati. Tā necieš no CS problēmas un var tikt izmantota, lai veidotu ieteikumus saturam, ar kuru ir mijiedarbojušies maz vai nemaz lietotāju. T.i., CB ļauj ieteikt saturu, kas TV vēl nav rādīts vai tādu saturu, kas nav bijis populārs (rādīts maz skatītā kanālā vai nepopulārā laikā).

Televīzijas ieteikumu sistēmu lielākie izaicinājumi – nekvalitatīvi un limitēti dati, kas iegūstami no EPG metadatiem – to piegādātājs ir TV kanālu izplatītājs un šajā laukā nav vienotu kvalitātes standartu. Datu tīrīšana, analīze un papildu datu ieguve ir liela daļa no ieteikumu sistēmas veidošanas procesa. Veidojot ieteikumu sistēmu, nonākts pie secinājuma, ka dažādu kanālu izplatītāju piegādātie metadati stipri atšķiras un daļu raidījumu ar nekvalitatīviem datiem ieteikumu sistēma ieteikt nevarēs – CB algoritmā šiem vienumiem būs pārāk mazs pazīmju skaits un tie tiks nofiltrēti. Produkcijā ieviesta RPG RS varētu veicināt kanālu pārstāvju interesi kvalitatīvi aizpildīt metadatus. Otrs ieteikumu sistēmas izaicinājums – īss satura pieejamības laiks. Informācija par nākotnes programmu pieejama nepilnu mēnesi uz priekšu, savukārt saturs vēsturē glabājas tikai pāris dienas. TV satura ieteikumi jāveido, piedāvājot gan saturu, kas būs skatāms tuvākajā laikā, gan tādu, kas pieejams skatīties tūlīt – gan tiešraidē, gan arhīvā. Trešais izaicinājums – ir pieejami tikai netieši skatīto vienību vērtējumi, respektīvi, skatījumu dati. Taču nav skaidras metodes, kā noteikt, vai lietotājam saturs interesē vai, piemēram, TV palicis ieslēgts un saturs nav interesants – interese ir jādefinē, analizējot lietotāju skatīšanās paradumus.

Būtisks TV ieteikumu aspekts ir lietotāja skatīšanās paradumi, konkrētāk – skatīšanās laiks un diena, kā arī kanāli un valoda. Ieteikumus var veidot, lietojot kādu no šiem parametriem kā papildus kontekstu, kas algoritmos izpaustos kā papildu dimensija vai dimensijas, tādējādi uzlabojot ieteikumu precizitāti. Kā novērots testējot ieteikumu sistēmu, izmantojot konteksta filtrus, var samērā precīzi noteikt, ko lietotājs skatīsies konkrētā dienā un laikā.

Atkarībā no RS pielietojuma mērķa, var tikt veidoti dažādi ieteikumi - var tikt ieteikts jauns un lietotājam interesējošs saturs lietotāja iecienītajos kanālos vai arī kanālos, ko lietotājs ikdienā neskatās. Var tikt veidoti “turpini skatīties” tipa ieteikumi, padarot piekļuvi interesējošiem raidījumiem vienkāršāku.

Apvienojot dažādas ieteikumu sistēmās lietotās metodes, iespējams veidot tiešraides un vēsturisko raidījumu ieteikumus. EPG RS palīdzētu esošiem arhīva un ieraksta funkciju lietotājiem ērtāk sameklēt saturu, savukārt, klientiem, kas šīs funkcijas nelieto, atvieglotu šo funkciju lietošanu, piedāvājot aktuālu saturu bez papildus meklēšanas. Tādā veidā EPG ieteikumu sistēma var celt pakalpojuma vērtību, palielinot klienta apmierinātību ar pakalpojumu.

Izvērtējot ieteikumu sistēmas rezultātus pēc 5.7 punktos aprakstītās loģikas, noskaidroti potenciāli ieteikumu sistēmas uzlabojumi un secināts, ka maģistra darba ietvaros izveidotā ieteikumu sistēma veido kvalitatīvus satura ieteikumus, kurus TV satura pakalpojuma sniedzējs var izmantot, lai lietotājam ieteiktu aktuālu saturu.

Lai pilnvērtīgi novērtētu RS darbību, kā arī pielāgotu filtrus, ieteikumu skaitu, noteiktu labāko secību, nepieciešams ieteikumu sistēmu testēt vispirms ar aptaujām un A/B testiem e-pastā un vēlāk - tiešsaistē EPG.

ATSAUCES

- [1] Wikipedia, the free encyclopedia, Wikipedia, Netflix_Prize, [Tiešsaiste]. Available: https://en.wikipedia.org/wiki/Netflix_Prize. [Piekļūts 22 05 2022].
- [2] C. Harrison, B. Amento and L. Stead, iEPG: an ego-centric electronic program guide and recommendation interface, *UXTV '08: Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*, Silicon Valley California USA, 2008.
- [3] N. Chang, M. Irvan and T. Terano, A TV Program Recommender Framework, *Procedia Computer Science*, no. 22, pp. 561-570, 2013.
- [4] M. W. Khan, G.-Y. Chain, F.-F. Chua, S.-C. Haw, M. Hassan and F. A. Saaid, Context-aware Ontological Hybrid Recommender System For IPTV, *6th International Conference on Information and Communication Technology (ICoICT)*, 2018.
- [5] J. Quan and S. Cho, A Hybrid Recommender System Based on AHP That Awares Contexts with Bayesian Networks for Smart TV, *Hybrid Artificial Intelligence Systems*, Cham, Springer, 2014, p. 527–536.
- [6] M. Ehrmantraut, T. Härder, H. Wittig and R. Steinmetz, The personal electronic program guide—towards the pre-selection of individual TV programs, *CIKM '96: Proceedings of the fifth international conference on Information and knowledge management*, Rockville Maryland USA, 1996.
- [7] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond and J. & Basilico, Deep Learning for Recommender Systems: A Netflix Case Study, *AI Magazine*, vol. 42, no. 3, pp. 7-18., 2021.
- [8] Centrālā statistikas pārvalde, Monetārā nabadzība, nevienlīdzība un sociālā atstumtība, stat.gov.lv, [Tiešsaiste]. Available: <https://stat.gov.lv/lv/statistikas-temas/iedzivotaji/monetara-nabadziba-nevienlidziba-un-sociala-atstumtiba/tabulas/nnn090>. [Piekļūts 22 05 2022].

- [9] MIDIA RESEARCH, TV Content Consumption by device in Q4 2020, miptrends, 22 01 2021. [Tiešsaiste]. Available: <https://www.miptrends.com/tv-business/tv-content-consumption-by-device-2020/>. [Piekļūts 22 05 2022].
- [10] L. Aloia, OTT: How the Rise of VOD Led to an Increase in Linear Streaming and How to Adapt, MEDIUM, 06 05 2021. [Tiešsaiste]. Available: <https://medium.com/float-left-insights/ott-how-the-rise-of-vod-led-to-an-increase-in-linear-streaming-and-how-to-adapt-2db028df4533>. [Piekļūts 22 05 2022].
- [11] Tele2, Pētījums: 63% iedzīvotāju televīzijas un video saturu skatās pārdomātāk, tas vairāk nekalpo kā fona troksnis, tele2, 29 10 2020. [Tiešsaiste]. Available: <https://www.tele2.lv/tele2/tele2-jaunumi/jaunums/petijums-63-iedzivotaju-televizijas-un-v/>. [Piekļūts 22 05 2022].
- [12] Nacionālā elektronisko plašsaziņas līdzekļu padome, neplpadome.lv, 2021. [Tiešsaiste]. Available: https://www.neplpadome.lv/lv/assets/documents/Petijumi/P%C4%93t%C4%ABjums_par_Latvijas_iedz%C4%ABvot%C4%81ju_medijprat%C4%ABbu_un_mediju_satura_lieto%C5%A1anas_paradumiem_2021.pdf. [Piekļūts 22 05 2022].
- [13] Kantar Media, Linear vs non-linear viewing:A qualitative investigation exploring viewers' behaviour and attitudes towards using different TV platforms and services providers, Ofcom, 2016. [Tiešsaiste]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0029/68816/km_report.pdf. [Piekļūts 22 05 2022].
- [14] R. Bambini, P. Cremonesi and R. Turrin, contentwise.tv, 03 2010. [Online]. Available: <https://silo.tips/download/a-recommender-system-for-an-iptv-service-provider-a-real-large-scale-production>.
- [15] I. Portugal, P. Alencar and D. Cowan, The use of machine learning algorithms in recommender systems: A systematic review, *Expert Systems With Applications*, vol. Volume 97, pp. Pages 205-227, 2018.

- [16] B. Rocca, Introduction to recommender systems, Towards data science, 03 06 2019. [Tiešsaiste]. Available: <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>. [Piekļūts 22 05 2022].
- [17] Wikipedia, the free encyclopedia, Wikipedia Recommender_system, [Tiešsaiste]. Available: https://en.wikipedia.org/wiki/Recommender_system#Hybrid_recommender_systems. [Piekļūts 22 05 2022].
- [18] G. Adomavicius, B. Mobasher, F. Ricci and A. Tuzhilin, Context-Aware Recommender Systems, *AI Magazine*, vol. 32, pp. 67-80., 2011.
- [19] S. R. Chavare, C. J. Awati and S. K. Shirgave, Smart Recommender System using Deep Learning, *6th International Conference on Inventive Computation Technologies (ICICT)*, 2021.
- [20] R. Bambini, P. Cremonesi and R. Turrin, Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environmen, *Recommender Systems Handbook*, Springer, 2011, p. 299–331.
- [21] F. Ullah, G. Sarwar, S.-C. Lee, Y. Park, K. Moon and J. Kim, Hybrid Recommender System with Temporal Information, *The International Conference on Information Network*, 2012.
- [22] P. Turrin and C. Roberto, Time-evolution of IPTV recommender systems, *Proceedings of the 8th European Conference on Interactive TV and Video (EuroITV '10)*, New York, 2010.
- [23] Z. Yuan, J. H. Lee and S. Zhang, Optimization of the Hybrid Movie Recommendation System Based on Weighted Classification and User Collaborative Filtering Algorithm, *Complexity*, no. 2021, p. 13, 2021.
- [24] B. Wang, Ranking Evaluation Metrics for Recommender Systems, towardsdatascience, 2021. [Tiešsaiste]. Available: <https://towardsdatascience.com/ranking-evaluation-metrics-for-recommender-systems-263d0a66ef54>. [Piekļūts 22 05 2022].

- [25] R. Turrin, A. Condorelli, P. Cremonesi un R. Pagano, Time-based TV programs prediction, *In Proceedings of the 1st Workshop on Recommender Systems for Television and Online Video (RecSysTV)*, Foster City, CA, USA, 2014.
- [26] D. Zibriczky, Z. Petres, M. Waszlavik and D. Tikk, EPG Content Recommendation in Large Scale: A Case Study on Interactive TV platform, *Proceedings of the 2013 12th International Conference on Machine Learning and Applications*, 2013.
- [27] I. Pilászy, D. Zibriczky and D. Tikk, Fast als-based matrix factorization for explicit and implicit feedback datasets, *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*, 2010.
- [28] N.-r. Kim, S. Oh and J.-H. Lee, A Television Recommender System Learning a User's Time-Aware Watching Patterns Using Quadratic Programming. *Appl. Sci., Appl. Sci.*, vol. 8, no. 8, 2018.
- [29] Z. Bahramian and R. A. Abbaspour, An Ontology-Based Tourism Recommender System Based on Spreading Activation Model, *International Conference on Sensors & Models in Remote Sensing & Photogrammetry*, Kish Island, Iran, 2015.
- [30] E. Kim, S. Pyo, E. Park and M. Kim, An Automatic Recommendation Scheme of TV Program Contents for (IP)TV Personalization, *IEEE TRANSACTIONS ON BROADCASTING*, vol. 57, no. 3, pp. 674.-684., 2011.
- [31] M. Rautiainen, A. Heikkinen, J. Sarvanko, V. Kostakos and M. Ylianttila, Kuukkeli-TV: Online content-based services and applications for broadcast TV with long-term user experiments, *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013.
- [32] B. Knijnenburg, M. Willemsen, Z. Gantner, H. Soncu and C. Newell, Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction*, vol. 22, p. 441–504, 2012.
- [33] Wikipedia, the free encyclopedia, Wikipedia Confusion_matrix, [Tiešsaiste]. Available: https://en.wikipedia.org/wiki/Confusion_matrix. [Pieklūts 22 05 2022].

- [34] Q. Lin, X. Wang, B. Hu, L. Ma, F. Chen, J. Li and C. A. Coello, Multiobjective Personalized Recommendation Algorithm Using Extreme Point Guided Evolutionary Computation, *Complexity*, no. 2018, p. 18, 2018.
- [35] H. Li, H. Xia, Y. Kang and U. M. Nashir, IPTV program recommendation based on combination strategies, *The Sixth International Multi-Conference on Engineering and Technology Innovation 2017*, Kunming, China, 2017.
- [36] H. Kriplani, Alternating Least Square for Implicit Dataset, towardsdatascience.com, 25 06 2019. [Tiešsaiste]. Available: <https://towardsdatascience.com/alternating-least-square-for-implicit-dataset-with-code-8e7999277f4b> . [Piekļūts 2022].
- [37] Saeima, Elektronisko plašsaziņas līdzekļu likums, likumi.lv, 12 07 2010. [Tiešsaiste]. Available: <https://likumi.lv/ta/id/214039>. [Piekļūts 22 05 2022].
- [38] Wikidata, query.wikidata.org, Wikidata, 2019. [Tiešsaiste]. Available: <https://query.wikidata.org/sparql>. [Piekļūts 22 05 2022].
- [39] J. Han, M. Kamber and J. Pei, 2.4.7 Cosine Similarity, *Data Mining (Third Edition): Concepts and Techniques*, 2012, pp. 39-82.
- [40] F. Pedregosa, G. Varoquaux, A. Gramfort and V. Michel, Scikit-learn: Machine Learning in {P}ython, *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [41] Y. Hu, Y. Koren and C. Volinsky, Collaborative Filtering for Implicit Feedback Datasets, *IEEE International Conference on Data Mining (ICDM)*, 2008.
- [42] B. Frederickson, github, 2022. [Tiešsaiste]. Available: <https://benfred.github.io/implicit/api/index.html>. [Piekļūts 22 05 2022].
- [43] E. Rosenthal, Intro to Implicit Matrix Factorization: Classic ALS with Sketchfab Models, ethanrosenthal, 19 10 2016. [Tiešsaiste]. Available: <https://www.ethanrosenthal.com/>. [Piekļūts 22 05 2022].

- [44] M. Malaeb, Recall and Precision at k for Recommender Systems, Medium, 2017. [Tiešsaiste]. Available: https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54. [Pieklūts 22 05 2022].
- [45] H. Lee, Fast Python Collaborative Filtering for Implicit Feedback Datasets, github, [Tiešsaiste]. Available: <https://github.com/benfred/implicit/blob/48457af7ae6117720ce13797bbabf3c96db96a72/implicit/evaluation.pyx#L262>. [Pieklūts 22 05 2022].
- [46] Z. Zolaktaf, R. Babanezhad and R. Pottinger, A Generic Top-N Recommendation Framework For Trading-Off Accuracy, Novelty, and Coverage, *IEEE 34th International Conference on Data Engineering (ICDE)*, 2018.
- [47] X. He, H. Zhang, M.-Y. Kan and T.-S. Chua, Fast Matrix Factorization for Online Recommendation with Implicit Feedback, *SIGIR '16: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016.

PIELIKUMI

1. pielikums. *Github* repozitorija apraksts

Kodu piemēri pieejami sekojošā github repozitorijā:

https://github.com/Hubabuberts/EPG_recommendations

Repozitorijā apkopoti python kodu paraugi, kuriem līdzīgi izmantoti TV ieteikumu sistēmu izstrādē:

- *EPG_movie_name_cleaning.ipynb* - nosaukumu tīrīšanas piemērs filmām.
- *EPG_movies_content_similar_movie_calculation.ipynb* - CB līdzīgo filmu aprēķina piemērs filmām.
- *MF_parameter_grid_search_movies.ipynb* - CF ieteikumu aprēķina piemērs un modeļa trenēšana filmām.
- *Popular_content_rek_movies.ipynb* - populārākā satura aprēķina piemērs filmām.
- *TV_viewer_cluster_analysis_KNN.ipynb* - klasteru analīzes piemērs paradumu analīzei.
- *Wikidata_movie_data_queries.ipynb* - Wikidata datu ieguves piemērs filmām.

2. pielikums. Iegūto datu piemērs filmu metadatiem no *Wikidata*

imdb	label_lv	label_en	label_ru	location	released	cast	directors
tt10223460	Precēsīmies	Marry Me	Первый встречный	United States of America	2022	Jimmy Fallon, Sarah Silverman, Owen Wilson...	Kat Coiro
tt11138512	Ziemeļnieks	The Northman	Варяг	United States of America, United Kingdom	2022	Willem Dafoe, Alexander Skarsgård, Kate Dickie...	Robert Eggers
tt11245972	Kliedziens	Scream	Крик	United States of America	2022	Dylan Minnette, Roger L. Jackson, Courteney Cox...	Tyler Gillett, Matt Bettinelli-Olpin
tt11252248	Suns	Dog	Лулу и Бригге	United States of America	2022	Q'orianka Kilcher, Jane Adams, Ethan Suplee...	Reid Carolin, Channing Tatum

3. pielikums. Ieteikumu paraugs filmām.

Ieteikumi filmām				
NR	A	B	C	D
1	frozen	отдел 44	madagascar escape 2 africa	frozen
2	moana	мammu es tevi mīlu	monsters vs aliens	ghostbusters
3	rio 2	владимирская 15	tower heist	vella kalpi
4	pārklenda	код константина	the house with a clock in its walls	hotel transylvania
5	onward	і будуть люди	about time	gemini man
6	izvēle	прокуроры	balta balta diena	zootopia
7	madagascar escape 2 africa	engela acis	007 координаты скайфолл	onward
8	mīlas nedienas	burvis vārdā rudzis	topsy turvy	ukrainas piezīmes pa asiņainām pēdām dokumentāls stāsts
9	labie meli	latviešu jaunie režisori lāsma bērtule augstas gudrības grāmata no pasaules un dabas	прогулка	владимирская 15
10	tu neesi tu	dejo kamēr vari	la delicatessen	home

Atbilstošs ieteikums

Neitrāli

Ieteikums nav

aktuāls

Komentāri:

4. pielikums. Ieteikumu paraugs seriāliem.

Ieteikumi seriāliem					
NR	A	B	C	D, ar redzēto	D, bez redzētā
1	criminal minds	gandrīz ideālas vakariņas	aizliegtais paņēmiens	gandrīz ideālas vakariņas	rīga 56°/24
2	paraugprāva spriedzes seriāls	sturm der liebe	ceturtnā studija	bez tabu	vikings
3	father brown	слепая	lielās patiesības	es mīlu tevi latvija	tēvs brauns
4	birminghamas banda kriminālseriāls	пав patrol	1 1 aktuālā intervija	superbingo	abandoned engineering
5	the brokenwood mysteries	nemīlētie	province	nekā personīga	zebra
6	the mentalist	Morden im Norden	šodienas jautājums	mini pa trim	manas mājas ar ikea
7	amatieris	bez tabu	bez tabu	murdoch mysteries	красная королева
8	banda kriminālseriāls	blaze and the monster machines	s w a t	rīga 56°/24	endeavour
9	caur erkšķiem uz	šodienas jautājums	nemīlētie	vikings	nākamā pietura
10	glabiet musu gimeni	the loud house	rīga 56°/24	tēvs brauns	miss fisher's murder mysteries

Atbilstošs ieteikums

Neitrāli

Ieteikums nav aktuāls

Komentāri:

Maģistra darbs “**Televīzijas satura personalizētu ieteikumu sistēmas izstrāde**” izstrādāts
LU Datorikas fakultātē.

Darba teksta galīgā versija izgatavota **2022.05.20**.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie
informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: _____

(Autora paraksts un datums)

Ar savu parakstu apliecinu, ka esmu lasījis augstāk minēto maģistra darbu un atzīstu to
par **p i e m ē r o t u / n e p i e m ē r o t u** (nevajadzīgo svītrot) aizstāvēšanai Latvijas
Universitātes datorzinātņu maģistrantūrā.

Darba vadītājs: _____

(Vadītāja paraksts un datums)

Darbs iesniegts **maģistratūras sekretariātā** _____.

(Iesniegšanas datums)

Ar šo es apliecinu, ka darba elektroniskā versija ir augšupielādēta LU informatīvajā sistēmā.

Studiju metodiķe: _____.

(Metodiķes paraksts)

Recenzents: _____

Docenta p.i, Dr. dat., Lelde Lāce

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

_____ prot. Nr. _____

(Darba aizstāvēšanas datums)

Komisijas sekretārs: _____

(Sekretāra paraksts)