TARTU ÜLIKOOL
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikule õppekava

Mait Lindpere

Loomuliku keele töötlemise algoritmide kasutamine muusikasarnasuse leidmisel

Magistritöö (15 EAP)

Juhendaja: Anna Aljanaki, PhD

Tartu 2022

Loomuliku keele töötlemise algoritmide kasutamine muusikasarnasuse leidmisel

Lühikokkuvõte:

Käesoleva magistritöö raames viiakse läbi eksperimentaalne uurimus leidmaks kas muusika semantilise kirjelduse (antud töö kontekstis albumikirjelduste) järgi on võimalik muusikat soovitada. Püstitatud ülesande lahendamiseks võrreldakse NLP meetoditega muusika metaandmetest avastatud tunnuste järgi leitud muusikasarnasust muusikakasutajatest inimeste sarnasuse hinnangutega. NLP meetodite sooritusele hinnangu andmiseks kasutatakse võrdluseks audio meetodite järgi leitud muusikasarnasust. Lisaks kombineeritakse uurimistöös NLP ja audio meetodeid. TagATune mängust kogutud inimekspertide hinnanguid muusikaklippide sarnasusele kasutati NLP ja audio meetodite tulemuste hindamisel tõe etalonina. Töö tulemusel selgus, et NLP meetoditega leitud muusikasarnasus on üksi kasutamiseks vähetäpne, paremasid tulemusi näitasid audio meetodid. Kombineerides mõlema lähenemise paremaid sooritusi näidanud meetodeid saavutati antud töö kontekstis kõige täpsem tulemus.

Võtmesõnad:

Andmeanalüüs, keeletöötlus, muusikasarnasus

CERCS:

P176 Tehisintellekt

Using natural language processing algorithms to find musical similarity

Abstract:

In this master's thesis, an experimental study was conducted to research whether it is feasible to recommend music according to it’s semantic description (in the context of this work, album descriptions). In order to solve the set task, NLP methods are used to extract the similarity of music from metadata and compare it to assessments of people who have rated the music in question. The music similarity found by audio methods is used for comparison to assess the performance of NLP methods. In addition, the research combines NLP and audio methods. The human expert’s similarity assessments of the music clips collected from the TagATune game were used as a benchmark (ground truth) in evaluating the results of NLP and audio methods. As a result of the research, it was found that the musical similarity

extracted with NLP methods alone is not accurate to use in music recommendation, audio methods showed better results. By combining the best-performing methods of both approaches, the most accurate result where obtained.

Keywords:

Data analysis, language processing, music similarity

CERCS:

P176 Artificial intelligence

Sisukord

Sissejuhatus

Digitaliseerimine on moel või teisel jõudnud kõikidesse valdkondadesse millega inimene tegeleb, raske on leida tegevusala, mida pole infotehnoloogia (IT) areng ja levik puudutanud. Gurjari ja Mooni [1] järgi on nii ka kunstimaailmas, mille valdkonnad on erineva sügavusega digitehnoloogiatega põimunud. Kui maalikunst ja skulptuur pole digitaliseerimisest eriti mõjutatud, siis tundmatuseni on muutnud kõik muusikaga seonduv, muutunud on kuidas muusikat kogutakse ja hoiustatakse, avardunud on muusikale ligipääs ning võib öelda, et enim on IT muutnud muusikatööstuse ärimudeleid. Muusikatööstuse märkimisväärseks osaks on muusikasoovitussüsteemid (täpsemalt punktis 3.2), millel on kaalukas mõju valdkonnale. Lisaks arvamuspõhisele filtreerimisele ehk *collaborative filtering* lähenemisele muusikasarnasuse leidmisel kasutatakse ka sisupõhiseid meetodeid. Sisupõhised meetodid võimaldavad erinevalt arvamuspõhisele filtreerimisele leida sarnasusi muusika vahel millel puudub kasutajainfo, selliseks muusikaks võib olla vähem populaarne või äsja avalikustatud heliteos. Antud uurimistöö otsib seoseid muusika semantilise kirjelduse ja muusika sarnasuse vahel ning uurib, kas muusika kirjaliku tõlgendamise järgi on võimalik sarnast muusikat leida ja soovitada?

Kirjelduste järgi muusikasarnasuse hindamine või kirjelduste lisamine muusikasarnasuse leidmise protsessi võib hinnatava sarnasuse täpsust oluliselt parendada. Mida rohkem on otsuste tegemiseks erinevatest allikatest ja erilaadsetel meetoditel kogutud informatsiooni, seda tõenäolisemalt on võimalik jõuda otsustamisprotsessis soovitud asjakohase lahenduseni. Muusikaga seotud informatsioon, mille järgi on muusikat võimalik iseloomustada, ei sisaldu ainult audiosignaalis. Erinevad tööd multimodaalse informatsiooni kasutamise kohta muusikasarnasuse leidmisel viitavad täpsuse paranemisele muusikasarnasuse leidmise tulemustes. Oramas *et al.* [2] on uurinud muusika klassifitseerimist audiosignaali ja albumi kaanepildi järgi, ning järeldanud, et audiosignaalist tuletatud ja visuaalselt leitud muusika tunnuste kombineerimine, parandas klassifitseerimise täpsust võrreldes mõlema meetodiga eraldi leitud tulemustega. Lisaks leidis Oramas *et al.* [3], et audiosignaalist ja artisti biograafiast leitud tunnuste põhjal tuletatud muusika sarnasuste kombineerimine tõstab samuti muusikasoovitussüsteemide täpsust. Erinevate meetodite kombineerimisel on paremaid tulemusi täheldanud ka Pandeya *et al.* [4] uurides muusika klassifitseerimist muusikavideotest saadud tunnuste järgi. Eelnimetatud uurimistööde tulemustele tuginedes saab väita, et erinevate meetoditega leitud muusikasarnasuse uurimine ning erinevate uurimismeetodite

kombineerimine on vajalik muusikasarnasuse täpsemaks määramiseks millest tulenevalt on võimalik luua paremini töötavaid soovitussüsteeme.

Alates 2000ndate aastate algusest on muusika andmeanalüütika (MIR - *Music Information Retrieval*) läbinud suure arengu kuid juba 2000ndate lõpuks oli muusikasarnasuse leidmisel antud valdkonda pidurdama hakanud nii-öelda „klaaslae efekt“ ehk audio signaalil põhinevate meetodite järgi leitud muusikasarnasuse täpsuse kasv oli oluliselt aeglustunud. Arengu peatumise põhjuseks võib pidada asjaolu, et muusika on enamat kui signaal, muusikal on kontekst, mida muusika esindab ning mis emotsioone tekitab. Seega ei tohiks muusikat hinnata ainult signaali järgi, muusikat peaks hindama tervikuna, milleks on signaal ja kontekst, mida muusikaga soovitakse väljendada. Probleemi, et muusikat ei saa täielikult määratleda audiosignaali järgi nimetatakse ka „semantika lõheks“ (*semantic gap*) [5]. Eelkirjeldatud takistuste ületamiseks on MIR edasiseks arenguks just muusikasarnasuse vaates, asjakohane uurida erinevaid muusika konteksti väljendamist võimaldavaid meetodeid, nagu näiteks loomuliku keele töötlemine (NLP - *Natural Language Processing*). NLP võimaluste kasutamine lisaks signaalipõhistele meetoditele võib olla lahenduseks nii-öelda „klaaslaest“ läbi murdmiseks.

Omanäolise vaatenurga MIR süsteemidele on avaldanud Sturm [6], kes leidis oma töös MIR süsteeme hinnates, et kõik tema töös uuritud MIR süsteemid on kallutatud ning pole täpsed. Autor tõstatas hüpoteesi, et MIR süsteemid on mõjutatud „Kavala Hansu Effekt“-i (*“Clever Hans Effect”*) poolt. Nimetatud efekt kirjeldab olukorda kus iseseisvalt probleemi lahendama loodud süsteemile on süsteemi loojate poolt eneseteadmata antud vihjeid kuidas probleemi lahendada ning tekib ekslik mulje, et loodud süsteem suudab iseseisvalt etteantud probleemi lahendada. “Kavala Hansu Efekt“ on oma nime saanud 20. sajandi alguses tegutsenud matemaatikaõpetaja Wilhelm von Osteni tegutsemisest, kes väitis, et tema hobune „Hans“ oskab arvutada ja demonstreeris oma hobuse võimeid laiale avalikkusele. Hobuse arvutamise võime seadis kahtluse alla Dr. Oskar Pfungst kes tõestas, et hobune oli treenitud küsimustele õigesti vastama varjatud märguannete peale. Kuigi Sturm leidis oma töös, et MIR süsteemid on kallutatud väitis ta samas, et see ei tähenda, et MIR mudelid on kasutud. Vigaste ja kallutatud mudelite täielikku kõrvale heitmist ei pidanud vajalikuks ka briti statistik George Box [7] kelle tööst on tekkinud tsitaat: *„All models are wrong but some are useful*" - kõik mudelid on valed, aga mõned on kasulikud.

Muusikasarnasuse hindamine erinevate meetoditega on relevantne muusikasoovitussüsteemide täpsuse parendamiseks, MIR valdkonna muusikasoovitussüsteemide nii-öelda „klaaslaest“ kõrgemale tõstmiseks kui ka olemasolevatele mudelite alternatiivide leidmiseks, et nimetatud valdkonna mudelite kvaliteeti tõsta. Eksperimenteerimine muusikasarnasuse leidmisel on vajalik, et katsetada erinevaid meetodeid leidmaks uusi teadmisi MIR valdkonnas. Antud magistritöö raames hinnatakse, kas muusika semantilise kirjelduse järgi on võimalik muusikat soovitada. Nimetatud hinnangu andmine toimub läbi uurimisküsimuse püstitamise: Kas muusika semantilise kirjelduse (muusika albumikirjelduse) järgi on otstarbekas muusikat soovitada?

Uurimisküsimusele vastuse leidmiseks võrreldakse NLP meetoditega muusika metaandmetest avastatud tunnuste järgi leitud muusikasarnasust muusikakasutajatest inimeste sarnasuse hinnangutega ning antakse ülevaade võrreldavate omavahelisest suhtest. Leitakse ka muusikasarnasus audio signaali põhjal saavutatud muusikasarnasuse ja inimekspertide hinnangu vahel, et anda võrdlev hinnang NLP meetodite sooritusele. Lisaks eksperimenteeritakse antud töös NLP meetodite ja audiosignaalist leitud tunnuste omavahelise kombineerimisega ning võrreldakse saadud muusikasarnasuse tulemusi inimekspertide sarnasushinnangutega. Uurimuse struktuuri ülevaade on lisatud töö lõppu (vt Lisa 1). Kahe lähenemise meetodite omavaheline kombineerimine annab võimaluse hinnata, kas muusika laiema konteksti hindamine muusikasarnasuse leidmisel on täpsust parendav mõju.

Töö koosneb kolmest osast. Esimeses tausta avavas osas antakse ülevaade MIR valdkonnast, muusikasarnasuse leidmise tähtsusest ja viisidest, tutvustatakse sarnasuse mõõdikuid ning antakse ülevaade sarnastest uuringutest. Teine osa tutvustab antud uurimustööks kasutatud metoodikat. Töö viimane osa kirjeldab saadud tulemusi.

1. Mõisted ja terminid

Metaandmed (ingl *metadata*) on andmed, mis kirjeldavad või määratlevad teisi andmeid[1]. Antud töö kontekstis võib metaandmeteks pidada ka muusika albumikirjeldusi.

Muusika andmeanalüütika - MIR (ingl *music information retrieval*) on interdistsiplinaarne uurimisvaldkond, mis tegeleb muusikast informatsiooni leidmisega[2].

Loomuliku keele töötlus – NLP (ingl *natural language processing*) on loomuliku keele arvutipõhine analüüs, mis põhineb intellektitehnikal ja matemaatilisel lingvistikal[3].

Tehisintellekt - AI (ingl *artificial intelligence*) on tehnilise süsteemi võime hankida, töödelda ja rakendada teadmust sarnaselt inimintellektiga[4].

Masinõpe – ML (ingl *machine learning*) on protsess, mis kasutab andmetest või kogemustest õppimise võimaldamiseks arvutustehnilisi meetodeid[5].

Süvaõpe – DL (ingl *deep learning*) on masinõppe haru, mis imiteerib ülesannete lahendamisel inimajule iseloomulikku närvivõrkude struktuuri ehk tehisnärvivõrke[6,7].

Tehisnärvivõrk ehk neurovõrk (ingl *artificial neural network*) on intellektitehniline andmetöötlusmudel, mis jäljendab bioloogilise neuronivõrgu omadusi[8].

Multimodaalsus on lähenemine mitme modaalsuse kaudu näiteks tekstiliselt, kuuldeliselt ja visuaalselt[9].

Veebikoorija (ingl *web scraper*) on andmete kaevandamine/andmekoorimine veebis[10]. Tegemist on algoritmiga mis, otsib ja kogub soovitud informatsiooni vastavalt sisestatud juhistele.

[1] https://akit.cyber.ee/term/3774-metadata
[2] https://en.wikipedia.org/wiki/Music_information_retrieval
[3] https://akit.cyber.ee/term/12606
[4] https://akit.cyber.ee/term/2183-tehisintellekt
[5] https://akit.cyber.ee/term/9968-masinope-automaatope
[6] https://en.wikipedia.org/wiki/Deep_learning
[7] https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network
[8] https://akit.cyber.ee/term/1638
[9] https://et.wikipedia.org/wiki/Multimodaalsus
[10] htthttps://akit.cyber.ee/term/6203-web-scrapingps://et.wikipedia.org/wiki/Multimodaalsus

2. Taust

2.1 MIR - muusika andmeanalüütika

Muusikaga seotud andmete kaeve, töötlemise ja andmetest informatsiooni leidmisega tegeleb MIR (*Music Information Retrieval*) uurimisvaldkond. Eesti keeles puudub laialtlevinud vaste terminile MIR, kuid Tartu Ülikooli MIR õppejõud Dr. Anna Aljanaki, kes tegeleb antud valdkonna uurimisega, on kasutusele võtnud eestikeelse termini „muusika andmeanalüütika". Downie [8] järgi on MIR teadusharu, mis uurib muusika eri tahke näiteks helikõrgust, tempot, harmooniat, tämbrit, noodikirja, laulusõnu ja muusika tekstilist kirjeldust, et leida tähenduslikke tunnuseid ja leida tunnuste seostest informatsiooni. Müller *et al.* [9] lisab, et MIR eesmärgiks on muusikast efektiivselt informatsiooni leidmiseks, organiseerimiseks ja mõistmiseks vajalike tehnikate ning tööriistade loomine. Kokkuvõtvalt võib öelda, et MIR on muusikaga seonduva mõtestamine andmeanalüütika meetodeid kasutades.

Termin *„Music Information Retrieval"* võeti kasutusele juba 1960ndail aastatel Michael Kassler'i poolt, kes oma uurimuses analüüsis klassikalise muusika noodikirja. Kuigi muusika on juba aastatuhandeid olnud inimkonda läbivalt saatvaks kaaslaseks, siis MIR kui uurimisvaldkond on oluliselt noorem, valdkonna alguseks võib pidada 2000ndate algust, kui peeti esimene Rahvusvaheline Muusika Andmeanalüütika Sümpoosion ehk *ISMIR – International Symposium on Music Information Retrieval* USA-s , Massachusettsi osariigis, Plymouthis[10].

Alates 2000ndate algusest on MIR valdkonnana pidevalt kasvanud ning populaarsust kogunud. Valdkonna populaarsuse kasvu põhjusteks saab tuua järgmisi põhjuseid: 1990ndail toimunud audio andmete pakkimise ehk andmete mahu vähendamise tehnoloogiate areng, arvutite arvutusvõimsuse suurenemisest tingitud arvutusaja vähenemine, erinevate muusikakandjate laialdane levik ja muusika voogedastusplatvormide tekkimine ning suur populaarsus [11]. Audio andmemahu vähendamise formaate on hulgaliselt, kuid siiani kõige levinum on MP3, mis võeti kasutusele juba 1991. aastal [12]. Arvutite arvutusvõimsuse kasvu väljendab ilmekalt asjaolu, et 1999. aastal oli suurim transistorite arv mikroprotsessoril 21.67 miljonit, aastaks 2017 oli sama näitaja juba 19.2 miljardit [13], mis on ligikaudu 900 kordne kasv. Kui lähtuda, et ligi 900 korda on vähenenud eeltoodud perioodi vältel ka arvutusaeg, võib näiteks tuua, et arvutusele, mis 2017. aastal võttis aega 1 minuti kulus 1999. aastal 15 tundi.

Wikipedia andmetel [14] on teemad mida MIR valdkonnas uuritakse näiteks muusika klassifitseerimine, muusika soovitussüsteemide parendamine, instrumentide automaatne tuvastamine, automaatne transkriptsioon ning isegi muusika automatiseeritud looming. Won *et al.* [15] on märkinud, et muusika klassifitseerimise võimalused on lõputud ning sobiva jaotamise viisi määrab soovitud tulemus, kuid enim on muusikat klassifitseeritud žanri, tuju ehk *mood,* kasutatavate instrumentide ja haaksõnade ehk *tag*-ide järgi. Muusika klassifitseerimise mudelite kasutamine aitab näiteks parendada erinevate rakenduste kasutajakogemust pakkudes kasutajatele täpsemaid soovitusi ja lihtsustades suurtes kogumikes sirvimist.

2.2 Muusika sarnasus ja rakendamine

Muusikatööstuse domineerivaks ärimudeliks on füüsiliste plaatide müügi asemel saanud muusika voogedastusteenused, mis pakuvad erinevate kokkulepete alusel oma klientidele miljoneid heliteoseid piiramatus mahus kuulamiseks. Voogedastusteenuse pakkuja äriline strateegia on hoida oma kliente võimalikult kaua pakutava teenuse maksva tellijana, selleks on vaja pakkuda tarbijale pidevalt teda huvitavat sisu. Konkurents muusika voogedastusteenuste pakkujate seas on ülimalt tihe. Piiratud hulga klientidele üritavad oma teenusega muljet avaldada „hiiglased“ nagu Apple Music, Google Play Music, Spotify, Youtube Music ja Amazon Music ning lisaks nendele on klientide eest võitlemas hulgaliselt väiksemaid turuosalisi.

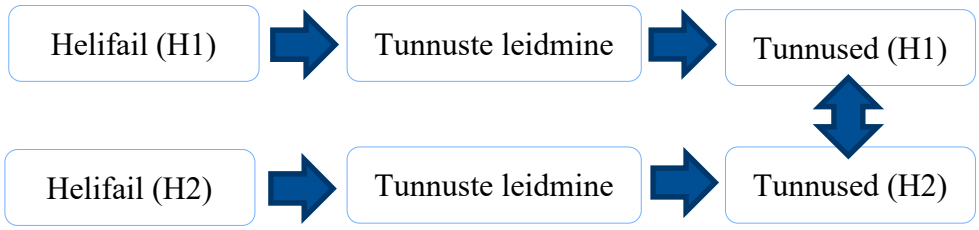
Voogedastusteenuse kasutajate iseseisev miljonite muusikapalade läbikuulamine endale meelepäraste muusikateoste leidmiseks on aeganõudev ülesanne, lisaks võib see kasutaja jaoks olla ebameeldiv tegevus, mis peletab kliendi teenuse juurest. Spotify tegevjuhi Daniel Eki sõnul laaditakse nende platvormile 2022. aasta veebruari seisuga igapäevaselt 60 000 uut lugu [16]. Tänapäeva mistahes valdkonna teenuste kliendid on järjest nõudlikumad ja eeldavad personaliseeritud kliendikogemust kohe, kui alustatakse teenuse kasutamist. Klientide hoidmiseks peab muusika voogedastuseteenuse pakkuja esimesest kokkupuutest oma kliendiga pakkuma talle koheselt personaliseeritud sisu, et vältida kliendi liikumist konkurendi teenusekasutajaks. Lisaks ei piisa antud valdkonnas ellujäämiseks ainult uute klientide leidmisest, klientide hoidmise pärast tuleb pidevalt pingutada ja selleks on vaja oma teenust pidevalt parendada.

Üheks võimaluseks muusika voogedastusteenuse pakkujatel klientide hoidmiseks on neile pakkuda neid huvitavat personaliseeritud sisu ning läbi selle hoida neid enda teenusest

huvitatuna. Seyerlehner [17] väidab oma doktoritöös, et muusika soovitussüsteemide peamine ülesanne on abistada muusika kuulajat, kui teenuse kasutajat endale meelepärase muusika otsimisel. Seega on täpselt toimiva muusikasoovitussüsteemi olemasolu igale muusikat teenusena pakkuvale ettevõttele elutähtis. Kasutajale varjatud soovitussüsteemi abil saab teenuse kasutaja hea kliendikogemuse, kuna tema vajadused saavad rahuldatud. Eelnevat saab tuua põhjuseks, miks muusika soovitussüsteemide arendamine ja uurimustöö antud suunal on väga aktiivne ja vajab pidevat edasiarendusi. Hea soovitussüsteem ei taga äriedu pikemas perspektiivis, kuid annab eduks eeldused ning kui pidevalt oma toimivat süsteemi parendada võib see anda eelise konkurentide ees.

Personaliseeritud sisu pakutakse kasutades erineva põhimõttega soovitussüsteeme. Lihtsamad muusika soovitussüsteemid töötavad näiteks: *user-user* - sarnastele kasutajatele soovitatakse sarnast sisu või *item-item* - kasutajale pakutakse tema poolt tarbitud sisule sarnast sisu põhimõttel. Sellest tulenevalt on muusika sarnasuse meetodite pidev parendamine soovitussüsteemide arendamiseks relevantne probleem. Kuigi *item-item* soovitussüsteemi lähenemine on küllaltki algeline ja juba võrdlemisi pikalt kasutusel olnud algoritm, on tegemist endiselt laialt kasutust leidva meetodiga. Selle tõestuseks on sündmus, et ajakiri *„IEEE Internet Computing“* valis 2017. aastal kõigi oma publikatsioonide seast välja artikli, mis on olnud enim ajatu ning selleks oli 2003. aastal Amazoni töötajate poolt avaldatud *„Amazon.com Recommendations: Item-to-Item Collaborative Filtering”*[18].

Item-item meetodi eelduseks on omavahel võrreldavate muusikapalade sarnasuse määramine erinevate valitud aspektide alusel, seda lähenemist kasutatakse ka antud töös. Muusikasoovitussüsteemi toimimiseks vajaliku sarnasuse saab leida nii sisupõhiselt *(content-based*) kui ka kontekstipõhiselt (*context-based*). Käesolevas töös kasutatakse sisupõhist muusikasarnasuse leidmise lähenemist. Sisupõhine sarnasus leitakse helifailist tuletatud tähenduslike tunnuste omavahelise sarnasuse hindamisel (vt Joonis 1) [17].



$$Sim_{(H1,H2)} = Sim_{(TH1;TH2)}$$

Joonis 1. Sarnasuse leidmine sisupõhiste tunnuste järgi (täiendatud [17] järgi)

Sisupõhine muusikasarnasus on leitav ka metaandmete põhjal. Eelnev joonis 1 kehtib ka metaandmete järgi sarnasuse leidmisel, kui sisendfail asendada metaandmetega. Sarnasuse leidmiseks vajalikud metaandmed võivad olla näiteks laulusõnad ja muusika tekstiline kirjeldus.

Igasuguse muusikasarnasuse valideerimise üheks meetodiks on tõepõhine lähenemine *(ground truth-based approach)*. Tõde ehk *ground truth* edaspidi GT, saavutatakse inimekspertide hinnangute kirjeldamisel ning seda saab kasutada hilisemate uuringute käigus tõe etalonina [1].

2.3 Muusikasarnasuse leidmine NLP ja audiosignaali kaudu

NLP lihtsustatud olemus on USA tehnoloogiaettevõtte IBM andmetel [19] järgmine: NLP on arvutiteaduse tehisintellekti (AI) haru, mille eesmärgiks on luua arvutitele võimekus mõista teksti ja kõne sarnaselt inimestele. Võiks arvata, et inimkeele arvutile arusaadavaks tegemine on kerge ülesanne, tuleb vaid tähed asendada numbritega ning probleem on lahendatud. Tegelikkuses on tegemist ülimalt keerulise ülesandega, mille lahendamisega ka tänapäeva teadus aktiivselt tegeleb. Inimkeele mõistmise teevad arvutile keeruliseks inimkeele iseärasused nagu homonüümid, sarkasm, idioomid, metafoorid, grammatika ja muud erisused, mida inimesed ise aastaid õpivad. NLP tegevussuundasid on rohkearvuliselt, kuid näitena võib välja tuua kõnetuvastuse, olemituvastuse, tekstist hoiakute leidmise ja inimesele arusaadava ning seostatud teksti automaatse genereerimise. Eelnimetatud ülesannete lahendamiseks on vabavarana kättesaadavad mitmed NLP tööriistad (tarkvara kontekstis võib öelda ka tööriistakomplektid – *toolboxes* või *toolkits*), näiteks NLTK ehk *Natural Language Toolkit,* mis on saadaval Pythoni programmeerimiskeele teegina. Tarkvaralised NLP tööriistad kasutavad ülesannete lahendamiseks statistika, masinõppe (ML) ja süvaõppe (DL) meetodeid. NLP rakendamise näideteks reaalses elus on rämpsposti tuvastamine, masintõlge, virtuaalsed agendid ja vestlusrobotid, sotsiaalmeedia sentimendi analüüs ja tekstidest automaatsete kokkuvõtete tegemine.

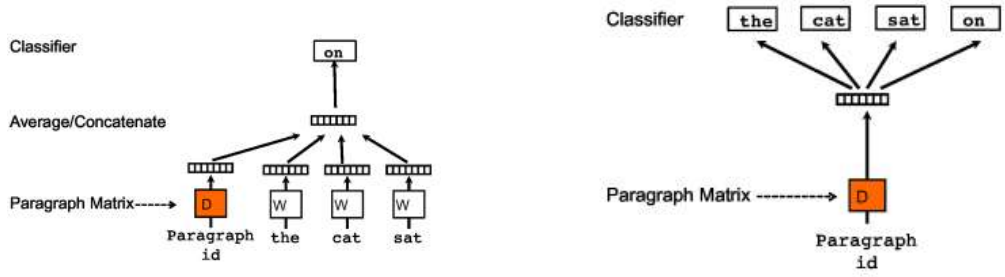
ML on sisuliselt automatiseeritud tähenduslike seoste leidmine andmehulkadest. ML-i eristab tavalisest, kindlat ülesannet täitvast algoritmist võime „õppida oma kogemusest“, see tähendab teha korrektuure käsiloleva probleemi lahendamise algoritmis selliselt, et paraneks probleemi lahendamise kvaliteet, kiirus või täpsus. ML meetodid edestavad oma sooritusvõime poolest lihtsaid algoritme õppimis- ja kohastumisvõime olemasolu pärast ning sellest tulenevalt on võimalik masinõpet rakendada inimesele omaste tegevuste käigus nagu näiteks

auto juhtimine, kõnest arusaamine ja piltide sisu mõistmine. ML eeliseks saab pidada ka võimet analüüsida inimesele hoomamatult suuri andmehulki sarnaselt inimese mõtlemisele. Digitaalse andmehulga kolossaalse kasvu ja inimese andmetöötlemise võime lõhe kasvab pidevalt, sellepärast avab ML meetodite kasutamine koos peaaegu piiramatu mälumahu ja pidevalt suureneva arvutite arvutuskiiruse kasvuga inimkonnale seninägematuid uusi võimalusi [20].

Käesolevas töös leitakse muusikasarnasust sisupõhiselt, kasutades erinevaid masinõppe mudeleid. Sisupõhise sarnasuse leidmiseks rakendatakse muusika kirjelduste omavaheliste seoste leidmisel NLP masinõppemudeleid nagu Doc2Vec ja BERT ning Pythoni programmeerimiskeele teeki musicnn.

2.3.1 Doc2Vec

Doc2Vec (D2V - *Document to Vector*) on Python'i Gensim *(Generate Similar)* [21] teegi Mikilovi ja Le [22] poolt välja pakutud juhendamata masinõppealgoritm, mis võimaldab mistahes mahus sisendteksti muuta vektoriks ehk numbriliseks representatsiooniks. Teksti vektoriline esitlus lubab ennustada, millised sõnad esinevad koos, sellest tulenevalt on tekstivektoritega võimalik edasi anda tekstis kirjeldatud konteksti ning saab võrrelda sõnade semantilist sarnasust. D2V võimaldab tekstivektoreid luua PV-DM (*Paragraph Vector – Distributed Memory*) ja PV-DBOW (*Paragraph Vector – Distributed Bag of Words)* (vt Joonis 2) meetodeid kasutades.



PV-DM PV-DBOW

Joonis 2. Doc2Vec PV-DM ja PV-DBOW lähenemised [22].

PV-DM lähenemise ennustab dokumendi vektori ja sõnade vektorite alusel järgmist sõna ja PV-DBOW lähtub ideest, et väljundsõnad ennustatakse sisendiks oleva dokumendi põhjal.

Mikilov ja Le [22] väidavad oma töös, et D2V meetod annab oluliselt paremaid tulemusi võrreldes teiste sarnaste meetoditega rakendades neid teksti klassifitseerimise ja konteksti mõistmise ülesannete lahendamisel. Lisaks märgivad autorid, et nende läbiviidud eksperimentides, kus hinnati PV-DM ja PV-DBOW sooritust erinevates aspektides, andis pidevalt paremaid tulemusi PV-DM.

2.3.2 BERT

BERT ehk *Bidirectional Encoder Representations from Transformers* on Google teadlaste Jacob Devlini, Ming-Wei Changi, Kenton Lee ja Kristina Toutanova [23] loodud ja 2018. aastal esitletud tehisnärvivõrkudel põhinev NLP mudel, mis on eeltreenitud erineva suurusega tekstikorpustel. Alates 2019. aastast on BERT rakendatud ka Google otsingumootori algoritmis. Erinevad BERT masinõppe mudelid on vabalt saadaval asjakohast informatsiooni koondaval TensorFlow [24] lehel. BERT mudelite produktiivsust ja jõudlust kirjeldavad ilmekalt tulemused, mida mudel erinevate ülesannete lahendamisel on saavutanud. Koroteev [25] on oma töös välja toonud, et GLUE *(General Language Understanding Evaluation)* testis, mille sisuks on hinnata NLP mudeli võimekust mõista loomulikku keelt, saavutas BERT teiste mudelitega võrreldes 4.5-7% paremaid tulemusi. Rakendades BERT-i SQuAD-l (*The Stanford Question Answering Dataset*) – mis on andmestik küsimustega, millele pakutakse vastusteks tekstilõike ja hinnatav mudel peab aru saama, millises pakutud lõigus on vastus küsimusele, näitas BERT parimat tulemust seni testitud mudelitest. BERT-i tulemus SQuAD-i testis oli 83.1, võrdluseks eelmine parim tulemus oli 78.0 ja inimese tulemus oli 89.5. Kolmanda testina läbis BERT SWAG (*The Situations With Adversarial Generations*) testi, kus tekstist arusaamise hindamiseks esitati nelja valikvastusega küsimusi, saavutades täpsuse 86.3, võrdluseks inimekspert saavutas sama testi tulemuseks 85.0.

2.3.3 Musicnn

Musicnn hääldatakse *musician* ehk eesti keeles muusik või moosekant, on eeltreenitud tehisnärvivõrkudel põhinevate masinõppe algoritmide teek. Musicnn teek sisaldab mitmeid mudeleid, mis on eeltreenitud kahel muusika andmestikul: MagnaTagATune (MTT) 19 000 laulu ja Million Song Dataset (MSD) 200 000 laulu, mis võimaldab audiofaili märgistada ehk *tag*-ida ja teostada andmekaevet, mille tulemusel on võimalik hinnatavat audiosignaali arvuliselt esitada. MTT ja MSD andmestikel treenitud mudeleid on hinnatud vastavalt: ROC-AUC 0.91; PR-AUC 0.38 (MTT) ja ROC-AUC 0.88; PR-AUC 0.29 (MSD) [26].

Musicnn dokumentatsioon ja kood on vabalt saadaval Github keskkonnas [27]. Antud meetodi kasutamiseks on sisendina vajalik audiofail, MP3 formaadis.

2.4 Sarnasuse mõõdikud

Sarnasus on abstraktne mõiste, lisaks võib selle hindamine olla subjektiivne. Näiteks saab kahte mistahes inimest kindlasti pidada sarnaseks näiteks bioloogiliselt, kuid neid võib ka pidada üksteisest erinevateks näiteks soo, rassi, pikkuse, kaalu, iseloomu, kultuurilise tausta ning teiste tunnuste järgi. Cambridge'i sõnastiku [28] järgi on sarnasus fakt, et kaks isikut või objekti on samad või näevad samad välja. Kuid kui on vaja anda hinnang sarnasusele, näites kui sarnased on omavahel euroopa mees ja austraalia naine võime väita, et kellegi vaates on nad identselt sarnased, samas on nad ka täiesti erinevad. Paljudes valdkondades pole selline ebamäärane hindamine aktsepteeritav ning üheselt arusaadav sarnasuse hinnang on vajalik. Üheselt mõistetavaks sarnasuse hinnanguks oleks numbriline hinnang skaalal 0–1 või (-1) – 1 kus 0 ja -1 märgiks minimaalset sarnasust ja 1 maksimaalset sarnasust. Objektidel võib olla üks kuni mitu tunnust ja mitme tunnusega objekte võib nimetada multidimensionaalseteks. Multidimensionaalseid objekte saab esitada vektorite kujul, kus objekti omadused on väljendatud vektori atribuutidena. Järgnevalt on esitatud kaks meetodit, kuidas andmeteaduse ja masinõppe kontekstis [29] vektorite vahelist sarnasust leida. Esiteks Eukleidese kaugus *(Euclidean distance*), mis mõõdab kahe punkti omavahelist kaugust N dimensionaalses ruumis ja avaldub:

$$d_{x,y} = \sqrt{\sum_{j=1}^{J} (x_j - y_j)^2}$$

kus d – kaugus,

J- dimensioonide arv,

x – punkti x koordinaat

Y – punkti y koordinaat [30].

Eukleidese kaugus on üks enimlevinud sarnasuse mõõtmise meetod, kuid ei sobi erineva mahuga objektide hindamiseks ja annab paremaid tulemusi, kui objektid on hinnatavad samal skaalal [31]. Erineva mahuga on antud uurimistöö kontekstis albumite sõnalised kirjeldused. Näiteks saab tuua 2 albumit, mille sarnasust hinnatakse albumikirjelduste järgi.

Esimene albumikirjeldus on 100 sõna pikkune ning albumi kirjelduses kasutatakse sõna S1 1 kord ja sõna S2 samuti 1 kord ning võib väita, et tekst on kahe sõna teemade vahel võrdselt jaotatud. Teine albumikirjeldus on 200 sõna pikkune ja albumi kirjelduses kasutatakse sõna S1 2 korda ja S2 samuti 2 korda ehk ka selle albumi kirjelduse sisu on jaotatud võrdselt kahe sama teema vahel. Seega näiteks toodud 2 albumit on mõlemad võrdselt jaotatud kahe teema vahel ehk 50% teema S1 ja 50% teema S2 vaatamata albumikirjelduse pikkusest. Asetades 2 albumit sõnade esinemise arvu järgi kahemõõtmelisse ruumi vastavalt: album 1 (1;1) ja album 2 (2;2) saame kahe albumi vahel Eukleidese kaugust mõõtes tulemuseks, et tegemist on omavahel erinevate albumitega, olgugi, et sisu on mõlemal sama.

Eelnimetatud sarnaste objektide omavahel eksitavalt erinevatena näitamise probleemi lahendamiseks saab kasutada Koosinus sarnasust (*Cosine similarity*), mis määrab kahe vektori omavahelise nurga ja avaldub:

$$cos_{sim}(u, v) = \cos(\theta) = \frac{\Sigma_{i=1}^{n} u_i v_i}{\sqrt{\Sigma_{i=1}^{n} {u_i}^2}\sqrt{\Sigma_{i=1}^{n} {v_i}^2}}$$

kus cos_{sim} – koosinus sarnasus,

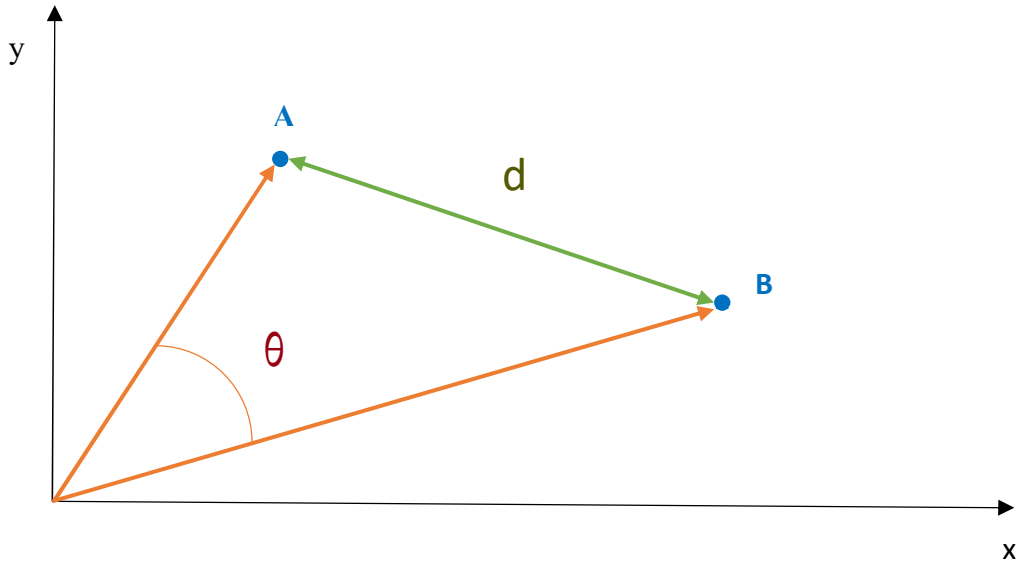
u – vektor 1,

v – vektor 2,

cos – koosinusfunktsioon,

θ – nurk vektorite vahel [32].

Koosinus sarnasus avaldub numbriliselt vahemikus -1 kuni 1 ehk cos(0°) = 1, mis näitab, et kahe vektori vaheline nurk on 0°, seega sarnasus on 1 ehk identne. Sarnasuse täielik puudumine avaldub -1 ehk cos(180°)= -1[31]. Eukleidese kauguse ja Koosinus sarnasuse omavaheline seos on illustreeritud järgneval joonisel (vt joonis 3), kus d = Eukleidese kaugus ja θ = Koosinus sarnasus.



Joonis 3. Eukleidese kauguse ja Koosinus sarnasuse omavaheline seos.

Antud uurimistöös kasutatakse sarnasuse leidmisel vektoreid, mida võib pidada atribuutide rohkuse järgi suuremõõtmelisteks. Eukleidese kauguse leidmise efektiivsus langeb, kui võrreldava vektori dimensionaalsus tõuseb [33], samas Koosinus sarnasus on vektori pikkusest sõltumatu [34], sellest tulenevalt kasutatakse antud töös sarnasuse leidmiseks Koosinus sarnasust.

2.5 Varasemad uuringud

Varasemates uuringutes on Gali ja Tiwari [35] leidnud, et tehisnärvivõrkudel põhineva NLP mudeli Word2Vec *(Word to Vector)* baasil loodud D2V-i võiks kasutada just muusikasarnasuse leidmise uuringutes. D2V võimaldab kogu uuritavat teksti esitada vektoriseeritud kujul ja sobib muusika tekstiliste kirjelduste omavaheliseks võrdluseks. Muusika tekstiline kirjeldus muudetakse vektoriks, vektoreid on võimalik omavahel kvantitatiivselt võrrelda ja sellest tulenevalt saab võrrelda ka muusikat, mis muidu igale hindajale subjektiivse objektina tundub. D2V mudelis võib muusika kontekstis dokumendiks pidada näiteks laulu sõnu või albumi tekstilisi kirjeldusi. NLP meetodite sooritusvõimet on hindanud Cunha *et. al* [36] võrreldes omavahel Word2Vec ja TF-IDF (*Term frequency–inverse document frequency*) mudeleid. Kuigi eelnimetatud mudelitest on Word2Vec hilisem ja peaks olema arenenum mudel leiti, et paremaid tulemusi andis hoopis TF-IDF. Lisaks märgiti eelnimetatud uuringus, et TF-IDF on oluliselt kiirem. Seega saab väita, et W2V-i kohta esineb erinevates

uuringutes vastuolulist informatsiooni ning selle edasiarenduse D2V kasutamine vajab lisauurimist.

Fell [37] väidab oma doktoritöös, et MIR on oma uuringutes suuresti olnud kaldu audio signaalist informatsiooni kaevandamisele, kuid muusika soovitamisteenuste arendamisel on hakatud järjest enam kasutama ka muid muusika metaandmeid nagu näiteks lüürikat. Felli doktoritöös osutus laulusõnade klassifitseerimisel edukaimaks BERT mudel 84,4% täpsusega, samas audiosignaalist leitud tunnuste põhjal muusika klassifitseerimise täpsuseks võib olla üle 90% [38]. Lisama peab, et BERT mudelite sooritusvõimet on mudelit peenhäälestades võimalik tõsta, seda on näitlikustanud Kelly *et al.* [39], kes oma katsetustes tõstsid BERT mudeli täpsuse peenhäälestades 72.7% -lt – 76.4% -ni. Sellest võib järeldada, et NLP ja audio mudelite kombineerimisel ja peenhäälestamisel võib saavutada täpseid muusikasarnasuse määramisi.

Omavahel audiosignaali põhjal tunnuseid leidvaid mudeleid kombineerinud Chathuranga ja Jayaratne [40] väidavad, et kombineeritud ehk *ensemble* meetodid (vt täpsemalt punktis 4) näitavad klassifitseerimise ülesannetes paremaid tulemusi kui üksikud meetodid. Erinevate uurimissuundade mudelite kombineerimist muusikasarnasuse leidmisel on uurinud Mayer ja Rauber [41] kombineerides NLP ja audiosignaali tunnuseid muusikasarnasuse määramisse leidsid, et *ensemble* meetodite sooritus on kuni 6 protsendipunkti parem kui üksikute mudelite kasutamine. Hindamaks muusikat kui tervikut oleks vajalik kasutada multimodaalset lähenemist, selleks peab kombineerima erinevate aspektide järgi muusikast tunnuseid leidvaid mudeleid. Kuigi antud valdkonda on palju uuritud ning jõutud ka võrdlemisi heade tulemusteni, võib eksperimentaalne muusikasarnasuse uurimine viia uute teadmiste leidmiseni. Uurimiseks on kasutada erinevad NLP ja audiosignaali järgi muusikasarnasuse leidmise meetodid ning võrdlemiseks erinevaid sobivad andmestikke. Iga uurimistöö antud valdkonnas on uurimismeetodite ja erineva võrdlusandmestiku kombineerimisel unikaalne ja toetab MIR valdkonna teadmisi uute avastustega.

3. Metoodika

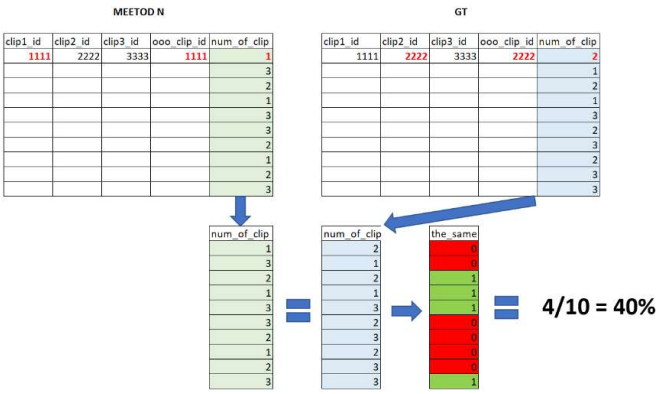
Uurimus teostatakse kasutades Pythoni programmeerimiskeelt, Google Colab keskkonnas, peamiseks tööriistaks Pandas andmeanalüüsi teek. Uurimuse ülesandeks on leida erinevate meetoditega muusika sarnasus ja võrrelda leitud sarnasusi valitud "tõe etaloniga"*(ground truth),* eesmärgiga anda hinnang erinevate meetodite efektiivsusele ning sellest järeldada, kas muusika albumite kirjelduste järgi on võimalik muusikat soovitada rahuldava täpsusega. Uurimisülesande täitmiseks vajalikud algandmed antud töö kontekstis on muusika metaandmed ja sama muusika audiosignaal. Töös kasutatav metoodika on uurimisülesannete kaupa jaotatud 4 mooduliks (vt Lisa 2):

1. Võrdluseks vajalike andmete kogumine ja sobivas formaadis "tõe etaloni" loomine;
2. Uurimuses kasutatavate muusika sarnasuse meetodite rakendamine;
3. Tulemuste kirjeldamine ja võrdlus;
4. Parimaid tulemusi näidanud mudelite koondamine *ensemble* meetodiks, rakendamine ja tulemuste kirjeldamine.

Uurimustöö 1. moodulis tutvutakse algandmetega, milleks kasutatakse MagnaTagATune andmestikku [42], mis põhineb vaba litsentsiga muusika kogumikul Magnatune, andmestik on saadaval London City ülikooli masinõppe ja meediainformaatika uurimisrühma koduleheküljel [43]. Kodulehelt laetakse alla võrdluses kasutatavate heliklippide meta-, sarnasuse- ja audioandmed. Meta- ja sarnasusandmed on saadaval CSV(*comma seperated value*) ning audioandmed MP3 formaadis. Andmestik kajastab TagATune mängu [44] tulemusel tekkinud muusikaklippide sarnasusandmeid. Sarnasusandmed väljenduvad muusikaklippide kolmikutena, mis on kasutajatele hindamiseks antud selliselt, et kasutajad peavad kolmest neile kuulata antud klipist leidma klipi, mis erineb kolmest enim. Kui sama kolmikut hindavad kaks mängijat valivad kolmikust sama klipi kõige erinevamaks, tekib sellekohane märge. Audiosignaali hakatakse võrdlema tekstiliste albumikirjelduste järgi, mis leitakse Magnatune.com veebilehelt, sellest tulenevalt luuakse *web scraper* ehk veebikraapija, et teostada andmekoorimine albumikirjelduste salvestamiseks. Magnatune veebilehe ja albumi kirjelduse näide on toodud järgneval joonisel:

Joonis 4. Ekraanitõmmis Magnatune lehelt albumi alamlehelt koos kirjeldusega (allbumi kirjeldus märgitud punasega).

Tõe etalon ehk GT, luuakse TagATune mängu kolmikute võrdluse tulemuste järgi. Kolmikute erinevuse tulemused on väljendatud järgmisel kujul: (clip1_id; clip2_id; clip3_id; clip1_numvotes; clip2_numvotes; clip3_numvotes), kus clipN_id viitab audioklipile ja clipN_numvotes näitab kui mitu korda leppisid 2 mängijat samaaegselt kokku, et antud muusikaklipp erineb ülejäänud kahest. GT väljendamise kujuks antud uurimuse kontekstis saab olema (clip1_id; clip2_id; clip3_id; ooo_clip_id; num_of_clip), kus clipN_id määrab klipi, ooo_clip_id tähistab klippi, mis on teistest enim erinev ja num_of_clip tähistab klipi järjekorra numbrit antud võrdluses vastavalt 1, 2 või 3. Järgneval joonisel on kujutatud töös kasutatud meetodite tulemuste võrdlus GT-ga:



Joonis 5. Meetodi võrdlemine GT-ga

Uurimuse 2. moodulis rakendatakse NLP meetodeid Magnatune leheküljelt andmekoorimise käigus kogutud albumikirjeldustel, et leida albumite omavaheline sarnasus. Oluline

on märkida, et antud uurimuse käigus võrreldakse omavahel muusika klippide põhjal loodud GT-d ja NLP meetoditega leitud sarnasust nende albumite kohta kus klipid on esindatud. NLP meetodid mida töös kasutatakse on D2V ja BERT ning nende modifikatsioonid vastavalt PV-DM ja PV-DBOW ning wiki_books ja wiki_books/qnli. BERT mudeli variatsioon wiki_books on eeltreenitud BooksCorpus ja Wikipedia andmestikel ning wiki_books/qnli on wiki_books versioon, mis on peenhäälestatud kasutades QNLI-d (*The Stanford Question Answering Dataset*) [45]. D2V PV-DM ja PV-DBOW modifikatsioone on antud töös kasutatud, et anda laiem vaade D2V meetodi järgi leitud muusikasarnasusele. Sarnaselt D2V meetodile on ka BERT meetodi laiema käsitluse huvides kasutatud kahte modifikatsiooni. Wiki_books on valitud kasutamiseks, sest tegemist on suure inglisekeelse tekstikorpuse peal treenitud mudeliga, mis sobib eelkõige just inglisekeelsete NLP ülesannete lahendamiseks [46]. Wiki_books/qnli kasutamist antud töös õigustab mudeli hea sooritusvõime võrreldes teiste BERT wiki_books mudelitega, wiki-books/qnli peenhäälestatud mudel näitas testimisel MRPC *(The Microsoft Research Paraphrase Corpus)* korpusel teiste wiki_books variatsioonide seas parimat tulemust [45].

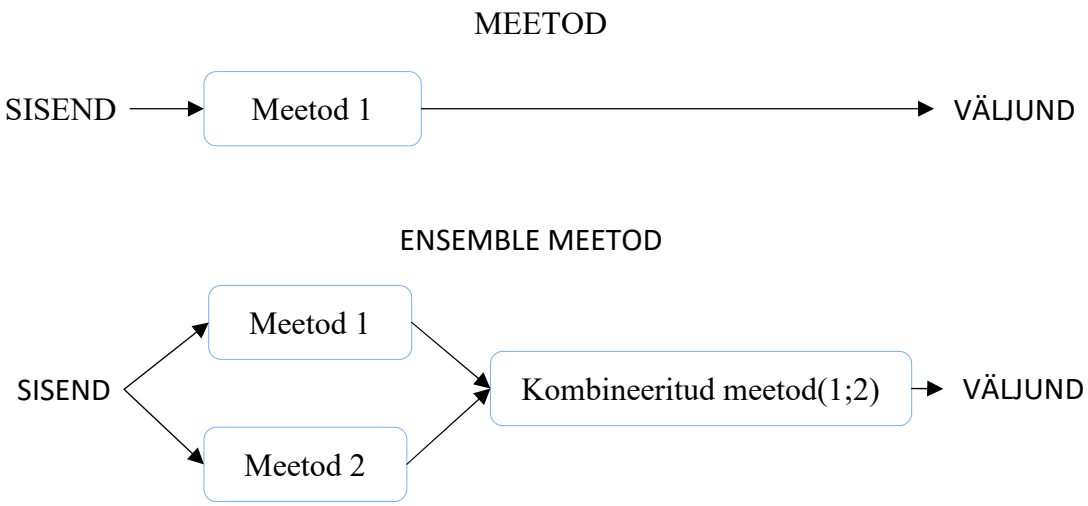
Audiosignaali põhjal muusikasarnasuse võrdlemiseks leitakse tunnused kasutades Pythoni Musicnn teeki. Töös kasutatakse kahel andmestikul eeltreenitud mudeleid MTT ja MSD, ning mõlema mudeli kaudu leitakse heliklippide tunnused kasutades Musicnn *music tagger* ja *music feature extractor* funktsioone. *Music tagger* võtab sisendiks audiofaili MP3 formaadis ja väljastab topN märgistust ehk *tag*-i (N: rock, pop, jazz, classical, slow, weird, happy, no voice jne.) failis oleva muusika kohta. Esimene ehk *top1 tag* iseloomustab enim hinnatavat muusikat ja kahanevas järjekorras väheneb ka kaal mille võrra *tag* muusikat karakteriseerib. TopN *tag*-ide järgi muudetakse hinnatav heliklipp vektoriliselt väljendatavaks ning vektorite omavahelist seost mõõdetakse Koosinus sarnasuse järgi. Kuigi *music tagger* võimaldab hinnatavat heliklippi iseloomustada kuni 50 *tag*-i järgi, kasutatakse antud töös top25 ehk 25 enim heliklippi iseloomustavat *tag*-i.

Music feature extractor võimaldab numbrilist väljavõtet Musicnn mudeli tehisnärvivõrkude erinevatest kihtidest, nimetatud numbrilisi väljavõtteid käsitletakse antud kontekstis muusika tunnustena *(music features).* Antud uurimuses kasutatakse Musicnn tehisnärvivõrgu lõpuosa kihti *penultimate. Penultimate* väljendab sisend audiosignaali (*mp3*) Pythoni programmeerimiskeele massiivina (ingl *array*), mis koosneb järjenditest (ingl *list*), *penultimate'i* väljundiks olev massiiv koosneb 9 järjendist, igaühes 200 elementi [47]. NLP ja audio

meetodite rakendamisest tekkinud muusika sarnasussanded konverteeritakse GT-ga samasse formaati ja võrreldakse eelkirjeldatud meetodil (vt Joonis 5).

Metoodika kolmas moodul keskendub NLP ja audio meetoditel põhinevate kaheksal viisil leitud muusikasarnasuste võrdlemist GT-ga. Määratakse kaheksal viisi protsentuaalne kokkulangevus GT-ga, leitakse NLP ja audio mudelite omavahelise võrdluse tulemusel mõlema meetodi täpsemad mudelid hilisemaks analüüsiks.

Neljandas metoodika moodulis kasutatakse NLP ja audio tõhusamaid mudeleid, et luua *ensemble* meetod, mis kombineerib mõlema mudeli tulemused. *Ensemble* meetod tähendab mitme meetodi kombineeritud kasutamist sama probleemi lahendamiseks (vt Joonis 6).



Joonis 6. Meetodi ja *ensemble* meetodi struktuur (täiendatud [48] järgi)

Lähenemisi *ensemble* meetodi loomisel on erinevaid, kuid antud töös kasutatakse kaalutud keskmise järgi loodud meetodite kombineerimist, mis tähendab meetodite mõjule *ensemble* meetodis osakaalu määramist ehk sisendmeetoditel on erinev efekt väljundile [48]. *Ensemble* meetodi kasutamine NLP ja audio meetodite baasil loodud sarnasuse hindamisel võimaldab muusikat hinnata kui tervikut nagu töö sissejuhatuses välja toodud, siis muusikal on mitmeid tahke ning ainult ühekülgne hindamine võib põhjustada uurimustööde tulemustele antud valdkonnas nii-öelda „klaaslae efekti“. Viimasena leitakse *ensemble* meetodiga

leitud muusikasarnasus ja võrreldakse GT-ga. Kokku antakse hinnang muusikasarnasuse leidmisele kasutades üheksat mudelit.

Kuna tegemist on eksperimentaalse uurimusega ja kasutatud NLP ning audio meetodeid rakendatakse uurimuses ilma, et nende usaldusväärsust oleks antud töös hinnatud, on oluline pöörata tähelepanu asjaolule, et antud uurimuse käigus leitud tulemused võivad olla juhuslikud. Antud töös leitakse NLP või audio meetodil loodud kolmikutest kõige erinevam klipp ja võrreldakse saadud tulemust GT-ga. Kolmikust juhuslikult GT-ga sama klipi leidmise tõenäosus on 1:3 (33%) ehk 122-st kolmikust leitakse sama erinev klipp *(odd-one-out)* ~ 40 korral. Kui leitud tulemusete täpsusprotsent on võrdne või ligilähedane 33% võib väita, et saadud tulemused võivad olla juhuslikud.

4. Tulemused

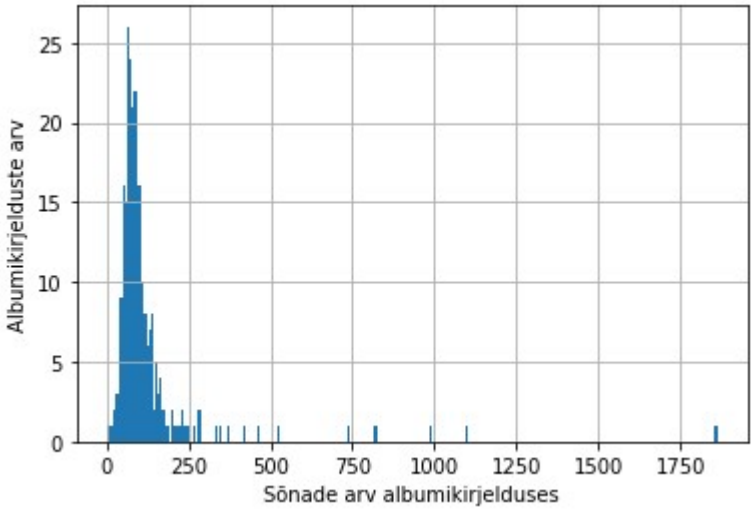
Uurimustöö läbiviimiseks loodud kood, töö käigus tekkinud failid ja audio analüüsiks kasutatud MP3 formaadis heliklipid on saadaval https://drive.google.com/drive/folders/15nMa7c-Eycy81wv-ZoyydsfBgkHbrtz0?usp=sharing.

GT loomiseks kasutatavad sarnasuseandmed MagnaTagATune andmestikust on saadaval CSV formaadis, klippide kolmikute kujul, kus on välja toodud kolmest klipist kõige rohkem erinevam klipp. Samast andmestikust on leitav ka andmestik kõikide kasutatavate klippide metaandmetega, kust on leitav link albumit tutvustavale kodulehele, kust heliklipp pärineb. Linki kasutatakse andmekoorimise teostamiseks, et luua andmestik, kus on iga klipi id-ga seostatud ka albumi tekstiline kirjeldus.

Heliklippe MagnaTagATune andmestikus on 31 382, mis pärinevad 517 albumilt. Seega on andmekoorimise eesmärk hankida 517 albumi kirjeldus. Andmekoorimise tulemusena selgus, et 114 albumil puudus kirjeldus ja 113 albumi kirjelduse *URL* andis „404“ veateate ehk lehekülge ei leitud või toimus muu tõrge. Andmekoorimise tulemusena saadi 290 albumi kirjeldus, millest 282 olid unikaalsed. Heliklippe mis pärinevad 290 albumilt ja millel on kirjeldused, on saadaval 17 046 (31 382-st). Erinevate albumite sama kirjeldus viitab sellele, et album koosnes mitmest plaadist ehk ühe albumi kaks plaati olid Megatune lehel kirjeldatud kui kaks erinevat albumit, aga nende tekstiline kirjeldus on sama. Leitud 290 albumikirjelduse jaotus pikkuse järgi on kirjeldatud järgmises tabelis (vt Tabel 1) ja joonisel (vt Joonis 7):

Tabel 1. Andmekoorimisel leitud albumikirjelduste sõnade pikkuse võrdlus.

Mõõdik	Sõnade arv
mean	120
std	155
min	3
25%	67
50%	87
75%	120
max	1862



Joonis 7. Andmekoorimisel leitud albumikirjelduste jaotus sõnade pikkuse järgi.

GT loomisel kasutatud kolmikuid on MagnaTagATune andmestikus 533, kuid kõik kolmikud GT loomiseks ei sobi, sest puuduvad kõikide kolmikus kasutatavate klippide albumi kirjeldused või on kolmikus vähemalt 2 ühesuguselt erinevaks hinnatud klippi, mis teeb kõige erinevama leidmise võimatuks. Pärast kolmikute eemaldamist, kus kõigi kolme võrreldava klipi kirjeldamiseks puudub kõikide klippide albumi kirjeldus jäi GT tekitamiseks sobilikke kolmikuid 152, millest omakorda eemaldati hinnangu järgi mittesobivad ning GT loomiseks kasutatakse 122 kolmikut, mis koosnevad 253 unikaalsest klipist.

NLP meetodite D2V ja BERT järgi leiti GT-s olevate muusikaklippidele vastav albumite sarnasus kasutades andmekraapimisel saadud albumikirjeldusi. Nii D2V kui ka BERT mudelile anti sisendiks 290 albumi kirjeldused ja saadi väljundiks albumite omavahelise sarnasuse maatriks (vt Joonis 8).

	0	1	2	3	4	5	6	7	8	9
0	1.000000	0.104162	0.416275	0.302859	0.375614	0.279265	0.285936	0.420616	0.197203	0.110919
1	0.104162	1.000000	0.162706	0.278896	0.125627	0.291543	0.307016	0.145389	0.151538	0.280911
2	0.416275	0.162706	1.000000	0.160720	0.323892	0.378315	0.185757	0.477883	0.114518	-0.009742
3	0.302859	0.278896	0.160720	1.000000	0.165356	0.039622	0.248512	0.530664	0.408268	0.209631
4	0.375614	0.125627	0.323892	0.165356	1.000000	0.116474	0.145516	0.414732	0.474321	0.490420

Joonis 8. Albumikirjelduste sarnasusmaatriks (tegemist osaga maatriksist, näitlikustamise eesmärgil)

Võrdlemaks saadud tulemusi loodi NLP meetodite järgi leitud muusikasarnasuse andmestik GT-ga samal kujul. GT-ga võrdlemiseks loodi andmestik, kus GT-s olevate klippide

sarnasusandmed muudeti vastavalt albumi sarnasusandmetele selliselt, et klipi andmed asendati albumi andmetega, kust heliklipp pärines. Kui GT-s võrreldakse omavahel kahe klipi erinevust, siis võrdlemiseks loodud andmestikus hinnatakse vastavaid albumeid.

NLP meetodite võrdlemisel GT-ga selgus, et parim tulemus saavutati D2V PV-DBOW mudelit kasutades (vt Tabel 2). Omavaheliste kolmikute hindamisel saavutati sama tulemus 55 korral 122-st ehk 45%.

Tabel 2. NLP meetoditel leitud kolmikute hindamise tulemused võrreldes GT-ga

Meetodid/andmestikud/mudelid		Tulemus	%
Ground truth (GT)		122	100%
D2V	PV-DM	53	43%
	PV-DBOW	**55**	**45%**
BERT	WIKI	52	43%
	WIKI/qnli	49	40%

Sama võrdlust korrati ka audiosignaali järgi hinnatud muusikasarnasuse võrdlemisel ning parima tulemuse saavutas Musicnn MTT (MagnaTagATune andmestik) eeltreenitud mudelist *tag*-ide järgi leitud muusikasarnasuse hindamine tulemus 78, 122-st ehk 64% (vt Tabel 3).

Tabel 3. Audio meetoditel leitud kolmikute hindamise tulemused võrreldes GT-ga

Meetodid/andmestikud/mudelid			Tulemus	%
Ground truth (GT)			122	100%
MUSICNN	**MTT**	**TAGS**	**78**	**64%**
		PENULTIMATE	69	57%
	MSD	TAGS	68	56%
		PENULTIMATE	71	58%

NLP ja audio meetodite omavahelise parima sarnasuse kasutades antud sarnasuse leidmise metoodikat andsid NLP – D2V PV-DBOW ja audio – Musicnn MTT *tag*-ide järgi leitud sarnasused (vt Lisa 3), näidates kokkulangevust 60, 122-st ehk 49%.

NLP ja audio järgi leitud muusikasarnasuse parimaid tulemusi näidanud mudeleid (vt Tabel 4) kasutades loodi *ensemble* meetod, et kombineerida kahe muusika aspekti: tekstilise kirjelduse ja audiosignaalil põhinev hindamine. *Ensemble* meetodi loomiseks on vajalik määrata meetodite kombineerimise osakaal ehk kui suurt mõju lõpptulemusel kombineeritud meetodid avaldavad. NLP parim tulemus oli 55, 122-st ning audio 78, 122-st. Kaalud määrati kahe meetodi tulemuste omavahelisest suhtest 55/78 ehk vastavalt .41 ja .59. *Ensemble*

meetodil leitud muusikasarnasuse võrdlemisel GT-ga saadi tulemuseks 79, 122-st ehk 65% sarnasus.

Tabel 4. NLP ja audio meetodite täpsus võrreldes GT-ga

Meetodid/andmestikud/mudelid			Vs. GT	GT %	Keskmine	Keskmine %	Kokku keskmine	Kokku keskmine %
NLP	D2V PV-DM		53	43%	52,25	43%	61,87	51%
	D2V PV-DBOW		55	45%				
	BERT WIKI		52	43%				
	BERT WIKI/qnli		49	40%				
Audio	MUSICNN-MTT	TAGS	78	64%	71,5	59%		
		PENULTIMATE	69	57%				
	MUSICNN-MSD	TAGS	68	56%				
		PENULTIMATE	71	58%				

Kasutatud mudelite sarnasuse tulemused (kokkulangevusi 122-st kolmiku erinevama klipi määramisel) võrreldes GT-ga olid NLP meetoditel 53, 55, 52, 49 ja audio meetoditel 78, 69, 68, 71, kokku ulatusega 29. Ulatus 29 näitab, et kõik saadud tulemused on koondunud 24% kogu 122 kolmiku võimalike variantide hulka. NLP meetodite keskmine täpsusprotsent võrreldes GT-ga on 43% ja audio meetodite vastav näitaja on 59%. Mõlemad eelnimetatud meetodite keskmised on suurema täpsusprotsendiga kui oleks juhuslikult (33%) leitud täpsus, vastavalt 10 ja 26 protsendipunkti täpsemad.

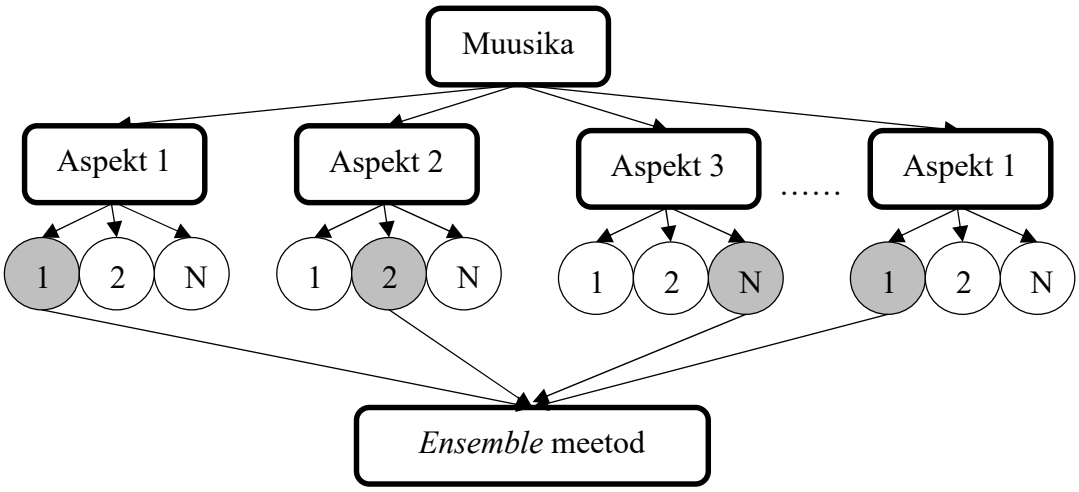
Meetodite sarnasuse tulemused on selgelt grupeeritavad NLP ja audio lähenemiste järgi (vt Lisa 4) ja täpsemad kui juhuslik valik. Kuna tulemused ei ümbritse juhuslikku täpsust, võib arvata, et tulemused ei ole juhuslikud.

Leitud tulemustele tuginedes saab väita, et audio meetoditel leitud muusikasarnasus on täpsem kui NLP meetodeid kasutades leitud muusikasarnasus. Kõik audio meetodid, mida töös kasutati olid täpsemad NLP mudelite järgi leitud sarnasusest. Parim NLP mudel näitas GT-ga sarnasust 45% ja parim audio mudel 64%, ehk antud töö kontekstis on audio meetod 42% parema täpsusega, mida iseloomustab võrreldavate 19 protsendipunktiline vahe. Kõikidest muusikasarnasuse hindamise meetoditest näitas parimat tulemust *ensemble* meetod, mis sai täpsuseks võrreldes GT-ga 65%. Lõpptulemusena leitud 65% muusikasarnasuse määramise

täpsust ei saa pidada rahuldavaks tulemuseks muusikasarnasuse määramisel, nagu antud töös eelnevalt mainitud, on võimalik NLP ja audio signaali põhjal saadud muusikasarnasuste leidmise täpsuseks pidada 80-90+%.

NLP meetodite kesise täpsuse tulemus, antud uurimustöös on 43% (vt Tabel 4), võib olla tingitud asjaolust, et võrreldi omavahel audioklippe ja albumeid, millelt klipid pärinesid. Album võib sisaldada väga erinevat muusikat näiteks, kui tegemist on erinevate artistide muusika kogumikuga ning see tekitab ebatäpsust. Lisaks võis NLP meetodite sooritust halvendada mõningate albumite kirjelduse väike maht (kõige vähima sõnade arvuga albumi kirjeldus oli 3 sõna). Mudelite madala sooritustulemuse taga võib olla ka asjaolu, et mudeleid kasutati nii-öelda *„out of the box"* lahendustena ning uurimistöö käigus ei toimunud mudelite atribuutide peenhäälestamist.

Parimat tulemust muusikasarnasuse leidmisel antud töö kontekstis näitas *ensemble* meetod, seega edasiste uuringute raames võiks antud valdkonna teadustöö käsitleda erinevaid muusika aspekte ja leida vastavate aspektide kaudu muusikasarnasus ja tulemused hiljem koondada *ensemble* meetodiks (vt Joonis 9). Oluline oleks just muusika aspektide ehk lähenemissuundade võimalikult laiapõhjaline esindatus *ensemble* meetodi loomisel.



Joonis 9. Muusikasarnasuse leidmine erinevate aspektide järgi kasutades iga aspekti paremaid tulemusi saavutavat meetodit.

Näiteks saab teadusartiklite põhjal leida, millised on senini parimad tulemusi näitavad ML meetodid erinevate muusikaaspektide järgi muusikasarnasuse leidmisel ja neid omavahel kombineerides. Tegemist oleks samuti eksperimentaalse uurimistööga, mille tulemused võivad anda MIR valdkonnale uusi teadmisi.

5. Kokkuvõte

Uurimuses leiti, et NLP meetoditega saavutatud muusikasarnasus pole üksi piisav, et tulemusi kasutada muusikasoovituste tegemiseks. Tegemist on hinnanguga, mis kehtib ainult antud töö kontekstis ehk muusikaklippe pole mõistlik madala täpsusprotsendi tõttu soovitada tekstilise albumikirjelduste järgi. Audiosignaali põhjal tekkinud muusikasarnasus on oluliselt parema täpsusega ja muudab ainult töös kasutatud NLP meetodite kasutamise üksi mõttetuks. Parim NLP meetod saavutas antud töös täpsuse 45%, samas kui parima audio meetodi täpsus oli 64%. Vaatamata NLP meetodite madalale sooritusvõimele pole NLP meetodite kasutamine soovitussüsteemides asjatu, töös leiti, et NLP ja audio lähenemiste meetodite kombineerimisel on võimalik tõsta muusikasarnasuse leidmise täpsust. Antud töös kasutatud *ensemble* meetod oli 1 protsendipunkti võrra täpsem (65%) kui NLP ja audio meetoditest parim (64%). Töös parimaid tulemusi näidanud meetodi (audio) kombineerimisel madalama täpsusprotsendiga meetodiga (NLP) saavutatud parem tulemus näitab, et kontekst on oluline ja muusikasarnasuse leidmine multimodaalselt parandab sarnasuse leidmise täpsust.

NLP ja audio meetodite üldist sooritusvõimet antud töö kontekstis saab pidada usaldusväärseteks, sest mõlemat lähenemist esindas töös neli meetodit, mis lähenemiste kaupa näitasid sarnaseid tulemusi. *Ensemble* meetodi käigus leitud sooritusvõime paranemine vajab antud kontekstis lisauuringuid, sest *ensemble* meetod koostati antud töös ühekordse katsena ja võrdluseks teisi kombinatsioone ei proovitud. Antud uurimistöö tulemus (40% - 65% täpsus) jääb alla juba antud valdkonnas leitavale täpsusele (90+%). Suure erinevuse põhjuseks võib olla antud töö eksperimentaalsus just muusika klippide sarnasuse leidmisel albumikirjelduste järgi. Seega saab väita, et ainult albumikirjelduste järgi leitud heliklippide sarnasus pole üksi piisav muusika soovitamiseks, madala täpsusprotsendi tõttu. Antud valdkonnas juba tavaliseks kujunenud täpsuseprotsendi tagab eelkõige muusikaklippide omavaheline võrdlemine, antud uuringus kasutatud klippide ja albumikirjelduste võrdlemine lisab sarnasuse leidmise protsessi keerukust ja ebamäärasust ning see kajastub ka uuringutulemustes.

Antud uurimistulemuste suurimaks panuseks antud valdkonda võib pidada kinnitust, et edaspidistes uuringutes peaks keskenduma just muusika erinevate aspektide leidmisele, mis moodustavad muusika konteksti ehk terviku. Kui muusikasarnasust leitakse modaalselt või kitsast konteksti arvestades, tähendab see sarnasuse leidmist, jättes protsessi palju

tundmatuid muutujaid ja see tekitab ebatäpsust. Muusikat peab sarnasuse leidmisel hindama võimalikult laia käsitlust kasutades, et oleks hõlmatud muusika tervikuna.

6. Viidatud kirjandus

[1] K. Gurjar and Y. Moon, "A comparative analysis of music similarity measures in music information retrieval systems", *Journal of Information Processing Systems*, vol. 14, no. 1, pp. 32-55, 2018. https://www.researchgate.net/publication/323704849_A_comparative_analysis_of_music_similarity_measures_in_music_information_retrieval_systems.

[2] S. Oramas, F. Barbieri, O. Nieto and X. Serra, "Multimodal Deep Learning for Music Genre Classification", *Transactions of the International Society for Music Information Retrieval*, vol. 1, no. 1, pp. 4-21, 2018. http://doi.org/10.5334/tismir.10

[3] S. Oramas, O. Nieto, M. Sordo and X. Serra, "A Deep Multimodal Approach for Cold-start Music Recommendation", *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, pp. 32-37, 2017. http://doi.org/10.1145/3125486.3125492

[4] Y. Pandeya and J. Lee, "Deep learning-based late fusion of multimodal information for emotion classification of music video", *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2887-2905, 2020. http://doi.org/10.1007/s11042-020-08836-3

[5] G. Wiggins, "Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music", *2009 11th IEEE International Symposium on Multimedia*, pp. 477-482, 2009. http://doi.org/10.1109/ism.2009.36

[6] B. Sturm, "A Simple Method to Determine if a Music Information Retrieval System is a “Horse”", *IEEE Transactions on Multimedia*, vol. 16, no. 6, pp. 1636-1644, 2014. http://doi.org/10.1109/tmm.2014.2330697

[7] "All models are wrong - Wikipedia", *En.wikipedia.org*, 2022. https://en.wikipedia.org/wiki/All_models_are_wrong. (27.04.2022).

[8] J. Downie, "Music information retrieval", *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 295-340, 2005. http://doi.org/10.1002/aris.1440370108

[9] M. Müller, Y. Özer, M. Krause, T. Prätzlich and J. Driedger, "Sync Toolbox: A Python Package for Efficient, Robust, and Accurate Music Synchronization", *Journal of Open Source Software*, vol. 6, no. 64, p. 3434, 2021. http://doi.org/10.21105/joss.03434

[10] C. Inskip, "Music information retrieval research", *Researchgate*, 2011. https://www.researchgate.net/publication/336676539_Music_information_retrieval_research.

[11] M. Schedl, E. Gómez and J. Urbano, "Music Information Retrieval: Recent Developments and Applications", *Foundations and Trends® in Information Retrieval*, vol. 8, no. 2-3, pp. 127-261, 2014. http://doi.org/10.1561/1500000042

[12] "MP3 - Wikipedia", *En.wikipedia.org*, 2022. https://en.wikipedia.org/wiki/MP3. (22.04.2022).

[13] "Moore's Law: The number of transistors per microprocessor", *Our World in Data*. https://ourworldindata.org/grapher/transistors-per-microprocessor. (22.04.2022).

[14] "Music information retrieval - Wikipedia", *En.wikipedia.org*, 2022. https://en.wikipedia.org/wiki/Music_information_retrieval. (22.04.2022).

[15] M. Won, J. Spijkervet and K. Choi, "Music Classification: Beyond Supervised Learning, Towards Real-world Applications", *Zenodo*, 2021. https://doi.org/10.5281/zenodo.5703780

[16] T. Ingham, "Over 60,000 tracks are now uploaded to Spotify every day. That's nearly one per second. - Music Business Worldwide", *Music Business Worldwide*, 2021. https://www.musicbusinessworldwide.com/over-60000-tracks-are-now-uploaded-to-spotify-daily-thats-nearly-one-per-second/#:~:text=That's%20approximately%20137%20million%20new,according%20to%20United%20Nations%20estimates (22.04.2022).

[17] K. Seyerlehner, "Content-Based Music Recommender Systems: Beyond simple Frame-Level Audio Similarity", Ph.D. dissertation, Johannes Kepler University Linz, 2010. http://www.cp.jku.at/research/papers/Seyerlehner_phd_2010.pdf

[18] L. Hardesty, "The history of Amazon's recommendation algorithm", *Amazon Science*, 2019. https://www.amazon.science/the-history-of-amazons-recommendation-algorithm (22.04.2022).

[19] "What is Natural Language Processing?", *Ibm.com*, 2020. https://www.ibm.com/cloud/learn/natural-language-processing (23.04.2022).

[20] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press, 2014, pp. 21-22. https://www.cs.huji.ac.il/w~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf

[21] R. Řehůřek, "Gensim: topic modelling for humans", *Radimrehurek.com*, 2021. https://radimrehurek.com/gensim/intro.html. (23.04.2022).

[22] Q. Le and T. Mikolov, "Distributed Representations of Sentences and Documents", *arXiv.org*, 2014. https://doi.org/10.48550/arXiv.1405.4053

[23] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv.org*, 2022. https://doi.org/10.48550/arXiv.1810.04805

[24] "TensorFlow Hub", *Tfhub.dev*, 2022. https://tfhub.dev/s?module-type=text-embedding (24.04.2022).

[25] M. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding", *ResearchGate*, 2021. https://www.researchgate.net/publication/350287107_BERT_A_Review_of_Applications_in_Natural_Language_Processing_and_Understanding

[26] J. Pons and X. Serra, "musicnn: Pre-trained convolutional neural networks for music audio tagging", *arXiv.org*, 2019. https://doi.org/10.48550/arXiv.1909.06654

[27] J. Pons, "musicnn/DOCUMENTATION.md at master · jordipons/musicnn", *GitHub*, 2019. https://github.com/jordipons/musicnn/blob/master/DOCUMENTATION.md#models. (24.04.2022).

[28] "similarity", *Dictionary.cambridge.org*, 2022. https://dictionary.cambridge.org/dictionary/english/similarity. (24.04.2022).

[29] "Measuring Similarity from Embeddings | Clustering in Machine Learning | Google Developers", *Google Developers*, 2022. https://developers.google.com/machine-learning/clustering/similarity/measuring-similarity (24.04.2022).

[30] M. Greenacre and R. Primicerio, *Multivariate Analysis of Ecological Data*. Fundacion BBVA, 2014, p. 50. https://www.fbbva.es/wp-content/uploads/2017/05/dat/DE_2013_multivariate.pdf

[31] O. Oduntan, I. Adeyanju, A. Falohun and O. Obe, "A Comparative Analysis of Euclidean Distance and Cosine Similarity Measure for Automated Essay-Type Grading", *Journal of Engineering and Applied Sciences*, vol. 13, no. 11, pp. 4198-4204, 2018. https://www.medwelljournals.com/abstract/?doi=jeasci.2018.4198.4204

[32] G. Mathur, M. Bundele, M. Lalwani and M. Paprzycki, *Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications*. Springer Nature, 2021, p. 170. https://books.google.ee/books?id=JhVfEAAAQBAJ&pg=PA170&dq=cosine+similarity&hl=en&sa=X&ved=2ahUKEwisp5Dj6K73AhU-Yiv0HHR6GC4YQ6AF6BAgHEAI#v=onepage&q=cosine%20similarity&f=false

[33] S. Xia, Z. Xiong, Y. Luo, WeiXu and G. Zhang, "Effectiveness of the Euclidean distance in high dimensional spaces", *Optik*, vol. 126, no. 24, pp. 5614-5619, 2015. http://doi.org/10.1016/j.ijleo.2015.09.093

[34] A. Shirkhorshidi, S. Aghabozorgi and T. Wah, "A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data", *PLOS ONE*, vol. 10, no. 12, p. 6, 2015. http://doi.org/10.1371/journal.pone.0144059

[35] N. Gali and V. Tiwari, "Speech and Lyric-based Doc2Vec Music Recommendation System", *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 8, pp. 67-70, 2021. https://www.ijert.org/speech-and-lyric-based-doc2vec-music-recommendation-system.

[36] R. Cunha, E. Caldeira and L. Fujii, *Determining Song Similarity via Machine Learning Techniques and Tagging Information*. 2017, p. 5. https://doi.org/10.48550/arXiv.1704.03844

[37] M. Fell, "Natural language processing for music information retrieval : deep analysis of lyrics structure and content", Université Côte d'Azur, 2021. https://tel.archives-ouvertes.fr/tel-03135082/file/2020COAZ4017.pdf

[38] Z. Fu, G. Lu, K. Ting and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation", *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 303-319, 2011. http://doi.org/10.1109/tmm.2010.2098858

[39] M. Kelly, K. Malloy, M. Moelgaard, R. Mannem and A. Rösti, *CS 274C Project: An Exploration of BERT for Song Classification and Recommendation*. 2021, p. 4. https://kaimalloy.com/172B_Project.pdf

[40] D. Chathuranga and L. Jayaratne, "Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches", *GSTF Journal on Computing (JoC)*, vol. 3, no. 2, 2013. https://www.researchgate.net/publication/313557774_Automatic_Music_Genre_Classification_of_Audio_Signals_with_Machine_Learning_Approaches

[41] R. Mayer and A. Rauber, "Music Genre Classification by Ensembles of Audio and Lyrics Features", in *12th International Society for Music Information Retrieval Conference, ISMIR 2011*, Miami, 2011, pp. 678-680. https://www.researchgate.net/publication/220723387_Music_Genre_Classification_by_Ensembles_of_Audio_and_Lyrics_Features

[42] E. Law, K. West, M. Mandel, M. Bay and J. Downie, "EVALUATION OF ALGORITHMS USING GAMES: THE CASE OF MUSIC TAGGING", in *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009, pp. 387-392. https://archives.ismir.net/ismir2009/paper/000019.pdf

[43] "The MagnaTagATune Dataset | City University MIRG", *Mirg.city.ac.uk*, 2013. https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset. (28.04.2022).

[44] E. Law and L. von Ahn, "Input-agreement: a new mechanism for collecting data using human computation games", in *CHI '09: CHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, 2009, pp. 1197-1206. https://www.cs.cmu.edu/~elaw/papers/tagatune.pdf

[45] "TensorFlow Hub", *Tfhub.dev*. https://tfhub.dev/google/collections/experts/bert/1. (29.04.2022).

[46] "TensorFlow Hub", *Tfhub.dev*, 2022. https://tfhub.dev/google/experts/bert/wiki_books/2. (29.04.2022).

[47] J. Pons, "musicnn/musicnn_example.ipynb at master · jordipons/musicnn", *GitHub*, 2019. https://github.com/jordipons/musicnn/blob/master/musicnn_example.ipynb. (30.04.2022).

[48] Z. Zhou, *Ensemble Methods Foundations and Algorithms*. Boca Ration: CRC Press, 2012, pp. 15, 16, 70. https://tjzhifei.github.io/links/EMFA.pdf

Lisad

I. Uurimistöö struktuur

Aspekt	Metoodika			Tulemused						
	Meetodid/andmestikud	Mudelid		Tulemus	%	Tulemuste võrdlus				
Ground Truth	MagnaTune andmestik, Web Scraper	TagATune Odd-One-Out		122	100%	122				
NLP	D2V	PV-DM		53	43%	PV-DBOW	PV-DBOW	Ensemble		
		PV-DBOW		55	45%			Kaal	Tulemus	%
	BERT	WIKI		52	43%	WIKI		.41	79	65%
		WIKI/qnli		49	40%					
AUDIO	MUSICNN	MTT	TAGS	78	64%	MTT TAGS	MSD PENULTIMATE	.59		
			PENULTIMATE	69	57%					
		MSD	TAGS	68	56%	MSD PENULTIMATE				
			PENULTIMATE	71	58%					

1 Andmestiku valik ja andmete töötlemine
2 NLP meetoditega andmete analüüs
3 Audio meetoditega andmete analüüs
4 Ensemble meetodi loomine ja andmete analüüs

II. Uurimuse struktuur jaotatuna 4 mooduliks

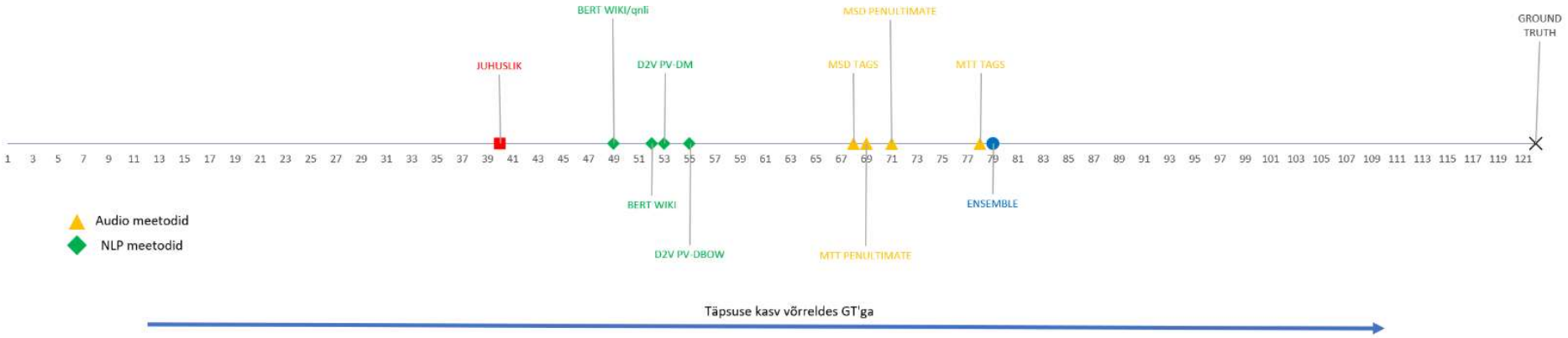
Aspekt	Metoodika			Tulemused						
	Meetodid/andmestikud	Mudelid		Tulemus	%	Tulemuste võrdlus				
Ground Truth	MagnaTune andmestik, Web Scraper	TagATune Odd-One-Out		122	100%	122				
NLP	D2V	PV-DM		53	43%	PV-DBOW	PV-DBOW	Ensemble		
		PV-DBOW		55	45%			Kaal	Tulemus	%
	BERT	WIKI		52	43%	WIKI		.41	79	65%
		WIKI/qnli		49	40%					
AUDIO	MUSICNN	MTT	TAGS	78	64%	MTT TAGS	MSD PENULTIMATE	.59		
			PENULTIMATE	69	57%					
		MSD	TAGS	68	56%	MSD PENULTIMATE				
			PENULTIMATE	71	58%					

1 moodul: Võrdluseks vajalike andmete kogumine ja sobivas formaadis "tõe etaloni" loomine.
2 moodul: Uurimuses kasutatavate muusika sarnasuse meetodite rakendamine
3 moodul: Tulemuste kirjeldamine ja võrdlus
4 moodul: Parimaid tulemusi näidanud mudelite koondamine ensemble meetodiks, rakendamine ja tulemuste kirjeldamine

III. NLP, audio ja *ensemble* meetodite omavahelise sarnasuse tulemused

	Ground truth (GT)	D2V PV-DM	D2V PV-DBOW	BERT	BERT/qnli	MTT tags	MDS tags	MTT penultimate	MDS penultimate	D2V & MTT tags ensemble
Ground truth (GT)	122	53	55	52	49	78	68	69	71	79
D2V PV-DM	53	122	74	58	55	56	55	57	45	63
D2V PV-DBOW	**55**	74	122	48	59	60	51	59	49	73
BERT	52	58	48	122	74	47	49	45	44	53
BERT/qnli	49	55	59	74	122	54	49	44	39	63
MTT tags	**78**	56	**60**	47	54	122	69	88	66	102
MDS tags	68	55	51	49	49	69	122	70	87	70
MTT penultimate	69	57	59	45	44	88	70	122	73	92
MDS penultimate	71	45	49	44	39	66	87	73	122	68
D2V & MTT tags ensemble	**79**	63	73	53	63	102	70	92	68	122

IV. Kolmikute sarnasus meetodite järgi



V. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Mait Lindpere,

1. Annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Loomuliku keele töötlemise algoritmide kasutamine muusikasarnasuse leidmisel“ mille juhendaja on Anna Aljanaki (PhD) reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Mait Lindpere

16.05.2022