

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Richard Nõmm

**Ülevaade Microsoft Power BI-st ja õppematerjali
koostamine**

Bakalaureusetöö (9 EAP)

Juhendaja: MSc Vambola Leping

Tartu 2020

Ülevaade Microsoft Power BI-st ja õppematerjali koostamine

Lühikokkuvõte:

Selles töös kirjeldatakse andmete analüüsimise programmi Microsoft Power BI ja koostatakse selle kohta õppematerjal. Antakse ülevaade programmi kujunemise ajaloost ja kirjeldatakse võimalusi andmete sissetoomiseks, hoidmiseks ja visualiseerimiseks. Lisaks sellele kirjeldatakse õppematerjali koostamist, mis tutvustab programmi Power BI Desktop.

Võtmesõnad:

Microsoft Power BI, Microsoft Power BI Desktop, Microsoft Power BI Service.

CERCS:

P175 Informaatika, süsteemiteooria.

Overview of Microsoft Power BI and composition of study material

Abstract:

This paper describes the data analysis program Microsoft Power BI and a study material is composed. An overview of its history is given and possibilities of data gathering, storing and visualizing are described. Also the development of the study material, which introduces Power BI Desktop is described.

Keywords:

Microsoft Power BI, Microsoft Power BI Desktop, Microsoft Power BI Service.

CERCS:

P175 Informatics, systems theory.

Sisukord

Sissejuhatus	4
Probleemi tutvustus	4
Bakalaureusetöö eesmärk ja teema olulisus	5
Töö ülevaade	5
Metoodika	5
1 Ülevaade Microsoft Power BI-st	6
1.1 Ajalugu	6
1.2 Power BI Desktopi lühitutvustus	7
1.3 Andmete toomine	7
1.4 Andmete hoidmine	9
1.5 Visualiseerimine	10
1.6 Power BI Service	12
1.7 Power BI ja tehisintellekt	12
1.7.1 K ja V	12
1.7.2 Kiirülevaated	13
1.7.3 Peamised mõjutajad	14
2 Õppematerjali loomine	16
2.1 ADDIE mudel	16
2.2 Analüüs	17
2.3 Kavandamine	17
2.4 Materjali väljatöötamine	18
2.5 Materjali kasutamine	19
2.6 Hindamine	25
3 Edasine arendus	26
4 Kokkuvõte	27
5 Viidatud kirjandus	28
Lisad	30

Sissejuhatus

Andmete visualiseerimine ja andmeanalüüs on muutumas väga olulisteks teemadeks ning andmeid kogutakse peaaegu igas valdkonnas. NewVantage Partners-i uuringus selgus, et 97,2% firmadest tegelevad andmete analüüsimisega, kusjuures suurema osa uuritavatest moodustasid finants- ja meditsiiniettevõtted [1]. Kogutud andmete analüüsimine on lisaks IT-teenustele oluline ka teistele tegevussfääridele, sest tänu andmeanalüüsile on võimalik teenuseid efektiivsemaks muuta.

2017. aasta IBM-i uuringust selgus, et kahe eelneva aastaga kogutud andmed moodustavad kõikidest andmetest 90% [2]. Gantzi ja Reisneli sõnul muutuvad maailmas kogutud andmed alates 2005. aastast kuni 2020. aastani 300 korda suuremaks, jõudes 40 triljoni gigabaidini [3]. Kuna kogutavate andmete mahud lähevad aina suuremaks, on vaja kasutada programme, mis saavad andmete visualiseerimise ja modelleerimisega hakkama.

Tartu Ülikoolis loetakse õppeaineid, mis selliseid programme tutvustavad. Arvutiteaduse instituudis on kursus "MTAT.03.010 Arvutikäsitusõpetus¹", mis sisaldab tabelarvutusprogrammide õpetamist. Autori kogemusest lähtuvalt ei pruugi aegunud programmid aina suuremate andmetega hakkama saada.

Microsoft on üks maailma suurimaid tehnoloogiaarenduse ettevõtteid, mis on läbi aegade tootnud palju populaarset tarkvara. Microsoft Power BI programmi peetakse heaks tööriistaks andmete visualiseerimiseks ja analüüsimiseks. Selle arendus erineb suuresti tuntud tabelarvutussüsteemist Excel, millel see põhineb.

Microsoft Power BI Desktop on tasuta kättesaadav kõikidele Tartu Ülikooli tudengitele. Samuti on võimalik kõik failid üles laadida Power BI veebikeskkonda, kus saab neid vaadata, jagada ja kommenteerida. Keskkonnas saab jagada tudengid gruppidesse, mis aitab praktikumi korraldajal tööd organiseerida.

Probleemi tutvustus

Töö autor ei leidnud töö koostamise ajal eestikeelset õppematerjali, mis tutvustaks Microsoft Power BI programmi. Leidub üksikuid tasulisi koolitusi, ent kõrge hinna tõttu ei ole see

¹ <https://ois2.ut.ee/#/courses/MTAT.03.010/details>

kõigile kättesaadav. Samuti pole Tartu Ülikoolis vastavaid õppematerjale ja autori hinnangul on programm nii õppejõudude kui ka tudengite jaoks üsna tundmatu.

Bakalaureusetöö eesmärk ja teema olulisus

Eesmärk on kirjeldada, missuguseid võimalusi pakub Power BI Desktop ning koostada õppematerjal, mis programmi tutvustab. Kirjeldatud on Power BI programmi 2020. aasta kevade seisuga, seega hilisema lugemise ajal võib olla tehtud programmis muudatusi.

Töö ülevaade

Käesoleva töö esimeses peatükis antakse ülevaade Power BI Desktopist. Esiteks tutvustatakse programmi ajalugu ja kujunemist. Teiseks kirjeldatakse programmi võimalusi ja antakse ülevaade tehisintellekti kasutatavatest tööriistadest.

Töö teine peatükk kirjeldab lähemalt õppematerjali loomist ja testimist. Esiteks tutvustatakse õppematerjali koostamiseks kasutatud mudelit ja hiljem antakse ülevaade õppematerjali koostamisest etappide kaupa. Samuti analüüsitakse õppematerjali testimise raames saadud tulemusi.

Töö kolmandas peatükis pakutakse välja ideid, kuidas saaks programmi õpetamist ja töö käigus loodud õppematerjali edasi arendada.

Metoodika

Bakalaureusetöö puhul on tegemist rakendusliku uurimusega ja selle raames koostatakse informaatikateemalisele ainele õppematerjal koos lahendustega. Töö koosneb nii teoreetilisest kui ka praktilisest osast.

Töö teoreetilises osas uuritakse Power BI programmi võimalusi ning selleks vajaliku info leidmiseks kasutatakse erinevaid kirjandusallikaid ja andmebaase.

Bakalaureusetöö praktilises osas koostatakse õppematerjal, mis tutvustab Power BI programmi ning annab ülevaate andmete analüüsimisest ja visuaalide loomisest. Kirjeldatakse õppematerjali valmimise protsessi ja õppematerjali sobivuse hindamiseks viiakse läbi näidispraktikum. Viimasest tagasiside saamiseks viiakse läbi küsitlus.

1 Ülevaade Microsoft Power BI-st

Järgnevides alapeatükkides kirjeldatakse Microsoft Power BI programmi. Esiteks kirjeldatakse programmi ajalugu ja hiljem tutvustatakse Power BI võimalusi. Antakse ülevaade Power BI funktsionaalsusest, keskendudes andmete sissetoomisele, hoidmisele ja visualiseerimisele. Samuti kirjeldatakse programmis kasutatavaid tehisintellekti funktsionaalsuseid.

1.1 Ajalugu

Aasta 2010 novembris toimus Seattle'is konverents nimega PASS Summit 2010. Bruckner [4] kirjutab Microsofti blogis, et konverents oli mõeldud Microsoft SQL Serveri professionaalidele ja seal tutvustati uut projekti nimega Crescent. Viimase loojateks võib pidada Thierry D'Hersi ja Amir Netzi, kes olid Microsoft SQL Server Reporting Services-i tiimi liikmed. Crescenti disainis Ron George 2010. aastal ja järgmine aasta toimus tarkvara väljalase [4].

Unkroth [5] kirjutas Microsofti blogis, et Crescent-i võis avalikult alla laadida alates 2011. aasta juulist ja sellega tuli kaasa SQL Server nimega Denali, mis oli edasimineku andmebaasi poolelt. Crescent tõi lisaks juurde uut funktsionaalsust võrreldes sellega, mis oli varem tehtud Excelis PowerPivot-iga. Crescentiga sai modelleeritud andmeid efektselt visualiseerida kasutades interaktiivseid visuaale ja animatsioone. Crescent oli mugavalt kasutatav ja palju funktsionaalsust lahendati pukseerimisega (ingl *drag and drop*), mis polnud samal ajal avaldatud Excelis eriti levinud [5]. Hiljem sai programm nimeks Power BI.

Albanesius [6] kirjutas ajakirjas PCMag ilmunud artiklis, et 2013. aastal teatas Microsoft uuendusest Office 365 tarkvarale, mille käigus võeti kasutusele Power BI. Viimati nimetatud programmi eelvaade koosnes Excelist ja samuti lisati visualiseerimisega seotud uuendusi. Ühe tähtsaima uuendusena loodud Power Map funktsioon võimaldas luua piirkonnapõhiseid visuaale, millega sai esile tõsta ZIP-koode, regioone, riike, maakondi ja linnu. Lisaks sellele räägiti edusammudest keelelistes päringutes. Microsofti sõnul suutis programm välja pakkuda interaktiivseid visuaale, lähtudes dialoogikasti kirjutatud äriteabe teema küsimustest [6].

Microsoft Power BI blogis [7] kirjutati, et Power BI Desktop avalikustati aastal 2014 koos Power BI teenusega, mis võimaldas loodud visuaale ja raporteid keskkonnas jagada. Selle abil sai töötada ühe failiga mitu inimest. Lisati funktsionaalsust koos töötavate inimeste

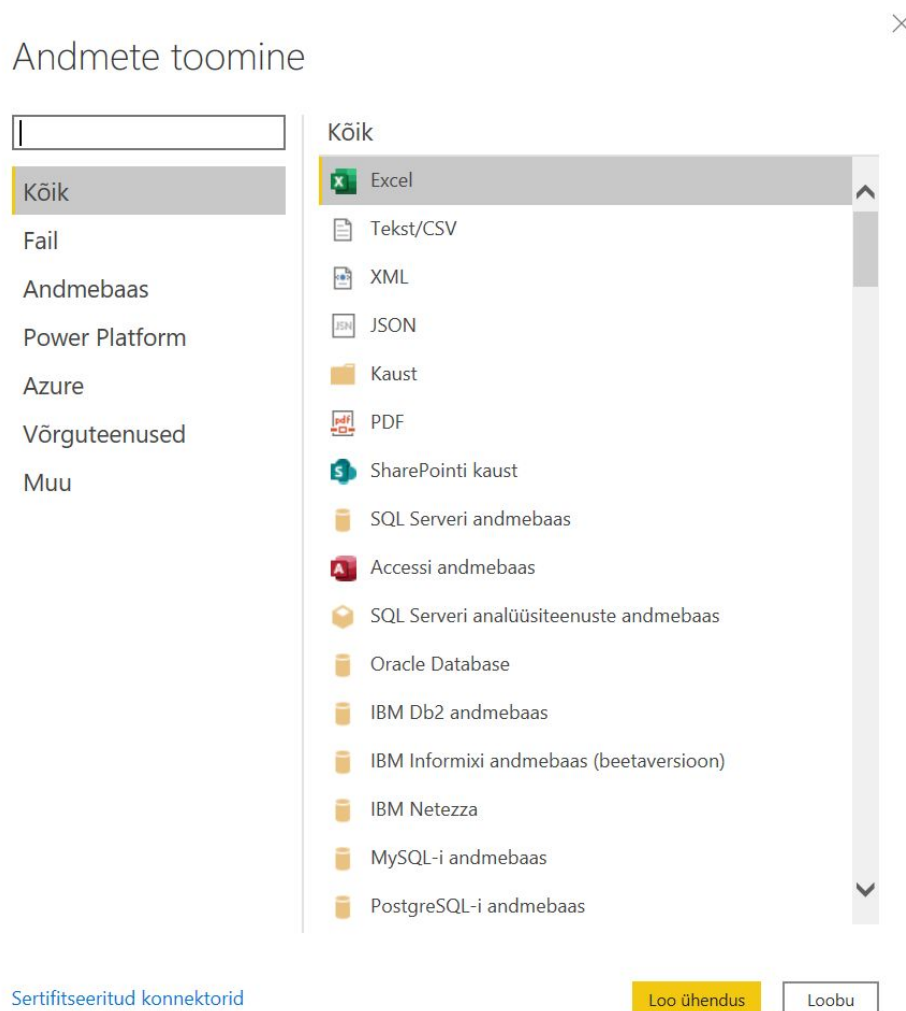
gruppidele, mis soodustas sisu jagamist grupisisest. Power BI-le arendati visualiseerimise funktsioone, näiteks lisati uusi diagramme ja värvi formaate. Lisaks sellele muudeti visualiseerimise raamistikku. Kui varem sõltusid kasutajad tarkvara loojate visuaalidest, siis uuendusena muudeti visualiseerimise raamistiku lähtekood avalikuks. Seega said nii kasutajad kui ka programmeerijad GitHub-i vahendusel luua ja jagada omavahel rohkem visuaale kui varem. Vastava projekti võib leida aadressilt <https://github.com/Microsoft/PowerBI-visuals>. Samuti lisati otseühendus SQL Server Analysis Services-i tabelikujulistele mudelitele, mis võimaldas teha reaalaja päringuid ja ligipääsu suurandmetele, kasutades Apache Spark-i avatud lähtekoodiga raamistikku. Uut Power BI teenust oli edaspidi võimalik ka veebipõhises süsteemis kasutada. Samuti parandati ühilduvust Exceliga, mis võimaldas importida Exceli töövihikud ja vaadata nende andmetega tehtud Power BI juhtpaneel Excel Online-is [7].

1.2 Power BI Desktopi lühitutvustus

Aspini [8] sõnul on Power BI pilveteenus, mille abil on andmeid analüüsides võimalik koostada ja jagada raporteid (ingl *report*) ja juhtpaneel (ingl *dashboard*). Larsoni ja Corley [9] arvates ühendab Power BI Desktop endas mootorit VertiPaq, tööriista Power Pivot ja visualiseerimise tööriista, mis oli varem tuntud nimega Power View. Raportite jagamiseks on tugi Windowsi, iOS-i ja Androidi seadmetele [8]. Power BI Desktopis loodud sisu on võimalik üles laadida ja jagada kasutades veebiteenust Power BI Service, mis on leitav aadressilt www.powerbi.com [9].

1.3 Andmete toomine

Aspini [8] ja Powelli [10] sõnul on Power BI Desktop hea konnektor (ingl *connector*) andmete sissetoomiseks, sest programmi saab ühendada paljude andmebaaside ja andmehoidlatega. Järgnevas lõigus on kirjeldatud erinevaid võimalusi andmete impordiks Power BI Desktop-is. Kirjeldatud on autori arvates enimkasutatavaid konnektoreid. Nii töö autori kui ka Aspini [8] arvates toimub programmis pidev andmeallikate uuendamine, mille käigus lisatakse uusi konnektoreid. Sellel põhjusel võivad järgnevas lõigus kirjeldatud andmeallikad erineda töö lugemise ajal olemasolevatest.



Joonis 1. Andmete toomine.

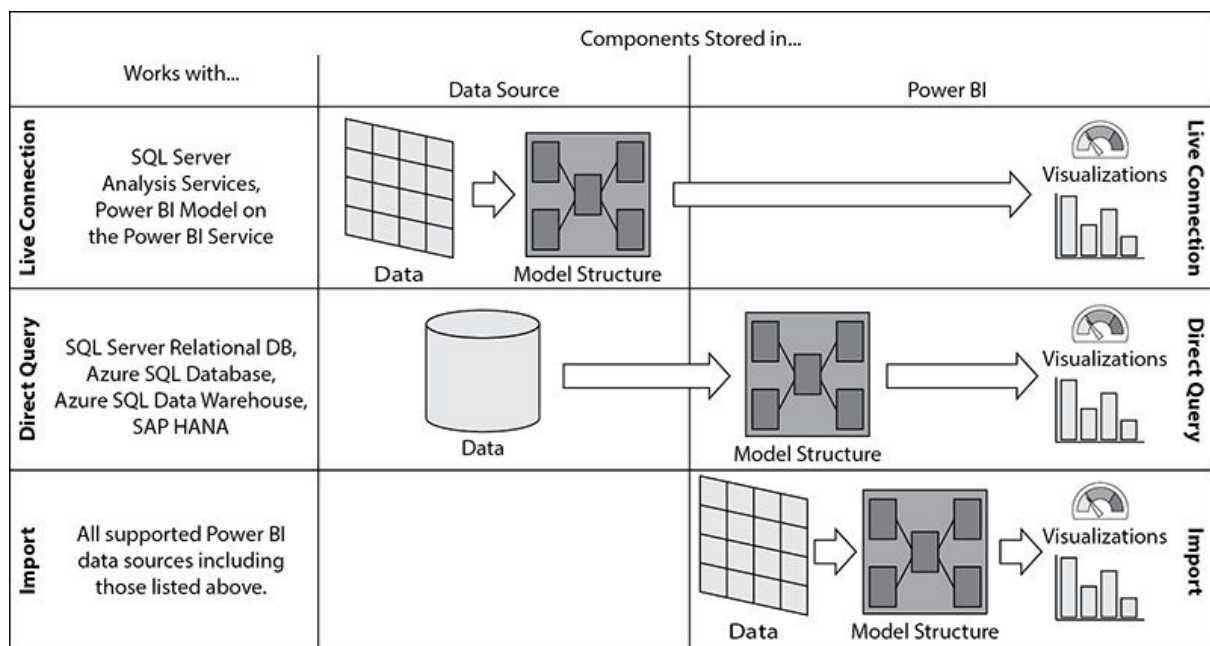
Joonisel 1 on välja toodud erinevaid konektoreid, mille abil andmeid Power BI Desktopi sisse tuua. Ühenduse saab luua levinud andmefailidega näiteks Exceli, CSV-, PDF- ja muude erinevate tekstifailidega. Aspini [8] sõnul on lisatud tugi paljudele enim kasutatavatele andmebaasisüsteemidele näiteks Oracle-i, Azure-i, SQL Serveri ja Accessi andmebaasidele. Samuti on võimalik luua ühendus erinevate võrguteenustega näiteks Google Analytics-iga või andmeimport otse veebist. Lisaks eelnevale on võimalik andmeid tuua Pythoni ja R-i skriptide abil.

Andmete sissetoomiseks kasutatakse võimsat funktsionaalkeelt ja päringumootorit M [10]. Iga kord kui andmeallika objektiga muutuseid tehakse, luuakse M-avaldisi (ingl *expression*). Need avaldised kirjutatakse vastavatesse dokumentidesse ja salvestatakse Power BI Desktopi faili päringud (ingl *queries*).

Powell [10] kirjutab, et pärast andmete sissetoomist kasutatakse nende talletamiseks salvestusmootorit (ingl *storage engine*) VertiPaq. Mälu kasutamise vähendamiseks ja jõudluse suurendamiseks kasutab VertiPaq kompressimisalgoritmi (ingl *compression algorithm*). Selle algoritmi abil salvestatakse unikaalsed väärtused ja pärast sarikodeerimist (ingl *run-length encoding*) salvestatakse korduvad väärtused ühekordselt. Selle tulemusena on pakitud korduvad väärtused näiteks kuude nimetused. Unikaalseid väärtuseid, näiteks primaarvõtmeid ei pakita.

1.4 Andmete hoidmine

Larsoni ja Corley sõnul [9] hoiustatakse andmeid erinevalt seoses nende tüübiga ja nendeks tüüpideks on: reaalaja ühendus (ingl *Live Connection*), otsene päring (ingl *Direct Query*) ja sisse toodud andmed (ingl *Import*).



Joonis 2. Andmete hoidmine Power BI programmis [9].

Joonisel 2 on kirjeldatud kolme erineva andmetüübi hoidmist Power BI süsteemis. Järgnevas lõikudes kirjeldatakse neid kolme.

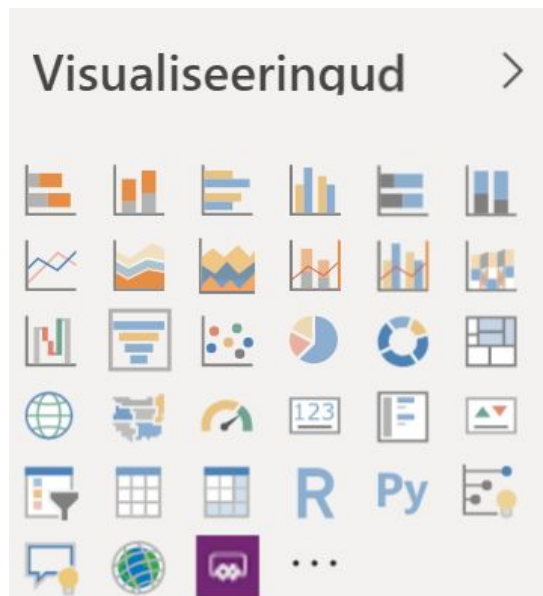
Reaalaja ühendust kasutatakse näiteks teenustega SQL Server Analysis Services ja Azure Analysis Services. Reaalaja ühenduse käigus ei salvestata andmeid Power BI mudelisse. Vajalike andmete saamiseks kasutatakse päringuid ja selle abil saadud andmeid saab Power BI programmis analüüsida ja visualiseerida [9].

Sisse toodud andmete hoidmine toimub täielikult Power BI siseselt ehk PBIX-failis. Selle üks suurimaid eeliseid on andmete analüüsimise suurem kiirus võrreldes reaalaaja ühenduse või otsese päringuga. Puudujäägiks võib pidada aeglast andmete uuendamist, sest uuenduse käigus tuuakse sisse kogu andmestik. Andmeid värskendatakse kaheksa korda päevas ja selle tulemusena võivad kasutatavad andmed aeguda. Lisaks sellele lülitatakse andmete värskendamiseks mudel välja ehk andmeid ei saa sel ajal kasutada [9].

Otsest päringut kasutatakse näiteks SQL Serveri ja Azure SQL andmebaasidega. Otsese päringu funktsionaalsus jääb eelnimetatud kahe võimaluse vahele. Sarnaselt reaalaaja ühendusega jäetakse andmed allikasse ehk neid ei laeta programmi sisse. Päringute abil saadakse analüüsitavad andmed reaalaajas. Samas võib see suurte andmemahutudega korral olla aeglane tegevus. Sarnaselt sisse toodud andmetega säilitatakse mudeli struktuur programmi siseselt, mis võimaldab mudeli struktuuri muuta [9].

1.5 Visualiseerimine

Larsoni [9] sõnul kasutatakse Power BI programmis visualiseerimiseks tööriista, mida tunti varem Power View nime all. Power View pärineb programmist Microsoft SharePoint ja seda on kasutatud ka Excelis. Power View abil on võimalik luua interaktiivseid esipaneele ja raporteid. Powelli [11] sõnul on oluline aru saada raporti ja esipaneeli erinevusest. Esipaneele kasutatakse tavaliselt tervikliku pildi kuvamiseks, ent raportis keskendutakse spetsiifilistele küsimustele. Lisaks sellele kasutatakse raportites analüütilisi võimalusi, näiteks tükeldi visuaali. Selle visuaali kasutamine võimaldab analüüsida soovitud infot, näiteks saab andmeid filtreerida.



Joonis 3. Power BI Desktopi visuaalid.

Joonisel 3 on kujutatud visualiseeringute paan, kus on 27 erinevat visuaali. Nende hulgas on võimalus kirjutada visuaali skripte programmeerimiskeeltes R ja Python. Lisaks standardsetele visuaalidele on võimalik leida visuaale AppSource-i abil. Selle kaudu avaneb keskkond nimega turuplats (ingl *marketplace*), kust on võimalik leida lisandmoodulitena allalaaditavaid visuaale. Neid visuaale pole Microsoft teinud, vaid on mõni kasutaja üles laadinud. Microsoft on visuaalid turvalisuse kaalutlustel üle vaadanud.

Powelli [11] sõnul on efektiivse raporti juures oluline, et see oleks visuaalselt meeldiv ja loogiliselt struktureeritud. Iga visuaal peaks olema kooskõlas raporti eesmärgiga ehk vastama mingile püstitatud küsimusele. Järgnevad Powelli [11] soovitusel raporti koostamiseks:

- Võimalusel eemaldada kõik visuaali elemendid, mis ei ole olulised konkreetse visuaali mõistmiseks. Ebaoluliseks võivad osutada näiteks legend, tekstiväli või telje nimed.
- Visuaalidel peaksid olema lihtsad ja selged pealkirjad.
- Kõige tähtsamad visuaalid peaksid olema üleval vasakus nurgas.
- Raporti taustaks sobib neutraalne värv, näiteks valge või helehall.
- Raportis ei tohiks olla suuri tühimikke ehk kogu kanvaa (ingl *canvas*) peaks olema täidetud [11].

Powell [11] pakub välja lihtsa võimaluse Power BI raporti testimiseks: sülearvuti ekraanilt raporti vaatamine umbes nelja meetri kauguselt. Kui sellelt kauguselt nähtavate visuaalide ja

kujutiste abil on võimalik tuletada raporti sisu, siis võib raportit pidada visuaalselt efektiivseks [11].

1.6 Power BI Service

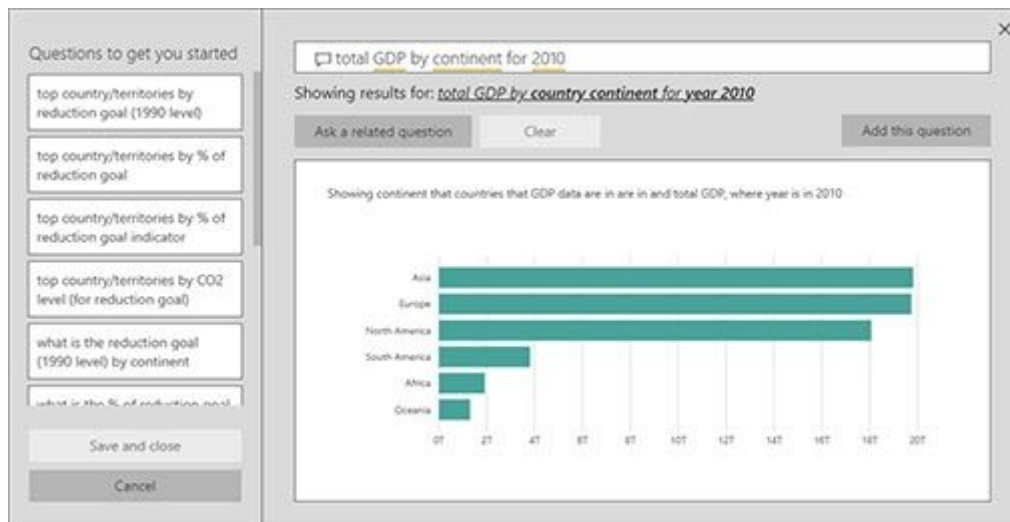
Larseni [9] sõnul kasutatakse Power BI Desktopi abil loodud sisu avaldamiseks ja jagamiseks veebikeskkonda Power BI Service. Bakalaureusetöö koostamise ajal oli eelnimetatud keskkonnas failide jagamiseks vajalik Power BI Pro litsents, millele pakuti tasuta prooviperioodi. Veebikeskkonda üles laetud Power BI failid salvestatakse kausta tööruumid (ingl *workspaces*). Vaikeseadistusena salvestatakse fail kausta minu tööruum, aga on võimalik luua ka oma tööruum. Autori arvates võib see olla vajalik, kui keskkonda laetakse üles erinevaid faile, mille paremaks haldamiseks saab neid grupeerida. Tööruumi üleslaetud faile on võimalik jagada ka organisatsiooni siseselt. Tartu Ülikooli töötajad ja tudengid saavad jagada faile kõikidele @ut.ee meiliaadressi kasutajatele. Vastava kasutajaga jagatud failid on leitavad kaustast minuga ühiskasutusse antud (ingl *shared with me*). Autori arvates võib see kaust olla oluline õppejõududele, sest seal on võimalik hõlpsasti õpilaste avaldatud raporteid kommenteerida ja neile tagasisidet anda.

1.7 Power BI ja tehisintellekt

Crittendeni [12] sõnul arendab Microsoft pidevalt tehisintellekti võimalusi Power BI programmis. Nendeks võimalusteks on tema arvates K ja V (küsimused ja vastused, ingl *Q&A*), kiirülevaated (ingl *Quick Insights*) ja peamised mõjutajad (ingl *Key Influencers*). Järgnevad alapeatükid kirjeldavad eelnevalt nimetatud võimalusi.

1.7.1 K ja V

K ja V on tööriist, mille abil on võimalik esitada oma andmete kohta küsimusi, kasutades selleks loomulikku keelt (ingl *natural language*).



Joonis 4. K ja V kasutamine [9].

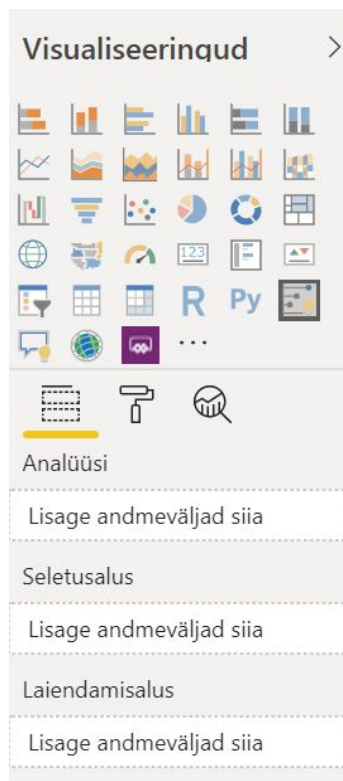
Joonisel 4 on kujutatud näidet tööriista K ja V kasutamisest. Küsimused võivad olla seotud erinevate tunnuste vaheliste seostega või näiteks tunnuse kohta käivate arvutustega. Aspini [8] arvates töötab see hästi siis, kui raportis kasutatud andmete veerupealkirjad on sõnastatud arusaadavalt. See on oluline, sest K ja V kasutamiseks tuleb vastavasse aknasse kirjutada küsimus, milles soovitakse teatud andmeveeru või tunnuse kohta infot saada. Joonisel 4 on küsimuse koostamiseks kasutatud veerupealkirjadeks GDP ja *country* ehk riik. Tööriist aitab küsimuste koostamisel, pakkudes ise välja küsimusi. Joonisel 4 näeb välja pakutud küsimusi vasakus veerus. Küsimusest aru saamiseks kasutab programm loomuliku keele töötlemist (ingl *natural language processing*). Küsimuse vastuseks väljastab programm visuaali, joonisel 4 esitatud näites on selleks lintdiagramm.

1.7.2 Kiirülevaated

Powelli [11] sõnul on kiirülevaated üks kõige arenenumatest tööriistadest Power BI programmis, sest selleks kasutatakse vastava andmestiku peal keerukaid masinõppe algoritme. Andmestik jagatakse väiksemateks ja spetsiifilistemateks alamhulkadeks, et leida korrelatsioone ja trende. Algoritmide abil leitakse andmetes seoseid ja selle tulemusena väljastatakse kõrgel tasemel vormistatud interaktiivseid visuaale. Tööriista võimekust tõestab asjaolu, et seda saab rakendada terve andmestiku peal. Selleks tuleb andmestik enne Power BI Service keskkonda üles laadida.

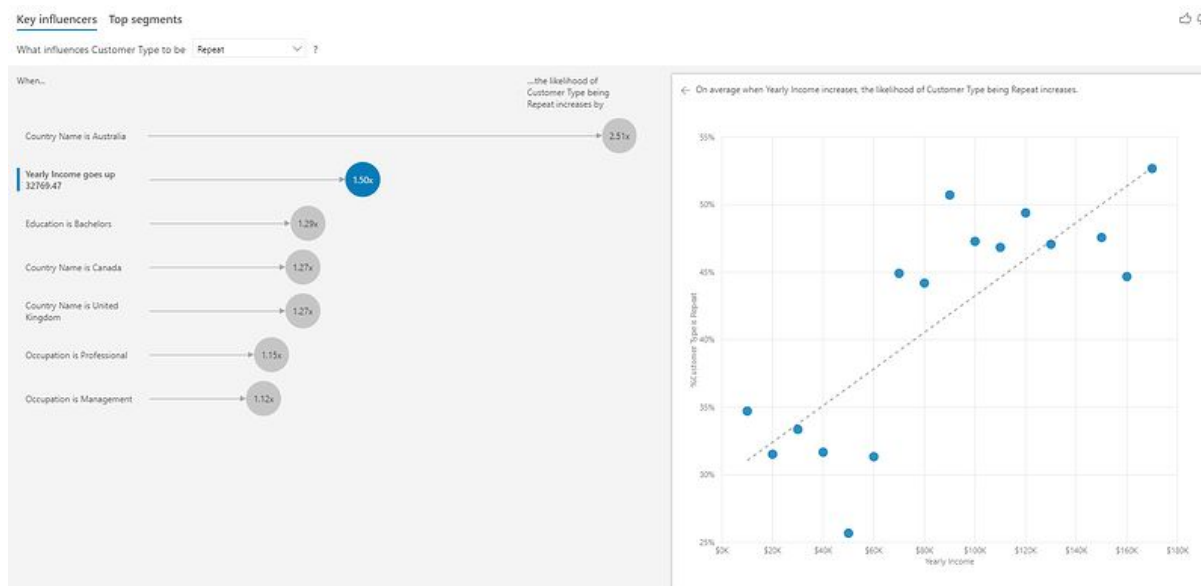
1.7.3 Peamised mõjutajad

Crittendeni [12] sõnul on peamised mõjutajad oluline visuaal, sest see annab võimaluse ise masinõpet kasutada, et andmestikust uuritavale küsimusele vastust leida. Visuaal on interaktiivne ehk tulemuste muutmiseks saab kasutada erinevaid raportis olevaid filtreid, tükkeldeid ja teisi visuaale.



Joonis 5. Visuaali Peamised mõjutajad loomine.

Joonisel 5 on kujutatud visuaali peamised mõjutajad loomist. Analüüsi atribuuti tuleb lisada andmeväli või tunnus, mille kohta soovitakse infot saada. Seletusalus on atribuut, kuhu tuleb lisada väljad, mis võivad mõjutada otsitavat ehk analüüsi lahtrisse lisatud tunnust. Laiendamisaluses saab täpsustada, millisel tasemel mõõte hinnatakse, kui otsitav tunnus on summeeritud väli või mõõt. Laiendamisalusesse lisatud välju ei peeta kunagi atribuudi “Analüüsi” mõjutajateks.



Joonis 6. Peamised mõjutajad visuaal [12].

Joonisel 6 on kujutatud Crittendeni [12] koostatud visuaali tulemust. Visuaal koosneb kahest paanist: peamised mõjutajad (visuaali nimi on sama) ja ülemised segmendid (joonisel vastavalt key influencers ja top segments).

Esimesel paanil uuritakse analüüsi atribuudis täpsustatud tunnust. Joonisel 6 kirjeldatud näites on selleks kliendi tüüp (ingl *customer type*), mis on kategoriseeritud gruppi ühekordne (ingl *one-time*) või korduv (ingl *repeat*). Otsitakse, mis tunnused mõjutavad seda, et kliendi tüüp oleks korduv. Kasutades logistilist regressiooni analüüsib programm andmestikku ja pakub välja aspekte, mis võivad kliendi tüüpi mõjutada. Joonisel 6 on need listina kujutatud vasakul pool ja paremale poole on koostatud illustreeriv joonis, mis iseloomustab aastase sissetuleku suurust ja kliendi tüüpi [12].

Teisel ehk paanil ülemised segmendid on kirjeldatud klasterdatud väärtused, kui valitud analüüsitava tunnuse tõenäosus on suur või väike. Klasterdatud väärtused on grupeeritud segmentidesse ja iga segmenti kohta on võimalik saada lisainfot [13].

2 Õppematerjali loomine

Järgnevat peatükkides on kirjeldatud bakalaureusetöö raames loodud õppematerjali koostamist ja testimist. Esiteks tutvustatakse materjali koostamiseks kasutatud mudelit ja hiljem kirjeldatakse õppematerjali loomist etappide kaupa. Antakse ülevaade loodud ülesannetest, analüüsitakse testimist ja sealt saadud tagasisidet. Töö raames koostatud materjalid on leitavad Google Drive-i kaustast (lisa 1). Pärast bakalaureusetöö kaitsmist on materjalidele juurdepääs lubatud ainult töö autori nõusolekuga.

2.1 ADDIE mudel

Õppematerjali loomisel on kasutatud ADDIE õpikavanduse mudelit. Kurti [14] sõnul põhines mudel esialgu USA õhujõudude kasutatud viie sammu lähenemisest (ingl *Five Step Approach*). Kuna sammudel oli hierarhiline struktuur, siis tuli neid täita järk-järgult. Hiljem muudeti varasemat mudelit, mistõttu on uus mudel interaktiivsem ja dünaamilisem. Tänapäevane ADDIE mudel on sarnane 1980. aastatel valminud versiooniga [14].

ADDIE mudeli viis sammu on HITSA [15] sõnastanud järgmiselt:

- analüüs (ingl *analyse*)

Analüüsi etapis tuleb välja selgitada õppematerjali teema ja osalejate taust. Samuti pannakse kirja õpiväljundid ja uuritakse, mis võib projekti eesmärkide täitmist takistada [14].

- kavandamine (ingl *design*)

Kavandamise faasis projekt kavandatakse ning määratletakse iga ülesande ajaraamid. Lisaks sellele selgitatakse välja kasutatavad meediumid ning vajalikud ressursid projekti valmimiseks [14].

- väljatöötamine (ingl *development*)

Väljatöötamise faasis algab materjalide koostamine, mille käigus kasutatakse analüüsi tulemusi ja kavandeid. Materjali loomisel tuleb kontrollida, et see täidaks analüüsi faasis seatud eesmärgid [14].

- kasutamine (ingl *implementation*)

Kasutamise faasis toimub materjali testimine ja sealt saadud tagasiside põhjal tehakse materjalis muudatusi. Samuti kirjeldatakse, kuidas kasutatakse juhul, kui materjali läbimisel tekkis probleeme [14].

- hinnangu andmine (ingl *evaluation*)

Hinnangu andmine on ADDIE mudeli viimane faas ja selle eesmärk on selgitada, kas said täidetud õpiväljundid ja anda põhjendusi, kuidas materjali saaks paremaks muuta [14].

2.2 Analüüs

Antud lõputöös loodud materjalide analüüsi etapis uuriti, millised on praegused võimalused Power BI õppimiseks. Seati eesmärgiks luua programmi tutvustav materjal, sest autori arvates seni Tartu Ülikoolis vastavad materjalid puuduvad ja internetist leitavad eestikeelsed materjalid on tasulised.

Materjali loomisel oli oluline välja selgitada tudengite taust, kelle peal õppematerjali testiti. Teatud huvi uuritava programmi vastu on oluline, sest programm on mõeldud andmeanalüüsi tegemiseks. Autori arvates on Power BI tudengite jaoks võõras programm, mistõttu ei ole programmi varasem kogemus õppematerjali läbimiseks nõutud.

Lisaks sellele selgitati välja, milliseid oskuseid pidi õppematerjal parandama. Arvestades bakalaureusetöö mahtu keskendus autor õppematerjali koostamisel visualiseerimisele, sest Power BI on mahukas programm.

Õpiväljundid on sõnastatud järgnevalt. Õppematerjali läbinu:

- oskab koostada interaktiivseid visuaale
- oskab probleemist lähtuvalt andmete abil mudeleid koostada
- oskab leida visuaalide abil andmemustreid

Samuti arvestati piiravaid faktoreid, mis võivad õppematerjali eesmärke takistada. Hoolimata sellest, et Power BI Desktop on Tartu Ülikooli tudengitele tasuta, on seal teatud piirangud. Tegemine on programmiga, mida saab kasutada ainult operatsioonisüsteemiga Windows. Materjali koostamise ajal puudus tugi teistele operatsioonisüsteemidele. Lisaks sellele on failide üleslaadimiseks vajaliku veebiteenuse Power BI Service kasutamine tasuta 30 päeva.

2.3 Kavandamine

Kavandamise etapis tehti esmalt selgeks erinevad vahendid, mida materjalis kasutatakse. Kuna Power BI kasutamine eeldab tudengilt oma ülikooli e-mailiga sisselogimist, siis on tudeng oma koostatud failiga seotud. Ülesannete lahendused jagab tudeng töö hindajaga Power BI Service veebiteenuse abil. Õppematerjal on koostatud PDF-failina, vajalikud failid on võimalik alla laadida õppematerjalis olevate linkide abil.

Õppematerjalis kasutatavate andmete valimisel oli oluline, et need oleks kasutajatele arusaadavad. Samuti oli oluline kasutada suuremahulisi andmeid, et demonstreerida Power BI tabelarvutuse võimekust.

Töös koostatud õppematerjal koosneb ülesannetest, mille koostamisel kasutati 4C/ID õppedisaini mudelile omast toetavat suunamist (ingl *scaffolding*) [16]. See tähendab, et ülesannete lahendamiseks pakutud juhised vähenevad iga ülesandega. Esimeses ülesandes on üksikasjalikult lahenduskäigu kava ette antud ja järgnevates ülesannetes tuleb kasutada varem saadud teadmisi ja vastu võtta täiesti iseseisvad otsused.

Ülesannete läbimiseks planeeriti orienteeruvalt 90 minutit.

2.4 Materjali väljatöötamine

Materjal algab sissejuhatusega, mis tutvustab Power BI-d ja annab ülevaate järgnevate ülesannete kohta. Samuti kirjeldatakse oodatavat õpiväljundit. Loodud õppematerjal koosneb kolmest ülesandest.

Esimeses ülesandes on vaja luua interaktiivne raport, kasutades viit erinevat visuaali. Nende moodustamiseks on antud selgitused ja põhjendatakse, mis oli visuaali loomise eesmärk ning mida see võimaldas. Samuti selgitatakse, kuidas visuaalid raportil paigutada ja antakse juhised, kuidas valida telgede, arväärtuste ja pealkirjade suuruseid. See on vajalik, et koostada professionaalne raport, mis annab kiire ülevaate loodud visuaalidele. Ülesande eesmärk on selgitada, kuidas luua visuaale. Peamiste mõjutajate visuaali abil näidatakse, kuidas leida visuaali abil andmemustreid.

Teises ülesandes tuleb vastata kaheksale küsimustele ja seda saab teha esimeses ülesandes loodud raporti abiga. Kui esimeses ülesandes piisas juhtnööride täitmisest, siis nüüd tuleb küsimustele vastamiseks sobivad visuaalid ise üles leida. Ülesande eesmärk on testida, kas varem loodud raport on korrektselt tehtud ja kas sealt osatakse andmeid välja lugeda.

Kolmanda ülesande tüüp on probleemülesanne, mille käigus tuleb koostada raport vajalike visuaalidega. Selles ülesandes pole lahenduskäiku ette antud, sest ülesande eesmärk on kontrollida, kas esimeses ja teises ülesandes saadud kogemusi osatakse kasutada. Ülesande lahendamiseks tuleb koostada andmemudel, lähtudes probleemülesandest.

2.5 Materjali kasutamine

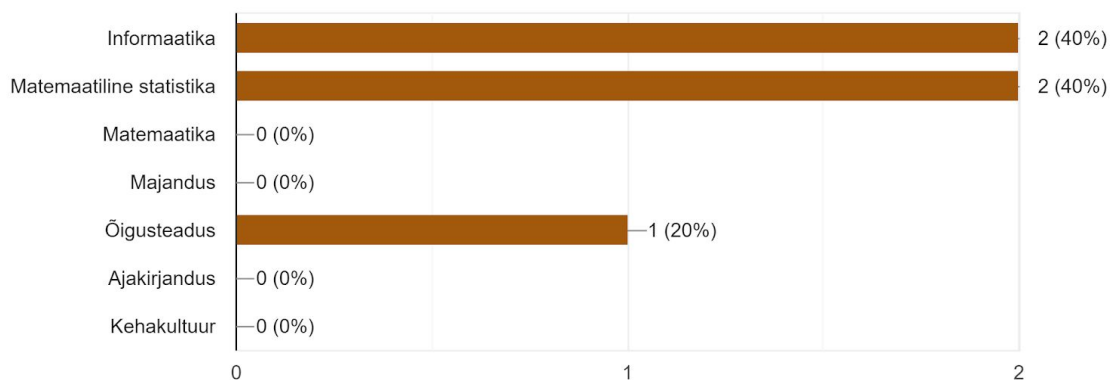
Materjali kasutamist testiti juhendatud õppeprotsessis. Materjali kavandamisel oli plaanis koostada näidispraktikum klassiruumis, et õppematerjali testida. Paraku polnud see võimalik seoses Eesti Vabariigis kehtestatud eriolukorraga. Näidispraktikum toimus videokõne vahendusel, kasutades keskkonda Zoom.

Näidispraktikum seisnes ülesannete täitmisel ja tagasiside andmisel. Praktikumis osales kaheksa inimest, ent kahel ei õnnestunud tehniliste probleemide tõttu programmi alla laadida ja praktikumis osaleda. Kokku osales programmi testimises kuus inimest ja tagasisidet jagas viis inimest. Järgnevalt analüüsitakse õppematerjali ja vastavat tagasisidet.

Praktikumis osales tudengeid kolmelt erinevalt bakalaureuseastme erialalt. Informaatika ja matemaatilise statistika erialasid esindasid kokku neli tudengit ning lisaks neile osales praktikumis üks õigusteaduse tudeng. Tagasisidet jaganud tudengite seas oli üks naine ja neli meest ning tudengite vanusevahemik oli 20-22 aastat.

Mis eriala tudeng oled?

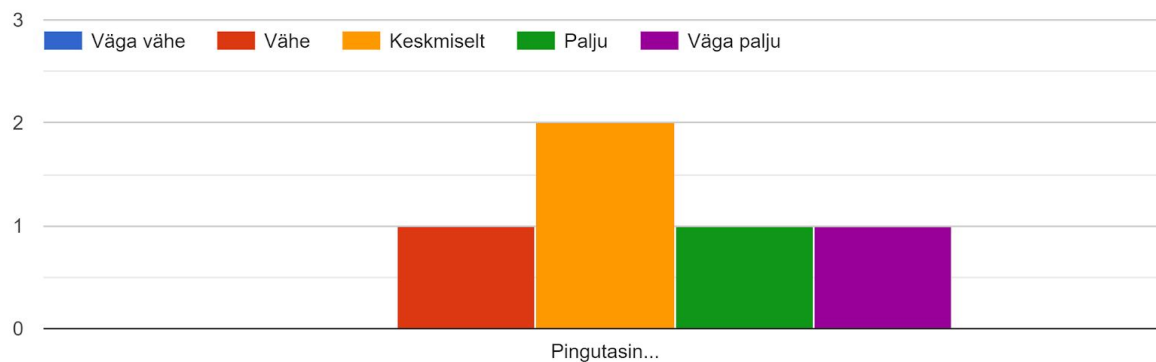
5 responses



Joonis 7. Tagasisidet andva tudengi eriala.

Kuna autoril puudus varasem kogemus õppematerjali läbiviimisel ja video kaudu oli keeruline tudengite motivatsioonist aru saada, siis uuriti, kui palju õppematerjali läbimisel pingutati.

Kui palju pingutasite õppematerjali läbimiseks?



Joonis 8. Tudengi pingutamise õppematerjali läbimisel.

Joonisel 8 esitatud graafiku põhjal võib väita, et ülesannete lahendamisel pingutati keskmiselt või isegi veidi rohkem.

Tagasiside küsimustikus uuriti iga ülesande püstituse, mahu ja kasulikkuse kohta.

Tabel 1. Tagasiside seoses esimese ülesandega.

	Üldse ei nõustu	Ei nõustu	Neutraalne	Nõustun	Väga nõustun
Ülesande püstitus oli selge	-	-	-	3	2
Ülesanne oli hästi organiseeritud ja planeeritud	-	-	1	3	1
Ülesande maht oli sobilik	-	-	1	2	2
Ülesanne võimaldas	-	-	1	2	2

täielikult kaasa töötada					
Sain ülesandest kasu	-	1	-	1	3

Tabelis 1 kirjeldatud vastuste põhjal võib väita, et esimese ülesande püstitus oli selge, sest nõustuti ülesande püstituse selgusega. Samuti oldi nõus, et ülesanne on hästi organiseeritud ja planeeritud. Ülesande mahtu hindasid tudengid sobilikuks, ehkki selle peale kulus rohkem aega, kui õppematerjali koostaja teada andis. Eelneva põhjustajaks oli osalejate erinev interneti allalaadimiskiirus, mis mõnel juhul muutis andmete kättesaamise aeglaseks. Tudengid olid nõus väitega, et ülesanne võimaldas täielikult kaasa töötada. Ülesande kasulikkusega üks tudeng ei nõustunud, ent ülejäänud neli nõustusid.

Tabel 2. Tagasiside seoses teise ülesandega.

	Üldse ei nõustu	Ei nõustu	Neutraalne	Nõustun	Väga nõustun
Ülesanne püstitus oli selge	-	-	-	2	3
Ülesanne oli hästi organiseeritud ja planeeritud	-	-	1	1	3
Ülesande maht oli sobilik	-	-	1	1	3
Ülesanne võimaldas täielikult kaasa töötada	-	1	1	-	3

Sain ülesandest kasu	-	-	2	1	2
----------------------	---	---	---	---	---

Tabelis 2 kirjeldatud vastustest selgub, et teise ülesande püstitusega olid tudengid rahul. Väitega, et ülesanne oli hästi organiseeritud ja planeeritud olid vastajad enamasti nõus. Samuti oli nende arvates sobilik ülesande maht. Küsimusele, mis uuris kas ülesanne võimaldas täielikult kaasa töötada vastas “Ei nõustu” üks, “Neutraalne” üks ja “Väga nõustun” kolm. Negatiivse vastuse andja ei ole hiljem tekstilises tagasisides sellest kirjutanud. Samas oli tema arvates probleemne ühe küsimuse sõnastus, mida hiljem muudeti. Ülesande kasulikkusega jäädi rahule.

Tabel 3. Tagasiside seoses kolmanda ülesandega.

	Üldse ei nõustu	Ei nõustu	Neutraalne	Nõustun	Väga nõustun
Ülesanne püstitus oli selge	-	1	-	2	2
Ülesanne oli hästi organiseeritud ja planeeritud	-	1	1	1	2
Ülesande maht oli sobilik	-	-	-	2	3
Ülesanne võimaldas täielikult kaasa töötada	-	-	-	3	2
Sain ülesandest kasu	-	-	-	2	3

Tabelis 3 kirjeldatakse kolmanda ülesanne tagasisidet. Ülesande selge püstitusega ei nõustunud üks vastaja, ent teised kas “Nõustusid” või “Väga nõustusid”. Ülesande hea organiseerimisega üks vastaja ei nõustunud. Selle põhjuseks oli asjaolu, et temal ei õnnestunud ülesannet Power BI Service-i abil esitada. Seega võttis tal ZIP-failiga ülesande edastamine rohkem aega. Ülesanne mahuga olid vastajad rahul. Samuti oldi kas nõus või väga nõus väitega, et ülesanne võimaldas täielikult kaasa töötada. Ülesande kasulikkusega oldi samuti nõus.

Tabel 4. Tagasiside seoses õppematerjali eesmärkide täitmisega.

	Üldse ei nõustu	Pigem ei nõustu	Neutraalne	Pigem nõustun	Väga nõustun
Õppematerjal tekitab huvi programmi vastu	-	-	-	3	2
Oskan koostada interaktiivseid visuaale	-	-	-	4	1
Oskan probleemist lähtuvalt andmete abil mudeleid koostada	-	-	-	4	1
Õppematerjalist saadud teadmistest on kasu edaspidiseks	-	-	-	2	3
Oskan leida	-	-	1	1	3

visuaalide abil andmemustreid					
Kasutaksin programmi veel		1	1	1	2
Sooviksin programmi veel edasi õppida		1	2	-	2

Tabel 4 kirjeldab õppematerjali eesmärkide täitmist. Vastustest lähtuvalt võib väita, et õppematerjal tekitas huvi programmi vastu. Samuti saab vastustest välja lugeda, et suurem osa vastanutest kasutaks programmi veel ja programmi edasise õppimise vastu oli ainult üks tudeng. Nõustuti õppematerjali sissejuhatuses sõnastatud eesmärkidega ehk osatakse koostada interaktiivseid visuaale ja probleemist lähtuvalt andmete abil mudeleid ning leida visuaalide abil andmemustreid.

Tagasisides küsiti ka kolme küsimusega tekstilist tagasisidet. Esimese küsimuse eesmärk oli uurida, mis õppematerjalis kõige rohkem meeldis ja millest kasu oli. Vastustes toodi välja andmemustrite leidmine. Samuti meeldis vastajatele loodud töö tulemus ja kiideti programmi kiirust näiteks andmete keskmise arvutamisel. Kiideti ka kolmeosalist ülesehitust ja raporti kujundust: “Kolmeosaline ülesehitus oli väga loogiline ja põnev. Esimene ülesanne oli üksikasjalik ja väga detailne, aga teine ja kolmas ülesanne aitasid programmi ja loodut kinnistada. Lisaks oli tore, et pealkirjad ja sildid muutsime eestikeelseks. Pärast on ülevaade palju arusaadavam ja loetavam”.

Teise küsimuse eesmärk oli uurida, mis antud õppematerjali osas kõige vähem meeldis ja kuidas seda parandada. Visuaalide loomisel kirjutati, et õiget visuaali oli raske leida. Lisaks sellele kirjutati, et oleks võinud seletada, kuidas keerulisemaid visuaale teha.

Kolmandas küsimuses sooviti teada mõtteid ja ettepanekuid seoses õppematerjaliga. Vastustest tuli välja, et õppematerjal võiks olla mõnes kohas paremini sõnastatud. Õppematerjali läbimise ajal oli avatud kommentaaride võimalus, kus sai kritiseerida sõnastust või pakkuda midagi paremat välja. Samuti tehti ettepanek andmestiku suuruse vähendamiseks ja soovitati, et Power BI Desktop võiks olla juba enne õppematerjali läbimist allalaetud olla.

2.6 Hindamine

Pärast materjali testimist viidi läbi mitmeid muudatusi. Muudeti õppematerjali sõnastust, lähtuvalt tudengite tagasisidest. Esimese ülesande andmefaili lingi juurde kirjutati soovitus, et aja kokkuhoiu mõttes võiks ka kolmanda ülesande andmefaili alla laadimisega seal alustada. Lisaks sellele tekkisid mõnel osalejal andmestiku sissetoomisega probleem, kus Power BI oli arvulistele väärtustele määranud tekstitüübi. Selle probleemi lahendamiseks koostati eraldi juhised, mis lisati vastavasse etappi.

Tagasiside kokkuvõtteks võib öelda, et programm võeti hästi vastu. Tagasisidest lähtuvalt võib väita, et tudengite arvates said õpiväljundid täidetud. Samuti hindas töö autor kõikide saadetud töid ja autori arvates oli tulemused head.

Materjali paremaks muutmiseks võiks koostada juurde keerulisemaid ülesandeid. Samuti tuleks materjali testida suurema testgrupiga, et saada veel rohkem tagasisidet.

3 Edasine arendus

Autori arvates on programmil väga suur potentsiaal ja autori kogemusest lähtuvalt pole Power BI veel eriti tuntud programm Tartu Ülikoolis. Järgnevalt pakub töö autor välja oma mõtteid ja ideid, kuidas õppematerjali edasi arendada.

Töös koostatud õppematerjal keskendus rohkem programmi tutvustamisele ja visualiseerimisele. Andmete sissetoomisel kasutati võrdlemisi mugaval kujul andmeid, ent andmete sissetoomiseks on veel mitmeid võimalusi. Õppematerjali edasiarendusena võiks programmi ühendada andmebaasiga ja sealt saadud andmetabelite vahel Power BI abiga seoseid luua. Samuti võiks koostada visuaale, kasutades reaalaaja päringute abil saadud andmeid.

Lisaks andmebaasidele võiks ühendada programmi ka mõne veebiteenusega näiteks Google Analytics-iga. Selle abil saaks koostada interaktiivse raporti, mis kasutab otse veebilehelt saadud andmeid.

4 Kokkuvõte

Töös kirjeldati andmete analüüsimise programmi Microsoft Power BI ja koostati selle kohta õppematerjal. Anti ülevaade programmi kujunemise ajaloost ning kirjeldati erinevaid võimalusi Power BI abil andmete sissetoomiseks, hoidmiseks ja visualiseerimiseks. Samuti kirjeldati programmi tööriistu, mis kasutasid tehisintellekti võimalusi. Töö raames koostati õppematerjal, mis tutvustas Power BI abil visualiseerimist ja andmete analüüsimist. Kirjeldati õppematerjali loomise protsessi, sealhulgas ka õppematerjali testimist. Koostati küsitlus ja selle abil analüüsiti testimisel saadud tagasisidet. Samuti pakuti välja ideid õppematerjali edasiseks arendamiseks.

5 Viidatud kirjandus

- [1] Davenport T. and Bean R. "Big Data Executive Survey," 2018, p. 9.
<http://newvantage.com/wp-content/uploads/2018/01/Big-Data-Executive-Survey-2018-Findings-1.pdf>. [19.03.2020].
- [2] IBM. "What is big data," 2017. [Online].
<https://www01.ibm.com/software/data/bigdata/what-is-big-data.html>. [19.03.2020].
- [3] Gantz J. and Reinsel D. "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," 2012, p. 1.
<https://www.cs.princeton.edu/courses/archive/spring13/cos598C/idc-the-digital-universe-in-2020.pdf>. [23.03.2020].
- [4] Bruckner R. "A Glimpse at Project Crescent," 2010. [Online].
<https://blogs.msdn.microsoft.com/sqlrsteamblog/2010/11/09/a-glimpse-at-project-crescent/>. [01.12.2019]
- [5] Unkroth K. "Announcing Microsoft SQL Server Code Name "Denali" Community Technology Preview 3 (CTP3)," 2011. [Online].
<https://blogs.msdn.microsoft.com/analysisservices/2011/07/12/announcing-microsoft-sql-server-code-name-denali-community-technology-preview-3-ctp3/>. [01.12.2019]
- [6] Albanesius C. "Office 365 Gets Colorful 3D Charts, Natural Language Search," PCMag, 2013. [Online].
<https://in.pcmag.com/productivity/70071/office-365-gets-colorful-3d-charts-natural-language-search>. [01.12.2019].
- [7] The Power BI Team. "Announcing Power BI general availability coming July 24th, 2015. [Online].
<https://powerbi.microsoft.com/en-us/blog/announcing-power-bi-general-availability-coming-july-24th>. [01.12.2019].
- [8] Aspin A. "Pro Power BI Desktop," Apress, 2018. ISBN 978-1-4842-3209-5.
- [9] Larson B. and Corley D. "Data analysis with Microsoft Power BI," McGraw-Hill Education, 2020. ISBN: 978-1-26-045862-6.
- [10] Powell B. "Microsoft Power BI Cookbook", Packt Publishing, Limited, 2017. ISBN: 978-1-78829-014-2.
- [11] Powell B. "Mastering Microsoft Power BI: expert techniques for effective data analytics and business intelligence," Packt Publishing, 2018. ISBN: 978-1-78829-723-3.
- [12] Crittenden J. "Exploring Power BI's Key Influencers," 2019. [Online].
<https://www.blue-granite.com/blog/exploring-power-bis-key-influencers>. [20.04.2020].
- [13] Deckler G. "Learn Power BI: a beginner's guide to developing interactive business intelligence solutions using Microsoft Power BI," Packt Publishing Ltd, 2019. ISBN: 978-1-83864-448-2.
- [14] Kurt S. "ADDIE Model: Instructional Design," 2018. [Online].
<https://educationaltechnology.net/the-addie-model-instructional-design/>. [15.04.2020].
- [15] HITSA. "Öppematerjalide väljatöötamise etapid," 2015. [Online].
<https://oppevara.hitsa.ee/kvaliteet/#oppematerjalide-valjatootamise-etapid>. [15.04.2020].
- [16] Van Merriënboer J., Clark R. & de Croock M. "Blueprints for complex learning: The 4C/ID-model," 2002, p. 45.

https://www.researchgate.net/publication/225798787_Blueprints_for_complex_learning_The_4CID-model. [13.04.2020].

Lisad

1. Juurdepääs õppematerjalide kaustale.
https://drive.google.com/open?id=1_v-UNmmCQNwKCLepTOrgAGV-jAjjF_6C
2. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Richard Nõmm,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

Ülevaade Microsoft Power BI-st ja õppematerjali koostamine,

(lõputöö pealkiri)

mille juhendaja on Vambola Leping,

(juhendaja nimi)

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 Commons'i 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.

Richard Nõmm

08.05.2020