

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond
Tarkvarateaduse instituut

Heleriin Ots 183138 IAPM

**ÕPISOORITUSE ENNUSTAMINE
MOODLE'I LOGIANDMETE JA
ENESEHINNANGULISTE ÕPPIMISEGA
SEOTUD PSÜHHOLOOGILISTE TEGURITE
PÕHJAL**

Magistritöö

Juhendaja: Ago Luberg

MSc

Tallinn 2020

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Heleriin Ots

13.05.2020

Annotatsioon

Tudengite suur väljalangevus ülikoolist on Eestis, sealhulgas ka Tallinna Tehnikaülikoolis, aktuaalne probleem. Eriti teravalt mõjutab suur väljalangevus infotehnoloogia valdkonna tudengeid, näiteks langeb Eestis ülikoolist esimesel aastal välja tervenisti 32% info- ja kommunikatsioonitehnoloogia erialadel õppivaid tudengeid [1, lk 4]. Seetõttu on oluline kasutada õpianalüütikat, mis aitab luua parema arusaama õppimise ja õpetamise protsessidest, et vähendada tudengite väljalangemist ülikoolist [2].

Käesolevas magistritöös vastati kolmele uurimisküsimusele. Selgitati välja, mil määral saab Moodle'i aktiivsuspõhistel logidel baseeruvat õpiedukuse ennustustäpsust parandada lisades küsitluse abil leitud õppimisega seotud psühholoogilised faktorid. Leiti, mil määral korreleeruvad Moodle'i aktiivsuslogid õppimisega seotud psühholoogiliste faktoritega. Uuriti, millised psühholoogilised faktorid näitavad statistiliselt märkimisväärset suhet aktiivsusmuustritega ning kui vara ilmnevad need näitajad õppeprotsessis.

Magistritöös leiti, et õppimisega seotud psühholoogiliste tegurite kasutamine tudengite klassifitseerimisel annab Moodle'i logiandmete kasutamisest suuremad ennustustäpsused. Osadel juhtudel aitab psühholoogiliste tegurite lisamine Moodle'i logiandmetele parandada ennustustäpsuseid. Aktiivsusega seotud logiandmed korreleerusid kõige enam õpimotivatsiooni, -strateegiate ning õppimisega seotud uskumuste hulka kuuluvate faktoritega. Hinnetega seotud logiandmed korreleerusid kõige enam õpistrateegiatega seotud psühholoogiliste faktoritega. Aktiivsusmuustrid ilmnevad Moodle'i logides juba semestri esimesel kuul ning semestri esimese kuu logiandmete ning õppimisega seotud psühholoogiliste faktorite vahel on võimalik leida korrelatsioone.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 93 leheküljel, 7 peatükki, 13 joonist, 35 tabelit.

Abstract

Predicting academic achievement based on Moodle log data and self-assessed learning-related psychological factors

The high university dropout rate is a topical issue in Estonia, which is also apparent in Tallinn University of Technology. Students in the field of information technology are particularly affected by the high drop-out rate, for example, in Estonia, up to 32% of students in the field of information and communication technology drop out of university already in their first year [1, lk 4]. Thus, it is important to use learning analytics, which help to create a better understanding of learning and teaching processes, to reduce the university drop-out rate [2].

In this master's thesis, three research questions were answered. It was found to which extent prediction accuracy based on Moodle activity logs can be improved by adding self-reported learning-related psychological factors. Also, it was found to which extent Moodle activity logs correlate with learning-related psychological factors. Furthermore, it was established which learning-related psychological factors show a significant statistical connection to activity patterns and how early these factors become evident in the learning process.

In this master's thesis, it was found that the use of learning-related psychological factors in the classification of students results in higher prediction precision scores than the use of Moodle log data. In some cases adding learning-related psychological factors to Moodle log data improves prediction precision scores. Activity related log data correlated the most with factors that belong to the categories of study motivation, learning strategy and learning-related beliefs. Grade related log data correlated the most with psychological factors that are related to learning strategy. Activity patterns become apparent in Moodle logs already during the first month of the semester and it is possible to find correlations between log data from the first month of the semester and learning-related psychological factors.

The thesis is in Estonian and contains 93 pages of text, 7 chapters, 13 figures, 35 tables.

Lühendite ja mõistete sõnastik

aBIC	<i>the adjusted Bayesian Information Criterion</i> . Korrigeeritud Bayesi informatsiooni kriteerium.
AIC	<i>the Akaike Information Criterion</i> . Akaike informatsiooni kriteerium.
BIC	<i>the Bayesian Information Criterion</i> . Bayesi informatsiooni kriteerium.
CRUD	<i>create, read update, delete</i> . Loo, loe, uuenda ja kustuta operatsioonid.
CSV	<i>comma-separated values</i> . Komadega eraldatud väärtusi sisaldav tekstifaili formaat.
JSON	<i>JavaScript Object Notation</i> . JavaScripti objektide notatsioonil põhinev failiformaat.
LMR	Lo-Mendell-Rubini test.
MEAP	<i>Moodle Engagement Analytics Plugin</i> . Moodle'i keskkonna kaasatuse analüütika pistikproramm.
MEAP+	<i>Moodle Engagement Analytics Plugin +</i> . Moodle'i keskkonna kaasatuse analüütika pistikproramm täiendatud versioon.
<i>Moodle Dashboard</i>	Moodle'i keskkonna poolt pakutav sisseehitatud töölaud.
<i>Moodle Reporting Tool</i>	Moodle'i keskkonna poolt pakutav sisseehitatud aruandluse tööriist.
PHP	Programeerimiskeel.
RegEx	<i>Regular expression</i> . Regulaaravaldis on sümbolite jada, mis defineerib tekstist otsitava mustri
TalTech	Tallinna Tehnikaülikool.
TSV	<i>tab-separated values</i> . Tabeldusmärkidega eraldatud väärtusi sisaldav tekstifaili formaat.
TXT	Tekstifaili formaat, mis sisaldab vormindamata teksti.
XLSX	Microsoft Excel'i failiformaat.

Sisukord

Autorideklaratsioon.....	2
Annotatsioon.....	3
Abstract Predicting academic achievement based on Moodle log data and self-assessed learning-related psychological factors	4
Lühendite ja mõistete sõnastik	6
Sisukord.....	7
Jooniste loetelu	11
Tabelite loetelu	13
1 Sissejuhatus	16
1.1 Magistritöö eesmärgid	17
1.2 Uurimismeetod.....	17
1.3 Ülevaade magistritööst	19
2 Ülevaade teooriast.....	21
2.1 Õpialalüütika.....	21
2.2 Moodle'i keskkond.....	23
2.2.1 Õpialalüütika Moodle'i keskkonnas	23
2.2.2 Moodle'i keskkonna logi seosed kursuste lõpptulemustega	26
2.3 Õpialalüütika Tallinna Tehnikaülikoolis	29
2.4 Õppimisega seotud enesehinnanguliste psühholoogiliste tegurite küsitlus	30
2.4.1 Õpipädevus.....	31
2.4.2 Õppimisega seotud emotsioonid	31
2.4.3 Õppimisega seotud ning enesekohased uskumused	32
2.4.4 Õpikäitumine.....	33
2.4.5 Õpimotivatsioon	35
2.5 Ennustamine.....	38
2.5.1 Otsustuspuu	38
2.5.2 Masinõppemudelite hindamine	39
2.6 Klasterdamine	39
2.6.1 <i>K-means</i> algoritm	41

2.6.2	Bayesi informatsiooni kriteerium	41
2.6.3	Keskmine siluettide koefitsient	42
3	Andmed	43
3.1	Andmekaitse	43
3.2	Õppeained	44
3.3	Kasutatud Moodle'i keskkonna logi andmed	45
3.3.1	Moodle'i logirida	46
3.3.2	Esmane logiandmete töötlus ning analüüs	47
3.3.3	Logiandmete töötlemine	50
3.3.4	Aktiivsus	52
3.3.5	Jooksvad tulemused	53
3.3.6	Keskmine jooksev tulemus	53
3.3.7	Ülesannete esitamise arv	53
3.4	Küsitlus	54
3.4.1	Küsitluse küsimused	54
3.4.2	Küsitluse koondtunnused	55
4	Klasterdamine	58
4.1	Sobivate klastrite arvude leidmine	59
4.1.1	Programmeerimise algkursus	59
4.1.2	Robotite programmeerimine	61
4.1.3	Sissejuhatus erialasse	62
4.2	Tudengite andmeid esitavad graafikud	63
4.3	Klasterdamise tulemused	64
4.3.1	Küsitluse koondtunnused	64
4.3.2	Jooksvad tulemused	65
4.3.3	Aktiivsus	66
4.3.4	Ülesannete esitamise korrad	66
5	Õppeaine sooritustulemuse ennustamine	68
5.1	Kursuste õppeedukuse klasside ennustamine	68
5.1.1	Ühe koondtunnuse kaupa ennustamine	70
5.1.2	Logi ning küsitluse andmete alusel ennustamine	71
5.1.3	Tunnuste valimisega mudelite treenimine	73
5.2	Võrdlusbaas	74
5.3	Küsitluse koondtunnuste ennustamine	74

5.4 Korrelatsioon Moodle'i logide ja psühholoogiliste faktorite küsitluse koondtunnuste vahel.....	75
6 Tulemused	77
6.1 Õppe edukuse ennustamine	77
6.1.1 Küsitluse andmete kasutamine programmeerimise algkursuse õppe edukuse klasside ennustamisel	78
6.1.2 Küsitluse andmete kasutamine robotite programmeerimise ja sissejuhatus erialasse õppe edukuse klasside ennustamisel	81
6.1.3 Moodle'i logiandmete kasutamine programmeerimise algkursuse õppe edukuse klasside ennustamisel	81
6.1.4 Moodle'i logiandmete kasutamine robotite programmeerimise õppe edukuse klasside ennustamisel	85
6.1.5 Moodle'i logiandmete kasutamine sissejuhatus erialasse õppe edukuse klasside ennustamisel	86
6.1.6 Programmeerimise algkursuse andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused	88
6.1.7 Robotite programmeerimise andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused	90
6.1.8 Sissejuhatus erialasse andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused	92
6.2 Küsitluse koondtunnuste seotus Moodle'i logiandmetega	94
6.2.1 Spearmani korrelatsioonikoefitsiendid	94
6.3 Soovitused.....	101
6.3.1 Küsitlus	101
6.3.2 Ennustamine.....	102
6.3.3 Moodle'i keskkonna kasutamine.....	103
6.3.4 Pistikprogrammi loomine Moodle'i keskkonnale	104
7 Kokkuvõte	105
7.1 Olulised tähelepanekud	108
Kasutatud kirjandus	109
Lisa 1 – Võrdlusbaasi mudelite õppe edukuse klasside ennustustulemused semestri esimese kuu andmetele tuginedes	112
Lisa 2 – Võrdlusbaasi mudelite õppe edukuse klasside ennustustulemused semestri esimese kahe kuu andmetele tuginedes.....	113

Lisa 3 – Õppeedukuse klasside ennustustulemused koondtunnusute kaupa.....	114
Lisa 4 – Õppeedukuse klasside ennustustulemused semestri esimese kuu andmete pealt	115
Lisa 5 – õppeedukuse klasside ennustustulemused semestri kahe esimese kuu andmete pealt õppeainete kaupa	117
Lisa 6 – Õppeedukuse klasside ennustustulemused semestri kahe esimese kuu andmete pealt üldiste mudelite korral	119
Lisa 7 – Õppeedukuse klasside ennustustulemused pärast parimate tunnuste valimist	121
Lisa 8 – Küsitluse koondtunnuste ennustustulemused programmeerimise algkursuse logiandmete pealt.....	127
Lisa 9 – Semestri esimese kuu Moodle'i andmete Spearmani korrelatsioonikoefitsiendid küsitluse koondtunnustega	129
Lisa 10 – Semestri esimese kahe kuu Moodle'i andmete Spearmani korrelatsioonikoefitsiendid küsitluse koondtunnustega.....	132
Lisa 11 – Projekt.....	135
Lisa 12 – Näide programmeerimise algkursuse jooksvate tulemuste graafikust positiivsete lõpphinnete korral.....	136
Lisa 13 – Näide programmeerimise algkursuse jooksvate tulemuste graafikust negatiivsete lõpphinnete korral.....	137

Jooniste loetelu

Joonis 1. Õpianalüütika tsükkel.....	21
Joonis 2. Programmeerimise algkursuse Moodle'i logirida.....	46
Joonis 3. Programmeerimise põhikursuse 20 kõige vähemaktiivset kasutajat logiandmete alusel. Vertikaalsel teljel kujutatakse logikirjete arvu ning horisontaalsel teljel kasutajate Moodle'i identifikaatoreid.....	48
Joonis 4. Programmeerimise põhikursuse logis 10 kõige sagedamini esinevat sündmust. Vertikaalsel teljel on esitatud protsentuaalne skaala ning horisontaalsel teljel sündmuste nimed.....	49
Joonis 5. Programmeerimise põhikursuse kõige suurema aktiivsusega kuud logiandmete alusel. Vertikaalsel teljel on esitatud protsentuaalne skaala ning horisontaalsel teljel kevadsemestri kuude numbrid.....	50
Joonis 6. Programmeerimise algkursuse semestri esimese kuu aktiivsuse näitajate andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.....	59
Joonis 7. Programmeerimise algkursuse semestri esimese kuu aktiivsuse näitajate andmehulga keskmiste siluettide koefitsientide väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel keskmisi siluettide koefitsiente.	60
Joonis 8. Programmeerimise algkursuse küsitluse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.....	61
Joonis 9. Robotite programmeerimise kursuse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.....	62
Joonis 10. Sissejuhatus erialasse kursuse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.....	63
Joonis 11. Robotite programmeerimise semestri kahe esimese kuu ülesannete jooksvate hinnete graafik protsentuaalsel skaalal, mis kujutab osa aine läbikukkunud tudengitest66	

Joonis 12. Näide programmeerimise algkursuse küsitluse koondtunnuseid kasutanud mudeli poolt genereeritud otsustuspuust.....	70
Joonis 13. Tudengite arv programmeerimise algkursuse hindeklassides küsitluse andmete korral.....	79

Tabelite loetelu

Tabel 1. Näidis logikirjet lahti seletav tabel.....	47
Tabel 2. Väljavalitud koondtunnuste Spearmani korrelatsioonikoefitsiendid programmeerimise algkursuse lõpphinnetega.....	56
Tabel 3. Programmeerimise algkursuse õppeedukuse klasside ennustamises parimaks osutunud koondtunnuste ennustustulemused.	78
Tabel 4. Programmeerimise algkursuse õppeedukuse klasside ennustustulemused küsitluse küsimuste ning koondtunnuste korral.....	80
Tabel 5. Programmeerimise algkursuse semestri esimese kuu Moodle'i logisid kasutavate andmehulkade ennustustulemused.....	82
Tabel 6. Programmeerimise algkursuse semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustulemused	84
Tabel 7. Programmeerimise algkursuse ennustustäpsused enne ja pärast parimate tunnuste valmist.	84
Tabel 8. Robotite programmeerimise semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.	86
Tabel 9. Sissejuhatus erialasse semestri esimese kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.	87
Tabel 10. Sissejuhatus erialasse semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.	87
Tabel 11. Programmeerimise algkursuse semestri esimese kuu andmete peal treenitud ning robotite programmeerimise semestri esimese kuu andmete peal testitud mudelite ennustustäpsused.....	88
Tabel 12. Programmeerimise algkursuse semestri esimese kahe kuu andmete peal treenitud ning robotite programmeerimise semestri esimese kahe kuu andmete peal testitud mudelite ennustustäpsused.	89
Tabel 13. Programmeerimise algkursuse semestri esimese kahe kuu andmete peal treenitud ning sissejuhatus erialasse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.	90
Tabel 14. Robotite programmeerimise semestri esimese kahe kuu andmete peal treenitud ning sissejuhatus erialasse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.	91

Tabel 15. Robotite programmeerimise semestri esimese kahe kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.	92
Tabel 16. Sissejuhatus erialasse semestri esimese kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.	93
Tabel 17. Sissejuhatus erialasse semestri esimese kahe kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.	93
Tabel 18. Programmeerimise algkursuse semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	95
Tabel 19. Programmeerimise algkursuse semestri esimese kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	95
Tabel 20. Programmeerimise algkursuse semestri esimese kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	95
Tabel 21. Programmeerimise algkursuse semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	96
Tabel 22. Programmeerimise algkursuse semestri esimese kahe kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	96
Tabel 23. Programmeerimise algkursuse semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	96
Tabel 24. Robotite programmeerimise semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	97
Tabel 25. Robotite programmeerimise semestri esimese kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	97
Tabel 26. Robotite programmeerimise semestri esimese kuu <i>number_of_active_weeks</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	97
Tabel 27. Robotite programmeerimise semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	98
Tabel 28. Robotite programmeerimise semestri esimese kahe kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	98

Tabel 29. Robotite programmeerimise semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.....	98
Tabel 30. Sissejuhatus erialasse semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	99
Tabel 31. Sissejuhatus erialasse semestri esimese kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	99
Tabel 32. Sissejuhatus erialasse semestri esimese kuu <i>number_of_active_weeks</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid	99
Tabel 33. Sissejuhatus erialasse semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.	100
Tabel 34. Sissejuhatus erialasse semestri esimese kahe kuu <i>number_of_active_days</i> andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.	100
Tabel 35. Sissejuhatus erialasse semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.....	100

1 Sissejuhatus

Tudengite suur väljalangevus ülikoolist on Eestis, sealhulgas ka Tallinna Tehnikaülikoolis (TalTech), aktuaalne probleem, millele pole kahjuks veel lahendust leitud. Eriti teravalt mõjutab suur väljalangevus inseneeria ning infotehnoloogia valdkonna tudengeid, näiteks langeb Eestis ülikoolist esimesel aastal välja tervelt 32% info- ja kommunikatsioonitehnoloogia erialadel õppivaid tudengeid [1, lk 4]. TalTechis on antud probleemi uuritud juba pikemat aega, näiteks selgus ühes 2016. aastal avaldatud aruandes, et väljalangenud tudengite hulk kõikidest TalTechi tudengitest oli 17–23%, lisaks sellele kukub juba esimesel ülikooliaastal välja tervenisti 15% kõikidest õpinguid alustanud tudengitest, kellest kolmandik olid saavutanud matemaatika riigieksamil konkurentsivõimelise tulemuse, mis oli suurem kui 75 punkti [3, lk 11–12].

Õpianalüütika on valdkond, mis üritab luua arusaama õppimise ning õpetamise protsessidest, et teha teadlikke ja informeeritud otsuseid nende protsesside parandamiseks. Õpianalüütika valdkonna keskmes on kõigi õpilaste, sealhulgas tudengite, õppe edukuse parandamine [2]. Selleks, et vältida tudengite ülikoolist väljakukkumist, tuleb sekkuda väljalangemisohus tudengite õppeprotsessi võimalikult vara, et seda korrigeerida ja üliõpilased õigele kursile suunata [4]. Ülikoolist väljalangemist mõjutab aga tugevalt õppeainete läbimine – mida rohkem kursuseid kukub tudeng läbi, seda tõenäolisem on, et üliõpilane ei lõpetagi ülikooli. Sellest johtuvalt on oluline võimalikult vara tuvastada tudengid, kes võivad kursuse läbi kukkuda, et õppejõud saaksid neile pakkuda lisatuge, mis aitaks omakorda üliõpilastel õppeaine edukalt läbida.

Nüüdisajal viiakse suurem osa ülikoolide kursustest läbi kasutades suuremal või vähemal määral veebipõhiseid õppekeskkondi. TalTechis on veebipõhise õppekeskkonnana kasutusel Moodle. Veebipõhised õppekeskkonnad talletavad igapäevaselt informatsiooni kõigi tegevuste kohta, mida tudengid ning õppejõud selles keskkonnas teevad, seega võivad Moodle'i-suguste õppekeskkondade andmebaasid sisaldada väärtuslikku informatsiooni tudengite õpikäitumise kohta erinevatel kursustel.

1.1 Magistritöö eesmärgid

Käesolev magistritöö aitab jõuda sammukese võrra lähemale tudengite ülikoolist väljalangemise vähendamisele, selgitades välja, millised Moodle'i keskkonna aktiivsuslogidest leitud õpikäitumist kirjeldavad näitajad ning enesehinnangulisel küsitlusel põhinevad õppimisega seotud psühholoogilised tegurid ennustavad kõige paremine tudengite õppeedukust.

Käesoleva magistritöö esimeseks eesmärgiks on selgitada välja, millised Moodle'i keskkonna logidest andmetöötluse käigus leitud õpikäitumisega seotud tunnused ennustavad kõige paremini tudengite õppeedukust kolmes erineva ülesehitusega Informaatika õppekava bakalaureuse taseme esimese kursuse õppeaines. Teiseks eesmärgiks on leida, millised samadel kursustel osalenud üliõpilaste hulgas tehtud õpikäitumise, -motivatsiooni, õppimisega seotud emotsioonide ning õppimisega seotud ja enesekohaste uskumuste temalise küsitluse näitajad ennustavad kõige paremini tudengite õppeedukust kolmes uuritud kursuses. Magistritöö kolmandaks eesmärgiks on uurida, kas õpikäitumist kirjeldavate Moodle'i logidest leitud näitajate ning küsitluse näitajate vahel on seoseid.

Eesmärkide täitmiseks otsib autor vastuseid järgnevatele uurimisküsimustele:

1. Mil määral saab Moodle'i keskkonna aktiivsuslohistel logidel baseeruvat õpiedukuse ennustamise täpsust parandada, lisades tudengite poolt küsitluse vastustena esitatud andmetest leitud õppimisega seotud psühholoogilised faktorid?
2. Mil määral korreleeruvad Moodle'i keskkonna aktiivsuslogid õppimisega seotud psühholoogiliste faktorite andmetega, mis pärinevad tudengite vastustest küsitlusele (uskumused õppimise kohta, õpimotivatsioon, akadeemilised emotsioonid, õpikäitumine ning õpistrateegiad)?
3. Millised küsitluse vastuste andmetest leitud psühholoogilised faktorid näitavad statistiliselt märkimisväärset suhet aktiivsusmuustritega ning kui varakult ilmnevad need näitajad õpiprotsessis?

1.2 Uurimismeetod

Uurimisküsimustele vastuste leidmiseks kasutab autor uurimismeetodit, mis koosneb kuuest sammust:

1. Probleemi tuvastamine ning magistritöö planeerimine: esimeses sammus tutvub autor probleemiga ning teemakohase teaduskirjandusega, et panna paika magistritöö eesmärgid, uurimisküsimused ja plaanid järgnevateks tegevusteks.
2. Andmetega tutvumine: selles sammus kogutakse magistritöös kasutatavad andmed kokku. Tuleb koostöös Ago Lubergi ning hariduspsühholoogidega Kati Aus ja Grete Arro läbi viia õppimisega seotud psühholoogiliste tegurite küsitlus. Autoril tuleb mõista, kuidas Moodle'i logides andmeid säilitatakse ning millist tudengite õpikäitumist kirjeldavat informatsiooni logid sisaldavad. Lisaks sellele tuleb autoril tutvuda ka läbiviidud õppimisega seotud psühholoogiliste tegurite küsitluse andmetega ja andmetel põhinevate koondtunnustega.
3. Andmetöötlus: kolmandas etapis töötleb ning puhastab autor Moodle'i logi ja küsitluse andmeid ning viib andmed sobivale kujule andmeanalüüsi jaoks.
4. Andmeanalüüs: neljandas sammus proovib autor leida Moodle'i logide ning küsitluse andmetest seoseid ning mustreid. Selleks kasutab autor mitmeid erinevaid meetmeid, sealhulgas andmete klasterdamist, korrelatsioonide leidmist erinevate andmehulkade vahel ning õppeedukuse ennustamist Moodle'i logi ning õppimisega seotud psühholoogiliste tegurite andmete pealt.
5. Järeldused: viiendas sammus vastab autor esitatud uurimisküsimustele tuginedes läbiviidud andmeanalüüsi tulemustele. Ennustustulemuste valideerimiseks võrdleb autor analüüsimise sammus leitud ennustustulemusi võrdlusbaasiga, milleks on logiridade arvu järgi õppeedukust ennustava mudeli tulemus.
6. Soovitused: viimases sammus annab autor soovitusi edasiseks tööks.

Magistritöö projektis (vt Lisa 11) kasutab autor programmeerimiskeelena Pythoni¹ versiooni 3.6. Andmete töötlemiseks kasutab autor peamiselt Pandase² teeki ning vähemal määral ka Numpyt³. Graafikute loomiseks kasutatakse Pythonile loodud Matplotlib⁴ visualiseerimisteeki. Masinõppe ning klasterdamisega seotud ülesanded

¹ <https://www.python.org/downloads/release/python-360/>

² <https://pandas.pydata.org/>

³ <https://numpy.org/#numpy>

⁴ <https://numpy.org/#numpyhttps://matplotlib.org/>

plaanitakse lahendada kasutades scikit-learn¹ ning SciPy² teeke. Otsustuspuude graafikute loomiseks võetakse kasutusele Graphviz³ tarkvara.

Magistritöö projekt jagatakse viide peamisse ossa. Esiteks luuakse skriptid logi ning teiste andmefailide õigetes formaatidesse konverteerimiseks ning esmase logianalüüsi graafikute koostamiseks. Teise gruppi kuuluvad kõik tudengite andmete töötlemisega ja puhastamisega seotud Pythoni klassid ning skript, mis käivitab andmetöötlust. Kolmandaks grupiks on masinõppemudelid sisaldavad Pythoni failid ning skript, mis käivitab masinõppemudelid. Neljandaks on eraldi failid ka andmete klasterdamiseks ning leitud tulemusi kirjeldavate graafikute loomiseks. Viimasena on projektis skript, mis käivitab andmehulkade vaheliste korrelatsioonide leidmise.

1.3 Ülevaade magistritööst

Peatükis „Ülevaade teooriast“ tutvustab autor õpianalüütika valdkonda ning annab põhjaliku ülevaate olemasolevatest õpianalüütika kogumise ning analüüsimise võimalustest Moodle'i keskkonnas. Lisaks eelnevale tutvustab autor õppeedukust ennustavaid teadustöid, mis on tehtud veebipõhiste õppekeskkondade, eelkõige Moodle'i keskkonna logide, andmetele tuginedes. Seejärel annab autor sisuka ülevaate läbiviidud küsitluse teoreetilisest taustast, mis põhineb hariduspsühholoogia alustaladel, seletades üksikasjaliselt lahti küsitluse küsimuste temavaldkonnad ning valdkonnasiseste näitajate õppeedukust ennustavad väärtused. Nendeks temavaldkondadeks on õpipädevus, õppimisega seotud emotsioonid, õpikäitumine, õpimotivatsioon ja õppimisega seotud ning enesekohased uskumused. Peatükk „Ülevaade teooriast“ lõpeb kasutatud masinõppe- ning klasterdamismeetodite ja kasutatud valideerimismeetodite teoreetilise tausta tutvustusega.

Peatükis „Andmed“ kirjeldab autor kasutatud Moodle'i keskkonna logisid ning andmeid, mida logid sisaldavad. Seejärel kirjeldatakse detailselt logiandmete töötlust ning analüüsi.

¹ <https://scikit-learn.org/stable/>

² <https://www.scipy.org/index.html>

³ <https://www.graphviz.org/>

Autor annab ka ülevaate läbiviidud küsitlusest, seal kasutatud küsimuste andmetest ning küsitluse vastuste põhjal leitud koondtunnustest.

Peatükis „Klasterdamine“ kirjeldatakse läbiviidud klasterdamise protsessi, klasterdamisel kasutatud andmeid ning andmete põhjal genereeritud graafikute loomist. Lisaks sellele annab autor ülevaate klasterdamise protsessi käigus andmetest leitud seostest ning muustritest.

Peatükk „Ennustamine“ annab ülevaate Moodle'i keskkonna logiandmete, küsitluse andmete ning küsitluse koondtunnuste põhjal tehtud õppeedukuse ja koondtunnuste väärtuste ennustuste tegemisest. Autor kirjeldab kasutatud andmehulki ning loodud masinõppe mudeleid.

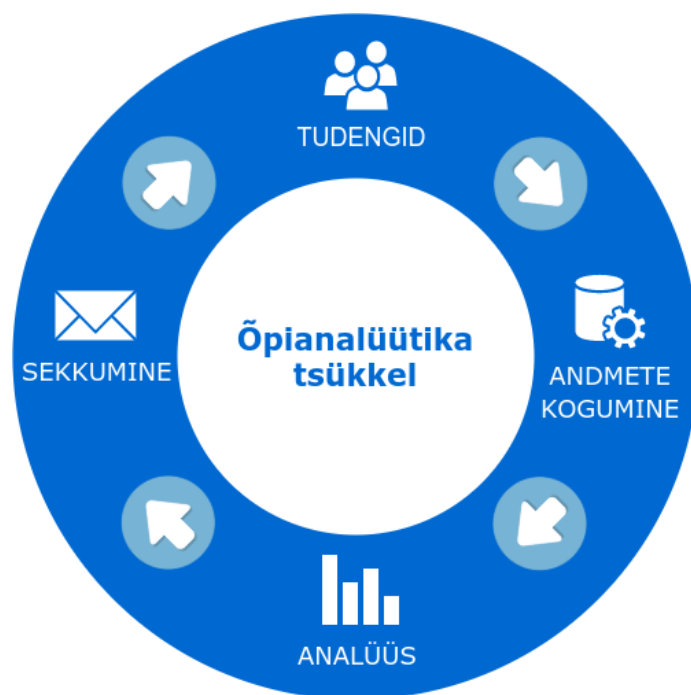
Peatükis „Tulemused“ annab autor põhjaliku ülevaate ennustamisel ning klasterdamisel saadud tulemustest. Lisaks sellele analüüsitakse saadud tulemusi ning antakse soovitusi Moodle'i keskkonna kasutamiseks ning edasiseks õppeedukuse ennustamisega seotud tööks.

2 Ülevaade teoriast

Käesolevas peatükis annab autor ülevaate õpianalüütika, Moodle'i keskkonna ning sealsete õpianalüütika võimaluste, läbiviidud õppimisega seotud psühholoogiliste tegurite küsitluse teoreetilise tausta ning valitud masinõppe- ning klasterdamisalgoritmidega seotud teaduskirjandusest.

2.1 Õpianalüütika

Õpianalüütika eesmärgiks on õpetamise ning õppimise mõistmine selleks, et teha informeeritud otsuseid õpiprotsessi parandamiseks. Suurem osa õpianalüütika valdkonnas läbiviidud uurimustest keskenduvad tudengite õpitulemuste parandamisele ja sellele, et vähem tudengeid kukuks ülikoolist välja. Selleks, et soovitud tulemused saavutada, tuleb leida näitajad, mis indikeerivad probleeme õpiedukusega ning viitavad tõenäolisemale väljakukkumisele [2], [5, lk 180]. Õpianalüütikat saab vaadelda tsüklina, mida kujutab Joonis 1 [4, lk 134–135].



Joonis 1. Õpianalüütika tsükkel.

Kõige esimeseks osaks tsüklist on ülikooli kontekstis tudengid, kes võtavad osa – kas täielikult või osaliselt – veebipõhistest kursustest, kus materjale jagatakse veebipõhises õpikeskkonnas näiteks Moodle'is [4, lk 134–135].

Teiseks sammuks tsüklis on tudengite kohta andmete kogumine ning saadud informatsiooni üldistamine. Näiteks võivad nendeks andmeteks olla tudengite demograafiline informatsioon, õppekeskkonda sisselogimisega seotud info, info tudengite tehtud foorumipostituste kohta, ülesannete ning kodutööde tulemused ja ka tudengi lõpetatud koolide ning ülikoolidega seotud informatsioon [4, lk 134–135].

Kolmandaks õpianalüütika tsükli osaks on andmete töötlemine, analüütika koostamine ning huvipakkuvate mõõdikute leidmine, mis aitavad kirjeldada õppimisprotsessi. Selle sammu alla kuulub näiteks andmete visualiseerimine, erinevate ülevaadete koostamine, väljakukkumisoshus või siis kursuse läbimisega raskustes olevate tudengite kaardistamine, andmete võrdluste koostamine jms. Osad nendest ülesannetest on triviaalsed ning neid on võimalik näiteks veebipõhises õpikeskkonnas automaatselt teha, teised on seevastu aga keerukad ning aja- ning ressursimahukad. Selles valdkonnas viiakse läbi ka suurim osa õpianalüütikaga seotud uurimustest [4, lk 134–135].

Neljandaks ning viimaseks sammuks õpianalüütika tsüklis on sekkumine. Sekkumine tähendab, et leitud mõõdikuid ning informatsiooni kasutatakse tudengite õppetöösse sekkumisel, et mõjutada nende õppimisprotsessi ning tulemusi. Ilma sekkumiseta ei ole õpianalüütika tsükkel lõpetatud ja täielik. Sekkumiseks võib olla näiteks tudengitele nende õppeprotsesside informatsiooni kuvamine, et nad saaks enda tulemusi võrrelda kursusekaaslaste omadega, või siis õppejõud võtmas ühendust tudengitega, kellel on suur oht ülikoolist välja kukkuda. Õpianalüütika tsükkel on lõpetatud ka siis, kui ei sekkuta nende tudengite tegevusse, kelle andmeid kasutati informatsiooni kogumiseks ja järelduste tegemiseks, vaid kasutatakse saadud teadmisi õppeprotsesside kohta selleks, et parandada näiteks kursuse läbiviimist uuel semestril [4, lk 134–135].

Õpianalüütika uurimused ei pea ilmtingimata hõlmama tsükli kõiki nelja sammu, kuid reaalseid õppeprotsessi parandavaid tulemusi näeb vaid juhul, kui rakendatakse ka õpianalüütika tsükli sekkumise samm. Käesolev magistritöö keskendub peamiselt teisele ning kolmandale õpianalüütika tsükli osale, milleks on andmete kogumine ning mõõdikute leidmine ja analüüsi koostamine. Selleks, et edukalt õpianalüütika tsükkel lõpuni viia, annab autor ka soovitusi, kuidas õppeprotsessi paremini läbi viia [4, lk 134–135].

Üldjoontes on tudengite õpitulemuste ennustamine sarnane tavalisele õpetamisele klassiruumis, kus õppejõud märkab, millised tudengid on hätta jäänud oma ülesannetega ja keda tuleks aidata. Õpitulemusi ennustavate mudelite loomine võimaldab need samad printsiibid kanda üle veebipõhisesse õppekeskkondadesse nagu Moodle. Samas erinevad õpitulemusi ennustavad mudelid mõneti traditsioonilisest õpetamismeetodist. Ennustavad mudelid koosnevad hinnangulistest tõeüksustest ning kohati on neid raske korrektselt mõista ning nende põhjal häid valikuid teha. Ennustavate mudelite informatsioon ei ole ainult klassiruumis tunde läbiviiva õppejõu teadmiseks, vaid seda saab jagada ka nende isikutega, kes ei ole õpetamise kontekstiga otseselt seotud. Viimasena saab õpiedukust ennustavate mudelite baasil võtta automaatselt ühendust abi vajavate tudengitega, ilma et õppejõud peaks protsessi ise sekkuma [6, lk 687].

2.2 Moodle'i keskkond

Kõrgharidusasutused pakuvad aina enam veebipõhiseid ning veebipõhist ja tavaõpet kombineerivaid kursuseid, mille puhul ei saa rakendada tavapärasele õppimis- ja õpetamismeetoditele omaseid õpiaktiivsust ning -edukust kirjeldavaid heuristikaid. Vähenenud kontakt tudengitega muudab õppejõudude jaoks keerukaks üliõpilaste kursusel osalemise aktiivsuse hindamise. Üldiselt viiakse kursuseid läbi veebipõhistes õppekeskkondades, näiteks Moodle'i¹ õpikeskkonnas, mis koguvad andmeid tudengite tegevuste kohta õpikeskkonnas logidesse ja andmebaasidesse [5, lk 180–181].

Moodle on avatud lähtekoodiga veebipõhine õppekeskkond, mida kasutatakse 238 riigis. Moodle'i baasil on maailmas loodud 154 639 veebilehte, millel on kokku üle 200 miljoni kasutaja. Ainuüksi Eestis on registreeritud 139 Moodle'i keskkonda [7], [5, lk 180]. TalTech on samuti aktiivne Moodle õppekeskkondade kasutaja.

2.2.1 Õpianalüütika Moodle'i keskkonnas

Informatsiooni tudengite tegevusest veebipõhises õpikeskkonnas salvestatakse suurtes kogustes süsteemi andmebaasi ning seetõttu on andmete analüüsimiseks ning sisukate näitajate kättesaamiseks algandmetest vaja informatsiooni eelnevalt töödelda. Salvestatud andmete hulgas on nii informatsiooni tudengite ja õppejõudude suhtlusest kui

¹ <https://moodle.org/>

ka erinevate õppematerjalide kasutamisest, millest võib leida vihjeid selle kohta, kuidas ja millal tudengid oma ülesandeid ning kodutöid lahendavad ning kuidas ja kui aktiivselt nad kursusest osa võtavad. Kahjuks on suurest andmehulgast sisulise informatsiooni eraldamine keerukas ülesanne [2, lk 50–52].

Õpianalüütika koostamiseks Moodle'i keskkonnas puudub ühtne universaalselt toimiv lahendus. Mitmed ülikoolid on eritellimusel loonud süsteemid õpianalüütika jälgimiseks. Näiteks on Lõuna-Austraalia Ülikool loonud oma töötajatele programmi, milleks on töölaud, kus kuvatakse veebipõhise õpikeskkonna ning teiste ülikoolidega seotud veebipõhiste tegevuste informatsiooni. Seevastu kasutab Lääne-Sydney Ülikool oma tudengite jälgimiseks hoopis äridele loodud indikaatorite põhjal riske ennustavat tööriista. Edith Cowani Ülikool ja New Englandi Ülikool kasutavad süsteemi, mis haldab nii analüütikat, identifitseerimist kui ka ülikooli siseseid saatekirju. Kesk-Queenslandi Ülikool, Sydney Ülikool ning Uus-Lõuna-Walesi Ülikool on loonud ise süsteemi, mis hõlmab nii analüütikat kui ka identifitseerimist ning võimaldavad võtta ka tudengitega elektrooniliselt ühendust. Konkreetne Austraalia ülikoolidel põhinev näide illustreerib õpianalüütika maastiku mitmekesisust ning kahjuks ka keerukust [5, lk 180–181].

Moodle Dashboard on Moodle'i keskkonna poolt pakutav sisseehitatud töölaud, mis lubab kasutajatel nii graafiliselt kui ka teksti kujul kuvada Moodle'is tehtud päringute vastuseid. Andmeid on võimalik visualiseerida nii tabelite, ajajoonte kui ka eri tüüpi graafikute, näiteks sektor- ning tulpdiagrammide, kujul. Lisaks sellele võimaldab *Moodle Dashboard* erinevaid andmeplokke kombineerida ning lubab kasutajal töölauda vastavalt oma vajadustele kohendada. *Moodle Dashboard* pakub ka suurepäraseid andmete filtreerimisvõimalusi ning automaatselt genereeritud andmeeksporte. Kahjuks toetab *Moodle Dashboard* ainult Moodle'i versioone kuni 2.5, seega on tegu vananenud süsteemiga, kuna Moodle'i uusim versioon on 3.8 ning Moodle'i keskkonna logimissüsteemi uuendati täielikult juba Moodle'i versioonis 2.7 [8]. Lisaks sellele on töölaua algne seadistamisprotsess väga keerukas ning programm ei luba teha paindlikke keerukamaid andmepäringuid [2, lk 52–59].

Lisaks *Moodle Dashboardile* pakub Moodle'i keskkond ka sisseehitatud aruandluse tööriista (*Moodle reporting tool*), mida võib samuti vaadelda töölauana, mis analüüsib kasutajate tegevusi Moodle'i keskkonnas erinevates kontekstides, milleks on veebileht, kursus või aktiivsus. Tööriist kuvab informatsiooni kasutajate kommentaaride ja kursuse

aktiivsusnäitajate kohta, näiteks kõige aktiivsemad tudengid ja kõige populaarsemad kursused, ning muid logiandmeid, sealhulgas reaajas lisatavaid logiandmeid. Tööriist suudab nende andmete kohta luua ka graafikuid. Moodle'i aruandluse tööriist pakub kasutajatele spetsiifilisi funktsionaalsusi, mis ei pruugi sobida kokku ülikooli, kursuse või õppejõu vajadustega. Lisaks sellele on tööriista kasutamine keerukas ning nõuab õppejõududelt palju käsitööd, kuna sobiva analüütika leidmiseks tuleb tihti andmed arvutustabelina alla laadida ning käsitsi vajalikud päringud kirjutada ja graafikud luua [2, lk 53–59].

Moodle'i keskkonna arendajate kogukond on loonud mõned õpianalüütika pistikprogrammid. Neist esimene on GISMO¹, mis on interaktiivne graafiline jälgimistöriist, mis aitab õppejõududel mõista, kuidas tudengid kasutavad kursuse materjale. Lisaks GISMO-le on olemas ka *Moodle Engagement Analytics Plugin* (MEAP), mis näitab, kuidas tudengid suhtlevad Moodle'i keskkonnaga, põhinedes erinevate indikaatoritel, näiteks sisselogimisaktiivsus, foorumipostituste vaatamine ning ülesannete esitamise aktiivsus. Nende indikaatorite põhjal näitab MEAP tudengi kursuse läbikukkumise riskihinnangut. MEAP-i jaoks on loodud ka edasiarendus MEAP+, millele on lisatud näiteks riskirühmas olevatele tudengitele e-kirjade saatmine ning eraldiseisvate andmefailide keskkonda üleslaadimine täiendavaks analüüsiks [5].

Kahjuks pole Moodle'i keskkonna jaoks loodud pistikprogrammid alati sobivaks lahenduseks õpianalüütika koostamisel. MEAP² on kahjuks kasutatav ainult Moodle'i versioonidega 2.2 kuni 2.7, seega puudub programmi kasutamise tugi uutes Moodle'i keskkonna versioonides. MEAP+ on uurimistöös käigus loodud edasiarendus MEAP-ile, mis pole leidnud laialdast kasutust, ning kuna programm sõltub täielikult MEAP-ist, siis ei ole võimalik loodud täiendavat programmi kasutada uute Moodle'i keskkonna versioonidega. GISMO puuduseks on see, et ta ei paku tudengite õppe edukuse ning väljakukkumisohtu kohta piisavalt analüütikat, kuna programm annab ülevaate ainult seitsme spetsiifilise kategooria kohta, näiteks ligipääs materjalidele, testide (*quiz* Moodle'i kontekstis) ülevaade ning ülevaade ülesannetest [2, lk 54–59], [5].

¹ https://moodle.org/plugins/block_gismo

² https://moodle.org/plugins/report_engagement

Praegusel hetkel pakutavad lahendused ei ole sobilikud Moodle'i keskkonna analüütika koostamiseks ega vaatlemiseks. Olemasolevad lahendused kuvavad tihtipeale andmeid, tabeleid ja graafikuid, mis vastavad lihtsatele ning üldistele päringutele, kuid ei ole piisavad keerukamate järeluste tegemiseks ning ülevaate saamiseks. Programmid on tihti aegunud ning ei sobi kokku Moodle'i keskkonna uuemate versioonidega, lisaks sellele on paljud programmid kasutajatele arusaamatud ning keerukad, nõudes tihtipeale täiendava töö tegemist näiteks arvutustabeleid kasutades. Paljud ülikoolid on loonud või tellinud endale isiklikud õpianalüütika programmid, mida teised õppeasutused kasutada ei saa [2], [5, lk 180–181].

2.2.2 Moodle'i keskkonna logi seosed kursuste lõpptulemustega

Suurem osa Moodle'i keskkonna sisseehitatud tööriistadest, pisitkprogrammidest ning ka platvormiülelestest rakendustest, mis töötavad ka koos Moodle'iga, kasutavad Moodle'i keskkonna logiandmeid *mdl_logstore_standard_log* andmebaasi tabelist. Mitmed teadusuuringud on näidanud, et veebipõhistest õppekeskkondadest ja nende logidest saab koguda erinevaid andmeid, mis omavad mõningat tudengi lõpphinnet ennustavat väärtust. Sellisteks andmeteks on näiteks aruteluforumitega seotud andmed, õppekeskkonda sisselogimise tihedus ning veebikeskkonnas ülesannete esitamine ja tegemine [2, lk 52], [5].

Mbeya Teadus- ja Tehnikaülikoolis läbiviidud uuring näitas, et Moodle'i õppekeskkonnas kaastudengitega suhtlemine, ülesannete sooritamine ning kursuse foorumis postituste tegemine omavad tugevat mõju tudengite õppeedukusele. Uuriti kaht õppeainet: neist esimeseks oli rakendusbioloogia kursus, mille tulemusi mõjutasid kõige enam suhtlus kaastudengitega ning foorumipostituste tegemine. Teiseks kursuseks oli teenused ning paigaldus, mille tulemusi mõjutasid kõige rohkem ülesannete sooritamine ning foorumipostituste tegemine. Uuringust selgus, et õppematerjalide allalaadimiste arv, keskkonda sisselogimise tihedus ning Moodle'i keskkonnas veedetud aeg ei omanud märkimisväärset mõju kummagi kursuse tudengite lõpptulemustele. Töö järelustes toodi välja, et ülikooli õppejõud peaksid tudengite lõpptulemuste parandamiseks andma üliõpilastele rohkem ülesandeid ning harjutusi. Uurimus näitas, et see, kuidas üliõpilased kasutavad Moodle'i õppekeskkonda, mõjutab nende õppetulemusi. Samas toodi uuringus välja, et logid salvestavad tudengite käitumist, kuid ei seleta, miks üks või teine faktor on

tulemuste ennustamisel oluline, ning selle väljaselgitamiseks tuleks läbi viia täpsemaid teadustöid [9].

Uuringus “*Predicting students’ final performance from participation in on-line discussion forums*” kasutati 114 esimese kursuse tudengi Moodle’i keskkonnas tehtud foorumipostitustega seotud andmeid, kes võtsid IT valdkonnaga seotud õppeainet, ning leiti, et foorumipostituste andmete põhjal on võimalik ennustada nii kursuse lõpphinnet kui ka teha varast ennustust õppeedukuse kohta enne kursuse lõppu [10, lk 458].

Moodle’i andmete pealt leitud tudengite õpiedukust ennustavate näitajate võrdlev uuring leidis, et vanusel, sool ning kultuurilisel taustal pole seost õpiedukusega. Kasutati ka MEAP-i indikaatoreid, mis üldiselt ei ennustanud tudengite lõpptulemusi, erandina toodi välja kursusele sisselogimiste arv, mis omas ennustavat väärtust. Moodle’i logidest leitud näitajatest kasutati ennustamiseks nelja: kursusele sisselogimiste arv, kursuse lugemismaterjali läbitöötamine, testidest osavõtmise aktiivsus ning testide tulemused. Üldiselt olid ennustustulemused nende näitajate põhjal väga suure varieeruvusega, kuid neist parimaks osutus testidest osavõtmise aktiivsus, mis oli näiteks tervelt 2,4 korda parem testitulemuste kasutamisest lõpptulemuste ennustamisel. Uurimisgrupis oli 228 New Yorgi Osariigi Ülikooli (SUNY) bakalaureusetaseme üliõpilast, kes õppisid kahel erineval kursusel: professionaalsuse seminaril ning inimressursside juhtimise kursusel [11, lk 922, 934–935].

Dietz-Uhler ja Hurn Ohio Miami Ülikoolist toovad oma analüüsis välja andmed, mida võib õpianalüütika koostamiseks kasutada, selleks et parandada ning ennustada tudengite õpiedukust. Nad jagavad andmed kahte gruppi: virtuaalse õpikeskkonna poolt genereeritud andmed ja õppejõu poolt genereeritud andmed, mida võib samuti leida virtuaalsest õpikeskkonnast. Virtuaalse õpikeskkonna poolt loodud andmeteks on: mitu korda on materjale kasutatud, materjalide kasutamise kellaaeg ning kuupäev, loodud arutelupostituste arv, loetud arutelupostituste arv ning see, millist tüüpi materjale on kasutatud. Õppejõu poolt genereeritud andmeteks on hinded postitustele aruteluforumis, hinded ülesannetele, hinded testidele, kursuse lõpphinded, kui palju küsimusi küsiti õppejõu poolt aruteluforumites, milliseid küsimusi küsis õppejõud aruteluforumis, kui palju e-kirju tudeng saatis õppejõule [12, lk 20–21].

Autorid tõid välja ka selle, et ühes nende uurimisrühma poolt läbiviidud uuringus, kus ennustati lõpphindeid veebipõhisele sissejuhatavale kursusele psühholoogias, selgus, et kursuse jooksul tehtud testide ning ülesannete tulemused ennustasid tugevalt lõpphinnet. Eriti tugevalt mõjutasid sellel kursusel lõpphinnet kaks esimest läbiviidud eksamit: test, mis viidi läbi enne kursust läbitud materjalide teemal, ning ülesanded, mida tudengid lahendasid kursuse teises pooles. Kokku moodustasid need näitajad 98% kursuse lõpphinnete dispersioonist. Virtuaalse õpikeskkonna poolt genereeritud andmetest mõjutas tulemusi kõige rohkem semestri teises pooles tehtud foorumipostituste arv [12, lk 20–21].

Artiklis *“Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success”* uuriti üheksat Moodle’i kursust ning leiti, et tudengid kasutavad erinevaid virtuaalsete õppekeskkondade funktsionaalsuseid erinevatel kursustel. Mitte ainult ei erinenud kasutatud funktsionaalsused, vaid ka see, kuidas ja mis mahus erinevaid funktsionaalsusi eri kursustel kasutati. Näiteks toodi artiklis välja, et vaadeldud bioloogia ja kommunikatsiooni kursuste tudengid kasutasid teiste kursuste tudengitest märgatavalt rohkem arutelufoorumeid. Lisaks tõid autorid välja ka selle, et kursustel, kus oli palju osalise õppekoormusega tudengeid, esines märksa rohkem seda, et tudengid hakkasid Moodle’i keskkonda kasutama tavapärasest hiljem või ei teinud seda üldse. Uuringust selgus, et kursusepõhiste ennustusmudelite täpsus lõpptulemuste leidmisel oli märgatavalt suurem üldistest mudelitest nii protsentuaalse hindamisskaalaga kui ka arvestuslikes õppeainetes [13].

Lisaks eelnevale selgus uuringust, et Moodle’i keskkonnast kogutud andmed ei olnud kõigi kursuste puhul headeks näitajateks õpitulemuste ennustamisel, erandiks osutus graafilise disaini kursus. Seitsemel juhul üheksast polnud kursusepõhiste ennustusmudelite korral olulisteks näitajateks kursuste sisselogimisandmed ning materjalide vaatamine, mis olid määravad üldistes mudelites. Seega on üldistes mudelites olulised tunnused, mis esinevad paljudes kursustes, kuid ei ole alati konkreetse kursuse aspektist olulised. Esinesid tunnuseid, mis ennustasid protsentuaalseid õpitulemusi hästi osadel vaadeldud kursustest, kuid mitte teistel kursustel. Uuringust selgus, et osade Moodle’i keskkonna funktsionaalsuste kasutamine oli positiivse mõjuga mõningatel kursustel, kuid mõjuvad hoopis negatiivselt õpitulemustele teistel. Seega on kursuseüleste õpiedukust ennustavate näitajate leidmine keerukas ülesanne. Autorid rõhutavad, et enne virtuaalse õpikeskkonna

logide peal ennustuste tegemist tuleb mõelda sellele, kuidas õppurid keskkonda kasutavad. Virtuaalse õpikeskkonna erisustega arvestamine on oluline ning vaeleavestuste korral võivad virtuaalsele õpikeskkonnale antud omadused mudeleid mõjutada [13].

2.3 Õpianalüütika Tallinna Tehnikaülikoolis

TalTechis on varasemalt keskendunud peamiselt tudengite ülikoolist väljakukkumise ennustamisele, mitte õppeainete läbimise ennustamisele, mis mõjutab tugevalt ka üldist ülikoolist väljalangemist. Tudengite Tallinna Tehnikaülikoolist väljalangemise ennustamiseks on varasemalt toetunud peamiselt tudengite üldisikuandmetele, sealhulgas näiteks sugu ning kodakondsus, ning ka sooritustulemustele, mille alla kuuluvad semestrite kaalutud keskmised hinded ning erinevatel semestritel kogutud EAP-de arv [3], [14].

Brenda Uga katsetas oma bakalaureusetöös “TTÜ tudengite väljalangemise ennustamine: tõenäosuse arvutamine masinõppe meetodite abil ning tulemuste kuvamine veebirakenduses” tudengite väljalangemise ennustamiseks mitmeid erinevaid masinõppealgoritme, millest kõige edukamaks osutus otsustuspuu algoritm. Erinevate masinõppe mudelite täpsused jäid vahemikku 60-90%. Kõige enam mõjutas ennustust tudengite kogutud EAP-de arv [14].

Kristin Ehala kasutas oma magistritöös “Kontekstipõhine õppeedukuse monitoorimise mudel” tudengite väljalangemise ennustamiseks kaht meetodit: üldist ning kontekstipõhist meetodit. Kontekstipõhises meetodis kasutas autor kontekstina õppekava ning vaatluse all olevat semestrit. Üldise meetodi korral kasutas autor tudengeid kirjeldavaid andmeid. Toetudes Brenda Uga tulemustele kasutas autor oma töös samuti otsustuspuude meetodit. Uuringust selgus, et kontekstipõhine mudel oli ennustamises halvem kui üldine mudel – kontekstipõhine mudel ennustas tudengite lõpetamist 82% täpsusega ning üldine mudel 91% täpsusega [3].

Kadri Umbleja tegi läbiviidud uurimuses tudengite õppeedukuse kursusepõhised ennustusmudelid, mis loodi kasutades andmekaebet koos regressioonanalüüsiga, peamiste komponentide analüüsi ning sümboolsete histogrammide väärtuste andmete hierarhilist klasterdamist. 35 tudengite tegevusi kirjeldavast tunnusest valiti välja 9

tähtsaimat, mida kasutati mudelis. Nendeks tunnusteks olid tehtud ülesannete arv, aktiivsete päevade arv, ülesannete korrektsus, ülesannete keerukus, abimaterjalide vaatamiste arv, ülesannete esitamise kellaaeg, käimasolev kursuse protsess, *HomeLabKiti* laenutamiste arv ning tunnis tehtud testide arv. Hierarhilise klasterdamise käigus avastati, et ei ole võimalik luua üht ennustumudelit kõigi kursuste jaoks, kuid on võimalik luua kursusekeskne mudel. Sarnase õpikäitumisega klastreid leiti kõigi viie loodud mudeli korral. Klasteriteks olid näiteks tudengid, kes alustavad ning lõpetavad ülesanded varakult; tudengid, kes teevad kõik ülesanded korraga kursuse lõpus; tudengid, kes teevad väga vähe; tudengid, kes katsetavad kursust semestri alguses ning hakkavad uuesti pingutama semestri teises pooles ja üliõpilased, kes hakkavad tegutsema alles semestri teises pooles. Tudengite kuuluvus klasteritesse muutus semestri jooksul. Mudelitest valiti täpsemaks testimiseks välja suurima andmehulgaga kursus, millel oli 14 erinevat klasterit. Alates semestri neljandast nädalast ennustati tudengite hindeid 80,91% täpsusega. Ülejäänud juhtudel oli ennustustulemus ühe hinde võrra madalam tegelikust sooritustulemusest [15, lk 187–188].

Kahes eelnevalt läbiviidud uuringus on tudengite väljalangemist ennustades keskendunud peamiselt tagantjärgi kogutud andmetele, näiteks semestri keskmine hinne ning kogutud EAP-de hulk. Seetõttu ei ole tegelikult võimalik loodud mudeleid kasutada selleks, et hoida ära tudengite väljalangemist ülikoolist. Kolmanda varem läbiviidud uuringu puhul on probleemiks keskendumine kursuse ülesehituse keskele sooritustulemuste ennustamisele [15, lk 187–188]. Varasemalt TalTechis läbiviidud uuringuid ei saa kahjuks kasutada tudengite jooksvaks abistamiseks õppeainetes, kus üliõpilastel on tekkinud probleemid [3], [14].

2.4 Õppimisega seotud enesehinnanguliste psühholoogiliste tegurite küsitlus

Autor kasutab magistritöös lisaks Moodle'i logide andmetele ka üliõpilaste hulgas läbi viidud õpipädevuse, -käitumise, -motivatsiooni ja õppimisega seotud emotsioonide ning õppimisega seotud ja enesekohaste uskumuste teemalise küsitluse tulemusi. Käesolevas alapeatükis annab autor ülevaate küsitluse teoreetilisest taustast, tuues välja valdkonnad, mille kohta läbiviidud küsitluses tudengitele küsimusi esitati.

2.4.1 Õpipädevus

Õpipädevus hõlmab tudengite oskust leida sobivat informatsiooni, planeerida oma õppetööd ja loodud plaani järgida, organiseerida oma õppekeskkonda, oskust kasutada õpitut erinevates olukordades ning kontekstides – sh probleeme lahendades – ja võimekust analüüsida oma tugevaid ning nõrku külgi. Eduka õppimise aluseks on nii kognitiivsete kui ka metakognitiivsete õpioskuste olemasolu. Kognitiivne õpistrateegia koosneb teadmiste säilitamise ning ammutamise oskustest. Metakognitiivne õpistrateegia koosneb aga õpiprotsessi korraldamisega seonduvatest oskustest [16].

2.4.2 Õppimisega seotud emotsioonid

Erinevad uuringud on tõestanud, et tudengid tunnevad akadeemilises kontekstis väga mitmekülgsid emotsioone. Traditsiooniliselt on uuritud peamiselt testidega seotud ärevust, kuid tegelikult tunnevad tudengid võrdselt nii negatiivseid kui ka positiivseid õppimisega seotud emotsioone. Emotsioonid on tugevalt seotud tudengite enesejuhitud õppimise komponentidega nagu huvi, motivatsioon, õpistrateegiad ning sisemine ja väline regulatsioon. Emotsioonid ennustavad ka tudengite õppeedukust, seega akadeemilisest saavutustes realistliku ülevaate saamiseks tuleks arvesse võtta ka tudengite emotsioone. Negatiivsed emotsioonid ei mõjuta õpiedukust ainult negatiivselt ning vastupidi ei mõjuta ka positiivsed emotsioonid akadeemilist edukust ainult positiivselt. Sellist käsitlust tuleks vältida, kuna positiivsed emotsioonid on aeg-ajalt kahjulike mõjudega ning negatiivsed emotsioonid nagu ärevus või häbi võivad osutada kasulikeks [17, lk 91, 102–103].

Kõige sagedamini esinevaks emotsiooniks oli ärevus, mida esines lausa 15% kuni 25% erinevates uuringutes osalenud tudengitest. Ärevust mainiti kõige rohkem seoses eksamitega, kuid tihtipeale ka seoses tööga klassiruumis või kodus. Tudengid väljendasid ka seda, et tihtipeale esinevad mitmed emotsioonid korraga, näiteks muretsemine läbikukkumise pärast käis käsikäes eksamitega seotud ärevuse, häbi ning lootusetusega. Lisaks ärevusele esines negatiivsetest emotsioonidest kõige rohkem viha, häbi ning igavust. Positiivsetest emotsioonidest mainiti kõige rohkem õppimisrõõmu, lootust, uhkust ning kergendust. Lisaks mainisid paljud tudengid metaemotsioone, näiteks et nad olid enda peale pahased, sest tundsid seoses eksamitega ärevust. Mõningatel nendest tudengitest aitas viha ärevustest üle saada. Seega võivad metaemotsioonid aidata tudengitel toime tulla negatiivsete emotsioonidega [17, lk 91–94].

Õppimisega seotud ärevus võib väljenduda ka konkreetsete teemadega seotud ärevuse näol. Näiteks on õpilaste ning ka tudengite hulgas levinud matemaatikaga seotud ärevus. Käesoleva magistritöö raames läbiviidud küsitluse tulemused viitasid sellele, et matemaatikaga seotud ärevus kattub suurel määral ka programmeerimisega seotud ärevusega. Väga suure matemaatikaärevusega õppurid väldivad matemaatikaga seotud tegevusi, mis vähendab omakorda nende matemaatikaalast võimekust ning võib mõjutada ka karjäärivalikuid. Matemaatikaalase ärevusega isikud eelistavad näiteks erialasid, kus viiakse läbi vähem matemaatikaga seotud kursuseid ning see võib viia ka juba ülikooli astunud tudengi eriala vahetamiseni. Matemaatikaärevusega tudengid valivad vähem matemaatikatunde kui teised üliõpilased ja saavad ka halvemaid tulemusi. Lisaks eelnevale on sellised tudengid ka matemaatika osas negatiivselt meelestatud ning neil puudub usk enda matemaatilistesse võimetes, seetõttu suudavad nad omandada neile õpetatavatest matemaatikateadmistest vähem [18].

Matemaatikaalane ärevus mõjutab kognitiivseid protsesse, segades parasjagu käimasolevaid tegevusi mälus. Ärevus võib avalduda ka emotsionaalselt, näiteks kartuse ja pingena või isegi füüsiliselt, näiteks värisevate käte näol. Uuringud on aga näidanud ka seda, et kasutades matemaatikaärevust leevendavaid meetmeid paranevad tudengite matemaatikatumused märgatavalt. Ärevusolekus matemaatikaülesandeid lahendavad inimesed keskenduvad rohkem pealetükkivatele negatiivsetele mõtetele kui parajasti lahendatavatele ülesannetele, mis mõjutab negatiivselt nende tulemusi [18].

Moodne ühiskond on aina enam andmete ning tehnoloogiale orienteeritud ning kuna kasutusel olevad õpetamismeetodid suurendavad tihti matemaatikaalast ärevust, ei ole piisavalt inimesi, kellel oleks adekvaatne matemaatikaoskus. Matemaatikaga seotud ärevus rohkem levinud naiste hulgas, mis selgitab teataval määral ka seda, miks naised on vähem tehnoloogia ning ITK valdkondadega seotud erialadel. Sooline erinevus on suurel skaalal küll väike, kuid avaldub selgelt väiksemaid gruppe, näiteks üliõpilasi vaadeldes [18].

2.4.3 Õppimisega seotud ning enesekohased uskumused

Ühiskonnas on levinud valearusaam, et inimestel on kaasasündinud võimed, mis määravad selle, kui palju ja mida nad on suutelised õppima. Seetõttu alahindab ühiskond indiviidide võimet saavutada oma õpieesmärged ja pühendunult treenides, harjutades ja kogemusi omandades paremaks saada. Pannes liiga palju rõhku kaasasündinud oskustele

ning alahinnates õppimisvõimet, usuvad õppurid tihti, et nende võimel on piirid ja nad alahindavad oma suutlikkust omandada uusi teadmisi. Õppuritel peab olema arenemisele orienteeritud mõtlemisviis, mitte fikseeritud mõttelaad [19, lk 436].

Samas ei saa eitada ka seda, et inimeste vahel on olulisi erinevusi, kuid need pole kaasasündinud vaid kogemustel põhinevad ning keskkonnast tingitud. Uued teadmised ehitatakse üles juba olemasoleva informatsiooni peale, seega on õppimise alguses olevate teadmiste hulk oluline faktor. Õppimisega seotud uskumustele, soovidele, eesmärkidele ning ootustele on tugev mõju ka perekonnal, kultuurilisel taustal ja traditsioonidel [19, lk 436].

Ühiskonnas levib ka valearusaam selle kohta, et õppimine peaks olema lihtne protsess. Efektiivne õppimine võib olla lõbus, oma saavutustega rahulolu tekitav ning isegi aega säästev protsess, kuid lihtne on korrektne õppimine harva. Pingutamine on õppimise ning uute teadmiste omandamise juures äärmiselt vajalik. Õigesti õppimine, piisav pingutamine ja pingutuse korrektne juhtimine tagavad akadeemilise edukuse, sest pingutamine ei näita mitte ainult eesmärkidele orienteeritust vaid ka pidevat õpistrateegiate kasutamist [19, lk 436], [20, lk 27].

Lisaks pingutuse olulisusele õppimisel on tähtis ka tudengite enesetõhusus. Enesetõhusus on enda võimekuse hindamine mingi eesmärgi saavutamiseks – uskumus, et seatud eesmärki või ülesannet on võimalik edukalt sooritada. Enesetõhusus hõlmab endas nii enda võimekuse hindamist kui ka usku juba varasemalt omandatud oskustesse, mis aitavad uusi probleeme lahendada [20, lk 13–14].

2.4.4 Õpikäitumine

Õpikäitumine kirjeldab seda, kuidas õpilased akadeemilises kontekstis käituvad ning oma ülesannetele lähenevad. Tudengid teevad seda väga paljudel erinevatel viisidel. Õpikäitumise valdkond uurib näiteks, kas tudengid väldivad ülesannete lahendamist või mitte ning kas tudengite õppimisstrateegiad on tõhusad [21].

Uuringutest on selgunud, et tudengite ootused oma akadeemilise edu osas ennustasid nende akadeemilisi saavutusi ning rahulolu oma sooritusega. Kui tudengitel oli usku enda akadeemilistesse oskustesse, siis olid nad ka edukamad. Lisaks sellele ennustab ülesannete tegemise vältimine madalat akadeemilist edukust ning madalat rahulolu

õpitulemustega. Ülesannete vältimine isikuomadusena ei mõjuta tudengite õpitulemusi, see väljendub vaid ainult juhul, kui ülesannete vältimine rakendub õpikäitumises. Madalad akadeemilised saavutused ning vähene rahulolu tulemustega soosivad aga omakorda jällegi ülesannete vältimist. Seega on tegu negatiivse käitumistsükliga, mida pole lihtne murda, sest negatiivsed akadeemilised tulemused ja rahulolu puudumine suurendavad õppimisega seotud ärevust ja juhivad tudengeid kordama käitumismustreid, mis neid kehvade tulemusteni viisid. Ülesannete vältimisega õpikäitumises kaasneb tihti ka teisi negatiivseid mõjusid õpikäitumisele, näiteks tudengid, kes väldivad ülesannete tegemist sageli, ei pinguta eriti oma tulemuste parandamiseks ka pärast akadeemilist läbikukkumist. Seevastu uskusid positiivsed tulemused saanud tudengid, et nende edu võtmeks olid nende endi oskused [21].

Kõrge sooritusmotivatsioon ning sellega seotud õpistrateegiad, optimism, meisterlikkusele orienteeritus, töökus ja ülesannetele keskendumine mõjutavad akadeemilisi tulemusi ning motivatsiooni positiivselt. Saavutatud akadeemiline edu innustab tudengeid ka tulevikus tegelema akadeemiliste väljakutsete ja keeruliste ülesannetega. Hirm läbikukkumise ees, madal panus õpingutesse ning ülesannete vältimine suurendavad läbikukkumise tõenäosust. Tihti väldivad end akadeemiliselt ebakindlalt tundvad tudengid teadlikult keerukamaid akadeemilisi olukordi ja ülesandeid, sest nad kardavad läbikukkumist [21].

Tõhusaks õpistrateegiaks informatsiooni meeldejätmisel on näiteks uute mõistete tähendusele keskendumine, uute ideede ja teadmiste seostamine juba omandatud kontseptidega ning õpitavate teadmiste organiseerimine. Hea õpistrateegia on ka näiteks õppimissessioonide laiali jaotamine pikema ajaperioodi peale teemade kaupa, mitte korraga suures mahus materjali õppimine. Head õppimisprotsessi iseloomustab ka oma töö pidev jälgimine ning selle põhjal korrektsete valikute tegemine järgnevate tegevuste osas, näiteks millist materjali tuleks järgmisena omandada ja kuidas seda teha [19, lk 421–422].

Tihti ei saa tudengid aru, et nende õpistrateegiad on mittetõhusad. Seda esiteks seetõttu, et õppimisoskustele ei pöörata haridussüsteemis piisavalt tähelepanu. Teisalt on oluliseks mõjutajaks ka tagasivaatav perspektiiv ning vaeleusaamad sellest, kuidas informatsiooni lihtsam meelde jätta oli. Näiteks väitis ühes läbiviidud uuringus 90% tudengitest õpingute ajal, et õppimissessioonide laiali jagamine pikema perioodi peale on kasulik, kui kõige

korraga õppimine. Pärast aktiivse õppetöö lõppu vastasid 71% samadest tudengitest, et materjali korraga õppimine oli kasulikum strateegia. Korraga õppimine tundub tudengitele lihtsam, kuid see pole sisuliste tulemuste osas parem, kuna strateegia ei soosi õpitu kordamist ning püsivate seoste loomist. Isegi õpikute koostamisel ei arvestata tihti heade õpistrateegiatega, näiteks tihti on kogu teemat käsitlev informatsioon ühes peatükis, mitte ei käsitleta seda korduvalt õppeprotsessi vältel [19, lk 434].

Tihti on õppurid ka liiga enesekindlad oma võimes õpitud informatsiooni tulevikus meelde tuletada ning seetõttu ei korrata õpitavat materjali piisavas mahus. Mida enesekindlam ollakse selles võimes, seda kehvemad on tulemused informatsiooni meeldetuletamisel testides. Lisaks sellele on tõenäoline, et kui tudengid teevad samasugust testi kaks korda, siis need teadmised, mille meeldetuletamiseks läks esimest testi sooritades kaua aega, on teist testi tehes paremini meeles [19, lk 427, 430–431, 433].

Tihti teevad õppurid valesid õpistrateegilisi otsuseid, kuna need otsused tuleb langetada õppimisprotsessi ajal. Seega kiputakse eelistama strateegiaid, mis tagavad parimad tulemused õppimise vältel. Seetõttu ei proovita näiteks probleemidele kõigepealt iseseisvalt vastuseid leida või siis ennast pärast materjali õppimist testida. Lisaks sellele välditakse õppides vigade tegemist. Oma vigadest õppimine on aga oluline õppeprotsessi osa ning vigade tegemine õppimise käigus aitab parandada õpitulemusi. Vead tekivad tihti ka keerukamaid probleeme lahendades ning näidisülesannetest pisut erinevaid probleeme lahendades. Selliste ülesannete lahendamine aitab aga paremini meeles hoida ning kasutada õpitud informatsiooni [19, lk 434–435].

Ka ühiskonnas levinud arusaamad ning arvamused mõjutavad inimeste õpistrateegiaid. Üheks ühiskonnas levinud valearusaamaks on, et lastele ning ka täiskasvanutele ei pea õpetama, kuidas hallata enda õpitegevust. Koolide ning ülikoolide sisseastumiskatsetel pannakse tihti proovile õpilaste teadmised mingis konkreetses valdkonnas, näiteks matemaatikas, kuid ei panda proovile õpilaste õppimisoskuseid, mida nad oma edaspidisel haridusteel kindlasti vajavad [19, lk 419].

2.4.5 Õpimotivatsioon

Uuringute tulemused on näidanud, et huvi mõjutab tähelepanu, eesmärkide seadmist ja õpistrateegiaid, seega on huvi oluliseks näitajaks hariduspraktika parandamisel. Huvi

ennustab eriala, mille tudengid endale valivad, ning kombineeritult väliste faktoritega aitab huvi ennustada ka õppeedukust [22, lk 121–123].

Üldiselt on uuringud tõestanud, et õppejõud saavad aidata tudengitel keerukate ülesannete korral hoida tähelepanu, näiteks pakkudes tuge, mis aitab tekitada situatsioonilist huvi, või andes tagasisidet, millest lähtuvalt saavad tudengid küsida uudishimust tekkinud küsimusi. Õppejõud võivad luua rohkem olukordi, kus tudengid saavad küsida uudishimust lähtuvaid küsimusi, või siis luua õppematerjale, mis soosivad probleemide lahendamise oskuse arendamist ning strateegiate genereerimist. Varastes huvi tekkimise faasides on äärmiselt olulisel kohal tegevusega seotud positiivsed tunded ning tugev teadmiste baas, mis aitavad tudengitel materjalist aru saada, eesmärged seada ning õppida [22, lk 121–123].

Seda illustreerib 257 muusiku hulgas läbiviidud uuring. Äärmiselt olulisel kohal oli huvi tekkimisel esimene muusikaõpetaja, kes pidi olema sõbralik, jutukas ning julgustav, kuid samaaegselt hea muusik, kes oskas hästi oma instrumenti mängida ning õigeid tehnikaid ette näidata. Sarnaseid tulemusi on uuringutes leitud ka klassiruumi kontekstis, õpetajad peavad olema julgustavad ning tugevad oma valdkonnas. Õpetajate tugi õpilaste huvidele on äärmiselt oluline. Huvi arengu algfaasis on oluline pakkuda tuge õpilaste erinevatele oskustele ning keskenduda nende rakendamisele. Seda saab teha näiteks valikülesandeid või innovatiivseid ülesandeid pakkudes. Oluline on pakkuda tuge ka ülesande lahendamise ajal, et ülesande valmides tekiks kompetentsi ning eduka soorituse tunne. Selleks, et õpilaste huvi püsiks, on vaja pidevalt nende arengut toetada. Soosiv keskkond muudab välised toed aja jooksul sisemiseks huviks. Näiteks muusikute hulgas läbiviidud uuringust selgus, et õpilased, kes said õpingute algfaasis mittedünaamiliselt ning ilma suuremate oskusteta tutvuda muusikaga, olid tulevikus talendikamad [22, lk 121–123].

Teises huvi tekkimise faasis oli muusikute uuringus olulisel kohal detailidele keskendumine ning oskuste omandamine. Kolmas huvi tekkimise faas on keerukaim periood, kuna õpilaste huvi kõikumise korral tuleb õpetajatel endiselt pakkuda tuge. Õpilased pidid selles faasis arendama oma ideid seoses sellega, kes nad on muusikuna, mis on nende spetsialiseerumine ning kuidas muusika nende ellu sobitub. See protsess väljendub õpilaste seas mitmetel viisidel ning eri aja vältel. Teised sarnased uuringud toetavad huvi tekkimise faase ning on toonud välja ka selle, et õpilased ei näita üles huvi, kui nad tunnevad ennast ebamugavalt või ohustatult [22, lk 121–123].

Huvi saab areneda, kuid on ebatõenäoline, et õppurite huvi areneb välja isoleeritud keskkonnas. Õpilased, kellel on tekkimas isiklikud huvid mingi valdkonna vastu, küsivad uudishimust ajendatud küsimusi, mis annavad võimaluse omandada uusi teadmisi ning kindlustada neid mõistes paremini seda, millest saadakse juba aru. See tagab omakorda püsivuse keerukates olukordades [22, lk 121–123].

Mitmed uuringud on näidanud, et sisemine ja integreeritud õpimotivatsioon erinevad oma mõju poolest psühholoogilisele heaolule ning eesmärkide täitmisele [23, lk 759].

Sisemiselt reguleeritud õpimotivatsioon ennustas ühes läbiviidud uuringus kindlat õpilaste vaimset heaolu, mis oli eraldiseisev õpilaste akadeemilistest saavutustest. Sama uuring tõestas ka, et integreeritud õpimotivatsioon ennustas hästi õpilaste õppeedukust, olles samaaegselt eraldiseisev sisemisest regulatsioonist, integreeritud õpimotivatsiooniga õpilased said paremaid hindeid. Lisaks eelnevale tõestati, et integreeritud õpimotivatsioon ennustas koostöös akadeemilise edukusega õpilaste vaimset heaolu, seega sidusid integreeritud õpimotivatsiooniga õpilased oma vaimset heaolu õpiedukusega. Õpilased, kellel olid paremad hinded, olid ka parema vaimse heaoluga [23, lk 759].

Teises sarnases uuringus leiti samuti, et sisemine õpimotivatsioon mõjutas tudengite psühholoogilist heaolu, sest tudengid, kes läbisid eksperimentaalse sisemise regulatsiooni koolituse, raporteerisid 10 päeva pärast eksameid paremat vaimset heaolu. Samuti leiti uuringust, et sisemise regulatsiooniga tudengite vaimse heaolu ning akadeemiliste saavutuste vahel polnud seost. Seevastu aktiivne panustamine oma õpieesmärkide saavutamiseks oli seotud psühholoogilise heaoluga, sest sisemise regulatsiooniga tudengite puhul oli õppimisele pühendatud aeg seotud rahuloluga elus toimunud muutuste aspektist. Uuring näitas samuti, et integreeritud õpimotivatsiooniga tudengid said kursuse lõpus paremad tulemused ning see näitaja oli eraldiseisev sisemisest regulatsioonist [23, lk 759].

Soorituse vältimisele orienteeritud tudengite õpieesmärkideks on oodatava läbikukkumise vältimine ehk negatiivne määratlemine. Soorituse vältimisele orienteeritud õpieesmärgid on seotud madalama motivatsiooni ning kehvemate sooritustega. Seevastu sooritusele orienteeritud õpieesmärgid võivad suurendada akadeemilist motiveeritust ning parandada õpikompetentsi [24, lk 365].

2.5 Ennustamine

Antud alapeatükis kirjeldab autor valitud masinõppealgoritmi ning põhjendab selle valikut käesoleva töö kontekstis. Lisaks sellele toob autor välja ka mudelite valideerimiseks kasutatud meetodid.

2.5.1 Otsustuspuu

Käesoleva magistritöö eesmärgiks ei ole leida parimat võimalikku masinõppealgoritmi õppeedukuse ennustamiseks, vaid uurida fikseeritud masinõppealgoritmi abil, millised sisendiks valitud andmed saavutavad parima ennustustäpsuse. Seega soovib autor teada saada, millised andmehulgad sisaldavad kõige rohkem kasulikku informatsiooni õppeedukuse ennustamiseks.

Otsustuspuu masinõppe algoritmi valiku peamiseks põhjuseks oli otsustuspuude interpreteeritavus. Otsustuspuud jagavad andmed iteratiivselt ja ahnelt alamhulkadeks. Klassifitseerimisülesannete korral, nagu seda on käesolevas magistritöös lahendatavad probleemid, valitakse puu hargnemised selliselt, et minimeerida tulemuseks saadud andmehulkade entroopiat või siis Gini ebapuhtust [25]. Interpreteeritavuse lihtsustamiseks genereeris autor ka kõigi loodud masinõppemudelite otsustuspuud kujutavad joonised. Genereeritud otsustuspuude graafidel on võimalik näha, kuidas ja milliseid näitajaid kasutas otsustuspuu andmete gruppideks jagamisel. Otsustuspuude interpreteerimiseks tuleb vaadata puu representatsiooni näiteks graafi kujul ning alustada juursõlmest. Seal saab liikuda edasi järgmiste sõlmede juurde ning servade abil saab tuvastada vaadeldavaid alamhulki. Lehtsõlm sisaldab ennustatud tulemust. Samuti saab otsustuspuude abil arvutada välja tunnuse olulisuse. Selleks tuleb läbida kõik hargnemised puus, kus kasutati tunnust, ning leida, kui palju on tunnus vähendanud dispersiooni või Gini indeksi väärtust võrreldes vanemsõlemega. Kõikide olulisuste summa tuleb viia protsentuaalsele skaalale, et tunnuse olulisuste summa moodustaks osa mudeli üldisest olulisusest [26].

Lisaks sellele, et otsustuspuud on interpreteeritavad, osutus otsustuspuude kasutamine edukaks eelnevalt TalTechis tehtud uuringutes, seda nii Brenda Uga bakalaureusetöös kui ka Kristin Ehala magistritöös [3], [14].

2.5.2 Masinõppemudelite hindamine

Masinõppemudelite ennustustulemuste hindamiseks kasutatakse scikit-learn'i *DecisionTreeClassifier* (otsustuspuu klassifikaator) klassi meetodit *score* (skoor), mis saab sisendiks testandmehulga ning testandmete klassid. Meetod skoor tagastab etteantud testandmete ja klasside keskmise ennustustäpsuse. Parimaks skoori väärtuseks on 1, mis tähistab 100% ennustustäpsust ning halvimaks on 0, mis tähistab ennustustäpsust 0% [27].

Lisaks *score*'ile kasutab autor masinõppealgoritmide ennustustulemuste võrdlemisel iga klassi jaoks leitud *precision score*'i ehk täpsusskoori. Täpsusskoori leidmiseks tuleb klassi tõeste positiivsete ennustusväärtuste arv jagada tõeste positiivsete ja valepositiivsete väärtuste arvude summaga. Klassi täpsusskoor näitab, kas mudel suudab klassi mitte kuuluvaid juhtumeid määratleda mitte klassi kuuluvatena või määrab nad klassi, kus need väärtused ei tohiks olla. Parimaks võimalikuks täpsusskoori väärtuseks on 1 ning halvimaks 0 [28].

Iga klassi väärtuse kohta leidis autor ka *recall score*'i ehk saagisskoori, mis näitab klassifitseerimismudeli võimet leida kõik klassi kuuluvad juhtumid. *Recall score*'i leidmiseks tuleb tõeste positiivsete ennustusväärtuste arv jagada tõesete positiivsete ja valenegatiivsete väärtuste arvude summaga. Sarnaselt täpsusskoorile on parimaks võimalikuks väärtuseks 1 ning halvimaks 0 [29].

Viimasena leidis autor iga mudeli kõigi klasside jaoks tõesed positiivsed, tõesed negatiivsed, valepositiivsed ja -negatiivsed ennustatud klasside arvud. Selleks kasutas autor *multilabel confusion matrix*'it [30].

2.6 Klasterdamine

Veebipõhistes õppekeskkondades on varasemalt klasterdatud tudengite poolt tehtud foorumipostitusi. Foorumeid kasutatakse kursuste raames näiteks selleks, et tudengid saaksid arutleda käsitletud teemade üle ning avaldada oma seisukohti seoses nendega. Tudengi poolt tehtud foorumipostituste arv mõjutab kursuse lõpphinnet. Põhja-Carolina Osariigi Ülikoolis läbiviidud uuringus püüti leida klastreid sarnastele foorumipostitustele. Kahjuks ei kasuta paljud TalTechi Moodle'is läbiviidavad kursused aktiivselt foorumeid [31, lk 146–147].

Seoulis asuvas suures naiste eraülikoolis läbiviidud uuringus klasterdati 612 Moodle'i kursuse andmeid, et tuvastada, kuidas õppejõud on kursused ülesehitanud ning milliseid võimalusi olemasolevatest tegelikult aktiivselt kasutatakse. Suurem osa kursustest jäid tudengite arvu poolest vahemikku 20 kuni 100 üliõpilast. Vaadeldi tudengite ja õppejõudude sisselogimiste tihedust, kas tudengid käisid kursuse lehel Moodle'i keskkonnas, lisaks sellele koguti ka informatsiooni selle kohta, milliseid Moodle'i poolt pakutavaid õppematerjale õppejõud ning tudengid kasutasid: teavitused, lingid, loengumärkmed, materjalid, küsimuste ja vastuste funktsionaalsus, arutelufoorumid, testid, grupitööd, vikid ning ülesannete esitamine [32, lk 1–10].

Kursused klasterdati uuringus nelja gruppi ning selgus, et tervelt 50% kursustest kasutas Moodle'i pakutavaid võimalusi mitteaktiivselt, 24,3% kursustest kasutasid suhtlust või grupitöid, 18% kursustest kasutasid info edastamist või arutelusid ning kõigest 7,2% kursustest ülesannete esitamist ning materjalide jagamist. Suhtlust ning grupitöid kasutavad kursused olid üldiselt väiksema tudengite arvuga võrreldes teiste kursustega. Peamiselt arutelufoorumeid kasutanud kursused julgustasid tudengid oma arvamust avaldama ning õpitu üle arutlema. 7,2% kursuste grupp oli vähem tudengikeskne ning toetus peamiselt materjalide jagamisele ning ülesannetele [31, lk 1–10].

Sobivate klastrite arvude määramisel kasutati erinevaid meetodeid – AIC-i (*the Akaike Information Criterion*), BIC-i (*the Bayesian Information Criterion*), aBIC-i (*the adjusted BIC*), Chi ruutu, LMR-i (Lo-Mendell-Rubin) ning entroopiat [32, lk 8].

Alapeatükis “Moodle'i keskkonna logide seosed kursuse lõpptulemustega” kirjeldatud New Yorgi Osariigi Ülikoolis (SUNY) läbiviidud uuringus pidid kõik testrühmas olnud tudengid kirjutama ka essee, milles üliõpilased pidid kirjeldama enda õpieesmärke kursusel. Esseedest koguti kokku 17 märksõna, mida kasutati koos tudengite lõpphinnetega klasterdamisel. Hinnete ning märksõnade alusel suudeti tudengid jagada viide erinevasse rühma. Märksõnarühmadeks olid (1) eesmärgile orienteeritud tudengid, (2) tudengid, kellele meeldis kursus ja selle materjalid, (3) tudengid, kellel oli IT-lahenduste kasutamisel probleeme, (4) tudengid, kes leidsid tuleviku jaoks midagi kasulikku kursusest, ning (5) tudengid, kes leidsid, et kursus oli liiga raske ning segane [11, lk 929–935].

Parimate hinnetega olid eesmärgile orienteeritud tudengid, kes moodustasid 11% kõikidest tudengitest ning kelle keskmiseks lõpphindeks oli 99,2 punkti. Tudengid, kellele meeldis kursuse sisu, olid õpitulemuste arvestuses teisel kohal keskmise hindega 95,7 punkti. See grupp moodustas 23% kõikidest uuringus osalenud tudengitest. 28% tudengitest oli IT-lahendustega probleeme, kuid nad ületasid need, saades oma keskmise tulemusega 92,3 punkti gruppidest paremuselt kolmanda koha. Neljas grupp keskmise tulemusega 91 punkti olid tudengid, kes leidsid kursuselt midagi kasulikku oma tuleviku tegemiste jaoks, selle rühma suuruseks oli 26%. Kõige kehvema tulemuse said tudengid, kes leidsid, et kursus oli liiga keerukas või segase korraldusega. Nende keskmiseks tulemuseks oli 76,3 punkti ning nad moodustasid 14% kõigist vaadeldud üliõpilastest [11, lk 929–935].

Klasterdamise peamiseks eesmärgiks käesolevas magistritöös oli leida tudengeid kirjeldavatest andmehulkadest käitumismustreid, mis ühendavad sarnaseid hindeid saavaid tudengeid. Autor soovis andmeid klasterdamise abil täiendavalt uurida, kuna varasemalt on TalTechis läbiviidud sooritusedukuse teemalistes uuringutes seda edukalt kasutatud [15, lk 187–188]. Tudengite klasterdamisel erinevate andmehulkade alusel kasutab autor *k-means* algoritmi ning koos sellega kaht erinevat meetodit sobivaima klastrite arvu valimiseks, et oleks võimalik võrrelda erinevaid lahendusi.

2.6.1 *K-means* algoritm

Tudengite klasterdamisülesande lahendamiseks valis autor *k-means* algoritmi. *K-means* algoritmis esindab igat klastrit tsentroid, mis on keskmine vaadeldava klastrite elementidest. Algoritm loob klastrid sarnasuse alusel kasutades tsentroide. Algoritmi peamiseks miinusteks on see, et ta toimib ainult arvandmetega, ning erandid mõjutavad algoritmi tööd [33, lk 10].

Autor otsustas selle algoritmi kasuks, kuna õpianalüütika valdkonnas läbiviidud klasterdamisalgoritme võrdlevas uuringus “*Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets*” osutus *k-means* testitud algoritmide hulgas parimaks [33, lk 15].

2.6.2 Bayesi informatsiooni kriteerium

Esimeseks klastrite arvu määramise meetodiks on BIC (*the Bayesian information criterion*) ehk Bayesi informatsiooni kriteerium. BIC-i arvutus põhineb empiirilisel

logaritmilisel tõenäosusel ning ei vaja eeldusena määratud näitajaid tõenäosuse jaotuse kohta. Seetõttu kasutatakse BIC-i tihti Bayesi modelleerimisprobleemides, mille puhul on keerukas määrata eeldusi tõenäosuse jaotuse kohta. BIC leitakse iga võimaliku mudeli jaoks ning leitud väärtuste hulgast valitakse välja minimaalne. Juhul kui lahendatava ülesande korral on BIC-i väärtused negatiivsed, leitakse maksimaalne BIC-i väärtus. Erinevalt alternatiivsetest mudeli valimise kriteeriumitest, näiteks AIC-ist (*Akaike information criterion*), eelistab BIC ökonoomsemaid mudeleid. Teiseks BIC-i eeliseks on tema arvutuslik lihtsus. Väikeste ning keskmise suurusega andmekogumite korral töötab BIC AIC-ist paremini, valides suuremal arvul kordadest õige struktuuriga mudeli. BIC ei toimi olukordades, kus andmehulga suurus ei ole palju suurem mudeli sisendparameetrite arvust. BIC ei ole sobiv ka keerukamate mudelite jaoks, näiteks kõrgedimensioonilised muutujate või funktsioonide valimise probleemid. Kuna tudengite logi ning küsitluse andmete puhul on tegu väiksemate ning keskmise suurusega andmehulkadega, otsustas autor BIC-i kasuks klastrite arvu määramisel [34, lk 50–53], [35, lk 199–202].

2.6.3 Keskmise siluettide koefitsient

Teiseks klastrite arvu leidmise meetodiks on siluettide (*silhouettes*) kasutamine, täpsemalt keskmise siluettide koefitsiendi kasutamine. Igat klastrit esitab siluett, mis tugineb võrdlusele tiheduse ning eraldatuse aspektis [36, lk 53]. Keskmise silueti laiuse abil saab hinnata klastrite valiidsust ning määrata andmehulgale korrektset klastrite arvu. Siluettide keskmise koefitsiendi arvutamiseks kasutatakse iga andmehulga klastrisisest keskmist kaugust ja keskmist lähima klastriga kaugust. Väärtused jäävad vahemikku -1 kuni 1, kus 1 tähistab parimat väärtust ning -1 halvimat. Nullilähedased siluettide keskmise koefitsiendi väärtused viitavad sellele, et klastrid on üksteisega kattuvad, ning negatiivsed väärtused näitavad, et vaadeldavad näited on pandud valedeesse klastritesse, sest teised klastrid on sarnasemad [37].

3 Andmed

Käesolevas magistritöös kasutas autor TalTechi *ained.ttu* Moodle'i andmebaasist pärinevaid logifaile, lisaks sellele ka 2019/2020 õppeaasta esimese kursuse, peamiselt Informaatika õppekaval õppivate tudengite hulgas läbi viidud õpikäitumise, -motivatsiooni, -uskumuste ning õppimisega seotud emotsioonide teemalise küsitluse andmeid. Sealhulgas kasutas autor nii tudengite vastuseid küsitlusele kui ka küsitluse põhjal leitud tudengeid kirjeldavaid koondtunnuseid. Kuna Moodle'i keskkonna logide põhjal ei ole võimalik eristada kursuse lõplikke punkte tavalistest jooksvatest hinnetest, tuli kokku viia ka tudengite Moodle'i keskkonna logiandmed ning nende lõpphinded, et oleks võimalik ennustada tudengite tulemusi ning leida seoseid klastrite ning tudengite lõpphinnete vahel. Selleks tuli kasutada eraldi andmefaile, mis sisaldasid kursuste lõpphindeid, ning faile, mis võimaldasid kokku viia tudengite e-posti aadressid nende Moodle'i keskkonna kasutaja identifikaatoritega.

3.1 Andmekaitse

Enne täpsemat ülevaadet andmetest on oluline tuua välja, kuidas töö autor, juhendaja ning teised projektis osalenud isikud tagasid tudengite andmete konfidentsiaalsuse ning turvalisuse. Kõikidele tudengitele viidatakse kasutades nende Moodle'i anonüümset identifikaatorit, mis tagab, et tudengite andmeid, mida kasutatakse antud töö andmeanalüüsis, ennustustes, klasterdamises ning tulemuste kirjeldamisel, ei ole võimalik kokku viia nende isikuga. Juhtudel, kus tudengite andmeid ei olnud võimalik koheselt esitada formaadis, kus tudengeid tähistati Moodle'i keskkonna anonüümse identifikaatori abil, anonümiseeriti andmed esimesel võimalusel, et tagada tudengite andmete maksimaalne kaitstus, samaaegselt võimaldades vajaliku informatsiooni kasutamist magistritöö raames.

Anonümiseerimiseks kasutati tudengite TalTechi e-posti aadresse, mis olid seotud Moodle'i keskkonna identifikaatoriga. Antud informatsiooni sisaldavad failid on hoiustatud turvaliselt TalTechi SharePointi keskkonnas ning on kättesaadavad ainult töö autorile ning juhendajale. Andmeid kasutati ainult tudengite anonümiseerimise eesmärgil ning ei jagatud muudes keskkondades, sealhulgas pole antud andmeid lisatud ka TalTechi GitLabi keskkonda koos logifailide ning muude anonüümsete andmekogumitega.

Autor soovib siinkohal kinnitada, et on sõlminud konfidentsiaalsuslepingu ning ei ole kuritarvitanud talle antud õigusi. Autor on teadlik käsitletavate andmete tundlikkusest ning on järginud konfidentsiaalsusega seotud reegleid ning teeb seda ka siis, kui töö andmetega on lõppenud.

3.2 Õppeained

Andmeanalüüsi jaoks valiti neli erinevat TalTechi Informaatika õppekava kursust – õppeaasta 2018/2019 kevadsemestril läbi viidud programmeerimise põhikursus ITI0202 ning õppeaasta 2019/2020 sügissemestril läbi viidud kursused sissejuhatus erialasse ITI0104, robotite programmeerimine ITI0201 ning programmeerimise algkursus ITI0102. Programmeerimise põhikursuse andmeid kasutati peamiselt logiandmete struktuuriga tutvumiseks ning, et välja mõelda, milliseid õpipädevuse ning -käitumisega seotud andmeid on võimalik Moodle'i keskkonna logidest kätte saada. Lõpptulemuste ennustamiseks ning andmete klasterdamiseks kasutati 2019/2020 sügissemestri kursuste andmeid, kuna antud tudengite hulgas viidi läbi ka õpipädevust ning -käitumist käsitlev küsitlus, mis pakub võimalust nii alternatiivseks kui ka logiandmetega kombineeritud tudengite õppe edukuse hindamiseks. Käesolevas töös analüüsitakse läbiviidud küsitluse efektiivsust tudengite õppetulemuste ennustamisel, võrreldakse seda logi andmetest pärineva informatsiooni abil ennustamisega ning ka kombineeritud küsitluse ning logi andmete abil ennustamisega. Lisaks sellele analüüsitakse ka logi ning küsitluse andmete klasterdamisel leitud tulemusi, et tuvastada erinevalt käituvate tudengite profiile. Antud kursused valiti, et saada võimalikult hea ülevaade tudengite käitumisest Moodle'i õppekeskkonnas erinevalt üles ehitatud kursuste korral.

Programmeerimise põhikursus (ITI0202) ning programmeerimise algkursus (ITI0102) on oma ülesehituselt suhteliselt sarnased õppeained: tudengid peavad nädala jooksul Moodle'i keskkonnas esitama rohkem kui ühe programmeerimisülesande, mida kontrollitakse Charon süsteemi üles seatud automaatse testimissüsteemiga [38]. Kodutöid saab esitada mitu korda, kuid saadud punktid sõltuvad sellest, kui palju automaatteste tudengi lahendus läbis ja kas ülesanne esitati õigeaegselt või pärast tähtaja möödumist. Lisaks sellele peavad tudengid oma kodutöid tunnis kaitsma. Juhul, kui nad seda ei tee, ei saa nad Moodle'i keskkonnas oma lahenduse eest punkte. Tudengid peavad antud

kursustel tavaliselt tegema ka paar tunnikontrolli ning ühe suurema kontrolltöö, mis viiakse samuti läbi Moodle'i keskkonnas ning mille tulemused kajastuvad logis. Kursused lõppevad suurema eksamiga, mis otsustab umbes 60% lõpphindest. Kursus on hindeline: hindega 0 on tudeng kursuse läbi kukkunud ning hinded 1 kuni 5 tähistavad positiivset tulemust.

Robotite programmeerimine (ITI0201) on samuti hindeline kursus. Tudengitel on vaja kursuse jooksul esitada neil kodutööd. Esitatud programmide koodistiili kontrollis Moodle'i keskkonna automaatne testimissüsteem, kodutöid tuli ka tunnis kaitsta, näidates toimivat robotit, et saada kätte punktid, mis lisati Moodle'isse. Kodutöid hinnati neljaastmelisel skaalal: tudengid, said kas pronksi, hõbeda, kulla või eliittaseme. Esitamata ülesande korral saadi 0 punkti. Robotite programmeerimise kursusel ei olnud eksamit ega kontrolltöid.

Sissejuhatus erialasse (ITI0104) on erinevalt eelnevalt kirjeldatud õppeainetest aga arvestuslik ehk tudengid hindeid ei saa ning lõpptulemusena on kursus kas arvestatud või mittearvestatud. Tudengid pidid kursuse vältel esitama erinevaid ülesandeid Moodle'i keskkonnas, sealhulgas tuli esitada iganädalane kommentaar, lugeda õppimisteemalisi artikleid ja materjale ning kirjutada ka üks teemakohane essee. Kõik tudengid, kes kogusid kursuse jooksul vähemalt 29 punkti, läbisid aine. Robotite programmeerimise kursusel toimusid semestri esimesel kuul ainult loengud, praktikumid algasid alles viiendal nädalal. Autor käsitleb aine Moodle'i logiandmeid samamoodi nagu teiste kursuste puhul, kuna oluline on tudengite läbikukkumise võimalus avastada võimalikult vara.

Programmeerimise algkursuse, sissejuhatus erialasse ning robotite programmeerimise kursuste näol on tegu üpris erinevate õppeainetega. On esindatud nii hindelised kui ka arvestusega ained, iganädalaste ülesannetega kui ka paari suurema kodutööga kursused, Moodle'i keskkonna automaatset testimissüsteemi kasutavad ja mittekasutavad õppeained ning ka eksamiga ja ilma eksamita kursused.

3.3 Kasutatud Moodle'i keskkonna logi andmed

Esimeseks ülesandeks oli Moodle'i logidest tudengite õpikäitumist kirjeldava informatsiooni leidmine.

Moodle'i logid pärinevad andmebaasitabelist *mdl_logstore_standard_log*. Logifailid olid algselt tekstfaili (TXT) formaadis, kuid nendes failides olid logiridade väärtused üksteisest eraldatud tabeldusmärk (*tab*) abi, seega sai failid lihtsalt konverteerida kas TSV- või CSV-formaati. Selleks kasutas autor Pythonile loodud Pandase teeki, mille abil tegi autor ka enamiku andmete töötlemisest ning analüüsist.

3.3.1 Moodle'i logirida

Üks Moodle'i logirida sisaldab rohkelt informatsiooni. Täpsemalt salvestatakse iga Moodle'i keskkonna sündmuse kohta kindlad metaandmed: sündmuse nimi (*eventname*), Moodle'i komponent, mis sündmuse genereeris (*component*), sündmuse tüübi klassifikaator (*action*), haridustase (*edulevel*), anonümiseeritud identifikaator (*anon_id*), mis võimaldab luua seoseid teiste allikate kirjetega, anonümiseeritud õppejõu identifikaator (*anon_teacher_id*), õppeaine identifikaator (*courseid*), sündmuse ajatempel (*timecreated*), sündumuse käivitanud kasutaja IP aadress (*ip*) ning täiendavad andmed (*other*), kui need on täpsustatud sündmuse kirjelduse all [39].

Lisaks Moodle'i keskkonna sündmuse informatsioonile sisaldab logirida ka tehnilisi metaandmeid – viide alamobjekti komponendile (*target*), päritolu (*origin*), anonüümsuse näitaja (*anonymous*), andmebaasi tabeli nimi, kus muudatus toimus (*objecttable*), rida andmebaasitabelis, mille poole pööruti (*objectid*), sündmust kirjeldav CRUD operatsioon (*crud*), Moodle'i sisene tehniline identifikaator (*contextid*), konteksti kirjeldav väärtus (*contextlevel*), mis ütleb, kas tegu on kursuse, tegevuse või kursuse kategooriaga ning vastavalt kas kursuse, mooduli või kategooria identifikaator (*contextinstanceid*) [39].

Üks tudengi hindamise sündmust kirjeldav Programmeerimise algkursuse logirida on esitatud Joonisel 2 (IP-aadress on muudetud). Selle logirea saab lahti mõtestada Tabel 1 abil, tabelis on IP aadressid muudetud.

```
13958113  \\core\\event\\user_graded core  graded user
grade_grades  851254 u 1 33482 50 275 0 275 5086 0
a:3:{s:6:"itemid";s:4:"7655";s:10:"overridden";b:0;s:10:"finalgrade";s
:7:"0.00000";} 1569999283 web 192.168.0.1 \N
```

Joonis 2. Programmeerimise algkursuse Moodle'i logirida.

Kuna eeltööna olid magistritöö juhendaja poolt loodud failid, mis sisaldasid ainult vaadeldavate kursuste logikirjeid, siis kõige olulisemateks andmeväljadeks logireas

osutused logi kirjes olev tegevus (*event*), tegevusega seotud kasutajad, õppejõud ning tudeng, logikirje loomise ajahetk UNIXi ajatempli formaadis ning lisainformatsiooni sisaldav andmeväli *other*.

Tabel 1. Näidis logikirjet lahti seletav tabel.

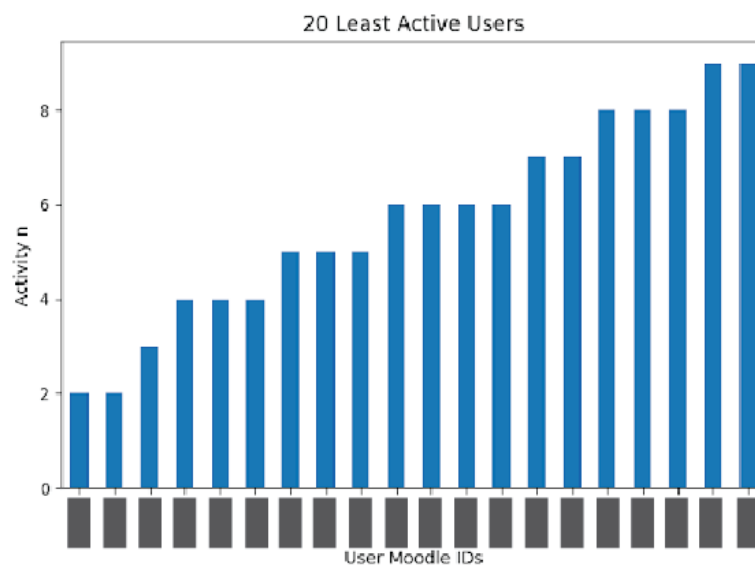
Metaandmete välja nimetus	Väärtus
<i>id</i>	13958113
<i>eventname</i>	\\core\\event\\user_graded
<i>component</i>	core
<i>action</i>	graded
<i>target</i>	user
<i>object_table</i>	grade_grades
<i>object_id</i>	851254
<i>crud</i>	u
<i>education_level</i>	1
<i>context_id</i>	33482
<i>context_level</i>	50
<i>context_instance_id</i>	275
<i>user_id</i>	0
<i>course_id</i>	275
<i>related_user_id</i>	5086
<i>anonymous</i>	0
<i>other</i>	a:3:{s:6:"itemid";s:4:"7655";s:10:"overridden";b:0;s:10:"finalgrade";s:7:"0.00000";}
<i>time_created</i>	1569999283
<i>origin</i>	web
<i>ip</i>	192.168.0.1
<i>real_user_id</i>	\N

3.3.2 Esmane logiandmete töötlus ning analüüs

Selleks, et logi andmetega tutvuda, tegi autor ülevaate saamiseks erinevaid graafikuid. Need graafikud koostas autor peamiselt eesmärgiga, et leida andmetes erinevaid huvitavaid trende, mis visualiseeritud kujul võivad paista rohkem silma, kui lihtsalt

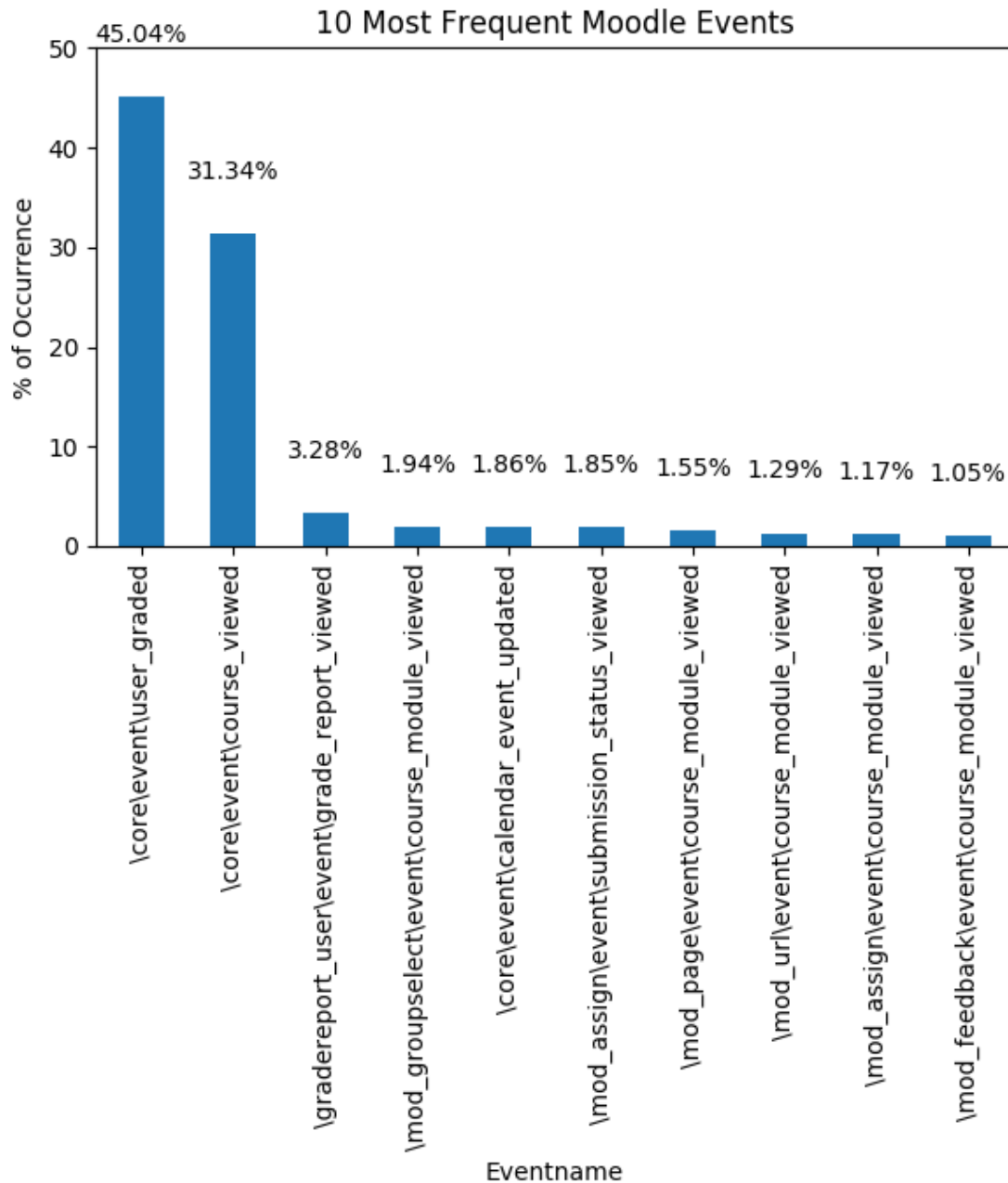
numbreid vaadates. Näiteks tegi autor graafikud kõige rohkem logikirjetes esinevate sündmuste, kursuse kõige aktiivsemate Moodle'i kasutajate kohta, kursuse kõige vähemaktiivsete kasutajate kohta ning kõige suurema aktiivsusega kuude, nädalapäevade ning kellaaegade kohta. Seda analüüsi tegi autor Programmeerimise põhikursuse andmete põhjal.

Graafikuid vaadeldes selgus näiteks, et kõige aktiivsemate ning vähemaktiivsete kasutajate logikirjete arvud on väga erinevad, nimelt oli kõige aktiivsematel kasutajate nimel ligi 200 000 logikirjet, seevastu oli aga kõige vähemaktiivsetel kasutajatel alla kümne logikirje (vt Joonis 3, joonisel on Moodle'i identifikaatorid anonümiseeritud).



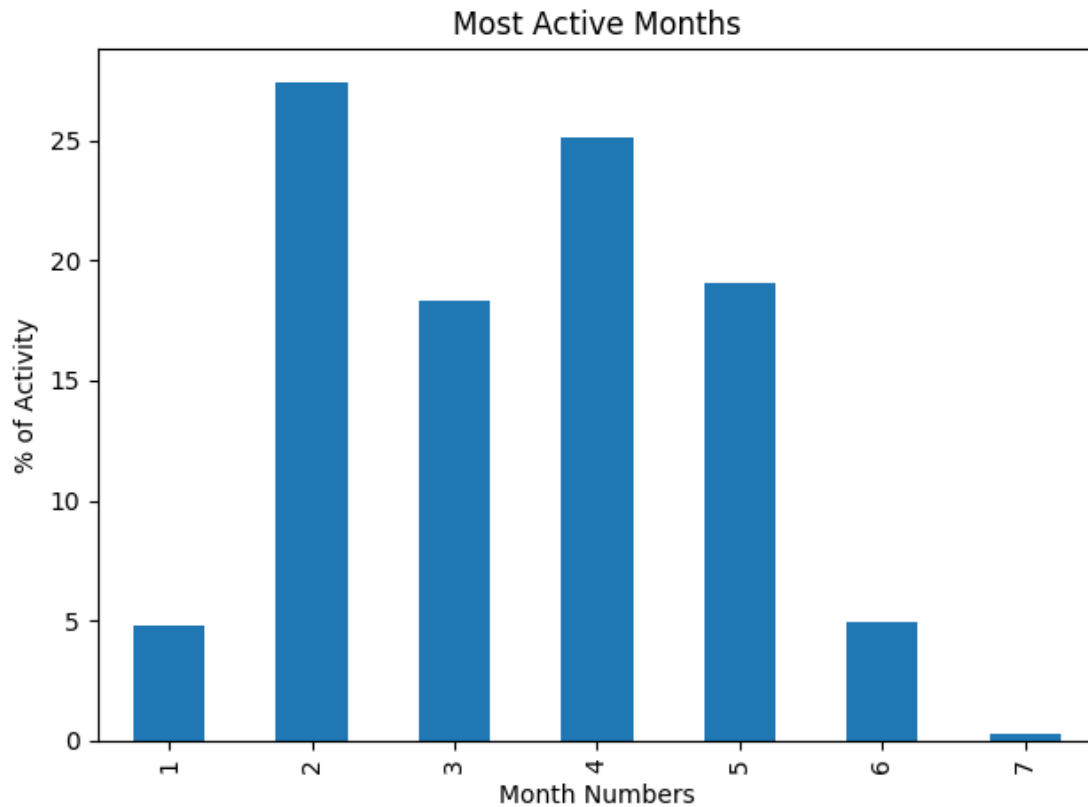
Joonis 3. Programmeerimise põhikursuse 20 kõige vähemaktiivset kasutajat logiandmete alusel. Vertikaalsel teljel kujutatakse logikirjete arvu ning horisontaalsel teljel kasutajate Moodle'i identifikaatoreid.

Samuti selgus, et kõige aktiivsemate Moodle'i keskkonna kasutajate hulgas on õppejõud ise, kuid leidub ka tudengeid, kes umbes sama aktiivselt õppekeskkonnas tegutsevad. Kõige sagedamini esinevateks logisündmusteks osutusid ülekaalukalt tudengi hindamise sündmus ning kursuse lehe külastamise sündmus (vt Joonis 4).



Joonis 4. Programmeerimise põhikursuse logis 10 kõige sagedamini esinevat sündmust. Vertikaalsel teljel on esitatud protsentuaalne skaala ning horisontaalsel teljel sündmuste nimed.

Kevadsemestril aset leidnud Programmeerimise põhikursuse logides oli kõige rohkem aktiivsust veebruarikuus ning iga kuuga jäi kirjeid logis vähemaks, kuigi aprillis ning mais oli kasutajate hulgas kerge aktiivsuse tõus. (vt Joonis 5). Saadud informatsioonist lähtudes jätkas autor tudengite kohta käivate andmekogumite koostamisega.



Joonis 5. Programmeerimise põhikursuse kõige suurema aktiivsusega kuud logiandmete alusel. Vertikaalsel teljel on esitatud protsentuaalne skaala ning horisontaalsel teljel kevadsemestri kuude numbrid.

3.3.3 Logiandmete töötlemine

Moodle'i logisündmuseid on mitmeid eri tüüpe ning kokku on neid Moodle'i keskkonnas 212. Esialgsete andmekogumite loomise käigus luges autor kokku nii üliõpilaste kõiki logikirjeid kui ka iga tudengi kursuse lehe külastuste arvu (logisündmus `\core\event\course_viewed`). Kuna kirjanduse ülevaates selgus, et kursuse lehe külastuste arv ei anna tihti sisukat informatsiooni tudengi lõpphinde kohta, otsustas autor seda näitajat mitte kasutada [9]. Seega kasutas autor peamiselt näitajana tudengi logikirjete arvu ning logikirjetes olevate tegevuste toimumise aegu. Üks kõige olulisemaid sündmusi logis, milleks on `\core\event\user_graded` ehk sündmus, mis kirjeldab tudengile hinde panemist, käib kokku mitte tudengi Moodle'i identifikaatoriga, vaid hoopis õppejõudude Moodle'i identifikaatoritega. Seega tuli hindamissündmused eraldiseisvalt õigete tudengitega kokku viia, et üliõpilastega oleks seotud ka nende hinnete ning ülesannete esitamise andmed [39].

Üheks kõige keerulisemaks osaks logikirjete töötlemisel oli andmeväljalt *other* vajaliku informatsiooni kättesaamine, mille sisu osutus äärmiselt oluliseks hindamissündmuse korral, kuna väli sisaldas informatsiooni hinnatud ülesande kohta ning tudengile ülesande eest antud punktide kohta. Kahjuks on Moodle'i logides andmeväli *other* salvestatud PHP objekti andmeformaadis, millest oli Pythonit ning Pandase teeki kasutades äärmiselt keeruline vajalikku informatsiooni kätte saada. Selleks, et mitte kasutusele võtta PHP näol teist programmeerimiskeelt ainult selleks, et üht andmevälja lahti pakkida, otsustas autor regulaaravaldisi kasutades konverteerida *other* andmevälja väärtused JSON formaati. See osutus võimalikuks, kuna PHP objekti sõne kuju vaadeldes ilmsed teatavad reeglipärasused andmete esituses. Kahjuks ei lõppenud töö korrektsete regulaaravaldise reeglite kirjutamisega, sest selgus, et üksikute logisündmuste korral oli PHP objekti formaadis vigu, mis ei vastanud üldiselt kehtivatele reeglitele. Õnneks polnud tegu tudengi hindamise informatsiooni sisaldavate andmeväljadega ning autor otsustas, et katkise formaadiga andmeväljade sisu polnud antud töö skoobis oluline ning seega võis need väljad JSON formaati konverteerimata jätta ning tuli lihtsalt korrektsetest andmetest eraldada.

Pärast andmete töötlemist oli iga tudengi kohta loodud Pythoni objekt, mis sisaldas kursuse Moodle'i logist leitud informatsiooni tudengi tegevuste kohta. Objekt sisaldas tudengi Moodle'i identifikaatorit, kõiki unikaalseid logisündmusi, mida konkreetne tudeng oli käivitanud oma tegevusega kursuse Moodle'i keskkonnas, tudengi kõige sagedamini esinevat logisündmust, tudengi kõige aktiivsemat kuud, kõige aktiivsemat päeva ning keskmiselt kõige aktiivsemat kellaaega kursuse Moodle'i keskkonnas. Lisaks eelnevale sisaldas objekt ka informatsiooni selle kohta, mitu korda tudeng on kursuse lehte külastanud, iga jooksva ülesande lõpphinnet, infot selle kohta, mitu korda tudeng ülesannet esitas ning millal tudeng ülesande esimest korda esitas ning millal ta ülesande lõplikult esitas. Kõigi logikirjete kohta oli tudengi kohta käivas objektis ka sündmus, mida logirida kirjeldas, ning sündmuse toimumise aeg. Viimasena sisaldas tudengit kirjeldav Pythoni objekt logist eraldiseisvate failide abil tudengi infole lisatud vastuseid küsitluse küsimustele, kui ta oli andnud nõusoleku seda informatsiooni jagada, ning vaadeldava kursuse lõpphinnet.

Autor valis esialgselt kokkukogutud tudengite andmete hulgast välja informatsiooni, mille kohta leiti täpsustavad näitajad, mida omakorda kasutati ennustamisel ning tudengi andmete klasterdamisel. Ennustamisel ning tudengite andmete klasterdamisel

kasutatavad näitajad salvestati kursuste ning näitajate tüüpide kaupa eraldi CSV-failidesse, et vältida andmete pidevat taastöötlmist, kuna andmetöötlus võtab kõige suurema osalejate arvuga kursuse logi korral natuke alla kahe minuti aega.

3.3.4 Aktiivsus

Esimeseks ennustamise ning tudengite klasterdamise jaoks väljavalitud näitajate komplektiks on tudengi kursusest osavõtmise aktiivsuse näitajad. Näitajad leitakse vaadeldava ajaperioodi logiandmete põhjal. Käesoleva magistritöö raames on vaadeldavateks perioodideks kas kursuse asetleidmise semestri esimene või esimesed kaks kuud. Vaadeldava ajaperioodi olulisust selgitatakse täpsemalt ennustamise peatükis.

Aktiivsusnäitajad on järgmised –

- *activity*
- *number_of_active_days*
- *number_of_active_weeks*
- Igale vaadeldavale nädalale vastav logikirjete arv (näiteks 36, 37 jne)
- *diff_from_average*
- *percentage_from_max*

Esimeseks aktiivsusnäitajaks on *activity* (aktiivsus), mis tähistab kõigi tudengiga seotud logikirjete arvu vaadeldaval perioodil. Teiseks aktiivsusnäitajaks on *number_of_active_days* (aktiivsete päevade arv), mis väljendab mitmel päeval vaadeldava ajaperioodi jooksul tudengi kohta oli logikirjeid. Kolmandaks üliõpilase aktiivsust kirjeldavaks näitajaks on *number_of_active_weeks* (aktiivsete nädalate arv), mis näitab mitmel nädalal vaadeldava perioodi jooksul tegutses tudeng Moodle'i keskkonnas. Lisaks sellele on eraldi näitajatena välja toodud iga vaadeldava perioodi nädala kohta, mitu logikirjet tudengil konkreetsel nädalal oli. Nädalaid tähistatakse nädalanumbritega. Viimased kaks aktiivsusnäitajat on *diff_from_average* (erinevus keskmisest) ning *percentage_from_max* (protsent maksimaalsest). Neist esimese jaoks leiti keskmine aktiivsus *activity* näitaja alusel, mis lahutati konkreetse tudengi aktiivsusest. Keskmisest aktiivsema tudengi näitaja *diff_from_average* on nullist suurem ning keskmisest madalama aktiivsusega tudengi *diff_from_average* on negatiivne. Näitaja *percentage_from_max* jaoks leiti *activity* näitajat kasutades kõige suurem võimalik tudengi aktiivsus ning *percentage_from_max* näitab, mitu protsenti konkreetse

tudengi aktiivsus moodustab maksimaalsest võimalikust aktiivsusest. Seega on tegu *activity* näitajaga, mis on normaliseeritud protsentuaalsele skaalale.

3.3.5 Jooksvad tulemused

Tudengitele antud kodutöid, suuremaid kursuse jooksul tehtud projekte, Moodle'i keskkonnas sooritatud teste või lihtsalt kontrolltöid ja tunnikontrolle, mille tulemused kanti õppejõu poolt kursuse Moodle'i keskkonda, ühendavad jooksvad hinded või punktid, mis kajastuvad Moodle'i logides. Selleks, et erinevate tudengite jooksvaid tulemusi võrrelda, tuleb kõigepealt leida kõik võimalikud ülesanded, mis vaadeldava perioodi jooksul tudengite poolt kursuse Moodle'i keskkonnas esitati. Seejärel saab leida iga tudengi kohta, kas ta on konkreetse ülesande eest punkte saanud. Jooksvaid tulemusi kirjeldab tabel, kus tabeli päises on Moodle'i ülesannete identifikaatorid ning igal real on ühe tudengi tulemused tulbas olevates ülesannetes.

3.3.6 Keskmise jooksev tulemus

Autor soovis leida ka üldiseid näitajaid, mida saaks kasutada masinõppemudelite korral, mida treenitakse näiteks programmeerimise algkursuse andmete peal, kuid millega ennustatakse robotite programmeerimise tulemusi. Sellisteks näitajateks on näiteks eelnevalt kirjeldatud aktiivsuse näitajad. Jooksvad ülesanded aga erinevad aine ainesse nii oma sisu, kasutatava hindamissüsteemi kui ka arvu poolest. Isegi sama õppeaine ülesannete arv ning sisu muutub semestrite ning õppeaastate jooksul. Autor soovis aga siiski kasutada leitud jooksvaid hindeid, kuna neil on eelnevalt läbiviidud uuringute tulemustest lähtuvalt lõpphinnet ennustav väärtus. Seega otsustas autor leida iga tudengi keskmise jooksva tulemuse. Tihti hinnatakse ülesandeid sarnastel punktiskaaladel – näiteks 0 kuni 5 või siis 0 kuni 100 – ja õppekavasiseselt kasutatakse sageli sarnaselt hindamissüsteemi, kuid keskmiste jooksvate hinnetega kaasnevaks riskiks on see, et kursustel kasutatakse erinevaid hindedkaalaid ning seetõttu ei ole võimalik treenimisel kasutatud õppeaine informatsiooni põhjal teise kursuse tulemuste edukas ennustamine.

3.3.7 Ülesannete esitamise arv

Sarnaselt jooksvatele tulemustele saab iga Moodle'i keskkonnas vaadeldava perioodi jooksul esitatud ülekande kohta leida ka selle, mitu korda tudeng konkreetset ülesannet esitas. Ülesannete mitmekordset esitamist esineb Moodle'i keskkonnas üpris palju, näiteks on tihti võimalik teste mitu korda sooritada, et saada parem tulemus. Tudengid

esitavad programmeerimisülesandeid pea alati mitu korda, kui kodutööde jaoks on Moodle'i kursuses üles seatud automaatne koodi testimissüsteem, mis iga kord annab tudengile vastuse selle kohta, kas ülesanne on juba täielikult õigesti lahendatud või on midagi veel puudu või katki.

3.4 Küsitlus

Alternatiivina Moodle'i keskkonna logifailidele ning nendest leitud õpikäitumist kirjeldavate andmetele viidi üliõpilaste hulgas läbi ka õpikäitumise, -motivatsiooni, õppimisega seotud emotsioonide ning õppimisega seotud ja enesekohaste uskumuste teemaline küsitlus. Küsitluse koostasid koostöös Ago Luberg ning Tallinna Ülikooli hariduspsühholoogid Kati Aus ja Grete Arro.

Küsitlus viidi läbi programmeerimise algkursuse tudengite hulgas jaanuaris 2020. Selleks, et võimalikult suur hulk tudengeid vastaks küsitlusele, viidi see läbi enne eksami sooritamist – ajal, mil tudengid tegelesid veel aktiivselt kursusega. Kõigilt küsitluses osalenutelt küsiti nõusolekut, kas tudeng on nõus, et ta vastuseid kasutatakse käesolevas magistritöös ja programmeerimise algkursuse läbiviimise tõhusamaks muutmise eesmärgil. Tudengid, kes vastasid sellele eitavalt, eemaldati uuritavate andmete hulgast. Lõppkokkuvõttes vastas küsitlusele 305 tudengit, kellest 16 ei andnud nõusolekut enda vastuste kasutamiseks, seega on valimi suuruseks 289 tudengit.

Läbi viidud küsitluses oli kokku 109 küsimust. Küsitluses oli ka paar korduvat küsimust, selleks et kontrollida, kas tudengid vastavad sarnastele küsimustele ühtemoodi, olles tähelepanelikud ning mõttega küsitlusele vastamise juures.

3.4.1 Küsitluse küsimused

Küsitluse vastustega andmefail oli algselt XLSX-formaadis, mille autor konverteeris xlr¹ teegi abil CSV-formaati, et seda oleks lihtsam edaspidi koos olemasolevate logiandmetega kasutada.

¹ <https://pypi.org/project/xlr/>

Iga tudengi vastused läbiviidud küsitlusele on samuti ennustamisel kasutatavaks tunnuste komplektiks. Eraldiseisvalt tudengite otsestest vastustest küsitlusele kasutatakse ka küsitluse pealt leitud koondtunnuseid.

3.4.2 Küsitluse koondtunnused

Küsitlusele antud vastuste pealt leitud koondtunnus kirjeldab üht teemavaldkonda esindavatele küsimustele tudengi poolt antud vastuste üldist väärtust. Koondtunnused on pidevad arvud vahemikus 0 kuni 6, sest tudengid said vastata küsimustele ja väidetele skaalal nullist kuueni, ühe koondtunnuse alla jäävate küsimuste vastused korrutati omavahel. Väärtus 0 tähendab, et tudengit ei iseloomusta üldse vaadeldav tunnus, ning väärtus 6 näitab, et tudengit iseloomustab väga tugevalt vaadeldav tunnus. Kõiki koondtunnuseid kasutati tudengite lõpphinnete ennustamisel, et välja selgitada, millised koondtunnused on kõige enam seotud tudengite lõpphinnetega vaadeldud kursustel.

Küsitluse põhjal leitud koondtunnused (sulgudes on esitatud kasutatud lühendid) on

1. Sisemine regulatsioon (*SisemineRegulatsioon*)
2. Omaksvõetud strateegiline regulatsioon (*OmaksvõetudStratRegulatsioon*)
3. Pealesurutud motivatsioon (*PealesurutudRegulatsioon*)
4. Soorituspõhine motivatsioonile lähenemine (*PerfApproach*)
5. Läbikukkumise vältimisele suunatud soorituseesmärgid (*PerfAvoidance*)
6. Uskumused matemaatikaalase võimekuse, loogilise mõtlemise ja programmeerimisalase võimekuse muudetavuse kohta (*MateUskumused*)
7. Uskumused ärevuse kontrollitavuse ja suhtlemispädevuse muudetavuse kohta (*SotsAnxUskumused*)
8. Pingutuse olulisus programmeerimiskursusel (*EffortinLearning*)
9. Enesetõhusus (*Enesetõhusus*)
10. Huvi (*Huvi*)
11. Õppimisega seotud ärevus (*Anxiety*)
12. Läbipõlemisele viitavad depressiivsed sümptomid (*Burnout*)
13. Huvikaotus (*Huvikaotus*)
14. Keeruliste ülesannete vältimine (*TaskAvoidance*)
15. Lihtsate ülesannete eelistamine (*Easytasks*)
16. Alustamisega viivitamine (*Prokkimine*)
17. Mittetõhusad õpistrateegiad (*MittetõhusStrat*)

18. Struktureeriv, seoseid loov õppimine ehk tõhusad strateegiad (*TõhusStrat*)

19. Sotsiaalne õppimine (*SotsStrat*)

20. Õppejõu pakutava emotsionaalse ja õppimisalase toe tajumine (*EmToetus*)

Täpsemateks ennustusteks said valitud koondtunnused, mis esindasid erinevaid õppimisega seotud psühholoogiliste tegurite valdkondi, näiteks enesekohaseid ning õppimisega seotud uskumusi ja õpimotivatsiooni. Valdkondadest valiti välja koondtunnused, millel oli märkimisväärne korrelatsioon programmeerimise algkursuse lõpphinnetega. Korrelatsioonide leidmiseks kasutasid hariduspsühholoogid Kati Aus ja Grete Arro Spearmani korrelatsioonikoefitsiente (vt Tabel 2). Väljavalitud tunnusteks osutusid pingutuse olulisus programmeerimiskursusel, enesetõhusus, sisemine regulatsioon, läbikukkumise vältimisele suunatud soorituseesmärgid, huvi, õppimisega seotud ärevus, keeruliste ülesannete vältimine ning mittetõhusad õpistrateegiad.

Tabel 2. Väljavalitud koondtunnuste Spearmani korrelatsioonikoefitsiendid programmeerimise algkursuse lõpphinnetega.

Koondtunnus	Spearmani korrelatsiooni koefitsient	p-väärtus
pingutuse olulisus programmeerimiskursusel	-0,343	0,000
enesetõhusus	0,605	0,000
sisemine regulatsioon	0,288	0,000
läbikukkumise vältimisele suunatud soorituseesmärgid	-0,241	0,000
huvi	0,368	0,000
õppimisega seotud ärevus	-0,328	0,000
keeruliste ülesannete vältimine	-0,509	0,000
mittetõhusad õpistrateegiad	-0,315	0,000

Tunnused pingutuse olulisus programmeerimiskursusel (tähistus graafikutel *EffortinLearning*) ning enesetõhusus (tähistus graafikutel *Enesetõhusus*) kuuluvad valdkonda õppimisega seotud ning enesekohased uskumused. Pingutuse olulisuse kohta kursusel oli küsitluses 3 küsimust ning enesetõhususe kohta oli väiteid kokku 3–6.

Õpimotivatsiooni valdkonda kuuluvateks tunnusteks on sisemine regulatsioon (tähistus graafikutel *SisemineRegulatsioon*) ning läbikukkumise vältimisele suunatud soorituseesmärgid (tähistus graafikutel *PerfAvoidance*). Sisemise regulatsiooni kohta oli küsitluses 4 erinevat väidet ning läbikukkumise vältimisele suunatud soorituseesmärkide kohta oli küsimusi 3.

Koondtunnused huvi (tähistus graafikutel *Huvi*) ning õppimisega seotud ärevus (tähistus graafikutel *Anxiety*) kuuluvad õppimisega seotud emotsioonide valdkonda. Tunnus huvi kohta oli küsimustikus 3 väidet ning õppimisega seotud ärevuse kohta oli väiteid 5.

Õpikäitumise valdkonda kuuluvad tunnused on keeruliste ülesannete vältimine (tähistus graafikutel *TaskAvoidance*) ning mittetõhusad õpistrateegiad (tähistus graafikutel *MittetõhusadStrateegiad*). Küsitluses oli 5 keeruliste ülesannete vältimise teemalist ning 5 mittetõhusate õppimisstrateegiate kohta käivat väidet.

4 Klasterdamine

Andmete klasterdamisel oli käesolevas magistritöös kaks peamist eesmärki. Esimeseks oli andmetega tutvumine ja andmeid kirjeldavate graafikute koostamine. Teiseks väljaselgitamine, kas Moodle'i logiandmete ning küsitluse koondtunnuste järgi on võimalik tudengeid klasterdada nõnda, et tekiks tudengeid kirjeldavad profiilid. Profiile saab kasutada, et edukamalt tuvastada väljakukkumis ohus olevaid tudengeid. Autor kasutas neid meetodeid seaduspärasuste ja trendide leidmiseks, millest võiks olla täiendavat kasu andmehulkade loomisel ennustamise jaoks.

Autor klasterdas kõigi kolme ennustamisel kasutatud kursuse andmeid, seda nii semestri esimese kui ka kahe esimese kuu kohta. Peamisteks vaadeldud andmehulkadeks olid tudengite aktiivsuse näitajad, tudengite jooksvad tulemused, tudengite ülesannete esitamise kordade arvud ning küsitluse koondtunnused.

Algselt kasutas autor klasterdamisel ka kombineeritud andmehulkasid, ülesannete jooksvad tulemused koos väljavalitud koondtunnustega, aktiivsuse näitajad koos väljavalitud koondtunnustega ja jooksvaid tulemusi koos aktiivsuse näitajatega, kuid nende andmehulkade korral ei õnnestunud BIC-ist ning keskmisest siluettide koefitsiendist lähtuvalt klasterdamine edukalt. Lisaks sellele oli andmehulkade kohta genereeritud graafikutel liiga palju informatsiooni, mis muutis klastrite visuaalse interpreteerimise ning mõistmise äärmiselt keeruliseks. Sarnastel põhjustel ei kasutanud autor klasterdatava andmehulgana ka küsimustiku vastuseid.

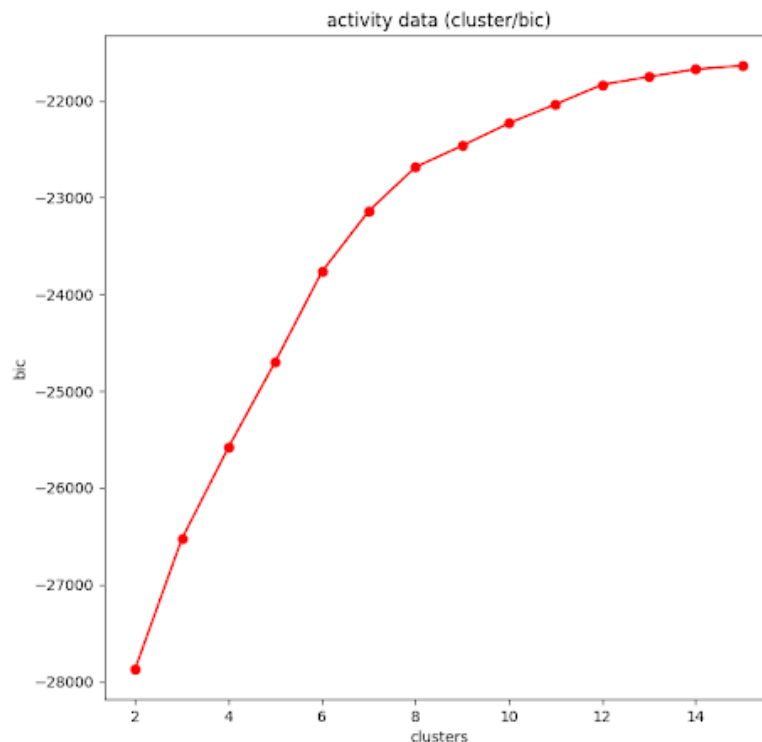
Kõigi klasterdamisel kasutatud andmehulkade kohta leidis autor parima klastrite arvu vahemikus 2 kuni 15 klastrit. Klastrite arvu leidmiseks kasutas autor BIC-i ning keskmist siluettide koefitsienti. Parima klastrite arvu väljaselgitamiseks genereeris autor graafikud BIC-i ja keskmiste siluettide koefitsiendi väärtustega 2 kuni 15 klastrit korral. BIC-i korral oli parimaks väärtuseks suurim BIC-i väärtus, kuna graafikul olid väärtused negatiivsed. Parim keskmiste siluettide koefitsient on võimalikult lähedal väärtusele 1.

4.1 Sobivate klastrite arvude leidmine

Järgnevas alapeatükis selgitab autor, kuidas leiti sobivad klastrite arvud programmeerimise algkursuse, robotite programmeerimise ning sissejuhatus erialasse õppeainete andmehulkadele.

4.1.1 Programmeerimise algkursus

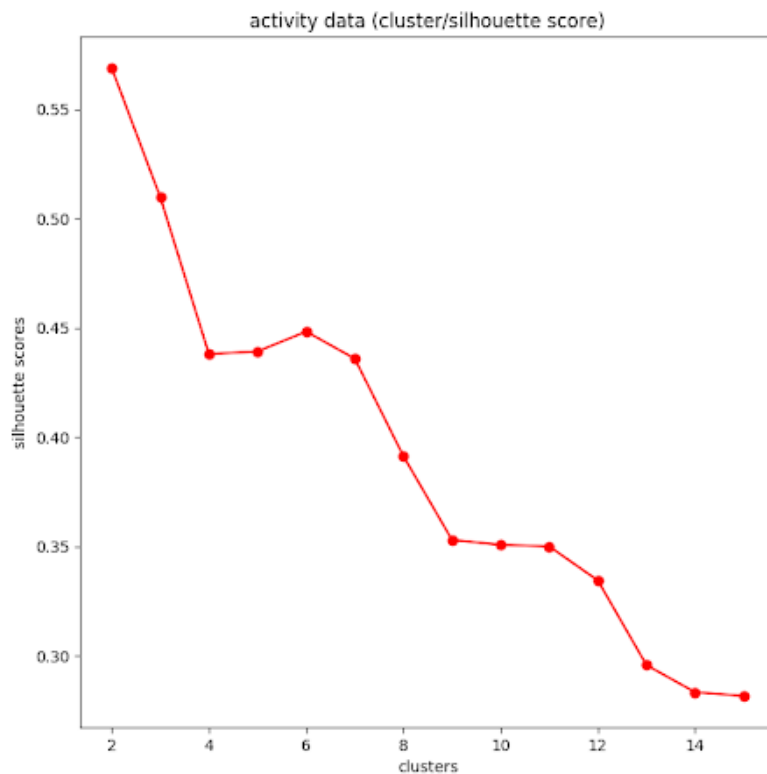
Programmeerimise algkursuse semestri esimese kuu aktiivsusnäitajate korral ei õnnestunud leida BIC-i abil sobivat klastrite arvu, kuna graafik oli tõusev ning negatiivsed väärtused lähenesid nullile (vt. Joonis 6). Keskmiste siluettide koefitsientide graafikut vaadates selgus, et klastrid erinevad üksteisest kõige enam, kui andmed jaotada kaheks, sest siis oli koefitsiendi väärtus kõige lähemal väärtusele 1, jäädes vahemikku 0,55 ja 0,60 vahel (vt Joonis 7). Seega valis autor klastrite arvuks kahe.



Joonis 6. Programmeerimise algkursuse semestri esimese kuu aktiivsusnäitajate andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.

Programmeerimise algkursuse semestri esimese kuu jooksvate tulemuste andmehulga korral osutusid BIC-i väärtused taaskord tõusvaks graafikuks, seega ei saanud BIC-i abil klastrite arvu määrata. Parimaks keskmiste siluettide koefitsiendi väärtuseks oli jooksvate

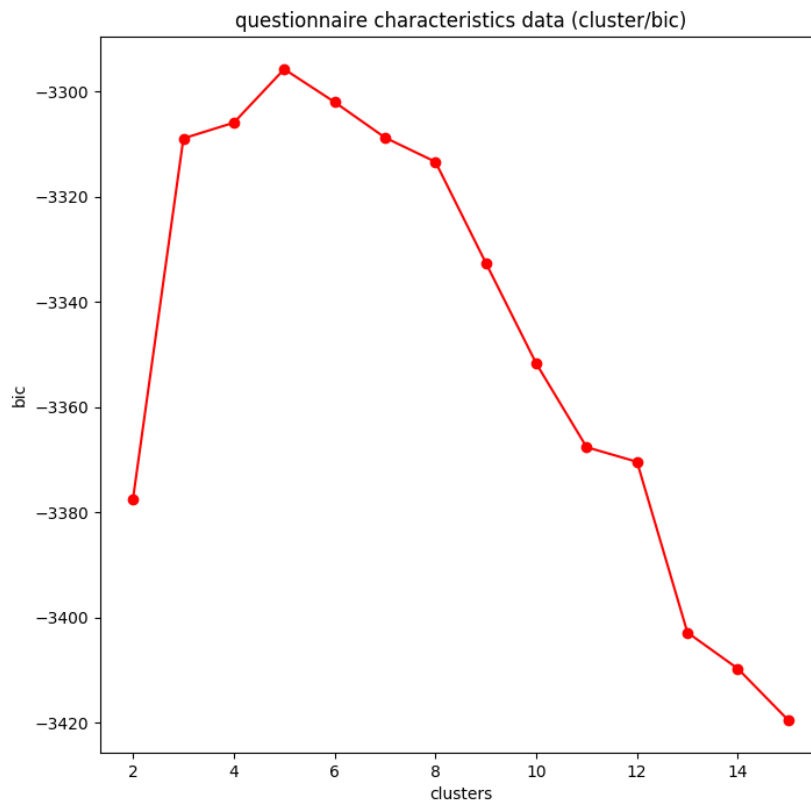
tulemuste andmehulga korral 13. Seega valis autor klastrite arvaks 13. Ülesannete esitamiste kordade andmehulga BIC-i väärtused olid samuti kasvaval graafikul, seega tuli taaskord klastrite arvu määramiseks kasutada siluettide koefitsiendi väärtust. Parimaks osutus siluettide koefitsiendi väärtus kahe klastrite korral, kuid koefitsiendi väärtuseks oli 0,215, mis ei ole ideaalsele väärtusele 1 eriti lähedal. Vaatamata sellele oli parimaks klastrite arvaks 2.



Joonis 7. Programmeerimise algkursuse semestri esimese kuu aktiivsuse näitajate andmehulga keskmiste siluettide koefitsientide väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel keskmisi siluettide koefitsiente.

Programmeerimise algkursuse semestri esimese kahe kuu Moodle'i aktiivsuse näitajate BIC väärtuste graafik oli tõusev, ning keskmiste siluettide koefitsientide alusel oli parimaks klastrite arvaks 2. Ka tudengite jooksvate tulemuste BIC-i graafik oli tõusev ning keskmiste siluettide koefitsiendid määrasid klastrite arvaks 2. Ülesannete esitamiste arvude BIC väärtuste graafik oli tõusev, seega tuli taaskord valida klastrite arv kasutades keskmiste siluettide koefitsiente, mis näitasid, et klastrid erinevad üksteisest kõige rohkem siis kui neid on kaks.

Viimaseks programmeerimise algkursuse andmehulgaks oli küsitluse koondtunnused. Koondtunnuste andmehulga korral sai klastrite arvu määrata kasutades BIC-i väärtust, parimaks klastrite arvuks osutus 5 (vt Joonis 8.). Keskmiste siluettide koefitsiendi alusel oleks klastrite arvuks pidanud valima kahe, kuid autor otsustas eelistada BIC-i, sest keskmiste siluettide koefitsiendi parim tulemus oli väärtusega 0,25.



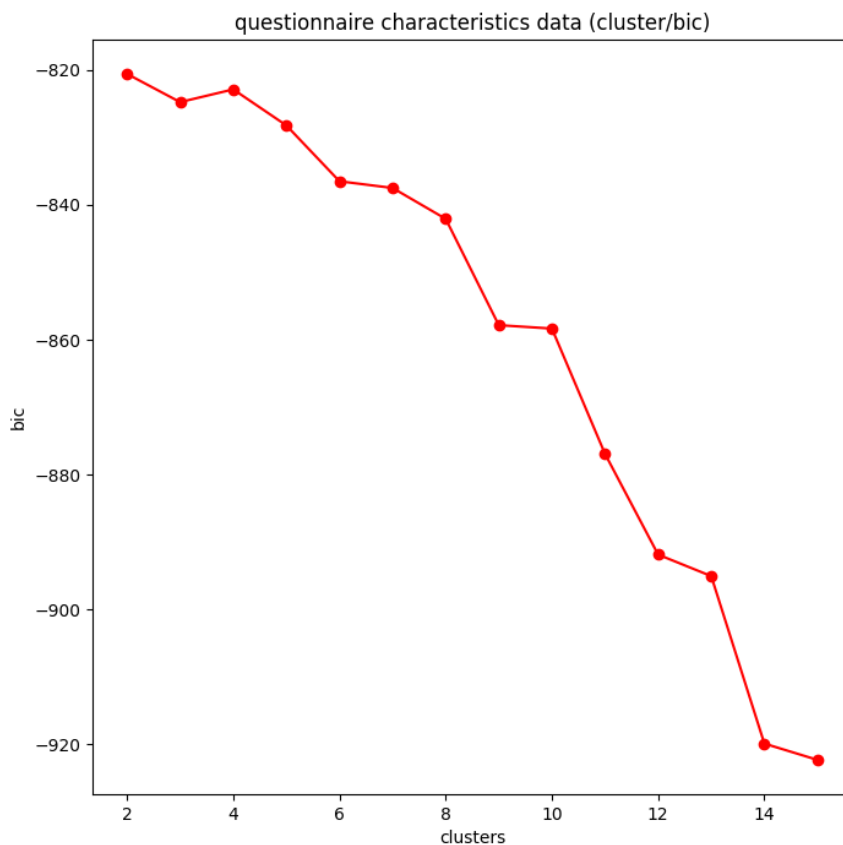
Joonis 8. Programmeerimise algkursuse küsitluse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.

4.1.2 Robotite programmeerimine

Robotite programmeerimise kursuse semestri esimese kuu aktiivusnäitajate andmehulkade BIC oli tõusva väärtusega, seega tuli klastrite arvu määramiseks kasutada keskmiste siluettide koefitsienti, mis määras klastrite arvuks 2 koefitsiendi väärtusega 0,59. Robotite programmeerimise kursuse semestri esimesel kuul ei olnud ülesannetega seotud andmeid.

Küsitluse koondtunnuste klastrite arvu määras BIC, mis oli maksimaalse väärtusega 2 klaster korral (vt Joonis 9), sama klastrite arvu andis ka keskmiste siluettide koefitsient.

Robotite programmeerimise semestri esimese kahe kuu aktiivsuse näitajate andmehulga klastrid erinesid üksteisest kõige rohkem kui andmed jagada kaheks. Parim võimalik klastrite arv oli kaks ka tudengite jooksvate tulemuste andmehulgal ning ülesannete esitamiste kordade andmehulgal. Kõigi kolme andmehulga korral ei saanud klastrite arvu määramiseks kasutada BIC-i ning optimaalne klastrite arv leiti keskmiste siluettide koefitsientide abil.

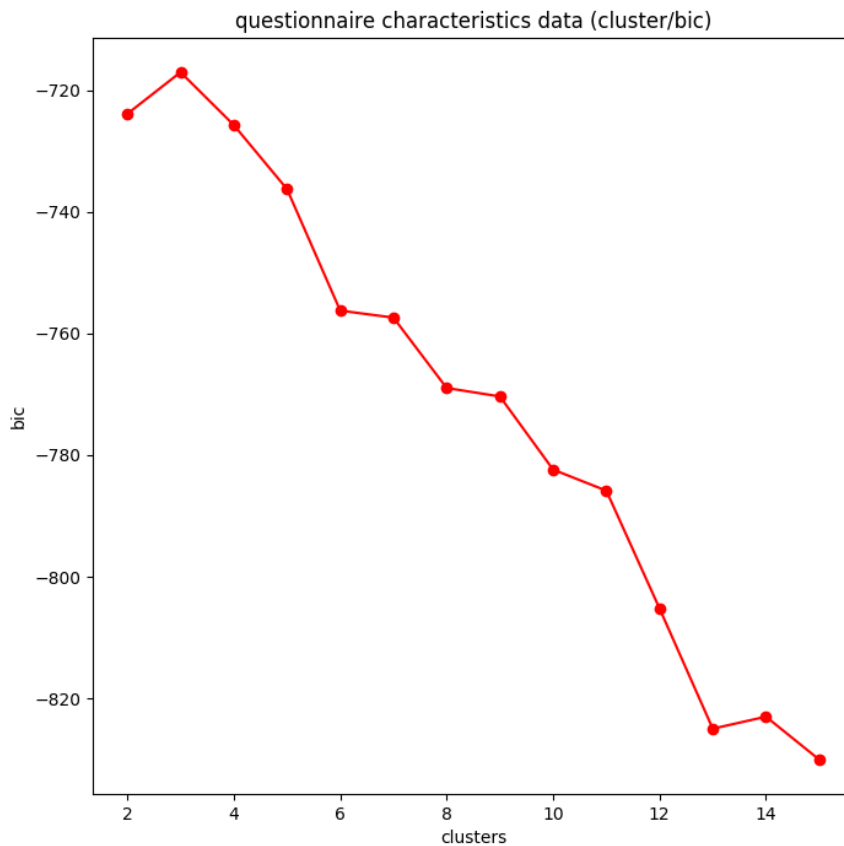


Joonis 9. Robotite programmeerimise kursuse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.

4.1.3 Sissejuhatus erialasse

Sissejuhatus erialasse kursuse esimese kuu aktiivsuse näitajate andmehulga BIC-i väärtus oli tõusva graafikuga ja parim keskmiste siluettide koefitsiendi väärtus oli 0,56 kahe klastri korral. Sarnaselt robotite programmeerimise kursusele polnud sissejuhatus erialasse õppeaines semestri esimesel kuul ülesannetega seotud andmeid.

Koondtunnuste parimaks klastrite arvuks määras BIC 3 klastrit (vt Joonis 10). Keskmiste siluettide koefitsient määras parimaks klastrite arvuks 2 klastrit, autor valis BIC-i järgi klastrite arvuks 3.



Joonis 10. Sissejuhatus erialasse kursuse koondtunnuste andmehulga BIC-i väärtuste graafik. Horisontaalteljel on kujutatud klastrite arvu ja vertikaalteljel BIC-i väärtuseid.

Sissejuhatus erialasse semestri esimese kahe kuu aktiivsuse näitajate andmehulkade BIC-i graafik oli taaskord tõusvate väärtustega ning keskmiste siluettide koefitsientide graafik määras klastrite arvuks 2. Sama trend kehtis ka tudengite jooksvate tulemuste ning esitatud ülesannete kordade andmehulkade kohta, mõlema andmehulga parim võimalik klastrite arv oli 2.

4.2 Tudengite andmeid esitavad graafikud

Kõikide andmehulkade kohta genereeris autor graafikud, kus horisontaalsel teljel on kujutatud andmete tüübid, näiteks ülesannete identifikaatorid või siis koondtunnuste

nimetused. Vertikaalsel teljel on kujutatud nende näitajate väärtuseid. Kõikide graafikute vertikaalsel teljel kujutatud andmete jaoks kasutas autor SciPy¹ teegi meetodit *gaussian_filter1d*², et muuta graafikute jooned sujuvaks ning paremini jälgitavaks.

Ühel graafikul kujutatakse loetavuse eesmärgidel maksimaalselt 15 tudengit, iga joon graafikul tähistab üht tudengit. Kõigi andmehulkade kohta genereeriti mitu 15 tudengiga graafikut. Graafikud on sorteeritud kursuste hinnete alusel alustades hindest 0 ning lõpetades hindega 5 või arvestuslike ainete korral hindega 1. Tudengi andmeid esitava graafiku joone kohta on ka märke graafiku legendis, kus on välja toodud joone värv, tudengi Moodle'i identifikaator, tudengi kursuse hinne ning klatri number, kuhu tudeng määrati.

Jooksvad tulemused normaliseeris töö autor enne klasterdamist protsentuaalsele skaalale, et erinevate punktisummadega ülesanded oleksid omavahel võrreldavad ning et graafikutel oleks visuaalselt paremini esindatud see, kas tudeng sai ülesande eest maksimaalsed punktid või siis maksimaalsest väiksema tulemuse. Sarnastel põhjustel normaliseeris autor ka ülesannete esitamise korrad protsentuaalsele skaalale, et genereeritud graafikuid vaadates oleks võimalik aru saada, kas tudeng esitas ülesannet keskmisest rohkem kordi või vähem kordi.

4.3 Klasterdamise tulemused

Käesolevas alapeatükiks toob autor välja klasterdamisel ilmnunud probleemid ja leitud trendid, samuti genereeritud graafikutelt tuvastatud tudengite õpikäitumist kirjeldavad mustrid kõigi kolme vaadeldava kursuse kohta.

4.3.1 Küsitluse koondtunnused

Kõigi kolme kursuse korral olid koondtunnused ainsateks andmehulkadeks, mille korral õnnestus andmed jaotada rohkem kui kaheks klatriks selliselt, et klatriite arv ei oleks ka liiga suur ja oleks võimalik leida seoseid andmete ja klatriite vahel. Selle peamiseks põhjuseks on autori arvates mõistliku suurusega sisendandmete hulk, mis oli sobilik BIC-

¹ <https://docs.scipy.org/doc/scipy/reference/index.html>

² https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian_filter1d.html#scipy.ndimage.gaussian_filter1d

i abil klastrite arvu leidmiseks ning üldiselt andmete edukaks klasterdamiseks. Sellest lähtuvalt usub autor, et edasistes teadusuuringutes tasuks põhjalikumal katsetada koondtunnuste abil tudengite klasterdamist ning leitud tulemuste põhjal tudengeid kirjeldavate profiilide loomist.

Programmeerimise algkursuse andmed on küsitluse koondtunnuste aspektist kõige usaldusväärsemad, kuna küsitlusele vastanud tudengite andmehulk oli kõige suurem. Programmeerimise algkursuse graafikutelt selgub, et tudengite hulgas, kes said hindeks 5 või 4, esineb kõige enam klastreid 3 ja 4. Tulemuseks nulli saanud tudengite klastriteks on kõige sagedamini 0 ja 1. Seega on klasterdamisel võimalik leida seoseid tudengite õppeedukusega, kuid kindlasti tuleks korrektselt põhjendatud järelduste tegemiseks viia läbi täiendavaid uuringuid. Aine läbikukkunud tudengite hulgas esineb rohkem madalamat sisemist regulatsiooni ning keskmisest rohkem mittetõhusate õpistrateegiate kasutamist ja ülesannete vältimist.

Robootika korral on väiksema uuritud tudengite hulga tõttu keerukam mustreid märgata, kuid hinde 1 ja 0 saanud üliõpilaste hulgas on mittetõhusate õpistrateegiate kasutamine samuti suurem kui ülejäänud kursuse tudengitel. Sissejuhatus erialasse aine korral ei eristunud selgeid mustreid.

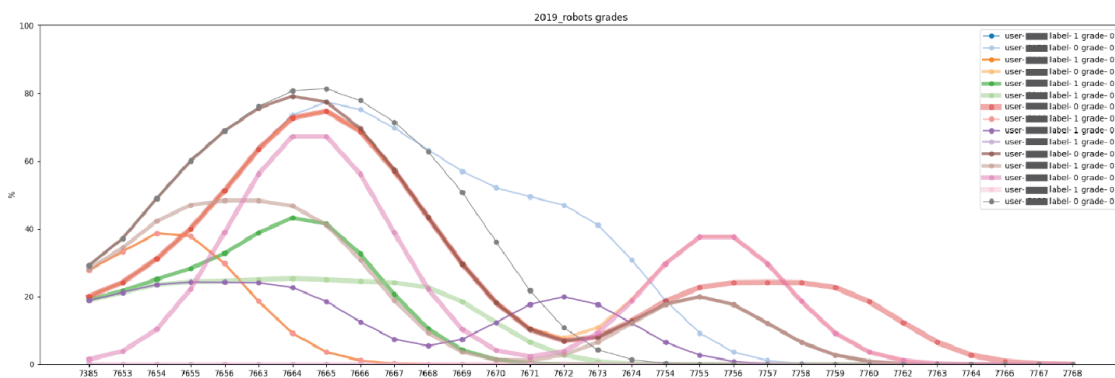
4.3.2 Jooksvad tulemused

Jooksvate tulemuste korral ei õnnestunud andmete klasterdamine eriti edukalt, sest klastrite arvu määravad algoritmid soovitasid üldjuhul jagada andmed kaheks grupiks. Autor usub, et klasterdamiseks valitud andmehulk sisaldas liiga palju tunnuseid ning seetõttu polnud valitud andmed sobilikud klasterdamisülesande lahendamiseks.

Vaatamata sellele on võimalik andmeid kirjeldavate graafikute pealt saada visuaalset informatsiooni tudengite õpikäitumise trendide kohta, kuna autor sorteeris tudengite tehtud ülesanded ajalisse järjekorda. Programmeerimise algkursuse tudengid, kes said õppeaine lõpphindeks viie, said just eriti kursuse esimesel kuul stabiilselt ülesannete tulemuseks kas 100% või siis selle lähedase tulemuse (vt Lisa 12). Graafikutel on ka selgelt näha, et kursuse läbi kukkunud tudengid said juba esimese kuu lõpus ülesannete eest peaaegu alati alla 50% punktidest (vt Lisa 13). Graafikutel on näha ka seda, et osad ülesanded on raskemad, sest isegi heade tulemustega tudengid saavad nende korral maksimumist vähem punkte, ning tudengid, kelle hinded on vahemikus 0 kuni 2 saavad

alla 50% punktidest. Graafikutelt on näha ka seda, et osad kursuse läbi kukkunud tudengid saavad juba esimese kuu lõpus kõigi ülesannete eest 0% punktidest, mis võib viidata sellele, et nad ei esitagi kodutöid. See trend on selgelt näha ka semestri kahe esimese kuu andmeid kujutavatel graafikutel ning kahjuks saavad need kursuse läbikukkunud tudengitest, kes endiselt ülesandeid esitavad, teise kuu lõpuks tihti alla 20% punktidest.

Robotite programmeerimise kursusel hakati ülesandeid esitama alles semestri teisel kuul, seega on andmeid vähem kui programmeerimise algkursuse korral. Ometigi joonistub välja see, et juba teise kuu lõpus said läbikukkunud tudengid ülesannete eest alla 40% punktidest ning leidsid ka neid, kes ülesandeid tõenäoliselt ei esitanud (vt Joonis 11, joonisel on Moodle'i identifikaatorid anonümiseeritud).



Joonis 11. Robotite programmeerimise semestri kahe esimese kuu ülesannete jooksvate hinnete graafik protsentuaalsel skaalal, mis kujutab osa aine läbikukkunud tudengitest

Kursusel sissejuhatus erialasse on kõige keerulisem leida selgeid ülesannete jooksvate tulemustega seotud trende, kuid osad läbikukkunutest ei esitanud ülesandeid.

4.3.3 Aktiivsus

Kahjuks ei õnnestunud ka aktiivsusnäitajate alusel klasterdamine edukalt, vaatamata sellele, et tunnuseid oli vähem kui jooksvate tulemuste andmehulgas. Taaskord erinesid andmehulkade klastrid üksteisest kõige rohkem, kui andmed jagada kahte gruppi. Kuna aktiivsusnäitajad on väga erinevate suurustega, ei ole kahjuks võimalik graafikute pealt tudengite õpikäitumist kirjeldavaid trende leida.

4.3.4 Ülesannete esitamise korrad

Sarnaselt jooksvate tulemuste andmete klasterdamisele oli ülesannete esitamise kordade andmehulkades liiga palju sisendeid, mis muutsid klasterdamise ebaedukaks,

kuna keskmine siluettide koefitsient pakkus parimaks klastrite arvuks alati kaht klastrit. Liigsete dimensioonide vastu kasutas Kadri Umbleja põhikomponentanalüüsi, mida võiks tulevikus kasutada klasterdamise tulemuste parandamiseks [15].

Programmeerimise algkursus graafikutelt selgub, et üldiselt esitavad nii heade kui halbade õpitulemustega tudengid ülesandeid umbes sama palju kordi. Ainsaks erinevuseks on kursuse läbikukkunud tudengid, kellest osa lõpetab üldse ülesannete esitamise. Robotite programmeerimise kursuse korral joonistus välja see, et teise kuu lõpus esitavad hinde 0 saanud tudengid ülesandeid teistest vähem. Sissejuhatus erialasse graafikute põhjal on taaskord keeruline leida õpikäitumise trende.

5 Õppeaine sooritustulemuse ennustamine

Masinõppe mudelite loomiseks otsustas autor kasutada scikit-learn teeki. Peamiseks põhjuseks antud masinõppe raamistiku valimisel oli see, et magistritöö eesmärgiks ei olnud luua üht võimalikult täpselt ennustavat lõplikku masinõppemudelit, vaid hoopis väga paljude erinevate võimalike mudelite katsetamine kasutades eri tüüpi andmeid. Autori eesmärgiks oli välja selgitada, millised logi ning küsitlusega seotud andmed võiksid potentsiaalselt olla parimateks algandmeteks tudengite õpiedukust ennustavate mudelite treenimisel. Scikit-learn sobis selleks eesmärgiks paremini, kuna teek võimaldab kiirelt ning mugavalt mudeleid kohandada ning saadud tulemusi hinnata vastavalt hetkel kasutusel olevatele andmetele.

Magistritöös tehtud tudengite õpiedukusega seotud ennustused saab jagada kahte peamisesse kategooriasse. Esimeseks kategooriaks on kursuse lõpphinnete ennustamine erinevate andmekogumite pealt ning teiseks kategooriaks on küsitluse koondtunnuste väärtuste ennustamine erinevate andmekogumite pealt. Käesolev peatükk annab detailse ülevaate ennustamiseks kasutatud andmetest, loodud mudelitest ning saadud ennustustulemustest.

5.1 Kursuste õppeedukuse klasside ennustamine

Magistritöös ennustas autor masinõppe mudelite abil kolme kursuse lõpphindeid – programmeerimise algkursus (ITI0102), robotite programmeerimine (ITI0201) ning sissejuhatus erialasse (ITI0104).

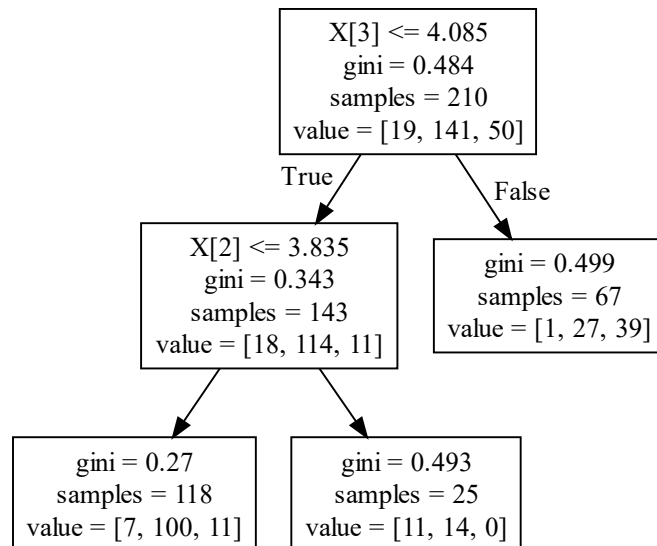
Algselt lõi autor masinõppemudelid, mis ennustasid tudengite lõpphindeid nullist viieni, kuid tegelikult on magistritöö eesmärgiks tudengite üldise õppeedukuse hindamine, mitte ühe konkreetse kursuse tulemuse ennustamine, seega sai tehtud otsus jagada hinded kolme gruppi, mis väljendavad paremini üliõpilaste õppeedukust. Esimesse hindegripi kuuluvad tudengid, kes said kursuse lõpptulemuseks hinde 4 või 5, sellesse gruppi kuuluvad tudengid, kelle õppeedukus on kas suurepärase või väga hea. Teise gruppi jäävad tudengid, kelle tulemus oli, kas hea, rahuldav või kasin (hinded 1–3). Kõik need tudengid läbisid õppeaine, kuid ei olnud parimate hulgas, seega võib nende õppeedukust lugeda keskpäraseks. Viimasesse gruppi kuuluvad tudengid, kes kukkusid kursuse läbi saades puuduliku tulemuse (hinne 0). Kuna tegemist on klassifitseerimisülesandega,

jaotab mudel tudengeid kolme gruppi: hinnetele 4–5 vastab klass 2, hinnetele 1–3 vastab klass 1 ning hindele 0 vastab klass 0 [40]. Arvestuslike kursuste korral – nagu sissejuhatus erialasse – ei ole võimalik kasutada samasugust gruppideks jaotamise süsteemi nagu hindeliste ainete korral, seega kasutavad ka masinõppemudelid tudengeid klassifitseerimisel kahte rühma jagades tulemusi vastavalt klassideks arvestatud (1) ning mittearvestatud (0).

Teiseks oluliseks aspektiks on kursuse läbikukkumisoskus olevate tudengite võimalikult kiire tuvastamine, et õppejõul oleks võimalik hätta jäänud üliõpilasi aidata juba näiteks semestri keskel. Alapeatükis “Õpialalüütika Tallinna Tehnikaülikoolis” tõi autor välja varasemalt TalTechis läbiviidud õpiedukust hindavate uuringute peamise miinuse, et ennustusi tehti tagasivaatavalt kasutades näiteks semestrite jooksul kogutud EAP-de arve või kursuste lõpphindeid. Kahjuks ei ole võimalik selliseid tudengite õppeedukust hindavaid masinõppemudeleid kasutada parajasti abivajavate õppurite tuvastamiseks ning toetamiseks, sest selleks hetkeks, kui neil on kogutud kokku näiteks kaastudengitega võrreldes vähem EAP-sid, on nad juba mitmeid aineid läbi kukkunud.

Sellest lähtuvalt oli käesoleva magistritöö kontekstis oluline kasutada Moodle'i logidest leitud jooksvaid andmeid, mis tekivad semestri vältel. Autor otsustas 2019/2020 õppeaasta sügissemestri logiandmetest kasutada ajaliselt kahte gruppi jagatud andmeid. Esimeseks grupiks on semestri esimesel kuul ehk septembris logisse kogunenud andmed, teiseks grupiks on kahe esimese kuu andmed ehk septembri ning oktoobri logiandmed. Seega treenis autor mudeleid ning ennustas nendega lõpptulemuste klasse logiandmete puhul alati nii semestri esimese kuu andmetega kui ka esimese kahe kuu andmetega. Loodud programmi abil on võimalik leida ka teiste huvipakkuvate ajavahemike andmed ning nende kasutamine mudelite treenimiseks.

Kõik loodud masinõppe mudelid kasutavad otsustuspuu algoritmi, selleks on kaks peamist põhjust. Esiteks on eelnevalt TalTechis läbiviidud uuringutest selgunud, et otsustuspuud on olnud heaks vahendiks tudengite õppeedukuse klasside ennustamisel. Teiseks põhjuseks on see, et otsustuspuude abil on võimalik paremini mõista, mille alusel mudel otsuseid teeb, ning kuna autor genereeris Graphviz-i abil mudelite poolt loodud otsustuspuude joonised, on võimalik neid ka analüüsida (vt Joonis 12).



Joonis 12. Näide programmeerimise algkursuse küsitluse koondtunnuseid kasutanud mudeli poolt genereeritud otsustuspuust.

5.1.1 Ühe koondtunnuse kaupa ennustamine

Ühe koondtunnuse kaupa õppeedukuse klasside ennustamise eesmärgiks oli välja selgitada, millised koondtunnused ennustavad õppeedukuse klasse kõige paremini. Autor ennustas ühe koondtunnuse kaupa õppeedukuse klasse kõigi kolme kursuse jaoks - programmeerimise algkursus, robotite programmeerimine ja sissejuhatus erialasse.

Kuna küsitluste koondtunnuste logiandmete hulk on väiksem üldisest logiandmete hulgast, siis algandmed jaotatakse test- ning treeningandmeteks nõnda, et kõik võimalikud klassid oleksid olemas nii treening- kui ka testandmetes, et tagada võimalikult kvaliteetne treenimine ning võimalikult tõetruu *score* väärtus.

Kõik Robotite programmeerimise ning sissejuhatus erialasse kursust võtnud tudengid ei täitnud programmeerimise algkursuse käigus läbiviidud küsitlust, seega oli nende ainete tegijate hulgas märksa vähem küsitlusele vastanud. Kahjuks tekkid seetõttu nii Robotite programmeerimise kui ka sissejuhatus erialasse puhul logiandmete test- ning treeningandmeteks jagamisel probleeme, kuna võrdlemisi väikeste andmehulkade korral ei olnud võimalik treeningandmehulka kõiki võimalikke klasse panna. Näiteks sissejuhatus erialasse andmehulga puhul olid treeningandmetes ainult tudengid, kes olid kursuse läbinud ning testandmetes oli üks kursuse läbikukunud tudeng. Robotite programmeerimise aine kukkus küsitlusele vastanud tudengitest läbi ainult kaks inimest, mis mõjutas samuti ennustustulemusi. Seetõttu ei ole küsitluse vastuste ning koondtunnuste kasutamine õppeedukuse klasside ennustamisel robotite

programmeerimine ja sissejuhatus erialasse kursuste korral usaldusväärne ning puudulikud andmed mõjutasid tugevalt ennustumudelite tööd ja tulemusi, mudelid pakkusid tudengitele alati sama õppeedukuse klassi.

Autor tegi iga küsitluse koondtunnuse jaoks eraldi otsustuspuu algoritmiga masinõppemudeli, mille käitumist muudeti *min_samples_split* sisendi abil. Sisend *min_samples_split* tähistab miinimum tudengite arvu, mis peab olema otsustuspuu sõlmes, et puu saaks hargneda. Autor leidis iga koondtunnuse masinõppemudeli jaoks parima võimaliku *min_samples_split* suuruse, et saavutada parim võimalik ennustustulemus. Kõikide erinevate treeningandmete põhjal loodud mudelite jaoks genereeriti ka otsustuspuud kujutav graaf. Programmeerimise algkursuse korral oli testandmehulga suuruseks N=175 ning treeningandmehulga suuruseks N=87 [27].

5.1.2 Logi ning küsitluse andmete alusel ennustamine

Loodud mudelid saab jagada kahte gruppi – esimesse gruppi kuuluvad ühe kursuse andmete põhjal loodud masinõppe mudelid, teise gruppi kuuluvad üldised masinõppe mudelid, mida treeniti ühe kursuse andmete peal ning testiti teise kursuse andmete peal. Kõiki üldiseid mudeleid treeniti kõigi kolme kursuse andmete peal, ning testiti leitud mudelit nende kahe kursuse andmete peal, millega mudelit parajasti ei treenitud. Lisaks sellele treeniti kõiki mudeleid nii esimese kuu logiandmetega kui ka kahe esimese kuu logiandmetega. Robotite programmeerimise kursusel ning aines sissejuhatus erialasse puuduvad semestri esimese kuu jooksvate hinnetega ja ülesannetega seotud andmed, sest tudengid ei olnud selleks hetkeks veel ühtegi ülesannet esitanud.

Ühe kursuse andmeid kasutavad mudelid jaotusid järgnevalt -

- Aktiivsuse näitajaid kasutav mudel
- küsitluse küsimuste vastuseid kasutav mudel
- jooksvate hinnete andmeid kasutav mudel
- keskmisi jooksvaid hindeid kasutav mudel
- ülesannete esitamiste kordade andmeid kasutav mudel
- küsitluse koondtunnuseid kasutav mudel
- jooksvaid hindeid ja aktiivsuse näitajaid kasutav mudel
- keskmisi jooksvaid hindeid ja aktiivsuse näitajaid kasutav mudel
- küsitluse küsimusi ja logiridade arve (aktiivsuse näitaja *activity*) kasutav mudel

- küsitluse koondtunnuseid ja logiridade arve (aktiivsusanäitaja *activity*) kasutatav mudel
- küsitluse koondtunnuseid ja jooksvaid hindeid kasutatav mudel
- küsitluse koondtunnuseid ja keskmiseid jooksvaid hindeid kasutatav mudel
- küsitluse koondtunnuseid ja aktiivsusanäitajaid kasutatav mudel

Üldised mudelid kasutavad järgmiseid andmeid –

- aktiivsusanäitajaid kasutatav mudel
- keskmisi jooksvaid hindeid kasutatav mudel
- keskmisi jooksvaid hindeid ja aktiivsusanäitajaid kasutatav mudel
- küsitluse koondtunnuseid ja keskmiseid jooksvaid hindeid kasutatav mudel
- küsitluse koondtunnuseid ja aktiivsusanäitajaid kasutatav mudel

Autor lõi iga kasutatud andmekomplekti kohta eraldi otsustuspuu algoritmil põhineva masinõppemudeli, mille tulemusi reguleeriti kasutades nii varem kirjeldatud *min_samples_split* sisendit kui ka *max_depth* sisendit, mis määrab otsustuspuu maksimaalse sügavuse. Kõigi treenitud mudelite kohta genereeris autor ka otsustuspuud kujutava joonise [27].

Ühe kursuse andmete põhjal loodud masinõppe mudelite sisendandmed on jagatud järgmiselt – 80% andmetest on treeningandmehulgas ning 20% andmetest on testandmehulgas. Andmed segatakse enne ennustamist koos seemnega (*seed*), et tulemusi oleks võimalik korrata.

Logi ning küsitluse andmete pealt tehti õppe edukuse klassi ennustusi kõigi kolme kursuse jaoks. Nagu eelmises alapeatükis kirjeldati, pole kahjuks küsitluse andmed robotite programmeerimise ning sissejuhatus erialasse kursuste korral usaldusväärsed, kuid autor leidis need tulemused võrdluse eesmärkidel siiski tulemuste tabelitesse Lisa 4 ja Lisa 5. Programmeerimise algkursuse korral oli kasutatavate mudelite treeningandmete suuruseks $N=384$ ja testandmete suuruseks $N=96$. Küsitluse andmeid sisaldavate andmehulkade suuruseks oli treeningandmete korral $N=210$ ja testandmete korral $N=52$. Robotite programmeerimise kursuse logi treening- ja testandmete hulkade suurused olid vastavalt $N=89$ ja $N=22$. Küsitluse andmete korral olid andmehulkade suurused vastavalt $N=50$ ja $N=13$.

Sissejuhatus erialasse kursuse logi treening- ja testandmete hulkade olid vastavalt $N=85$ ning $N=21$. Küsitluse andmete korral olid sissejuhatus erialasse andmehulkade suurused treeningandmete korral $N=49$ ja testandmete korral $N=12$.

Üldiste mudelite korral olid andmehulkade suuruse järgmised –

- Programmeerimise algkursus $N=480$ ja küsitlusandmetega $N=262$
- Robotite programmeerimine $N=111$ ja küsitlusandmetega $N=63$
- Sissejuhatus erialasse $N=106$ ja küsitlusandmetega $N=61$

5.1.3 Tunnuste valimisega mudelite treenimine

Õppeedukust ennustavate mudelite lihtsustamiseks ning täpsustamiseks otsustas autor kasutada tunnuste valimist (*feature selection*). Tunnuste valimisel kasutas autor andmehulkasid, mida sai kasutada üldistes mudelites, mida treeniti ühe kursuse andmete peal, kuid testiti teise õppeaine andmete peal. Kuna tunnuste valimisest kasutatakse mudelite korral, mille sisendiks on mitu tunnust, siis kasutas autor andmehulkadest aktiivsusanäitajaid, küsitluse koondtunnuseid, aktiivsusanäitajaid koos keskmiste jooksvate hinnetega, aktiivsusanäitajaid koos küsitluse koondtunnustega ning küsitluse koondtunnuseid koos keskmiste jooksvate hinnetega.

Autor kasutas parimate tunnuste valmiseks scikit-learn meetodit *SelectKBest*, mis valib välja k väärtusega määratud arvu tunnuseid, millel olid parimad skoorid. Skoori leidmiseks kasutas autor Chi ruudu meetodit. Scikit-learn teegi meetod *chi2* leiab Chi ruudu väärtuse kõigi mittenegatiivsete tunnuse ning klassifitseerimisülesande klasside vahel. Chi ruudu meetod suudab tuvastada tunnused, mis on tõenäoliselt kõige vähem seotud klassidega ning seega pole vajalikud ka klassifitseerimisülesande lahendamisel. Kuna Chi ruudu korral peavad sisendandmete väärtused olema positiivsed, pidi autor aktiivsusanäitaja *diff_from_average*, mis võis olla ka negatiivne, muutma positiivseks [41], [42].

Autor leidis tunnuste valimisel parima võimaliku tunnuste arvu, kasutades programmeerimise algkursuse andmeid, eraldi nii semestri esimese kuu andmehulkadele kui ka semestri kahe esimese kuu andmehulkadele. Samu parimate tunnuste arve kasutas autor ka robotite programmeerimise ning sissejuhatus erialase andmehulkade korral. Leitud parimad k arvud on esitatud Lisas 7 esitatud tunnuste valimisega treenitud mudelite tulemuste tabelites.

5.2 Võrdlusbaas

Selleks, et tudengite õppeedukuse ennustustulemusi saaks omavahel võrrelda otsustas autor luua võrdlusbaasi. Võrdlusbaasi ennustustulemuste saamiseks treenitakse mudel ainult tudengi logiridade arvu andmetega (aktiivsusnäitajate üks tunnus *activity*). Võrdlusbaas leitakse eraldi kõigi erinevate ennustustüüpide jaoks, võrdlusbaas on eraldi nii semestri esimese kuu kui ka esimese kahe kuu andmete kohta, lisaks sellele on eraldi võrdlusbaasid leitud ka kõigi kolme uuritava kursuse jaoks – sissejuhatus erialasse, robotite programmeerimine ning programmeerimise algkursus. Võrdlusbaasis on leitud ka mudelite jaoks, mida treeniti ühe kursuse andmete peal ning testiti teise õppeaine andmehulkade peal.

Võrdlusbaaside leidmise eesmärgiks on teha kindlaks, kas täpsemad logidest leitud tunnused ning küsitlusega seotud õpikäitumist kirjeldavad tunnused parandavad masinõppemudelite tulemusi võrreldes ainult logiridade arvu arvesse võtvate mudelitega. Võrdlusbaasi masinõppemudeliks oli ilma täiendavate seadistusteta otsustuspuu algoritmi kasutatav mudel. Võrdlusbaasi ennustustulemused leiab Lisadest 1 ja 2.

5.3 Küsitluse koondtunnuste ennustamine

Küsitluse koondtunnused on pideva väärtusega ning väärtused jäävad vahemikku 0 kuni 6, seega tuli koondtunnuste väärtused jagada klassideks. Selleks otsustas autor ümardada koondtunnuste väärtused täisarvudeks, saades 7 erinevat klassi vahemikus 0 kuni 6. Masinõppe mudelitega ennustati igat küsitluse koondtunnust eraldi.

Küsitluse koondtunnuseid ennustas autor ainult programmeerimise algkursuse aine andmete pealt, kuna küsitlus viidi läbi seda kursust võtnud tudengite hulgas. Seetõttu eeldas autor, et seos logiandmete ning koondtunnuste vahel on kõige tugevam just programmeerimise algkursuse korral. Autori eesmärgiks koondtunnuste ennustamisel oli välja selgitada, kas leidub seoseid logiandmete ning konkreetsete koondtunnuste vahel. Sarnaselt kursuse lõpptulemuste klasside ennustamisele ennustati küsitluse koondtunnuseid nii logi esimese kuu andmete kui ka esimese kahe kuu andmete pealt.

Ennustamisel kasutatud andmehulkadeks olid tudengite aktiivsuse näitajad, jooksvad tulemused, ülesannete esitamise arvud ja keskmised jooksvad tulemused. Lisaks sellele

kasutas autor ka kombineeritud andmehulkasid - aktiivsusnäitajad koos keskmiste jooksvate tulemustega ja jooksvad hinded koos aktiivsusnäitajatega.

Autor tegi iga kasutatud andmehulga jaoks eraldi otsustuspuu algoritmil põhineva masinõppemudeli ja muutis nende mudelite käitumist sisendi *min_samples_split* abil. Kõigi mudelite kohta genereeriti ka joonis otsustuspuust [27].

Ühe kuu programmeerimise algkursuse andmetega loodud masinõppemudelite korral oli treeningandmehulga suuruseks N=210 ja testandmehulga suuruseks N=52. Kahe kuu andmete kasutamisel oli treeningandmete suuruseks N=210 ning testandmete suuruseks N=52. Seega kõik küsitlusest osavõtnud tudengid olid aktiivsed nii esimesel kui ka teisel semestri kuul.

5.4 Korrelatsioon Moodle'i logide ja psühholoogiliste faktorite küsitluse koondtunnuste vahel

Koondtunnuste ennustamine Moodle'i logide andmete pealt oli autori poolt eksperimentaalne lähenemine kolmandale uurimusküsimusele vastuste leidmiseks, mille eesmärgiks oli tuvastada seoseid andmete vahel, mida võib-olla ei pruugi traditsioonilisi meetodeid kasutades leida. Seega otsustas autor, et koondtunnuste ennustamine polnud piisav ning vaja on leida ka korrelatsioonid Moodle'i logide andmehulkade ning kõigi õppimisega seotud psühholoogiliste faktorite küsitluse koondtunnuste vahel. Korrelatsioonide leidmiseks kasutas autor õppeainete sissejuhatus erialasse, robotite programmeerimine ning programmeerimise algkursus semestri esimese ning esimese kahe kuu andmeid.

Moodle'i logide pealt leitud näitajatest kasutas autor üldiseid näitajaid, mida sai kasutada treenides andmeid ühe kursuse peal ning testides treenitud masinõppemudelit teise õppeaine andmete peal. Autor kasutas aktiivsusnäitajatest korrelatsioonide leidmiseks *activity* näitajat, mis tähistab logiridade arvu ning mis oli kasutusel ka võrdlusbaasi leidmiseks masinõppe mudelitele. Lisaks näitajale *activity* kasutas autor aktiivsusnäitajatest korrelatsioonide leidmisel küsitluse koondtunnustega ka tunnuseid *diff_from_average*, *percentage_from_max*, *number_of_active_days* ja *number_of_active_weeks*. Lisaks aktiivsusnäitajatele kasutas autor ka keskmist jooksvat tulemust ehk *average_grade*. Siinkohal on oluline märkida, et kursustel robotite

programmeerimine ning sissejuhatus erialasse ei olnud logides semestri esimese kuu kohta jooksvate hinnetega seotud andmeid, seega ei saanud ainult semestri esimese kuu andmeid kasutades leida ka korrelatsiooni küsitluse koondtunnuste ning tudengite keskmiste jooksvate hinnete vahel.

Korrelatsioonide leidmiseks kasutas autor SciPy¹ teegis olevat meetodit *scipy.stats.spearmanr*, mille abil saab leida Spearmani korrelatsioonikoefitsiendi. Spearmani korrelatsioon on mitteparameetiline viis kahe andmehulga vahelise monotoonsuse mõõtmiseks. Spearmani korrelatsiooni korral ei pea mõlemad andmehulgad olema normaaljaotusega. Spearmani korrelatsioonikoefitsiendi väärtused jäävad vahemiku -1 kuni 1. Väärtus 0 näitab, et korrelatsiooni ei esine üldse, positiivne korrelatsioonikoefitsiendi väärtus näitab, et kui ühe andmehulga väärtused suurenevad, siis suurenevad ka teise andmehulga väärtused. Negatiivne väärtus näitab, et kui ühe andmehulga väärtused kahanevad, siis teise omad suurenevad [43].

¹ <https://www.scipy.org/index.html>

6 Tulemused

Autor esitab antud peatükis magistritöö tulemused. Leitud ennustustulemusi kõrvutatakse võrdlusbaasiga, lisaks sellele toob autor välja ka kasutatud andmehulkadega seotud probleemid. Autor annab ka soovitusi tulevikus läbiviidavate uuringute jaoks.

6.1 Õppedukuse ennustamine

Õppeedukuse mudelite ennustustulemuste hindamisel ning võrdlemisel on äärmiselt oluline saadud tulemuste kõrvutamine võrdlusbaasi tulemustega. Võrdlusbaasi tulemused leiti kõigi õppeedukuse ennustusmudelite tüüpide jaoks kasutades täiendavate seadistustusteta otsustuspuu algoritmil põhinevaid masinõppe mudeleid, mida trenniti vaadeldava õppeaine tudengite Moodle'i aktiivsuslogide logiridade arve kasutades. Seejärel testiti trennitud masinõppemudeleid, kas sama kursuse tudengite andmetest koosneva testandmehulgaga või siis teise kursuse tudengite andmetega. Õppeedukust ennustavad masinõppemudelid klassifitseerivad tudengeid hindeliste ainete korral kolme gruppi – (1) tudengid, kes kukkusid aine läbi ja said tulemuseks hinde 0, (2) tudengid kes said lõpphindeks, kas 1, 2 või 3 ja (3) üliõpilased, kes said kursuse hindeks, kas 4 või 5. Arvestuslike kursuste korral jagati tudengid kahte klassi: need, kes kukkusid õppeaine läbi või siis üliõpilased, kes said tulemuseks „arvestatud“. Võrdlusbaasi ühe kuu logi andmete ennustustulemused leiab Lisast 1 ning semestri kahe esimese kuu ennustustulemused Lisast 2.

Autor kõrvutab saadud ennustustulemusi võrdlusbaaside tulemusega erinevate vaadeldud ennustusmudelite tüüpide kaupa. Ennustusmudelite tüüpideks on kolme uuritud kursuse Moodle'i logidel põhinevad andmed. Logi andmed omakorda jagunevad kahte gruppi: semestri esimese kuu logiandmed ning semestri esimese kahe kuu logiandmed. Lisaks logiandmetele kasutab autor ka küsitluse küsimuste ning koondtunnuste andmeid, mida kasutati koos Moodle'i aktiivsuslogidest leitud andmetega. Eraldi mudelitüüpideks on masinõppemudelid, mida trenniti ühe kursuse andmete peal, kuid testiti teise kursuse andmetega.

6.1.1 Küsitluse andmete kasutamine programmeerimise algkursuse õppeedukuse klasside ennustamisel

Autor ennustas iga koondtunnuse pealt eraldi õppeedukuse klasse õppeaine programmeerimise algkursus tudengitele. Üksikute koondtunnuste pealt ennustamise eesmärgiks oli välja selgitada, millised koondtunnused ennustavad õppeedukust kõige paremini. Kõik koondtunnuste abil leitud ennustustäpsused jäid vahemikku 57,5%–67,8% (vt Lisa 3). Kõige paremad ennustustulemused saavutati kasutades koondtunnuseid enesetõhusus ning keeruliste ülesannete vältimine (vt Tabel 3). Nendele koondtunnustele järgnesid 65,5% suuruse ennustustäpsusega sooritusele orienteeritud motivatsioon, läbikukkumise vältimisele suunatud soorituseesmärgid, matemaatikaalased uskumused, pingutus programmeerimiskursusel, huvi, lihtsate ülesannete eelistamine ning sotsiaalne õppimine.

Tabel 3. Programmeerimise algkursuse õppeedukuse klasside ennustamises parimaks osutunud koondtunnuste ennustustulemused.

Koondtunnus	Ennustustäpsus	Hinde 0 täpsusskoor
1 kuu andmete võrdlusbaas	39,6%	40,5%
2 kuu andmete võrdlusbaas	49%	56,8%
enesetõhusus	67,8%	12,5%
keeruliste ülesannete vältimine	67,8%	12,5%

Üldiselt ennustasid koondtunnused halvasti läbikukkunud tudengite tulemusi, sest hinde 0 täpsusskoorid jäid vahemiku 0%–12,5%. Hinde 0 halvad täpsusskoorid võivad olla tingitud kasutatud andmetest (vt Joonis 13), kuna psühholoogiliste tunnuste küsitlus viidi tudengite hulgas läbi semestri lõpus, siis kahjuks ei vastanud mitmed kursuse läbi kukkunud tudengid küsitlusele, mistõttu puudub osade väljalangenute kohta informatsioon. Teiseks põhjuseks võib olla see, et tudengite käest küsiti nõusolekut küsitluse vastuste kasutamiseks. On võimalik, et tudengid, kes ei tundnud end oma tulemustes kindlalt, ei andnud ka nõusolekut oma vastuste kasutamiseks käesolevas magistritöös. Seetõttu ongi küsitluse andmeid sisaldavates andmehulkades vähem

tudengeid, kes kukkusid kursuse läbi, mis omakorda muudab täpsema ennustamise keerukamaks.



Joonis 13. Tudengite arv programmeerimise algkursuse hindeklassides küsitluse andmete korral.

Kõige paremini ennustasid hindeklassi 0 koondtunnustest pealesurutud motivatsioon, sotsiaalse ärevusega seotud uskumused, enesetõhusus, huvikaotus, keeruliste ülesannete vältimine ning sisemine regulatsioon. Seega olid üldiselt parimateks koondtunnusteks ennustamisel enesetõhusus ning keeruliste ülesannete vältimine. Koondtunnuseid kasutavate masinõppemudelite ennustustulemustega täismahus tabeli leiab Lisades 4 ja 5.

Küsitluse küsimusi kasutanud masinõppemudeli tulemused olid paremad nii semestri esimese kui ka esimese kahe kuu võrdlusbaasi tulemustest (vt Tabel 4). Eriti märkimisväärne on täpsusskoor, mis on tervelt 100% aine läbi kukkunud tudengite korral ning ka saagisskoor, mis on samuti 100%. Tudengite väljalangemise probleemi korral on äärmiselt oluline tuvastada just neid tudengeid, kes võivad aine läbi kukkuda.

Autor leidis eraldi ka logiridade arvu ning küsitluse küsimuste peal treenitud mudeli ning logiridade arvu ja küsitluste koondtunnuste peal treenitud mudeli ennustustulemused, et selgitada välja, kas ainult võrdlusbaasimudelil kasutavate logiridade lisamine küsitluse andmetele parandab tulemust.

Küsitluse küsimusi ning semestri esimese kuu logiridade arvu kasutanud mudeli tulemus oli pisut parem ainult küsitluse küsimusi kasutanud mudeli ennustustäpsusest ning

märgatavalt parem võrdlusbaasi tulemustest. Semestri esimese kuu logiridade arvu ning koondtunnuseid kasutanud mudeli ennustustäpsus oli halvem ainult koondtunnuseid kasutanud mudeli tulemusest, kuid samuti märgatavalt parem ainult ühe kuu logiridade arve kasutanud masinõppemudelilist.

Semestri kahe esimese kuu logiridade arve ning küsitluse küsimusi kasutanud mudeli ennustustäpsus oli parem võrdlusbaasi tulemusest ning võrdne ainult küsitluse küsimusi kasutanud mudeli tulemusega. Küsitluse koondtunnuseid koos logiridade arvuga kasutanud mudeli ennustustäpsus oli kahe kuu andmeid kasutades madalam ühe kuu andmeid kasutanud sama mudeli tulemusest. Lisaks sellele oli tulemus madalam ainult koondtunnuseid kasutanud masinõppemudeli tulemusest. Küll aga oli mudeli tulemus parem kahe kuu andmete võrdlusbaasist ning saavutas 25% täpsusskoori hinde 0 klassifitseerimisel.

Tabel 4. Programmeerimise algkursuse õppeedukuse klasside ennustustulemused küsitluse küsimuste ning koondtunnuste korral.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor
1 kuu andmete võrdlusbaas	39,6%	40,5%
2 kuu andmete võrdlusbaas	49%	56,8%
küsitluse küsimused	67,3%	100%
küsitluse koondtunnused	78,9%	0%
küsitluse küsimused koos 1 kuu logiridade arvudega	69,2%	0%
küsitluse küsimused koos 2 kuu logiridade arvudega	69,2%	0%
küsitluse koondtunnused koos 1 kuu logiridade arvudega	71,2%	0%
küsitluse koondtunnused koos 2 kuu logiridade arvudega	67,3%	25%

Küsitluse koondtunnuste andmehulga (kokku 8 koondtunnust) masinõppemudeli ennustustäpsus oli suurim programmeerimise algkursuse jaoks saadud tulemustest (vt

Tabel 4), kuid hinde 0 täpsusskooriks oli 0%, mis on madalam mõlema ajaperioodi võrdlusbaasi tulemustest. Teisalt oli hinnete 1 kuni 3 täpsusskooriks 87,5% ning hinnete 4 kuni 5 täpsusskooriks 81,3%. Chi ruutude abil leitud parimate tunnuste valimine ei muutnud ennustustulemust koondtunnuste puhul paremaks

Küsitluse küsimuste ning küsitluse koondtunnuste täielikud ennustustulemused leiab Lisast 4 koos semestri esimese kuu Moodle'i logisid kasutavate mudelite ennustustulemustega ning Lisast 5 koos kahe esimese kuu Moodle'i logisid kasutavate mudelite ennustustulemustega.

6.1.2 Küsitluse andmete kasutamine robotite programmeerimise ja sissejuhatus erialasse õppeedukuse klasside ennustamisel

Magistritöö käigus selgus, et robotite programmeerimise ning ka sissejuhatus erialasse kursusel osalenud tudengitest ei vastanud programmeerimise algkursuse raames läbiviidud õppimisega seotud psühholoogiliste tegurite küsitlusele piisavalt üliõpilasi, et ennustustulemused, mis kasutavad küsitluse küsimusi või küsitluse koondtunnuseid, oleks usaldusväärsed. Robotite programmeerimise kursuse treeningandmehulk oli suurusega N=50 ja testandmehulga suuruseks oli N=13, lisaks sellele oli kõikidest küsitlusele vastanud tudengites aine läbi kukkunud ainult kaks üliõpilast. Sissejuhatust erialasse andmete korral oli treeninghulga suuruseks kõigest N=49 ja testandmehulga suuruseks N=12, kõikidest küsitlusele vastanud tudengitest kukkus õppaine läbi ainult üks tudeng. Kahjuks ei saa seetõttu robotite programmeerimise ning sissejuhatus erialasse kursuste korral küsitluse andmeid kasutavate mudelit ennustustulemusi kasutada järelduste tegemiseks. Tulemused on siiski leitavad koos teiste robotite programmeerimise, sissejuhatus erialasse ning programmeerimise algkursuse ennustustulemustega Lisades 4 ja 5.

6.1.3 Moodle'i logiandmete kasutamine programmeerimise algkursuse õppeedukuse klasside ennustamisel

Tabelid semestri esimese kuu andmete peal treenitud mudelite täielike tulemustega leiab magistritöö lõpus olevast Lisast 4.

Semestri esimese kuu aktiivsusnäitajaid kasutanud ennustusmudeli, mille sisendandmed sisaldasid nii võrdlusbaasis kasutatud tudengite logiridade arve kui ka teisi aktiivsusega seotud tunnuseid näiteks aktiivsete nädalate arvu. Ennustuskoor oli võrdlusbaasist

suurem, kuid tuleb märkida, et täpsusskoor aine läbikukkunud tudengite klassifitseerimisel oli väiksem võrdlusbaasi omast (vt Tabel 5).

Jooksvate tulemuste andmeid kasutanud masinõppemudel oli parem võrdlusbaasist, ennustustulemuse täpsuseks oli 68,8% ning täpsusskoor hinde 0 ennustamisel oli erinevalt aktiivsuse näitajatest parem võrdlusbaasi täpsusskooriga. Keskmisi jooksvaid tulemusi kasutanud masinõppemudeli ennustustäpsus oli taaskord parem võrdlusbaasist, kuid väiksem lihtsalt jooksvaid hindeid kasutanud mudeli ennustustulemusest. Seega võib järeldada, et keskmise jooksva tulemuse leidmisel läks osa informatsioonist kaduma. Ülesannete esitamiste kordade andmehulga ennustustulemuse täpsuseks oli 60,4%, mis oli samuti parem võrdlusbaasi tulemusest.

Tabel 5. Programmeerimise algkursuse semestri esimese kuu Moodle'i logisid kasutavate andmehulkade ennustustulemused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor
1 kuu andmete võrdlusbaas	39,6%	40,5%
aktiivsuse näitajad	52,1%	32,4%
jooksvad tulemused	68,8%	55,6%
keskmised jooksvad tulemused	58,3%,	48,7%
ülesannete esitamiste korrad	60,4%,	48,7%
koondtunnused aktiivsuse näitajatega	78,9%	0%
koondtunnused keskmiste jooksvate tulemustega	76,9%	0%
koondtunnused jooksvate tulemustega	75%	0%
aktiivsuse näitajad jooksvate tulemustega	65,63%	54,05%
aktiivsuse näitajad keskmiste jooksvate tulemustega	58,33%	45,95%

Kahe andmehulga kombineerimisel saadud andmehulkade mudelitest osutus parimaks küsitluse koondtunnused koos aktiivsusanäitajatega, mis saavutas täpselt sama tulemuse ainult aktiivsusanäitajaid kasutanud mudeliga. Kõik küsitluse koondtunnuseid esimese kuu logiandmetega kombineerinud mudelid said hinde 0 täpsusskooriks 0% (vt Tabel 5). Logiandmed ei parandanud koondtunnuste mudelite tulemust, vaid tegid selle pigem halvemaks (vt Tabel 5). Aktiivsusanäitajaid ja jooksvaid tulemusi kasutanud mudeli ennustustäpsus oli parem võrdlusbaasist ja aktiivsusanäitajaid kasutavast masinõppemudelist, kuid halvem ainult jooksvaid tulemusi kasutanud mudelist. Kombineeritud andmehulkasid kasutanud mudelitest oli halvim aktiivsusanäitajate ja keskmiste jooksvate tulemuste mudel, mille tulemus oli sama ainult semestri esimese kuu keskmisi jooksvaid tulemusi kasutanud mudeli omaga.

Kõikide semestri esimese kahe kuu andmeid kasutavate masinõppemudelite ennustustäpsused olid paremad võrdlusbaasi omast (vt Tabel 6). Aktiivsusanäitajate mudel saavutas üpris suure hinde 0 saagisskoori, milleks oli 74,1%. Sellest aga veel parema hinde 0 täpsus- ja saagisskoori (82,9%) sai jooksvate tulemuste peal treenitud mudel. Keskmisi jooksvaid tulemusi kasutav mudel sai taaskord kehvema tulemuse jooksvaid hindeid kasutanud mudelist. Ülesannete esitamiste kordade mudel sai sama hinde 0 täpsusskoori, mis jooksvate hinnete masinõppemudel.

Sarnaselt semestri esimese kuu logiandmeid ning küsitluse koondtunnuseid kombineerinud mudelitele, ei parandanud küsitluste koondtunnuste masinõppemudeli ennustustäpsust ka semestri kahe esimese kuu andmed. Küsitluse koondtunnused koos aktiivsusanäitajatega ja küsitluse koondtunnused koos jooksvate tulemusega said täpselt sama ennustustulemuse ainult koondtunnuseid kasutanud mudeliga. Koondtunnuseid ja keskmisi jooksvaid tulemusi kasutanud mudel saavutas aga isegi kehvema ennustustäpsuse. Aktiivsusanäitajad kombineeritult keskmiste jooksvate tulemustega andsid nii ainult aktiivsusanäitajaid kui ka ainult keskmisi jooksvaid tulemusi kasutanud mudelitest parema tulemuse. Seevastu aktiivsusanäitajad koos jooksvate tulemusega andsid ainult jooksvaid tulemusi kasutavast masinõppemudelist halvema ennustustäpsuse.

Tabelid semestri esimese kahe kuu andmete peal treenitud mudelite täielike ennustustulemustega leiab magistritöö lõpus olevast Lisast 5.

Tabel 6. Programmeerimise algkursuse semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustulemused

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsuskoor
2 kuu andmete võrdlusbaas	49%	56,8%
aktiivsuse näitajad	58,3%	54,1%
jooksvad tulemused	76%	78,4%
keskmised jooksvad tulemused	67,7%	75,7%
ülesannete esitamiste korrad	71,9%	78,4%
koondtunnused aktiivsuse näitajatega	78,9%	0%
koondtunnused jooksvate keskmiste tulemustega	73,1%	0%
koondtunnused jooksvate tulemustega	78,9%	0%
aktiivsuse näitajad jooksvate tulemustega	70,8%	75,7%
aktiivsuse näitajad keskmiste jooksvate tulemustega	74%	78,5%

Autor kasutas ka Chi ruudu abil parimate tunnuste valimist aktiivsuse näitajate, küsitluse koondtunnuste, aktiivsuse näitajad koos keskmiste jooksvate hinnetega, aktiivsuse näitajad koos küsitluse koondtunnustega ning küsitluse koondtunnused koos jooksvate keskmiste hinnete andmehulkade korral. Tabelis 7 on esitatud mudelite ennustustulemused enne ja pärast parimate tunnuste valimist ja kasutatud tunnuste arv k . Detailselt esitatud ennustustulemused leiab Lisast 7.

Tabel 7. Programmeerimise algkursuse ennustustäpsused enne ja pärast parimate tunnuste valmist.

Ennustamisel kasutatud andmed	k	Algne ennustustäpsus	Ennustustäpsus pärast parimate tunnuste valmist
koondtunnused	1	78,9%	78,9%

Ennustamisel kasutatud andmed	<i>k</i>	Algne ennustustäpsus	Ennustustäpsus pärast parimate tunnuste valmist
aktiivsusnäitajad (1 kuu)	9	52,1%	53,1%
aktiivsusnäitajad keskmiste jooksvate tulemustega (1 kuu)	2	58,3%	62,5%
koondtunnused aktiivsusega (1 kuu)	5	78,9%	78,9%
koondtunnused keskmiste jooksvate tulemustega (1 kuu)	8	76,9%	76,9%
aktiivsusnäitajad (2 kuud)	6	58,3%	61,5%
aktiivsusnäitajad keskmiste jooksvate tulemustega (2 kuud)	8	70,8%	68,8%
koondtunnused aktiivsusega (2 kuud)	10	78,9%	78,9%
koondtunnused keskmiste jooksvate tulemustega (2 kuud)	8	73,1%	73,1%

6.1.4 Moodle'i logiandmete kasutamine robotite programmeerimise õppeadukuse klasside ennustamisel

Robotite programmeerimise sügissemestri esimese kuu aktiivsusnäitajaid kasutanud masinõppemudel klassifitseeris kõik tudengid klassi hinded 1 kuni 3. Logiridade arvu kasutanud võrdlusbaasi mudel seda ei teinud ning sai ennustustäpsuseks 54,6% (vt Lisa 1). Robotite programmeerimise kursuse logides polnud semestri esimese kuu kohta jooksvate hinnetega seotud informatsiooni ning nagu eelnevalt sai välja toodud pole küsitluse küsimuste ning küsitluse koondtunnuste andmed usaldusväärsed.

Üllataval kombel oli hinde 0 täpsusskoor kõigi robotite programmeerimise semestri esimese kahe kuu andmehulkade korral 0% ning mitmed mudelid paigutasid kõik tudengid hindeklasi 1 kuni 3, mistõttu pole ennustustulemused kasutatavad järelduste tegemiseks (vt Tabel 8). Aktiivsusnäitajate ja aktiivsusnäitajate ning keskmiste jooksvate tulemuste mudelid paigutasid kõik tudengid ühte klassi. Tõenäoliselt oli mudelitel raskusi

läbikukkunud tudengite tuvastamisega, kuna läbikukkujaid oli robotite programmeerimise aines märksa vähem kui programmeerimise algkursusel.

Robotite programmeerimise kursuse keskmiste jooksvate tulemuste mudeli ennustustäpsus oli parem jooksvate tulemuste omast, mõlemad mudelid olid paremad võrdlusbaasi tulemustest. Keskmiste jooksvate hinnete mudeli hindeklassi 1 kuni 3 korral oli täpsusskoor 93,8% ja saagisskoor 75% ning hinnete 4 kuni 5 täpsusskooriks oli 25% ja saagisskooriks suisa 100%. Ülesannete esitamiste kordade mudel saavutas aga võrdlusbaasi ennustustäpsusest kehvema tulemuse. Aktiivsuse näitajate ja jooksvate tulemuste masinõppemudeli ennustustäpsus oli üllataval kombel väiksem võrdlusbaasi täpsusest. Chi ruutude kasutamine parimate tunnuste valimiseks ei parandanud masinõppemudelite tulemusi.

Tabel 8. Robotite programmeerimise semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.

Ennustustulemus	Ennustustäpsus	Mudel paigutas kõik tudengid ühte klassi
2 kuu andmete võrdlusbaas	59,1%.	Ei
aktiivsuse näitajad	72,7%	Jah
jooksvad tulemused	68,2%,	Ei
keskmised jooksvad tulemused	72,7%	Ei
ülesannete esitamiste korrad	54,6%	Ei
aktiivsuse näitajad jooksvate tulemustega	54,6%	EI
aktiivsuse näitajad keskmiste jooksvate tulemustega	72,7%	Jah

6.1.5 Moodle'i logiandmete kasutamine sissejuhatus erialasse õppeedukuse klasside ennustamisel

Sissejuhatus erialasse semestri esimese kuu logiridade arvul põhinev võrdlusbaas oli ainsaks õnnestunud ennustuseks, kuna aktiivsuse näitajaid kasutav võrdlusbaasist detailsem mudel klassifitseeris kõik tudengid hindeklassi „arvestatud“ (vt Tabel 9). Selline käitumine oli tõenäoliselt tingitud väikesest treeningandmete hulgast ja sellest, et aine kukkus läbi võrreldes kursuse läbinud üliõpilastega palju vähem tudengeid. Mudel

saavutas võrdlusbaasist parema tulemise klassifitseerides kõik tudengid ühte kahest arvestusliku aine hindeklassist, seega võis parema ennustustäpsuse saavutamisele orienteeritud mudeli konfiguratsioon käesoleval juhul pigem kahju teha.

Tabel 9. Sissejuhatus erialasse semestri esimese kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.

Ennustustulemus	Ennustustäpsus	Hinde „mittearvestatud“ täpsusskoor	Hinde „arvestatud“ täpsusskoor
1 kuu andmete võrdlusbaas	71,4%	33,3%	77,8%
aktiivsusanäitajad	85,7%	0%	100%

Semestri kahe esimese kuu logiridade arvu kasutatav võrdlusbaasi mudel oli üllataval kombel madalama ennustustäpsusega kui ainult esimese kuu sissejuhatus erialasse andmeid kasutanud võrdlusbaasi masinõppemudel.

Sarnaselt semestri esimese kuu aktiivsusanäitajaid kasutavale mudelile klassifitseerisid mitmed semestri kahe esimese kuu Moodle'i andmete peal treenitud mudelid kõik tudengid hindeklassi „arvestatud“, seda tõenäoliselt ka eelnevalt kirjeldatud põhjustel. Ainult ühte klassi tudengite klassifitseerimine toimus nii aktiivsusanäitajate, jooksvate tulemuste, keskmiste jooksvate tulemuste kui ka aktiivsusanäitajaid koos keskmiste jooksvate tulemustega kasutavate masinõppemudelite korral. Ainsaks erandiks oli mudel, mis kasutas aktiivsusanäitajaid koos jooksvate hinnetega, mis saavutas nii semestri esimese kui ka esimese kahe kuu võrdlusbaasidest parema ennustustäpsuse (vt Tabel 10). Mudel oli just eriti hea läbikukkunud tudengite klassifitseerimisel, sest täpsusskoori paranemine võrreldes semestri esimese kuu võrdlusbaasiga läbikukkunud tudengite paremal klassifitseerimisel, kuna mõlema mudeli hinde „arvestatud“ täpsusskoor oli sama. Chi ruutude abil parimate tunnuste valmine ei muutnud ennustustulemusi kahjuks paremaks.

Tabel 10. Sissejuhatus erialasse semestri esimese kahe kuu Moodle'i logisid kasutavate andmehulkade ennustustäpsused.

Ennustustulemus	Ennustustäpsus	Hinde „mittearvestatud“ täpsusskoor	Hinde „arvestatud“ täpsusskoor
1 kuu andmete võrdlusbaas	71,4%	33,3%	77,8%

Ennustustulemus	Ennustustäpsus	Hinde „mittearvestatud“ täpsusskoor	Hinde „arvestatud“ täpsusskoor
2 kuu andmete võrdlusbaas	66,7%	33,3%	72,2%
aktiivsusanäitajad jooksvate tulemustega	76,2%	66,7%	77,8%

6.1.6 Programmeerimise algkursuse andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused

Järgnevalt annab autor ülevaate ühe kursuse semestri esimese või esimese kahe kuu andmete peal treenitud ja teise kursuse semestri esimese või esimese kahe kuu andmete peal testitud masinõppemudelite ennustustulemustest (vt Lisa 4 ja Lisa 6).

Semestri esimese kuu programmeerimise algkursuse andmete peal treenitud ning robotite programmeerimise kursuse peal testitud võrdlusbaasi tulemus oli võrreldes eelmistes alapeatükkides välja toodud ennustustäpsustega väga madal, kõigest 27,9%. Ainult aktiivsusanmeid kasutatav mudel sai võrdlusbaasist veelgi madalama tulemuse, kuna klassifitseeris kõik tudengid läbikukkunute hulka. Selle peamiseks põhjuseks on tõenäoliselt robotite programmeerimise kursuse madalam tegevuste arv Moodle'is semestri esimesel kuul võrreldes programmeerimise algkursusega, näiteks polnud robotite programmeerimise kursusel esimesel kuul vaja esitada ülesandeid Moodle'i keskkonnas. Küsitluse koondtunnused koos aktiivsusanäitajatega saavutasid aga võrdlusbaasist ning ainult aktiivsusanäitajaid kasutanud mudelist palju parema tulemuse (vt Tabel 11). Tulemuseks oli 61,9%, mitte rohkem peamiselt seetõttu, et saagisskoorid olid hinnete 0 ning 4 kuni 5 korral madalad.

Tabel 11. Programmeerimise algkursuse semestri esimese kuu andmete peal treenitud ning robotite programmeerimise semestri esimese kuu andmete peal testitud mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor	Hinnete 1 kuni 3 täpsusskoor	Hinnete 4 kuni 5 täpsusskoor
1 kuu andmete võrdlusbaas	27,9%	100%	2,8%	0%
aktiivsusanäitajad	26,1%	100%	0%	0%
koondtunnused aktiivsusega	61,9%	100%	56,6%	88,9%

Semestri kahe esimese kuu programmeerimise algkursuse andmete peal treenitud ning robotite programmeerimise kursuse peal testitud võrdlusbaasi tulemus (31,5%) oli parem ainult ühe kuu andmeid kasutanud mudeli omast. Kõik täpsemaid Moodle'i logiandmeid kasutavad mudelid saavutasid seekord võrdlusbaasist parema tulemuse (vt Tabel 12). Aktiivsusnäitajate masinõppemudel ning aktiivsusnäitajad koos keskmiste jooksvate hinnetega said tulemuseks täpselt sama ennustustäpsuse, mis on üllatav, arvestades seda, et keskmiste jooksvate tulemuste ennustustäpsus oli aktiivsusnäitajate mudeli omast suurem. Parimad tulemused saavutati kasutades küsitluse koondtunnuseid koos aktiivsusnäitajatega ning küsitluse koondtunnuseid koos keskmiste jooksvate hinnetega, seega pakkusid küsitluse koondtunnused mudelile vajalikku lisainformatsiooni, mis aitas märkimisväärselt parandada hinnete 4 kuni 5 klassifitseerimist, aga ka teiste hindegruppide täpsusskoore. Koondtunnused mõjutasid ennustustulemusi rohkem kui aktiivsusnäitajad ja keskmised jooksvad tulemused, kuna mõlema koondtunnusega kombineeritud andmehulga korral saavutati sama ennustustäpsus.

Tabel 12. Programmeerimise algkursuse semestri esimese kahe kuu andmete peal treenitud ning robotite programmeerimise semestri esimese kahe kuu andmete peal testitud mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor	Hinnete 1 kuni 3 täpsusskoor	Hinnete 4 kuni 5 täpsusskoor
2 kuu andmete võrdlusbaas	31,5%	79,3%	16,7%	0%
aktiivsusnäitajad	42,3%	93,1%	27,8%	0%
keskmised jooksvad tulemused	51,4%	96,6%	36,1%	30%
aktiivsusnäitajad keskmiste jooksvate tulemustega	42,3%	93,1%	27,8%	0%
koondtunnused aktiivsusega	61,9%	100%	56,6%	88,9%
koondtunnused keskmiste jooksvate tulemustega	61,9%	100%	56,6%	88,9%

Programmeerimise algkursuse esimese kuu Moodle'i aktiivsuslogi andmete peal treenitud ning sissejuhatus erialasse andmete peal testitud võrdlusbaasi ennustustäpsuseks oli 55,7%, millega täpselt sama ennustustäpsuse saavutas ka aktiivsusnäitajaid kasutav mudel. Küsitluse koondtunnuseid koos aktiivsusnäitajatega kasutanud mudel saavutas võrdlusbaasis kehva tulemuse. Sissejuhatus erialasse on arvestuslik kursus, seega ei

ole õppainel sama palju hindeklaasse kui programmeerimise algkursusel, mis põhjendab ka madalaid ennustustäpsuseid. Mudel määrab tudengitele klasse, mida andmehulgas ei esine. Mudeli tulemusi saaks parandada, muutes programmeerimise algkursuse andmed samuti arvestuslikuks, kus „mittearvestatud“ klassi saavad tudengid, kelle hindeks on 0 ning „arvestatud“ klassi üliõpilased, kelle tulemused oli positiivsed.

Hindamisklassidega seotud probleemid tekkisid mõistagi ka kahe kuu Moodle'i programmeerimise algkursuse andmete peal treenitud ning sissejuhatus erialasse peal testitud mudelite korral. Ometi on oluline märkida, et aktiivsuse näitajaid kasutanud mudel saavutas üpris hea tulemuse, mis oli parem ka kahe kuu andmete võrdlusbaasist (vt Tabel 13). Täpselt sama ennustustäpsuse sai ka aktiivsuse näitajaid ning keskmisi jooksvaid tulemusi kasutanud mudel. Ainult keskmiseid jooksvaid hindeid kasutanud mudel saavutas võrdlusbaasist palju madalama tulemuse. Võrdlusbaasi tulemusest saavutasid halvema ennustustäpsuse ka koondtunnuseid ning logiandmeid kombineerinud andmehulkasid kasutanud masinõppemudelid.

Tabel 13. Programmeerimise algkursuse semestri esimese kahe kuu andmete peal treenitud ning sissejuhatus erialasse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde „mittearvestatud“ täpsusskoor	Hinde „arvestatud“ täpsusskoor
2 kuu andmete võrdlusbaas	61,13%	68,2%	59,5%
aktiivsuse näitajad	75,5%	90,9%	71,14%
aktiivsuse näitajad keskmiste jooksvate tulemustega	75,5%	90,9%	71,14%

6.1.7 Robotite programmeerimise andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused

Robotite programmeerimise semestri esimese kuu Moodle'i andmete peal treenitud ning sissejuhatus erialasse andmete peal testitud mudelite puhul on samuti probleemiks see, et sissejuhatus erialasse on arvestuslik aine ning tudengeid paigutatakse klassidesse, mida andmehulgas pole. Esimese kuu võrdlusbaasi mudel paigutas kõik tudengid „arvestatud“ hindeklassi, seda tegi ka aktiivsuse näitajaid koos küsitluse koondtunnustega kasutanud mudel. Tõenäoliselt on see tingitud kursuse läbikukkunud tudengite väiksest arvust andmehulgas. Ainult aktiivsuse näitajaid kasutanud mudel saavutas ennustustäpsuse

58,5%, hinde „mittearvestatud“ täpsusskooriks oli 68,2% ning hinde „arvestatud“ täpsusskooriks oli 56%.

Semestri kahe esimese kuu andmeid kasutanud võrdlusbaasi ennustusmudel sai ennustustäpsuseks 76,4% ning ei paigutanud kõiki tudengeid samasse klassi. Aktiivsuse näitajaid kasutanud mudel saavutas võrdlusbaasist paremad tulemused (vt Tabel 14.).

Tabel 14. Robotite programmeerimise semestri esimese kahe kuu andmete peal treenitud ning sissejuhatus erialasse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde „mittearvestatud“ täpsusskoor	Hinde „arvestatud“ täpsusskoor
2 kuu andmete võrdlusbaas	76,4%	18,2%	91,7%
aktiivsuse näitajad	84,9%	59,1%	91,7%

Keskmissi jooksvaid tulemusi treenimiseks ja testimiseks kasutanud mudeli tulemus oli võrdlusbaasi ennustustäpsusest palju madalam ning küsitluse koondtunnuseid koos aktiivsuse näitajatega kasutanud mudel ning küsitluse koondtunnuseid koos keskmiste jooksvate hinnetega kasutanud mudel paigutasid kõik tudengid „arvestatud“ klassi, tõenäoliselt seetõttu, et treeningandmehulgas oli küsitluse andmete kasutamise korral kaks aine läbikukkunud tudengit ning testandmehulgas kõigest üks.

Robotite programmeerimise esimese kuu andmete peal treenitud ning programmeerimise algkursuse peal testitud mudeli võrdlusbaas sai tulemuseks 47,9%, kuid mudel paigutas peaaegu kõik tudengid hindeklassi 1 kuni 3, hinde 0 täpsusskooriks oli 1,8%. Sama probleem ilmnis ka aktiivsuse näitajaid kasutanud mudeli korral ning koondtunnuseid koos aktiivsuse näitajatega kasutanud mudeli puhul, kus kõik tudengid paigutati hindeklassi 1 kuni 3. Tõenäoliselt on need ennustustulemused tingitud programmeerimise algkursuse ning robotite programmeerimise õppeainete ülesehituste erinevusest. Programmeerimise algkursuse semestri esimese kuu Moodle'i logides on märksa suurem aktiivsus, kuna tuli esitada näiteks juba mitmeid kodutöid, erinevalt robotite programmeerimise õppeainest, kus hinnetega seotud andmed jõudsid logidesse alles kursuse teisel kuul.

Kahe kuu andmeid kasutavate mudelite puhul tulemused paranevad – kõik mudelid saavutasid kahe kuu võrdlusbaasist parema ennustustäpsuse (vt Tabel 15). Kahjuks paigutas küsitluse koondtunnuseid ja keskmisi jooksvaid tulemusi kasutanud mudel kõik tudengid hindeklassi 1 kuni 3. Parima ennustustäpsusega mudeliks osutus koondtunnuseid koos aktiivsusanäitajatega kasutanud mudel, kuid mudel ei suutnud sarnaselt paljudele teistele koondtunnuseid kasutavatele mudelitele klassifitseerida kursuse läbikukkunud tudengeid.

Tabel 15. Robotite programmeerimise semestri esimese kahe kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor	Hinnete 1 kuni 3 täpsusskoor	Hinnete 4 kuni 5 täpsusskoor
2 kuu andmete võrdlusbaas	47,5%	17,3%	81,5%	16,5%
aktiivsusanäitajad	58,8%	33,3%	99,6%	0%
keskmised jooksvad tulemused	61%	31,5%	97,8%	21,2%
aktiivsusanäitajad keskmiste jooksvate tulemustega	58,8%	33,3%	99,6%	0%
koondtunnused aktiivsusega	71,8%	0%	97,1%	30,3%

6.1.8 Sissejuhatus erialasse andmete peal treenitud ning teise kursuse andmete pealt testitud mudelite ennustustulemused

Järgmised kaks mudelite gruppi on treenitud sissejuhatus erialasse aine andmete peal ning testitud programmeerimise algkursuse ja robotite programmeerimise andmete peal, seega ei oska ükski mudelitest korrektselt klassifitseerida hinderühma 4 kuni 5.

Semestri esimese kuu andmeid kasutanud ja robotite programmeerimise kursuse peal testitud mudelid paigutasid kõik tudengid ühte klassi. Võrdlusbaasi masinõppemudel ning aktiivsusanäitajaid kasutanud mudel paigutasid kõik üliõpilased läbikukkujate hulka. Selle peamiseks põhjuseks on tõenäoliselt kahe vaadeldava kursuse väga suur aktiivsuse erinevus Moodle'i logides semestri esimese kuu vältel. Küsitluse koondtunnuseid ning aktiivsusanäitajaid kasutanud mudel paigutas aga kõik tudengid hindeklassi 1 kuni 3.

Mudel oli tõenäoliselt taaskord tugevalt mõjutatud hinde 0 saanud tudengite vähesusest küsitluse koondtunnuste andmehulgas.

Semestri esimese kuu andmeid kasutanud ning programmeerimise algkursuse andmete peal treenitud mudelid olid robotite programmeerimise andmete peal treenitud mudelitest paremate ennustustulemustega. (vt Tabel 16) Erandiks oli küsitluse koondtunnuseid ning aktiivsuse näitajaid kasutanud mudel, mis klassifitseeris kõik üliõpilased klassi 1 kuni 3.

Tabel 16. Sissejuhatus erialasse semestri esimese kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor	Hinnete 1 kuni 3 täpsusskoor
1 kuu andmete võrdlusbaas	47,3%	36,9%	72,7%
aktiivsuse näitajad	54%	47,6%	78,9%

Semestri esimese kahe kuu andmete kasutuselevõtmine parandas pisut ennustustäpsusi (vt Tabel 17), kuid keskmisi jooksvaid tulemusi, koondtunnused koos aktiivsuse näitajatega ning koondtunnuseid koos keskmiste jooksvate tulemustega kasutanud mudelid paigutasid kõik tudengid hindeklassi 1 kuni 3. Parimateks mudeliteks olid aktiivsuse näitajaid ning aktiivsuse näitajaid koos keskmiste jooksvate tulemustega kasutanud mudelid, mis saavutasid sama ennustustäpsuse tõenäoliselt tingitud aktiivsuse näitajate kasutamisest.

Tabel 17. Sissejuhatus erialasse semestri esimese kahe kuu andmete peal treenitud ning programmeerimise algkursuse semestri esimese kahe kuu andmete peal testitud võrdlusbaasist paremate mudelite ennustustäpsused.

Ennustamisel kasutatud andmed	Ennustustäpsus	Hinde 0 täpsusskoor	Hinnete 1 kuni 3 täpsusskoor
2 kuu andmete võrdlusbaas	54%	47,6%	78,9%
aktiivsuse näitajad	59,2%	59,5%	81,1%
aktiivsuse näitajad keskmiste jooksvate tulemustega	59,2%	59,5%	81,1%

6.2 Küsitluse koondtunnuste seotus Moodle'i logiandmetega

Koondtunnuste ennustamisel programmeerimise algkursuse Moodle'i logide andmete põhjal selgused, et logide andmed ennustavad kõige paremini sisemist regulatsiooni. Semestri esimese kuu jooksvate hinnete andmehulga pealt õnnestus ennustada sisemise regulatsiooni väärtust 50% täpsusega. Semestri esimese kuu keskmisi jooksvaid tulemusi kasutades oli ennustustäpsuseks 51,9%.

Kahe esimese kuu andmeid kasutades sisemise regulatsiooni ennustustulemused muutusid: keskmiste jooksvate tulemuste andmehulga ennustustäpsus vähenes ning nii jooksvate tulemuste kui ka ülesannete esitamise kordade andmehulkadel treenitud mudelid ennustasid koondtunnust 50% täpsusega. Enesetõhusust ennustasid kahe kuu jooksvad hindid täpsusega 51,9%. Läbikukkumise vältimist ennustasid 50% täpsusega semestri esimese kahe kuu jooksvad tulemused. Teiste koondtunnuste korral olid ennustustäpsused madalamad. Ennustustäpsused on leitavad Lisas 8 olevates tabelites.

6.2.1 Spearmani korrelatsioonikoefitsiendid

Spearmani korrelatsioonikoefitsientide (vt Lisad 9 ja 10) leidmisel oli esimeseks tähelepanekuks see, et korrelatsioonikoefitsiendid koondtunnuse ja logiridade arvu (*activity*), koondtunnuste ja *diff_from_average* aktiivsustunnuse ning koondtunnuste ja *percentage_from_max* aktiivsustunnuse korral olid alati samad. Seda seletab fakt, et nii erinevus keskmisest aktiivsuses (*diff_from_average*) kui ka protsent maksimum aktiivsusest (*percentage_from_max*) leiti *activity* tunnuse ehk logiridade arvu pealt.

Programmeerimise algkursuse esimese kuu *activity* näitajaga oli suurim korrelatsioon koondtunnusel alustamisega viivitamine, millele järgnesid sisemine regulatsioon ja huvi (vt Tabel 18). Esimese kuu andmehulga *number_df_active_days* (aktiivsete päevade arv) suurim korrelatsioon oli koondtunnusega alustamisega viivitamine, millele järgnesid taaskord sisemine regulatsioon ning huvi (vt Tabel 19). Aktiivsuse näitaja *number_of_active_weeks* (aktiivsete nädalate arv) korral olid peaaegu kõik korrelatsioonikoefitsiendid nullilähedased, parimaks osutus koefitsiendiga 0,1 sotsiaalne õppimine. Keskmistel jooksvatel tulemustel oli mitmete koondtunnustega üpris suured korrelatsioonid, neist parimaks osutus enesetõhusus, millele järgnesid alustamisega viivitamine ning keeruliste ülesannete vältimine (vt Tabel 20).

Tabel 18. Programmeerimise algkursuse semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>activity</i> ehk logiridade arv	-0,221
sisemine regulatsioon	<i>activity</i> ehk logiridade arv	0,215
huvi	<i>activity</i> ehk logiridade arv	0,194

Tabel 19. Programmeerimise algkursuse semestri esimese kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>number_of_active_days</i>	-0,222
sisemine regulatsioon	<i>number_of_active_days</i>	0,215
huvi	<i>number_of_active_days</i>	0,194

Tabel 20. Programmeerimise algkursuse semestri esimese kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
enesetõhusus	keskmised jooksvad tulemused	0,416
alustamisega viivitamine	keskmised jooksvad tulemused	-0,309
keeruliste ülesannete vältimine	keskmised jooksvad tulemused	-0,297

Semestri kahe esimese kuu programmeerimise algkursuse andmete korral oli *activity* tunnuse suurim korrelatsioon koondtunnusega alustamisega viivitamine, millele järgnesid sisemine regulatsioon ning omaksõetud strateegiline regulatsioon (vt Tabel 21). Aktiivsuse näitaja *number_of_active_days* Spearmani korrelatsioonid olid suurimad koondtunnustega alustamisega viivitamine, sisemine regulatsioon ning huvi (vt Tabel 22). Tunnuse *number_of_active_weeks* korrelatsioonikoefitsiendid oli taaskord üldiselt nullilähedased, leidsid ka mõned erandid, millest suurimaks oli taaskord sotsiaalne õppimine (0,232). Kahe kuu keskmiste jooksvate tulemuste korrelatsioon oli suurim enesetõhususe, keeruliste ülesannete vältimise ning alustamisega viivitamisega (vt Tabel 23).

Tabel 21. Programmeerimise algkursuse semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>activity</i> ehk logiridade arv	-0,227
sisemine regulatsioon	<i>activity</i> ehk logiridade arv	0,231
omaksvõetud strateegiline regulatsioon	<i>activity</i> ehk logiridade arv	0,219

Tabel 22. Programmeerimise algkursuse semestri esimese kahe kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>number_of_active_days</i>	-0,239
sisemine regulatsioon	<i>number_of_active_days</i>	0,236
huvi	<i>number_of_active_days</i>	0,193

Tabel 23. Programmeerimise algkursuse semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
enesetõhusus	keskmised jooksvad tulemused	0,401
keeruliste ülesannete vältimine	keskmised jooksvad tulemused	-0,308
alustamisega viivitamine	keskmised jooksvad tulemused	-0,278

Robotite programmeerimise kursuse semestri esimese kuu logiridade arvude Spearmani korrelatsioonikoefitsient oli suurim alustamisega viivitamise korral, millele järgnesid uskumused matemaatikaalase võimekuse kohta ja enesetõhusus (vt Tabel 24). Suurem osa *number_of_active_days* tunnuse korrelatsioonikoefitsientidest olid nullilähedased, erandisteks olid alustamisega viivitamine, uskumused matemaatikaalase võimekuse kohta ning enesetõhusus (vt Tabel 25). Aktiivsuse näitaja *number_of_active_weeks* korreleerus kõige rohkem uskumustega matemaatikaalase võimekuse kohta, pingutuse olulisusega programmeerimiskursusel ning läbikukkumise vältimisele suunatud soorituseesmärkidega (vt Tabel 26).

Tabel 24. Robotite programmeerimise semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>activity</i>	-0,177
uskumused matemaatikaalase võimekuse kohta	<i>activity</i>	0,155
enesetõhusus	<i>activity</i>	0,141

Tabel 25. Robotite programmeerimise semestri esimese kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>number_of_active_days</i>	-0,179
uskumused matemaatikaalase võimekuse kohta	<i>number_of_active_days</i>	0,17
enesetõhusus	<i>number_of_active_days</i>	0,145

Tabel 26. Robotite programmeerimise semestri esimese kuu *number_of_active_weeks* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
uskumused matemaatikaalase võimekuse kohta	<i>number_of_active_weeks</i>	0,224
pingutuse olulisusega programmeerimiskursusel	<i>number_of_active_weeks</i>	0,2
läbikukkumise vältimisele suunatud soorituseesmärgid	<i>number_of_active_weeks</i>	0,135

Robotite programmeerimise semestri kahe esimese kuu *activity* tunnuse korrelatsioon oli suurim soorituspõhise motivatsioonile lähenemise korral, millele järgnesid enesetõhusus ning uskumused ärevuse kontrollitavuse ja suhtlemispädevuse muudetavuse kohta (vt Tabel 27). Aktiivsusanäitaja *number_of_active_days* korrelatsioonid olid enamjaolt nullilähedased, kuid kõige enam korreleerusid andmehulgaga koondtunnustest soorituspõhine motivatsioonile lähenemine, uskumused ärevuse kontrollitavuse ja suhtlemispädevuse muudetavuse kohta ning enesetõhusus (vt Tabel 28). Tunnus

number_of_active_weeks korrelatsioonikoefitsiendid olid samuti enamjaolt nullilähedased, tunnus korreleerus kõige rohkem uskumustega matemaatikaalase võimekuse kohta (0,193), sotsiaalse õppimisega (0,186) ning õppejõu pakutava emotsionaalse ja õppimisalase toe tajumisega (0,168). Keskmiste jooksvate hinnete korrelatsioonikoefitsient on suurim alustamisega viivitamise, keeruliste ülesannete vältimise, pingutuse olulisusega programmeerimiskursusel ning enesetõhususe korral (vt Tabel 29).

Tabel 27. Robotite programmeerimise semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
soorituspõhine motivatsioonile lähenemine	<i>activity</i>	-0,235
enesetõhusus	<i>activity</i>	0,175
uskumused ärevuse kontrollitavuse ja suhtlemispädevuse muudetavuse kohta	<i>activity</i>	0,173

Tabel 28. Robotite programmeerimise semestri esimese kahe kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
soorituspõhine motivatsioonile lähenemine	<i>number_of_active_days</i>	-0,249
uskumused ärevuse kontrollitavuse ja suhtlemispädevuse muudetavuse kohta	<i>number_of_active_days</i>	0,192
enesetõhusus	<i>number_of_active_days</i>	0,162

Tabel 29. Robotite programmeerimise semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viitamine	keskmised jooksvad tulemused	-0,32
keeruliste ülesannete vältimine	keskmised jooksvad tulemused	-0,315
pingutuse olulisus programmeerimiskursusel	keskmised jooksvad tulemused	-0,307
enesetõhususe	keskmised jooksvad tulemused	0,301

Sissejuhatus erialasse sügissemestri esimese kuu logiridade arvu Spearmani korrelatsioonikoefitsient oli suurim alustamisega viivitamise, keeruliste ülesannete vältimise ning uskumuste matemaatikaalase võimekuse kohta korral (vt Tabel 30). Aktiivsuse näitaja *number_of_active_days* korrelatsioon oli samuti suurim alustamisega viivitamise, keerulise ülesannete vältimise ning uskumustega matemaatikaalase võimekuse kohta korral (vt Tabel 31). Semestri esimese kuu aktiivsuse näitaja *number_of_active_weeks* korrelatsioon oli suurim õppejõu pakutava emotsionaalse ja õppimisalase toe tajumise, huvikaotuse ning huvi korral (vt Tabel 32).

Tabel 30. Sissejuhatus erialasse semestri esimese kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>activity</i>	-0,29
keeruliste ülesannete vältimine	<i>activity</i>	-0,284
uskumused matemaatikaalase võimekuse kohta	<i>activity</i>	0,241

Tabel 31. Sissejuhatus erialasse semestri esimese kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
alustamisega viivitamine	<i>number_of_active_days</i>	-0,307
keeruliste ülesannete vältimine	<i>number_of_active_days</i>	-0,285
uskumused matemaatikaalase võimekuse kohta	<i>number_of_active_days</i>	0,244

Tabel 32. Sissejuhatus erialasse semestri esimese kuu *number_of_active_weeks* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
õppejõu pakutava emotsionaalse ja õppimisalase toe tajumine	<i>number_of_active_weeks</i>	-0,303
huvikaotus	<i>number_of_active_weeks</i>	0,282
huvi	<i>number_of_active_weeks</i>	-0,259

Sissejuhatus erialasse sügissemestri esimese kahe kuu Moodle'i logiridade arvu korrelatsioon oli suurim koondtunnustega keeruliste ülesannete vältimise, alustamisega viivitamise ja tõhustate õpistrateegiate korral (vt Tabel 33). Aktiivsuse näitaja *number_of_active_days* korreleerus samuti kõige enam keeruliste ülesannete vältimise, alustamisega viivitamise ja tõhustate õpistrateegiate korral (vt Tabel 34). Tunnuse *number_of_active_weeks* korrelatsioonikoefitsientidest olid suurimad õppejõu pakutava emotsionaalse ja õppimisalase toe tajumine (-0,201), huvikaotus (0,174) ning omaksvõetud strateegiline regulatsioon (-0,153).

Tabel 33. Sissejuhatus erialasse semestri esimese kahe kuu Moodle'i logiridade arvude suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
keeruliste ülesannete vältimine	<i>activity</i>	-0,34
alustamisega viivitamine	<i>activity</i>	-0,295
tõhusad õpistrateegiad	<i>activity</i>	0,228

Tabel 34. Sissejuhatus erialasse semestri esimese kahe kuu *number_of_active_days* andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
keeruliste ülesannete vältimine	<i>number_of_active_days</i>	-0,329
alustamisega viivitamine	<i>number_of_active_days</i>	-0,314
tõhusad õpistrateegiad	<i>number_of_active_days</i>	0,211

Semestri kahe esimese kuu keskmiste jooksvate tulemuste andmehulk korreleerub kõige rohkem mittetõhusate õpistrateegiatega, lihtsate ülesannete eelistamisega ning tõhustate õpistrateegiatega (vt Tabel 35).

Tabel 35. Sissejuhatus erialasse semestri esimese kahe kuu keskmiste jooksvate tulemuste andmehulga suurimad Spearmani korrelatsioonikoefitsiendid.

Koondtunnus	Andmehulk	Spearmani korrelatsioonikoefitsient
mittetõhusad õpistrateegiad	keskmised jooksvad tulemused	-0,226
lihtsate ülesannete eelistamine	keskmised jooksvad tulemused	0,201
tõhusad õpistrateegiad	keskmised jooksvad tulemused	-0,189

6.3 Soovitused

Järgnevalt annab autor soovitusi küsitluse paremaks läbiviimiseks, sooritustulemuste ennustamiseks ning tudengeid rohkem kaasvaks Moodle'i keskkonna kasutamiseks.

6.3.1 Küsitlus

Magistritöö tulemuste peatükist saab järeldada, et esimese kursuse Informaatika õppekava bakalaureusetudengite hulgas läbiviidud küsitlus osutus heaks indikaatoriks tudengite võetud kursuste lõpptulemuste ennustamisel. Lisaks sellele osutusid küsitluse koondtunnused ka parimaks andmehulgaks tudengite klastriteks jaotamisel. Seega saab soovitada TalTechi õppejõududele oma tudengite hulgas semestri alguses küsitluse läbiviimist, et leida üles riskirühma kuuluvad tudengid, kellele saab vajaduse korral semestri jooksul lisatuge kursuse läbimiseks pakkuda. Veelgi parem oleks kui viidaks tudengite hulgas läbi vaid üks küsitlus, mille tulemusi saaksid kõik õppejõud kasutada, et tudengid ei peaks täitma liiga palju sarnaseid küsitlusi. Kuna tudengite uskumused ning emotsionaalne olukord muutuvad ajas, oleks soovituslik küsitlust korrata. Näiteks võiks küsitlust läbi viia nii semestri alguses kui ka lõpus, et näha, kuidas läbitud õppeained tudengeid mõjutanud on. Küsitlusi võiks viia läbi ka erinevatel semestritel, et pakkuda tudengitele tuge kogu nende ülikoolitee jooksul.

Tudengite väljakukkumise ennetamise aspektist on kõige olulisem viia küsitlus läbi just semestri alguses, kuna on tõenäoline, et siis vastavad sellele ka tudengid, kes kukuvad välja juba poole kursuse pealt. Käesolevas magistritöös läbiviidud küsitlus viidi läbi alles semestri lõpus ning seetõttu ei olnud võimalik kasutada nende tudengite andmeid, kes jätsid kursuse pooleli või siis katkestasid oma õpingud ülikoolis. Kursuste läbikukkumise ennetamiseks ning tudengite ülikoolist väljalangemise leevendamiseks on just nende tudengite aitamine kõige suuremaks prioriteediks, kuid kahjuks on selliste tudengite kohta alati kõige vähem informatsiooni. Küsitluse läbiviimine semestri alguses aitab tulevikus kujundada parema arusaama nende tudengite õpikäitumisest ning õppimisega seotud psühholoogilistest faktoritest.

Kindlasti saab küsitlust ka täiendada. Läbiviidud küsitluse juures märkasid tudengid, et osad küsimused olid korduvad, ning nad avaldasid küsitluse lõpus olnud vabateksti lahtris frustratsiooni selle aspekti osas. Seega ei tasuks küsitlusse panna korduvaid küsimusi,

kuna tudengid tunnevad, et see raiskab nende aega, ning seega võivad nende poolt antud vastused olla vähem keskendunult vastatud ning sellest lähtuvalt ka vähem tõesed.

Teiseks muudatuseks võiks olla tudengite käest täiendava isikliku informatsiooni küsimine, mis võib aidata parandada tudengite profiilideks jagamist. Näiteks on esmakursuslaste hulgas nii värskelt keskkooli lõpetanud tudengeid kui ka juba töökogemust või suisa kõrgharidust omavaid vanemaid õppureid, kes soovivad ennast täiendada. Õppeedukust mõjutab tugevalt ka see, kas tudeng töötab õpingutega paralleelselt. Tõenäoliselt erineb töötavate tudengite ajakasutus märkimisväärselt ainult õpingutega tegelevate üliõpilaste omast ning see mõjutab tugevalt õpikäitumist ja võib-olla ka õpioskusi. Töötavate tudengite ametil võib samuti olla roll õpitulemuste mõjutamisel, näiteks võivad erialase töökohaga tudengid omada kasulikke oskusi, mis annavad neile kaasüliõpilaste ees eelise. Profileerimisel võivad lisaks eelmainitud tunnustele kasuks tulla ka informatsioon tudengi koduse keele, soo või läbitud keskkooli ning eelneva haridustaseme kohta, kuna ka sotsiaalne ning kultuuriline keskkond omavad teatud mõju tudengite uskumuste ning oskuste väljakujunemisel [13], [18], [22].

6.3.2 Ennustamine

Esimeseks probleemiks õppeedukuse ennustamisel osutus andmete vähesus. Robotite programmeerimise ja sissejuhatus erialasse õppeainete korral polnud piisavalt suur hulk tudengeid vastanud küsitlusele ning seetõttu ei olnud tulemused kasutatavad järelduste tegemiseks. Lisaks sellele ilmnes, et sissejuhatus erialasse kursus oli üldiselt väiksema osavõtjate arvuga ning seetõttu kippusid otsustuspuu algoritmi kasutavad masinõppemudelid tudengeid klassifitseerima ainult nendesse klassidesse, mida esines kõige rohkem.

Üheks võimaluseks selle probleemi lahendamisel oleks täiendavate kursuse konteksti sobivate treeningandmete genereerimine. Ülesanne pole aga lihtsate killast, kuna iga kursuse jaoks eraldiseisva andmeid genereeriva programmi loomine pole mõistlik ning ühesuguste treeningandmete genereerimine kõigile kursustele ei võta arvesse kursuste eripärasid ning Moodle'is tehtud tegevuste nüansse. Seega võiks tulevikus selgitada välja, millised on eri tüüpi kursused ja nende omadused, et oleks võimalik üldiste meetodite abil genereerida andmeid erinevalt ülesehitatud kursuste tarbeks.

Küsitluse koondtunnused andsid parimaid ennustustulemusi, kuid kahjuks ei olnud koondtunnuseid kasutavad masinõppemudelid eriti edukad just kursuse läbikukkunud tudengite klassifitseerimisel. Autor usub, et seda probleemi aitaks lahendada küsitlusele vastanud tudengite hulga suurendamine või siis genereeritud täiendavate treeningandmete kasutamine. Kindlasti aitaks ennustustulemust parandada ka kõrgetasemelisemate ennustusmudelite algoritmide kasutamine, mis suudaks tabada rohkem andmetega seotud nüansse.

Keskmete jooksvate tulemuste kasutamisel läks mitmete masinõppemudelite korral õppeedukust ennustavat informatsiooni kaduma, võrreldes kõiki jooksvaid tulemusi arvesse võtvate andmehulkadega. Tulevikus võiks jooksvate tulemuste keskmise leidmisel kõigepealt viia ülesannete punktid protsentuaalsele skaalale ning seejärel leida keskmise.

Juhul kui mudeleid treeniti arvestusliku kursuse andmete peale ning mudelit testiti hindelise õppeaine andmetega, siis ei osanud mudel arusaadavatel põhjustel nelja- ja viielistele tudengitele klasse panna. Tõenäoliselt oleks edaspidi mõistlik arvestuslike ainete treeningandmeid mitte kasutada hindeliste kursuste korral. Teiseks võimaluseks oleks hindeliste kursuste tudengid jaotada kahte klassi – kursuse läbinud tudengid ning aine läbikukkunud üliõpilased, et andmehulkades oleks sama arv klasse arvestuslike kursustega. Probleem tekib ka vastupidises olukorras, kus hindelise kursuse andmete peal treenitud mudelit testitakse arvestusliku aine andmetega ning mudel pakub tudengitele klasse, mida arvestusliku ainete tulemustes pole. Autor usub, et hindelise aine andmeid saab kasutada arvestuslike ainete treenimisel, kuid selleks tuleks eelnevalt kõik positiivsed hinded (1–5) muuta üheks klassiks „arvestatud“.

6.3.3 Moodle'i keskkonna kasutamine

Üldiselt ei kasuta õppeasutused veebipõhistes õppekeskkondades ühtlast süsteemi kursusi luues ning tudengeid õpetades. Lisaks sellele ei kasutata tihti kogu potentsiaali, mida veebipõhised õpikeskkonnad õppejõududele ning tudengitele pakuvad, jättes tihti kasutamata võimaluse õpianalüütika kogumiseks. Õppejõud loovad veebipõhised kursused üldiselt iseseisvalt ning kasutavad tihti ainult elementaarseid funktsionaalsusi, laadides õppekeskkonda üles loengute märkmeid ning ainekava. Õppejõud saavad edaspidi senisest aktiivsemalt kasutada Moodle'isse sisseehitatud võimalusi, et muuta enda kursused mitmekesisemaks ning tudengeid kaasavamaks [32, lk 1].

Autor soovitaks Moodle'it edaspidi kasutada mitmekesisemalt kui seni. Isegi mõned üksikud lisategevused Moodle'i keskkonnas suurendavad logiridade arvu, mis omakorda on väärtuslikuks sisendiks õpianalüütika koostamisel. Õppejõud, kes seni on kasutanud Moodle'it ainult loenguslaidide või märkmete hoiustamiseks, võiksid edaspidi näiteks panna suuremat rõhku kodutööde esitamisele Moodle'i keskkonnas, sest mitmed uuringud on näidanud, et veebipõhistes õppekeskkondades pidevalt jooksvate ülesannete lahendamine parandab tudengite sooritustulemusi [9], [11], [12]. Tudengeid kaasavad õppetöösse suuremas mahus ka Moodle'i keskkonnas läbiviidavad testid ning kursuse foorumite kasutamine [12], [13].

6.3.4 Pistikprogrammi loomine Moodle'i keskkonnale

Peatükis "Ülevaade teooriast" kirjeldas autor praegusel hetkel Moodle'i keskkonnale loodud õpianalüütikaga seotud sisseehitatud vaateid ning pistikprogramme, tuues välja nende omadused ning peamised puudused. Ülevaatest selgus, et puudub universaalne hästitoimiv tööriist, mis võimaldaks õppejõududel hinnata tudengite õppeedukust erineva ülesehitusega kursustel ning näha seda, kas üliõpilane on läbikukkumise ohus. Kuna käesoleva magistr töö fookuseks oli välja selgitada, millised logidest leitavad õpikäitumist kirjeldavad tunnused ennustavad tudengite lõpptulemusi kõige paremini eri tüüpi ülesehitusega kursuste puhul, siis saaks käesoleva töö tulemusi kasutada uue universaalse Moodle'i logidel põhineva õppeedukust hindava ning tudengeid riskirühmadesse jagava pistikprogrammi loomiseks.

Loodav Moodle'i pistikprogramm võiks olla võimalikult universaalne ning sobilik mitte ainult TalTechi Moodle'i keskkondadele, vaid Moodle'i kogukonnale üldiselt. Üldisele Moodle'i kogukonnale orienteeritud programmi loomine tagab selle, et ka juhul, kui TalTech muudab enda Moodle'i keskkondade kasutust, saab pistikprogrammi ka edaspidi kasutada. Lisaks sellele annab see võimaluse teistel pistikprogrammi vastu huvi tundvatel õppeasutustel programmi katsetada ning kasutada. Programm võiks olla sobilik eri tüüpi ülesehitusega kursustele, näiteks nii hindelistele kui ka arvestuslikele kursustele. Programmi kasutatavus eri tüüpi kursuste korral tagab selle, et ka õppejõud, kelle Moodle'i kursused erinevad pisut kõige tüüpilisematest õppeainetest, saaksid siiski programmi kasutada ning oma tudengeid seeläbi rohkem jälgida, toetada ning aidata.

7 Kokkuvõte

Käesoleva magistritöö käigus leidis autor Moodle'i keskkonna logidest ning TalTechi Informaatika eriala esimese kursuse bakalaureusetudengite hulgas läbiviidud õppimisega seotud psühholoogiliste tegurite küsitluse andmetest tunnused, mis ennustasid kolme vaadeldud õppeaine, programmeerimise algkursus, sissejuhatus erialasse ning robotite programmeerimine, õpisooritust kõige paremini. Lisaks sellele leidis autor seosed küsitluse vastustest leitud õppimisega seotud psühholoogiliste tegurite ning Moodle'i aktiivsuslogide andmete vahel. Magistritöö tulemused aitavad kaasa informatiivsema õpianalüütika koostamisele TalTechis, mis omakorda aitab tulevikus paremini tuvastada tudengeid, kes võivad õppained läbi kukkuda. Abivajavate tudengite õigeaegne toetamine aitab aga vähendada tudengite väljalangemist inseneeria ning IKT valdkonna erialadelt.

Autor leidis magistritöös vastused kolmele uurimisküsimusele. Neist esimeseks oli „Mil määral saab Moodle'i keskkonna aktiivsus põhistel logidel baseeruvat õppeedukuse ennustamise täpsust parandada, lisades tudengite poolt küsitluse vastustena esitatud andmetest leitud õppimisega seotud psühholoogilised faktorid?“.

Töö tulemustest selgus, et piisava suurusega andmehulkade korral saavutavad nii Moodle'i logidel põhinevate andmete peal treenitud masinõppemudelid kui ka küsitluse küsimusi või siis küsitluse koondtunnuseid kasutavad masinõppemudelid märkimisväärse suurusega ennustustäpsuseid tudengite õppeedukuse klasside ennustamisel. Lihtsakoelistest Moodle'i logiridade arve kasutavatest mudelitest on piisava suurusega treeningandmehulga korral alati paremad detailsemaid logiandmehulkasid kasutanud mudelid, näiteks aktiivsusnäitajaid või kursuse jooksvaid tulemusi kasutanud mudelid. Õppimisega seotud psühholoogiliste tegurite andmed olid kõige usaldusväärsemad programmeerimise algkursuse korral. Küsitluse koondtunnuste peal treenitud mudel ennustas selle kursuse sooritustulemusi kõige paremini, saavutades ennustustäpsuse 78,8%. Moodle'i logiandmete lisamine küsitluse koondtunnustele ei parandanud masinõppemudelite ennustustäpsusi. Küsitluse küsimusi kasutanud masinõppemudeli ennustustäpsus (67,3%) ei olnud alati parem Moodle'i logiandmeid kasutanud mudelite tulemustest, kuid küsitluste küsimustele tudengite logiridade arvude lisamine parandas masinõppetulemust. Uue mudeli ennustustäpsus (69,2%) muutus paremaks semestri esimese kuu Moodle'i logiandmeid kasutanud mudelite tulemustest. Seega saab autor

soovitada õppimisega seotud psühholoogiliste tegurite andmete kasutamist sooritustulemuste ennustamiseks, sest küsitluse koondtunnuseid kasutav mudel on piisavalt suure treeningandmehulga korral parem kõikidest logide andmeid kasutanud mudelitest ning küsitluste küsimuste lisamine Moodle'i logide andmetele muudab ennustustäpsused samuti suuremaks.

Teiseks uurimisküsimuseks oli „Mil määral korreleeruvad Moodle'i keskkonna aktiivsuslogid õppimisega seotud psühholoogiliste faktorite andmetega, mis pärinevad tudengite vastustest küsitlusele (uskumused õppimise kohta, õpimotivatsioon, akadeemilised emotsioonid, õpikäitumine ning õpistrateegiad)?“.

Logiridade arvuga korreleerusid kõige rohkem alustamisega viivitamine, sisemine regulatsioon, omaksvõetud strateegiline regulatsioon, huvi, enesetõhusus, keeruliste ülesannete vältimine ja uskumused matemaatikaalase võimekuse kohta. Sissejuhatus erialasse semestri esimese kuu logiridade arvu ning keeruliste ülesannete vältimise vaheline Spearmani korrelatsioonikoefitsient oli -0,34. Alustamisega viivitamise Spearmani korrelatsioonikoefitsient oli programmeerimise algkursuse semestri kahe esimese kuu logiridade arvu korral -0,227. Seega kajastab aktiivsus logides näiteks seda, et tudengid alustavad ülesannete tegemist teistest hiljem ja, et osad tudengid väldivad keerulisi ülesandeid. Mida väiksem on aktiivsus logides, seda tõenäolisemalt on tudengitel need omadused.

Aktiivsete päevade arv korreleerus kõige enam alustamisega viivitamisega, sisemise regulatsiooniga, enesetõhususega, uskumustega matemaatikaalase võimekuse kohta, soorituspõhise motivatsioonilise lähenemisega, keeruliste ülesannete vältimisega ning ka tõhusate õpistrateegiatega. Näiteks robotite programmeerimise semestri esimese kahe kuu aktiivsete päevade arve sisaldava andmehulga korrelatsioonikoefitsient alustamisega viivitamisega oli -0,307.

Aktiivsete nädalate arv korreleerus üldiselt kõige vähem psühholoogiliste faktoritega, aga leidis siiski psühholoogilisi faktoreid, millega esines korrelatsioon, näiteks uskumused matemaatikaalase võimekuse kohta, pingutuse olulisus programmeerimiskursusel, huvikaotus, huvi, läbikukkumise vältimisele suunatud soorituseesmärgid, õppejõu pakutava emotsionaalse ja õppimisalase toe tajumine ning sotsiaalne õppimine. Sissejuhatus erialasse semestri esimese kuu aktiivsete nädalate arvu alusel joonistus välja,

et andmehulga korrelatsioonikoefitsient huviga oli $-0,259$ ning huvikaotusega $0,282$, mis on märkimisväärne, sest suurem aktiivsete nädalate arv vastab üllataval kombel madalale huvile. Õppejõu pakutava emotsionaalse ja õppimisalase toe tajumise korrelatsioonikoefitsient nii sissejuhatus erialasse esimese kuu kui ka kahe esimese kuu aktiivsete nädalate andmetega näitas, et mida suurem on tudengi aktiivsete nädalate arv, seda vähem tajuvad nad õppejõu pakutavat emotsionaalset ja õppimisalast tuge.

Korrelatsioonikoefitsiendid kõigi õppeainete keskmiste jooksvate tulemustega olid üldiselt psühholoogiliste teguritega kõige suuremad. Näiteks korreleerusid selle Moodle'i logiandmete hulga enesetõhusus, alustamisega viivitamine, keeruliste ülesannete vältimine, pingutuse olulisus programmeerimiskursusel, lihtsate ülesannete vältimine ning tõhusad- ja mittetõhusad õpistrateegiad. Näiteks nii esimese kui ka kahe esimese kuu programmeerimise algkursuse andmete korrelatsioonikoefitsient oli enesetõhususe korral suurem väärtusest $0,4$.

Seega korreleerusid aktiivsusega seotud logiandmed õpimotivatsiooni, õpistrateegiate ning õppimisega seotud uskumuste faktoritega. Hinnetega seotud logiandmed korreleerusid kõige enam õpistrateegiatega seotud psühholoogiliste faktoritega. Aktiivsete nädalate arv omas tugevamat seost õppimise sotsiaal-emotsionaalsete aspektidega.

Viimaseks uurimisküsimuseks oli „Millised küsitluse vastuste andmetest leitud psühholoogilised faktorid näitavad statistiliselt märkimisväärset suhet aktiivsustritega ning kui varakult ilmnevad need näitajad õpiprotsessis?“

Autor uuris magistritöös kaht logiandmete gruppi, semestri esimese kuu andmed ning semestri esimese kahe kuu andmed. Aktiivsustrid avalduvad selgelt juba esimese kuu andmetest, seda eriti kursuste korral, kus Moodle'i keskkonnas toimub iganädalane aktiivne tegevus, näiteks koduülesannete esitamine. Teise kuu andmete lisamine täpsustas sooritustulemuste ennustustulemusi ning ka korrelatsioone psühholoogiliste faktoritega. Näiteks oli alustamisega viivitamise Spearmani korrelatsioonikoefitsient programmeerimise algkursuse esimese kuu logiridade arvuga $-0,221$ ning esimese kahe kuu andmete korral $-0,227$. Moodle'i logide andmetega kõige enam korreleerunud psühholoogilised tegurid olid kas samad või sarnase sisuga (näiteks huvi ja huvikaotus) nii esimese kuu kui ka kahe esimese kuu andmeid kasutades. Seega võib järeldada, et

aktiivsustrid avalduvad juba kursuse esimesel kuul ning teise kuu andmete lisamine aitab täpsustada tulemusi.

7.1 Olulised tähelepanekud

Kõige olulisemaks tähelepanekuks on andmehulkade suurustest lähtuvalt tekkinud probleemid. Robotite programmeerimise ja sissejuhatus erialasse kursuste tudengitest ei vastanud piisavalt suur hulk õppimisega seotud psühholoogiliste tegurite küsitlusele. Puudulikud andmed mõjutasid tugevalt ennustustulemusi, kuna vaid üksikud küsitlusele vastanud tudengid olid need kursused läbi kukkunud ning masinõppemudelid klassifitseerisid tudengeid ainult kõige rohkem esinenud klassidesse. Sellest johtuvalt polnud neid andmeid kasutades leitud tulemused usaldusväärsed ning kasutavad järelduste tegemiseks.

Probleemi lahendamiseks on kaks peamist võimalust. Magistritöö raames viidi õppimisega seotud psühholoogiliste tegurite küsitlus läbi alles semestri lõpus, seega ei olnud vastanute hulgas tudengid, kes olid juba programmeerimise algkursuse või üldse ülikooli pooleli jätanud. Tulevikus oleks mõistlik õppimisega seotud psühholoogiliste tegurite küsitlus viia läbi semestri alguses, et võimalikult suur hulk tudengeid vastaks küsimustele. Teiseks võimaluseks oleks täiendavate kursuse konteksti sobivate treeningandmete genereerimine.

Autor tooks välja ka selle, et paremate mitmete kursuste jaoks sobivate masinõppemudelite loomiseks tuleks täiendada ka keskmiste jooksvate tulemuste leidmist. Tulemused võiks viia protsentuaalsele skaalale, et üldistada erinevaid hindamissüsteeme kasutavate kursuste jooksvad tulemused.

Kasutatud kirjandus

- [1] K. Mardo, "Info- ja kommunikatsioonitehnoloogia erialade tudengite õpingud kõrgkoolis: esimesel aastal väljalangemine ja õpingute jätkamine," Tartu, 2016.
- [2] M. Á. Conde, F. J. García-Peñalvo, Á. Hernández-García and M. L. Séin-Echaluze, "Exploring Student Interactions: Learning Analytics Tools," *Springer International Publishing Switzerland*, vol. 9192, lk 50-61, 2015.
- [3] K. Ehala, "Kontekstipõhine õppeedukuse monitoorimise mudel," Tallinn, 2018.
- [4] D. Clow, "The learning analytics cycle: closing the loop effectively," *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, lk 134–138, 2012.
- [5] A. Atif, J. C. Froissard, D. Y. T. Liu and D. Richards, "An enhanced learning analytics plugin for Moodle: Student engagement and personalised intervention," *Australasian Society for Computers in Learning and Tertiary Education, Conference Proceedings*, lk 180-189, 2019.
- [6] D. Clow, "An overview of learning analytics," *Teaching in Higher Education*, vol. 18, no. 6, lk 683-695, 2013.
- [7] "Moodle Statistics," [Online]. Available: <https://stats.moodle.org/>. [Accessed 1. mai 2020].
- [8] "Moodle Releases," [Online]. Available: <https://docs.moodle.org/dev/Releases>. [Accessed 1. mai 2020].
- [9] I. Mwalumbwe and J. S. Mtebe, "Using Learning Analytics to Predict Students' Performance in Moodle Learning Management System: A Case of Mbeya University of Science and Technology," *The Electronic Journal of Informations Systems in Developing Countries*, vol. 79, no. 1, lk 1-13, 2017.
- [10] C. Romero, M. López, J. Luna and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Computers & Education*, vol. 68, lk 458-472, 2013.
- [11] K. D. Strang, "Beyond engagement analytics: which online mixed-data factors predict student learning outcomes?," *Education and Information Technologies*, vol. 22, lk 917–937, 2017.
- [12] B. Dietz-Uhler and J. E. Hurn, "Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective," *Journal of Interactive Online Learning*, vol. 12, no. 1, lk 17-26, 2013.
- [13] S. Dawson, D. Gašević, D. Gasevic and T. Rogers, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet and Higher Education*, vol. 28, lk 68-84, 2016.
- [14] B. Uga, "TTÜ tudengite väljalangemise ennustamine: Tõenäosuse arvutamine masinõppe meetodite abil ning tulemuste kuvamine veebirakenduses," Tallinn, 2017.
- [15] K. Umbleja, "Competence Based Learning – Framework,," Tallinn, 2017.

- [16] B. Petjärv, “Enesejuhitud õppimise toetamine võimekususkumuste mõjutamisega,” Tallinn, 2018.
- [17] T. Goetz, P. Reinhard, R. P. Perry and T. Wolfram, “Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research,” *Educational Psychologist*, vol. 37, no. 2, lk 91–105, 2010.
- [18] M. H. Ashcraft, “Math Anxiety: Personal, Educational, and Cognitive Consequence,” *Current Directions in Psychological Science*, vol. 11, no. 5, lk 181–185, 2002.
- [19] R. A. Bjork, J. Dunlosky and N. Kornell, “Self-Regulated Learning: Beliefs, Techniques, and Illusions,” *Annual Review of Psychology*, vol. 64, lk 417–444, 2012.
- [20] T. Garcia, M. W. J., P. R. Pintrich and D. A. F. Smith, “A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ),” Michigan, 1991.
- [21] K. Aunola, M. Lindroos, J. Nurmi and K. Salmela-Aro, “The role of success expectation and task-avoidance in academic performance and satisfaction: Three studies on antecedents, consequences and correlates,” *Contemporary Educational Psychology*, vol. 28, lk 59–90, 2003.
- [22] S. Hidi and K. A. Renninger, “The Four-Phase Model of Interest Development,” *Educational Psychologist*, vol. 41, no. 2, lk 111-127, 2006.
- [23] K. D. Burton, D. U. D’Alessandro, J. E. Lydon and R. Koestner, “The Differential Effects of Intrinsic and Identified Motivation on Well-Being and Performance: Prospective, Experimental, and Implicit Approaches to Self-Determination Theory,” *Journal of Personality and Social Psychology*, vol. 91, no. 4, lk 750–762, 2006.
- [24] C. Abraham, M. Richardson and R. Bond, “Psychological correlates of University students' academic performance: a systematic review and meta analysis,” *Psychological Bulletin*, vol. 138, no. 2, lk 353-387, 2012.
- [25] G. Tam, “Interpreting Decision Trees and Random Forests,” Pivotal Engineering Journal, [Online]. Available: <http://engineering.pivotal.io/post/interpreting-decision-trees-and-random-forests/>. [Accessed 6. mai 2020].
- [26] C. Molnar, *Interpretable Machine Learning*, 2020.
- [27] “Decision Tree Classifier,” [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>. [Accessed 1. mai 2020].
- [28] “Precision Score,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html. [Accessed 6. mai 2020].
- [29] “Recall Score,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html. [Accessed 6. mai 2020].
- [30] “Multilabel Confusion Matrix,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.multilabel_confusion_matrix.html. [Accessed 6. mai 2020].

- [31] A. Ezen-Can, K. E. Boyer, S. Kellogg and S. Booth, “Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach,” *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, lk 146–150, 2015.
- [32] Y. Park, J. H. Yu and I. Jo, “Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute,” *The Internet and Higher Education*, vol. 29, lk 1-11, 2016.
- [33] A. M. Navarro and P. Moreno-Ger, “Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, lk 9-16, 2018.
- [34] C. Giraud, Introduction to high-dimensional statistics, Boca Raton: CRC Press, 2015, lk 50–53.
- [35] J. E. Cavanaugh and A. A. Neath, “The Bayesian information criterion: background, derivation, and applications,” *WIREs Computational Statistics*, vol. 4, no. 2, lk 199-202, 2012.
- [36] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, lk 53, 1987.
- [37] “Silhouette Score,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html. [Accessed 1. mai 2020].
- [38] J. F. M. Alviste, “Programmeerimisülesannete automaattestimissüsteemi liidestus Moodle’i keskkonnaga ja mugav kasutajaliides tudengite hindamiseks,” 2017.
- [39] “What are events,” [Online]. Available: https://docs.moodle.org/dev/Events_API#What_are_events.3F. [Accessed 1. mai 2020].
- [40] “Õppekorralduse eeskiri,” [Online]. Available: <https://www.ttu.ee/tudengile/oppeinfo/oppekorraldus/oppetegevuse-juhendid-ja-oigusaktid/oppee/>. [Accessed 1. mai 2020].
- [41] “Select k Best,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html. [Accessed 9. mai 2020].
- [42] “Chi-squared,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html. [Accessed 9. mai 2020].
- [43] “Spearman Correlation Coefficient,” [Online]. Available: <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html>. [Accessed 10. mai 2020].

Lisa 1 – Võrdlusbaasi mudelite õppedukuse klasside ennustustulemused semestri esimese kuu andmetele tuginedes

Õppedukuse klasside võrdlusbaasi ennustustulemused esimese kuu andmete pealt								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
python	benchmark	0.396	0.405	0.465	0.188	0.385	0.426	0.3
robots	benchmark	0.545	0	0.75	0	0	0.706	0
intro	benchmark	0.714	0.333	0.778	-	0.2	0.875	-
train python test robots	benchmark	0.279	1	0.028	0	0.266	1	1
train python test intro	benchmark	0.557	0.545	0.56	-	0.343	0.855	-
train robots test intro	benchmark	0.792	0	1	-	1	0.792	-
train robots test python	benchmark	0.479	0.018	1	0	1	0.476	1
train intro test robots	benchmark	0.261	1	0	0	0.261	1	1
train intro test python	benchmark	0.473	0.369	0.727	0	0.419	0.497	1

Õppedukuse klasside võrdlusbaasi ennustustulemused esimese kuu andmete pealt													
Course	Dataset	true negatives			false positives			false negatives			true positives		
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
python	benchmark	35	26	73	22	23	13	24	27	7	15	20	3
robots	benchmark	16	1	17	2	4	4	4	5	1	0	12	0
intro	benchmark	14	1	-	2	4	-	4	2	-	1	14	-
train python test robots	benchmark	2	39	101	0	70	10	80	0	0	29	2	0
train python test intro	benchmark	61	14	-	10	37	-	23	8	-	12	47	-
train robots test intro	benchmark	84	0	-	22	0	-	0	22	-	0	84	-
train robots test python	benchmark	312	3	395	165	0	85	0	250	0	3	227	0
train intro test robots	benchmark	0	39	101	0	72	10	82	0	0	29	0	0
train intro test python	benchmark	226	86	395	106	62	85	86	167	0	62	165	0

Lisa 2 – Võrdlusbaasi mudelite õppedukuse klasside ennustustulemused semestri esimese kahe kuu andmetele tuginedes

Õppedukuse klasside võrdlusbaasi ennustustulemused esimese kahe kuu andmete pealt								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
python	benchmark	0.49	0.568	0.558	0.125	0.5	0.522	0.25
robots	benchmark	0.591	0	0.75	0.25	0	0.75	0.5
intro	benchmark	0.667	0.333	0.722	-	0.167	0.867	-
train python test robots	benchmark	0.315	0.793	0.167	0	0.256	0.706	0
train python test intro	benchmark	0.613	0.682	0.595	-	0.405	0.877	-
train robots test intro	benchmark	0.764	0.182	0.917	-	0.667	0.856	-
train robots test python	benchmark	0.475	0.173	0.815	0.165	0.784	0.505	0.182
train intro test robots	benchmark	0.315	0.966	0.097	0	0.275	0.778	1
train intro test python	benchmark	0.54	0.476	0.789	0	0.571	0.526	1

Õppedukuse klasside võrdlusbaasi ennustustulemused esimese kahe kuu andmete pealt													
Course	Dataset	true negatives			false positives			false negatives			grade 4-5		
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
python	benchmark	38	31	74	16	19	14	21	22	6	21	24	2
robots	benchmark	16	2	17	2	4	3	4	4	1	0	12	1
intro	benchmark	13	1	-	2	5	-	5	2	-	1	13	-
train python test robots	benchmark	15	34	97	6	60	10	67	5	4	23	12	0
train python test intro	benchmark	62	15	-	7	34	-	22	7	-	15	50	-
train robots test intro	benchmark	82	9	-	18	7	-	2	13	-	4	77	-
train robots test python	benchmark	304	72	332	139	42	71	8	181	63	29	185	14
train intro test robots	benchmark	8	37	101	1	65	10	74	2	0	28	7	0
train intro test python	benchmark	252	92	395	88	48	85	60	161	0	80	179	0

Lisa 3 – Õppeedukuse klasside ennustustulemused koondtunnusute kaupa

Õppeedukuse klasside ennustamine koondtunnuste kaupa								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
	SisemineRegulatsioon	0.621	0.125	0.912	0.045	0.143	0.667	0.5
	OmaksvõetudStratRegulatsioon	0.621	0	0.912	0.091	1	0.65	0.286
	PealesurutudRegulatsioon	0.609	0.125	0.895	0.045	0.5	0.646	0.167
	PerfApproach	0.655	0	1	0	1	0.655	1
	PerfAvoidance	0.655	0	1	0	1	0.655	1
	MateUskumused	0.655	0	1	0	1	0.655	1
	SotsAnxUskumused	0.575	0.125	0.807	0.136	0.5	0.639	0.231
	EffortinLearning	0.655	0	1	0	1	0.655	1
	Enesetõhusus	0.678	0.125	0.842	0.455	0.333	0.727	0.556
	Huvi	0.655	0	1	0	1	0.655	1
	Anxiety	0.621	0	0.789	0.409	1	0.682	0.429
	Burnout	0.586	0	0.895	0	1	0.638	0
	Huvikaotus	0.644	0.125	0.965	0	0.333	0.655	1
	TaskAvoidance	0.678	0.125	0.912	0.273	1	0.693	0.545
	Easytasks	0.655	0	1	0	1	0.655	1
	Prokkimine	0.644	0	0.912	0.182	1	0.675	0.4
	TõhusStrat	0.586	0	0.877	0.045	0	0.641	0.143
	MittetõhusStrat	0.609	0	0.859	0.182	1	0.662	0.308
	SotsStrat	0.655	0	1	0	1	0.655	1
python	EmToetus	0.632	0	0.895	0.182	0	0.662	0.5

Õppeedukuse klasside ennustamine koondtunnuste kaupa													
Course	Dataset	true negatives			false positives			false negatives			true positives		
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
	SisemineRegulatsioon	73	4	64	7	5	21	6	26	1	1	52	1
	OmaksvõetudStratRegulatsioon	79	2	60	8	5	20	0	28	5	0	52	2
	PealesurutudRegulatsioon	78	2	60	7	6	21	1	28	5	1	51	1
	PerfApproach	79	0	65	8	0	22	0	30	0	0	57	0
	PerfAvoidance	79	0	65	8	0	22	0	30	0	0	57	0
	MateUskumused	79	0	65	8	0	22	0	30	0	0	57	0
	SotsAnxUskumused	78	4	55	7	11	19	1	26	10	1	46	3
	EffortinLearning	79	0	65	8	0	22	0	30	0	0	57	0
	Enesetõhusus	77	12	57	7	9	12	2	18	8	1	48	10
	Huvi	79	0	65	8	0	22	0	30	0	0	57	0
	Anxiety	79	9	53	8	12	13	0	21	12	0	45	9
	Burnout	79	1	58	8	6	22	0	29	7	0	51	0
	Huvikaotus	77	1	65	7	2	22	2	29	0	1	55	0
	TaskAvoidance	79	7	60	7	5	16	0	23	5	1	52	6
	Easytasks	79	0	65	8	0	22	0	30	0	0	57	0
	Prokkimine	79	5	59	8	5	18	0	25	6	0	52	4
	TõhusStrat	77	2	59	8	7	21	2	28	6	0	50	1
	MittetõhusStrat	79	5	56	8	8	18	0	25	9	0	49	4
	SotsStrat	79	0	65	8	0	22	0	30	0	0	57	0
python	EmToetus	77	4	61	8	6	18	2	26	4	0	51	4

Lisa 4 – Õppe edukuse klasside ennustustulemused semestri esimese kuu andmete pealt

Õppe edukuse klasside ennustamine esimese kuu andmete pealt								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
python	activity	0.521	0.324	0.84	0	0.667	0.487	1
	questionnaire	0.673	1	0.735	0.556	1	0.735	0.556
	grades	0.688	0.556	0.93	0.438	0.76	0.625	1
	average grades	0.583	0.486	0.721	0.438	0.62	0.596	0.467
	submission number	0.604	0.486	0.767	0.438	0.667	0.589	0.538
	questionnaire characteristics	0.788	0	0.875	0.813	1	0.8	0.765
	activity + grades	0.656	0.541	0.791	0.563	0.69	0.667	0.563
	activity + average grades	0.583	0.459	0.698	0.563	0.773	0.577	0.409
	questionnaire + log rows	0.692	0	0.813	0.625	1	0.743	0.588
	questionnaire characteristics + log rows	0.712	0	0.813	0.686	0	0.765	0.688
	questionnaire characteristics + activity	0.788	0	0.875	0.813	1	0.8	0.765
questionnaire characteristics + average grades	0.75	0	0.844	0.75	0	0.771	0.8	
python	questionnaire characteristics + grades	0.769	0	0.844	0.813	0	0.794	0.765
robots	activity	0.727	0	1	0	1	0.727	1
	questionnaire	0.923	0	1	0	1	0.923	1
	questionnaire characteristics	0.538	0	1	0	1	0.538	1
robots	questionnaire characteristics + activity	0.462	0	1	0	1	0.462	1
intro	activity	0.857	0	1	-	1	0.857	-
	questionnaire	0.917	0	1	-	1	0.917	-
	questionnaire characteristics	0.583	0	1	-	1	0.583	-
	questionnaire characteristics + activity	0.333	0	1	-	1	0.333	-
train python test robots	activity	0.261	1	0	0	0.261	1	1
	questionnaire characteristics + activity	0.619	1	0.566	0.889	0.333	0.968	0.276
train python test intro	activity	0.566	0.409	0.607	-	0.333	0.836	-
	questionnaire characteristics + activity	0.492	0	0.5	-	0	0.968	-
train robots test intro	activity	0.585	0.682	0.56	-	0.288	0.87	-
	questionnaire characteristics + activity	0.984	0	1	-	1	0.984	-
train robots test python	activity	0.473	0	1	0	1	0.473	1
	questionnaire characteristics + activity	0.66	0	1	0	1	0.66	1
train intro test robots	activity	0.261	1	0	0	0.261	1	1
	questionnaire characteristics + activity	0.841	0	1	0	1	0.841	1
train intro test python	activity	0.504	0.571	0.643	0	0.468	0.531	1
	questionnaire characteristics + activity	0.66	0	1	0	1	0.66	1

Oppeetukuse klasside ennumamine esimese kuu andmete pealt

Course	Dataset	true negatives					false positives					false negatives					true positives								
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5									
python	activity	53	13	80	25	5	16	6	40	0	12	38	0	48	11	28	4	7	6	0	9	8	0	25	10
	questionnaire	53	29	80	18	3	9	6	24	0	19	40	7	48	32	72	19	12	9	11	21	8	18	72	7
	average grades	50	30	74	19	10	9	9	23	6	18	33	7	50	30	74	19	10	9	9	23	6	18	33	7
	submission number	48	13	32	4	4	3	0	7	4	0	28	13	48	13	32	4	4	3	0	7	4	0	28	13
	questionnaire characteristics	50	36	73	17	9	7	9	17	7	20	34	9	50	36	73	17	9	7	9	17	7	20	34	9
	activity + average grades	54	31	67	20	13	7	5	22	13	17	30	9	54	31	67	20	13	7	5	22	13	17	30	9
	questionnaire + log rows	48	11	29	4	6	6	0	9	7	0	26	10	48	11	29	4	6	6	0	9	7	0	26	10
	questionnaire characteristics + log rows	46	12	31	4	6	5	2	8	5	0	26	11	46	12	31	4	6	5	2	8	5	0	26	11
	questionnaire characteristics + activity	48	13	32	4	4	3	0	7	4	0	28	13	48	13	32	4	4	3	0	7	4	0	28	13
	questionnaire characteristics + average grades	46	12	33	4	5	4	2	8	3	0	27	12	46	12	33	4	5	4	2	8	3	0	27	12
robotics	questionnaire characteristics + grades	47	13	32	4	5	3	1	7	4	0	27	13	47	13	32	4	5	3	1	7	4	0	27	13
	activity	20	0	18	2	0	4	0	6	0	16	0	20	0	18	2	0	4	0	6	0	16	0		
	questionnaire	13	0	12	0	0	1	0	1	0	12	0	13	0	12	0	0	1	0	1	0	12	0		
	questionnaire characteristics	13	0	7	0	0	6	0	6	0	0	7	0	13	0	7	0	0	6	0	6	0	0	7	0
	questionnaire characteristics + activity	12	0	7	1	0	6	0	7	0	0	6	0	12	0	7	1	0	6	0	7	0	0	6	0
intro	activity	18	0	-	3	0	-	0	3	-	0	18	-	18	0	-	3	0	-	0	3	-	0	18	-
	questionnaire	11	0	-	1	0	-	0	1	-	0	11	-	11	0	-	1	0	-	0	1	-	0	11	-
	questionnaire characteristics	12	0	-	0	0	-	0	5	-	0	7	-	12	0	-	0	0	-	0	5	-	0	7	-
	questionnaire characteristics + activity	11	0	-	1	0	-	0	8	-	0	4	-	11	0	-	1	0	-	0	8	-	0	4	-
	activity	0	39	101	0	72	10	82	0	0	29	0	0	0	39	101	0	72	10	82	0	0	29	0	0
train python test robots	questionnaire characteristics + activity	60	9	33	0	23	1	2	1	21	1	30	8	60	9	33	0	23	1	2	1	21	1	30	8
	activity	66	12	-	13	33	-	18	10	-	9	51	-	66	12	-	13	33	-	18	10	-	9	51	-
train python test intro	questionnaire characteristics + activity	58	0	-	1	30	-	2	1	-	30	-	58	0	-	1	30	-	2	1	-	30	-		
	activity	47	15	-	7	37	-	37	7	-	15	47	-	47	15	-	7	37	-	37	7	-	15	47	-
train robots test intro	questionnaire characteristics + activity	60	0	-	1	0	-	0	1	-	0	60	-	60	0	-	1	0	-	0	1	-	0	60	-
	activity	312	0	395	168	0	85	0	253	0	227	0	312	0	395	168	0	85	0	253	0	227	0		
train robots test python	questionnaire characteristics + activity	239	0	196	23	0	66	0	89	0	173	0	239	0	196	23	0	66	0	89	0	173	0		
	activity	0	39	101	0	72	10	82	0	0	29	0	0	0	39	101	0	72	10	82	0	0	29	0	
train intro test robots	questionnaire characteristics + activity	62	0	54	1	0	9	0	10	0	53	0	62	0	54	1	0	9	0	10	0	53	0		
	activity	203	124	395	72	81	85	109	129	0	96	146	0	203	124	395	72	81	85	109	129	0	96	146	0
train intro test python	questionnaire characteristics + activity	239	0	196	23	0	66	0	89	0	173	0	239	0	196	23	0	66	0	89	0	173	0		

Lisa 5 – õppedukuse klasside ennustustulemused semestri kahe esimese kuu andmete pealt õppeainete kaupa

Õppedukuse klasside ennustamine kahe esimese kuu andmete pealt ainete kaupa								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
	activity	0.583	0.541	0.837	0	0.741	0.522	1
	questionnaire	0.673	0	0.781	0.625	1	0.735	0.556
	grades	0.76	0.784	0.814	0.563	0.829	0.714	0.75
	average grades	0.677	0.757	0.674	0.5	0.7	0.69	0.571
	submission number	0.719	0.784	0.744	0.5	0.806	0.681	0.615
	questionnaire characteristics	0.788	0	0.875	0.813	1	0.8	0.765
	activity + grades	0.708	0.757	0.698	0.625	0.848	0.682	0.526
	activity + average grades	0.74	0.784	0.744	0.625	0.725	0.744	0.769
	questionnaire + log rows	0.692	0	0.813	0.625	1	0.743	0.588
	questionnaire characteristics + log rows	0.673	0.25	0.719	0.688	0.2	0.742	0.688
	questionnaire characteristics + activity	0.788	0	0.875	0.813	1	0.8	0.765
	questionnaire characteristics + average grades	0.731	0	0.813	0.75	1	0.765	0.667
python	questionnaire characteristics + grades	0.788	0	0.875	0.813	0	0.824	0.765
	activity	0.727	0	1	0	1	0.727	1
	questionnaire	0.923	0	1	0	1	0.923	1
	grades	0.682	0	0.875	0.25	0	0.737	1
	average grades	0.727	0	0.938	0.25	0	0.75	1
	submission number	0.545	0	0.688	0.25	0	0.733	1
	questionnaire characteristics	0.538	0	1	0	1	0.538	1
	activity + grades	0.455	0	0.625	0	0	0.667	1
	activity + average grades	0.727	0	1	0	1	0.727	1
	questionnaire characteristics + activity	0.462	0	1	0	1	0.462	1
	questionnaire characteristics + average grades	0.462	0	1	0	1	0.462	1
robots	questionnaire characteristics + grades	0.462	0	1	0	1	0.462	1
	activity	0.857	0	1	-	1	0.857	-
	questionnaire	0.917	0	1	-	1	0.917	-
	grades	0.857	0	1	-	1	0.857	-
	average grades	0.857	0	1	-	1	0.857	-
	submission number	0.857	0	1	-	1	0.857	-
	questionnaire characteristics	0.583	0	1	-	1	0.583	-
	activity + grades	0.762	0.667	0.778	-	0.333	0.933	-
	activity + average grades	0.857	0	1	-	1	0.857	-
	questionnaire characteristics + activity	0.333	0	1	-	1	0.333	-
	questionnaire characteristics + average grades	0.333	0	1	-	1	0.333	-
intro	questionnaire characteristics + grades	0.333	0	1	-	1	0.333	-

Oppeedukuse klasside ennustamine kahe esimese kuu andmete pealt alinate kaupa

Course	Dataset	true negatives					false positives					false negatives					true positives									
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5										
python	activity	52	20	80	17	7	16	7	33	0	20	36	0	48	11	28	4	7	6	0	9	8	0	25	10	
	questionnaire	48	11	28	4	7	6	0	9	8	0	25	10	53	39	77	8	8	7	6	14	3	29	35	9	
	grades	47	40	74	9	14	8	12	13	6	28	29	8	52	38	75	8	11	8	7	15	5	29	32	8	
	average grades	48	13	32	4	4	3	0	7	4	0	28	13	54	39	71	9	13	6	5	14	9	28	30	10	
	questionnaire characteristics	48	42	77	8	11	6	11	11	11	3	32	10	activity + average grades	48	13	32	4	4	3	0	7	4	0	28	13
	activity + average grades	48	13	32	4	4	3	0	7	4	0	28	13	questionnaire characteristics + activity	48	11	29	4	6	6	0	9	7	0	26	10
	questionnaire characteristics + activity	48	11	29	4	6	6	0	9	7	0	26	10	questionnaire + log rows	44	12	31	3	9	5	4	8	5	1	23	11
	questionnaire + log rows	44	12	31	3	9	5	4	8	5	1	23	11	questionnaire characteristics + average grades	48	12	30	4	6	4	0	8	6	0	26	12
	questionnaire characteristics + average grades	48	12	30	4	6	4	0	8	6	0	26	12	questionnaire characteristics + grades	47	14	32	4	4	3	1	6	4	0	28	13
	questionnaire characteristics + grades	47	14	32	4	4	3	1	6	4	0	28	13	activity	20	0	18	2	0	0	0	6	0	0	16	0
	activity	20	0	18	2	0	0	0	6	0	0	16	0	questionnaire	13	0	12	0	0	1	0	12	0	0	12	0
	questionnaire	13	0	12	0	0	1	0	12	0	0	12	0	grades	18	1	18	2	2	3	2	5	0	0	14	1
	grades	18	1	18	2	2	3	2	5	0	0	14	1	average grades	19	1	18	2	1	3	3	1	5	0	15	1
	average grades	19	1	18	2	1	3	3	1	5	0	15	1	submission number	14	2	18	2	5	3	6	4	0	0	11	1
	submission number	14	2	18	2	5	3	6	4	0	0	11	1	questionnaire characteristics	13	0	7	0	0	6	0	6	0	0	7	0
questionnaire characteristics	13	0	7	0	0	6	0	6	0	0	7	0	activity + grades	13	1	18	2	6	4	7	5	0	0	10	0	
activity + grades	13	1	18	2	6	4	7	5	0	0	10	0	activity + average grades	20	0	18	2	0	0	0	6	0	0	16	0	
activity + average grades	20	0	18	2	0	0	0	6	0	0	16	0	questionnaire characteristics + activity	12	0	7	1	0	6	0	7	0	0	6	0	
questionnaire characteristics + activity	12	0	7	1	0	6	0	7	0	0	6	0	questionnaire characteristics + average grades	12	0	7	1	0	6	0	7	0	0	6	0	
questionnaire characteristics + average grades	12	0	7	1	0	6	0	7	0	0	6	0	questionnaire characteristics + grades	12	0	7	1	0	6	0	7	0	0	6	0	
questionnaire characteristics + grades	12	0	7	1	0	6	0	7	0	0	6	0	activity	18	0	-	3	0	0	0	3	3	0	18	-	
activity	18	0	-	3	0	0	0	3	3	0	18	-	questionnaire	11	0	-	1	0	0	0	1	1	0	11	-	
questionnaire	11	0	-	1	0	0	0	1	1	0	11	-	grades	18	0	-	3	0	0	0	3	3	0	18	-	
grades	18	0	-	3	0	0	0	3	3	0	18	-	average grades	18	0	-	3	0	0	0	3	3	0	18	-	
average grades	18	0	-	3	0	0	0	3	3	0	18	-	submission number	18	0	-	3	0	0	0	3	3	0	18	-	
submission number	18	0	-	3	0	0	0	3	3	0	18	-	questionnaire characteristics	12	0	-	0	0	0	0	5	5	0	7	-	
questionnaire characteristics	12	0	-	0	0	0	0	5	5	0	7	-	activity + grades	14	2	-	1	4	4	4	1	1	2	14	-	
activity + grades	14	2	-	1	4	4	4	1	1	2	14	-	activity + average grades	18	0	-	3	0	0	0	3	3	0	18	-	
activity + average grades	18	0	-	3	0	0	0	3	3	0	18	-	questionnaire characteristics + activity	11	0	-	1	0	0	0	8	8	0	4	-	
questionnaire characteristics + activity	11	0	-	1	0	0	0	8	8	0	4	-	questionnaire characteristics + average grades	11	0	-	1	0	0	0	8	8	0	4	-	
questionnaire characteristics + average grades	11	0	-	1	0	0	0	8	8	0	4	-	questionnaire characteristics + grades	11	0	-	1	0	0	0	8	8	0	4	-	
questionnaire characteristics + grades	11	0	-	1	0	0	0	8	8	0	4	-	intro	11	0	-	1	0	0	0	8	8	0	4	-	
intro	11	0	-	1	0	0	0	8	8	0	4	-														

Lisa 6 – Õppeedukuse klasside ennustustulemused semestri kahe esimese kuu andmete pealt üldiste mudelite korral

Hindeklasside ennustamine 2 kuu andmete pealt - treenitud teise aine andmete peal								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
train python test robots	activity	0.423	0.931	0.278	0	0.325	0.714	1
	average grades	0.514	0.966	0.361	0.3	0.394	0.813	0.375
	activity + average grades	0.423	0.931	0.278	0	0.325	0.714	1
	questionnaire characteristics + activity	0.619	1	0.566	0.889	0.333	0.968	0.276
	questionnaire characteristics + average grades	0.619	1	0.566	0.889	0.333	0.968	0.276
train python test intro	activity	0.755	0.909	0.714	-	0.488	0.968	-
	average grades	0.17	0.682	0.036	-	0.268	0.75	-
	activity + average grades	0.755	0.909	0.714	-	0.488	0.968	-
	questionnaire characteristics + activity	0.492	0	0.5	-	0	0.968	-
	questionnaire characteristics + average grades	0.492	0	0.5	-	0	0.968	-
train robots test intro	activity	0.849	0.591	0.917	-	0.65	0.895	-
	average grades	0.217	0.591	0.119	-	0.241	0.667	-
	activity + average grades	0.849	0.591	0.917	-	0.65	0.895	-
	questionnaire characteristics + activity	0.836	0	0.85	-	1	0.981	-
	questionnaire characteristics + average grades	0.984	0	1	-	1	0.984	-
train robots test python	activity	0.588	0.333	0.996	0	0.982	0.534	1
	average grades	0.61	0.315	0.978	0.212	0.946	0.551	0.857
	activity + average grades	0.588	0.333	0.996	0	0.982	0.534	1
	questionnaire characteristics + activity	0.718	0	0.971	0.303	1	0.709	0.8
	questionnaire characteristics + average grades	0.66	0	1	0	1	0.66	1
train intro test robots	activity	0.306	0.724	0.181	0	0.239	0.565	1
	average grades	0.649	0	1	0	1	0.648	1
	activity + average grades	0.306	0.724	0.181	0	0.239	0.565	1
	questionnaire characteristics + activity	0.841	0	1	0	1	0.841	1
	questionnaire characteristics + average grades	0.841	0	1	0	1	0.841	1
train intro test python	activity	0.592	0.595	0.811	0	0.633	0.571	1
	average grades	0.473	0	1	0	1	0.473	0
	activity + average grades	0.592	0.595	0.811	0	0.633	0.571	1
	questionnaire characteristics + activity	0.66	0	1	0	1	0.66	1
	questionnaire characteristics + average grades	0.66	0	1	0	1	0.66	1

Hindeklasside ennustamine 2 kuu andmete pealt - treenitud teise alme andmete peal

Course	Dataset	true negatives					false positives					false negatives					true positives				
		grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5					
train python test robots	activity	26	31	101	2	52	10	56	8	0	27	20	0								
	average grades	39	33	96	1	46	7	43	6	5	28	26	3								
	activity + average grades	26	31	101	2	52	10	56	8	0	27	20	0								
	questionnaire characteristics + activity	60	9	33	0	23	1	2	1	21	1	30	8								
train python test robots	questionnaire characteristics + average grades	60	9	33	0	23	1	2	1	21	1	30	8								
	activity	63	20	-	2	24	-	21	2	-	20	60	-								
	average grades	43	21	-	7	81	-	41	1	-	15	3	-								
	activity + average grades	63	20	-	2	24	-	21	2	-	20	60	-								
train python test intro	questionnaire characteristics + activity	58	0	-	1	30	-	2	1	-	0	30	-								
	questionnaire characteristics + average grades	58	0	-	1	30	-	2	1	-	0	30	-								
	activity	77	13	-	9	7	-	7	9	-	13	77	-								
	average grades	43	17	-	9	74	-	41	5	-	13	10	-								
train robots test intro	activity + average grades	77	13	-	9	7	-	7	9	-	13	77	-								
	questionnaire characteristics + activity	60	0	-	1	9	-	0	1	-	0	51	-								
	questionnaire characteristics + average grades	60	0	-	1	9	-	0	1	-	0	60	-								
	activity	311	56	395	112	1	85	1	197	0	56	226	0								
train robots test python	average grades	309	72	392	115	5	67	3	181	3	53	222	18								
	activity + average grades	311	56	395	112	1	85	1	197	0	56	226	0								
	questionnaire characteristics + activity	239	20	191	23	5	46	0	69	5	0	168	20								
	questionnaire characteristics + average grades	239	0	196	23	0	66	0	89	0	0	173	0								
train intro test robots	activity	15	29	101	8	59	10	67	10	0	21	13	0								
	average grades	82	0	101	29	0	10	0	39	0	0	72	0								
	activity + average grades	15	29	101	8	59	10	67	10	0	21	13	0								
	questionnaire characteristics + activity	62	0	54	1	0	9	0	10	0	0	53	0								
train intro test python	questionnaire characteristics + average grades	62	0	54	1	0	9	0	10	0	0	53	0								
	activity	254	115	395	68	43	85	58	138	0	100	184	0								
	average grades	312	0	395	168	0	85	0	253	0	0	227	0								
	activity + average grades	254	115	395	68	43	85	58	138	0	100	184	0								
train intro test python	questionnaire characteristics + activity	239	0	196	23	0	66	0	89	0	0	173	0								
	questionnaire characteristics + average grades	239	0	196	23	0	66	0	89	0	0	173	0								

Lisa 7 – Õppeedukuse klasside ennustustulemused pärast parimate tunnuste valimist

Õppeedukuse klasside ennustamine esimese kuu andmete pealt koos tunnuste valimisega									
Course	Dataset	k	score	precision score			recall score		
				grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
	activity	9	0.531	0.405	0.837	0	0.6	0.507	1
	questionnaire characteristics	1	0.788	0	0.875	0.813	1	0.8	0.765
	activity + average grades	2	0.625	0.405	0.837	0.563	0.789	0.581	0.6
	questionnaire characteristics + log rows	9	0.712	0	0.813	0.686	0	0.765	0.688
	questionnaire characteristics + activity	5	0.788	0	0.875	0.813	1	0.8	0.765
	questionnaire characteristics + average grades	8	0.769	0	0.875	0.75	0	0.778	0.8
python	activity	9	0.727	0	1	0	1	0.727	1
robots	questionnaire characteristics	1	0.538	0	1	0	1	0.538	1
	questionnaire characteristics + activity	5	0.462	0	1	0	1	0.462	1
	activity	9	0.857	0	1	-	1	0.857	-
intro	questionnaire characteristics	1	0.583	0	1	-	1	0.583	-
	questionnaire characteristics + activity	5	0.333	0	1	-	1	0.333	-
	activity	9	0.261	1	0	1	0.261	1	1
train python test robots	questionnaire characteristics + activity	5	0.238	1	0.17	0.556	0.25	0.75	0.106
train python test intro	activity	9	0.311	0.955	0.143	-	0.226	0.923	-
	questionnaire characteristics + activity	5	0	0	0	-	1	1	-
train robots test intro	activity	9	0.792	0	1	-	1	0.792	-
	questionnaire characteristics + activity	5	0.984	0	1	-	1	0.984	-
train robots test python	activity	9	0.473	0	1	0	1	0.473	1
	questionnaire characteristics + activity	5	0.66	0	1	0	1	0.66	1
train intro test robots	activity	9	0.261	1	0	0	0.261	1	1
	questionnaire characteristics + activity	5	0.841	0	1	0	1	0.841	1
train intro test python	activity	9	0.504	0.571	0.643	0	0.468	0.531	1
	questionnaire characteristics + activity	5	0.66	0	1	0	1	0.66	1

Oppedukuse klasside ennumamine esinuse kuu andmete pealt koos tunnuste valimisega

Course	Dataset	k	true negatives					false positives					false negatives					true positives				
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5					
python	activity	9	49	18	80	22	7	16	10	35	0	15	36	0	0	0						
	questionnaire characteristics	1	48	13	32	4	4	3	0	7	4	0	28	13	0	0						
	activity + average grades	2	55	27	74	22	7	7	4	26	6	15	36	9	0	0						
	questionnaire characteristics + log rows	9	46	12	31	4	6	5	2	8	5	0	26	11	0	0						
	questionnaire characteristics + activity	5	48	13	32	4	4	3	0	7	4	0	28	13	0	0						
python	questionnaire characteristics + average grades	8	47	12	33	4	4	4	1	8	3	0	28	12	0	0						
	activity	9	20	0	18	2	0	4	0	6	0	0	16	0	0	0						
robots	questionnaire characteristics	1	13	0	7	0	0	6	0	6	0	0	7	0	0	0						
	questionnaire characteristics + activity	5	12	0	7	1	0	6	0	7	0	0	6	0	0	0						
	activity	9	18	0	-	3	0	-	0	3	-	0	18	-	0	0						
intro	questionnaire characteristics	1	12	0	-	0	0	-	0	5	-	0	7	-	0	0						
	questionnaire characteristics + activity	5	11	0	-	1	0	-	0	8	-	0	4	-	0	0						
train python test robots	activity	9	0	39	101	0	72	10	82	0	0	29	0	0	0	0						
	questionnaire characteristics + activity	5	59	7	12	0	44	4	3	3	42	1	9	3	0	0						
train python test intro	activity	9	12	21	-	1	72	1	72	1	-	21	12	-	0	0						
	questionnaire characteristics + activity	5	60	1	-	1	60	-	0	0	-	0	0	-	0	0						
train robots test intro	activity	9	84	0	-	22	0	-	0	22	-	0	84	-	0	0						
	questionnaire characteristics + activity	5	60	0	-	1	0	-	1	1	-	0	60	-	0	0						
train robots test python	activity	9	312	0	395	168	0	85	0	253	0	227	0	0	0	0						
	questionnaire characteristics + activity	5	239	0	196	23	0	66	0	89	0	173	0	0	0	0						
train intro test robots	activity	9	0	39	101	0	72	10	82	0	10	29	0	0	0	0						
	questionnaire characteristics + activity	5	62	0	54	1	0	9	0	10	0	53	0	0	0	0						
train intro test python	activity	9	203	124	395	72	81	85	109	129	0	96	146	0	0	0						
	questionnaire characteristics + activity	5	239	0	196	23	0	66	0	89	0	173	0	0	0	0						

Õppeedukuse klasside ennustamine esimese kahe kuu andmete pealt koos tunnuste valimisega

Course	Dataset	k	score	precision score			recall score		
				grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
	activity	6	0.615	0.568	0.86	0.063	0.778	0.561	0.333
	questionnaire characteristics	1	0.788	0	0.875	0.813	1	0.8	0.765
	activity + average grades	8	0.688	0.784	0.628	0.625	0.644	0.711	0.769
	questionnaire characteristics + log rows	2	0.731	0.25	0.906	0.5	0.5	0.725	0.8]
	questionnaire characteristics + activity	10	0.788	0	0.875	0.813	1	0.8	0.765
python	questionnaire characteristics + average grades	8	0.731	0	0.813	0.75	1	0.765	0.667
	activity	6	0.727	0	1	0	1	0.727	1
	questionnaire characteristics	1	0.538	0	1	0	1	0.538	1
	activity + average grades	8	0.727	0	1	0	1	0.727	1
	questionnaire characteristics + activity	10	0.462	0	1	0	1	0.462	1
	robots	questionnaire characteristics + average grades	8	0.462	0	1	0	1	0.462
	activity	6	0.857	0	1	-	1	0.857	-
	questionnaire characteristics	1	0.583	0	1	-	1	0.583	-
	activity + average grades	8	0.857	0	1	-	1	0.857	-
	questionnaire characteristics + activity	10	0.333	0	1	-	1	0.333	-
	intro	questionnaire characteristics + average grades	8	0.333	0	1	-	1	0.333

Oppeedukuse Klasside ennustamine esimese kahe kuu andmete pealt koos tunnuste valimisega

Course	Dataset	k	true negatives				false positives				false negatives				true positives			
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	
	activity	6	53	24	78	16	6	15	6	29	2	21	37					
	questionnaire characteristics	1	48	13	32	4	4	3	0	7	4	0	28					
	activity + average grades	8	43	42	77	8	16	6	16	11	3	29	27	10				
	questionnaire characteristics + log rows		47	9	34	3	3	8	1	11	2	1	29	8				
python	questionnaire characteristics + activity	10	48	13	32	4	4	3	0	7	4	0	28	13				
	questionnaire characteristics + average grades	8	48	12	30	4	6	4	0	8	6	0	26	12				
	activity	6	20	0	18	2	0	4	0	6	0	0	16	0				
	questionnaire characteristics	1	13	0	7	0	0	6	0	6	0	0	7	0				
robots	activity + average grades	8	20	0	18	2	0	4	0	6	0	0	16	0				
	questionnaire characteristics + activity	10	12	0	7	1	0	6	0	7	0	0	6	0				
	questionnaire characteristics + average grades	8	12	0	7	1	0	6	0	7	0	0	6	0				
	activity	6	18	0	-	3	0	-	0	3	-	0	18	-				
intro	questionnaire characteristics	1	12	0	-	0	0	-	0	5	-	0	7	-				
	activity + average grades	8	18	0	-	3	0	0	0	3	-	0	18	-				
	questionnaire characteristics + activity	10	11	0	-	1	0	-	8	-	0	4	-					
	questionnaire characteristics + average grades	8	11	0	-	1	0	-	8	-	0	4	-					

Hindeklasside ennustamine 2 kuu andmete pealt - treenitud teise aine andmete peal								
Course	Dataset	score	precision score			recall score		
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5
train python test robots	activity	0.423	0.931	0.278	0	0.325	0.714	1
	average grades	0.514	0.966	0.361	0.3	0.394	0.813	0.375
	activity + average grades	0.423	0.931	0.278	0	0.325	0.714	1
	questionnaire characteristics + activity	0.619	1	0.566	0.889	0.333	0.968	0.276
train python test intro	activity	0.755	0.909	0.714	-	0.488	0.968	-
	average grades	0.17	0.682	0.036	-	0.268	0.75	-
	activity + average grades	0.755	0.909	0.714	-	0.488	0.968	-
	questionnaire characteristics + activity	0.492	0	0.5	-	0	0.968	-
train robots test intro	activity	0.849	0.591	0.917	-	0.65	0.895	-
	average grades	0.217	0.591	0.119	-	0.241	0.667	-
	activity + average grades	0.849	0.591	0.917	-	0.65	0.895	-
	questionnaire characteristics + activity	0.836	0	0.85	-	1	0.981	-
train robots test python	activity	0.588	0.333	0.996	0	0.982	0.534	1
	average grades	0.61	0.315	0.978	0.212	0.946	0.551	0.857
	activity + average grades	0.588	0.333	0.996	0	0.982	0.534	1
	questionnaire characteristics + activity	0.718	0	0.971	0.303	1	0.709	0.8
train intro test robots	activity	0.306	0.724	0.181	0	0.239	0.565	1
	average grades	0.649	0	1	0	1	0.648	1
	activity + average grades	0.306	0.724	0.181	0	0.239	0.565	1
	questionnaire characteristics + activity	0.841	0	1	0	1	0.841	1
train intro test python	activity	0.592	0.595	0.811	0	0.633	0.571	1
	average grades	0.473	0	1	0	1	0.473	0
	activity + average grades	0.592	0.595	0.811	0	0.633	0.571	1
	questionnaire characteristics + activity	0.66	0	1	0	1	0.66	1
train intro test python	questionnaire characteristics + average grades	0.66	0	1	0	1	0.66	1

Oppeedukuse klasside ennustamine esimese kahe kuu andmete pealt koos tunnuste valimisega

Course	Dataset	k	true negatives					false positives					false negatives					true positives				
			grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5	grade 0	grade 1-3	grade 4-5					
Course	activity	6	63	24	101	6	18	10	19	15	0	23	54	0	0	0						
	activity + average grades	8	36	33	101	2	40	10	46	6	0	27	32	0	0	0						
	questionnaire characteristics + activity	10	59	9	8	1	48	1	3	1	46	0	5	8	8	0						
train python test robots	questionnaire characteristics + average grades	8	60	9	33	0	23	1	2	1	21	1	30	8	8	0						
	activity	6	26	21	-	1	58	-	58	1	-	21	26	-	-	-						
	activity + average grades	8	63	20	-	2	21	-	21	2	-	20	63	-	-	-						
train python test intro	questionnaire characteristics + activity	10	58	1	-	1	59	-	2	0	0	1	1	-	-	-						
	questionnaire characteristics + average grades	8	58	0	-	1	30	-	2	1	-	0	30	-	-	-						
	activity	6	74	17	-	5	10	-	10	5	-	17	74	-	-	-						
train robots test intro	activity + average grades	8	77	13	-	9	7	-	7	9	-	13	77	-	-	-						
	questionnaire characteristics + activity	10	60	0	-	1	0	-	0	1	-	0	60	-	-	-						
	questionnaire characteristics + average grades	8	60	1	-	1	26	-	0	0	-	0	34	-	-	-						
train robots test python	activity	6	303	93	395	77	7	85	9	160	0	91	220	0	0	0						
	activity + average grades	8	311	56	395	112	1	85	1	197	0	56	226	0	0	0						
	questionnaire characteristics + activity	10	239	0	196	23	0	66	0	89	0	0	173	0	0	0						
train intro test robots	questionnaire characteristics + average grades	8	239	52	114	23	79	17	0	37	82	0	94	49	0	0						
	activity	6	41	24	101	10	36	10	41	15	0	19	36	0	0	0						
	activity + average grades	8	30	30	101	6	45	10	52	9	0	23	27	0	0	0						
train intro test python	questionnaire characteristics + activity	10	62	0	54	1	0	9	0	10	0	0	53	0	0	0						
	questionnaire characteristics + average grades	8	62	0	54	1	0	9	0	10	0	0	53	0	0	0						
	activity	6	254	115	395	68	43	85	58	138	0	100	184	0	0	0						
train intro test python	activity + average grades	8	254	115	395	68	43	85	58	138	0	100	184	0	0	0						
	questionnaire characteristics + activity	10	239	0	196	23	0	66	0	89	0	0	173	0	0	0						
	questionnaire characteristics + average grades	8	239	0	196	23	0	66	0	89	0	0	173	0	0	0						

Lisa 8 – Küsitluse koondtunnuste ennustustulemused programmeerimise algkursuse logiandmete pealt

Koondtunnuste ennustamine programmeerimise algkursuse andmete pealt						
		SisemineRegulatsioon	OmaksvõetudStratRegulatsioon	PealesurutudRegulatsioon	PerfApproach	PerfAvoidance
python (1 month)	activity	0.481	0.288	0.212	0.288	0.423
	grades	0.5	0.365	0.269	0.288	0.385
	submissions	0.365	0.269	0.212	0.288	0.385
	average grades	0.519	0.365	0.288	0.308	0.404
	average grades + activity	0.308	0.269	0.173	0.288	0.365
	grades + activity	0.346	0.327	0.25	0.269	0.404
python (2 months)	activity	0.481	0.385	0.423	0.288	0.462
	grades	0.5	0.365	0.269	0.308	0.5
	submissions	0.5	0.365	0.25	0.192	0.308
	average grades	0.481	0.442	0.288	0.25	0.462
	average grades + activity	0.288	0.327	0.288	0.269	0.308
	grades + activity	0.442	0.423	0.288	0.212	0.327

Koondtunnuste ennustamine programmeerimise algkursuse andmete pealt						
		MateUskumused	SotsAnxUskumused	EffortinLearning	Enesetõhusus	Huvi
python (1 month)	activity	0.269	0.288	0.346	0.365	0.308
	grades	0.327	0.231	0.423	0.423	0.327
	submissions	0.288	0.231	0.173	0.346	0.346
	average grades	0.212	0.269	0.365	0.442	0.365
	average grades + activity	0.404	0.288	0.173	0.346	0.231
	grades + activity	0.365	0.25	0.308	0.385	0.346
python (2 months)	activity	0.288	0.288	0.423	0.327	0.327
	grades	0.231	0.288	0.327	0.519	0.442
	submissions	0.25	0.269	0.288	0.25	0.211
	average grades	0.327	0.25	0.327	0.423	0.269
	average grades + activity	0.327	0.25	0.308	0.231	0.346
	grades + activity	0.365	0.365	0.327	0.423	0.404

Koondtunnuste ennustamine programmeerimise algkursuse andmete pealt						
		Anxiety	Burnout	Huvikaotus	TaskAvoidance	Easytasks
python (1 month)	activity	0.269	0.231	0.346	0.288	0.192
	grades	0.288	0.269	0.269	0.346	0.231
	submissions	0.231	0.135	0.3461	0.269	0.308
	average grades	0.404	0.269	0.25	0.327	0.173
	average grades + activity	0.25	0.192	0.308	0.25	0.154
	grades + activity	0.231	0.173	0.288	0.346	0.327
python (2 months)	activity	0.346	0.231	0.288	0.404	0.192
	grades	0.423	0.269	0.231	0.327	0.308
	submissions	0.25	0.327	0.25	0.308	0.288
	average grades	0.308	0.25	0.288	0.365	0.327
	average grades + activity	0.288	0.231	0.346	0.231	0.192
	grades + activity	0.346	0.288	0.308	0.269	0.25

Koondtunnuste ennustamine programmeerimise algkursuse andmete pealt						
		Prokkimine	TõhusStrat	MittetõhusStrat	SotsStrat	EmToetus
python (1 month)	activity	0.212	0.346	0.308	0.308	0.327
	grades	0.269	0.192	0.308	0.269	0.346
	submissions	0.269	0.154	0.308	0.25	0.365
	average grades	0.308	0.231	0.404	0.288	0.346
	average grades + activity	0.173	0.135	0.212	0.288	0.327
	grades + activity	0.308	0.288	0.327	0.269	0.346
python (2 months)	activity	0.288	0.231	0.308	0.327	0.385
	grades	0.346	0.346	0.404	0.327	0.404
	submissions	0.173	0.231	0.327	0.346	0.385
	average grades	0.192	0.231	0.385	0.346	0.423
	average grades + activity	0.135	0.288	0.365	0.308	0.308
	grades + activity	0.308	0.308	0.346	0.192	0.385

Lisa 9 – Semestri esimese kuu Moodle'i andmete Spearmani korrelatsioonikoefitsiendid küsitluse koondtunnustega

Spearmani korrelatsioonikoefitsiendid esimese kuu andmetega								
course	dataset	benchmark / activity	diff. from average (activity)	percentage from max (activity)	number_of_active_days	number_of_active_weeks	average_grade	
python	SisemineRegulatsioon	0.215	0.215	0.215	0.215	0.034	0.21	
	OmaksvoeludStratRegulatsioon	0.171	0.171	0.171	0.173	0.005	0.061	
	PealesurutudRegulatsioon	-0.027	-0.027	-0.027	-0.024	-0.019	0.02	
	PerfApproach	-0.026	-0.026	-0.026	-0.03	0.04	0.152	
	PerfAvoidance	-0.054	-0.054	-0.054	-0.052	-0.049	-0.205	
	MateUuskumused	0.088	0.088	0.088	0.085	0.049	-0.125	
	SotsAnUuskumused	0.063	0.063	0.063	0.058	0.046	-0.079	
	EffortInl-earning	0.086	0.086	0.086	0.084	0.029	-0.161	
	Eneasetohusus	0.079	0.079	0.079	0.078	0.02	0.416	
	Huvi	0.194	0.194	0.194	0.196	0.068	0.213	
	Anxleily	-0.02	-0.02	-0.02	-0.017	-0.025	-0.183	
	Burnout	-0.003	-0.003	-0.003	-0.008	-0.013	-0.154	
	Huvikaetus	-0.159	-0.159	-0.159	-0.165	0.063	-0.154	
	TaskAvoidance	-0.118	-0.118	-0.118	-0.121	-0.003	-0.297	
	Easytasks	0.088	0.088	0.088	0.089	-0.006	-0.052	
	Prokkimine	-0.221	-0.221	-0.221	-0.222	-0.006	-0.309	
	TohusStrat	-0.023	-0.023	-0.023	-0.015	-0.048	-0.037	
	MitteohusStrat	0.111	0.111	0.111	0.118	0.001	-0.202	
	SotsStrat	-0.047	-0.047	-0.047	-0.043	0.1	0.106	
	EmToetus	-0.008	-0.008	-0.008	-0.007	0.049	0.144	

Spearmani korrelatsiooni koefitsendid esimese kuu andmetega

course	dataset	benchmark / activity	diff_from_average (activity)	percentage_from_max (activity)	number_of_active_days	number_of_active_weeks
robots	SisemineRegulatsioon	-0.052	-0.052	-0.052	-0.007	-0.093
	OmaksvõetudStratRegulatsioon	0.086	0.086	0.086	0.127	-0.11
	PealesunutudRegulatsioon	-0.073	-0.073	-0.073	-0.053	-0.005
	PerfApproach	-0.067	-0.067	-0.067	-0.033	-0.001
	PerfAvoidance	-0.035	-0.035	-0.035	-0.026	0.135
	MateUskumused	0.155	0.155	0.155	0.17	0.224
	SotsAnxUskumused	-0.009	-0.009	-0.009	-0.002	0.107
	EffortinLearning	0.093	0.093	0.093	0.077	0.2
	Enesetõhusus	0.141	0.141	0.141	0.145	0.038
	Huvi	-0.027	-0.027	-0.027	-0.011	0.064
	Anxiety	0	0	0	0.019	0.098
	Burnout	0.051	0.051	0.051	0.064	-0.034
	Huvikaotus	-0.048	-0.048	-0.048	-0.04	0.018
	TaskAvoidance	-0.035	-0.035	-0.035	-0.049	0.055
	Easytasks	-0.113	-0.113	-0.113	-0.085	0.085
	Prokkimine	-0.177	-0.177	-0.177	-0.179	-0.126
	TohusStrat	0.031	0.031	0.031	0.067	-0.047
	MitteõhusStrat	-0.072	-0.072	-0.072	-0.04	-0.024
	SotsStrat	0.061	0.061	0.061	0.066	0.118
	EmTõetus	0.072	0.072	0.072	0.075	0.101

Spearmani korrelatsiooni koefitsendid esimese kuu andmetega

course	dataset	benchmark / activity	diff_from_average (activity)	percentage_from_max (activity)	number_of_active_days	number_of_active_weeks
intro	SisemineRegulatsioon	-0.079	-0.079	-0.079	-0.076	0.006
	OmaksvoetudStratRegulatsioon	-0.005	-0.005	-0.005	0.005	0.195
	PealesunrutudRegulatsioon	-0.019	-0.019	-0.019	-0.017	0.126
	PerfApproach	-0.106	-0.106	-0.106	-0.105	0.184
	PerfAvoidance	0.082	0.082	0.082	0.078	-0.009
	MateUskumused	0.241	0.241	0.241	0.244	-0.043
	SotsAnxUskumused	0.2	0.2	0.2	0.193	-0.008
	EffortInLearning	0.113	0.113	0.113	0.095	-0.09
	Enesetõhusus	0.131	0.131	0.131	0.133	-0.205
	Huvi	-0.062	-0.062	-0.062	-0.075	-0.259
	Anxiety	-0.082	-0.082	-0.082	-0.093	0.165
	Burnout	0.025	0.025	0.025	0.019	0.142
	HuviKaotus	0.017	0.017	0.017	0.001	0.282
	TaskAvoidance	-0.284	-0.284	-0.284	-0.285	0.197
	Easytasks	-0.004	-0.004	-0.004	-0.014	0.118
	Prokkimine	-0.29	-0.29	-0.29	-0.307	0.168
TõhusStrat	0.153	0.153	0.153	0.158	-0.036	
MittetõhusStrat	0.043	0.043	0.043	0.041	0.077	
SotsStrat	-0.136	-0.136	-0.136	-0.145	0.077	
EmTõetus	0.04	0.04	0.04	0.042	-0.303	

Lisa 10 – Semestri esimese kahe kuu Moodle'i andmete Spearmani korrelatsioonikoefitsiendid küsitluse koondtunnustega

course		dataset	benchmark / activity	diff. from average (activity)	percentage from_max (activity)	number_of_active_days	number_of_active_weeks	average_grade
python	SisemineRegulatsioon	0.231	0.231	0.231	0.236	0.095	0.195	
	OmaksõuetudStratRegulatsioon	0.219	0.219	0.219	0.192	0.097	0.143	
	PealesuritudRegulatsioon	0.066	0.066	0.066	0.018	0.072	0.076	
	PerfApproach	-0.001	-0.001	-0.001	-0.035	0.072	0.105	
	PerfAvoidance	0.017	0.017	0.017	-0.017	-0.041	-0.143	
	MateLuskumused	-0.001	-0.001	-0.001	0.023	-0.021	-0.141	
	SoisAnxSkumused	0.045	0.045	0.045	0.052	0.012	-0.074	
	EffortInLearning	0.098	0.098	0.098	0.074	-0.033	-0.171	
	Enesetõhusus	0.061	0.061	0.061	0.059	0.131	0.401	
	Huvi	0.182	0.182	0.182	0.193	0.143	0.201	
	Anxiety	0.024	0.024	0.024	0.027	0.034	-0.166	
	Burnout	0.067	0.067	0.067	0.038	0.048	-0.085	
	Huvikaotus	-0.121	-0.121	-0.121	-0.129	-0.05	-0.141	
	TaskAvoidance	-0.102	-0.102	-0.102	-0.105	-0.136	-0.308	
	EasyTasks	0.072	0.072	0.072	0.069	-0.023	-0.11	
	Proklimine	-0.227	-0.227	-0.227	-0.239	-0.11	-0.278	
TõhusStrat	0.002	0.002	0.002	-0.014	0.008	-0.032		
MitteTõhusStrat	0.2	0.2	0.2	0.184	0.01	-0.158		
SoisStrat	-0.018	-0.018	-0.018	-0.042	0.232	0.111		
EmTõetus	0.037	0.037	0.037	0.009	0.094	0.136		

Spearmani korrelatsiooni koefitsendid kahe esimese kuu andmetega

course	dataset	benchmark / activity	diff_from_average (activity)	percentage_from_max (activity)	number_of_active_days	number_of_active_weeks	average_grade
robots	SisemineRegulatsioon	0.027	0.027	0.027	-0.044	-0.004	0.228
	OmaksvoetudStratRegulatsioon	-0.077	-0.077	-0.077	-0.036	-0.053	0.179
	PealesurutudRegulatsioon	-0.075	-0.075	-0.075	-0.049	-0.03	-0.117
	PerfApproach	-0.235	-0.235	-0.235	-0.249	0.019	-0.153
	PerfAvoidance	-0.034	-0.034	-0.034	-0.057	0.076	-0.19
	MateUskumused	0.102	0.102	0.102	0.156	0.193	0.031
	SoitsAnUskumused	0.173	0.173	0.173	0.192	0.058	0.094
	EffortfulLearning	-0.112	-0.112	-0.112	-0.064	0.149	-0.307
	Enesetõhusus	0.175	0.175	0.175	0.162	0.138	0.301
	Huvi	0.146	0.146	0.146	0.1	0.077	0.174
	Anxiety	-0.041	-0.041	-0.041	0.012	0.04	0.052
	Burnout	-0.098	-0.098	-0.098	-0.07	-0.03	-0.041
	Huvikaotus	-0.099	-0.099	-0.099	-0.033	-0.11	-0.147
	TaskAvoidance	-0.148	-0.148	-0.148	-0.074	-0.045	-0.315
	Easytasks	0.022	0.022	0.022	0.078	0.07	0.023
	Prokkimine	-0.053	-0.053	-0.053	-0.016	-0.143	-0.32
	TõhusStrat	0.079	0.079	0.079	0.094	0.008	0.177
MitteTõhusStrat	-0.081	-0.081	-0.081	-0.028	-0.04	-0.029	
SoitsStrat	-0.009	-0.009	-0.009	0.031	0.186	-0.025	
EmToetus	-0.026	-0.026	-0.026	-0.046	0.168	0.022	

Spearmani korrelatsiooni koefitsendid kahe esimese kuu andmetega

course	dataset	benchmark / activity	diff_from_average (activity)	percentage_from_max (activity)	number_of_active_days	number_of_active_weeks	average_grade
intro	SisemineRegulatsioon	0.03	0.03	0.03	0.017	0.085	-0.071
	OmaksõuetudStratRegulatsioon	0.011	0.011	0.011	0.006	0.163	0.015
	PealesurutudRegulatsioon	-0.064	-0.064	-0.064	-0.04	-0.006	-0.045
	PerfApproach	-0.117	-0.117	-0.117	-0.11	0.071	0.093
	PerfAvoidance	0.081	0.081	0.081	0.089	-0.045	-0.106
	MateUskumused	0.154	0.154	0.154	0.174	-0.054	0.12
	SotsAnxUskumused	0.141	0.141	0.141	0.148	0.069	-0.011
	EffortInL_earning	0.152	0.152	0.152	0.15	-0.131	-0.096
	Ereasetõhusus	0.167	0.167	0.167	0.174	-0.109	-0.056
	Huvi	0.066	0.066	0.066	0.047	-0.153	-0.071
	Anxiety	-0.113	-0.113	-0.113	-0.114	0.088	0.086
	Burnout	0.009	0.009	0.009	0.004	0.089	0.143
	Huvikaotus	-0.022	-0.022	-0.022	-0.002	0.174	-0.184
	TaskAvoidance	-0.34	-0.34	-0.34	-0.329	0.037	0.095
	EasyTasks	0.098	0.098	0.098	0.121	0.052	0.201
	Prokkimine	-0.295	-0.295	-0.295	-0.314	0.104	0.041
	TõhusStrat	0.228	0.228	0.228	0.211	0.04	-0.189
MittelõhusStrat	0.081	0.081	0.081	0.071	0.101	-0.226	
SotsStrat	-0.029	-0.029	-0.029	-0.051	0.05	0.087	
EniToetus	0.137	0.137	0.137	0.111	-0.201	-0.123	

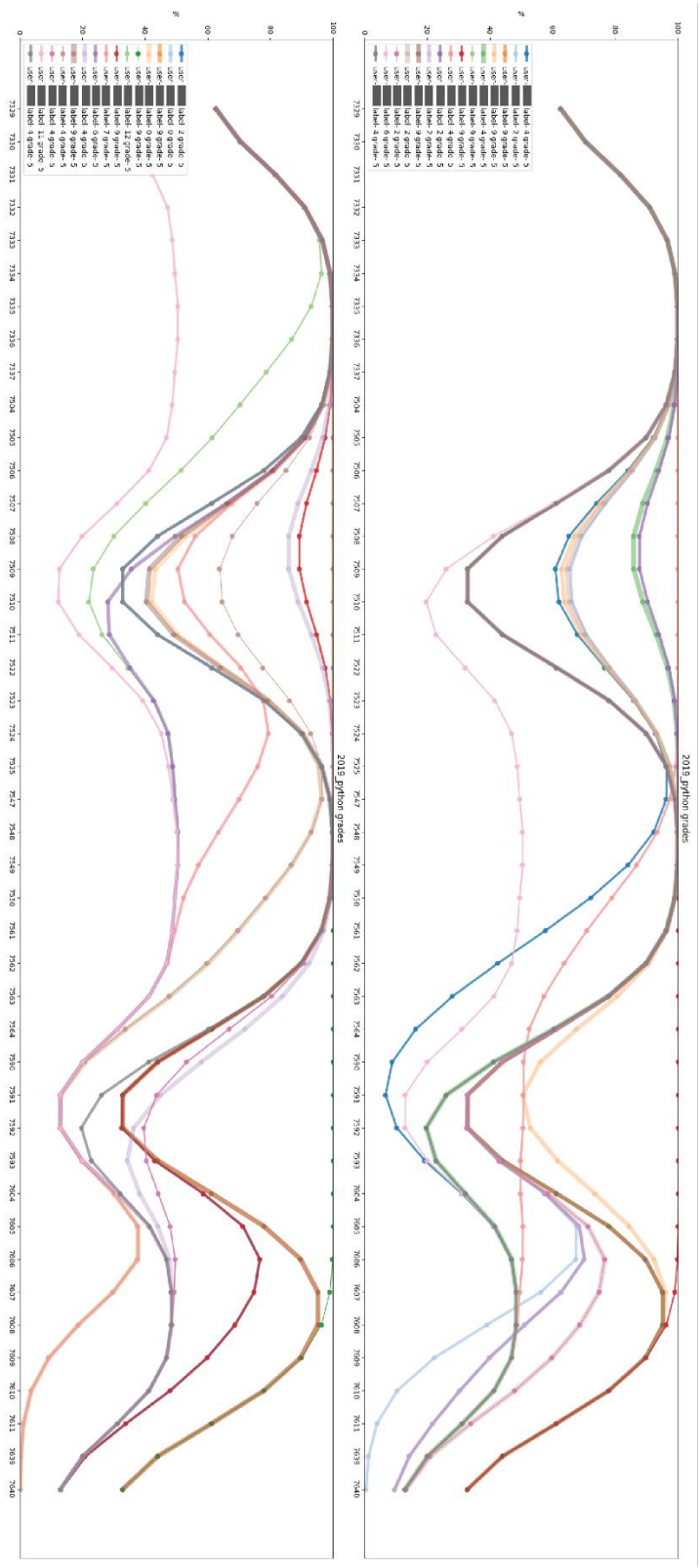
Lisa 11 – Projekt

Projekti kood asub TalTechi Gitlabi koodisalves¹. Koodisalvele pääseb ligi TalTechi kontoga.

Kaustas *student_analysis* asuvad neli kõige olulisemat projekti skripti. Esimeseks on *user_controller.py*, mille abil saab käivitada andmetöötlust. Teiseks skriptiks on *model_controller.py*, mida käivitades saab trennida masinõppemudeleid. Kolmas skript *user_profile_controller.py* käivitab tudengite klasterdamise. Viimaseks skriptiks on *correlation_controller.py*, mille abil saab leida Spearmani korrelatsioonikoefitsiente andmehulkade vahel. Projekti sisu täpsem kirjeldus on leitav README.md failist.

¹ <https://gitlab.cs.ttu.ee/heleot/student-predictions>

Lisa 12 – Näide programmeerimise algkursuse jooksvate tulemuste graafikust positiivsete lõpphinnete korral



Lisa 13 – Näide programmeerimise algkursuse jooksvate tulemuste graafikust negatiivsete lõpphinnete korral

