TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Kätrin Grauberg 192180IABM

Optimaalseima RFM mudeli leidmine otseturundustegevusteks telekommunikatsioonieettevõttes

Magistritöö

Juhendaja: Ants Torim
PhD

Tallinn 2022

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: Kätrin Grauberg

03.01.2022

Annotatsioon

Otseturundustegevuste jaoks on oluline kliente segmenteerida. Üheks segmenteerimise viisiks on RFM mudel.

Lõputöö eesmärgiks on parandada telekommunikatsioonіettevõtte otseturunduses klientide segmenteerimiseks loodud RFM mudelit. RFM mudel hindab kliente kolme tunnuse järgi: kui hiljuti on ost tehtud (*recency)*, kui tihti ostetakse (*frequency)* ja kui palju raha kulutatakse (*monetary value)*.

Eesmärgi täitmiseks on autor analüüsinud varasemate autorite töid ning rakendanud nendes kasutusel olevaid praktikaid. Praktilises osas on loodud 6 üksteisest veidi erinevat RFM andmemudelit, mida on võrreldud üksteise ja esialgse andmemudeliga.

Töö tulemusena eristus varasemast parem andmemudel, mis leidis kliendigrupid, kelle otseturunduslike pakkumiste määr on kõrgem. Mudeli olulisim täiendus oli ARPA (*Average Revenue Per Account*) ehk kliendi keskmise käibe tunnuse lisamine.

Loodud mudel on ettevõttes rakendatav.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 53 leheküljel, 4 peatükki, 22 joonist, 12 tabelit.

Abstract

Finding the most optimal RFM model for direct marketing activities in a telecommunications company

It is important to segment customers for direct marketing activities. One way to segment customers is to use RFM model.

The purpose of this master's thesis is to improve the RFM model created for the segmentation of customers for direct marketing of a telecommunications company. The RFM model evaluates customers based on three characteristics: how recently a purchase is made (recency), how often are purchases made (frequency) and how much money is spent (monetary value).

In order to achieve this goal, the author has analyzed the works of previous authors and applied the practices used in them. The process on creating these data models contains of: data preprocessing, adding ARPA (Average Revenue Per Account) component, replacing outlier values, applying Jenks Breaks algorithm to determine RFM scores, data normalization, data standardization, applying Tukey test to find distinct groups, applying k-means and k-means++ algorithm to find clusters.

In the practical part, 6 slightly different RFM data models have been created, which have been compared to each other and to the original data model.

As a result of the work, a better data model has been made, which found customer groups with a higher rate of accepting direct marketing offers. The most important addition to the model was the addition of the ARPA component.

The created model can be used in the company.

The thesis is in Estonian and contains 53 pages of text, 4 chapters, 22 figures, 12 tables.

Lühendite ja mõistete sõnastik

ARPA	*Average Revenue Per Account*, keskmine käive kliendi kohta
ARPU	*Average Revenue Per User,* keskmine käive kasutaja kohta
F	*Frequency*, ostude sagedus
M	*Monetary* value, ostude rahaline väärtus
R	*Recency*, ostu ajaline värskus
RFM	Turunduses klientide segmenteerimiseks kasutatav meetod. Lühend mõistetest *recency*, *frequency, monetary*.

Sisukord

Jooniste loetelu

Tabelite loetelu

1 Sissejuhatus

Magistritöö põhiülesandeks on leida RFM mudel, mille järgi kliendibaasi segmenteerides on võimalik leida kliendigrupid, kes on otseturundustegevuste kaudu saadud pakkumistele vastuvõtlikumad. Eesmärgi täitmiseks on analüüsitud varasemate autorite töid, rakendatud parimaid praktikaid, täiendatud ettevõttes varasemalt loodud RFM mudelit erinevatel viisidel ning võrreldud saadud tulemusi ajaloolisel otseturunduse tulemuslikkuse andmestikul.

1.1 Taust

Otseturundus on turunduse liik, kus sõnum on mõeldud otse konkreetsele kliendile. Sõnumi saatmist on võimalik teostada läbi erinevate kanalite, näiteks e-mail, sms, kõne, teavitused veebilehtedel, rakendustes või isiklikus kasutuses olevates seadmetes. Mida laiem on turundussõnumi publik, seda tõenäolisemalt väiksema efektiivsusega on sõnum. Laiemale sihtgrupile saadetud sõnum võib mõnele kliendile korda minna, kuid valimit kitsendamata võivad paljud kliendid näha seda hoopis segajana. Kõige efektiivsem turundus on võimalikult täpselt sihitud, seetõttu on oluline ka otsekommunikatsiooni tegevusi sihtida [1].

Üks otseturunduse väljakutseid on leida isikud, kes on konkreetsest turundussõnumist reaalselt huvitatud, kellest võiksid saada uued kliendid või kes sooviksid teha korduva ostu või võtta lisateenuseid.

Kliendid, kes on ettevõttele andnud loa turunduslike pakkumiste saamiseks, soovivad sõnumeid saada. Ettevõtte jaoks on väärtuslik, et kliendid oleksid turundusloa andnud, sest see annab võimaluse kahesuunaliseks suhtluseks [2]. Kui sõnum ei ole hoolikalt vastava kliendigrupi jaoks disainitud, võib antud tegevus hoopis suurendada klientide arvu, kes soovivad turunduslikest pakkumistest loobuda [3]. Kliendid eelistavad kommunikatsiooni, mis on seotud nende huvidega ja pakub asjakohast teavet [4]. Kui

kliendid saavad pakkumisi, mis neid ei huvita, siis nad ei ava ka teisi kirju, sest nende kogemus on näidanud, et selle ettevõtte saadetud sisu neid ei huvita.

Laialdane klientide sihtimine võib tuua ettevõttele ka otsest rahalist kahju, näiteks kui tegemist on kõnevalimiga, kus kõnesid teevad töötajad, kellele on nende töö ja aja eest vaja tasu maksta.

Selleks, et ettevõtte kliente paremini tundma õppida ja teha efektiivsemaid pakkumisi, võib kasutada klasterdamist [5]. Üheks võimalikuks klientide klasterdamise või segmenteerimise viisiks on RFM mudel. RFM on turundusanalüüsi tööriist, mida kasutatakse ettevõtte parimate klientide väljaselgitamiseks. Traditsiooniliselt hinnatakse kliente kolmes kategoorias - *recency, frequency, monetary value* [6]. Täpsemalt on RFM mudel kirjeldatud peatükis 2.2.

1.2 Probleem

Selleks, et turundustegevustes hoida klientide kõrget rahulolu, täita ärieesmärke ja ressursse parimal viisil kasutada, on vaja turunduslike pakkumiste jaoks teha andmetel põhinevaid otsuseid. Vajadus on kliente segmenteerida, et pakkumisi oleks võimalik sihtida vastavalt kliendigruppidele. Klientide segmenteerimiseks on ettevõttes arendatud esialgne RFM mudel, kuid on vaja testida ja aru saada, kas väljatöötatud mudel on piisavalt efektiivne.

Varasemalt teostatud tööde ja katsetatud metoodikate hulk, millele on läbivalt käesoleva töö jooksul viidatud, kinnitab, et klientide segmenteerimine on laiem probleem ka teistes ettevõtetes üle maailma.

1.3 Eesmärk

Lõputöö eesmärgiks on leida ettevõtte kliendibaasile kõige efektiivsem ja optimaalsem segmenteerimise viis, mis aitaks kaasa ärieesmärkidele. Olemasoleva RFM mudeli puhul on tarvis hinnata, kas kasutusel olevad tunnused ja tehnika on ettevõtte

ärispetsiifikast lähtuvalt kõige optimaalsemad. Parim RFM mudel on selline, mis leiab kliendibaasist kõige täpsemalt sellised grupid, mille otsekommunikatsiooni pakkumiste vastu võtmise määr on kõige kõrgem.

1.4 Struktuur

Käesolev töö on jaotatud neljaks peatükiks. Esimeses peatükis on kirjeldatud üldine taust, konkreetne probleem, töö eesmärk ja struktuur. Teises peatükis on toodud ülevaade teiste autorite varasematest töödest, mis haakuvad antud töö teemaga ning on kirjeldatud kogu töö läbiviimise metoodika ja protsess. Kolmandas peatükis on esitletud töö tulemused ning neljandas peatükis on välja toodud põhilised järeldused, millele järgneb töö kokkuvõte.

2 Metoodika

Käesolevas peatükis on esmalt välja toodud kasutatud tööriistad. Seejärel selgitatud RFM tähendust ja tutvustatud varasemate autorite panust RFM mudeli täiendamisel. Sellele järgneb andmete eeltöötlusprotsessi kirjeldav peatükk, peale mida on kirjeldatud töös kasutusel olevad klasterdamisalgoritmid. Viimasena on selgitatud Tukey testi.

2.1 Tööriistad

RFM mudeli loomiseks on kasutatud Pythoni programmeerimiskeelt. Pythoni jooksutamiseks on kasutusel Jupyteri notebook, mis töötab Anaconda Navigatori abil. Algandmestiku kokkupanemiseks on kasutatud SQL struktuurpäringukeelt.

2.2 RFM

RFM on turundusanalüüsi tööriist, mida kasutatakse ettevõtte jaoks kasumlikumate klientide väljaselgitamiseks.

RFM hindab kliente numbriliselt (kvantitatiivselt) kolmes kategoorias ning üldjuhul skaalal 1 kuni 5 (mida suurem number, seda parem tulemus). Kategooriad on järgmised:

- *Recency* - kui hiljuti on ost tehtud
- *Frequency* - kui tihti ostetakse
- *Monetary value* - kui suur on ostu rahaline väärtus

Recency - Mida värskemalt on ost sooritatud, seda tõenäolisemalt on kliendil järgmiste ostude puhul sama pakkuja mõttes. Võrreldes klientidega, kes pole ettevõttest ostnud pikemat aega, on hiljutiste klientidega tulevastes tehingutes osalemise tõenäosus suurem. Et mitte kaotada kliente, kellel on viimasest ostust rohkem aega möödunud, on võimalik teha meeldetuletavat turundust.

Frequency - Selle tunnuse ennustamine võib aidata turundustegevustel, mille eesmärk on meelde tuletada klienti uuesti ettevõtet külastama.

Monetary - Rahaline väärtus tuleneb sellest, kui palju klient kulutab. Loomulik kalduvus on rõhuda sellele, et julgustada kliente, kes kulutavad kõige rohkem raha, seda jätkama. Ehkki see võib turundusse ja klienditeenindusse tehtud investeeringutelt paremat tulu toota, on oht ka võõrandada kliente, kes on olnud järjepidevad, kuid ei pruugi iga tehingu jaoks nii palju kulutada.

Vaatamata RFM-i analüüsi käigus kogutud kasulikule teabele, peavad ettevõtted arvestama sellega, et isegi parimad kliendid ei soovi üleliigset nõudmist ning madalama asetusega kliente võidakse kasvatada täiendavate turundustegevustega [6].

RFMi kolme indikaatori olulisus ettevõtte tegevusvaldkonniti erineb, mistõttu võib olla oluline leida indikaatoritele kaalud või lisada mudelisse täiendavaid muutujaid [7].

2.2.1 RFMi täiendamine muutujatega

Viimastel aastatel on RFMi mudelit palju erinevates valdkondades rakendatud ja toodud välja uusi viise selle parandamiseks või täiendamiseks. Mitmed autorid on täiendanud RFM mudelit selliselt, et lisanud sellele täiendavaid kompenente või rakendanud andmekaeve meetodeid [8].

Näiteks RFMC, kus on lisaks arvestatud klientide ostude kategooriaid (C - *category*). Mudeli autor on välja toonud, et kliendid, kes muidu RFMi mudeliga oleksid sarnaste profiilidega, võivad tegelikult olla täiesti erinevate ostuharjumustega. RFMC plussina on välja toodud, et see mudel aitab klientidele teha järgmise parima pakkumise (NBO - *next best offer*) [8].

LRFM on täiendatud mudel, kus L (*length*) tähistab päevade arvu, alates esimesest kontaktist kuni viimase kontaktini [9] või teisiti öelduna kliendisuhte pikkust telekommunikatsioonieettevõttes [5].

Kuna mõnes valdkonnas on perioodilisus oluline mõõdik, siis on kasutatud LRFMP mudelit, kus L tähistab kliendisuhte pikkust ja P perioodilisust tähendusega keskmine päevade arv, mis on jäänud kahe ostu vahele [10].

Lisaks on loodud varasemalt ka RFMTC mudel, mida on täiendatud kahe muutujaga T (*time since first purchase*) - esimesest ostust möödunud aeg ja C (*churn probability*) - churni tõenäosus. [11]

Kaubanduses on kasutatud RFMOC mudelit, kus O (*offer*) tähistab, kui tihti klient ostab tooteid kampaaniaperioodil ja C (*category variance*), kui palju varieeruvust on kliendi ostudes tootekategooriate lõikes. Käsitletud olukorras töötas RFMOC mudel paremini kui klassikaline RFM mudel [12].

On ka välja jäetud muutujaid, kui vastavalt valdkonnale ei anna see väärtust juurde. Näiteks jäeti välja ühes hambaravikliinikus klientide M *monetary value*, sest suur osa kasumist tuleb riiklikust ravikindlustusest [9]. RFMi rakendamisel on oluline jälgida tegevusvaldkonda ning vastavalt sellele teha otsuseid, milliseid komponente lisada või välja jätta.

Varasemalt on telekommunikatsiooni valdkonnas kasutatud otsese *monetary* väärtuse kui kaubamüügi tulemuste asemel ARPU väärtust (*Average Revenue Per User*), sest see on põhiline teenuseoperaatori edu mõõdik. [13] Käesolevas töös asendatakse monetary väärtus kliendi 3 kuu kesmise ARPAga (*Average Revenue Per Account*).

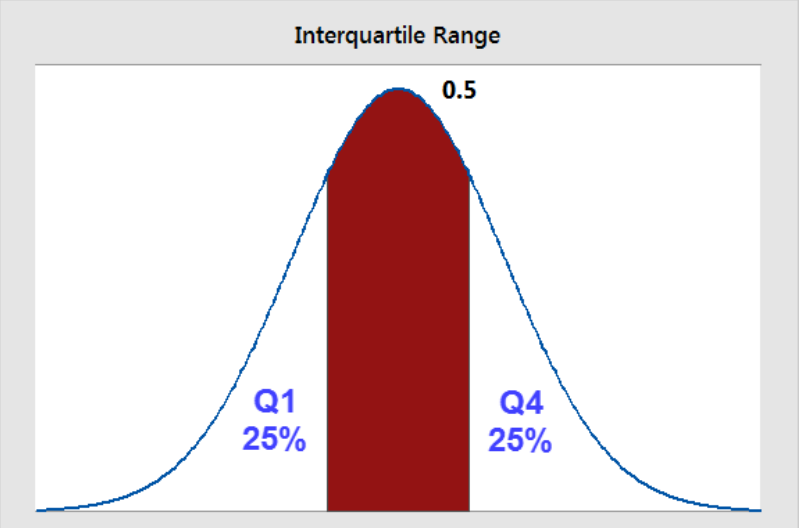
2.3 Andmete eeltöötlus

Kliendile segmendi määramiseks kasutatakse eelneva 1 aasta ja 5 kuu toodete ning teenuste andmeid ja eelneva 3 kuu keskmise ARPA andmeid. Iga kuu arvutatakse kliendile segment uuesti külge, mis tähendab mõne teise kuu puhul ka vastavalt teise perioodi ajaloolisi andmeid.

Andmete eeltöötluse käigus on eemaldatud duplikaadid ning kuna oli kliente, kellel ei ole ARPA arvutatud, siis nendel klientidel asendati ARPA väärtus 0-ga.

2.3.1 Erindid

Andmete väärtustest on asendatud erindid. Erindid on andmestiku ebaharilikud väärtused ja võivad moonutada tulemusi, mistõttu on need oluline eelnevalt eemaldada [14]. Kvartiilide vahe (*IQR Interquartile range*) on ülemise ja alumise kvartiili erinevus [15] ning seda kasutatakse varieeruvuse iseloomustamiseks [16]. Kõrgemad IQR väärtused näitavad, et andmete keskmik ulatub kaugemale ja väiksemad väärtused näitavad, et keskmised väärtused koonduvad tihedamalt. [17]



Joonis 1. Kvartiilide vahe [17].

Kõrgeima otsa erindid on leitud valemiga

$a = (Q3 + 1.5 \cdot IQR)$,

kus Q3 on kolmas kvartiil ja IQR on kvartiilide vahe.

Kui väärtus on suurem kui a, siis asendatakse see a väärtusega.

Madalaima otsa erindid on leitud valemiga

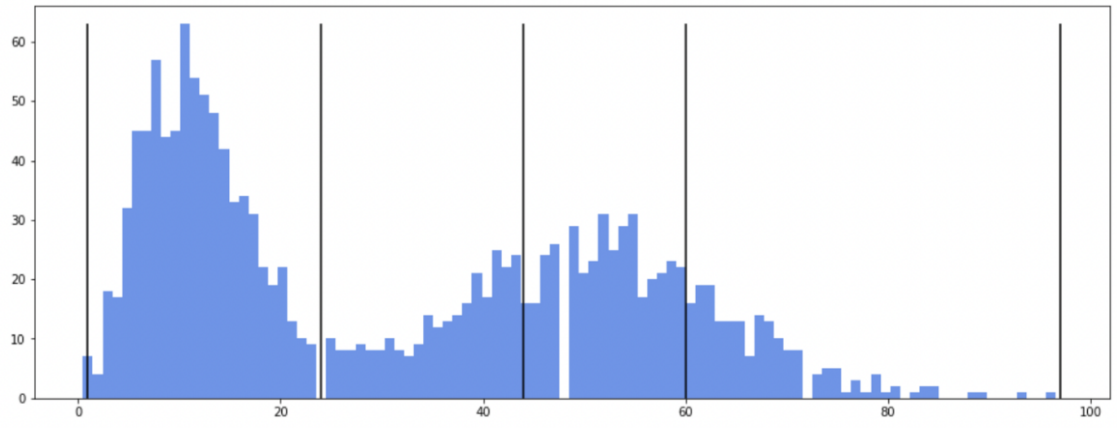
$b = (Q3 - 1.5 \cdot IQR)$.

Kui väärtus on väiksem kui b, siis asendatakse see b väärtusega.

2.3.2 Jenks Breaks ja skooride määramine

Jenks Breaks algoritm (Jenksi optimeerimise meetod või Jenks Natural Breaks optimeerimise meetod) on klassifitseerimismeetod, kus klassid leitakse vastavalt andmete loomulikele gruppidele. Klassid on jaotatud selliselt, et sarnased väärtused on kokku rühmitatud ja klasside vahelised erinevused maksimeeritud. Klassidevahelised piirid on kohtades, kus andmeväärtustes on suured erinevused [18]. Sarnaselt k-means algoritmile tuleb Jenks Breaks puhul klastrite arv esmalt ette öelda [19].

Joonisel on illustreeritud Jenks Breaksi klasterdamise tulemus, kus sinised tulbad tähistavad väärtuste esinemist ning horisontaalsed jooned gruppide piire. Joonisel ei ole kujutatud selles töös kasutusel olev andmestik ja sellel on vaid selgitav ja illustreeriv eesmärk.



Joonis 2. Jenks Breaks klasterdamine [20].

Jenks Breaks algoritmi selgitamiseks on siia töösse toodud teise autori loodud näide [19]**:**

1. Esmalt arvutatakse listi väärtuste keskmise hälvete ruudu summa (SDAM *sum of squared deviations for array mean*)

list = [4, 5, 9, 10]

mean = (4 + 5 + 9 + 10) / 4 = 7

SDAM = (4-7)^2 + (5-7)^2 + (9-7)^2 + (10-7)^2 = 9 + 4 + 4 + 9 = 26

2. Igale vahemiku kombinatsioonile arvutatakse klassi keskmise hälvete ruudu summa (SDCM_ALL)

For [4][5,9,10]

SDCM_ALL = (4-4)^2+(5-8)^2+(9-8)^2+(10-8)^2 = 0 + 9 + 1 + 4 = 14

For [4,5][9,10]

SDCM_ALL = (4-4.5)^2+(5-4.5)^2+(9-9.5)^2+(10-9.5)^2 = 0.25 + 0.25 + 0.25 + 0.25 = 1.

For [4,5,9][10]

SDCM_ALL = (4-6)^2+(5-6)^2+(9-6)^2+(10-10)^2 = 4 + 1 + 9 + 0 = 14.

3. Arvutatakse *goodness of variance fit* GVF, mis on defineeritud kui (SDAM - SCDM) / SDAM. 1 tähistab ideaalset sobivust ja 0 kõige halvemat.

GVF for [4,5][9,10] is (26 - 1) / 26 = 25 / 26 = 0.96

GVF for the other 2 ranges is (26 - 14) / 26 = 12 / 26 = 0.46

GVF [4,5][9,10] puhul on kõige kõrgem, mis tähistab, et see kombinatsioon on parim viis listi väärtustega [4, 5, 9, 10] jaotamiseks [19].

Jenks Breaks algoritmi kasutatakse töös *recency*, *frequency* ja ARPA väärtuste vahemike jagamiseks, et nendele vahemikele määrata RFM skoorid. Ettevõttes on esialgse RFM mudeli puhul otsustatud, et RFM skooride vahemik peaks olema 1-4. Kuna RFM puhul traditsiooniliselt kasutatakse vahemikku 1-5, siis töö üheks eesmärgiks on välja selgitada, kas vahemiku 1-4 valimine on põhjendatud või sobiks traditsiooniline lähenemine paremini.

Mida “parem” väärtus, seda kõrgem RFM skoor määratakse. *Recency* väärtusteks on päevade arv viimasest ostust, seega mida suurem on väärtus, seda “halvem” on tulemus ehk RFM skooriks saab 1. Mida kõrgem on ARPA, seda “parem” on tulemus ehk RFM skooriks saab kõrgemate väärtuste korral 4 või 5. RFM puhul määratakse igale kliendile 3 skoori. Näiteks klient, kes on ostu teinud hiljuti (*Recency* 5), teeb oste harva (*Frequency* 1) ja on kulutanud keskmiselt raha (*Monetary* 3), saab vastavalt külge RFM grupi 513. Sarnaseid kombinatsioone tekib palju olenevalt atribuutide ja atribuutide väärtuste vahemike arvust.

Alternatiiv Jenks Breaks algoritmi kasutamisele vahemike leidmiseks oleks kvartiilide kasutamine. Kuigi nende meetodite kasutamisel tulemused võivad erineda, siis antud töö skoopi ei jää võrdlus, kumb lahendus oleks parem. Jenks Breaks on valitud seetõttu, et ettevõttes on esialgne mudel selle algoritmiga loodud. Varasemalt on tehtud töö, kus on võrreldud RFM puhul Jenks Breaks ja kvartiilide rakendamist ning selles töös on väidetud, et valiku tegemine sõltub olukorrast [20].

2.3.3 Normaliseerimine

Normaliseerimine on tehnika, mida kasutatakse masinõppe rakendamisel andmete eeltöötlusel [21]. Selle andmetöötluse sammu eesmärgiks on, et andmestiku väärtused oleksid samal skaalal, kus miinimumväärtus on 0 ja maksimumväärtus on 1.

Normaliseerimiseks on kasutatud Pythoni *sklearn.preprocessing* moodulist *MinMaxScaler* funktsiooni [22], mis teisendab andmed järgmise valemi järgi:

Normaliseeritava veeru x ja rea i elemendi uus väärtus x^i_{norm} :

$$x^{i}_{norm} = \frac{xi - xmin}{xmax - xmin}$$

x^i : vana väärtus;

x_{max}, x_{min} : veeru x maksimaalne ja minimaalne väärtus [23].

Töös on testitud erinevaid mudeleid ning katsetatud, kas andmete normaliseerimine annab paremaid lõpptulemusi. Seetõttu on *recency*, *frequency* ja ARPA väärtused normaliseeritud, mida on teinud ka teised autorid [12][24].

2.3.4 Standardiseerimine

Standardiseerimise eesmärk on viia andmestiku väärtused sellisele kujule, kus keskväärtus on 0 ja standardhälve 1.

Standardiseerimiseks on kasutatud Pythoni *sklearn.preprocessing* moodulist *StandardScaler* funktsiooni [25], mis teisendab andmed järgmise valemi järgi:
Standardiseeritava veeru x ja rea i elemendi uus väärtus *xistd* :
$x_{std}^{\ i} = \frac{xi-\mu x}{\sigma x}$ [23].

Töös on testitud, kas andmete standardiseerimine annab paremaid lõpptulemusi.

2.4 Klasterdamise algoritmid

Käesolevas töös on rakendatud kliendigruppide leidmiseks kahte klasterdamise algoritmi - k-means ja k-means++.

2.4.1 K-means

Kõige populaarsem klasterdamise algoritm on k-keskmiste klasterdamine (*k-means*) [26] [27], mida on ka RFMi rakendmise puhul mitmel juhul kasutatud [28] [29] [30] [31] [32]. K-keskmiste klasterdamise algoritmi on kasutatud erinevates uurimisvaldkondades, sest see on kiire ja lihtne rakendada [27].

K-keskmiste klasterdamise puhul on tegemist juhendamata masinõppe algoritmiga [26], kus k väärtus ehk klastrite arv tuleb eelnevalt ette öelda. Esmalt genereeritakse

juhuslikud klastrite keskpunktid. Järgmisena arvutatakse vahemaa iga elemendi ja keskpunkti vahel ning iga element määratakse kõige lähema keskpunkti juurde [27].

Peale klastrite moodustamist arvutatakse ümber iga klastri keskmine väärtus klastri praeguste elementide põhjal. Protsessi korratakse, kuni keskpunktid ei muutu. Kuna k-means algoritmi sooritus sõltub k väärtusest, tuleb meetodit katsetada erinevate k väärtustega, et leida optimaalseim [27].

K-keskmiste klasterdamises "keskmised" (*means*) viitabki keskpunkti leidmisele [26].

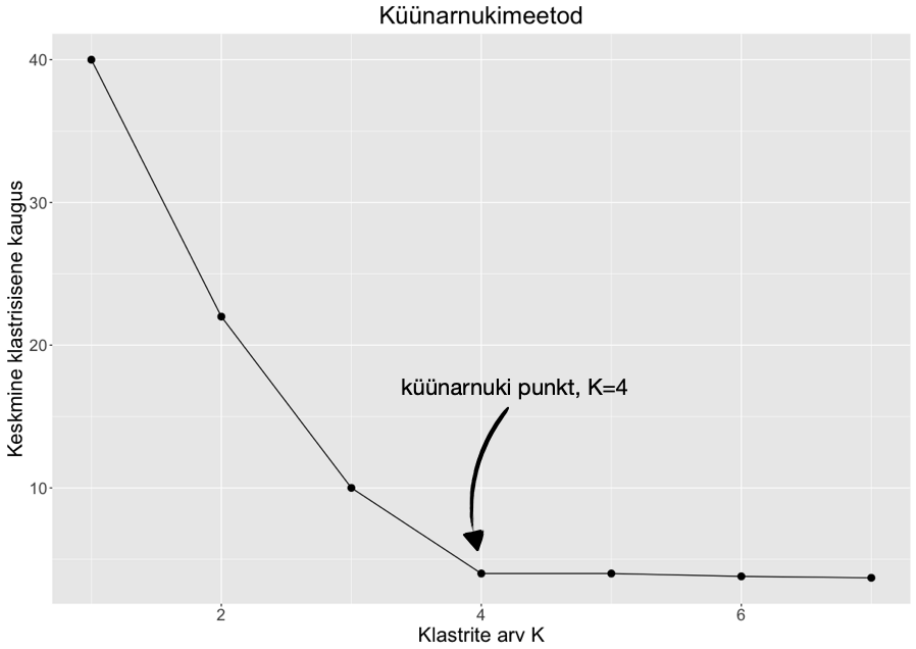
2.4.2 K-means++

Klasterdamisel on kasutatud ka k-means++ algoritmi [33] [34]. K-means++ on edasiarendus traditisioonilisest k-means klasterdamise algoritmist. Selle algoritmi puhul on optimiseeritud algse klastri keskpunkti valiku meetodit, mille tulemusena elimineeritakse algse keskpunkti valiku mõju klastri tulemusele ja saavutatakse parem klasterdamise efekt [33].

2.4.3 Küünarnuki meetod

Juhendamata masinõppe algoritmi puhul on oluline kindlaks määrata optimaalne klastrite arv, kuhu saab andmeid rühmitada. Üks populaarsemaid meetodeid optimaalseima k väärtuse leidmiseks on küünarnuki meetod (*Elbow method*) [35].

Parim k väärtus on võimalik leida graafikult, kus keskmise klastrisisese kauguse vähenemise kiirus järsult aeglustub. Seda punkti nimetatakse küünarnuki punktis ning peetakse üldjuhul optimaalsemaks k väärtuseks. Näiteks on Joonis 3 optimaalseim klastrite arv 4 ehk küünarnuki punkt k=4 [36].



Joonis 3. Küünarnuki meetod [36].

2.5 Tukey test

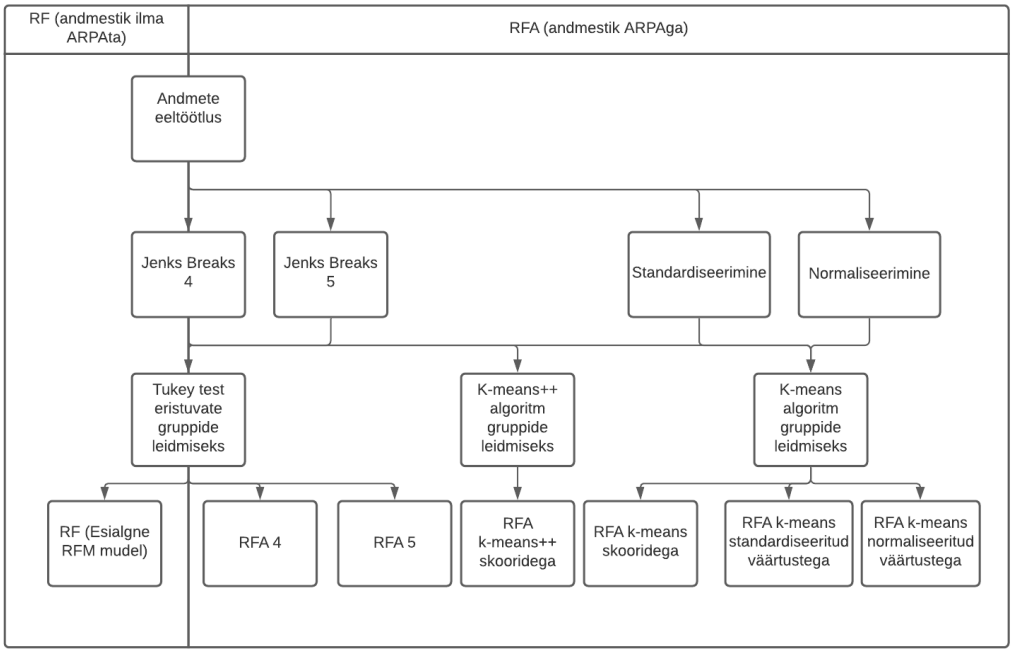
Tukey test on statistiline tööriist, mida kasutatakse, et välja selgitada, kas erinevus kahe andmehulga vahel on statistiliselt oluline. Tukey test võrdleb kõiki võimalikke paaride keskmisi [37]. Seda on kasutatud varem ka RFMi rakendamisel [38].

Tukey test võrdleb väärtuste keskmiste erinevusi mitte väärtuspaare. Tukey testi väärtus saadakse, võttes keskmiste paaride vahe absoluutväärtuse ja jagades selle keskmise standardveaga (SE), mis on määratud ühesuunalise ANOVA testiga. SE on omakorda ruutjuur (variatsioon jagatud valimi suurusega). Tukey test on post hoc test, kuna muutujaid võrreldakse pärast andmete kogumist [37].

Tukey test oli varasemalt ettevõttes loodud esialgses RFM andmemudelis kasutatud.

3 Peamised tulemused

Töö käigus on loodud lisaks ettevõttes esialgselt arendatud RFM mudelile lisaks 6 erinevat täiendatud RFM mudelit, mille loomisprotsessi kirjeldab Joonis 4. Täpsemalt on protsessid vastavate mudelite alapeatükkides kirjeldatud.



Joonis 4. RFM mudelite loomise protsess.

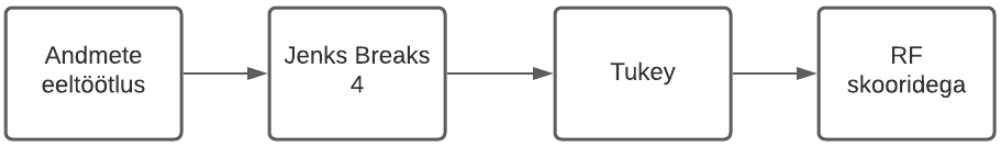
Eelnevalt kirjeldatud IQR valemite abil on leitud andmestikus *Frequency* ja ARPA väärtusete hulgast erindid ja need asendatud. Kliendid, kes teevad väga palju oste, näiteks kampaaniaperioodil, võivad omada väärtuseid, mis lähevad antud juhul erindi alla. Selline käitumine ei iseloomusta kliendibaasi üldiselt. Käesolevas töös kirjeldatud mudelid on kõik loodud samasuguse andmestiku peale.

Kõikide mudelite testimiseks ja valideerimiseks on kasutatud sama andmestikku, mis sisaldab 3 kuu otsekommunikatsiooni tulemuste andmeid. Testimiseks on klientidele iga mudeli abil külge saadud konkreetne grupp. Igale grupile on arvutatud keskmine

otsekommunikatsiooni pakkumiste vastuvõtlikkuse määr, mida on võrreldud kogu valimi keskmise pakkumiste vastuvõtlikkuse määraga.

3.1 Esialgne RFM mudel ehk RF

Esialgne RFM mudel (edaspidi RF), mis on eelnevalt ettevõtte andmeteadlaste poolt loodud, sisaldab kahte väärtust *recency* ja *frequency,* kuna varasemalt oli ettevõttes otsustatud, et *monetary*, mis peaks olema ostude andmed, ei anna ettevõtte tegevusvaldkonna tõttu piisavalt kasulikku informatsiooni. Joonis 5 illustreerib RF mudeli loomise protsessi.

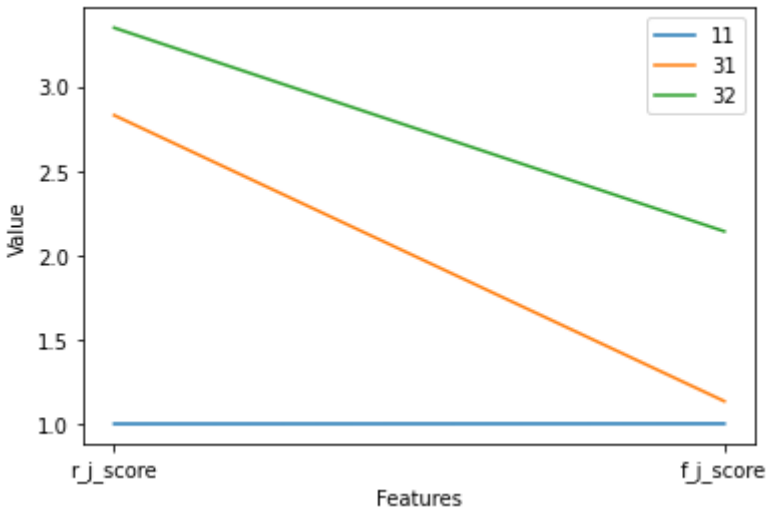


Joonis 5. RF mudeli loomise protsess.

Andmete eeltöötluse järgselt on rakendatud Jenks Breaks algoritmi. Algoritm jagas *recency* ja *frequency* väärtused 4 vahemikku, mille järgi on igale kliendile määratud nii *recency* kui *frequency* tunnusele skoorid vahemikus 1-4. Täpsemalt on skooride määramine kirjeldatud peatükis 2.3.2. Erinevate skooride kombinatsioone tekkis 16. Kuna 16 gruppi on liiga palju, siis on leitud Tukey testi järgi grupid, mille erinevus on statistiliselt oluline. Tukey test on kirjeldatud peatükis 2.7.

Peale Tukey testi jäi järgi 3 eristuvat gruppi, kus kahel grupil on vastuvõtlikkuse määr kogu valimi keskmisest väiksem ning ühel grupil keskmisest vastuvõtlikkuse määrast 9% kõrgem.

Joonis 6 illustreerib RF mudeli gruppide tunnuste väärtuseid. Gruppide nimes tähistab esimene number *recency* skoori ja teine number *monetary* skoori (nt kui grupp on 11, siis *recency* 1 ja *frequency* 1).



Joonis 6. RF mudeli grupid ja tunnuste väärtused

Tabelis 1 on kirjeldatud RF mudeli gruppide vastuvõtlikkuse määrad.

Tabel 1. RF mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
11	keskmisest 40% madalam
31	keskmisest 8% madalam
32	keskmisest 9% kõrgem

Grupp 11 on kliendid, kelle viimasest ostust on kaua aega möödas ning kes teevad oste harva. Selle kliendigrupi pakkumiste vastuvõtlikkuse määr on keskmisest 40% madalam.

Grupp 31 on kliendid, kes on üsna hiljuti ostu teinud, kuid teevad sarnaselt grupile 11 oste harva. Selle kliendigrupi pakkumiste vastuvõtlikkuse määr on keskmisest 8% madalam.

Grupp 32 on selle mudeli järgi kõige kõrgemate R ja F väärtustega. Need on kliendid, kes on väga värskelt ostu teinud ning on oste sooritanud ka varem teatud ajavahemiku

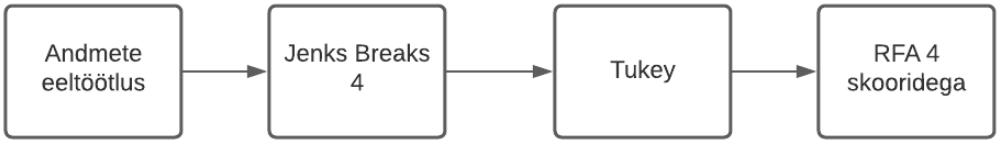
tagant. Selle kliendigrupi vastuvõtlikkuse määr on keskmisest 9% kõrgem, mis on ühtlasi selle mudeli kõige kõrgema vastuvõtlikkuse määraga grupp.

3.2 RFM mudel ARPAga ehk RFA

Esialgse RFM mudeli täiendamisel on lisatud uus tunnus ARPA (*average revenue per account*), mis võiks asendada RFMis *monetary* väärtust. Andmestikule, mis on võtnud arvesse ARPA tunnust, on loodud 6 veidi erinevat andmemudelit.

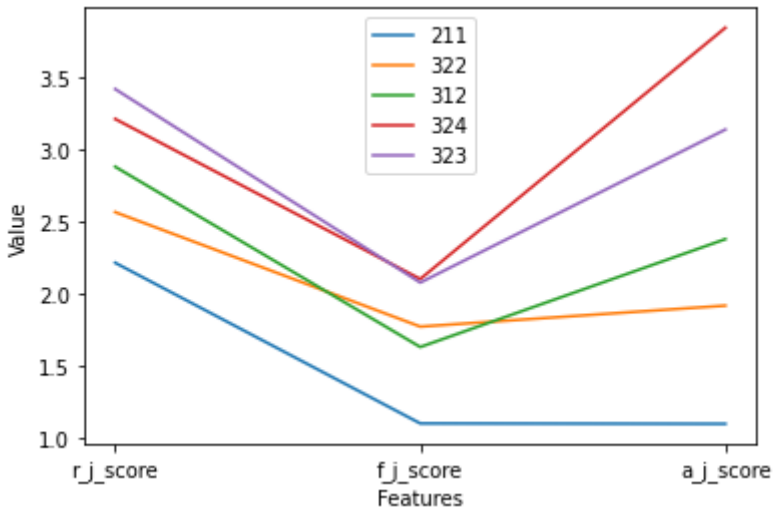
3.2.1 RFA skoorid 1-4 ehk RFA 4

RFA skooridega mudel on loodud sarnasel põhimõttel nagu RF mudel. Joonis 7 illustreerib RFA 4 mudeli loomise protsessi.



Joonis 7. RFA 4 mudeli loomise protsess.

Andmete eeltöötluse järgselt on rakendatud Jenks Breaks algoritmi, mille järgselt tekkis 61 erinevat gruppi. Peale Tukey testi jäi järgi 5 eristuvat gruppi. Kõigil gruppidel tähistab esimene number *recency* (R), teine *frequency* (F) ja kolmas ARPA (A) skoori. Joonis 8 illustreerib RFA 4 mudeli gruppe ja tunnuste väärtuseid.



Joonis 8. RFA 4 mudeli grupid ja tunnuste väärtused

Tabelis 3 on RFA 4 mudeli gruppide vastuvõtlikkuse määrad.

Tabel 2. RFA 4 mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
211	keskmisest 46% madalam
322	keskmisest 14% madalam
312	keskmisest 7% kõrgem
324	keskmisest 39% kõrgem
323	keskmisest 24% kõrgem

Grupp 211 on kliendid, kelle viimasest ostust on juba pikem aeg möödas. Nende ostude sagedus ja ARPA on madal. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 46% madalam.

Grupp 322 on kliendid, kelle viimasest ostust on ka üsna palju aega möödas, kuid nad on kunagi varem ka ostu teinud ja nende ARPA on võrreldes teiste gruppidega madal. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 14% madalam.

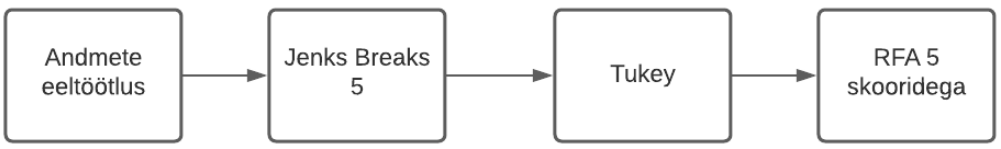
Grupp 312 on sarnane grupile 322. 312 kliendid on ostu teinud värskemalt ja nende ARPA on suurem, kui 322 grupi omad, kuid nende klientide ostude sagedus on madalam. Kuigi grupp 312 ja grupp 322 on sarnased, siis need kliendid on pakkumistele erinevalt reageerinud. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 7% kõrgem.

Grupp 324 on viimase ostu teinud üsna värskelt ning nad sooritavad kordusoste. Nende klientide ARPA on kõige kõrgem. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 39% kõrgem, mis on ühtlasi selle mudeli kõige kõrgema vastuvõtlikkuse määr.

Grupp 323 on sarnane grupile 324. Need kliendid on üsna värskelt teinud viimase ostu ning nad sooritavad kordusoste sarnaselt grupile 324, kuid nende klientide ARPA on väiksem. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 24% kõrgem.

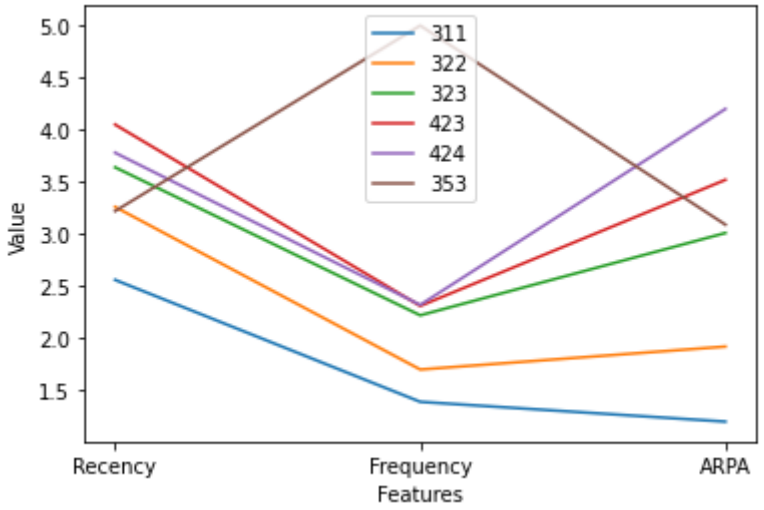
3.2.2 RFA skoorid 1-5 ehk RFA 5

Kuna RFM puhul traditsiooniliselt kasutatakse skooride vahemikku 1-5 ja esialgselmudelil on skoorid vahemikus 1-4, siis on loodud võrdlemiseks mudel ka traditsiooniliste vahemikega. RFA 5 skooridega mudel on loodud sarnasel põhimõttel nagu RF ja RFA 4 mudel, kuid Jenks Breaks algoritm jagas R, F, A väärtused 5 vahemikku, mille järgi on igale kliendile määratud tunnusele skoorid vahemikus 1-5. Joonis 9 illustreerib RFA 5 mudeli loomise protsessi.



Joonis 9. RFA 5 mudeli loomise protsess.

Joonis 10 illustreerib RFA 5 mudeli gruppe ja tunnuste väärtuseid. Kõigil gruppidel tähistab esimene number recency (R), teine frequency (F) ja kolmas ARPA (A) skoori.



Joonis 10. RFA 5 mudeli grupid ja tunnuste väärtused

Tabelis 3 on RFA 5 mudeli gruppide vastuvõtlikkuse määrad.

Tabel 3. RFA 5 mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
311	keskmisest 59% madalam
322	keskmisest 31% madalam
323	keskmisest 9% madalam
423	keskmisest 13% kõrgem
424	keskmisest 53% kõrgem
353	keskmisest 290% kõrgem

Grupp 311 on kliendid, kes on varem küll oste teinud, aga viimasest ostust on võrreldes teiste gruppidega kõige rohkem aega möödas. Nende ostude sagedus ja ARPA on väga madal. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 59% madalam.

Grupp 322 on kliendid, kes on veidi väärtuslikumad kui 311, kuid kelle viimasest ostust on ka üsna kaua aega möödas. Need kliendid on kunagi varem ka oste teinud ja nende ARPA on võrreldes teiste gruppidega madal. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 31% madalam.

Grupp 323 on kliendid, kes on viimase ostu teinud keskmiselt värskemalt, ostude sagedus ja ARPA on ka keskmine. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 9% madalam.

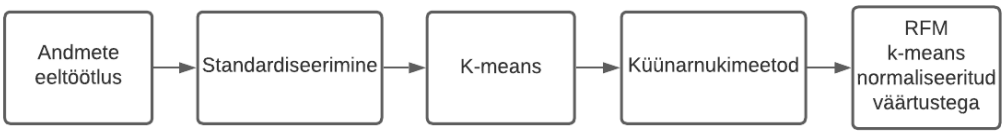
Grupp 423 on kliendid, kes on kõige värksemalt teinud viimase ostu. Võrreldes teiste gruppidega teevad nad oste sagedasti ja ARPA on kõrge. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 13% kõrgem.

Grupp 424 on sarnane grupile 423. Grupp 424 viimasest ostust on veidi kauem aega möödas, kui grupp 423 klientidest, kuid nende ARPA on kõrgem. Ostude sagedus on keskmisest kõrgem ja sama, mis grupil 423. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 53% kõrgem.

Grupp 353 on kliendid, kes teevad väga sagedasti oste. Nende viimasest ostust on möödas keskmine aeg võrreldes teiste gruppidega ja ARPA on samuti keskmine. Kuigi selle grupi pakkumiste vastuvõtlikkuse määr on keskmisest 290% kõrgem, ei ole see grupp piisavalt suur et annaks otseturundusele olulist väärtust.

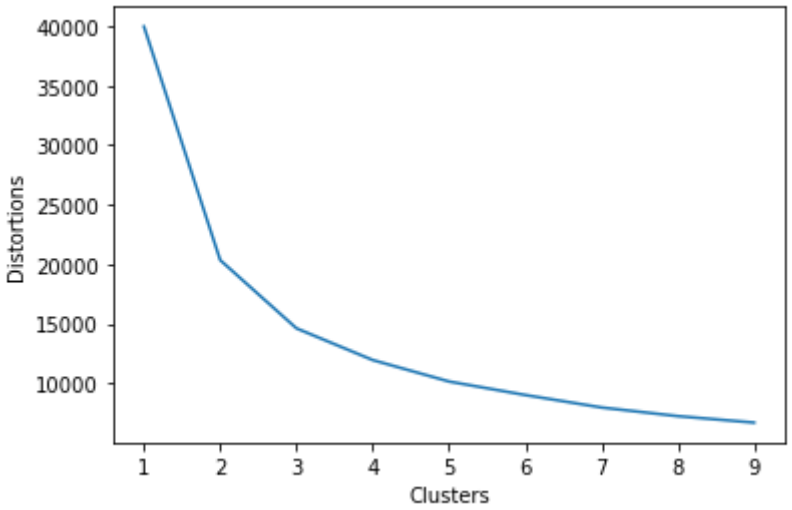
3.2.3 K-means standardiseeritud RFA väärtustega

Kuna RFMi rakendamisel on teiste autorite poolt kasutatud erinevaid klasterdamise algoritme, siis rakendati selle töö puhul kahte populaarsemat, milleks üks oli k-means. Andmete eeltöötluse järgselt on R, F ja A väärtused standardiseeritud ja rakendatud k-means klasterdamisalgoritmi. Joonis 11 illustreerib RFA standardiseeritud k-means mudeli loomise protsessi.



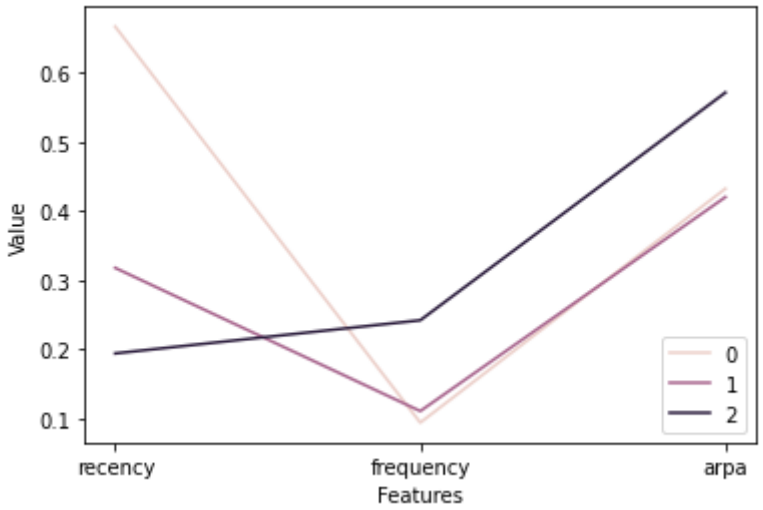
Joonis 11. RFA standardiseeritud k-means mudeli loomise protsess.

Optimaalseima k väärtus on leitud küünarnuki meetodil. Jooniselt 12 paistab, et optimaalne k väärtus ehk klastrite arv on 4.



Joonis 12. RFA standardiseeritud k-means küünarnuki meetod.

Gruppide nimed 0, 1, 2 on juhuslikud ja ei oma tähendust. Joonis 13 illustreerib RFA standardiseeritud k-means mudeli gruppe ja tunnuste väärtuseid.



Joonis 13. RFA standardiseeritud k-means mudeli grupid ja tunnuste väärtused

Tabelis 4 on RFA standardiseeritud k-means mudeli gruppide vastuvõtlikkuse määrad.

Tabel 4. RFA standardiseeritud k-means mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
0	keskmisest 9% kõrgem
1	keskmine
2	keskmisest 9% madalam

Grupp 0 on kliendid, kes on värskelt teinud ostu, kuid nende ostude sagedus on madal. Nende ARPA on keskmine. Selle grupi vastuvõtlikkuse määr on keskmisest 9% kõrgem.

Grupp 1 on kliendid, kellel on viimasest ostust juba mõnda aega möödas. Nad teevad oste väga harva ja nende ARPA on keskmine. Selle grupi vastuvõtlikkuse määr on samuti keskmine.

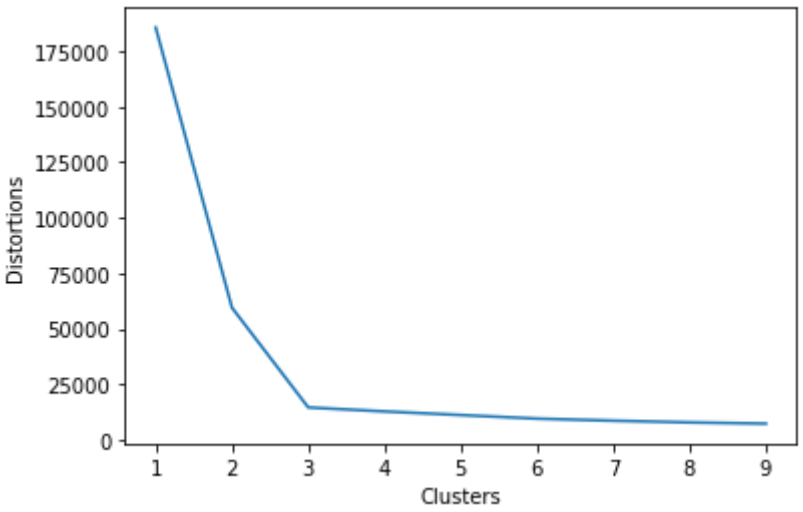
Grupp 2 on kliendid, kellel on viimasest ostust kõige kauem aega möödas, kuid nad teevad keskmisest tihedamini oste ning nende ARPA on kõrge. Selle grupi vastuvõtlikkuse määr on keskmisest 9% madalam.

3.2.4 K-means normaliseeritud RFA väärtustega

RFA normaliseeritud k-means mudel on loodud sarnaselt nagu eelmises peatükis kirjeldatud k-means standardiseeritud RFA väärtustega mudel, kuid selle mudeli puhul on andmete eeltöötluse järgselt on R, F ja A väärtused normaliseeritud ja rakendatud k-means klasterdamisalgoritmi. Joonis 14 illustreerib RFA normaliseeritud k-means mudeli loomise protsessi.

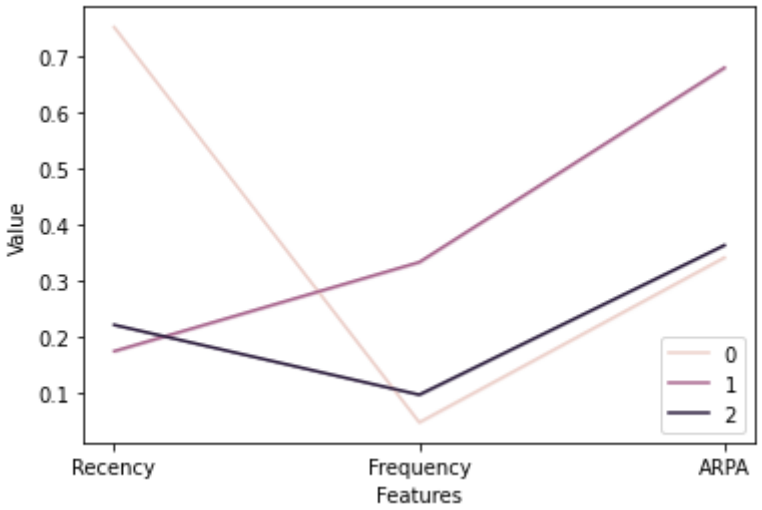
Joonis 14. RFA normaliseeritud k-means mudeli loomise protsess.

Optimaalseima k väärtus on leitud küünarnuki meetodil. Jooniselt 15 paistab, et optimalne k väärtus ehk klastrite arv on 3.



Joonis 15. RFA normaliseeritud k-means küünarnuki meetod.

Joonis 16 illustreerib RFA normaliseeritud k-means mudeli gruppe ja tunnuste väärtuseid. Gruppide nimed 0, 1, 2 on juhuslikud ja ei oma tähendust.



Joonis 16. RFA normaliseeritud k-means mudeli grupid ja tunnuste väärtused

Tabelis 5 on RFA normaliseeritud k-means mudeli gruppide vastuvõtlikkuse määrad.

Tabel 5. RFA normaliseeritud k-means mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
0	keskmisest 9% madalam
1	keskmisest 14% madalam
2	keskmisest 23% kõrgem

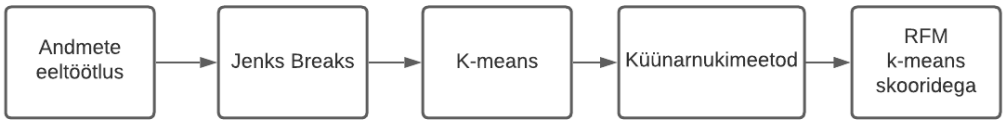
Grupp 0 on kliendid, kes on väga värskelt teinud oste. Nende ostude sagedus on väga madal ning ARPA on keskmine. Selle grupi vastuvõtlikkuse määr on keskmisest 9% madalam.

Grupp 1 on kliendid, kellel on viimasest ostust kõige kauem aega möödas, kuid nad teevad oste suhteliselt sagedasti ja nende ARPA on kõige kõrgem. Selle grupi vastuvõtlikkuse määr on keskmisest 14% madalam.

Grupp 2 on kliendid, kellel on viimasest ostust pikem aeg möödas, nende ostude sagedus on üsna madal ja ARPA on keskmine. Selle grupi vastuvõtlikkuse määr on keskmisest 23% kõrgem.

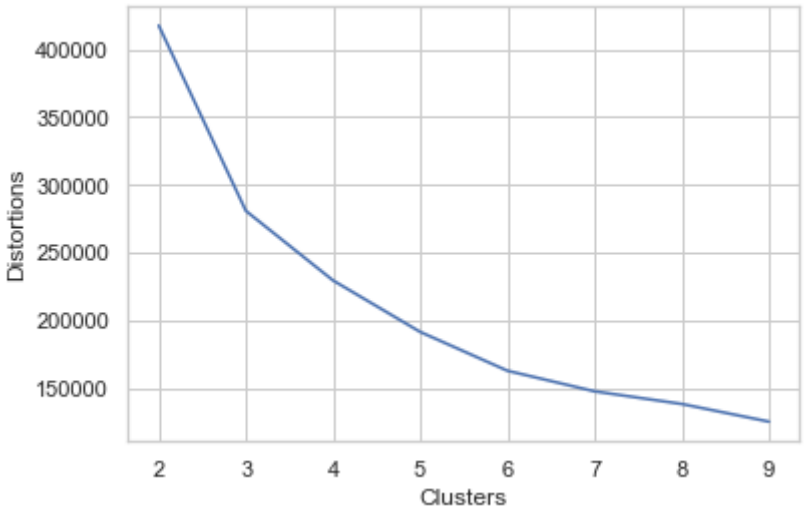
3.2.5 K-means RFA skooridega

Et mõista, kas normaliseerimine annab paremad tulemused, on tehtud teine mudel mitte normaliseeritud RFM väärtustega, vaid Jenks Breaksi abil leitud skooridega. Jenks Breaks algoritm jagab RFA väärtused 4 vahemikku nagu kõige esialgsema mudeli puhul. Segmentide leidmiseks on rakendatud k-means klasterdamisalgoritmi. Joonis 17 illustreerib RFA skooridega k-means mudeli loomise protsessi.



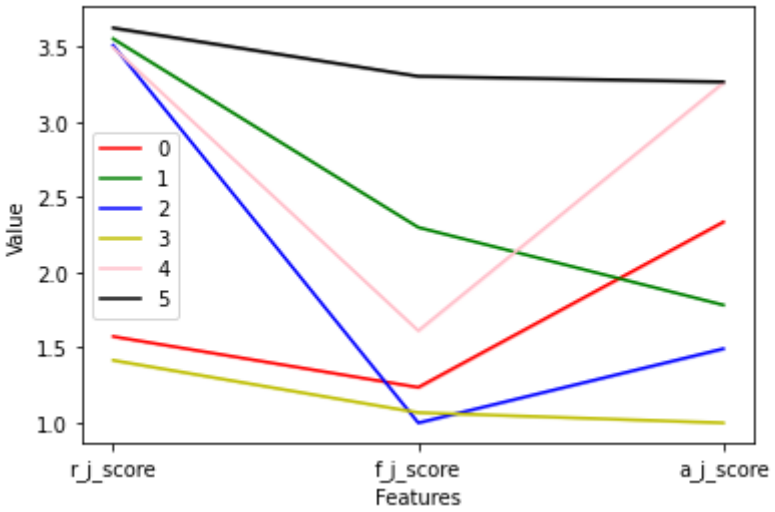
Joonis 17. RFA skooridega k-means mudeli loomise protsess.

Optimaalne k väärtus on leitud küünarnuki meetodil. Joonise 18 põhjal otsustades on klastrite arvuks valitud 6.



Joonis 18. RFA skooridega k-means mudeli küünarnuki meetod.

Mudel, mis kasutab RFM skoore ja k-means algoritmi annab tulemuseks väga erinevad grupid. Joonis 19 illustreerib RFA skooridega k-means mudeli gruppe ja tunnuste väärtuseid. Gruppide nimed 0, 1, 2 on juhuslikud ja ei oma tähendust.



Joonis 19. RFA skooridega k-means mudeli grupid ja tunnuste väärtused

Tabelis 6 on RFA skooridega k-means mudeli gruppide vastuvõtlikkuse määrad.

Tabel 6. RFA skooridega k-means mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
0	keskmine
1	keskmisest 5% madalam
2	keskmisest 9% madalam
3	keskmine
4	keskmisest 8% kõrgem
5	keskmisest 4% kõrgem

Grupp 0 on teinud oma viimase ostu rohkem aega tagasi ja nad ei tee oste sagedasti. Samas on ARPA nendel klientidel üle keskmise. Selle grupi pakkumiste vastuvõtlikkuse määr on keskmine.

Grupp 1 on kliendid, kes on väga värskelt teinud ostu ja on varem ka oste sooritanud, kuid nende ARPA on pigem madal. Selle kliendigrupi pakkumiste vastuvõtlikkuse määr on keskmisest 5% madalam.

Grupp 2 on kliendid, kes on väga värskelt teinud ostu, kuid see on nende esimene ost. ARPA on madal. Selle kliendigrupi pakkumiste vastuvõtlikkuse määr on keskmisest 9% madalam.

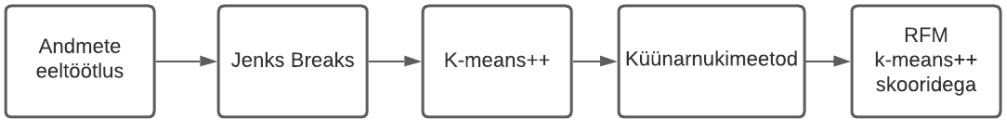
Grupp 3 on kliendid, kellel on viimasest ostust kõige kauem aega möödas. Nad teevad oste harva ja ARPA on madal. Selle kliendigrupi pakkumise vastuvõtlikkuse määr on keskmine.

Grupp 4 on kliendid, kes on värskelt teinud ostu. See on kordusost, kuid nad ei tee oste väga sagedasti. Nende klientide ARPA on kõrge. Selle kliendigrupi pakkumise vastuvõtlikkuse määr on keskmisest 8% kõrgem.

Grupp 5 on kliendid, kes on viimase ostu teinud hiljuti, nad teevad oste sageli ning nende ARPA on kõrge. Selle kliendigrupi pakkumise vastuvõtlikkuse määr on keskmisest 4% kõrgem.

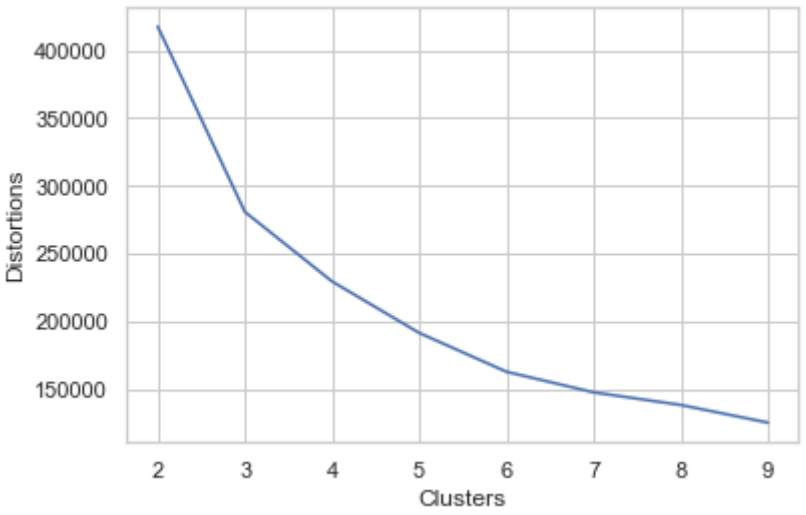
3.2.6 K-means ++ RFA skooridega

Et võrrelda, kas k-means ja k-means++ algoritmide rakendamine annab tulemustes erinevusi, on mõlemaid algoritme rakendatud. Joonis 20 illustreerib RFA skooridega k-means++ mudeli loomise protsessi.



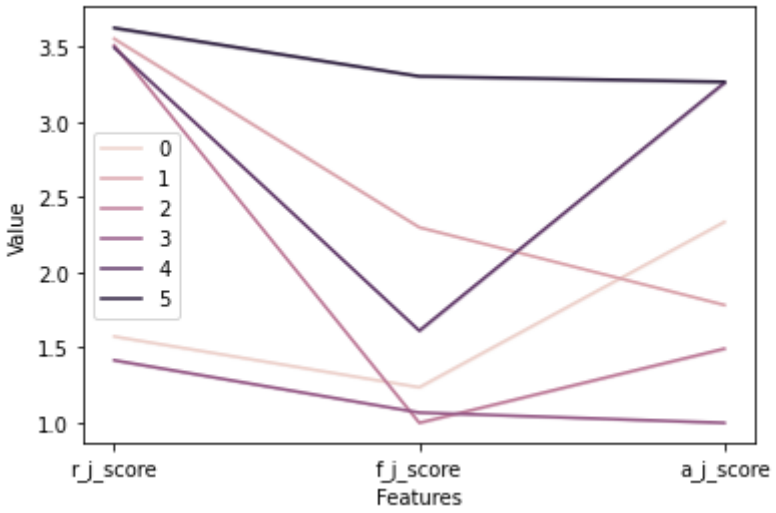
Joonis 20. RFA skooridega k-means++ mudeli loomise protsess.

Optimaalseimaks klastrite leidmideks on kasutatud küünarnuki meetodit. Joonise 21 põhjal otsustades on klastrite arvuks valitud 6.



Joonis 21. RFA skooridega k-means++ mudeli küünarnuki meetod.

Joonis 22 illustreerib RFA skooridega k-means++ mudeli gruppe ja tunnuste väärtuseid.



Joonis 22. RFA skooridega k-means++ mudeli grupid ja tunnuste väärtused.

Tabelis 7 on RFA skooridega k-means++ mudeli gruppide vastuvõtlikkuse määrad.

Tabel 7. RFA skooridega k-means++ mudeli gruppide vastuvõtlikkuse määrad.

Grupp	**Vastuvõtlikkuse määr**
0	keskmine
1	keskmisest 5% madalam
2	keskmisest 9% madalam
3	keskmine
4	keskmisest 8% kõrgem
5	keskmisest 4% kõrgem

Kuna tulemused on täpselt samad RFA skooridega k-means mudeliga, siis on neid võimalik detailsemalt lugeda eelmisest peatükist.

4 Analüüs ja olulised järeldused

Selles peatükis on esmalt toodud mudelite üldine võrdlus ning kõrvutatud erinevaid metoodikaid. Seejärel on võrreldud esialgset andmemudelit mudeliga, kuhu on lisatud ARPA. Järgmisena on toodud välja, milline erinevus on sellel, kui RFM skoorid jaotada 4 või 5 vahemikku. Võrreldud on, kas segmendid tulevad erinevad, kui algandmed standardiseerida, normaliseerida või kasutada skoore. Lisaks on võrreldud mudeleid, kus on rakendatud k-means ja k-means++ algoritme. Peatüki lõpus on mudelite kokkuvõte ning võimalikud edasiarendused. Lisaks on välja toodud mudeli rakendamise majanduslik kasu ettevõttele, hinnang ettevõtte spetsialistidelt ja käesoleva töö kasu laiemale publikule.

4.1 Mudelite võrdlus

Tulemused hinnatakse heaks, kui on leitud grupid, mida turundus saab oma tegevustes efektiivselt kasutada. Kõige olulisem on leida üles kõrge vastuvõtlikkuse määraga kliendigrupid, kuid kasulikku infot annavad ka keskmisest madalama vastuvõtlikkusega grupid, sest neid on võimalik valimitest välistada. Joonisel on näha töö käigus loodud erinevate mudelite gruppide jaotus ja nende gruppide klientide keskmine turunduslike pakkumiste vastuvõtlikkuse määr. Tabelis 8 on kõiki tulemusi võrreldud kogu valimi keskmisega.

Tabel 8. Mudelite võrdlustabel.

	RF	**RFA 4**	**RFA 5**	**RFA k-means skoorid**	**RFA k-means++ skoorid**	**RFA k-means standard**	**RFA k-means norm**
Jenks Breaks	4	4	5	4	4	-	-
Andmed	skoorid	skoorid	skoorid	skoorid	skoorid	standard	normal
Tukey	X	X	X	-	-	-	-
Klasterdamine	-	-	-	k-means	k-means++	k-means	k-means
	9% kõrgem	39% kõrgem	290% kõrgem	8% kõrgem	8% kõrgem	9% kõrgem	23% kõrgem
	8% madalam	24% kõrgem	53% kõrgem	4% kõrgem	4% kõrgem	keskmine	14% madalam
	40% madalam	7% kõrgem	13% kõrgem	keskmine	keskmine	9% madalam	9% madalam
		14% madalam	9% madalam	keskmine	keskmine		
			31% madalam	5% madalam	5% madalam		
			59% madalam	9% madalam	9% madalam		

4.1.1 ARPA lisamine mudelisse

Ettevõtte andmeteadlased olid varasemalt otsustanud, et rahalise väärtuse kui kauba ostude summa arvestamine mudelis ei ole telekommunikatsiooni valdkonnas asjakohane, mistõttu loodi RF mudel, mis arvestab vaid *recency* ja *frequency* väärtuseid.

Varasemalt on teise autori poolt telekommunikatsiooniettevõttes kasutatud ARPU väärtust [13], millest tulenevalt tekkis ka hüpotees selle töö jaoks, et ARPA lisamine võiks mudelit parandada. Kuna ühel kliendil võib olla mitmeid kasutajaid ja eesmärk on kliendile mitte kasutajale segment leida, siis kasutatakse ARPU (*Average Revenue Per User*) asemel ARPA (*Average Revenue Per Account*) väärtust. Mudeli sisendiks on võetud kliendi 3 kuu keskmine ARPA väärtus.

Tabel 9 kirjeldab RF ja RFA 4 mudelite tulemusi. RF mudelis tekkis 3 eristuvat gruppi, kus üks grupp on keskmisest 9% kõrgema vastuvõtlikkusega ja 2 gruppi keskmisest madalama vastuvõtlikkusega (9% ja 40%). RFA mudelis tekkis 5 eristuvat gruppi, kus 3 gruppi on keskmisest kõrgema vastuvõtlikkusega (39%, 24% ja 7%) ning 2 gruppi keskmisest madalama vastuvõtlikkusega (14% ja 46%). Kummalgi juhul ei tekkinud gruppe, mille vastuvõtlikkuse määr oleks võrdne kogu valimi keskmisega

Tabel 9. RF ja RFA 4 mudelite võrdlus.

	RF	RFA 4
Gruppide arv	3	5
Keskmisest kõrgema vastuvõtlikkusega grupid	1 (keskmisest **9%** kõrgem)	3 (keskmisest **39%** kõrgem, keskmisest **24%** kõrgem ja keskmisest **7%**)
Keskmisest madalama vastuvõtlikkusega grupid	2 (keskmisest **9%** madalam ja keskmisest **40%** madalam)	2 (keskmisest **14%** madalam ja keskmisest **46%** madalam)

Turunduslikust vaatest on RFA 4 mudel oluliselt kasulikum, sest selle mudeli järgi on võimalik üles leida lausa 3 keskmisest kõrgema vastuvõtlikkusega gruppi. Tänu sellele on võimalik täpsemalt sihtida kliente, kes oleks turunduslikest pakkumistest huvitatud. Seejuures kaks gruppi nendest eristuvad kõikide mudelite kõikidest gruppidest märgatavalt, sest vastuvõtlikkuse määr on kõigist nendest kõrgem (keskmisest 39% ja 24% kõrgem).

Kolm “head” gruppi ehk gruppi, millel oli keskmisest kõrgem vastuvõtlikkuse määr on kasulikud ka seetõttu, et sageli peavad otseturundustegevustes valimite mahud jääma teatud piiridesse. Kolme “hea” grupiga on rohkem võimalusi. Kui on pakkumine, mille puhul soovitakse sihtida laiemalt, on võimalik kasutada kõiki “häid” gruppe ning kui on kommunikatsioonitegevus, mis on mõeldud väiksemale valimile, on võimalik valida kliendid ainult ühest kõige kõrgemast.

Kui võrrelda gruppide pakkumiste vastuvõtlikkuse määrasid ja ARPA väärtuseid, siis need on korrelatsioonis. Mida kõrgem on ARPA, seda kõrgem on ka grupi pakkumiste vastuvõtlikkuse määr. Pearsoni korrelatsioonikordaja on 0.998 ehk ARPA ja vastuvõtlikkuse määra vahel on väga tugev positiivne korrelatsioon.

Kuigi siin ei saa kindlalt tuua otsest põhjuslikku seost, et kõrge ARPAga kliendid on tingimata alati pakkumistele vastuvõtlikkumad, sest kehtib ka vastupidine seos - pakkumisi vastu võtnud klientide ARPA üldiselt kasvab, sest antud juhul on otseturunduse eesmärgiks olnud hange või ülesmüük. Siiski võib uskuda, et kui klient on varasemalt ettevõtte teenuseid usaldanud selliselt, et on olnud nõus võtma kallimaid pakette või rohkem teenuseid (mõlemad mõjutavad ARPAt kasvavalt), siis on tekkinud piisav usaldus ja lojaalsus, et klient võiks ka edaspidi olla huvitatud.

Otseturunduses võiks RFA 4 mudeliga saadud tulemusi võrrelda valimiga, kuhu on käsitsi võetud kõrgeima ARPAga kliendid, et näha, kas *recency* ja *frequency* on üldse olulised.

4.1.2 RFM skoorid 4 vs 5

Võrdlemaks, kas mudel annab paremaid tulemusi, kui RFA väärtused jagada 4 või 5 vahemikku, on mudelid loodud kahel viisil. Tabel 10 kirjeldab RFA 5 ja RFA 4 mudelite tulemusi.

Tabel 10. RFA 5 ja RFA 4 mudelite võrdlus.

	RFA 4	RFA 5
Gruppide arv	5	5
Keskmisest kõrgema vastuvõtlikkusega grupid	3 (keskmisest **39%** kõrgem, keskmisest **24%** kõrgem ja keskmisest **7%**)	3 (keskmisest **290%** kõrgem, keskmisest **53%** kõrgem, keskmisest **13%** kõrgem)
Keskmisest madalama vastuvõtlikkusega grupid	2 (keskmisest **14%** madalam ja keskmisest **46%** madalam)	3 (keskmisest **9%** madalam, keskmisest **31%** madalam, keskmisest **59%** madalam)

RFA 5 leidis kliendigrupi, mille turunduslike pakkumiste vastuvõtlikkuse määr on keskmisest 290% kõrgem, kuid selle grupi klientide arv on niivõrd väike, et see grupp jääb hetkel võrdlusest välja.

RFA 5 järgmise parima grupi vastuvõtlikkuse määr on 53% keskmisest kõrgem, kui RFA 4 puhul oli parim keskmisest 39% kõrgem ehk RFA 5 leiab paremini vastuvõtlikud kliendid üles.

Mõlemad mudelid sobivad otseturunduses segmenteerimiseks ja annavad häid tulemusi ning siin tuleks mudeli valikul lähtuda sellest, kui suurt valimit on tarvis sihtida ning arvestada nende gruppide valimite suurustega. Käesolevas töös ei ole võimalik gruppide suuruseid avalikustada, kuna ettevõtte kliendibaasi kirjeldavad andmed on konfidentsiaalsed.

4.1.3 Standardiseeritud väärtused vs normaliseeritud väärtused vs skoorid

Võrdlemaks, kas normaliseerimine annab paremaid tulemusi loodi mudel k-means algoritmi kasutades kahel viisil. Ühel juhul olid sisendandmeteks standardiseeritud *recency*, *frequency* ja *monetary* väärtused. Teisel juhul olid sisendiks RFA väärtuste skoorid. Tabelis 11 on standardiseeritud, normaliseeritud ja skooridega mudelite võrdlus.

Tabel 11. Standardiseeritud, normaliseeritud ja skooridega mudelite võrdlus.

	RFA k-means standardiseeritud	RFA k-means normaliseeritud	RFA k-means skoorid
Gruppide arv	3	3	6
Keskmisest kõrgema vastuvõtlikkusega grupid	1 (keskmisest 9% kõrgem)	1 (keskmisest 23% kõrgem)	2 (keskmisest 8% kõrgem ja keskmisest 4% kõrgem)
Keskmisest madalama vastuvõtlikkusega grupid	1 (keskmisest 9% madalam)	2 (keskmisest 9% madalam ja keskmisest 14% madalam)	2 (keskmisest 5% madalam ja keskmisest 9% madalam)

Kolmest mudelist eristub kõige rohkem hetkel RFA k-means normaliseeritud andmetega mudel, sest see andmemudel leidis ühe kliendigrupi, kelle otseturunduse pakkumiste vastuvõtlikkuse määr on keskmisest 23% kõrgem. Võrdluseks, RFA k-means standardiseeriud andmetega mudel leidis ühe kliendigrupi, kelle otseturunduse pakkumiste vastuvõtlikkuse määr on keskmisest 9% kõrgem. Standardiseeritud andmetega mudel leidis ühe kliendigrupi, mille vastuvõtlikkuse määr on keskmisest madalam, aga normaliseeritud andmetega mudel leidis kaks kliendigruppi, kelle vastuvõtlikkuse määr on keskmisest madalam.

Skooridega ja normaliseeritud andmetega mudelid on andnud erineva arvu kliendigruppe, vastavalt 6 ja 3. Skooridega loodud andmemudel leidis kaks keskmisest kõrgema vastuvõtlikkusega gruppi (keskmisest 8% kõrgem ja keskmisest 4% kõrgem) ja normaliseeritud andmetega mudel leidis ühe (keskmisest 23% kõrgem). Kuna oluline oli leida kliendigrupid, kes on kõige kõrgema vastuvõtlikkuse määraga, siis on võimalik öelda, et normaliseeritud andmetega mudel on parem.

4.1.4 K-means vs k-means++

Kuigi k-means ja k-means++ on sarnased algoritmid, siis need ei ole päris samad ja võivad anda erinevaid tulemusi. Tabelis 12 on välja toodud RFA k-means ja k-means++ mudelite võrdlus. Antud juhul selgub, et tulemused tulid täpselt samad ja hetkel ei ole erinevust, kumba algoritmi kasutada.

Tabel 12. RFA k-means ja k-means++ mudelite võrdlus.

	RFA k-means	RFA k-means++
Gruppide arv	6	6
Keskmisest kõrgema vastuvõtlikkusega grupid	2 (keskmisest 8% kõrgem, keskmisest 4% kõrgem)	2 (keskmisest 8% kõrgem, keskmisest 4% kõrgem)
Keskmisest madalama vastuvõtlikkusega grupid	2 (keskmisest 5% madalam, keskmisest 9% madalam)	2 (keskmisest 5% madalam, keskmisest 9% madalam)

4.1.5 Mudelite kokkuvõte

Mudeli tulemuste analüüsil on selgunud, et:

- ARPA tunnuse arvestamine parandab mudelit;
- andmete normaliseerimine k-meansi puhul annab paremad tulemused;
- ei ole vahet, kas rakendada k-means või k-means++ klasterdamisalgoritmi;
- mudel ARPAga, kus on kasutatud RFM skoore ja Tukey testil leitud eristuvad grupid, on kõige parem;
- RFM tunnuste jaotamine 5 vahemikku jaotamine on veidi parem, kuid 4 vahemikku jaotatud mudel töötab ka hästi.

4.2 Võimalikud edasiarendused

Otseturunduses võiks RFA mudeliga saadud tulemusi võrrelda valimiga, kuhu on käsitsi võetud kõrgeima ARPAga kliendid, et näha, kas *recency* ja *frequency* on üldse olulised.

Lisaks võiks katsetada veel erinevaid klasterdamisalgoritme, näiteks DBSCAN, Fuzzy C-means või mean shift.

RFA mudelile võiks lisada O (*offer*) tunnuse, mis tähistab, kui palju pakkumisi klient on varasemalt vastu võtnud. Siin võiks kontrollida seose olemasolu. Autor esitab hüpoteesina, et kliendid, kes on varasemalt rohkem pakkumisi vastu võtnud, võiksid olla ka edaspidi vastuvõtlikumad. Seda hüpoteesi võiks kontrollida.

Samuti võib proovida M (*marketing*) tunnuse lisamist, mis tähistab, kui palju klient on turundussõnumitele avatud olnud, kuid ei ole tingimata ostu veel teinud. See tunnus võiks tähistada näiteks, kas klient on e-maili teel saadud kirju avanud või pakkumisele edasisuunavale nupule vajutanud.

4.3 Majanduslik kasu

Varasemalt on RF mudelit ettevõttes turundustegevuste jaoks katsetatud ja kasutatud. Mudel on andnud paremaid tulemusi, kui spetsialistil oma parima teadmise ja tunnetuse järgi valimeid koostades.

Kuna töö käigus on selgunud, et ARPAga loodud RFM mudel on turundustegevusteks efektiivsem ning see on hetkel parim teadmine, siis plaanis on mudelit tulevastel otseturundustegevustel testida ning kui ajalooliste andmete peal loodud testid ja reaalne tulemus on sarnased, siis on võimalik mudel töös rakendada.

Efektiivselt leitud kliendigrupid suurendavad klientide rahulolu ja aitavad otseturunduses täita kampaania eesmärke, mis on otsene rahaline kasu ettevõttele.

4.4 Hinnang ettevõtte spetsialistidelt

Autor on töö käigus saanud nõu käesoleva projektiga seotud ettevõtte andmeteadlaselt Siim Tarbelt, kes on andnud hinnangu ka sellele tööle: “Autori panus ettevõtte otsekommunikatsiooni edendamiseks on olnud märkimisväärne. Klientide segmenteermiseks õnnestus autoril luua parem versioon RFM mudelist võrreldes senisega. Samuti on autor välja toonud mitmeid häid mõtteid edasisteks sammudeks mudeli täiendamisel ja rakendamisel. Otsekommunikatsioon meie ettevõttes võikski potentsiaalselt tulevikus toetuda suuresti RFM mudelile.”

4.5 Kasu üldsusele

Suur hulk publitseeritud töödest, kus on rakendatud RFMi, on teostatud kaubandusvaldkonnas, kus *monetary* väärtuseks on kaubaostude rahaline väärtus. Kuigi *monetary* väärtuseid on erinevates tegevusvaldkondades erinevalt rakendatud, siis on jäädud üsna üldsõnaliseks, mida *monetary* väärtus endas sisaldab. Kuigi see võib olla töödes kasutatud ka kui teenustele kulutatud summa, siis töö autor ei ole leidnud publitseeritud töid, kus oleks *monetary* väärtuse asemel konkreetselt ARPA väärtust

kasutatud. Autori panus valdkonnas on selle nüansi täpsustamine ning ettepanek, et ettevõtetes, kus on põhitegevusalaks teenuste pakkumine, võiks kasutada kaubaostude rahalise väärtuse asemel alternatiivina ARPA väärtust.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli leida ettevõtte otseturundustegevuste paremaks sihtimiseks efektiivsem mudel. Turundustegevuste sihtimine on oluline, sest see hoiab kliendi rahulolu ja annab ettevõttel võimaluse efektiivsemalt täita ärieesmärke.

Ettevõttes oli esialgne RFM mudel koostatud, kuid see vajas testimist ja täiendamist. Esialgne RFM mudel sisaldas endas vaid *recency* ja *frequency* väärtuseid. Töö käigus on valminud 6 üksteisest veidi erinevat andmemudelit.

Olulisim täiendus oli *monetary* väärtuse lisamine mudelisse, mis tulenevalt ettevõtte tegevusvaldkonnast on asendatud kliendi kolme kuu keskmise ARPA väärtusega. Seejärel võrreldi, kas ja kuidas muudab tulemusi RFM skooride erinevad vahemikud; andmete standardiseerimine, normaliseerimine või skooride kasutamine; k-means või k-means++ algoritmi rakendamine segmenteerimiseks.

Mudeleid on võrreldud omavahel ning esialgse RFM mudeliga. Nendest mudelitest on eristunud parim mudel, mis leidis kliendigrupid, kelle turunduslike pakkumiste vastuvõtlikkuse määr on keskmisest 1) 53% kõrgem, 2) 13% kõrgem, 3) 9% madalam 4) 31% madalam, 5) 59% madalam (ja väike grupp, kelle vastuvõtlikkuse määr on keskmisest 290% kõrgem). Väga kasulik teadmine on 1. ja 2. grupp, kuhu kuuluvad kliendid, kellele tasub teha turunduslikke pakkumisi ja samas ka grupp 4. ja 5., sest nad on madalama vastuvõtlikkuse määraga ning selliseid kliente tasub turunduslike pakkumiste tegemisel valimist eemaldada.

Kogu protsess on dokumenteeritud ja ettevõttes on võimalik mudel kasutusse võtta ning edasi arendada.

Käesolev magistritöö pakub välja, et ettevõtetes, kus on põhitegevusalaks teenuste pakkumine, võiks kasutada kaubaostude rahalise väärtuse asemel alternatiivina ARPA väärtust.

Kasutatud kirjandus

[1] „Direct Marketing: What You Need to Know“, *Investopedia*. https://www.investopedia.com/terms/d/direct-marketing.asp (vaadatud okt 27, 2021).

[2] V. Kumar, X. (Alan) Zhang, ja A. Luo, „Modeling Customer Opt-In and Opt-Out in a Permission-Based Marketing Context“, *J. Mark. Res.*, kd 51, nr 4, lk 403–419, aug 2014, doi: 10.1509/jmr.13.0169.

[3] A. Marinova, J. Murphy, ja B. L. Massey, „Permission E-mail Marketing as a Means of Targeted Promotion“, *Cornell Hotel Restaur. Adm. Q.*, kd 43, nr 1, lk 61–69, veebr 2002, doi: 10.1177/0010880402431006.

[4] M. Krafft, C. M. Arden, ja P. C. Verhoef, „Permission Marketing and Privacy Concerns — Why Do Customers (Not) Grant Permissions?“, *J. Interact. Mark.*, kd 39, lk 39–54, aug 2017, doi: 10.1016/j.intmar.2017.03.001.

[5] vahide babaiyan ja S. A. Sarfarazi, „Analyzing customers of South Khorasan Telecommunication Company, with the expansion of the RFM to LRFM model“, *J. AI Data Min.*, nr Online First, mai 2018, doi: 10.22044/jadm.2018.6035.1715.

[6] T. Segal, „Recency, Frequency, Monetary Value (RFM) Definition“, *Investopedia*. https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp (vaadatud okt 16, 2021).

[7] K. Chen, Y.-H. Hu, ja Y.-C. Hsieh, „Predicting customer churn from valuable B2B customers in the logistics industry: a case study“, *Inf. Syst. E-Bus. Manag.*, kd 13, nr 3, lk 475–494, aug 2015, doi: 10.1007/s10257-014-0264-1.

[8] S. Allegue, T. Abdellatif, ja K. Bannour, „RFMC: a spending-category segmentation“, *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Bayonne, France, sept 2020, lk 165–170. doi: 10.1109/WETICE49692.2020.00040.

[9] J.-T. Wei, S.-Y. Lin, C.-C. Weng, ja H.-H. Wu, „A case study of applying LRFM model in market segmentation of a children’s dental clinic“, *Expert Syst. Appl.*, kd 39, nr 5, lk 5529–5533, apr 2012, doi: 10.1016/j.eswa.2011.11.066.

[10] A. Sheikh, T. Ghanbarpour, ja D. Gholamiangonabadi, „A Preliminary Study of Fintech Industry: A Two-Stage Clustering Analysis for Customer Segmentation in the B2B Setting“, *J. Bus.--Bus. Mark.*, kd 26, nr 2, lk 197–207, apr 2019, doi: 10.1080/1051712X.2019.1603420.

[11] I.-C. Yeh, K.-J. Yang, ja T.-M. Ting, „Knowledge discovery on RFM model using Bernoulli sequence“, *Expert Syst. Appl.*, kd 36, nr 3, lk 5866–5871, apr 2009, doi: 10.1016/j.eswa.2008.07.018.

[12] N. Jha, D. Parekh, M. Mouhoub, ja V. Makkar, „Customer Segmentation and Churn Prediction in Online Retail“, *Advances in Artificial Intelligence*, kd 12109, C. Goutte ja X. Zhu, Toim Cham: Springer International Publishing, 2020, lk 328–334. doi: 10.1007/978-3-030-47358-7_33.

[13] J. Tri Wibowo ja M. Suryanegara, „Segmenting the Subscribers of An Indonesian 4G Service Operator Using RFM Method“, *2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*,

Purwokerto, Indonesia, juuli 2021, lk 280–283. doi: 10.1109/COMNETSAT53002.2021.9530772.
[14] „Guidelines for Removing and Handling Outliers in Data“, *Statistics By Jim*, okt 23, 2019. https://statisticsbyjim.com/basics/remove-outliers/ (vaadatud dets 07, 2021).
[15] „scipy.stats.iqr — SciPy v1.7.1 Manual“. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.iqr.html (vaadatud dets 01, 2021).
[16] T. Kaart, „Sagedused ja osakaalud – diskreetne tunnus“, lk 13, 2010.
[17] „Interquartile Range (IQR): Definition and Uses“, *Statistics By Jim*, aug 29, 2021. https://statisticsbyjim.com/basics/interquartile-range/ (vaadatud dets 07, 2021).
[18] „Data classification methods—ArcGIS Pro | Documentation“. https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm (vaadatud dets 17, 2021).
[19] R. Ahmad, „Jenks Natural Breaks — Best range finder algorithm.“, *Analytics Vidhya*, aug 05, 2019. https://medium.com/analytics-vidhya/jenks-natural-breaks-best-range-finder-algorithm-8d1907192051 (vaadatud nov 03, 2021).
[20] I. Felton, „RFM Segmentation using Quartiles and Jenks Natural Breaks“, *Medium*, juuli 15, 2019. https://towardsdatascience.com/rfm-segmentation-using-quartiles-and-jenks-natural-breaks-924f4d8baee1 (vaadatud dets 17, 2021).
[21] U. Jaitley, „Why Data Normalization is necessary for Machine Learning models“, *Medium*, apr 09, 2019. https://medium.com/@urvashilluniya/why-data-normalization-is-necessary-for-machine-learning-models-681b65a05029 (vaadatud dets 15, 2021).
[22] „sklearn.preprocessing.MinMaxScaler“, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html (vaadatud dets 17, 2021).
[23] „IDN1605: Teema ja ülesanne 5. Andmete eeltöötlus. (.html)“. https://moodle.taltech.ee/mod/resource/view.php?id=216098 (vaadatud jaan 03, 2022).
[24] Hshan.T, „Exploring Customers Segmentation With RFM Analysis and K-Means Clustering.“, *The Startup*, märts 20, 2021. https://medium.com/swlh/exploring-customers-segmentation-with-rfm-analysis-and-k-means-clustering-93aa4c79f7a7 (vaadatud dets 15, 2021).
[25] „sklearn.preprocessing.StandardScaler“, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (vaadatud dets 17, 2021).
[26] D. M. J. Garbade, „Understanding K-means Clustering in Machine Learning“, *Medium*, sept 12, 2018. https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 (vaadatud nov 03, 2021).
[27] S. Peker, A. Kocyigit, ja P. E. Eren, „LRFMP model for customer segmentation in the grocery retail industry: a case study“, *Mark. Intell. Plan.*, kd 35, nr 4, lk 544–559, mai 2017, doi: 10.1108/MIP-11-2016-0210.
[28] A. J. Christy, A. Umamakeswari, L. Priyatharsini, ja A. Neyaa, „RFM ranking –

An effective approach to customer segmentation“, *J. King Saud Univ. - Comput. Inf. Sci.*, lk S1319157818304178, sept 2018, doi: 10.1016/j.jksuci.2018.09.004.
[29] J. Wu *et al.*, „An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm“, *Math. Probl. Eng.*, kd 2020, lk 1–7, nov 2020, doi: 10.1155/2020/8884227.
[30] Y. Parikh ja E. Abdelfattah, „Clustering Algorithms and RFM Analysis Performed on Retail Transactions“, *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York, NY, USA, okt 2020, lk 0506–0511. doi: 10.1109/UEMCON51285.2020.9298123.
[31] A. G. Aggarwal ja S. Yadav, „Customer Segmentation Using Fuzzy-AHP and RFM Model“, *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, juuni 2020, lk 77–80. doi: 10.1109/ICRITO48877.2020.9197903.
[32] Y. Huang, M. Zhang, ja Y. He, „Research on improved RFM customer segmentation model based on K-Means algorithm“, *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*, Beijing, China, juuni 2020, lk 24–27. doi: 10.1109/ICCIA49625.2020.00012.
[33] X. Hu, Z. Shi, Y. Yang, ja L. Chen, „Classification Method of Internet Catering Customer Based on Improved RFM Model and Cluster Analysis“, *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, Chengdu, China, apr 2020, lk 28–31. doi: 10.1109/ICCCBDA49378.2020.9095607.
[34] R. Zhao ja C. Li, „Research on E-commerce Customer Segmentation Based on RFAC Model“, *2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, juuli 2021, lk 439–444. doi: 10.1109/ICPICS52425.2021.9524108.
[35] „Elbow Method for optimal value of k in KMeans“, *GeeksforGeeks*, juuni 06, 2019. https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/ (vaadatud dets 08, 2021).
[36] „Tehisintellekti Algkursus - Kursused - Arvutiteaduse instituut“. https://courses.cs.ut.ee/2020/Tehisintellekti_algkursus/spring/Main/PARTVISegmenteerimine (vaadatud dets 08, 2021).
[37] „What Is the Tukey HSD Test?“, *Sciencing*. https://sciencing.com/what-is-the-tukey-hsd-test-12751748.html (vaadatud dets 02, 2021).
[38] M. F. Bacila, R. Adrian, ja M. Ioan, „RFM Based Segmentation An Analysis of a Telecom Companys Customers MID2012.pdf“.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks[1]

Mina, Kätrin Grauberg

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose , mille juhendaja on
   1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
   1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

1 Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.