

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Infotehnoloogia mitteinformaatikutele õppekava

**Kaisa Ehrpais**

**Tehisintellekti tutvustavate materjalide  
kogumiku väljatöötamine STACC OÜ näitel**

**Magistritöö (15 EAP)**

Juhendajad: Kerli Mooses PhD

Karl-Oskar Masing BSc

Tartu 2022

## **Tehisintellekti tutvustavate materjalide kogumiku väljatöötamine STACC OÜ näitel**

### **Lühikokkuvõte:**

Üha populaarsust koguvatel tehisintellektil põhinevatel tehnoloogiatel meie ümber on paratamatult oma keerukuse ning kiire arengu tõttu ka varjukülgi, mis paneb ettevõtjaid sellise investeeingu mõistlikkuses kahtlema. Öeldakse, et kõik uus tundubki alguses hirmuäratav - paljud ettevõtjate hirmud on võimalik selgitustööga hajutada. Selleks on aga vaja esialgu ettevõtete hirmud masinõppe lahenduste ees kaardistada ning hetkel puudub andmeteaduse ettevõttes STACC OÜ vastav kaardistus. Antud magistritöö eesmärgiks on STACC OÜ vajaduste põhjal kaardistada ning avada potentsiaalsete klientide jaoks masinõppel põhinevate süsteemide kohta esinevaid küsimusi. Intervjuude, ajurünnaku ning olemasolevate dokumentide analüüsist kogutud materjal koondati ning kategoriseeriti vastavalt sisule. Kogutud andmete põhjal analüüsiti kõige sagedamini esinevaid masinõppega seotud teemasid, küsimusi ning hirme, mille põhjal moodustati 15 eelnevalt käsitletud kategooriat katvat küsimust. Seejärel vormistati küsimustele koostöös andmeteaduse spetsialistiga kogutud materjalide põhjal sobivad vastused. Koostatud materjali eesmärk oleks võimaldada tehisintellekti rakendamist kaaluval ettevõtjal teostada iseseisvalt esmane analüüs, mis käsitleks lahenduse võimalikkust, vastaks esmastele küsimustele ning maandaks ettevõtja hirme.

### **Võtmesõnad:**

Tehisintellekt, turundus.

**CERCS:** P176, Tehisintellekt

## **The development of a introductory materials on Artificial Intelligence as illustrated by the example of STACC OÜ**

### **Abstract:**

Due to its complexity and fast development the artificial intelligence based technologies inevitably have their downsides which makes the entrepreneurs doubt the reasonableness of such an investment. It is said that everything unknown and new seems scary at first - many of the fears of entrepreneurs can be decreased by giving explanations about the technology. However, for the time being, it is necessary to map companies' fears about machine learning solutions. Currently there is no such mapping in the data science company STACC OÜ. The aim of this thesis is to map the needs of STACC and give answers to the questions about machine learning based systems to its potential customers. The information collected from the interviews, brainstorming session and analysis of existing documents was compiled and categorized according to content. Based on the collected data, the most common topics, questions and fears related to machine learning were analyzed and 15 questions covering the previously discussed categories were formed based on the analysis. Appropriate answers to the questions were then formulated on the basis of the materials collected in cooperation with a data science specialist. The purpose of the compiled material is to enable the potential client to carry out an initial analysis of the feasibility of a solution, to answer the most frequent questions and to decrease the fears of the entrepreneur.

### **Keywords:**

Artificial intelligence, marketing

**CERCS:** P176, Artificial intelligence

# Sisukord

<b>Sissejuhatus</b>	<b>6</b>
<b>1 Tausta kirjeldus</b>	<b>10</b>
1.1 Tehisintellekti olemus	10
1.1.1 Tehisintellekti kategooriad ja alamvaldkonnad	11
1.1.2 Tehisintellekti kasutusvaldkonnad	12
1.1.3 Tehisintellekti kasutamine Eestis	12
1.2 Masinõpe	13
1.2.1 Masinõpe vs tehisintellekt	14
1.3 Tehisintellekti mõjuvõimu kasvuga kaasnevad probleemid	15
1.3.1 Tehisintellekti mõjuvõimu kasv	15
1.3.2 Praktilised probleemid	16
1.3.3 Eetilised probleemid	17
1.3.4 Inimeste suhe tehisintellektiga	18
1.4 Probleem	19
<b>2 Metoodikad</b>	<b>20</b>
2.1 Andmete kogumise meetodid	20
2.1.1 Ajurünnak	20
2.1.2 Intervjuu	21
2.1.3 Koolitustel osalemine	22
2.1.4 Masinõppe lõuendi läbitöötamine	22
2.2 Analüüsimeetodid	23
<b>3 Erinevatelt sihtrühmadelt kogutud sisend</b>	<b>25</b>
3.1 STACCI sisene ajurünnak	25
3.1.1 Peamised põhjused masinõppe lahenduste kasutamiseks	25
3.1.2 Klientide hirmud ning takistused masinõppe projektiga alustamiseks	25
3.1.3 Klientide küsimused seoses masinõppega	27
3.2 Protsessijuhtimise küpsuse mudeli koolitus	28
3.3 Intervjuu tegevjuhiga	29
3.3.1 Masinõppe Lõuend	31
3.4 Intervjuud andmeteaduse tiimijuhiga	31
3.5 Kratikoolitused	33
3.5.1 Kratisõbralik tarkvaraarendus	33
3.5.2 Tehisintellekti rakendamise 1. töötuba	34
3.5.3 Tehisintellekti rakendamise 2.töötuba	35
3.6 Analüüsi kokkuvõte	37

<b>4 Küsimuste koostamine</b>	<b>38</b>
4.1 Projekti võimalikkus	38
4.2 Ressursside kulu	39
4.3 Lahenduse kvaliteet ja toimimine	39
4.4 Detailid	40
<b>5 Koostatud küsimustele vastamine</b>	<b>41</b>
5.1 Projekti võimalikkus	41
5.1.1 Kas meil on masinõpet vaja või saab sama tulemust ka muude lahendustega?	41
5.1.2 Kas masinõpe aitab optimeerida mingit olulist näidikut (aeg, raha jne)?	43
5.1.3 Kas masinõpe on võimeline antud juhul lahendama probleemi?	43
5.2 Ressursside kulu	44
5.2.1 Kui kallis masinõppe lahendus on?	44
5.2.2 Millal lahendus end ära tasub? Mis on potentsiaalne kasum?	44
5.2.3 Millal me lahenduse saame?	45
5.2.4 Kui palju andmeid on selleks vaja?	45
5.2.5 Milliseid andmeid on vaja?	46
5.2.6 Kui palju meie inimeste aega selle projekti jaoks vaja on? Mis peaks olema meiepoolne panus?	47
5.3 Lahenduse kvaliteet ja toimimine	48
5.3.1 Kuidas garanteerite töökindla lahenduse?	48
5.3.2 Kui hästi see lahendus töötab?	48
5.3.3 Kuidas ma tean, miks mudel just sellise ennustuse tegi?	48
5.3.4 Kuidas ma tean kui hästi see lahendus töötab?	49
5.4 Detailid	49
5.4.1 Miks Python?	49
5.4.2 Milline meetod valida?	50
5.5 Töö tulemuste kokkuvõte	51
<b>6 Kokkuvõte</b>	<b>52</b>
<b>7 Viidatud kirjandus</b>	<b>53</b>
<b>8 Lisad</b>	<b>57</b>
8.1 Lisa 1. STACCI sisese info kogumise originaalmaterjalid.	57
8.2 Lisa 2. STACCI sisese info kogumise materjalid eesti keelde tõlgituna.	59
8.3 Lisa 3. Masinõppe Lõuend [37]	61
8.4 Lisa 5. Masinõppe projektiks valmistumine: kontroll küsimustik [54]	62

## Sissejuhatus

Mõni aasta tagasi teostatud turu-uuringu põhjal on tehisintellekt vankumatult populaarseim kasutatav ning arendatav strateegiline tehnoloogia, mille populaarsuse kasv viimastel aastatel on ilmselge [1]. Samast uuringust selgus, et ligi 60% ettevõtetest mõtlevad tehisintellekti kasutusele võtmisele, aga neid takistavad teadmatuses tulenevad hirmud, mistõttu ollakse hetkel veel informatsiooni kogumise faasis [1]. Kuigi tehisintellekti on arendatud ja protsesside automatiseerimiseks rakendatud juba aastakümneid, tõstatub endiselt inimestel palju küsimusi seoses tehisintellekti- ja masinõppepõhiste lahendustega. Inimesed ja ettevõtted kõhklevad, kas nad saaksid masinõppe süsteeme probleemide lahendamiseks rakendada ning millistes protsessi osades oleks seda võimalik teha. Samuti tekib tihti küsimus miks üldse masinõppel põhinevaid lahendusi kasutada. Lisaks spetsiifiliste äriprobleemidega seotud hirmudele kaasnevad selle kiiresti areneva ja võimeka tehnoloogia rakendamisega ka eetilised probleemid. Eesti tehisintellekti kasutuselevõtu eksperdirühma aruandes on välja toodud, et 67% eestlastest suhtub tehisintellekti kasutuselevõttu pigem ettevaatlikkusega [2]. Tänu võrdlemisi keerukale ning kiirelt arenevale valdkonnale on ettevõtjate hirmud ja küsimused seoses tehisintellekti lahendustega kaardistamata. Lähtuvalt eelnevalt kirjeldatud probleemist on käesoleva magistritöö eesmärgiks aidata andmeteaduse ettevõttel STACC OÜ kaardistada ning avada potentsiaalsete klientide jaoks masinõppel põhinevate süsteemide kohta esinevaid küsimusi ning hirme. Eesmärgi saavutamiseks on püstitatud järgmised uurimisküsimused:

1. Mis on ettevõtete peamised hirmud seoses masinõppega?
2. Mis on ettevõtete sagedasemad küsimused seoses masinõppega?
3. Kuidas efektiivistada STACCi klientide värbamise protsessi seoses potentsiaalsete klientide hirmude ja küsimustega masinõppe kohta?

Uurimisküsimustest tulenevalt rakendatakse vastuste leidmiseks järgmisi tegevusi:

1. Peamiste masinõppega seotud hirmude kaardistamine.
2. Sagedasemate masinõppega seotud küsimustega kaardistamine.
3. Kogutud andmete põhjal kõige olulisemate küsimuste moodustamine.
4. Moodustatud küsimustele vastamine.

Andmete kogumiseks viiakse läbi intervjuusid ja tööseminare ning analüüsitakse olemasolevaid dokumente. Kogutud andmete põhjal kaardistatakse kõige sagedamini esinevad masinõppega seotud teemad, küsimused ning hirmud, mille põhjal moodustatakse olulisemaid masinõppe projektidega seotud kategooriaid käsitlevad küsimused.

Käesoleva magistritöö oodatavaks tulemuseks on kõige sagedamini esinevate küsimustega materjali esmase versiooni välja töötamine, mida saab peale valideerimisprotsessi läbiviimist kasutada STACCi teenuste efektiivsemaks turundamiseks. Materjal peaks aitama iseseisvalt potentsiaalse kliendi hirme maandada ja klienti veenda, et pakutavad tehisintellekti lahendused aitavad tal oma äriprobleeme efektiivselt lahendada. Teisalt peaks materjal

potentsiaalsele kliendile andma aimu, kui lahendust otsitavat probleemi ei ole võimalik masinõppe abil lahendada ning lisaks tuleb katta teemad, mille peale klient ise ei pruugi tulla.

Töö koosneb viiest peatükist. Esimeses peatükis tutvustatakse probleemi sisu ning sellele eelnevalt antakse ülevaade tehisintellekti olemusest, rakendusvõimalustest ning sellega kaasnevatest probleemidest. Teises peatükis avatakse sisendi kogumiseks kasutatud meetodikaid ning kogutud info jaoks rakendatavaid analüüsimeetodeid. Kolmandas peatükis esitletakse erinevatelt sihtrühmadelt kogutud sisendit ning teostatakse materjali analüüs. Analüüsis paigutatakse kogutud masinõppega seotud küsimused, hirmud ning teemad eelnevalt eksperdi poolt valideeritud kategooriatesse. Neljandas peatükis moodustatakse kõige sagedamini esinenud küsimuste, hirmude ning teemade põhjal kõige olulisemad küsimused, millele leitakse vastused viiendas peatükis. Töö lisadest on leitavad ettevõtte sisese ajurünnaku tulemused nii originaalis inglise keelsena kui tõlgituna eesti keelde. Lisaks masinõppe lõuend ning Majandus- ja Kommunikatsiooniministeeriumi poolt välja antud masinõppe projektiks valmistumise kontrollküsimustik.

## Mõisted ja terminid

Andmestik	<i>Dataset</i> - Andmete kogum
Andmeteadus	<i>Data Science</i> - Ükskõik milline tegevus, mis aitab andmete põhjal kasulikke otsuseid teha.
Andmetriiv	<i>Data drift</i> - Andmete uuenemisel tekkiv paratamatu olukord, kus osad andmed kaotavad oma asjakohasuse ning mõjutavad masinõppe süsteemide tööd negatiivselt
GPU	<i>Graphics Processing Unit</i> - Graafikaprotsessor
Kitsas tehisintellekt	<i>Artificial Narrow Intelligence (ANI)</i> - Tehisintellekti liik, mida saab kasutada kindla raamistikuga piiratud ülesande lahendamiseks
Masinõpe	<i>Machine Learning (ML)</i> - Protsess, millega funktsionaalsus täiustab oma talitlust, omandades uut teadmist või uusi oskusi või korraldades ümber seniseid
Masinõppe projekti lõuend	<i>Machine Learning Canvas</i> - Masinõppe süsteemide kujundamise ja dokumenteerimise mall
Minimaalne töötav toode	<i>Minimum Viable Product (MVP)</i> - Toote algeline kuju, mis rahuldab klientide esmaseid vajadusi
Pidevintegratsioon	<i>Continuous Integration (CI)</i> - rakenduse pideva integratsiooni protsess
Pidevjuurutus	<i>Continuous Deployment (CD)</i> – rakenduse pideva juurutuse protsess (ilma inimese sekkumiseta)
Protsessijuhtimise küpsuse mudel	<i>Business Process Maturity Model (BPMM)</i> - Organisatsiooni juhtimisküpsust hindav mudel
Rakendusliides	<i>An Application Programming Interface (API)</i> - liides, mis võimaldab luua ühenduse erinevate programmide vahel
Superintelligentsus	<i>Artificial Superintelligence (ASI)</i> - Tehisintellekti liik, mille intelligentsus on igas valdkonnas kõrgem kui inimese oma
Suurandmed	<i>Big Data</i> - Väga suure mahuga andmestikud, mis ei mahu ühe arvuti mälli.
Tehisintellekt	<i>Artificial Intelligence (AI)</i> - Informaatika haru, tegeleb inimõistusele omaste funktsioonidega andmetöötlussüsteemide arendusega.

Tulemuslikkuse  
võtmemõõdik

*Key Performance Indicator (KPI)* - mõõdik, mis näitab  
ettevõtte eesmärkide saavutamise efektiivsust

Ärimudeli lõuend

*Business Model Canvas* - Ettevõtte väärtusloome  
kirjeldamiseks mõeldud mall

Üldine tehisintellekt

*Artificial General Intelligence (AGI)* - Tehisintellekti liik, mis  
on suuteline lahendama inimesele võimele vastavaid  
ülesandeid

## 1 Tausta kirjeldus

Alljärgnevas peatükis on avatud antud töös käsitletava probleemi tausta selgitades tehisintellekti ning masinõppe olemust ja võimalikke kasutusvaldkondi. Kuna töö üheks eesmärgiks on potentsiaalsete klientide masinõppeteemaliste hirmude ning küsimuste kaardistamine, avatakse antud peatükis ka tehisintellekti rakendamisega seotud mõjusid ning probleeme nii tehnoloogilisel kui eetilisel tasemel.

### 1.1 Tehisintellekti olemus

Esimest korda mainiti terminit “tehisintellekt” Ameerika arvutiteadlase John McCarthy poolt organiseeritud Dartmouth’i konverentsil aastal 1956. Tehisintellekti valdkond on viimase 60 aasta jooksul teinud läbi väga suure arengu [1].

Tehisintellekt on lai termin, millega viidatakse üha populaarsemaks muutuvale informaatika harule, kus kasutatakse nii arvutiteadust kui statistikat ning mille abil on võimalik kasvatada ettevõtetele loodavaid lisandväärtusi ja tõhustada tööprotsesside voogu nii avalikus kui erasektoris. Antud haru tegeleb suures pildis andmetöötlussüsteemide arendusega, mille eesmärgiks on täita inimõistusele omaseid funktsioone näiteks õppimine või arutlemine, märgitakse Eesti Keele Instituudi poolt hallatavas IT terministandardi sõnastikus [2]. Eesti tehisintellekti kasutuselevõtu ekspertrühma aruandes on tehisintellekti definitsiooni märgitud järgmiselt: *“ELis kasutusel oleva definitsiooni kohaselt mõeldakse tehisintellekti all süsteeme, mis ilmutavad intelligentset käitumist, analüüsid ümbritsevat keskkonda ja tehes teatud ulatuses iseseisvalt otsuseid, et saavutada teatud eesmäärke [2].”*

Tehisintellekti sünni põhjuseks võib pidada probleeme, mida on keeruline või isegi võimatu reeglite või registrite abil lahendada. Läbi aastakümnete on tehisintellektil olnud tehnoloogia arengus märkimisväärne roll. Mõned eksperdid eristavad nendes arenguetappides aktiivsemaid ja edukamaid aegu ning perioode, kus edulugude vähenemine põhjustas uue valdkonna populaarsuse seiskumist või isegi tõusvat ebapopulaarsust - antud perioode nimetatakse vastavalt “tehisintellekti talv” ja “tehisintellekti kevad” [1]. Laias laastus eristatakse kahte suuremat mõõnaperioodi, mida iseloomustab huvi kadumine, usalduse langus, investorite passiivsus ning sellest tulenev arengu aeglustumine [5]. Hetkel on tõenäoliselt maailmas taas toimumas tehisintellekti arengu positiivne periood [1]. Seda eeskätt tänu edasiarenenud algoritmidele ning uute tehnoloogiate nagu loomuliku keele töötamise, sügavate närvivõrkude ja suurandmete töötamise paljutõotavusele. Tõestus praegu käimasolevast nõ “tehisintellekti kevadest” ehk populaarsuse tõusust on ka see, et tehisintellekti tehnoloogiatel põhinevad funktsionaalsused on saanud loomulikuks osaks meie igapäevaelu ümbritsevas tehnoloogilistes süsteemides näiteks sotsiaalmeedias reklaamide suunamises, muusika rakendustes olevates personaalsetes kuulamissoovitustes ning digitaalsetes postkastides rämpsposti eraldamises. See omakorda tõstab nende populaarsust ja väärtust üha enam.

Tehisintellekti areng on muutnud ning muudab meie maailma nii praegu kui tulevikus. The Economist avalikustas 2017. aastal artikli sellest, kuidas maailma kõige väärtuslikum ressurss ei ole enam mitte nafta, vaid andmed [6]. Sellest saati on läinud laialt kasutusse mõte, et andmed on tänapäeva uus nafta. Gartneri 2018. aastal tehtud uuringu põhjal on tehisintellekt vaieldamatult kõige populaarsem strateegiline tehnoloogia [1]. Sama uuring näitas, et 59% osalenud ettevõtetest koguvad siiani informatsiooni selle kohta, kuidas tehisintellekti kõige efektiivsemalt enda kasuks rakendada, ülejäänud on aga juba kaugemale jõudnud ning tegelevad projektide arenduse või on jõudnud juba adapteerumise faasi [1]. Andmeid koguneb iga päevaga üha rohkem ning nähakse palju vaeva, et neist kätte saada seal peituv väärtus. Suurte andmekogude töötlemisel on praegu tehisintellektil põhinevad lahendused oma tehnoloogiaid tõestanud ning sellest tuleneb ka ettevõtete suurenev nõudlus antud tehnoloogiliste lahenduste järgi.

### 1.1.1 Tehisintellekti kategooriad ja alamvaldkonnad

Tehisintellekti valdkond ja tehnoloogia on nii lai, et üldjoontes jaotatakse see kohanemisvõime ja intelligentsuse taseme järgi kolme erinevasse kategooriasse: kitsas tehisintellekt, üldine tehisintellekt ja superintelligentsus [7].

Kitsas ehk “nõrk” tehisintellekt (ingl k *Artificial Narrow Intelligence (ANI)*) suudab lahendada üht konkreetset probleemi ning on võimeline õppima ja etteantud probleemi üha edukamalt lahendama [7]. Kõik tänapäeval rakendatavad tehisintellekti süsteemid liigituvad siia kategooriasse. Näiteks näotuvastus fotodelt, rämpsposti filtreerimine ja isejuhtivad autod. Kuigi programmid saavutavad neis ülesannetes väga häid ja kasulikke tulemusi, suudavad need siiski lahendada ainult konkreetset programmi ette antud probleemi ning enamjaolt vajavad need inimese supervisiooni [8]. Kitsas tehisintellekt jaotub mitmeks alamvaldkonnaks. Nende hulka kuuluvad näiteks loomuliku keele töötlus (ingl k *NLP-Natural Language Processing*), masinõpe (ingl k *ML-Machine Learning*), tehisnärvivõrgud (ingl k *ANN-Artificial Neural Network*), automatiseeritud planeerimine (ingl k *AI Planning/Automated planning*) ning ekspertsüsteemid või teadmistepõhised süsteemid (ingl k *Expert Systems/Knowledge-based systems*) [1].

Üldine ehk “tugev” tehisintellekt (ingl k *Artificial General Intelligence (AGI)*) suudab lahendada inimese intelligentsusele vastavaid ülesandeid [7]. Inglise keeles viidatakse tihti antud süsteemile kui inimtasemele vastavale tehisintellektile (ingl k *humal-level AI*) [8]. Erinevalt kitsast versioonist suudab see luua seoseid ning arvestada ühe probleemi lahendamisel ka selle ümber olevat konteksti. Selline tehisintellekt suudab mõelda laiemalt ühest konkreetsest probleemist, õppida, teha järeldusi ning kanda lahendust ühelt probleemilt teisele [7]. Praeguses maailmas sellist tehisintellekti ei eksisteeri [7].

Superintelligentsus (ingl k *Artificial Superintelligence (ASI)*) on kõrgeim tase, mil tehisintellekti oskused ja võimed ületavad kõiki inimvõimeid kaasaarvatud loomingulisus ja sotsiaalsus [8]. Superintelligentsus kõlatab ulmeliselt ja hirmuäratavalt. Küll aga tasub meeles hoida, et hetkel eksisteerib meie maailmas vaid kitsas tehisintellekt.

### 1.1.2 Tehisintellekti kasutusvaldkonnad

Andmeid kogutakse tänapäeval igas valdkonnas ning levib arusaam, et andmete abil on võimalik luua ettevõttele rohkem lisaväärtust. Erinevad tehisintellektil põhinevad lahendused on tänapäeval juba väga laialt levinud - neid kasutatakse erinevate suuri ressursse nõudvate protsesside automatiseerimiseks, müügikanalite optimeerimiseks, toodangu mahu ja kvaliteedi suurendamiseks, klienditeeninduse tõhustamiseks ning energia säästmiseks. Taoliseid lahendusi võetakse üha enam kasutusele nii era- kui avalikus sektoris - nii hariduses, meditsiinis, seadusandluses kui kaubanduses, tööstuses, turisminduses ja finantssektoris [9].

Tänapäeval kogutavad virtuaalsed andmemahud nõuavad võimsaid ressursse, mille abil andmeid töödelda ning nendest seal leiduv väärtus kätte saada - seda võimaldavad erinevad tehisintellektil põhinevad lahendused. Samuti hakkab aina rohkem riike kasutama tehisintellekti kuritegevusega võitlemisel [10]. Suur osa elust on kolinud veebi ning hetkel tehakse jõulisi samme ka veebi kuritegevuse piiramiseks. Näiteks paljud veebiplatvormid kasutavad erinevaid tehisintellekti lahendusi soovimatu veebikäitumise tuvastamiseks [10].

Suur tehisintellekti haru tegeleb tervishoiu valdkonnas - kõrgete sissetulekutega riikides juba kasutatakse meditsiinis erinevaid nutikaid tehnoloogilisi lahendusi [1]. Neid protsesse analüüsid on leitud, et näiteks aastaks 2026 säästetakse Ameerika Ühendriikides taoliste lahenduste toel igal aastal riiklikus tervishoius ligi 150 miljardit dollarit ning nüüd tahetakse saavutada taolisi lahendusi rakendades samu tulemusi ka vaesemates riikides [1]. Seda soovitakse teha läbi digitaalsete terviseluude (ingl k *EMR-Electronic medical record*), pilveteenuste kasutamise ning mobiilse tervise (ingl k *mHealth*) [1].

### 1.1.3 Tehisintellekti kasutamine Eestis

Eestis võeti riiklike tehisintellekti kaasavate projektide raames kasutusele mütooloogilistest pajatustest tuntud krati metafoor, mille eesmärgiks on lihtsustada kommunikatsiooni [11]. Riigikantselei digiinnovatsiooni nõunik Marten Kaevats selgitas otsust järgmiselt: *“Kui tehisintellekt on mitte-eksperdi jaoks miskit äärmiselt keerukat ja salapärast, siis muinasjuttudest tuttav kratt aitab seda uute võimaluste välja hõlpsamini mõista. Ja nagu ütlesin, peame oma kratid tegema õiglasel ja kasulikud ning hoidma silma peal, et need kurja ei teeks [12].”* Kratik nimetatakse antud kontekstis erinevaid infotehnoloogilisi tehisintellekti süsteeme, mis põhinevad õppimisvõimelistel algoritmidel ning mille eesmärgiks on automatiseerida rutiinseid toiminguid, mida inimese asemel saaks teha masin [2]. Tehisintellekti lahendustel puudub kindel programmiloogika, mis on esindatud klassikalistes tarkvaralahendustes, kus arvuti täidab talle programmeerija poolt määratud käsked - erinevus seisneb selles, et kratile ei pea programmeerija täpset koodi lahendamiseks ette andma [2]. Tema eesmärk on jõuda erinevaid masinõppe algoritmide kasutades õige lahenduseni ise.

On väga selge, et tehisintellekti lahendused on tulevikus laialt kasutusel ning need lihtsustavad meie igapäevaelu juba täna. Eestis loodi Riigikantselei juurde tehisintellekti ehk

krati töörühm, mille eesmärgiks on ühelt poolt leida võimalusi, kuidas tehisintellekti tehnoloogiate läbi meie digiriigile kasu tuua ning teisalt kaitsta Eestit taoliste lahenduste rakendamiseга seotud ohtude eest [11]. Tehisintellektil põhinevaid lahendusi rakendavad hetkel avalikus sektoris näiteks Riigi Infosüsteemi amet, Maanteeamet, Põllumajanduse Registrite, Informatsiooni Amet, Politsei- ja Piirivalveamet ning lisaks on ka palju alles väljatöötamisel olevaid projekte [11].

## 1.2 Masinõpe

Tehisintellekti tehnoloogiatest väärib oma rohke kasutatavuse ja laia levikuga enim tähelepanu masinõppe meetod, mille käigus on süsteem võimeline andmete põhjal õppides lahendama teatud ülesandeid ilma, et lahendamise algoritm oleks süsteemi sisse programmeeritud ehk siis ilma reegleid teadmata [7]. Nii riigi- kui eraettevõtete huvi masinõppe lahenduste vastu on märkimisväärselt suur - sellelt oodatakse suurt kasumlikkust. Masinõppe programmid suudavad eristada parima võimaliku valiku, kuna tänu kindlatele reeglitele ja suurele arvutusvõimsusele kaaluvad need läbi väga suurel hulgal erinevaid variatsioone ning teevad siis kõige parema võimaliku valiku [7]. Läbi selle on need võimelised mingites valdkondades ka inimesest efektiivsemaid otsuseid tegema. Lihtsustatult on masinõppe arvutitele täpseid käitumisjuhiseid andmata nende targemaks muutmine ehk õpetamine.

Kui vaadata ajalukku, siis kõige esimese iseõppiva arvutiprogrammi töötas välja Arthur Lee Samuel juba 1949–1959 aastatel - see seisnes kabemängu õppimises [13]. Algoritmi vastaseks oli algoritm ise ning tuhandeid kordi mängides õppis see kabe nii hästi ära, et võitis 1962. aastal ka inimesest kabetšempionit [13]. Samuelist on veel tuntum IBMi poolt loodud masinõppel põhinev algoritm Deep Blue, mis saavutas ülemaailmse tuntuse 1996. aastal võites males kuulsat omaala meistrit Garri Kasparovit [13]. Ajalooliseks hetkeks masinõppe arengus sai 2015. aastal Google DeepMind'i poolt loodud arvutiprogrammi AlphaGo võidud strateegilises kabest ja malest keerukamaks peetavas lauamängus Go Euroopa meistri Fan Huid'i vastu [14]. AlphaGo tehisnärvivõrkudel põhinev arvutiprogramm jäljendab tehislkult inimaju mõtlemisprotsessi ning on võimeline ilma inimese sekkumiseta õppima, analüüsima ning koostama strateegiat ja teha taktilisi otsuseid ning kuulsas 2015.aastal toimunud mängus võitis arvuti meistrit tulemusega 5:0 [14].

Masinõppes kasutatakse programmeerimist, matemaatikat ja statistikat. Seda jaotatakse omakorda juhendatud (ingl k *supervised learning*) ning juhendamata õppeks (ingl k *unsupervised learning*) ja stiimulõppeks (ingl k *reinforcement learning*) [7]. Joonisel 1 on illustreeritult ära toodud tehisintellekti ja masinõppe jagunemised.

Järgnevalt on masinõppe jaotused lahti seletatud toetudes Taivo Pungase [15] avaldatud materjalidele. Laialt võib eristada kahte eraldiseisvat alamvaldkonda - reeglipõhised ja masinõppel põhinevad süsteemid. Reeglipõhised süsteemid vajavad erinevalt masinõppest ette antud loogikareegleid, mille põhjal lahenduseni jõuda. Juhendatud õppe eesmärk on luua

mudel, mis ennustab midagi andmepunkti põhjal. Näiteks kui mudel teab korteri asukohta, pindala ning hinnangut seisukorra kohta, on võimalik ennustada korteri hinda. Juhendamata õppe puhul ei ole mudelile ette antud, mida konkreetselt ennustada - eesmärgiks on leida andmetest mingit teatud struktuuri. Näiteks kas nähtust esineb tihti või harva? Või kas nähtusi saab kuidagi loogiliselt gruppidesse jaotada? Stiimulõppe läbi on võimalik arvutit panna mingit sooritust paremini tegema. Näiteks õpetada arvutit Tetrise mängu efektiivsemalt mängima. Stiimulõppe käigus toimuvad erinevalt juhendamata ja juhendatud õppes nii õppimisfaas kui õpitu rakendusfaas samal ajal.



Joonis 1. Tehisintellekti alamvaldkonnad [7]

Kui vaadelda maailmas virtuaalseid assistente näiteks Siri, Alexa või Google Assistant, siis tajub ka tehnoloogiast mitte väga teadlik inimene, et neis kasutatav tehnoloogia on väga arenenud. Need rakendused on võimelised kuulama ning mõistma inimkeelt ning seeläbi leidma inimese probleemidele lahendusi ning täita inimese poolt ette antud ülesandeid. Viimaste aastate enim kasutatud masinõppe meetod on aga sügavõpe (ingl k *DL* ehk *deep learning*) ehk tehisnärvivõrkude rakendamine, mida on võimalik rakendada nii juhendatud, juhendamata kui stiimulõppes [16]. Tehisnärvivõrgud on inimese ajast inspireeritud ning meetodi eesmärgiks on panna masin õppima väga suurtest andmehulkadest [16].

### 1.2.1 Masinõpe vs tehisintellekt

Mõisteid tehisintellekt ja masinõpe kasutatakse uute tehnoloogiate arengul üha tihedamini. Siiski ei ole need kaks mõistet sünonüümid ning tihti kasutatakse neid läbiseigi täpset erinevust teadvustamata.

Masinõppe ja tehisintellekti mõistete ebakorrektselt kasutamist soodustab nende kahe termini vaheline tugev seos. Tehisintellekt on laiem termin, millega tähistatakse masinate võimekust viisil, mida inimesed peavad intelligentseks ning masinõpe on selle tehnoloogia üks

rakendamise võimalusi, mis paneb masinad iseseisvalt õppima [6]. Ehk siis tehisintellekt on arvuti võimekus jäljendada inimese kognitiivseid võimekusi nagu näiteks õppimine ja probleemide lahendamine. Masinõpe on tehisintellekti rakendamise üks viisidest. Masinõppe käigus kasutatakse matemaatilisi andmemudeleid, eesmärgiga panna arvuti ise õppima ilma sellele konkreetseid juhiseid andmata [17].

### **1.3 Tehisintellekti mõjuvõimu kasvuga kaasnevad probleemid**

Tehisintellektil põhinevad lahendused ümbritsevad juba praegu meie igapäevarutiini. Ekspertid leiavad, et lähiaastatel saab antud tehnoloogiate mõju ühiskonnas olema üha suurenev, kuna selle täpsus ja töökindlus paranevad tänu andmete kvaliteedi ning kättesaadavuse paranemisele [18].

#### **1.3.1 Tehisintellekti mõjuvõimu kasv**

Alljärgnev lõik tugineb väljaande *The New York Times Magazine* 2012. aastal avaldatud artiklile, milles käsitleti aasta varem Ameerika Ühendriikides toimunud vahejuhtumit, mis on tekitanud palju kõneainet just seoses tehisintellekti mõjuvõimuga turundusvaldkonnas ning eetikaküsimustes [19]. Andrew Pole, kauplusketi Target andmeteadlane, lõi mudeli, mis suutis läbi kliendi isikuandmete ning ostuandmete ennustada naiste rasedust eesmärgiga pakkuda personaliseeritud reklaame. Poekett oli mudelit personaalseks turundustöök kasutanud umbes aasta, kui kurikuulus vahejuhtum aset leidis. Minneapolisise läheduses asuvasse Targeti kauplusesse astus maruvihane meesterahvas, kelle pahameel oli suunatud poeketi turunduse pihta. Target oli saatnud tema teismelisele tütrele personaalse sooduskupongide pakkumise, mis lubas soodsamalt soetada väga kindla temaatikaga kaupu nagu näiteks lutid, mähkmed ja beebiriided. Asja uuriti ning kui poe juhataja paari päeva pärast mehega vabandamiseks ühendust võttis, tuli välja, et tema teismeline tütar oli tõepoolest rase. Antud vahejuhtumi puhul tuleb märkida, et nooruk ise ei olnud rasedusega seotud sooduskaupu otsinud ega uudistanud. Tema seisundi ennustas mudel läbi keerukate algoritmide, mis analüüsisid isiku ostukäitumisi. Seal esinesid muustrid, mis lubasid mudelil ennustada tooteid, mida klient potentsiaalselt võiks vajada. Targeti mudelil oli ennustuse tegemiseks kasutada kolm allikat: e-posti andmed, krediidiinfo ning sooduskupongide kasutamise andmed [19].

Johnson ning Gutman käsitlevad oma raamatus “Pimenägemine. (Enamasti) varjatud viisid, kuidas turundus meie aju kujundab” tehnoloogia, turunduse ja psühholoogia vahelisi põimumisi, mida nähakse tuleviku turundusstrateegiana [20]. Hetkel peetakse Suure Viisiku isiksusejoonte mudelit isiku kujutamises ja ennustustäpsuses kõige võimekamaks [20]. Mudeli inglise keelne tuntud akronüüm OCEAN võtab kokku viis kõige olulisemat aspekti: avatus kogemusele (ingl k *openness*), meeledindlus (ingl k *conscientiousness*), ekstraversus (ingl k *extraversion*), sotsiaalsus (ingl k *agreeableness*) ning neurootilisus (ingl k *neuroticism*) [20]. Suure Viisiku isiksuseomaduste eestikeelsed tõlkevasted inglise keelsetele omadustele on võetud Sirli Kanguri [21] magistritööst. Antud mudel on ettevõtete jaoks suure

väärtusega, kuna edukama turundustöö aluseks on kliendi parem tundmine [20]. Mida rohkem detaile teatakse kliendi kohta, seda rohkem andmeid saab kasutada ennustusmudelites ning seda täpsemad on ennustused. Üks tuntumaid ning efektsemaid kuid negatiivsemaid näiteid taoliste andmete kasutamisel oli 2016. aasta Ameerika Ühendriikide presidendivalimiste kampaania, kus Donald Trumpi turundusmeeskonna palgatud andmeteaduse ettevõtte Cambridge Analytica andmete abil Trumpi kampaaniat toetama ja võimendama. Läbi küsimustiku, millele vastas 270 000 inimest, ning tänu Facebooki nõrgale turvalisusele koguti seaduslikult 87 miljoni inimese andmeid - alates Facebooki üles laetud kasutajainfost kuni meeldimiste ning isiku tutvusringkonnani [20]. Kogutud andmete ning Suure Viisiku isiksuse profiili põhjal paigutati inimesed kolme suurde kategooriasse: kindlad toetajad, tugevad vastased ning need, kes ei pooldanud ega olnud vastu [22]. Kuigi kõigile gruppidele suunati läbimõeldud sisuga reklaame, pandi kõige suurem rõhk kahe äärmuse keskele jäävale massile, kuna nende hääled olid need, mis osutusid määravaks [22]. Taolised juhtumid suurendavad märgatavalt hirmu ja usaldamatust niivõrd võimsate tehnoloogiate ees.

### **1.3.2 Praktilised probleemid**

Vaadates tehisintellekti massilist kohalolekut meid ümbritsevad maailmas, võib kergesti tekkida tunne, et masinad on varsti nii targad, et inimene ei peagi enam midagi tegema. Tegelikult jäävad praegu eksisteerivad ja rakendatavad algorütmid pigem nõrga tehisintellekti valdkonda. Hetkel ei ole veel jõutud selleni, et masinad suudaksid päris ise mõelda. Kuid igal asjal on ka oma varjukülg. Tehisintellekti rakendavates tarkvaraprojektides on toodud välja erinevaid keerukusi majanduslikes, poliitilistes, materiaalses ning organisatoorses aspektides [23].

Kuigi antud tehnoloogiad on tegelikult juba aastaid arendatud, on need ikkagi võrdlemisi uued ning väga kiiresti edasi arenevad. Sellest tulenevalt puuduvad seaduslikud kontrollraamistikud ning järevalve - neid puudusi peetakse võrdlemisi suureks riskiteguriks taoliste tehnoloogiate arendamisel [14]. On selge, et kui tehisintellekt teeb mingeid otsuseid, siis masin ise nende otsuste eest ei vastuta. Siinkohal tekib küsimus, et kes lõpuks täpselt vastutab juhul, kui midagi valesti peaks minema [14]? Suureks ohuks hinnatakse puudulikku arusaama tehisintellekti otsustest ehk tulemuste tõlgendamisest [23]. Tihti on viidatud taoliste mudelitele kui "mustadele kastidele", mille sisu ei ole hoolimata õigest vastusest tõlgendatav ei kasutajale ega loojale - see omakorda mõjutab märkimisväärselt taolise tehnoloogia usaldusväarsust selle rakendamisel [23].

Majanduslikult on põhiliseks pudelikaelaks nii uue ja kalli tehnoloogia rahastamine. Eriti keerukaks teeb selle tihti asjaolu, et erinevalt klassikalisest tarkvaraarendus projektist, ei ole enamasti tehisintellekti projekti tulemus nii selge ja kindel. Tulemuste jaoks on vaja näha ja tunda projektis kasutatavaid andmeid, neid aga ei saada juriidilistel põhjustel enne koostööd kinnitavate lepingute allkirjastamist avaldada. Tihti on enne projektiga alustamist teadmata olemasolevate andmete kogus ja kvaliteet. Siit järgnevad ka organisatoorsed ja juriidilised küsimused tundlike andmete turvalisuse kohta [14]. Probleemi lahendamiseks hakkas Euroopas alates aastast 2018 kehtima Euroopa Liidu isikuandmete kaitse üldmäärus, mis loob

isikuandmete kaitse normidele õigusliku raamistiku, millega kehtestatakse suunised isikuandmete töötlemiseks Euroopa Liidus ja teatud tingimustel ka väljaspool Euroopa Liidu territooriumi [24]. Määruses on nõutud, et kõikidele automeeritud tehisintellekti poolt tehtud otsustele on teenuse tarbijal õigus saada antud valiku põhjendus. Ekspertide hinnangul on selliselt sõnastatult antud määrust pea võimatu täita, seega ei teeni vastu võetud määrus oma eesmärki ning vajaks põhjalikku analüüsi ning ümber sõnastamist [24].

### **1.3.3 Eetilised probleemid**

Tehisintellekti rakendamisega seotud eetilised küsimused tõstatuvad üha enam esile ning mõned teadlased peavad neid juba praegu selle tehnoloogia arengu kõige suuremaks probleemiks [1]. Siinkohal arvatakse, et riikide valitsuste mõju on tehisintellekti arenguga seotud riskide maandamisel kriitilise tähtsusega [1].

Alljärgnevas lõigus tuuakse eetilistesse probleemide näited Tara Qian Sun'i ning Rony Medaglia [14] läbiviidud uuringu põhjal, mille käigus kaardistati tehisintellektiga seotud murekohti avalikus tervisesektoris. Enamus projektidega seotud erinevate valdkondade osapooli toovad tehisintellekti lahenduste rakendamisel eraldi välja kaks põhilist tegurit. Esiteks, avalik üldsuse usaldus taolise keeruka tehnoloogia suhtes. Kui arst paneb patsiendile diagnoosi tehisintellekti tehnoloogia abil saadud tulemuse põhjal, on arsti usaldusväärsus tehnoloogia suhtes äärmiselt oluline. Samuti on oluline, et patsiendil oleks taolise süsteemi vastu usalduslik tunne - enamasti soovivad patsiendid mure korral siiski isiklikult arstiga suhelda, kuna see tundub neile kindlam viis oma tervise eest hoolitsemiseks ning abi saamiseks. Usaldus tekib tänu kommunikatsioonile ning siinkohal ei aita usaldust tehisintellekti süsteemidesse tõsta ka fakt, et masinatel puuduvad inimlikud omadused nagu empaatia, eetika ning kaastunne. Nii nende kui ka tehnoloogiliste põhjuste tõttu täidavad tehisintellekti süsteemis toetavaid rolle - viimane otsus on ikkagi arsti, mitte masina vastutusel. Teiseks, ebaeetiline andmete jagamine. Tehisintellektil põhinevate lahenduste üheks kõige olulisemaks alustalaks on kvaliteetsed andmed. Tänapäeval on andmete kogumisest saanud viis, kuidas oma ärile suur lisaväärtus luua - seega on andmete väärtus väga suur ning sellega kaasneb omakorda suur risk andmete kuritarvitamiseks.

Järgnevad mõtted tuuakse Johnsoni ja Ghumani [20] raamatust, milles nad kasutavad sotsiaalmeediarakendustest kirjutades mõistet reklaamijate võidurelvastumine. Võetakse kasutusele erinevaid uusi tehisintellekti rakendavaid tehnoloogiaid, mille psühholoogiline mõju on tarbijale üha suurem. Kuna konkurents on tihe, tuleb massist eristumiseks teinekord ka ekstreemsustega silma paista. Arvatakse, et isikustatud reklaamid on järgmine areng turunduses. Süvavõltsinguga on ühiskond kokku puutunud juba mõned aastad tagasi nii piltide, videote kui valeuudiste näol ning see on ühiskonnas tekitanud pahameelt, hirmu ja hoogustanud arutelu taoliste tehnoloogiate eetilise üle. Hirm on üks tugevamaid manipulaatoreid, millega inimest mõjutada saab. Kui kasutada inimese enda nägu videos, kus inimene põleb magades majja sisse, on tõenäoline, et suitsuandurit on sellele konkreetsele inimesele võrdlemisi lihtne maha müüa.

Austria, Saksamaa ning Ameerika Ühendriikide teadlaste koostöös tehtud uurimus näitas, et kui avalike teenuseid pakkuvad tehnoloogiad, mis on disainitud visuaalselt inimesest inspireerituna, siis mõjutab see inimeste käitumist nende suhtes, teisalt aga hinnatakse robotit ebaeetilise käitumise puhul karmimalt, kui seda tehakse inimeste puhul [9]. Samuti tuletavad robotite ebainimlikud käitumismaneerid inimesele veelgi teravamalt meelde, et tegu on tehnoloogiaga [9]. Uuringu tulemusena tõdeti, et robotite käitumise eetilisus peaks olema osa suurest avalikust debatist ning inimesed peaksid selle üle laiemalt arutlema [9].

### **1.3.4 Inimeste suhe tehisintellektiga**

Eesti tehisintellekti kasutuselevõtu eksperdirühma aruandes on välja toodud, et 67% eestlastest suhtub tehisintellekti kasutuselevõttu pigem ettevaatlikkusega [2]. Teadusuuring on näidanud, et inimesed suhtuvad tehisintellekti sarnaselt nagu teistesse inimestesse [9]. See tähendab, et tehnoloogialt oodatakse usaldusväärust ning eetilisust. Tõdetakse, et inimeste usaldus taolise tehnoloogia vastu on oluliseks mõjutavaks faktoriks selle tehnoloogia arengus [9]. Usalduse tõstmiseks tuleb uue tarkvara arendamisel silmas pidada turvalisust ning lahenduse kasutajakogemuse disaini [9].

Teaduse ja tehnoloogia maailma suurkujud Stephen Hawking, Elon Musk ja Bill Gates on tehisintellekti edasiste arenguetappide suhtes pigem skeptilised ning ettevaatlikud [25, 26]. Professor Hawkingu sõnul oleks tehisintellekti täielik areng väga potentsiaalne oht inimkonna tulevikule juhul, kui ühel päeval leiutatakse inimesega võrdne või isegi inimese intelligentsi ületav tehnoloogia [25]. Siinkohal mõtleb Hawking tugeva ehk üldise tehisintellekti arengut, mida tänapäeval maailmas veel ei eksisteeri. Nõrga ehk kitsa tehisintellekti senise arenguga on professor pigem rahul - tema sõnul on see tehnoloogia inimkonnale palju kasu toonud ning üldist heaolu parandanud [25]. Oma muret tehisintellekti arengu pärast on väljendanud ka Microsofti asutaja Bill Gates, kes oponeeris sel teemal Microsofti tehisintellekti valdkonna juhtivteadlase Eric Horvitz'iga [26]. Horvitz, kes ei pea tehisintellekti arengut absoluutselt mingisuguseks ohuks inimese elule Maal [26]. Esialgu teevad lihtsamatel tehnoloogiatel töötavad masinad inimese eest palju tööd ära ning see on positiivne - küll aga tõuseb masinate intelligentsuse probleem päevakorda mõnekümne aasta pärast, arvab Gates [26].

Stamfordis asuva tehnoloogiliste uuringute ja konsultatsioonifirma Gartneri asepresident ning tunnustatud analüütik Whit Andrews [1] leiab, et on võrdlemisi loomulik ja oodatav, et tehisintellekti areng toob palju küsimusi, takistusi ning hirme. Ta seletab seda paratamatusega, mis kaasneb iga uue tehnoloogia arenguga. Küll aga on ta positiivselt meelestatud, kuna praegused piloot projektidest saadud õppetunnid tulevad tulevikus suurematele ja ametlikutele tehisintellekti arendamise projektidele palju kasuks.

## **1.4 Probleem**

Andmeteaduse tähtsus tänapäevases ettevõttlusmaailmas tõuseb iga päevaga. Üha enam püüeldakse tõenduspõhise otsustamise poole (ingl k *Data-Driven decision making* or *DDDM*), mis näib olevat justkui edu võti. Tõenduspõhises otsustamises rakendatakse valikute

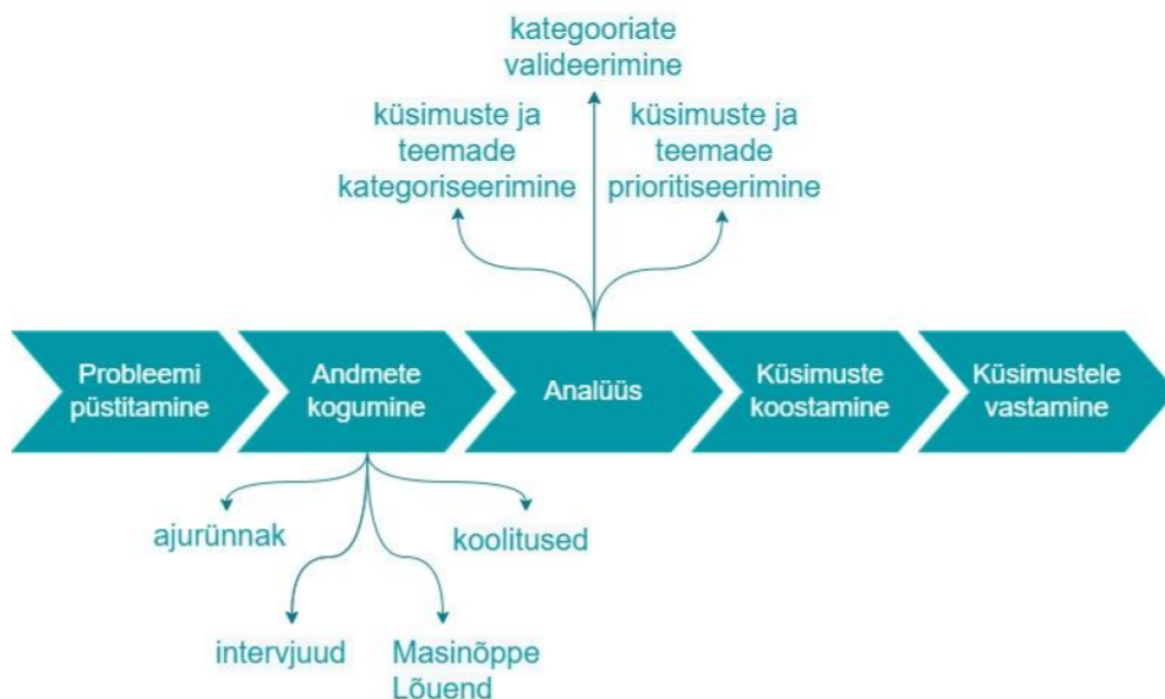
tegemise põhjendamiseks intuitsiooni ning vaatluse asemel reaalseid andmeid ja statistikat [27]. Kuna andmeid tekib tänapäeva ühiskonnas hiiglaslikes kogustes, tuleb nendest kasu saamiseks appi võtta uusimad tehnoloogiad. Küll aga võivad taolised lahendused olla ka võrdlemisi riskantsed. Lisaks kõlab see suurele osale inimestele võõralt ja seetõttu ka veidi hirmuäratavalt ning väga kulukalt. Eelnimetatu on põhjuseks, miks enne andmeteaduse ja tehisintellekti rakendamist soovitakse enne teema ja võimalustega tutvuda ning sellega kaasnevatest võimalikest kasu- ja kahjutegurites kindlad olla.

STACC on firma, kes pakub ettevõtjale andmeteaduse lahendusi ning aitab firmadel oma kogutud andmeid rakendada ning nendest suuremat kasumit toota. Asutuse klientideks on ettevõtteid erinevatest sektoritest näiteks e-kaubandus, tööstus ja tootmine, online-meedia, telekommunikatsioon, metsamajandus, finantstehnoloogia ja personaalmeditsiin [28]. STACC kasutab oma arendusmeetodina agiilset meetodit, mille eesmärk seisab töökorralduses, mis rõhub töötava lahenduse väljatöötamist ning võimekust tulla toime erinevate paratamatute muutuste ning ootamatustega projekti vältel [29]. Kiirelt areneva andmeteaduse firma tegeleb STACC hetkel väga aktiivselt uute potentsiaalsete klientidega suhtlemisega. Iga kord, kui STACC tegeleb uute klientide värbamisega, tekivad klientidel tehisintellekti ning masinõppe kohta küsimused. Tihti need küsimused korduvad. STACCis puudub hetkel analüüs taolistest korduvatest küsimustest ning enamasti tegeletakse küsimuste avamisega erinevate inimeste poolt nõ puusalt ning puudub ühtne viis, kuidas klientidele põhitõdesid selgitada.

Antud töö eesmärk on kaardistada tehisintellekti lahendustega kaasnevad olulisemad teemad ja küsimused, mida oleks võimalik tulevikus kasutada STACCi toodete ja teenuste efektiivsemaks turustamiseks loodud materjalides. Loodud materjalid peaksid aitama iseseisvalt potentsiaalse kliendi hirme maandada ja klienti veenda, et pakutavad tehisintellekti lahendused aitavad tal oma äriprobleeme lahendada.

## 2 Metoodikad

Antud töö on teostatud kvalitatiivse uuringuna, mille käigus kasutatakse andmete kogumiseks intervjuusid, tööseminare ning olemasolevate dokumentide analüüsi. Järgneval joonisel 2 on näha töö koostamiseks läbi viidud etapid, mille käigus rakendatud erinevatest meetoditest andmete kogumiseks ning sisendi analüüsimiseks antakse allolevas peatükis täpsem ülevaade.



Joonis 2. Magistritöö etapid

### 2.1 Andmete kogumise meetodid

#### 2.1.1 Ajurünnak

Ajurünnak (ingl k *brainstorming*) on tänapäeval rohkelt rakendatav meetod, mida enamasti kasutatakse koostöös mitme inimesega, kuid ei ole välistatud teostada ajurünnakut vaid ühe inimesega [30]. Ajurünnaku eesmärk on julgustada inimeste loovust ning tuua kiiresti välja palju erinevaid probleemi võimalikke lahendusi [31]. Antud meetodit saab kasutada spetsiifilise probleemi või küsimuse lahendamiseks, uue teema tutvustamiseks, inimestes huvi äratamiseks või inimeste teadmiste ja suhtumise tajumiseks [31]. Ajurünnaku käigus pakutavad lahendused ei pea olema valideeritud, kuna eesmärgiks on kvaliteedi asemel kvantiteet. Mida rohkem ideid välja pannakse, seda suurema hulga seast on võimalik valida päriselt sobiv lahendus. Parima lahenduse leidmiseks kasutatakse pakutud ideede prioritseerimist või klassifitseerimist. Ajurünnak toimib efektiivsemalt, kui osalejatele on loodud võimalikult kriitikavaba õhkkond - see ajendab inimesi julgelt välja pakkuma ka

ebareaalsete lahendusideid, millest omakorda teised osalejad võivad inspiratsiooni saada ning veel ideid genereerida [30].

Antud magistritöö raames korraldati 20.01.2022 STACC OÜ kontoris ettevõttesisene ajurünnak, mille käigus selgitati slaidiesitluse saatel lahti lahendatav probleem ning töö eesmärk. Ürituse käigus viidi läbi küsitlus, mille anonüümsetest vastustest saadi väärtuslikku infot eristamiseks kõige olulisemaid pidevalt korduvaid küsimusi, millele vastamine võiks aidata STACCil tulevikus efektiivsemalt kliente värvata. Üritusel osales 15 inimest, kellel on kogemus andmeteaduse projektides. Kohal olid andmeteaduse tiimijuhid, andmeteadlased, andmeinsenerid ning turundus- ja kommunikatsioonijuht. Küsitlusele vastamiseks kasutasid osalejad arvuteid või nutitelefoni ning soovi korral oli võimalik sisestada mitu vastust. Üritus viidi läbi inglise keeles.

### 2.1.2 Intervjuu

Intervjuu (ingl. k *interview*) on kvalitatiivses uurimuses üks enim kasutatud andmekogumismeetod. Bonnie S. Brennen on intervjuud defineerinud kui „*mingile teemale keskenduvat eesmärgipärast vestlust kahe või enama inimese vahel [32]:27*“. Intervjuu osalisteks on intervjuueeritav ning intervjuueerija - sõltuvalt eesmärgist võib mõlemad osapooli olla üks või mitu [33]. Intervjuu eesmärk on vastata ühele või mitmele uurimisküsimusele [32].

Struktureerimata individuaalintervjuu STACCi tegevjuhi Kalev Koppeliga viidi läbi STACCi kontoris 20. jaanuaril aastal 2022. Intervjuu eesmärgiks oli saada ülevaade ja tagasiside plaanitavast materjalide kogumikust, mille sisu antud töö välja töötab. Tegevjuhilt sooviti uurida tema arvates kõige tihedamini küsitavaid küsimusi kliendi poolt, millele vastamisel saaks välja arendatav materjal abiks olla. Tegevjuht puutub palju klientidega kokku just projektide algfaasis ning seega on temal kogemus just uuritavate teemadega. Intervjuu kestis umbes 40 minutit.

Lisaks tegevjuhile viidi antud töö raames läbi struktureerimata individuaalintervjuud ka STACCi andmeteaduse tiimijuhi Karl-Oskar Masinguga. Intervjuud viidi läbi STACCi kontoris 3. detsembril 2021. aastal ning 19. jaanuaril aastal 2022. Masingul on aastatepikkune kogemus nii andmeinseneri, andmeteaduse projektijuhi, analüütiku kui andmeteaduse tiimijuhi rollis ning ta omab seetõttu laia vaadet sagedamini tekkivatele kliendi hirmudele ja küsimustele. Töö koostamise hetkel juhib ta loomuliku keeletötlusega tegelevat meeskonda ning on projektijuhi rollis nii suuremates kui väiksemates masinõppe projektides. Antud töö raames on oluline saada lisaks tegevjuhile, andmeteadlastele ja andmeinseneridele ka projektijuhi vaatenurk, kuna selles rollis puututakse tihti kokku klientide hirmude, murede ning küsimustega.

### **2.1.3 Koolitustel osalemine**

Koolitustel ja ettekannetel osalemine on hea võimalus laiendada oma teadmistaasi erinevatel teemadel ilma selleks mingeid eelteadmisi omamata. Pikad ja madala interaktiivsusega loenguid peetakse vähem efektiivseks kui lühikesi ning kuulajaskonnaga rohkem suhtlemist pakkuvaid koolitusi ning esitlusi [34]. Antud magistritöö raames osaleti tehisintellekti projekti läbiviimise koolitussarjas ning lisaks läbiti iseseisvalt veebis protsessijuhtimise küpsuse mudeli koolitus, mille käigus käsitleti ka projektide alustamisel tehtud tüüpilisi vigu.

Protsessijuhtimise küpsuse mudeli koolituse viis läbi Mihkel Lauk 2015. aastal toimunud Nortal TechDay 2015 ürituse raames. Mihkel Lauk oli ettekande tegemise ajal PwC Advisors IT valdkonna konsultatsioonide osakonna juht ja e-valitsemise, IT-teenuste haldamise ja tulemuslikkuse juhtimise ekspert. Koolitust oli võimalik hiljem läbida, kuna see on internetist järelvaadatav ning toimus inglise keeles. Koolituse eesmärgiks oli lisaks protsessijuhtimise küpsuse mudeli käsitlemisele ka projektide alustamisel tehtud kõige tüüpilisematest vigadest - need teemad aitasid antud magistritöö sisu laiemalt avada ja probleeme kaardistada.

Lisaks osaleti ajavahemikus 11.01 - 14.02.2022 kolmeosalisel STACC OÜ poolt läbi viidud Kratikoolituste seerial, mis keskendus masinõppega varem mitte kokkupuutunud projektijuhide informeerimisele ning juhendamisele. Koolituste eesmärk oli teha ülevaade andmeteaduse protsessidega masinõppe projektides, tutvustada taoliste projektide tehnilise poole tähtsamaid nüansse, demonstreerida masinõppe projekti ärilise poole analüüsi ning selgitada andmetega seotud protsesse. Lisaks teoreetilisele osale, kaasnes iga koolitusega ka praktiline pool, kus osalejad said proovida erinevaid andmeteadusega seotud ülesandeid. Koolitusseerial osalemine avas spetsiifilisemalt problemaatilisi aspekte just masinõppega seotud projektides. Kokkuvõttes oli koolitussari antud magistritöö raames väga kasulik, kuna seal anti ülevaade masinõppe töövoost, tähtsamatest terminitest ning kajastati üldiseid masinõppe projektidega seotud probleeme. Sari andis lisaväärtust koolitajate laia kogemustebaasi ning isiklikel kogemustel põhinevate näidetega.

### **2.1.4 Masinõppe lõuendi läbitöötamine**

Erinevad avaldatud materjalid pakuvad palju väärtuslikku informatsiooni, mille analüüsist on võimalik uurimusele vajalikku teadmist lisada. Olemasolevad materjalid kvalifitseeruvad erinevate teabekandjate alla olenemata oma vormist tingimused, et jäädvustatud teave on säilinud nii, et selle sisu ei ole oluliselt muudetud ega moonutatud [35].

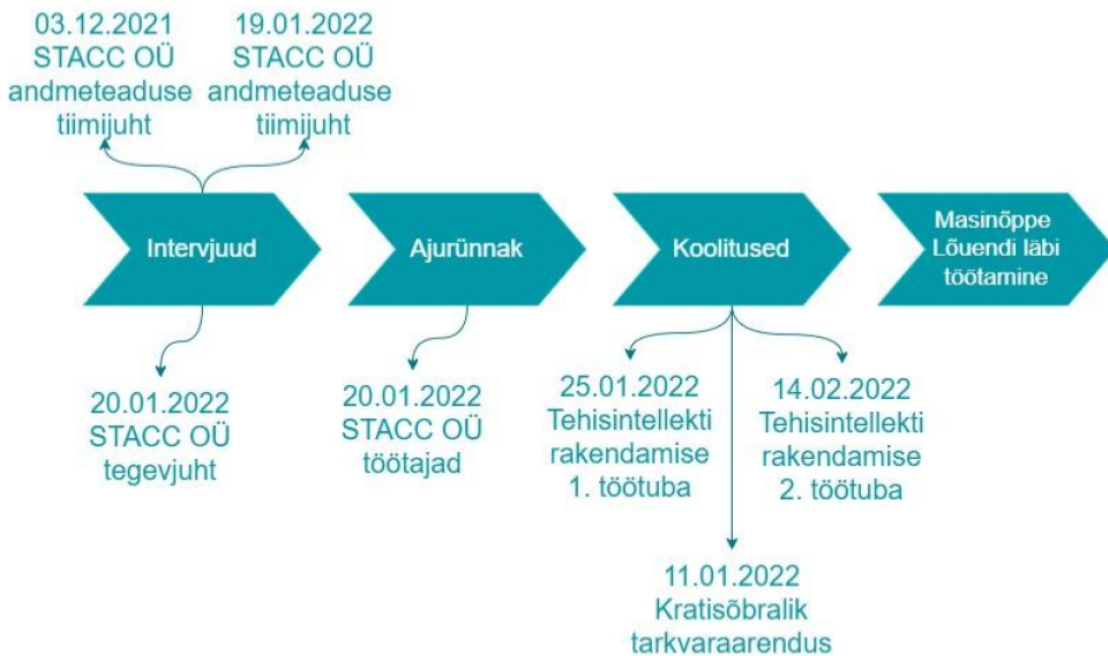
Masinõppe projektid on tihti keerukad nii tehnilises kui projekti juhtimise mõttes - eriti, mis puutub kommunikatsiooni kliendi ja lahenduse tarnija vahel. Probleemiks on süsteemi eelanalüüsi koostamine ilma, et oleks teada kui suur võiks olla loodava süsteemi mõju ja kasum. Antud magistritöö raames kasutati olemasolevate materjalide läbitöötamist masinõppe lõuendi puhul. Masinõppe lõuendi [37] on välja töötanud aastatepikkuse masinõppe projektide kogemusega Louis Dorard, kes esitleb lõuendit ning avab selle sisu ja

nippe kasutamiseks oma raamatus “Masinõppe Lõuend” (ingl k “*Machine Learning Canvas*”). Masinõppe projektid on tihti keerukad nii tehnilises kui projekti juhtimise mõttes - eriti, mis puutub kommunikatsiooni kliendi ja lahenduse tarnija vahel. Masinõppe lõuend aitab kaardistada probleemi olemust ja olulisi detaile lahenduseni jõudmiseks ehk annab hea ülevaate projekti eesmärkidest, mistõttu on seal esitletavad aspektid olulised ka antud uurimisprobleemi lahendamisel [36].

Uute ärimudelite väljatöötamisel või olemasolevate dokumenteerimiseks kasutatakse ärimudeli lõuendit (ingl k *Business Model Canvas*), mille erinevaid näidiseid on internetist laialt leida. Aastatepikkuse masinõppe projektide kogemusega Louis Dorard [36] kasutab oma raamatus “Masinõppe Lõuend” (ingl k “*Machine Learning Canvas*”) sarnast lõuendit ka masinõppe projektide jaoks. Masinõppe lõuendi [37] üks versioonidest on leitav antud töö lisadest. Masinõppe projektide puhul on taoline lõuend abiks, et kõik projektiga kokku puutuvad osapooled - nii klient, projektijuht, andmeteadlane kui andme insener - saaksid projekti eesmärkidest ja põhiprobleemidest ühtmoodi aru [36].

## 2.2 Analüüsimeetodid

Andmete kogumise protsess on visualiseeritud alloleval joonisel 3. Rakendatud meetoditeks valiti lahendatava probleemi põhjal ajurünnak, intervjuud, koolitustel osalemine ja olemasoleva materjali läbitöötamine.



Joonis 3. Andmete kogumine

Kogutud materjal koondati ning kategoriseeriti vastavalt sisule ning prioritseeriti vastavalt teemade esinemissagedusele. Kategoriate valimisel lähtus töö autor kogutud materjali sisust ning paigutas sarnase sisuga teemad ühte kategooriasse. Autori valitud kategooriad ning kategoriseerimise protsessi valideeris STACCI andmeteaduse spetsialist ning pikaajne

andmeteaduse projektijuht. Paremate analüüsitulemuste saamiseks tehti valideerimisel vajalikud sisulised parandused ning jagati üks liiga mahukas kategooria kaheks.

Küsimustele vastates kasutati rakendatud uurimismeetodite abil kogutud informatsiooni, otsiti lisainfot kättesaadavast kirjandusest ning konsulteeriti STACCI poolse andmeteaduse spetsialistiga. Vastuses kasutatud kogutud informatsioon pärineb eelnevalt läbi viidud ajurünnaku, intervjuu, koolituste ning materjali läbitöötamise protsesside tulemustest. Vajalik lisainformatsioon leiti kättesaadavast kirjandusest ning tunnustatud andmeteadlaste avaldatud töödest. Vastuste valideerimiseks kohtuti STACCI andmeteaduse spetsialisti ning projektijuhiga ning kohtumise käigus tehti autori poolt koostatud sisus parandusi ning lisati vajalik puuduolev informatsioon.

### **3 Erinevatelt sihtrühmadelt kogutud sisend**

Allolevas peatükis on välja toodud erinevate meetodite rakendamine, rakendamise tulemused ning saadud tulemuste analüüs. Analüüsis rakendatud protsessid on kirjeldatud peatükis 2.2. Antud peatükis saadud tulemustest moodustatakse siinse analüüsi põhjal kõige olulisemaid teemasid katvad küsimused, millele hiljem vastused leitakse.

#### **3.1 STACCI sisene ajurünnak**

Järgnevalt on välja toodud STACCI sisese ajurünnaku tulemused ning neist kogutud andmete analüüs koos kategoriseerimisega. Saadud tulemuste esitlemine jaotati kolme peatükki: peamised põhjused masinõppe lahenduste kasutamiseks, klientide hirmud ning takistused masinõppe projektiga alustamiseks ning klientide küsimused seoses masinõppega.

##### **3.1.1 Peamised põhjused masinõppe lahenduste kasutamiseks**

Põhjuseid, miks klient peaks soovima kasutada masinõppe lahendusi, toodi ajurünnaku käigus erinevaid. Nii selleks, et oma kliente aidata kui neid enda teenuse juures säilitada, kui ka turunduslikel eesmärkidel. Rahalise ressursi temaatikaga seoses toodi välja, et ettevõtetel on võrdlemisi kallis panna kokku oma andmeteaduse meeskond - mõistlikum on see teenus sisse osta. Eriti, kuna seda võib vaja minna vaid hooajaliselt, lühikese ajaperioodi vältel või ühekordselt. Samuti kasutatakse masinõppet rahastuse saamiseks läbi erinevate riiklike fondide, mis rahastavad andmeteadust panemata ranget survet tulemustele.

##### **3.1.2 Klientide hirmud ning takistused masinõppe projektiga alustamiseks**

Kõige enam peljatakse suurt rahalist kulu, mis selliste tehnoloogiliste lahendustega võib kaasneda. Tuntakse muret, kui palju lahendus maksma läheb ning soovitakse teada saada millal lahendus võiks umbes ära tasuda. Lisaks ei ole mõnede ettevõtete jaoks soositud STACCis kasutusel olev tunnipõhine maksesüsteem.

Rahaline ressurss eraldi põhjustab selgelt kõige suuremat hirmu. Sellele järgnevad kahtlused teiste masinõppe projektiks vajalike ressursside olemasoluks. Neist kõige olulisemad on andmed ning nende kvaliteet ja kvantiteet. Teinekord on vaja projekti alustamist edasi lükata, kuna ettevõtetel on vaja täiendada olemasolevat andmemudelit või koguda rohkem andmeid. Lisaks võib takistuseks saada ka äriprobleemi ebakonkreetsus ning ettevõtte töötajate ajaline ressurss, mida ei ole võimalik omalt poolt projekti pühendada. Tihti ei mõista küpsusel madalamatel tasemetel olevad ettevõtted, et masinõppe projekti edukaks läbiviimiseks on vaja ka neil endil sellesse võrdlemisi palju aega panustada, eriti kuna kasutatakse agiilset arendusmeetodit. Mis puutub äriprobleemi, siis ebakindluse tegelikud põhjused saavad alguse just sealt - ei olda kindlad, kas antud probleem üldse on masinõppega lahendatav. Või siis ei ole probleemi piisavalt selgelt sõnastatud, et seda üldse hinnata saaks. Kuna populaarsus masinõppe rakendamisel muudkui kasvab, kardetakse ka konkurentide kiiremat tulemusteni

jõudmist, mis tekitab arvestatavaid äririske, mille vähendamine suurendab omakorda kulutusi. Kardetakse, et on juba hiljaks jäänud.

Masinõppe lahenduste üks oluline eripära on see, et tihti ei ole võimalik kliendile anda täpset seletust, miks mudel just valitud ennustuse tegi. Teisisõnu kasvatab limiteeritud tõlgendatavus umbusaldust ja hirmu taoliste lahenduste ning süsteemide osas. Teine eripära, mis potentsiaalsetes klientides kõhklusi tekitab, on tihti ettetulev olukord, kus masinõppe projekti alustades ei ole võimalik garanteerida täpset tulemuse kvaliteeti. See selgub töö käigus ning oleneb andmete kvaliteedist ning kvantiteedist. Arusaadavalt tunnevad ettevõtted ärilistel põhjustel hirmu ebakindla tulemuse ees.

Üksikute takistustena toodi välja välja liiga kõrged ootused, liiga karmid tingimused kliendi enda poolt, parem pakkumine konkurendi poolt ning Pythoni programmeerimiskeele ebasobivus. Peamised hirmud ja takistused on koondatud tabelis 1.

Tabel 1. Klientide peamised hirmud ja takistused masinõppe kasutamiseks.

Küsimus / Hirm / Teema	Kategooria
Konkurentidel on toode olemas, meie aga alles hakkaksime arendama.	Aeg
Peamine konkurent jõuab heade tulemusteni enne mind ja siis olen temast maha jäänud.	Aeg
Enne meie palkamist peavad nad oma andmemudelit täiendama.	Andmed
Pole piisavalt andmeid.	Andmed
[STACCI] Töötajatel puudub kogemus teatud valdkondades.	Inimressurs
Liiga kõrged ootused.	Lahendus
Süsteemi hooldamisel tekkivad probleemid.	Lahendus
Liiga karmid tingimused kliendi poolt.	Muu
Python ei ole aktsepteeritav.	Muu
Liiga kallis.	Raha
Parem pakkumine konkurendi poolt.	Raha
Raha.	Raha
Tunnipõhine maksesüsteem.	Raha
Millal see [lahendus] minu jaoks ära tasub?	Raha
Liiga kallis [lahendus] - Kui palju see maksab?	Raha
Millal see minu jaoks ära tasub?	Raha
Ebakindel tulemus.	Tõlgendatavus
Tõlgendatavus - Kui nad [kliendid] kasutavad masinõpet, siis on neil tihti keeruline nendest süsteemidest aru saada ja see tekitab pahameelt ja hirmu.	Tõlgendatavus
Ei olda kindlad, kas me saame neid aidata.	Vajadus
Masinõppe ei ole rakendatav.	Vajadus
Pole kindel, kas nende probleem on masinõppega lahendatav.	Vajadus

### 3.1.3 Klientide küsimused seoses masinõppega

Ajurünnaku käigus uuriti enne täpsemate küsimuste juurde jõudmist osalejatelt, millised on masinõppega seonduvad teemad, mis klientide seas kõige rohkem küsimusi tekitavad ning mida on STACCI töötajad kõige tihedamini pidanud klientidele selgitama. Parema ülevaate saamiseks visualiseeriti tulemused nii, et sarnased vastused kuvati suuremas kirjas. Tulemused on leitavad jooniselt 4.



Joonis 4. STACCI sisese ajurünnaku tulemused. Enim küsimusi tekitanud teemad seoses masinõppega.

Ajurünnaku käigus toodi välja mitmeid erinevaid küsimusi, mida töötajad on klientidelt saanud. Osad neist võib sisu järgi grupeerida ning nii tekib kolm erinevat kategooriat: projekti võimalikkus, ressursid, lahenduse süsteem ja selle kasutamine. Projekti võimalikkuse juurde käib peaaugjalikult küsimus, et kas antud äriprobleemile on võimalik masinõppe lahendust efektiivselt rakendada. *Kas üldse vajatakse masinõpet või saaks tegelikkuses odavamate vahenditega sarnase tulemuse?* Ressursside kategooria alla jäävad küsimused, mis puudutavad rahalisi, ajalisi kui andmete ressursse. *Millal on lahendus valmis? Millal lahendus ära tasub? Milline on potentsiaalne kasum? Kui palju see maksab? Kui palju andmeid on vaja?* Lahenduse kategooriasse liigituvad küsimused on sisult süsteemi töö kohta. Klientid on soovinud teada saada, kuidas lahendus täpselt töötab. *Kui hästi lahendus töötab? Kas võite kindlad olla, et see lahendus töötab? Mis andmeid pean ma lisama, et masin tööle hakkaks? Miks mudel selle plastiku alla klassifitseerib? Kuidas ma seda lahendust kasutan? Kuidas see lahendus töötab?* Alljärgnevas tabelis 2 on välja toodud ajurünnaku käigus kaardistatud klientide peamised küsimused ja selgitusi vajavad teemad seoses masinõppega.

Tabel 2. Klientide peamised küsimused ja teemad seoses masinõppega.

Küsimus / Hirm / Teema	Kategooria
Millal me selle [lahenduse] saame?	Aeg
Kui palju andmeid on vaja?	Andmed

Kui hästi see [lahendus] töötab?	Lahendus
Kuidas see [lahendus] töötab?	Lahendus
Miks mudel selle plastiku alla klassifitseerib?	Lahendus
Kas mudel suudab ennustada järgmist pandeemiat?	Lahendus
Kas võite kindlad olla, et see [lahendus] töötab?	Lahendus
Mulle ei meeldi see soovitus, ülejäänud on normaalsed.	Lahendus
Kui palju see [lahendus] maksab?	Raha
Kui palju see [lahendus] maksab?	Raha
Millal see [lahendus] ennast ära tasub?	Raha
Potentsiaalne kasum?	Raha
Kuidas ma seda [lahendust] kasutan?	Tõlgendatavus
Kas me üldse vajame masinõpet?	Vajadus

### 3.2 Protsessijuhtimise küpsuse mudeli koolitus

Mihkel Lauk [38] viis 2015.aastal toimunud Nortal TechDay 2015 raames läbi protsessijuhtimise küpsuse mudeli koolituse. Järgnevas peatükis on esitatud tema mõtted ja selgitused ettevõtete erineva küpsusastmete ning nende mõjude kohta ühisprojektides. Sagedased vead, mida projekti alustamisel tehakse, on näiteks kliendi soovide üdini usaldamine, kehtivad ka andmeteaduse projektide juures. Tihti on kliendil soov ja visioon, aga puuduvad kindlalt paika pandud ärilised väärtused, mida tegelikult saavutada soovitakse. Teisalt aga võib ühe olulise veana välja tuua kliendi nõudmiste mittetäitmise. Sageli juhtub, et arendaja või andmeteadlane väänab veidi kliendi soove, kuna nii tundub efektiivsem ja tõhusam. See lähenemine aga kulutab kõigi osapoolte tööaega, kuna isegi kui nii tõepoolest olekski efektiivsem probleemi lahendada, on lõpuks on klient see, kes lahenduse eest maksab ja seda kasutama hakkab. Pigem pöörata rõhku tihedamale kommunikatsioonile enne muudatuste isepäist rakendamist.

Protsessijuhtimise küpsuse mudeli (ingl k *BPMM - Business Process Maturity Model*) põhjal on võimalik kaardistada projekti võimalikkust mõlema poole jaoks. Seda saab teha läbi viie ettevõtte küpsust hindava küsimuse vastuste analüüsimise. On oluline aru saada, kui valmis ettevõtte projekti alustamiseks on, kuna see määrab tugevalt projekti kulgu ning tulemust.

Mudelil on 5 taset, kus ettevõtted oma küpsuselt asetsevad.

1. *Ad hoc* tase - Kõige madalam tase, kus puudub strateegilise planeerimise võimekus ning tahe arenguks.
2. Arenev tase - On olemas vajadus muutuseks ning põhiarusaam eksteerivatest protsessidest. Protsessi juhtimise süsteemi ei eksisteeri.
3. Kontrollitud tase - Protsessid on kirjeldatud ning vastutavad isikud määratud. Protsessi juhtimise süsteemi ei eksisteeri.

4. Mõõdetud tase - Protsessidel on olemas tulemuslikkuse võtmemõõdikud. Eksisteerib protsessi juhtimise süsteem.
5. Optimeeritud tase - Toimub pidev areng. Protsessidel on olemas suurepäraselt töötavad tulemuslikkuse võtmemõõdikud, mis on seotud protsessi juhtimise süsteemiga.

Küpsusastet mudelil aitavad määrata järgmised küsimused:

- 1. Mis muutub firmas kui kasumlikkus muutub?**
  - a. selgitus on ähmane ning üldine - tasemed 1 ja 2
  - b. muudatus KPI väärtustes - tasemed 3, 4 ja 5
- 2. Kes otsustab firmas tulemuse üle?**
  - a. omanik või juhatuse liige - tasemed 1 ja 2
  - b. erinevatel protsessidel on kindlad volitustega vastutajad - tasemed 3, 4 ja 5
- 3. Mis on projekti õnnestumise skaala?**
  - a. juhtkonna isiklik arvamus (kas töö on hästi tehtud või mitte) - tasemed 1 ja 2
  - b. mõõdetav muutus KPI väärtustes - tasemed 3, 4 ja 5
- 4. Kellel on volitused anda hinnanguid?**
  - a. vaid juhtkond/projektijuht - tasemed 1 ja 2
  - b. suurem protsent projekti osapooltest - tasemed 3, 4 ja 5

Kõige keerulisem on koostööd teha firmadega tasemetel 1 ja 2 ning kõige tõhusam on tõenäoliselt koostöö firmadega tasemetel 3, 4 ja 5. Esineja märgib ära, et enamus ettevõtteid, kellega ta on oma pika karjääri jooksul ärianalüütikuna kokku puutunud, asuvad maksimaalselt 3. tasemel. Ehk siis suure tõenäosusega tuleb enamus ettevõtete puhul arvestada puudustega, mis projekti kulu jooksul vajavad rohkem ressursse, kui esialgu planeeritud oli. Järgnevas tabelis 3 on ära toodud koolituselt üles märgitud küsimused, mis aitavad projekti alguses saada aimu ettevõtte küpsusastmest ning sellest tulenevalt kaaluda projekti võimalikkust.

Tabel 3. Protsessijuhtimise küpsuse taseme koolituselt üles märgitud küsimused.

Küsimus / Hirm / Teema	Kategooria
Kas ettevõttel on olemas projektiks vajalik inimressurss?	Inimressurss
Kas ettevõttes on määratud protsesside eest vastutavad isikud?	Inimressurss
Kuidas toimub ettevõttes otsuste langetamine erinevates projektides?	Inimressurss
Kas ettevõttel on olemas strateegilise planeerimise võimekus?	Inimressurss
Kas ettevõttel on ülevaade toimivatest protsessidest?	Inimressurss

### 3.3 Intervjuu tegevjuhiga

Struktureerimata individuaalintervjuu tegevjuhi Kalev Koppeliga viidi läbi 20.01.2022. Ettevõtte tegevjuht nõustus suuremas jaos eelnevas peatükis avatud Mihkel Laugu mõtete ja strateegiaga koostöövõimaluste efektiivsemaks hindamiseks. Samuti tõi ta välja enda poolt

taolistes olukordades tihti kasutatavaid küsimusi, mis aitavad tal enda jaoks tunnetada ettevõtte küpsust alustamaks koostööd STACCiga andmeteaduse projektis. Järgnevaid küsimusi ei esitata otsesõnu klientidele, pigem on need küsimused, millele tegevjuht sooviks enne projektiga edasi liikumist enda jaoks vastuseid saada.

*Kas ettevõttel on määratud kindlate eesmärkidega projektijuht või tooteomanik?* Intervjuueeritav tõi välja, et ettevõttega suheldes on oluline aru saada tugeva projektijuhi või tooteomaniku olemasolust. Vastutav isik, kes omab ülevaadet töövoost ning kellel on selge visioon soovitatavast tulemusest. Projektide õnnestumiseks on antud faktor äärmiselt oluline. Äärmiselt tähtis on paika panna vastutuse jaotumine. Kiputakse arvama, et kui klient tellib, siis tarkvaraettevõtte tarnib lahenduse ning sellega piirduvad kliendi kohustused. Tegelikkus on keerukam. Eriti tänu andmeteaduse projektide muutlikule kulule on vajalik pidev koostöö kliendiga. Koostööaldis klient saab märksa kiiremini just teda rahuldava lahenduse. Siinkohal on oluline roll ka kommunikatsioonil.

*Kas on piisavalt andmeid?* Andmeteaduse projekti puhul on oluline andmete hulk. Sageli katkeb koostöö mingis läbirääkimiste faasis, kuna selgub, et masinõppe mudelite rakendamiseks ei ole kliendil piisavalt andmeid. Sealt tekib alati ka küsimus - kui palju andmeid vaja on? *Kui suur on pakutava lahenduse vastutus?* Lisaks on oluline hoomata, mis on kliendi infotehnoloogilised võimekused, kuna ka need määravad projektide tulemuslikkust. *Milline on ettevõtte tehnoloogilise pädevuse tase?* *Kui laiad on teadmised andmeteadusest?* Kui kliendil on kasvõi umbkaudne aimdus andmeteadusest, on projekti läbiviimine märgatavalt kergem ja sujuvam.

Tegevjuht juhindub suuresti analüüsi tehes Masinõppe Lõuendist (*Machine Learning Canvas* ingl k) [37]. Läbi selle on mugav visualiseerida kogupilti alustades äriprobleemi defineerimisega kuni lõpetades masinõppe rakendamise võimalustega. Alljärgnevas tabelis 4 on välja toodud tegevjuhi intervjuust üles märgitud küsimused ja selgitusi vajavad masinõppega seotud teemad.

Tabel 4. Tegvjuhi intervjuust üles märgitud küsimused ja teemad.

Küsimus / Hirm / Teema	Kategooria
Kas on piisavalt andmeid?	Andmed
Kas andmed on kättesaadavad?	Andmed
Kas on tugev projektijuht või tooteomanik, kes saab aru, mida tahab?	Inimressurss
Milline on ettevõttesisene IT tase?	Inimressurss
Kas on eelteadmisi andmeteadusest?	Inimressurss
Kui suur on süsteemi vastutus?	Lahendus
Kas on äriprobleem on?	Vajadus

### 3.3.1 Masinõppe Lõuend

Dorard'i [36] selgitused Masinõppe Lõuendi kohta kaetakse alljärgnevas peatükis. Autori sõnul peaks masinõppe projekti lõuendi täitmist alustama masinõpet rakendama hakkava süsteemi väärtuspakkumise (ingl k *Value Propositions*) sõnastamisest - *mida, milleks ja kes teeb?* Järgmiseks blokiks tuleks mõelda andmetest õppimise peale. *Kust saadakse toorandmed? Kust tulevad uued andmed? Millal uuendatakse mudelite treeningandmeid ja kaua selleks aega on?* Kolmandasse blokki on autor paigutanud ennustamise. *Millise info masinõppe mudelile anname ja mida tahame tagasi saada? Kuidas ennustuste põhjal otsuseid tehakse? Kuidas toimib töö uute andmetega?* Ning viimaseks blokiks on hinnangud. *Kuidas mõõta süsteemi töökindlust? Kuidas süsteemi monitooritakse?* Masinõppe Lõuendi analüüsist üles märgitud küsimused on välja toodud tabelis 5.

Masinõppe lõuend aitab inimestel, kellel on vähene kokkupuude, avastada erinevaid kasutegureid, mida masinõppe meetodite rakendamine pakkuda võib. Kommunikatsioon on projektides väga tähtsal kohal ning tihti kipuvad probleemid algama just sealt, kui inimesed ei väljenda end piisavalt selgelt ja osapooled mõistavad teineteist valesti. Masinõppe lõuend aitab seda vältida - see kasulik abivahend hoidmaks kõiki projekti osalisi ühisel arusaamisel projekti eesmärkidest ja kulust.

Tabel 5. Masinõppe lõuendi analüüsist üles märgitud küsimused ja teemad.

Küsimus / Hirm / Teema	Kategooria
Süsteemi väärtuspakkumise sõnastamine	Andmed
Kust saadakse toorandmed?	Andmed
Kust tulevad uued andmed?	Andmed
Kuidas hinnata edukust?	Lahendus
Millal uuendatakse mudelite treeningandmeid ja kaua selleks aega on?	Lahendus
Millise info masinõppe mudelile anname ja mida tahame tagasi saada?	Lahendus
Kuidas ennustuste põhjal otsuseid tehakse?	Lahendus
Kuidas toimib töö uute andmetega?	Tõlgendatavus
Kuidas mõõta süsteemi töökindlust?	Vajadus
Kuidas süsteemi monitooritakse?	Vajadus

### 3.4 Intervjuud andmeteaduse tiimijuhiga

Intervjuu sissejuhatuseks alustatakse kaugemalt - iga probleemi lahendamiseks tõesti ei ole vaja rakendada niivõrd keerulist tehnoloogiat nagu masinõpe. Masingu kogemuse põhjal on rutiinsete ning aeganõudvate probleemide lahenduseks enamjaolt mõistlik leida mõni lihtsam tehnoloogiline võimalus, kui masinõpe. Küll aga tasub selle mõelda, kui on probleem, millele inimene ise ei oska vastust leida. Või kui otsuste tegemiseks puudub kindel reeglistik, mille põhjal neid otsuseid langetada. Sellistel juhtudel võiks masinõppe peale mõelda.

Intervjueeritava arvates on ettevõtjatel enamasti üks põhiline küsimus, millest omakorda kasvavad välja kõik muud hirmud ja küsimused. *Kuidas vähendada kulusid ja kasvatada tulusid võimalikult efektiivselt?* Sealt edasi lähevad ettevõtjate küsimused tavaliselt spetsiifilisemaks, kuna üritatakse aru saada, kuidas masinõpe just konkreetses ärimudelis võiks kasulik olla ja kulusid optimeerida. Riskianalüüsi läbiviimiseks soovitakse teada ohukohti ja riske. *Millised on sagedased riskid? Kas on riske, mida oleme suutelised ette nägema?* Masingu sõnul tuntakse muret ka eeldatavate tulemuste pärast, kuna masinõpet rakendades on eeldatavat kvaliteeti võrdlemisi keeruline prognoosida. Mõnikord uuritakse eelnevate sarnaste projektide kohta - *Kuidas teil eelmiste klientidega läinud on?* Kogenumatel ettevõtetel tekib küsimusi ka lahenduste integreerimise, juurutamise ning reaalse rakendamise kohta. Lisaks ei saa tänapäeval mööde turvalisusest - andmed on kallid vara ning nendega peab ettevaatlikult ümber käima.

Kardetakse andmete jagamisega seotud riske. Lõpuks taandub põhiline murekoht rahalisel ressursil - *kui palju masinõppe rakendamine mulle maksma läheb? Kui palju me siis võidame? Kui palju rohkem me teenime selle arvelt?* Masing leiab, et sagedamini esitatavate küsimuste välja sorteerimisel ning nendele vastuste otsimisel on oluline silmas pidada, kellele täpselt on vaja midagi seletada. Näitlikustamiseks toob ta ettevõtte kommunikatsioonijuhi ja finantsijuhi - selleks, et hirme maandada on vaja teada ka sihtgruppi, kuna samad argumendid ei puugi mõlema osapoole peal töötada. Alljärgnevas tabelis 6 on välja toodud andmeteaduse tiimijuhi intervjuust üles märgitud küsimused ja selgitusi vajavad masinõppega seotud teemad

Tabel 6. Tiimijuhi intervjuust üles märgitud küsimused ja teemad.

Küsimus / Hirm / Teema	Kategooria
Kaua lahendusega läheb?	Aeg
Andmete turvalisus	Andmed
Kui hea on eeldatav lahenduse kvaliteet?	Lahendus
Kuulsin, et on võimalik saada 95% täpsust, kas me saaks ka sellise lahenduse?	Lahendus
Millist arendust kasutada? (agiilne/waterfall)	Lahendus
Mis on suurimad riskid?	Muu
Milliseid riske suudate ette näha?	Muu
Kuidas teil eelmiste klientidega läinud on?	Raha
Kui palju masinõppe rakendamine mulle maksma läheb?	Raha
Kui palju me siis võidame?	Raha
Kui palju rohkem me teenime selle arvelt?	Riskid
Masinõppe lahenduse reaalne rakendamine on keeruline	Riskid
Masinõppe juurutamine on keeruline	Tõlgendatavus
Süsteem tundub keeruline ja arusaamatu	Tõlgendatavus
Kuidas ma seda kasutama hakkan?	Tõlgendatavus

Kas on vaja rakendada masinõpet?	Vajadus
Milliseid probleeme saab masinõpe selles valdkonnas lahendada?	Vajadus
Kuidas saab masinõpe minu kulusid optimeerida?	Vajadus

## 3.5 Kratikoolitused

### 3.5.1 Kratisõbralik tarkvaraarendus

Tiit Sepp [39] keskendus koolitusel põhiliselt masinõppe projektide üldise ülevaate andmisele ning projektijuhi vaates kõige tähtsamatele ning kriitilisematele aspektidele. Lähenemisi on erinevaid, kuid koolitaja jaotab masinõppe projekti kümneks etapiks, mis on ära toodud alloleval joonisel 5.



Joonis 5. Masinõppe projekti etapid [39].

Sepa sõnul on ettevõtetal tihti kahtlusi automatiseerimise võimalikkuse osas. Kõige lihtsam vastus sellele küsimusele ilmneb, kui mõelda, kas antud protsessi on võimalik teha tabelitöötuse formaadis ehk exceliga. Kui jah, on protsess kindlasti automatiseeritav. Kui aga mitte, peab lähemalt detailidesse süvenema.

Koolitaja enda esmased kõige paremini informeerivad küsimused enne andmeteaduse projekti alustamist on:

1. Kui palju võib lahendus tõsta teenuse kvaliteeti?
2. Kui palju võib selle töö säästa inimeste tööaega?
3. Kas andmed on olemas?

Kõige olulisemaks peab koolitaja kindla, selge ja defineeritava eesmärgi seadmist. Seda siis näiteks raporti, prototüübi või rakenduse näol. Projektijuhi seisukohast oleks enne projektiga alustamist vajalik defineerida minimaalne töötav toode ehk MVP (ingl k *MVP Minimum Viable Product*) ning selle vastand ehk unistuste töötav lahendus. Nii toimides püstitatakse meetrikud projekti edukuse mõõtmiseks. Minimaalse töötava lahenduse jaoks defineeritakse minimaalsed nõuded, mis on kindlasti vaja täita, et lahenduse esimene versioon töötaks ning suudaks püstitatud probleemi lahendada kaasates ainult hädavajalikud featuurid ehk tunnused [40]. Samuti tasuks ära määrata olulisemad verstapostid ja hetked, millal on võimalik veel projekti katkestada - selline toimimine aitab vältida projekti keskel tekkivaid kommunikatsiooniprobleeme. Lisaks minimaalse ning unistuste lahenduse defineerimisele on tehisintellekti rakendatavates projektides oluline roll ka andmetel. Ehk siis on vaja võimalikult kiiresti aru saada andmete kvaliteedist ja kvantiteedist. Eriti hea, kui saaks mõista, kes ja kuidas kogub andmeid. Lisaks aru saada, kuidas toimitakse, kui selgub, et andmeid pole piisavalt - kas neid saab juurde tekitada?

Oma kogemuse põhjal liigitas koolitaja enamlevinud projektide ebaõnnestumise põhjusteks andmete kvaliteedi ja/või kvantiteedi, valed eeldused projekti alustades ning puuduliku minimaalse töötava toote ehk MVP (*Minimum Viable Product* ingl k). Samuti tõi ta välja suhtumise, mille eesmärgiks on lahendada kõik probleemid korraga ja võimalikult ressursi säästlikult. Antud lähenemisega kaasneb enamasti fookuse hajumine, mille tulemusel on soovitatavale hoopis vastupidine efekt. Ettekandes märgiti ära ka MKM-i poolt välja töötatud kontrollküsimused andmeteaduse projekti alustamiseks, mis on leitav antud töö lisadest. Sepa sõnul võiks sellest masinõppega vähe kokupuuteid omavatel projektijuhtidel palju kasu olla, kuna see võib aidata vältida tüüpilisemaid vigu.

### **3.5.2 Tehisintellekti rakendamise 1. töötuba**

Tiit Sepp [41] tõi masinõppe algoritme seletades palju elulisi näiteid soovitusüsteemidest, kuna on mitmetes masinõpet rakendavaid soovitusüsteeme arendanud. Ta tõi välja ühe klientide poolt tihti esineva küsimuse: Miks soovitab süsteem just seda toodet? Reaalsus on aga see, et masinõppe puhul on enamasti võimatu mõista soovituse andmise põhjust, kuna kasutatavad algorütmid on väga keerulised. Seetõttu levibki nimetus nõ “musta kasti” meetod [23]. Koolitaja sõnul on statistiliste meetoditega palju kergem taolitele küsimustele vastata.

Dage Särg [42] tõi välja näited, mil masinõpe võiks inimese elu mugavamaks ning säästa aega ja raha: nutikad tekstiotsingud, tekstide automaatne liigitamine, automaatsed sisukokkuvõtted, nimeüksuste tuvastamine, anonümiseerimine ja dokumendimallide genereerimine. Masinõppe projekti jaoks peavad andmed olema kättesaadavad, mõistetavad, hallatavad, mahukad ja kasutatavad. Kättesaadavuse all mõtleb koolitaja ka juriidilisi takistusi, mis andmete jagamisel tihti takistuseks on. Hallatavuse all aga seda, et nii andmed kui metaandmed peavad olema korduvalt käsitletavad ning kasutatavus vihjab struktureeritusele ja puhtusele. Masinõppel on veel tähtsaid nõudmisi andmetele. Tihti tekib ettevõtetel masinõppe lahendusi kaaludes küsimus, et kui palju on andmeid vaja? Selle jaoks ühest vastust ei ole. Koolitaja leiab aga, et kui andmeid on piisavalt vähe, et neid käsitsi läbi vaadata, siis ei ole mõtet masinõppe lahendusi rakendada. Üldiselt kehtib andmeteadlaste seas reegel, et mida rohkem andmeid, seda parem - mida keerulisem mudel, seda enam andmeid heade tulemuste saavutamiseks vajatakse. Lisaks on oluline ka andmete kvaliteet, kuna vigaste andmete tulemuseks on vigane mudel ja kehvad tulemused. Ja viimase olulise nõudena toob koolitaja välja tunnuste relevantsuse sõltuva muutuja suhtes. Näiteks ei ole võimalik mõne kunstiteose žanrit vaid autori abil ennustada - selleks on vaja veel andmeid. Lisaks käsitleb esineja eraldi erinevaid mudeleid ning selgitab, mille põhjal tuleks erinevate mudelite rakendamist kaaluda.

### **3.5.3 Tehisintellekti rakendamise 2.töötuba**

Anna Laaneväli [43] iseloomustas andmete kvaliteeti andmeteadlaste seas levinud ütlusega, mis iseloomustab ebakvaliteetsete või vigaste sisendandmete põhjal saadavat projekti olematut väärtuspakkumist. Laaneväli selgitas andmete kvaliteeti järgmiste oluliste märksõnadega: andmete täpsus, terviklikkus, vorming, järjepidevus, duplikaatide olemasolu ning integreeritavus. Lisaks käsitles koolitaja tundlike andmete lekkeid, mis on masinõppe projektide alustamisel sagedane murekoht. Tänapäeval lähevad andmete kogumise, hoiustamise ja jagamise üle kehtivad seadused aina karmimaks ning andmeid peetakse üha väärtuslikemaks. Andmelekete vältimiseks kasutatakse pseudonümiseerimist, hashimist ja krüpteerimist. Pseudonümiseerimise põhimõte seisneb isikuandmete töötlemises taoliselt, et andmetes olevat teavet ei oleks võimalik seostada konkreetse andmesubjektiga. Tegelikult ei ole mures ainult ettevõtted, kelle andmeid andmeteaduse firmad kasutavad, vaid ka andmeteaduse firmad ise - neil langeb suur vastutus nende käes olevate andmete eest. Seetõttu järgitakse tundlike andmete kogust põhimõtte järgi: nii vähe kui võimalik, kuid nii palju kui vajalik. Samuti ollakse ettevaatlikud õiguste ja ligipääsude jagamisega süsteemides - seda tehakse nii palju, kui andmeteaduse teenust pakkuval ettevõttel projekti raames vaja on, aga mitte rohkem.

Järgnevas tabelis 6 on välja toodud kolmelt koolituselt kogutud sisend, mis on vastavalt sisule kategoriseeritud.

Tabel 6. Kratikoolitustelt üles märgitud küsimused ja teemad.

Küsimus / Hirm / Teema	Kategooria
Kas andmed on hallatavad?	Andmed
Kas andmed on kvaliteetsed?	Andmed
Kas andmed on kättesaadavad?	Andmed
Kas andmed on mõistetavad?	Andmed
Kas andmed on olemas?	Andmed
Kas andmed on täpsed?	Andmed
Kas andmeid on piisavalt?	Andmed
Kes ja kuidas kogub andmeid?	Andmed
Kui palju andmeid vaja on?	Andmed
Kuidas tohib andmeid vahetada?	Andmed
Kus võib andmeid hoida?	Andmed
Millised on hetkel kasutusel olevad ressursid ja kas neid saab juurde?	Andmed
Kas on koostatud kommunikatsiooniplaan?	Inimressurss
Kas projekti on kaasatud lõppkasutaja?	Inimressurss
Kes tohib tulemusi näha?	Lahendus
Kui palju lahendus võib tõsta teenuse kvaliteeti?	Lahendus
Kui palju võib töö säästa inimeste tööaega?	Lahendus
Kuidas tulemust hinnata saab?	Lahendus
Millised on sisulised piirangud rakendusele?	Lahendus
Milliste ajaliste ja täpsuslike parameetritega peab mudel töötama?	Lahendus
Mis on minimaalne töötav lahendus?	Lahendus
Mis on unistuste töötav lahendus?	Lahendus
Fookuse hajumine	Riskid
Kas on paika pandud projekti katkestamise tingimused?	Riskid
Kui andmeid pole, siis kas neid saab juurde tekitada?	Riskid
Mis on suurimad riskid?	Riskid
Puudulik "vähim töötav lahendus"	Riskid
Puudulik inimressurss	Riskid
Vaja lahendada kõik olemasolevad probleemid ühe lahendusega.	Riskid
Valed eeldused	Riskid
Miks mulle soovitati seda toodet?	Tõlgendatavus
Kas on defineeritud MVP?	Vajadus
Kas on defineeritud projekti tähtsamad etapid?	Vajadus
Kas on defineeritud unistuste toode?	Vajadus

Kas on määratud täpne eesmärk, mida tulemuseks oodatakse?	Vajadus
Kas on vajadus masinõppeks?	Vajadus
Kas probleem on automatiseeritav?	Vajadus
Kas äriprobleem on olemas?	Vajadus

### 3.6 Analüüsi kokkuvõte

Kõige enam küsimusi ning hirme seoses masinõppe rakendamise võimalustega tekib lahenduse süsteemi, kvaliteedi ja võimekusega seotud teemade ning erinevate projektiks vajaminevate ressursside kulu kohta. Soovitakse teada ennustuse täpsuse protsenti, lahenduse töö protsesside kulgu ning süsteemi paremaks hoomamiseks mõista, kuidas lahendus töötab. Rahalise poole pealt soovitakse teada saada projekti umbkaudset või täpset maksumust ning ajalise poole pealt tahetakse teada prognoositavat lahenduse ära tasuvuse aega ja umbkaudset lahenduse väljatöötamise aega. Samuti on olulisel kohal äriprobleemi olemuslikud küsimused - probleemi sõnastus, lahenduse vajadus ning võimalikkus. Soovitakse teada saada, millised on masinõppe kasutegurid ning kas oleks võimalik saavutada sama tulemust ka muude meetoditega. Lisanduvad konkreetsed küsimused antud ettevõtte tegevusvaldkonna ja masinõppe kasumlikkuse kohta. Lisaks ollakse kursis, et masinõppe lahendustes on olulisel kohal andmed, seega tõstatub ka andmete koguse ning kvaliteedi temaatika.

## 4 Küsimuste koostamine

Töö eesmärgiks oli kaardistada kõige sagedamini esinevad küsimused ning seletusi vajavad teemad. Eesmärgi täitmiseks koguti kokku kõik intervjuudest, ajurünnakust ning koolitustelt välja tulnud küsimused ning teemad. Parema ülevaate saamiseks määrati igale küsimusele ja teemale sisu järgi otsustades vastav kategooria. Autori valitud ning andmeteaduse tiimijuhil valideeritud 9 eristavat kategooriat ning nende selgitused koos tulemustega on ära märgitud allolevas tabelis 7. Igale küsimusele ja teemale lisati sisuga vastav kategooria. Kokku analüüsiti 113 küsimust. Vastavalt intervjuudest, ajurünnakust ning koolitustelt välja tulnud kõige sagedamini esinevate küsimustele ning välja toodud seletusi vajavatele teemadele sõnastati antud töö raames 15 küsimust, mis jaotati nende paremaks hoomamiseks eraldi nelja kategooriasse: projekti võimalikkus, ressursside kulu, lahenduse kvaliteet ja toimimine ning detailid.

Tabel 7. Kogutud andmete analüüsi tulemused

Kategooriad	Selgitus	Kokku
Lahendus	Kõik, mis on seotud lahenduse süsteemi, selle töökindluse ja võimekusega.	24
Andmed	Andmete ressurss	21
Vajadus	Kõik, mis puutub äriprobleemi olemusse ning küsimuste ja hirmudega seoses lahenduse võimalikkusega.	17
Raha	Rahaline ressurss	14
Inimressurss	Inimeste aeg ja oskused projekti panustamiseks. Ettevõtte küpsusastmega seotud	11
Riskid	Teemad, mis on seotud masinõppe projektide riskidega.	10
Tõlgendatavus	Kõik, mis puutub masinõppe süsteemide raskesti mõistetavusse ning tõlgendatavusse.	8
Aeg	Ajaline ressurss	5
Muu	Muud küsimused, mis ei kvalifitseeru muude teemadega.	3
<b>Küsimusi kokku:</b>		<b>113</b>

### 4.1 Projekti võimalikkus

Antud kategooriasse paigutati kõige sagedamini esinenud küsimused, mis on seotud üldise projekti võimalikkusega. Need küsimused tekivad tavaliselt kõige esimestes idee faasides ning on üldised ja laiamahulised jättes välja spetsiifilised tehnilised detailid. Enamasti kiputakse projekti võimalikkuse kaalumist alahindama - tegelikult tuleks sellele pühendada piisavalt palju aega, et jõuda konkreetse ja hästi sõnastatud probleemi ning vähemalt

minimaalse võimaliku lahenduseni. Seejärel saab otsustada, kas antud lahenduse saavutamiseks on optimaalsemaid viise või oleks kõige mõistlikum kaaluda masinõppe lahendust. Projekti võimalikkuse osas sõnastati järgmised kõige sagedasemad ning olulisemad küsimused:

1. *Kas meil on masinõpet vaja või saab sama tulemust ka muude lahendustega?*
2. *Kas masinõppe aitab optimeerida mingit olulist näidikut (aeg, raha jne)?*
3. *Kas masinõppe on võimeline antud juhul lahendada probleemi?*

## **4.2 Ressursside kulu**

Masinõppe projektidega seotud vajalikke ressursse on mitmeid. Siia kategooriasse on kokku pandud üles kerkinud küsimused ja teemad raha, aja, andmete ning inimressursi teemadel. Kogutud materjalide analüüsist selgus, et ettevõtted uurivad tihti rahalise ressursi kohta - soovitakse teada umbkaudset projekti maksumust ning tasuvusperioodi pikkust, et hinnata oma võimalusi ja kaasnevaid riske. Inimressursi kulude kohta ei taipa ettevõtted tavaliselt küsida, see tuli tugevamalt välja teenusepakkuja poolsetest vastustest.

1. *Kui kallis masinõppe lahendus on?*
2. *Millal lahendus end ära tasub? Mis on potentsiaalne kasum?*
3. *Millal me lahenduse saame?*
4. *Kui palju andmeid on selleks vaja?*
5. *Milliseid andmeid on vaja?*
6. *Kui palju meie inimeste aega selle projekti jaoks vaja on? Mis peaks olema meiepoolne panus?*

## **4.3 Lahenduse kvaliteet ja toimimine**

Kõige keerukamasse kategooriasse kuuluvad küsimused masinõppe lahenduse tõlgendatavuse, kvaliteedi ja toimimise kohta. Keerukaks teeb selle asjaolu, et ilma konkreetset tausta teadmata ja andmenäidiseid nägemata on väga raske niisugustele küsimustele vastata. Siiski tuleb antud küsimused ja teemad töö raames kaasata, kuna analüüs näitas selgelt nende kõrget esinemissagedust.

1. *Kuidas garanteerite töökindla lahenduse?*
2. *Kui hästi see lahendus töötab?*
3. *Kuidas ma tean, miks mudel just sellise ennustuse tegi?*
4. *Kuidas ma tean kui hästi see lahendus töötab?*

## 4.4 Detailid

Analüüsist selgus ka mõni tehnoloogiliselt spetsiifilisem teema, mis vajaks käsitlemist, kuid ei kvalifitseeru ühegi teise kategooria alla. Programmeerimiskeel Python on vaieldamatult kõige populaarsem keel, mida andmeteaduse valdkonnas kasutatakse [44]. See aga tekitab potentsiaalsetes klientides mõnikord ebamugavust, kuna erinevate keelte integreerimine on tülikas. Lisaks soovitakse parima tulemuse saavutamiseks lisateavet erinevate meetodite kohta.

1. *Miks Python?*
2. *Milline meetod valida?*

## **5 Koostatud küsimustele vastamine**

Alljärgnevas peatükis on eelnevalt välja valitud küsimuste sisu avatud ning lisatud neile vastav kontekst, mille eesmärgiks on potentsiaalsete klientide esmastele küsimustele ja hirmudele vastata ning võimalusel ka kliendi värbamine. Küsimustele vastates on kasutatud antud töös rakendatud uurimismeetodite abil kogutud informatsiooni ning otsitud lisainfot kättesaadavast kirjandusest. Vastused on valideeritud andmeteaduse spetsialisti ning andmeteaduse projektijuhi poolt.

### **5.1 Projekti võimalikkus**

Projekti võimalikkuses kindlate vaadete omamine on kahtlemata projekti õnnestumiseks äärmiselt oluline faktor. See on kogu projekti vundament - mõrasid ei tohiks endale antud etapist läbi kiirustades lubada, kuna hiljem võivad need kalliks maksma minna. Siia kategooriasse mahub küsimusi ning murekohti, millega STACC tihti klientide poolt kokku puutub, kuid samas ka aspektid, mille peale ettevõtte sagedasti ise ei tule, kuid mis on projekti õnnestumiseks hädavajalikud. Mida paremini informeeritud ettevõtja oluliste nüansside suhtes on, seda teadlikumalt oskab ta ka konkurentidelt erinevate võimaluste kohta uurida ning seeläbi parima võimaliku lahenduseni jõuda.

#### **5.1.1 Kas meil on masinõpet vaja või saab sama tulemust ka muude lahendustega?**

Fakt on, et iga probleemi lahendamiseks ei ole mõistlik rakendada niivõrd keerukaid süsteeme nagu näiteks masinõpe. Kas sama tulemus on võimalik saada kasutades tabelitöötlustes andmeid filtreerides, sorteerides või teisi tabelitöötluste funktsioone kasutades? - Kui selline tööviis annab soovitud vastused, siis saab probleemi lahendada ka ilma masinõppeta. Lahendamise viis sõltub probleemi sisust ning keerukusest. Probleemi kaardistamisel pakub STACC ettevõtetele nõu. Alljärgneval joonisel 6 on välja toodud kõige sagedasemad masinõppe abi lahendatavad probleemid.

Kui on olemas konkreetne probleem ning vajalikud ressursid (andmed, raha, aeg ja inimesed), võib siiski olla mõistlik masinõppe lahenduste kohta uurida. Lisaks on tehisintellekti rakendamine tänapäeval hea motivatsioon investoritele, mis näitab toote või teenuse innovaatilisust ning loob läbi turueelise sellele lisaväärtust. Parimat lahendust otsides tasub siiski meeles pidada, et masinõpe on hinnanguline - ei ole olemas kindlat 100% täpsusega ennustavat mudelit, mille kätte inimelu usaldada.



Joonis 6. Peamised masinõppe abil lahendatavad probleemid [45].

Kui on selge, et masinõppe rakendamist oleks mõistlik lahendusena kaaluda, tasub edasi mõelda, kas probleemi lahendamiseks on olemas kindlad reeglid, mille abil suudab masin soovitud lahenduseni jõuda? Kui jah, siis on võimalik, et probleem on lahendatav reeglipõhiste süsteemide abil - tehisintellekti intuiitivseima liigi näol [7]. Reeglipõhiseid tehisintellekti süsteeme kasutatakse näiteks e-kirjade sorteerimiseks mingite etteantud märksõnade alusel. Reeglipõhised süsteemid võivad olla lihtsamad, kui masinõppe põhine lähenemine, aga ei pruugi - see sõltub reeglite keerukuse tasemest, kuna see omakorda määrab kogu lahenduse mahu ning keerukuse. Lisaks tuleb kaardistada reeglite stabiilsus - kui on olemas reeglid, mis tiheda intervalli tagant (näiteks iga päev või iga nädal) muutuvad, siis on see hoopis tugevaks argumendiks masinõppe lahenduse kaalumisel, kuna muud lahendused ei ole nii optimaalsed.

Juhul, kui lahenduseni jõudmiseks kindlaid reegleid ei eksisteeri, tasuks uurida masinõppe rakendamise võimalikkuse kohta, kuna masinõppes on süsteem võimeline ise olemasoleva info põhjal õppima ning järeldusi tegema. Minnes tagasi e-kirja näite juurde - masinõppe meetod oleks erinevalt eeltoodud lihtsamast variandist võimeline vanemate e-kirjade põhjal ise sobivaid võtmesõnu tuvastama ning selle põhjal kirju sorteerima - masinõppe lahendus ei vaja selleks ettekirjutatud reegleid ega reeglite haldamist [7].

STACCil on kogemus nii reeglipõhiste kui masinõppel põhinevate süsteemide arendamisega. Oma kauaaegse kogemuse põhjal leitakse, et praktikas on kõige edukamad hübriidsüsteemid, milles kasutatakse mõlemat varianti - nii reegleid kui masinõpet.

### **5.1.2 Kas masinõpe aitab optimeerida mingit olulist näidikut (aeg, raha jne)?**

Olenevalt konkreetsest probleemist võimaldab masinõpe läbi protsesside automatiseerimise ning tootmise kvaliteedi tõstmise optimeerida erinevaid näidikuid näiteks aega, raha, toote kvaliteeti ja kvantiteeti. Masinõppe abil on võimalik ette näha ja tuvastada rikkeid, mis omakorda vähendab rikete peale tehtavaid kulutusi [46]. Erinevates tööstusettevõtetes kasutatakse masinõppe lahendusi ressursside kokkuhoiduks, mis omakorda langetab toote omahinda ning annab seega ettevõttele selge eelise võrreldes turul olevate konkurentidega [46]. Masinõppel põhinevat pilditöötlust kasutatakse näiteks puidutööstuses, kus selle abil koostatakse võimalikult optimaalne saekava. Saekava on täpselt läbi arvutatud plaan palgi lahtisaagimiseks eesmärgiga teha seda võimalikult optimaalselt - saades maksimaalselt toodet ning minimaalselt jäätmeid. Tallinna Tehnikaülikool professor Argo Rosina leiab, et tehisintellekti kasutuselevõtt intensiivistub tööstuse keerukuse kasvu ning karmistuvate keskkonnanõuete tõttu [46]. STACC on metsandusvaldkonnas loonud Riigimetsa Majandamise Keskusele raie planeerimise automaatse tööriista, mis aitab inimtööd automatiseerida ning võimalikult optimaalselt raieid planeerida [47]. Lisaks arendati Muinsuskaitseametile lahendus, mis prognoosib museaalide säilivust ning teostab ajalooliste fotode ja dokumentide pildianalüüsi. Antud lahenduse abil säästetakse palju kulukaid inimtöötunde ning väldidakse inimlikke eksimusi. Veelgi suurema kaaluga näite võib tuua meditsiinivaldkonnast, kus SATCC arendas Ravimiametile apteegi hinnakokkulepete rikkumise riskimudeli ning kommunikatsiooni valdkonnast, kus arendati õhtulehele kommentaaride klassifitseerija [47]. Klassifitseerija aitab veebilehe omanikul tuvastada halvustava sisuga kommentaare, mille tuvastamata jätmiseks oleks veebilehe omanikul seadusest tulenev kohustus vastustus võtta ning trahvi maksta. Need lahendused aitavad ära hoida seadusest tulenevate rikkumisi - tehisintellekt on siinkohal firmadele suureks abiks. Populaarse näitena võib tuua ka turunduse, kus kasutatakse tihti masinõppe algoritme, et reklaamide suunamine oleks võimalikult optimaalne ning efektiivne. Kuna turunduskulud on paljude ettevõtete kuludes võrdlemisi suure osakaaluga, on oluline, et tänapäeva laias turundusmaailmas enda ettevõttele kõige kasulikum platvorm leida [48]. Nii on võimalik masinõppe abil leida erinevate mõjutegurite, nagu näiteks sooduskampaaniad või ilm, positiivset mõju müügile või näha mustreid konkurentide tegevusest tulenevast negatiivsest mõjust müügile [48]. Seega vastus küsimusele on jah - tehisintellekt võib edukalt mõnda olulist mõõdikut optimeerida ning ettevõttele seeläbi suurt väärtust pakkuda.

### **5.1.3 Kas masinõpe on võimeline antud juhul lahendama probleemi?**

Ilma täpset konteksti teadmata on keeruline sellele küsimusele otsest ja ühest vastust anda. Küll aga on kõige olulisem selge arusaam probleemi olemusest, mille lahendamiseks soovitakse masinõpet rakendada. Väga oluline on määratleda tellijapoolsed ootused masinõppe otsuse täpsusele. Probleemi püstitades tasub meeles pidada, et masinõpe on toetav, mitte absoluutne lahendus. STACC on valmis ettevõtteid probleemi püstitamisel konsulteerima ning kiirete näidete põhjal arutlema kas ja mida on võimalik probleemi lahendamiseks ära teha. On võimalik teha ka kiire kontseptsiooni tõestus (ingl k *Proof Of*

*Concept (POC)*), et testida probleemi lahendamise ettepanekute teostamist ning saada ülevaade erinevatest valiku variantidest [49]. STACCI põhimõte ei ole igale potentsiaalsele kliendile soovitada masinõpet - kui seda lahenduse saamiseks vaja ei ole, siis seda ka väljendatakse. Eesmärk on lahendada päris probleemi. Tasuks vältida suhtumist, mille eesmärgiks on lahendada kõik olemasolevad probleemid ühekorraga ning võimalikult väheste ressurssidega. Kogenud andmeteadlaste sõnul viib taoline suhtumine enamasti fookuse hajumise ning projekti ebaõnnestumiseni [39]. Mõistlikum on defineerida üks konkreetne probleem ja hiljem edasi liikuda teiste juurde. Kuigi esialgne ülesande püstitus võib töö käigus muutuda ja täpsemaks minna, annab täpsem soovitava tulemuse definitsioon võimaluse projektiga kiiremini edasi areneda [50]. Lõpuks on aga iga projekt erinev ning enamasti vajab spetsiifilisemaid analüüse ning arutluskäike.

## **5.2 Ressursside kulu**

Kõige teravamalt paneb enamasti ettevõtjaid muretsema ressursside kulu - mis ei ole riskihinnanguid tehes asjatu hirm. Masinõppe projekti olulisemate ressursside hulka kuuluvad andmed, raha, aeg ning inimressurss.

### **5.2.1 Kui kallid masinõppe lahendus on?**

Projekti hinda valemis on olulised muutujad probleemi keerukus, ootused lahendusele, integreerimiskeerukused ning sobivate andmete hulga olemasolu. Mida rohkem on võimalik masinal otsuseid teha, seda keerulisem on süsteem. Lisaks tuleb arvestada probleemi lahendamise keskkonna muutusi - see lisab samuti projektile keerukust juurde. Integreerimiskeerukuste all mõeldakse kliendipoolseid tehnilisi valikuid seoses lahenduse tarnimise ning hooldamisega, mis mõjutavad otseselt ka kogumaksumust.

Masinõppe projekti puhul on lahendusele kuluvate ressursside nagu aja, kulude ja jõudluse tasakaalustamine skaleeritava lahenduse kavandamisel ülioluline [51]. Kas valida 98% täpsusega mudeli, mille treenimine maksab 100 000 eurot, või 97% täpsusega mudeli, mis maksab 10 000 eurot [51]? Loomulikult sõltub vastus konkreetsest probleemist ning kontekstist, kuid siinkohal on oluline teada, mis on lahendatav probleem ning milline on soovitatav tulemus - sealt saab edasi arendada umbkaudset projekti maksumust ning ajakulu. Reaalsuses on võimalik esmane minimaalsetele nõuetele vastav lahendus tarnida võrdlemisi kiiresti - hinnanguline prognoos on paar nädalat. See lahendus on aga väga robustne ning tõesti äärmiselt minimalistlik. Mida täpsemat tulemust soovitakse, seda rohkem tööd ning ressursse on tarvis.

### **5.2.2 Millal lahendus end ära tasub? Mis on potentsiaalne kasum?**

Projekti tasuvuse arvutamise valem on välja toodud alloleval joonisel 7. Potentsiaalse kasumi hindamiseks on vaja defineerida kolm põhilist muutujat: esialgne arendamise maksumus, perioodi hooldamise maksumus ning perioodi kasum. Esialgse arenduse maksumuse jaoks on võimalik kasutada fikseeritud hinda ning arendada toodet või teenust edasi seni, kuni eelarve võimaldab ning seejärel otsustada edasised sammud. Arendusega seotud protsessidele on

võimalik fikseeritud hinda kergemini ette näha, andmeteaduse poole pealt on hindade prognoosimine keerukam. Läbi mingi perioodi hooldamise maksumuse saab prognoosida edasist hooldamise kulu. Lahenduse poolt tuleneva perioodi kasumi kohta saab infot läbi monitoorimise, peale mida on võimalik anda hinnanguline kasumi prognoos. STACCI kogemus näitab, et enamasti on lihtsamatel lahendustel kergem suurema kasumini jõuda.

$$\text{Perioodi potentsiaalne kasum} = \text{Esialgne arendamise maksumus} + \text{Perioodi hooldamise maksumus} - \text{Perioodi kasum}$$

Joonis 7. Potentsiaalse kasumi hinnanguline valem

### 5.2.3 Millal me lahenduse saame?

Samuti nagu on pea võimatu ilma kontekstita ette ennustada projekti maksumust, ei ole võimalik anda ühest valemit projekti ajakulu teada saamiseks. Küll aga on enamasti võimalik paari nädala jooksul tarnida esialgne lahendus, millele klient saab tagasisidet anda. Pidevjuurutuse (ingl k *Continuous Deployment*) eesmärgiks on teha valmis pigem kiire esialgne lahendus ja anda see kliendile tutvumiseks ning tagasisidestamiseks selle asemel, et arendada pika ajaperioodi vältel valmis lahendus, mida klient enne lõpliku valmimist näinud ei ole [52]. Viimast meetodit kasutades on suur oht kommunikatsiooniprobleemideks, tähtaegade ületamiseks ning kliendi huvidega mitte kattuva lahenduse välja töötamiseks. Lisaks on pidevjuurutuse plussiks kiirem arendus, kuna arendajad laevad muudatusi järjepidevalt üles ega pea selleks ootama kokkulepitud kuupäeva [52]. Tänu tihedale muutuste üles laadimistele on ka väiksem risk suurte vigade tekkeks, mille parandamine võtab palju aega [52]. STACC kasutab oma arendusmeetodina agiilset meetodit, mis on saanud tänapäeval üldtunnustatuks ning enim kasutatuks. Praktika on näidanud, et agiilset arendusmeetodi kasutades on loodav tarkvara paremas vastavuses kliendi vajadustega ning risk tähtaegade ületamise ja suurema maksumuse jaoks on väiksem [53]. Üldiselt seisab agiilse arenduse eesmärk töökorralduses, mis rõhub töötava lahenduse väljatöötamist ning võimekust tulla toime erinevate paratamatute muutuste ning ootamatustega projekti vältel [50]. Nii toimuvad nõuete kaardistamine, lahenduse kavandamine, tarkvaraarendus ja lahenduse testimine vaheldumisi [40]. Kui esmane lahendus on võimalik tarnida paari nädalaga, siis projekti kogu kestus on kindlasti pikem - esmase lahenduse jaoks on vaja kvaliteetseid andmeid ning enne lahenduse tarnimist optimeeritakse lahendust arvestades kliendi äriprotsesse ning tutvutakse valdkonna spetsiifiliste eripäradega, mis on projekti teostamisel olulised.

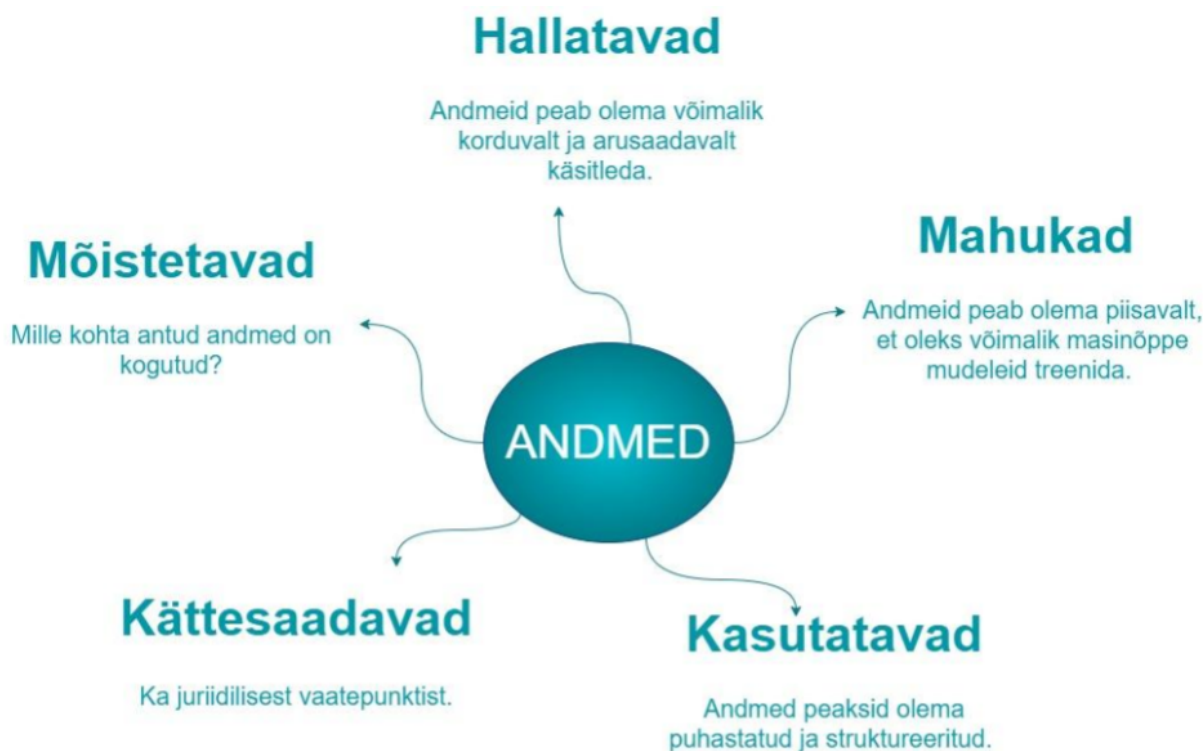
### 5.2.4 Kui palju andmeid on selleks vaja?

Tihti tekib masinõppe lahendusi kaaludes küsimus, kui palju on edukaks projektiks andmeid vaja? Tegelikult peaks alustama algusest - andmete kogumisest. Paljud ettevõtted soovivad kasutada probleemide lahendamiseks masinõpet. Isegi kui probleem on väga sobiv, võib olla probleeme andmetega. Neid kas ei ole üldse, või on need ebasobivad - igal juhul tuleb

sellistel puhkudel alustada andmete kogumisest. Teinekord on vaja projekti alustamist edasi lükata, kuna ettevõtetel on vaja täiendada olemasolevat andmemudelit või koguda rohkem andmeid. Kui andmed on olemas, saame pöörduda tagasi algse küsimuse juurde - kui palju andmeid on vaja? Andmeteaduse rusikareegel ütleb, et kui andmeid on piisavalt vähe, et neid käsitsi läbi vaadata, ei ole optimaalne masinõppe lahendusi rakendada [42]. Andmete piisavuse hindamiseks ei ole olemas üht konkreetset valemit või töötavat süsteemi. Küll aga on Majandus- ja Kommunikatsiooniministeeriumi koostatud masinõppe projekti juhendmaterjalis mõned juhtnõõrid, mis võivad aidata küsimusele umbkaudset aimdust saada. *“Iga tunnuse kohta vähemalt 20 näidet ning arvuliste väärtuste ennustamisel vähemalt 50 näidet. Ennustava mudeli loomise puhul on soovituslik kasutada andmeid vähemalt  $n+2$  printsübil ehk näiteks liiklusõnnetustes hukkunute prognoosimiseks järgneva 3 aasta jooksul võiks kasutada andmeid eelnevast 5 aastast [55].”* Üldiselt kehtib andmeteadlaste seas reegel, et mida rohkem eriilmelisi ning reaalselt mõõdetavaid andmeid, seda parem, sest mida keerulisem on projekti jaoks valitud mudel, seda enam andmeid heade tulemuste saavutamiseks vajatakse [42]. Reaalselt mõõdetavate andmete all mõeldakse andmeid, mis kirjeldaksid masinale tegelikku maailma, milles tuleb otsuseid teha. STACCI kogemusel on ettevõtted tavaliselt paari kirjega lahendatavate probleemidega võimelised ise hakkama saama. Küll aga jõutakse küsimustega STACCini, kui on probleeme paljude kirjade hulgast probleemi lahendamiseks kõige sobivamate leidmisega ning siinkohal on STACCil piisav kogemustebaas ettevõtete aitamiseks.

### **5.2.5 Milliseid andmeid on vaja?**

Masinõppe projekti õnnestumisel on andmete kvaliteedil tähtis roll - kui andmed on ebakvaliteetsed, ei saa oodata, et mudel suurepäraselt töötab ja soovitud vastuseid annab. Andmeteadlaste seas liigub väga levinud ütlus: “prügi sisse, prügi välja” - mis iseloomustab seda, et ebakvaliteetsete või vigaste sisendandmete põhjal ei saa oodata projektist mingit olulist väärtuspakkumist [43]. Alljärgneval joonisel 8 on välja toodud olulisemad aspektid andmete kvaliteedi osas. Vigaste või puudulike andmete tulemuseks on halvasti töötav mudel ja kehvad tulemused. Lisaks alltoodud viiele aspektile on oluline jälgida tunnuste relevantsust muutuja suhtes [42]. Näiteks ei ole võimalik mõne kunstiteose žanrit vaid autori abil ennustada - selleks on vaja veel tunnuseid. Ehk siis ei ole vaja rohkem autoreid, vaid mingit muud tüüpi tunnuseid. Liiklusõnnetuse toimumise ennustamiseks ei ole vajalikud sõidukijuhtide nimed, vaid hoopis andmed teelõigu seisukorrast, piirkiirusest, ilmast ja muudest mõjutajatest. Samuti on masinõppe projektis palju rohkem potentsiaali kiiremini efektiivse tulemuseni jõudmiseks, kui andmeteadlasteni jõudnud andmed on juba märgendatud. Märgendatud andmestikus on igal näitel kaks väärtust, mille abil masin ennustama õpib - tunnuse ning märgendi oma [7]. Nii saab üht tunnust teise põhjal ennustada. Märgendamine on keeruline protsess, aga STACC on võimeline ettevõtteid vajadusel konsulteerima ning aitama ka andmete märgendamisega



Joonis 8. Andmete vajalikud omadused masinõppe projekti jaoks [42].

### 5.2.6 Kui palju meie inimeste aega selle projekti jaoks vaja on? Mis peaks olema meiepoolne panus?

Eduka masinõppe projekti üheks võtmesõnaks on kommunikatsioon, seega on väga oluline lahenduse tellija poolne kaasatus. Tellija poolt on oluline oma kollektiivist määrata projektile projektijuht või tootemanik, kes oleks vastutav ning omaks ülevaadet töövoost ning selget visiooni eesmärkidest. Lisaks on vajalik valdkonna spetsialist, kellel on olemas valdkonnaspetsiifiline sisend otsuste tegemise kohta. Vähem oluline ei ole ka tehniline personal, kes vastutab süsteemide integratsiooni ning tehniliste valikute tegemise eest. Need kolm kõige olulisemat rolli on minimaalsed, mida projektis vaja on. Ei ole oluline, et nende valdkondade eest vastutaksid eraldi inimesed, vaid see, et need kolm rolli oleksid tellija poolt täidetud.

Kogemus on näidanud, et koostööaltid kliendid võivad nii lahenduse arendamise ajas ning kvaliteedis, kuna pidev kommunikatsioon annab arendajale järjepidevat tagasisidet, mis võimaldab täpsemalt kliendi soove mõista ning rakendada. Projektiks vajaliku inressursi hulk sõltub, nagu muudki ressursid, konkreetsest projektist, selle mahukusest ja keerukusest, kuid kindlasti on projekti õnnestumiseks vajalik määrata tellijapoolne projektijuht või tootemanik, valdkonna spetsialist ning tehniline personal.

### **5.3 Lahenduse kvaliteet ja toimimine**

Spetsiifilisemad küsimused seoses lahenduse süsteemi toimimise ning tulemuse kvaliteetsusega on avatud järgnevates alapeatükkides. Siiski on siin esitatud küsimustele ilma kontekstita võrdlemisi keeruline vastuseid ning hinnanguid anda. Kõige parem oleks küsimuste korral otse STACCI poole pöörduda.

#### **5.3.1 Kuidas garanteerite töökindla lahenduse?**

Sellist garantiid on vaid väga väheseid projekte alustades võimalik anda, aga kasutades pidevjuurutust (ingl k *Continuous Deployment*), on projekti õnnestumise tõenäosus suurem [52]. Antud meetodi abil keskendutakse sellele, et klient saaks võimalikult kiiresti töötavaid ja integreeritud funktsioone katsetada ja tagasisidestada selle asemel, et võtta pikk aeg ning töötada välja suur lahendus, mis lõpuks kliendi huvidega ei ühti [52]. Pidevjuurutusega käib käsikäes tihe suhtlus lahendust arendaja ning tellija vahel - see omakorda vähendab riske kommunikatsiooni probleemide tekkeks. Lisaks on annab agiilsete arendusmeetodite kasutamine tugeva võimekuse tulla toime erinevate paratamatute muutuste ja ootamatustega projekti kestel [53]. Töökindluses aitab veenduda ka pidev lahenduse monitoorimine. Kõige olulisem projekti töökindluse alustala on projekti võimalikkuses veendumine enne, kui projektiga alustatakse - seda faasi ei tasu võtta kergekäeliselt, kuna sellest sõltub projekti edukus.

#### **5.3.2 Kui hästi see lahendus töötab?**

Lahenduse kvaliteet ja töökindlus sõltub suuresti projekti planeerimise käigus tehtud analüüsist ning plaanidest. Kui projekti planeerides on paika pandud selge projekti skoop, kaardistatud nii minimaalne kui maksimaalne oodatav tulemus ning kokku lepitud kommunikatsiooniplaan, paranevad lahenduse tarnimise aeg ja kvaliteet märgatavalt. Minimaalse ning unistuste lahenduse kaardistamine vähendavad riske projekti ebaõnnestumiseks. Veel efektiivsem ebaõnnestumise vältimiseks oleks kõrvutada neid kahte varianti hetkeolukorraga. Lisaks on projekti edukuse mõõtmiseks efektiivne kasutada inimvõimekuse testimist probleemi lahendamiseks - antud mõõdikut saab edukalt kasutada hetkevõimekuse hindamiseks ning ootuste määramiseks [56]. Tabelandmete puhul saab seda efektiivsemalt rakendada, struktureerimata andmete puhul võib inimvõimekuse testimine keeruline olla, kuna masin on optimaalsem ühetaoliste andmete hindamiseks, inimene aga suudab paremini läbi paljude seoste piltidelt kassi või koera tuvastada [56]. Küll aga sõltub küsimuse täpsem vastus konkreetsest äriprobleemist, olemasolevate andmete kogusest ja kvaliteedist.

#### **5.3.3 Kuidas ma tean, miks mudel just sellise ennustuse tegi?**

Kui vaadata mõningaid konkreetseid juhtumeid, on olukordi, kus on võimalik näha, mis pani mudeli mingit ennustust tegema. Kergesti mõistetava struktuuriga mudeleid, nagu lineaarne regressioon või otsustuspuud, on tavaliselt võimalik tõlgendada lähtuvalt mudeli disainist ning struktuurist [57]. Teisalt on ka olukordi, mil mudelite tõlgendatavuses kasutatakse

väljendit “must kast” - isegi perfektselt tehtud ennustuse puhul on ka mudeli loojal võimatu aru saada, kuidas mudel antud tulemuseni jõudis [24]. Taolised mudelid teevad keerukaks tuhandete või isegi miljonite parameetritega sügavad närvivõrgud ning nende tõlgendamiseks on vaja keerulisi väliseid mudeli diagnostika meetodeid [57]. On välja töötatud meetodeid, mille abil on võimalik taastada miks mudel tegi just selle konkreetse otsuse, kuid keeruliste mudelite puhul on see enamasti võrdlemisi kallis ning ajamahukas protsess ning kulutatud ressurss ei pruugi lõpuks ikkagi mingit tulemust tuua [53]. Seega on endiselt palju nõ “musta kasti” mudeleid, mille töökulgu ei osata seletada [57]. Üks on aga kindel - mudelid teevad ennustusi konkreetsetele neile ette antud andmete põhjal, seega põhjendusi saab otsida andmetest. Kui mudeli ennustus ei ole soovipärane, tuleb teha andmetes korrekture - olenevalt juhtumist kas andmeid lisada, eemaldada või korrigeerida.

### **5.3.4 Kuidas ma tean kui hästi see lahendus töötab?**

Lahenduse tööd on võimalik hinnata hetkeseisuna või pikema ajaperioodi vältel. Selleks, et hinnata lahenduse töökindlust praeguses hetkes, on seda võimalik teha kahe mõõdiku abil. Kõigepealt kvantitatiivse ennustamise läbi - panna treeningandmetest kõrvale osa andmeid, mida hiljem kasutada tulemuse ennustatavuse hindamisel. Teine variant on kvalitatiivne - on võimalik lasta kliendil endal mudeli tööd testida ning hinnata. Siinkohal võib saada takistavaks asjaoluks inimese emotsionaalsus, kuna antud meetod ei ole metoodiline ning inimene hindab mudelit eelarvamuslikult, mitte neutraalselt. Lahenduse töö mõõtmisel pikema perioodi vältel on oluline lahenduse tööd monitoorida, et andmetriiv (ingl k *data drift*) ei mõjutaks läbi andmete aegumise lahenduse tulemusi. Andmete aegumine võib juhtuda olukordades, kus ettevõttel tekivad uued kliendid, uued tooted või toimuvad ennustuskeskkonna sisesed muudatused. Selliste olukordade vältimiseks on äärmiselt oluline kasutada hästi välja töötatud monitoorimise lahendusi, mida pakub ka SATCC, eesmärgiga hoida arendatud lahendused efektiivselt töös ka homses maailmas.

## **5.4 Detailid**

Lahenduse toimimisest veelgi detailsemate murede hulka kuuluvad tehnilisemad küsimused nagu näiteks projekti jaoks programmeerimiskeele valimine, selle integreerimine olemasoleva süsteemiga ning projekti jaoks kõige paremini sobiva meetodi valimine. Meetod on justkui projekti süda, kuna sellel on tulemusel kõige määravam roll.

### **5.4.1 Miks Python?**

Python on juba aastaid olnud eelistatuim programmeerimiskeel teaduslikus andmetöötuses, andmeteaduses ning masinõppes ning lihtsalt õpitava ja kasutatava keelena jääb see antud valdkondade kasutatavaks programmeerimiskeeleks ka lähiaastatel [58]. Suurimaks takistuseks Pythoni kasutamisel näevad ettevõtted tavaliselt integratsiooniprobleeme juhu, kui ülejäänud süsteemid põhinevad mõnel muul programmeerimiskeelel. Küll aga on integreerimiskeerukused lahendatavad API-de ning konteinerite abil. Pythoni eelisteks peetakse kõige laiemaid masinõppe jaoks saadaolevaid teke ning suurt olemasolevat

kogukonda, mis võimaldab tugineda teiste kogemustele ning kasutada juba loodud lahendusi. Lisaks on tänu kogukonnale hea lahendada projekti jooksul ette tulevaid probleeme, kuna alati on võimalik küsida ning teiste koodinäiteid vaadata. Selline toimimine on masinõppe lahenduse tellijale väga kasulik, kuna nii hoitakse kokku palju arendusaega ja sellega ka tellija raha. Kokkuvõtlikult on Python masinõppe jaoks kõige kiirem ja odavam programmeerimiskeel. Tänu edusammudele GPU võimsustes ning üha kasvavatele kasutajate kogukondadele ja loodavatele teekidele arvatakse, et Python jääb paljudeks aastateks teadusarvutuste domineerivaks keeleks [58].

#### 5.4.2 Milline meetod valida?

Masinõppe algoritme on väga erinevad ning ei ole olemas ühte “parimat”, mille võiks valida - kõik sõltub probleemist, mida soovitakse lahendada. Masinõppe algoritme eristatakse vastavalt sellele, kuidas need õpivad ning missuguseid detaile seeläbi omandavad [7]. Andes erinevatele algoritmidele samad andmed, jõuavad need erinevatele tulemustele, kuna optimeeritakse erinevaid aspekte [7]. Ei ole olemas ühte meetodit, mida saaks rakendada kõikide erinevate probleemide lahendamiseks, parimani jõutakse olulisi aspekte arvestades [51]. Meetodi valikul olulised aspektid, mida peaks arvesse võtma, on ära toodud joonisel 9.



Joonis 9. Mudeli valimine [48]

Näiteks ei ole ilma ennustades suur probleemi, kui ennustust tehes eksitakse vähesel määral, küll aga juhul, kui kasvõi ühe korra väga suuresti eksitakse - inimese usaldus langeb [7]. Sporditulemuste ennustamisel on olukord vastupidine - eksimuste suurus ei ole määrav, vajalik on teada võitjat ning kaotajat [7]. Üldiselt on mõistlik kasutada lihtsamaid ning paremini tõlgendatavaid meetodeid juhuks, kui probleemi olemus seda võimaldab [18]. Tihti aga pakuvad keerulised meetodid suuremat ennustustäpsust kui lihtsamad ja paremini tõlgendatavad variandid ning sellistel juhtudel tasuks võtta olulisimaks valikukriteeriumiks meetodi täpsus [51]. Konkreetse probleemi jaoks kõige sobivama meetodi leidmiseks analüüsib andmeteadlane erievaadid algotirme, nende parameetreid ning mida ja kuidas tuleb õppida [7]. Kokkuvõtteks võib öelda, et võimalikult lihtne, kiire ja vähe ressursse nõudev meetod, mis saavutab aktsepteeritava täpsuse, on kõige mõistlikum lahendus.

## **5.5 Töö tulemuste kokkuvõte**

Antud magistritöö fookuseks võetud sisendi kogumine ning masinõpet tutvustavate materjalide kogumiku esmase versiooni väljatöötamise käigus selgitati välja 15 küsimust, mis oleks kõige efektiivsemad potentsiaalsete STACCI klientide küsimustele ning hirmudele vastama. Antud töö puuduseks võib tuua puuduliku tulemuste valideerimise sihtrühma peal. Teisalt saab tugevustena näha sisendit pikaajase kogemusega osapooltelt, kelle hulgas on nii andmeteadlasi, andmeinsenere, projektijuhte kui ärilise poole spetsialiste - see annab töö sisendile tugeva aluse, mis omakorda tõstab tulemuse kvaliteeti. Edaspidi oleks mõistlik viia läbi esmase materjali põhjalik valideerimine sihtrühma peal, analüüsida valideerimise tulemusi ning viia sisse analüüsist selgunud muutmist vajavad aspektid materjalide kogumiku esmases versioonis. Valideerimine on oluline teada saamaks, kuidas inimesed vastavad antud mahus vastustele - kas selline informatsiooni hulk on paras, liiga väike või hoopis liiga mahukas. Lisaks on vaja kinnitust, et koostatud materjal tõepoolest vastab sagedasematele potentsiaalse kliendi küsimustele. Soovi korral oleks võimalik valminud materjali sisu abil kujundada füüsilised infoaldikud ning digitaalsed kirjutised sotsiaalmeedias avaldamiseks. Pärast valideerimis- ning täiendusprotsessi oleks materjali võimalik kasutada STACCI teenuste turundamisel, kus oleks võimalik antud materjali kasutada näiteks erinevatel üritustel flaierite näol, STACCI enda inimeste turundusalase silmaringi laiendamiseks või potentsiaalse kliendiga läbirääkimisi pidades.

## 6 Kokkuvõte

Tehisintellekti lahenduste rakendamise kaalumise tekitab ettevõtjates kahtlusi, kas investering uudsesse ning kallisse lahendusse on ikka seda väärt. Antud töö raames kaardistati ning avati andmeteaduse ettevõtte STACC OÜ vajaduste põhjal potentsiaalsete klientide jaoks masinõppel põhinevate süsteemide kohta esinevaid küsimusi. Kuna hetkel puudub STACCis antud teemade süstemaatiline kaardistamine ning analüüs, oli töö eesmärgiks nimetatud probleemid kaardistada ning nende põhjal töötada välja kõige sagedamini esinevate küsimustega materjali esmane versioon, mida oleks võimalik peale selle valideerimist kasutada STACCi teenuste efektiivsemaks turundamiseks. Järgnevalt on välja toodud töö alguses püstitatud uurimisküsimuste vastused:

1. Ettevõtete peamisteks hirmudeks seoses masinõppega saab välja tuua rahaliste ressursside puuduse ning selge ja konkreetset piiritletud äriprobleemi olemasolu. Lisaks muretsetakse andmete piisava mahu ja kvaliteedi pärast.
2. Kõige enam küsimusi seoses masinõppe rakendamise võimalustega tekib lahenduse süsteemi, kvaliteedi ja võimekusega seotud teemade ning erinevate projektiks vajaminevate ressursside suuruse kohta. Lisaks ka erinevate ressursside nagu aeg, raha, andmed ning inimressursi vajaduse kohta.
3. STACCi klientide protsessi efektiivistamiseks töötati välja klientide peamisi hirme, küsimusi ning olulisi teemasid katva materjali esmane versioon, mis võimaldab iseseisvalt tekkinud hirme maandada ning küsimustele vastata. Samuti kaeti materjalisis teemad, mille peale kliendid ise tihti ei tule, kuid mis spetsialistide sõnul võivad projekti kulgu takistada.

Töö alguses püstitatud uurimisküsimused said töö käigus vastatud ning käesoleva magistritöö tulemuseks on STACCi potentsiaalsete klientide kõige sagedamini esinevaid küsimusi ning hirme katva materjali esmane versioon, mille abil on tulevikus võimalik STACCi pakutavaid teenuseid efektiivsemalt turundada. Koostatud materjal kannab kaht põhilist funktsiooni: aitab iseseisvalt potentsiaalse kliendi hirme maandada ja klienti veenda tehisintellekti lahenduste võimekuses efektiivselt äriprobleeme lahendada. Teiseks kaetakse materjalisis projekti õnnestumise seisukohast olulisi küsimusi, mille peale kliendid ise tihti ei tule ning tuuakse välja näited, millistel juhtudel probleemile ei ole võimalik masinõppe abil lahendust pakkuda.

Tööd oleks võimalik edasi arendada tulemuste valideerimise näol. Valideerimine oleks võimalik läbi viia STACCi turundustöö käigus, kaasates välja töötatud materjale üritustel, läbirääkimistel ning üldises turundustöös.

## 7 Viidatud kirjandus

- [1] Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/J.IJINFOMGT.2019.01.021> (31.03.2022)
- [2] Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne. (2019). [https://www.kratid.ee/files/ugd/0b32e3\\_9e397d14453b454db0b8d3615a7012ba.pdf](https://www.kratid.ee/files/ugd/0b32e3_9e397d14453b454db0b8d3615a7012ba.pdf) (22.03.2022)
- [3] Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? Handling editor Seye Abimbola. *BMJ Glob Health*, 3, 798. <https://doi.org/10.1136/bmjgh-2018-000798> (12.04.2022)
- [4] IT terministandardi sõnastik. (2021). *Eesti Keele Instituut*. <http://www.eki.ee/dict/its/index.cgi?Q=tehisintellekt&F=M&C06=et&C10=1> (05.04.2022)
- [5] Muthukrishnan, N., Maleki, F., Ovens, K., Reinhold, C., Forghani, B., & Forghani, R. (2020). Brief History of Artificial Intelligence. *Neuroimaging Clinics of North America*, 30(4), 393–399. <https://doi.org/10.1016/J.NIC.2020.07.004> (31.03.2022)
- [6] Bhageshpur, K. (2019) Data Is The New Oil - And That's A Good Thing. *Forbes*, 15.november. <https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=6b5bbecd7304> (14.04.2022)
- [7] Sügis, E., Allikivi, M.-L., Fišel, M., Kolde, R., Kruusamäe, K., Leontjeva, A., Luhtaru, A., Matiisen, T., Paales, M., Reisberg, S., Sharma, R., Tampuu, A., Tommasini, R. (2020). *Tehisintellekti algkursus*. Tehisintellekti algkursuse materjalid veebis. Tartu Ülikool. [https://courses.cs.ut.ee/2020/Tehisintellekti\\_algkursus/spring/Main/PARTITehisint](https://courses.cs.ut.ee/2020/Tehisintellekti_algkursus/spring/Main/PARTITehisint) (20.04.2022)
- [8] Batin, M., Zhila, A., & Denkenberger, D. (2017). Artificial Intelligence in Life Extension: from Deep Learning to Superintelligence. In *Informatica*, 41, 401-417. [https://www.researchgate.net/publication/322482979\\_Artificial\\_intelligence\\_in\\_life\\_extension\\_From\\_deep\\_learning\\_to\\_superintelligence](https://www.researchgate.net/publication/322482979_Artificial_intelligence_in_life_extension_From_deep_learning_to_superintelligence) (21.04.2022)
- [9] Willems, J., Schmidhuber, L., Vogel, D., Ebinger, F., & Vanderelst, D. (2022). Ethics of robotized public services: The role of robot design and its actions. *Government Information Quarterly*, 39(2), 101683. <https://doi.org/10.1016/J.GIQ.2022.101683> (14.04.2022)
- [10] (2020). Tehisintellekti võimalused ja ohud. *Euroopa Parlament Uudised*, 20.oktoober. <https://www.europarl.europa.eu/news/et/headlines/society/20200918STO87404/tehisintellekti-voimalused-ja-ohud> (12.04.2022)
- [11] Velsberg, O. (2018). Mida teevad Eesti kratid? *Medium*, 23.juuli. <https://medium.com/digiriik/kuidas-kratt-eesti-paremaks-teeb-cfebc526cb47> (22.03.2022)
- [12] Kaevats, M. (2018). Kuidas kratt Eesti paremaks teeb? *Medium*, 27.detsember. <https://medium.com/digiriik/kuidas-kratt-eesti-paremaks-teeb-cfebc526cb47> (22.03.2022)

- [13] Hern, A. (2016) Google AI in landmark victory over Go grandmaster. *The Guardian*, 27.jaanuar. <https://www.theguardian.com/technology/2016/jan/27/google-hits-ai-milestone-as-computer-beats-go-grandmaster> (14.04.2022)
- [14] Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383. <https://doi.org/10.1016/J.GIQ.2018.09.008> (12.04.2022)
- [15] Pungas, T. (2017). Masinõppe mittetehniline ülevaade. *Pungas*, 29.jaanuar. <https://pungas.ee/masinope-mittetehniline-ulevaade/> (05.04.2022)
- [16] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*, 8, 53. <https://doi.org/10.1186/s40537-021-00444-8> (12.04.2022)
- [17] Microsoft Azure (2022). Microsoft. <https://azure.microsoft.com/en-us/overview/artificial-intelligence-ai-vs-machine-learning/#introduction> (14.04.2022)
- [18] Marr, B. (2020) The Top 10 Artificial Intelligence Trends Everyone Should Be Watching In 2020, 6. jaanuar. <https://www.forbes.com/sites/bernardmarr/2020/01/06/the-top-10-artificial-intelligence-trends-everyone-should-be-watching-in-2020/?sh=5bb9bea6390b> (11.05.2021)
- [19] Duhigg, C. (2012) How Companies Learn Your Secrets, *The New York Times Magazine*, 16.veebbruar. <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> (19.04.2022)
- [20] Johnson, M., Ghuman, P. (2021). Pimenägemine. (Enamasti) varjatud viisid, kuidas turundus meie aju kujundab. Tallinn.: AS Äripäev.
- [21] Kangur S. Meeste ja naiste isiksuseomaduste erinevused. TÜ psühholoogia instituudi magistritöö. 2012. [http://dspace.ut.ee/bitstream/handle/10062/29834/kangur\\_sirli.pdf](http://dspace.ut.ee/bitstream/handle/10062/29834/kangur_sirli.pdf) (19.04.2022)
- [22] Noujaim, J., Amer, K. (2019, 26. jaanuar). The Great Hack [Video]. Netflix. <https://www.netflix.com/ee/title/80117542> (16.02.2022)
- [23] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). *Explainable AI: A Review of Machine Learning Interpretability Methods*. <https://doi.org/10.3390/e23010018> (13.04.2022)
- [24] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. <http://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf> (14.04.2022)
- [25] Cellan-Jones, R. (2014) Stephen Hawking warns artificial intelligence could end mankind, *BBC News*, 2. detsember. <https://www.bbc.com/news/technology-30290540> (10.04.2022)
- [26] Rawlinson, K. (2015) Microsoft's Bill Gates insists AI is a threat, *BBC News*, 29. jaanuar. <https://www.bbc.com/news/31047780> (19.04.2022)
- [27] Koosapoeg, K. (2019). Tõenduspõhine otsustamine – kuidas alustada BI-ga? *TWN Blog*, 26.oktoober. [https://blog.twn.ee/et/Toenduspohine\\_otsustamine](https://blog.twn.ee/et/Toenduspohine_otsustamine) (30.03.2022)
- [28] STACC OÜ koduleht. (2022). *STACC OÜ*. <https://stacc.ee/et/meist/> (22.03.2022)
- [29] Tepandi, J. (2021) *Tarkvara protsessid ja kvaliteet*. TTÜ kursuse lühiülevaade. <http://tepani.ee/tnks-loeng.pdf> (26.03.2022)

- [30] Jokinen, T. (s.a.) Tootelooma ja leiundus. Kursuse materjal TTÜ Elektriainite ja jõuelektroonika instituudis.  
<http://www.ene.ttu.ee/elektriamid/oppeinfo/AAV3310/335.htm> (22.03.2022)
- [31] Method 1: Brainstorming. The Human Rights Education Handbook: Effective Practices for Learning, Action, and Change. (2000)  
<http://hrlibrary.umn.edu/edumat/hreduseries/hrhandbook/methods/1.htm> (22.03.2022)
- [32] Brennen, B. S. (2013). Qualitative research methods for media studies. New York and London: Routledge
- [33] Lepik, K., Harro-Loit, H., Kello, K., Linno, M., Selg, M., Strömpl, J. (2014). *Sotsiaalse analüüsi ja metodoloogia õpibaas - Intervjuu*. Ühiskonnateaduste instituudi sotsiaalteaduslike andmekogumise ja -analüüsi meetodite ning vahendite veebiõpik. Tartu Ülikool. <https://samm.ut.ee/intervjuu> (22.03.2022)
- [34] Method 21: Presentations. The Human Rights Education Handbook: Effective Practices for Learning, Action, and Change. (2000)  
<http://hrlibrary.umn.edu/edumat/hreduseries/hrhandbook/methods/21.htm> (22.03.2022)
- [35] Rootamm-Valter, J (2019). Dokumendid. Kvalitatiivsed uurimismeetodid sotsiaalteadustes. Tartu Ülikool. <https://sisu.ut.ee/kvalitatiivne/dokumendid> (27.04.2022)
- [36] Dorard, L. (2019). The Machine Learning Canvas. Prantsusmaa: OwnML.
- [37] Dorard, L. (2022). The Machine Learning Canvas. Masinõppe Lõuend - projekti läbiviimise abimaterjal internetist. <https://www.ownml.co/machine-learning-canvas> (28.04.2022)
- [38] Lauk, M. (2015). Business analyst's role in PricewaterhouseCoopers. Nortal's TechDay 2015. Tallinn, 22.05.2015. Nortal. <https://vimeo.com/129636896> (31.03.2022)
- [39] Sepp, T. (2022). Kratisõbraliku tarkvaraarenduse töötuba. Tark e-riik. Tallinn, 11.01.2022. Riigi Infosüsteemi Amet. [https://youtu.be/UjMEVbZKQ\\_A](https://youtu.be/UjMEVbZKQ_A) (31.03.2022)
- [40] Milani, F. (2019). Digital Business Analysis (1st ed.). Cham: Springer International Publishing
- [41] Sepp, T. (2022). Tehisintellekti rakendamise 1. töötuba (1. osa). Tark e-riik. Tallinn, 25.01.2022. Riigi Infosüsteemi Amet. <https://youtu.be/2HfOHHUq8VE> (31.03.2022)
- [42] Särg, D. (2022). Tehisintellekti rakendamise 1. töötuba (2. osa). Tark e-riik. Tallinn, 25.01.2022. Riigi Infosüsteemi Amet. [https://youtu.be/f6y2nK9P\\_sg](https://youtu.be/f6y2nK9P_sg) (31.03.2022)
- [43] Laaneväli, A. (2022). Tehisintellekti rakendamise 2. töötuba (1. osa). Tark e-riik. Tallinn, 14.02.2022. Riigi Infosüsteemi Amet. <https://youtu.be/OUBLQm7N3-g> (31.03.2022)
- [44] Luna, J. C. (2022). Top programming languages for data scientists in 2022. Datacamp blog, 21. jaanuar.  
<https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022> (03.05.2022)
- [45] Zachariasz, K. (2018). When to Use Machine Learning - Does Your App Really Need ML? Netguru Blog, 28. veebruar.  
<https://www.netguru.com/blog/when-to-use-machine-learning> (14.05.2022)
- [46] TalTech Inseneriteaduskond (2022, 26. aprill). AIRE - AI & Robotics Estonia korraldab sel nädalal koolituse „Masinõpe ja selle rakendused tööstusele ressursi- ja

- energiatõhususe tõstmiseks“, mille üheks koolitajaks on meie prodekaan professor Argo Rosin. [Facebooki postitus]. Facebook.  
<https://www.facebook.com/InseneriTeaduskond/posts/2156386604510818>  
 (04.05.2022)
- [47] STACC OÜ koduleht. (2022). Avaliku sektori lahendused.  
<https://stacc.ee/et/ai-lahendused/avalik-sektor/> (14.05.2022)
- [48] Arandi, I. (2019). Kuidas optimeerida turunduskulusid – masinõppega muidugi. Masinõpe, 2. mai.  
<https://masinope.ee/masinoppimine/turunduskulude-optimeerimine-masinoppega/>  
 (04.05.2022)
- [49] Malsam, W. (2021). What Is Proof of Concept (POC)? Definition, Steps & Best Practices. ProjectManager, 1. detsember.  
<https://www.projectmanager.com/blog/proof-of-concept-definition> (14.05.2022)
- [50] Arandi, I. (s.a.).Kuidas alustada masinõppe rakendamise? Masinõpe, 2. mai.  
<https://masinope.ee/rakendamise/> (04.05.2022)
- [51] Valdarrama, S. (2021) Considerations when choosing a machine learning model, Towards Data Science, 13. juuli.  
<https://towardsdatascience.com/considerations-when-choosing-a-machine-learning-model-aa31f52c27f3> (11.05.2021)
- [52] Pittet, S. (s.a.) How to get started with Continuous Integration. *Atlassian Blog*.  
<https://www.atlassian.com/continuous-delivery/continuous-integration/how-to-get-to-continuous-integration> (26.04.2022)
- [53] Kippar, J., Petuhhov, I., Rinde, A. (2020) Tarkvaraarenduse projekt. Tallinna Ülikooli kursuse slaidid.  
[http://www.cs.tlu.ee/~inga/TTP/Agiilsed\\_meetodid\\_2019.pdf](http://www.cs.tlu.ee/~inga/TTP/Agiilsed_meetodid_2019.pdf) (26.04.2022)
- [54] Tepandi, J. (2021) Tarkvara protsessid ja kvaliteet. TTÜ kursuse lühiülevaade.  
<http://tepandi.ee/tks-loeng.pdf> (26.03.2022)
- [55] Masinõppe projektiks valmistumine:kontroll küsimustik. (2020).  
[https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f\\_3f8941981bb743a68c1d64406e29b4df.pdf](https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f_3f8941981bb743a68c1d64406e29b4df.pdf) (05.05.2022)
- [56] Rushiti, B. (2021). Human-Level Performance and Bayesian Optimal Error. Medium, 18. juuni.  
<https://bardhrushiti.medium.com/human-level-performance-and-bayesian-optimal-error-fadf4f55cd48> (14.05.2022)
- [57] Perteneder, F. (2021) Understanding Black-Box ML Models with Explainable AI, *Dynatrace Research Blog*, 29. aprill.  
<https://engineering.dynatrace.com/blog/understanding-black-box-ml-models-with-explainable-ai> (10.05.2022)
- [58] Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, 11(4), 193. <https://doi.org/10.3390/info11040193>  
 (10.05.2022)

## 8 Lisad

### I. 8.1 Lisa 1. STACCI sisese info kogumise originaalmaterjalid.

#### 1. Why would a customer want an ML solution?

1. To help their customers.
2. To keep their customers.
3. Building a Data Science team on your own costs much and you might only need them seasonally or for a short period of time.
4. For marketing purposes - this is not a good reason to start a project for us.
5. For funding (Government funds to learn whether the solution is working or not)
6. Often they want ML but don't actually need it - we can sell them just some data science parts.

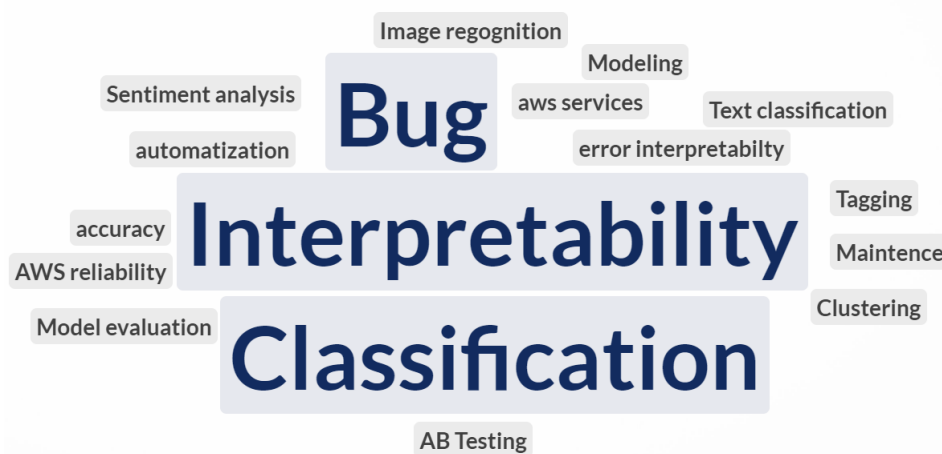
#### 2. What are the clients afraid of?

1. Too expensive - How much does it cost?
2. When will it repay me?
3. The main competitor gets good results before me and I will be too late.
4. When they use ML, then they are frustrated bc they dont understand - interpretability
5. Maintenance - random errors

#### 3. What special values do we offer to our customers?

1. Years of experience.
2. Knowledge about competitors.
3. Diversity between the teams and backgrounds.
4. We use open source software.
5. Stability, security and budget friendly (because there are no hidden fees) - GPL and MIT license
6. We give the rights to the code.
7. Infrastructure costs are low - HPC is much cheaper than Amazon.

#### 4. Name a topic about ML that you have been explaining to a customer.



**5. What are the main questions you have received from a customer about ML?**

1. Do we even need ML?
2. When can we have it?
3. Can our model predict another upcoming pandemic?
4. Can you be sure it works?
5. I don't like this recommendation but all the others are fine.
6. What data do I need to add to make the machine work?
7. Potential profit?
8. Why does the model classify it as "plastics"?
9. How to use it?
10. How does it work?
11. When will it pay for itself?
12. How much does it cost?
13. How much data is needed?
14. When will it repay me?
15. How well does it work?
16. How much does it cost?

**6. What do you think are the main causes why the potential clients back down from cooperation with us?**

1. Uncertain outcome.
2. Hourly payment structure.
3. Python is not acceptable.
4. They need to mature their data model before they can hire us.
5. The competitors have a product, whereas we would have to start developing.
6. Not sure if it is solvable with ML
7. Too expensive.
8. Too strict requirements from the client side.
9. ML is not applicable.
10. Too high expectations.
11. Better offer from a competitor.
12. They are not sure if it will help them
13. The employees have no experience in certain fields.
14. Not enough data.
15. Money.

## II. 8.2 Lisa 2. STACCI sisese info kogumise materjalid eesti keelde tõlgituna.

### 1. Miks peaks klient soovima kasutada masinõppe lahendust?

1. Selleks, et oma kliente aidata.
2. Selleks, et oma kliente säilitada.
3. Kuna enda andmeteaduse meeskonna ehitamine on liiga kallis ning seda võib vaja minna vaid hooajaliselt või lühikese ajaperioodi jooksul.
4. Turunduslikel eesmärkidel - mis ei ole iseenesest hea põhjus meiega projekti alustamiseks.
5. Rahastuse saamiseks (Näiteks riiklikud fondid rahastavad teadust olenemata sellest, kas tulemus töötab või mitte).
6. Tihti tahetakse masinõpet, kuid tegelikult seda ei vajata - sellisel juhul saame me müüa vaid mõnda andmeteaduse osa.

### 2. Mida kliendid kardavad?

1. Liiga kallis [lahendus] - Kui palju see maksab?
2. Millal see minu jaoks ära tasub?
3. Peamine konkurent jõuab heade tulemusteni enne mind ja siis olen temast maha jäänud.
4. Tõlgendatavus - Kui nad [kliendid] kasutavad masinõpet, siis on neil tihti keeruline nendest süsteemidest aru saada ja see tekitab pahameelt ja hirmu.
5. Süsteemi hooldamisel tekkivad probleemid.

### 3. Milliseid erilise väärtusi meie oma klientidele pakume?

1. Aastatepikkust kogemust.
2. Oma konkurentide tundmist.
3. Meeskondadesisest mitmekesisust ja inimeste erinevat tausta.
4. Me kasutame avatud avatud lähtekoodiga tarkvara.
5. Stabiilsus, turvalisus, rahakotisõbralikkus (kuna pole peidetud lisatasusid) - GPL, MIT litsensid.
6. Anname õigused koodile.
7. Infrastruktuuri tasud on madalad - HPC on oluliselt odavam kui Amazon.

### 4. Nimeta masinõppega seotud teema, mida oled pidanud kliendile üle seletama.



**5. Millised on peamised küsimused, mida oled kliendilt masinõppega seoses saanud?**

1. Kas me üldse vajame masinõpet?
2. Millal me selle [lahenduse] saame?
3. Kas mudel suudab ennustada järgmist pandeemiat?
4. Kas võite kindlad olla, et see [lahendus] töötab?
5. Mulle ei meeldi see soovitus, ülejäänud on normaalsed.

**6. Mis andmeid pean ma lisama, et masin tööle hakkaks?**











1. Potentsiaalne kasum?
2. Miks mudel selle plastiku alla klassifitseerib?
3. Kuidas ma seda [lahendust] kasutan?
4. Kuidas see [lahendus] töötab?
5. Millal see [lahendus] ennast ära tasub?
6. Kui palju see [lahendus] maksab?
7. Kui palju andmeid on vaja?
8. Millal see [lahendus] minu jaoks ära tasub?
9. Kui hästi see [lahendus] töötab?
10. Kui palju see [lahendus] maksab?


**7. Mis on Sinu arvates peamised põhjused, miks potentsiaalne klient meiega koostööd ei alusta?**

1. Ebakindel tulemus.
2. Tunnipõhine maksesüsteem.
3. Python ei ole aktsepteeritav.
4. Enne meie palkamist peavad nad oma andmemudelit täiendama.
5. Konkurentidel on toode olemas, meie aga alles hakkaksime arendama.
6. Pole kindel, kas nende probleem on masinõppega lahendatav.
7. Liiga kallis.
8. Liiga karmid tingimused kliendi poolt.
9. Masinõpe ei ole rakendatav.
10. Liiga kõrged ootused.
11. Parem pakkumine konkurendi poolt.
12. Ei olda kindlad, kas me saame neid aidata.
13. [STACCI] Töötajatel puudub kogemus teatud valdkondades.
14. Pole piisavalt andmeid.
15. Raha.

### III. 8.3 Lisa 3. Masinõppe Lõuend [37]

**Masinõppe Lõuend (v0.4)** Kellele disainitud: \_\_\_\_\_ Kelle poolt disainitud: \_\_\_\_\_ Kuupäev: \_\_\_\_\_ Iteratsioon: \_\_\_\_\_

<p><b>Otsused</b> </p> <p>Kuidas kasutatakse ennustusi selliste otsuste tegemiseks, mis pakuvad lõppkasutajale soovitud väärtust?</p>	<p><b>Masinõppe ülesanne</b> </p> <p>Sisend, prognoositav väljund, probleemi tüüp.</p>	<p><b>Väärtus-pakkumised</b> </p> <p>Mida püütakse ennustussüsteemi lõppkasutaja(te) heaks teha? Milliseid eesmärke tarnitakse?</p>	<p><b>Andmeallikad</b> </p> <p>Milliseid töötlemata andmeallikaid on võimalik kasutada (sisemisi ja välised)?</p>	<p><b>Andmete kogumine</b> </p> <p>Kuidas saadakse uusi andmeid, millest õppida (sisendeid ja väljundeid)?</p>
<p><b>Ennustuste tegemine</b> </p> <p>Millal me tehakse prognoos uute sisendite kohta? Kui kaua kulub uue andmevormide teisendamiseks ja ennustuste tegemiseks?</p>	<p><b>Hindamine ilma võrgu-ühenduseta</b> </p> <p>Meetodite ja moodsikute kaardistamine süsteemi hindamiseks enne juurutamist.</p>		<p><b>Tunnused</b> </p> <p>Toorandmete allikatest eraldatud sisendi esitusvorm.</p>	<p><b>Mudelite loomine</b> </p> <p>Millal luuakse või uuendatakse mudeleid uute treeningandmetega? Kui pika aja tagant treenitakse uus mudel?</p>
	<p><b>Otseülekanne ja monitoorimine</b> </p> <p>Meetodid ja moodsikud süsteemi hindamiseks pärast kasutuselevõttu ja väärtuse loomise kvantifitseerimiseks.</p>			

[machinelearningcanvas.com](http://machinelearningcanvas.com) Louis Dorard, Ph.D. Litsentsitud Creative Commons Attribution-ShareAlike 4.0 rahvusvahelise litsentsi alusel. 

#### IV. 8.4 Lisa 5. Masinõppe projektiks valmistumine: kontroll küsimustik [54]

- Andmete ettevalmistel on kõik toimingud (teisendused, treening- ja testandmeteks jagamine, puuduvate tunnuste täitmine jne) dokumenteeritud.
- Algsed (toor)andmed on võimalikult muutumatul kujul salvestatud ja kättesaadavad. Oluline kuna võimaldab vigade korral alustada nn puhtalt lehelt.
- Kas andmed on tõesed/usaldusväärsed? (Mudeli treenimisel on oluline, et andmed, sealhulgas märgendid, oleksid õiged. Vastasel juhul ei saa mudel anda häid tulemusi.)
- Kas andmed on hõlpsasti juurdepääsetavad ja projektis kasutatavad?
- Kas andmed on kirjeldatud? (Andmete kirjeldamisel on oluline, et andmeelemendi kirjeldus kataks järgnevat: andmeelemendi tähis, nimetus, lühikirjeldus, andmeelemendi tüüp, andmeelemendi kasutamise/kogumise algus ning vajadusel mõõtmise kirjeldus ning mõõtühik.)
- Kas andmed on läbivalt kokkulepitud kirjaviisiga? (Näiteks kirjavigade puudumine, ühtne suure ja väikese algustähe kasutus, ühesugune asutuste nimede kasutus jne.)
- Kas kõigil andmeväljadel on läbivalt sama tähendus?
- Kas andmed sisaldavad sihtmärgi ennustamist võimaldavat teavet ehk kas olemasoleva põhjal on võimalik ennustada sihtmärki?
- Kas andmed on ennustamiseks piisava sageduse/kaetusega?
- Kas puuduvad väärtused moodustavad sisendist väikese osa?
- Kas puuduvate väärtuste põhjused on teada?
- Kas andmed on märgendatud?
- Kas andmeid on piisavalt?
- Kas andmed on nihketa?

## **Litsents**

### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Kaisa Ehrpais,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Tehisintellekti tutvustavate materjalide kogumiku väljatöötamine STACC OÜ näitel”, mille juhendajad on Kerli Mooses PhD ja Karl-Oskar Masing BSc, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kaisa Ehrpais

**17.05.2022**