

# Juhendmaterjal krattide hankimiseks

November 2019

## Sisukord

Terminoloogia .....	3
Juhendmaterjalist .....	4
Soovitused – lühikokkuvõte .....	5
Soovitused detailsemalt kirjeldatult .....	8
1. Krati testprojekti tulemuseks saab olla ainult teadmine (sh. raport, info algoritmidest kohta, mis töötavad ja mis ei tööta, nimekiri andmete kohta).....	8
2. Hankimisel ei tasu keskenduda ühele kindlale lahenduskäigule, vaid tuua välja probleemid ning võimaldada projekti skoopi iteratiivselt muuta. ....	9
3. Andke ülevaade olemasolevatest andmetest.....	10
4. Hankimisel arvestage turuolukorraga.....	11
5. Võimaldage modernsete tehnoloogiate ja infrastruktuuri (pilvandmetöötlus) kasutamist.....	11
6. Testprojektide puhul jätke oma võimalus projekt katkestada. ....	11
7. Arvestage, et pakkumiste koostamine on pakkujate jaoks töö ning nõuded pakkumises ja riskid töö teostamisel vähendavad pakkujate hulka ja tõstavad hinda ...	12
8. Arvestage, et krati projektid nõuavad tellijalt aktiivset panustamist kogu projekti vältel.....	12
Näide andmeteadeuse testprojekti hankedokumendi „Tehniline kirjeldus“ ülesehituse kohta .....	13

# Terminoloogia

Alljärgnevas osas on väljatoodud olulisemad masinõppe ja andmeteaduse valdkonna terminid:

**Andmetunnus** (*Feature*) – sisendatribuut, mida kasutatakse ennustamiseks

**Hajusarvutus** - suurandmete töötlemiseks kohandatud arvutusrhitektuur

**Masinõpe** - arvutiteaduse haru, mis annab arvutitele õppimisvõime ilma neid otseselt selleks programmeerimata

**Neurovõrk** – algoritm, mis üritab imiteerida inimaju toimetusi kasutades selleks omavahel ühendatud neuroneid, mis asuvad kihiti ja saadavad üksteisele infot

**Suurandmed** (*Big Data*) - andmed, mis on väga suure mahuga, kiiresti muutuvad/tekkivad või piisavalt keerulised, et nende töötlemiseks on vaja eraldi tehnilist lahendust.

**Sügavõpe** (*Deep Learning*) – masinõppe valdkond, mis kasutab kunstlikke neurovõrke, mis asetsevad mitmetes kihtides.

**Tehisintellekt** (Kratt) - populaarteaduslik termin

- Kitsas tehisintellekt – algoritm, mis on mõeldud lahendama ühte konkreetset ülesannet
- Üldine tehisintellekt – algoritm, mis võimaldab lahendada ükskõik millist probleemi. Üldine tehisintellekt on antud dokumendi kirjutamise hetkel veel suund, kuhu teadus liigub ning praktilisi lahendusi on siiski vähe

**Treeningandmestik** - andmed masinõppe mudeli treenimiseks

**Valideerimisandmestik** - andmed masinõppe mudeli valideerimiseks

**Pilvandmetöötlus** ehk **pilvraalindus** (**Cloud computing**) - suuremahulise arvutusvõimsuse pakkumine kasutajatele sellisel kujul, et seda saab suhteliselt kiirelt ja mugavalt kasutusele võtta, vabastada ja arvestada. Pilvandmetöötluse teenusteks on näiteks Microsoft Azure, Amazon Web Services, Google Cloud ja teised.

**Testprojekt** (Proof-of-Concept) - projekt, mille käigus katsetatakse erinevaid lähenemisi ja lahendusi ning mille eesmärgiks on jõuda otsuseni, kas ja kuidas mingi lahendus töötab või ei tööta.

# Juhendmaterjalist

## Head andmeteaduse huvilised!

Eesti avaliku sektori digi-lahenduste ehitamine toetab visiooni nähtamatust riigist, kus teenused toimivad ilma lõppkasutajat koormamata. Kodaniku jaoks väljendub see teenuste saamises ilma selleks otseselt sekkumata. Sündmuspõhine ja taustal toimiv riik vajab visiooni elluviimiseks uusi tööriistu, kus teadmuspõhised otsused võetakse süsteemi poolt iseseisvalt vastu. Lihtsamate protsesside automatiseerimises oleme alustanud sündmuspõhiste teenuste pakkumisega. Siiski, kui soovime digitaliseerida oma riigi, tuleb automatiseerida ka keerukamate rutiinsete otsuste vastuvõtmine krati lahenduste abil.

Pärast põhjalikku tööd riigiasutustega, oleme jõudnud arusaamale, et eksisteerib puudulik teadmine, kuidas käivitada andmeteaduse projekte, seda eriti krati projektide osas. Käesoleva juhendmaterjali eesmärk ongi anda ülevaade sagedasematest probleemidest ning võimalikest lahendustest, mida võiks andmeteaduse projektide puhul arvestada.

Antud juhend on loodud koostöös ning sellesse saavad kõik huvipooled ka tulevikus panustada! Läbi selle tagame, et valdkonna arenedes saame omavahel uusi kogemusi ja teadmisi jagada. Dokumendile saab tagasisidet anda [siin](#).

Ott Velsberg  
Riigi andmete juht

# Soovitused – lühikokkuvõte

Soovitus	Põhimõtted
<ul style="list-style-type: none"> <li>• Krati testprojekti tulemuseks peaks suuresti olema teadmine (sh. raport, info algoritmide kohta, mis töötavad ja mis ei tööta, nimekiri andmete kohta, ülevaade projekti protsessist ja vahetulemitest, visiooni dokument) ja nõuded (näiteks andmetele) produktsioonikõlbulikule lahendusele. Samas tulemus peaks sisaldama ka nõ. käega katsutavat lahendust (näiteks algeline tööriist, uuenev raport või tootestatud mudel).</li> </ul>	<ul style="list-style-type: none"> <li>• Testprojekti eesmärgiks ei ole mõistlik võtta töötava süsteemi loomist. Töötav süsteem on mõistlik luua eraldi projektiga peale testprojekti lõppu, kasutades testprojekti saadud teadmisi.</li> <li>• Ärge ostke testprojekti töötavat infosüsteemi ega integratsioone teiste süsteemidega.</li> <li>• Ärge nõudke treening- ja valideerimisandmestiku hankimiseks ning kasutamiseks integratsioone eksisteerivate andmebaasidega, vaid kasutage andmete väljavõtet.</li> <li>• Ärge nõudke testprojekti tootmissüsteemide nõuetele vastavust. sh. juhendeid, installeerimist.</li> <li>• Käega katsutav lahendus on vajalik, et tulemiks olev lahendus ka praktikas kasutust leiaks ja oleks tugi organisatsioonis projektiga jätkamiseks.</li> <li>• Jätke omale võimalus muuta hüpoteesi ja hankige pigem töö mahtu, spetsialiste ja metoodikat kui ühe konkreetse hüpoteesi testimist.</li> </ul>
<ul style="list-style-type: none"> <li>• Hankimisel ei tasu keskenduda ühele kindlale lahenduskäigule, vaid tuua välja probleemid ning võimaldada projekti skooopi iteratiivselt muuta.</li> </ul>	<ul style="list-style-type: none"> <li>• Kasutage paindlikke hankemeetodeid</li> <li>• Tooge välja konkreetne probleem ja ootus projektile, mitte lahenduskäik</li> <li>• Kasutage agiilset arendust.</li> <li>• Ärge nõudke detailset projektiplaani vaid projekti läbiviimise metoodika kirjeldust, sealhulgas nimekirja milliseid masinõppe meetodeid/algoritme on plaanis katsetada ning millistega on olemas varasem kogemus. Samas ei tasu</li> </ul>

	<p>nõuda varasemat kogemust kõigi meetoditega.</p> <ul style="list-style-type: none"> <li>• Suurte andmemahutude puhul nõudke varasemat kogemust andmekogudega töötamisel.</li> </ul>
<ul style="list-style-type: none"> <li>• Andke ülevaade olemasolevatest andmetest</li> </ul>	<ul style="list-style-type: none"> <li>• Andke ülevaade, millistele andmetele on ligipääs olemas</li> <li>• Tooge välja maht, kvaliteedinäitajad, periood jmt.</li> <li>• Keerukamate andmekomplektide puhul võiks lähteülesande osa olla kirjeldava analüüsi läbiviimine.</li> <li>• Välise andmekogude puhul küsige kinnituskirjad</li> <li>• Väga hea oleks kui olulistest andmetabelitest oleks esitatud mõned read. Näiteks Exceli tabelina.</li> </ul>
<ul style="list-style-type: none"> <li>• Hankimisel arvestage turuolukorraga.</li> </ul>	<ul style="list-style-type: none"> <li>• Arvestades, et Eesti andmeteaduse sektor on võimekas, aga väike, siis tasub lubada välismaa spetsialistide kasutamist hangete teostamisel. See tähendab lubada teha tööd inglise keeles ning teenusepakkuja asukohas, vajadusel välismaal.</li> <li>• Kvalifitseerimistingimustes ei tasu seada kõrgeid nõudeid pakkuvate firmade käibe (käibenõue ei peaks ületama 100 000eur/a) või ajaloo (nõue ei peaks olema suurem kui 2 aastat) osas, sest valdkond on noor ja firmad väikesed.</li> <li>• Andmeteaduse tööd pakuvad nii väikesed spetsialiseerunud firmad kui ka suured ja pika ajaloo IT majad, kus on väikesed andmeteaduse tiimid. Kvalifitseerimise ja hindamiskriteeriumites tasub tähele panna, et need seataks firma andmeteaduse valdkonna kohta (mitte kogu firma käibe ja ajaloo põhised). Lisaks tuleks nõuda, et masinõppe algoritmid oleksid enda rakendatud ja/või edasi arendatud.</li> </ul>
<ul style="list-style-type: none"> <li>• Võimaldage modernsete tehnoloogiate ja infrastruktuuri (pilvandmetöötlus) kasutamist</li> </ul>	<ul style="list-style-type: none"> <li>• Ärge seadke (testprojektide puhul) nõudeid IT tööriistade, infrastruktuuri osas. Nii saavad</li> </ul>

	<p>pakkujad pakkuda teile kõige sobivamat, kvaliteetsemat ja soodsaimat lahendust.</p> <ul style="list-style-type: none"> <li>● Pilvandmetöötlus ei ole automaatselt välistatud andmeteaduse projektide puhul.</li> </ul>
<ul style="list-style-type: none"> <li>● Testprojektide puhul jätke omale võimalus projekt katkestada kui iteratiivne protsess enam uusi teadmisi ei lisa või kui on selge, et testi tulemus on negatiivne ja on selge, et kättesaadavate andmete pealt toimivat kratti ei ole võimalik arendada</li> </ul>	
<ul style="list-style-type: none"> <li>● Arvestage, et pakkumiste koostamine on pakkujate jaoks töö ning nõuded pakkumises ja riskid töö teostamisel vähendavad pakkujate hulka ja tõstavad hinda</li> </ul>	<ul style="list-style-type: none"> <li>● Küsige nii vähe dokumente kui vähegi võimalik.</li> <li>● Vältige ülesandeid, mille puhul ei ole võimalik teada, kuidas seda teha, aga peab esitama hinna ja tähtaja. (Näiteks kohustus integreerida kratt kasutusel oleva IT-süsteemiga, aga ilma, et süsteemil oleks API-t või muud teada olevat integratsiooni viisi.)</li> <li>● Andmeallikate otsimine, andmete hankimine ja nende analüüs võiks olla projekti skoobis, mitte osa pakkumise koostamisest.</li> </ul>
<ul style="list-style-type: none"> <li>● Arvestage, et krati projektid nõuavad tellijalt aktiivset panustamist kogu projekti vältel</li> </ul>	<ul style="list-style-type: none"> <li>● Arvestage, et kratt ei ole tavapärane IT arendus, kus piisab lähteülesande kirjeldamisest ja hilisemast tööde vastuvõtmisest. Tellija peab olema valmis panustama märkimisväärse osa oma ajast läbi terve projekti.</li> <li>● Tellija poolelt peab olema teada, kes on äripoolelt probleemi omanik ning kes on tehnilise poole pealt projektijuht.</li> </ul>

# Soovitused detailsemalt kirjeldatult

- 1. Krati testprojekti tulemuseks peaks suuresti olema teadmine (sh. raport, info algoritmide kohta, mis töötavad ja mis ei tööta, nimekiri andmete kohta, ülevaade projekti protsessist ja vahetulemitest, visiooni dokument). Samas tulemus peaks sisaldama ka nõ. käega katsutavat lahendust.**

Seda on oluline hanke teostamisel arvestada, et saada maksimaalselt soodsalt võimalikult väärtuslik ja õige tulemus.

Andmeteaduse testprojekti eesmärgiks on:

- a) Leida hüpoteese (näiteks, et andmete X, Y, Z järgi on võimalik ennustada, et ettevõtte A tegevuses on maksurisk.)
- b) Hüpoteese täiendada (näiteks lisades andmeallikaid) ning
- c) Kontrollida kas hüpoteesid peavad paika.
- d) Kaardistada edasised tegevused, sealhulgas kuidas masinõppe osa tuleks edasi arendada, mis andmeid tuleks kaaluda tulevikus, kas masinõppe võiks sobitada tulevikus arendatavasse süsteemi, kuidas andmete kvaliteeti peaks parendama, milliseid lisaandmeid tuleks koguda, kuidas tulevikus tagada, et masinõppe kvaliteet paraneks või säiliks.
- e) Anda ülevaade projekti protsessist ja vahetulemitest, millised projekti käigus katsetatud meetodid töötasid, millised mitte.
- f) Luua visioon, kuidas kratti tulevikus kasutada, milliseid probleeme saab (või ei saa) lahendada, milliseid eesmärke täita aitab.

Enne testprojekti on teada üks või mitu hüpoteesi kuid ei ole teada kas ta on tõene või mitte. Selle välja selgitamine on üks testprojektide eesmärgiks.

Testprojekti käigus ei ole mõtet osta valmis IT-süsteemi, sest:

- a) kui töötavat hüpoteesi ei leidu, siis ei ole mõtet teha IT-süsteemi, mitte-töötava hüpoteesi jaoks.
- b) Juhul kui nõuda hankes töötava hüpoteesi jaoks ka IT-süsteemi ehitamist sama ajakava ja eelarve raames, siis ei ole töö teostajatel paratamatult motivatsiooni leida töötavat hüpoteesi.
- c) Töötava IT-süsteemi loomine andmeteaduse testprojekti sees võib tähendada, et andmeteaduse osa jääb projekti töömahus alla 30% ja seega ei ole projekti fookus, hind ja töömaht õiges proportsioonis. Mõistlik oleks testprojektis kasutada ca 30% ajast andmete ettevalmistamisele ehk data engineer-i tööle ning 70% andmeteadusele.

Pigem on mõistlik saada tulemiks käega katsutav lahendus, mis aitaks valideerida, kas lahendus ka praktikas kasutust leiaks ja saada ülejäänud organisatsioonis pardale projektiga jätkamiseks.



## **2. Hankimisel ei tasu keskenduda ühele kindlale lahenduskäigule, vaid tuua välja probleemid ning võimaldada projekti skooopi iteratiivselt muuta.**

Näiteks oleks piisavalt täpne ja samas piisavalt üldine lahendatav probleem: „112 kõnede arvu ennustamine kasutades ajaloolisi andmeid, avalike ürituste andmebaasi, ilma andmeid ja muid võimalikke andmeallikaid.“

### **Miks on see oluline?**

See on oluline kuna määratledes eesmärgi liiga täpselt võib projekti tulemuseks olla küll projekti alguses kirjeldatud, aga eba-optimaalselt töötav kratt. Näiteks võib ülal kirjeldatud probleemi lahendamisel selguda, et parema 112 kõnede ennustamise saavutamiseks on mõistlik kasutada ka andmestut sotsiaaltoetuste maksmise kohta kuid kuna seda ei olnud hankes nii küsitud, siis seda kasutusele ei võeta.

#### **a. Kasutage paindlikke hankemeetodeid**

Läbirääkimistega hankemenetlus on parem kui lihtsalt pakkumiste küsimine:

- Pakkujate poolt annavad läbirääkimised võimaluse saada täpsemat infot projekti eesmärkide ja teostamise võimaluste kohta. Näiteks selle kohta, milliseid andmeid on võimalik kasutada ning mida peab tegema, et neid kasutada. (Võib-olla on vaja andmeid anonümiseerida, võib-olla on vaja andmeid osta jne.)
- Ostja poolt annavad läbirääkimised võimaluse tutvuda pakkuja tiimiga ning kuulata ära erinevaid ideid lahenduse loomisel. Kuna lahendused võivad olla vägagi erinevad, siis on hea need koos läbi käia, mitte ainult edastada neid kirjalikult. Näiteks on võimalikud projektid kus mingi etapi tegemiseks (pildituvastus, andmete korrastus) on mõistlik kasutada krattide asemel inimesi.
- Kindlasti oleks mõistlik peale hanke väljakuulutamist teha avalik hanke tutvustamise ja küsimuste-vastuste kohtumine.

#### **b. Tooge välja konkreetne probleem, mitte lahenduskäik**

Kui paljude IT-projektide puhul on üsna selgelt võimalik määratleda, et mis projekti käigus ehitatakse, siis andmeteaduse projektide ja eriti testprojektide puhul on eesmärgiks teadmiste hankimine. Peamiselt selle kohta, et kas mingi probleemi või teema lahendamisel saab andmeteadust ja kratti kasutada või mitte. Seega on probleemile lahenduskäigu leidmine ka üks projekti/hanke eesmärkidest ning seega ei ole mõistlik seda enne kirjeldada.

Küll aga on väga oluline defineerida projekti ja hanke alguses, et millist probleemi või valdkonda andmeteaduse töövahenditega lahendada tahetakse.

#### **c. Kasutage agiilset arendust**

Teadmiste hankimise projektis on agiilne arendusmeetod ainuvõimalik. Projekti käigus saadakse nimelt iga päev või iga nädal uusi teadmisi ja seetõttu on mõistlik teostatavaid töid ka tihedalt muuta. Näiteks saab projektis pidevalt teadmisi andmete kasutatavuse kohta ning selle kohta kas ja kuidas erinevad andmeteaduse algoritmid andmete peal töötavad. Vastavalt sellele on mõistlik jooksvalt projekti töid muuta. Kui näiteks otsustada juba projekti alguses, et milliste andmete peal, milliseid algoritme katsetatakse, siis on väga suur

tõenäosus, et projektis tehakse valed tööd ära, õigeid töid ei tehta. See võib olla teada juba enne nende tööde teostamist, kuid kuna projektiplaan oli näiteks hankelepingu osa, siis tuleb valed tööd ära teha.

d. Tooge välja, kuidas on hetkel olukord lahendatud  
Hetkeolukorra kirjeldamine on oluline andmeteaduse projektide puhul. See annab võimaluse hinnata, kuidas peaks tulevikus süsteem toimima ja mida saaks olemasolevast olukorrast probleemi lahendamiseks kasutada.

### 3. Andke ülevaade olemasolevatest andmetest

#### Miks see on oluline:

Andmeteaduse projekti puhul on sisendandmed ülimalt olulised ja nende olemasolust ning kvaliteedist sõltub projekti edukus täielikult. Piltlikult öeldes võib öelda, et andmeteaduse projekti käigus loodav „kratt“ on nagu väike imik, kes alguses ei tea maailmast midagi ja kellele me tahame õpetada selgeks oskuse tunda ära pildil olev koer. Õpetamise meetodiks on see, et me näitame talle erinevaid pilte ja ütleme, millisel pildil on koer ning millisel mitte. Erinevalt imikust on tänapäeva masinõppe süsteemid võimelised õppima väga palju kiiremini ning käsitlema sisendina väga suuri ning mitme-dimensioonilisi andmestuid, kuid kui meie õppematerjalides on osadel koerapiltidel märkimata, et seal on koer või on hobusepilte, mille kohta on märgitud, et see on koer, siis õpib meie masin valesti.

Seega on andmeteaduse projekti puhul andmestu osas hea järgida järgmist:

- Andmeid, mille pealt õppida, peab olema palju. Pigem sadu tuhandeid ja miljoneid ridu. Samas lihtsamate ülesannete lahendamiseks piisab oluliselt väiksemas mahus andmetest.
- Kui andmed on pärit erinevatest andmebaasidest, siis peavad nad olema korrektselt ja üheselt seotud, et oleks olemas üks andme „pilt“, mille kohta saab väita üheselt, et kas see on või ei ole meie otsitav nähtus.
- Peab olema palju treeningandmeid, mille kohta on teada, kas sellisel juhul meie otsitav asi oli või ei olnud. Näiteks ei saa me õpetada kratti tundma pildil ära koera kui meie treeningpiltide puhul ei ole märgitud, et kas pildil on koer või ei ole koera. Võimalik on ka projekti raames märgendamine, aga ka siis on hea, et tellija ise märgendab domeeniteadlikkuse tõttu.
- Hea on krati treenimiseks kasutada võimsaid GPU-dega servereid, mida on mõistlik rentida avalikest pilvandmetöötlus keskkondadest. Seega võiks olla võimalik treeningandmete viimiseks ja hoidmiseks nendes keskkondades. Sealhulgas arvestades andmeturbe nõuetega.
- Anonümiseerimine ja muu andmete modifitseerimine on sageli võimalik kuid läbi anonümiseerimise võivad andmed kaotada palju oma väärtusest. Seda võib võrrelda olukorraga kus õpetame masinat piltidelt ära tundma koera, aga teeme pildid piisavalt uduseks, et ei ole võimalik ära tunda, milline konkreetne koer (või hobune) pildil on.
- Sageli on mõistlik kasutada sisendandmetena ka teiste asutuste, ettevõtete või muud avalikku andmestut. Hea oleks eelnevalt teada, millised on võimalused ja tasud nende andmete kätte saamisel.

#### **4. Hankimisel arvestage turuolukorraga.**

##### **Miks see on oluline:**

Hankija eesmärgiks võiks olla kvaliteetse ja soodsaima pakkumise leidmine. Selle saavutamiseks on oluline, et hanke tingimustega ei välistataks tiime, kes oleks võimelised töö kvaliteetselt teostama, kuid kellel näiteks ei ole pikka referentside nimekirja, käivet vms. Eesti andmeteaduse valdkond on väga väike näiteks võrreldes tarkvara arenduse mahuga ning seega ei ole mõistlik rakendada samu nõudeid nagu IT-projektide puhul üldiselt.

- Valdav osa andmeteaduse tööst tehakse Eestis täna sisemise arendusena, sh. startuppides, finants- ja laenuettevõtetes ning ülikoolides.
- Andmeteaduse tööd teenusena pakub aastal 2019 Eestis suurusjärgus 10 firmat, kus töötavate Eesti andmeteadlaste hulk jääb enamasti alla 10.
- Esimesi andmeteaduse projekte teenusena hakati Eestis pakkuma aastatel 2015-2016.
- Nimekiri andmeteaduse ettevõtetest on leitav [www.kratid.ee](http://www.kratid.ee)

#### **5. Võimaldage modernsete tehnoloogiate ja infrastruktuuri (pilvandmetöötlus) kasutamist**

Andmeteaduse süsteemide puhul võivad nõuded IT-riistvarale kõikuda väga suuresti projekti käigus. Nii öelda mudeli õpetamise puhul võib olla vaja väga võimsaid graafilisi protsessoreid kuid seda ainult mõneks tunniks. Sellisel juhul on pilvandmetöötluse teenuste kasutamine palju mõistlikum, sest seal on võimalik vajalikku võimsust rentida ning seda ei pea välja ostma.

Eriti testprojektide, aga ka tootmissüsteemide puhul, läheb süsteemi jaoks vaja erinevaid töövahendeid ja teenuseid, mida on mõistlik kasutada teenusena pilvandmetöötluse keskkonnast ning mitte osta nende jaoks litsentse.

Seega kokkuvõttes on mõistlik mitte seada (eriti testprojektide puhul) nõudeid IT tööriistade, infrastruktuuri osas. Nii saavad pakkujad pakkuda teile kõige sobivamat, kvaliteetsemat ja soodsaimat lahendust.

#### **6. Testprojektide puhul jätke omale võimalus projekt katkestada.**

##### **Miks see on oluline:**

Testprojektides testitakse erinevaid võimalusi mingi töö efektiivsemaks teostamiseks. Näiteks, kas on võimalik ehitada kratt, mille abil parandada Ehisregistri andmestut kasutades eri aegadel tehtud aerofotosid. Selleks seatakse projekti käigus erinevaid hüpoteese ja proovitakse saada olevate andmete puhul erinevaid algoritme ning proovitakse leida selline lahendus, mis töötab.

Väga sageli võib aga juba peale esimesi teste selguda, et kasutada olevate andmete põhjal on täiesti lootusetu leida hüpoteesi, mis töötab. Sellisel juhul ei ole mõtet projekti jätkata,

aega ning raha kulutada ja projekt tasub katkestada. Samuti ei ole mõistlik selliste mitte-töötavate lahenduste tootmissüsteemiks arendamine.

## **7. Arvestage, et pakkumiste koostamine on pakkujate jaoks töö ning nõuded pakkumises ja riskid töö teostamisel vähendavad pakkujate hulka ja tõstavad hinda**

### **Miks see on oluline:**

Kuigi ostja seisukohalt on pakkumiseks tegemise töö tasuta, siis pakkujad on äriettevõtted ning pakkumiste koostamisel arvestatakse võidu tõenäosust, pakkumise tegemise mahtu, projekti riske ning otsustatakse selle põhjal, kas ja milline pakkumine esitada. Kuna ostja huvides on saada palju ja häid pakkumisi, mille vahel valida, siis on mõistlik vältida pakkumise koostamist takistavaid asjaolusid.

Pakkumise koostamist raskendavad järgmised nõuded:

- Ebamäärane või teostamatu skoop.
- Töö teostaja jaoks kontrollimatu suurusega tööd ja riskid. Näiteks nõue integreerida lahendus olemasoleva IT-süsteemiga, mille kohta ei ole mingit täpsemat infot, ega hinnangut kas ja kuidas integratsioon üldse võimalik on.
- Ebaselgus andmete olemasolu, kasutatavuse ja kvaliteedi osas
- Mida rohkem dokumente (tõendeid, CV-sid, referentse jms.) peab esitama, seda rohkem tööd pakkumise koostamine tähendab.
- Nõuded projekti- ja tehnilise plaani osas, mis ei ole antud projekti puhul vajalikud või on ebasobivad. Näiteks nõue kasutada agiilset arendust ja esitada detailne projektiplaan koos spetsialistide tööplaaniga nädalate lõikes.

## **8. Arvestage, et krati projektid nõuavad tellijalt aktiivset panustamist kogu projekti vältel**

### **Miks see on oluline:**

- Arvestage, et kratt ei ole tavapärane IT arendus, kus piisab lähteülesande kirjeldamisest ja hilisemast tööde vastuvõtmisest.
- Tellija peab olema valmis panustama märkimisväärse osa oma ajast ja muudest ressurssidest terve projekti vältel, et projekt õnnestuks.

# Näide andmeteaduse testprojekti hankedokumendi „Tehniline kirjeldus“ ülesehituse kohta

## 1. Hanke ese

/Mõne lausega kirjeldada projekti eesmärki/

## 2. Taustainfo

/Kirjeldada praegust olukorda. Optimeerimise võimalusi./

## 3. Eesmärk

/Kirjeldada millist tööprotsessi või protsesse võiks testprojekti optimeerida. Millised on prioriteedid. Millist säästu, suuremat kvaliteeti või muud soovitakse saavutada./

## 4. Töö meetod

/Töö meetod (ja tiim) võiks jääda Pakkujaga valida ja kirjeldada. Töö meetod võiks olla üks hindamiskriteeriumitest. Siin võiks kirjeldada seda kuidas pakkuja peab pakkumises oma töömeetodit kirjeldama, et hiljem oleks võimalik pakkumisi võrrelda. Samuti tuua välja Ostja poolsed ressursid, spetsialistid ja tegevused. Nõuded töö teostamisele (nt. asukoht, keel, andmeturve, masinõppe meetodid/algoritmid, litsentsid)./

## 5. Töö sisendinfo ehk andmed

/Kirjeldus saada olevatest andmetest/

## 6. Töö tulemid

/Töö tulemituks võiksid olla:

- Raport, mis sisaldab järgmist:
  - Milliseid hüpoteese testiti ja milliseid andmeid selleks kasutati
  - Kui hästi kratt töötas erinevate hüpoteeside korral
  - Milliseid andmeid võiks veel kasutada ja kuidas neid peaks töötlemata.
  - Tulemus ja ettepanekud edasiseks
  - Juhul kui tulemus oli positiivne, siis ka ettepanekud süsteemi arhitektuuri ja realiseerimise kohta ning välja tooma mittefunktsionaalsed nõuded
- Dokumenteeritud ja kommenteeritud masinõppe ja andmete töötamise tarkvara./

## 7. Tähtajad ja tingimused

/Ootused projekti teostamise aja kohta. Projekti etapid ja võimalus projekt katkestada kui pooled leiavad, et seatud eesmärgid ei ole võimalik täita./