

Visioon

Aero-, satelliit- ja lidaripiltide pealt huvipakkuva tuvastamine on teema millega vähemal või suuremal määral tegelevad kõik maailma riigid. Eesti näitel ka mitmed riigiasutused korraga, masinõppe seisukohast lahendatakse samu ülesandeid paralleelselt. Siin on meeletu raisatud potentsiaal, kui kõik (riigid ja asutused) omaette nokitsevad neid lahendusi. Tulevikus võiks olla "geograafiline baasmudel", mis on treenitud riikide üleselt. Näiteks kogu Euroopal võiks olla mõned baasmudelid, mis on treenitud ühise andmestiku pealt ühisel arvutusressursil. Selline lähenemine annaks selgelt kõige parema tulemuse. Küll aga selle organisatoorne pool on keerukam, mistõttu valdkonna algusaegadel kipub töö silodesse. Sellist ühist Euroopa baasmudeli visiooni silmas pidades ning selle poole nügides oleme ka käesoleva projekti visiooni loonud. Hetkel võib öelda, et [Prantsuse Maaamet](#) on selle visiooni eestvedajad ning plaanime ka ise siinkohal töö sobitada nende formaati. Eestil on seega hea võimalus "Euroopa baasmudelite" initsiatiivi kaas-vedada, minnes nõnda geoinformaatika ajalukku.

Näeme, et ülesande 5 klassi tasub jaotada nõnda:

Mudel 1 - Üldine segmentatsioonimudel mis ühildab endas teed, ehitised ja vooluveekogud

Mudel 2 - Hoonete korrused

Mudel 3 - Hoonete sissepääsud

Selline jaotamise põhjendus:

1. Sisend, väljund, tunnused. Näiteks selleks, et pildilt tuvastada teid või ehitisi, nii inimene kui tehisintellekt vaatavad "sama asja" (pilti), "sama moodi" (visuaalsete sarnase abstraktsioonitasemega tunnuste alusel) ja väljundiks on "sama asi" (segmentatsioon). Küll aga näiteks hoonete korruste tuvastamine on fundamentaalselt erinev asi, kus pigem fookus kõrgusandmetel ning väljund pole segmentatsioon, vaid arv. Samuti sissepääsu loogikat võib väga erinevalt vaadata, aga sellest juba all konkreetse peatüki juures.
2. Kui võimalik ja põhjendatud, siis masinõppes kipub olema rohkem asju koos vaadata parem:
 - a. Jagatud tunnused - madalama taseme tunnused nagu ääred, nurgad, tekstuurid jne on kasulikud üle kõigi klasside
 - b. Kontekstuaalsed seosed - mudel on targem, kui ta saab lisaks õppida seoseid erinevate klasside vahel, mitte lihtsalt "silma klappidega" otsida ühte konkreetset klassi ja mitte-klassi.

- c. Treeningandmete haldamine, mudeli treenimine/haldamine/ juurutamine on märgatavalt lihtsam. Seda nii juhtudel kui treeningandmeid tuleb juurde (see on kindlalt MaRu puhul nii), kui ka kui uusi klasse lisandub (tõenäoliselt leiab samuti aset MaRu puhul).
- d. Arvutuslik efektiivsus - ühe mitme klassiga, veidi suurema mudeli treenimine on palju ressursi-efektiivsem kui mitme binaarse mudeli treenimine.
- e. Tõenäoliselt lähevad üldisesse segmentatsioonimudelisse ka kõik alamklassid kokku, kuigi siin võib juhtuda et teatud juhtudel on hierarhiline lähenemine (alam-klassifikaator) põhjendatud.
- f. Masinõppe üldine trend on ikkagi *end-to-end*. Näideteks võib tuua mudelid nagu YOLO, Clip, sisuliselt kogu multimodaalsete fundamentaalmudelite teema (kogu suurte keele- ja pildimudelite valdkond). Ning ka baasmudelid mida siin kasutada plaanime võtavad sellise laiapõhjalise lähenemise ja eelduse.
- g. Soomlaste ATMU projekt on tähelepanuväärne ja panustas vastupidisele lähenemisele. Ühtpidi seda ei saa eirata, aga teist pidi tunneme, et suunis minna mitu-klassi-ühes-mudelis teed on siin kogu eelkirjeldatu põhjal liialt tugev. Meie soovitatud lähenemine on natuke keerukam, et alguses käima saada, näiteks klasside ebaühtlase jaotuse probleemiga tuleb rohkem tegeleda, aga ta on suures plaanis tõenäoliselt edukam.

Metoodika

Metoodika peatükis on järgmised teemad lahti mõtestatud:

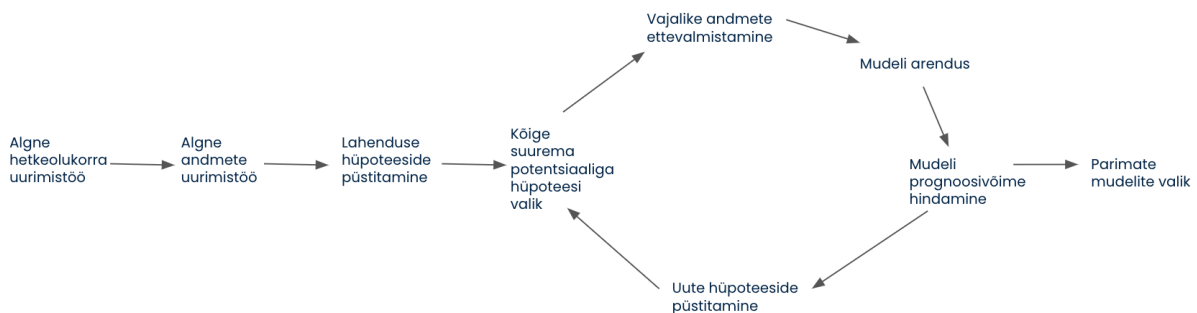
1. **Andmeteaduse arendusmetoodika** - kirjeldame meie üldist andmeteaduse ja masinõppe arendusmetoodikat. Tegemist on meie poolt välja töötatud ja rafineeritud metoodikaga, mis on end palju aastaid hästi tõestanud. Metoodika on inspireeritud teaduslikust lähenemisest ja agiilsest tarkvaraarendusest.
2. **Andmete eeltöötlus** - käsitleme mõningaid olulisi nüansse andmete ja nende eeltötluse teemal.
3. **Testandmestike ja testprotseduuride defineerimine** - rõhutame, et iga masinõppe projekti puhul on üks olulisemaid asju lähtuda kasutajate ja kasutusloogika poolelt, mis siin peegeldub testprotseduuride näol. Viimased on omakorda aluseks kogu masinõppe arendusele.
4. **Ühine segmentatsioonimudel: hooned/teed/vooluveekogud** - selgitame metoodikat ja visiooni antud mudeli loomiseks.
5. **Hoone korruste mudel** - selgitame metoodikat ja visiooni antud mudeli loomiseks.

6. **Hoone sissepääsu mudel** - selgitame metoodikat ja visiooni antud mudeli loomiseks.

Siinkohal mainime ka, et käesolev on esmane visioon ning töö käigus lähevad ideed täpsemaks ja ülesande reaalsusega kooskõlla.

Andmeteaduse arendusmetoodika

Iteratiivne ja agiilne andmeteaduse/masinõppe arendusmetoodika on kirjeldatud järgneval joonisel.



Joonisel esindatud komponendid:

- **Algne hetkeolukorra uurimistöö** - uuritakse probleemi olemust ja seda, kuidas seda praegu lahendatakse või on proovitud lahendada. Tuvastatakse praeguste lähenemiste pudelikaelad ja selgitatakse välja, kust on võimalik saavutada suurimaid võite.
- **Algne algandmete uurimistöö** - kaardistatakse, millised andmed on olemas, milliseid andmeid on võimalik juurde hankida/koguda, mis on andmete olukord.
- **Lahenduste hüpoteeside püstitamine** - toimub koostöös tellija ja pakkujaga. Tellija ja pakkuja pakuvad mõlemad välja ideid, mis võiksid olukorra lahendada. Kui tellija pakub lahenduse, hindab pakkuja, kas tegu on lähenemisega, mis kuulub masinõppe alla või on see lihtsamini tehtav. Kui hankija pakub lahendus, hindab tellija, kas selline lähenemine päriselt lahendaks probleemi ärilisest vaatepunktist ning kas selleks eksisteerib andmeid.
- **Kõige suurema potentsiaaliga hüpoteeside valik** - Ülesannet saab üldiselt lahendada rohkem kui ühte moodi ning piiratud ressursside korral on mõistlik neid üksikhaaval testida. Arvestatakse ärilist mõju, tehnilist teostatavust, arendamiseks kuluvat aega, andmete olemasolu ning üheskoos valitakse kõige sobilikum lahenduse hüpotees, mida minna masinõppega lahendada
- **Vajalike andmete ettevalmistamine** - andmeinseneri ja andmeteadlase ühistöö, kus valmistatakse andmed ette, et nende pealt oleks võimalik õpetada masinõppe

modelit. Võib sisaldada tööd, kus on tarvis andmete kvaliteeti parandada, neid juurde koguda või andmeallikaid lisada.

- **Mudeli arendus** - teostab andmeteadlane, kasutades andmeid ning üritades ehitada modelit, mis optimeerib mõõdikut, mis on äripoollega varem kokku lepitud.
- **Mudeli valideerimine ja hindamine** - andmeteadlane selgitab ning visualiseerib tulemusi. Äripool hindab, kui hästi model on püstitatud ülesannet lahendanud.
- **Uute hüpoteeside püstitamine** - isegi, kui ei saavutatud arendustsüklis paremat tulemust, õpitakse paremini õppima andmeid ning probleemi olemust. Olles varustatud uute teadmiste, võib tellija ja pakkuja koostöös välja tulla uute lahenduse hüpoteesidega.
- **Parimate mudelite valik** - hinnatakse mudelite sooritust mõõdikute järgi, mudeli ennustuskiiirust, vajalike andmete kogust, vajalike integratsioonide hulka ja keerukust. Loomulikult võib olla ka teisi äriole olulisi kriteeriumeid, kuid alati ei pruugi olla optimaalseim valida kõige parema sooritusega model, sest selle ülevaheldamine võib kujuneda kulukaks.

Iteratiivsus on kujutatud arendusprotsessi graafi tsüklilises osas. Üldisemas plaanis itereeritakse hüpoteeside prioriteedijärjekorra üle. Andmeteadlane ja projektijuht koostöös optimeerivad, et iga iteratsioon tegeletakse kõige optimaalsemate arendustga. Agiilsus väljendub selles, et iga tsükli läbimisel kogunenud teadmised muudavad järgmiste tsüklite kavandamist.

Andmete eeltöötlus

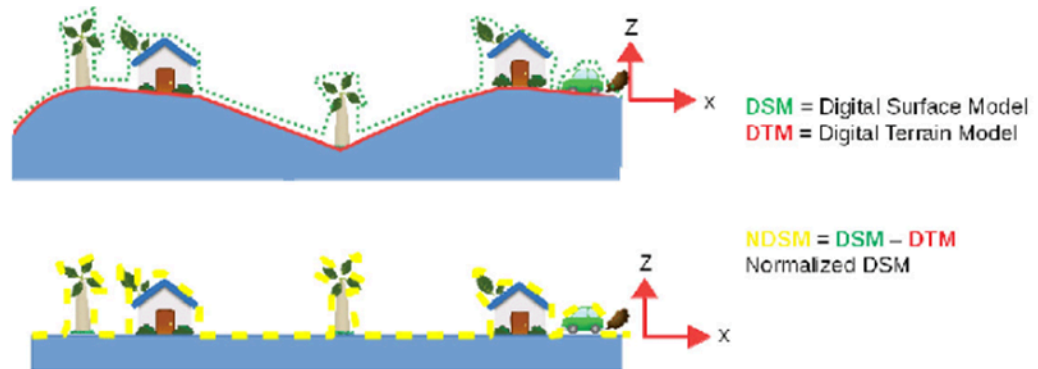
MaRu ortofotod on hea alguspunkt edasiseks analüüsiks, sest aerofotod on juba töödeldud sobivale kujule: ortorektifitseeritud, georeferentseeritud, jne.

Kindlasti plaanime kasutusele võtta ka Lidari kõrgusandmed:

- a. ALS kõrguspunktid
- b. Maapinna kõrgusmodel (DTM) lahutusega 1m, GeoTIFF
- c. Maakatte kõrgusmodel (nDSM=DSM-DTM) lahutusega 1 m, GeoTIFF

- d. Taimkatte kõrgusmudel (CHM) lahutusega 4-10 m, GeoTIFF

$$nDSM = DSM - DTM$$



Väga oluline aspekt on andmete sobivale kujule struktureerimine, ehk kui suurte alam-piltide peal treenida. Tegemist on väga arvutusmahuka ülesandega ning alam-piltide suurus määrab nii jõudluse kui soorituse tulemust. Sestap lahendatakse küsimus arvestades mõlemat nüansi ning eksperimentaalselt.

Ortofotode ja ALS integratsioon

ALS andmete lisamine toimub tõenäoliselt just nDSM profiili abil. See võiks olla väga kasulik segmenteerimise jaoks, sest hoonetel, teedel ja vooluveekogudel on kõigil selge kõrgusprofiil. Hoonete korruste määramise jaoks panustame põhiliselt kõrgusandmetele. Sissepääsu tuvastamisel on kõrgusandmete tõenäoliselt samuti olemas, kuna hoone ja ümbritsevate teede kuju joonistab teatava intuiitse avause, ehk sissepääsu.

Süvaõppe seisukohast on selline lisamine samuti igati hästi võimaldatud, sest erinevad konvolutsioonilised närvivõrgud (ja muud pilditötluse meetodid) on võimalik peaaegu triviaalselt viia 3'lt sisendkanalilt (RGB) 4'le (RGB+nDSM). Käesolevas projektis saame selle osa tegemiseks kasutada meetodeid, mis praegu PRIA põllupiiride tuvastamise projektis teinud oleme.

GIS tarkvara

Erinevates andmete ettevalmistamise, eel- ja järeltötluse etappides võib olla väga kasulik kasutada GIS tarkvara. Sarnastes projektides on seni kõige edukamaks valikuks osutunud QGIS. Lisaks tuleb käsitleda GIS tarkvaraga liidistamise küsimusi, et sobitada lõplikesse töövoogudesse tulevikus.

Testandmestike ja testprotseduuride defineerimine

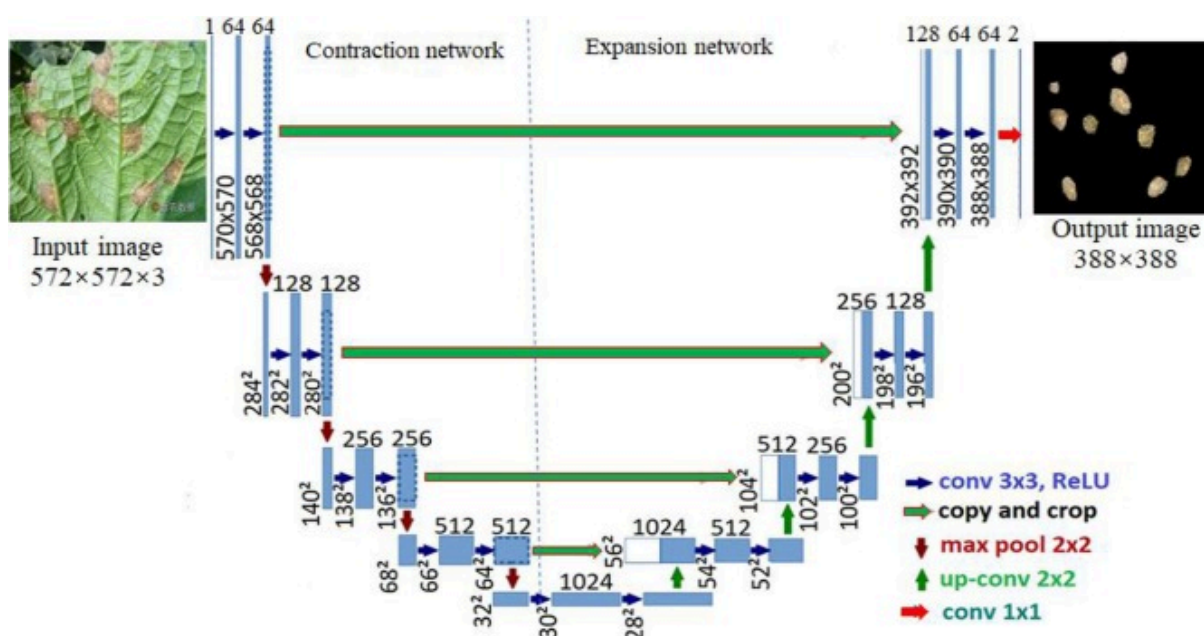
Käesoleva projekti fookus pole mitte ainult täpsel pildituvastusel, vaid konkreetse ajavõidu ja efektiivsuse saavutamisel erinevates andmehõive protsessides. Sellest lähtuvalt koostatakse hanke töö alguses konkreetne testandmestik ning sellele tuginevad

testimise protseduurid. Need peavad näitlikustama reaalsel tänast tööprotsessi ning ajavõitu. Kuna tegemist on eksperimentaalse uue töövooga ning samuti, et katsetuste iteratsiooni kiirust maksimeerida, siis seadistame lõpliku töövooga sisuliselt ekvivalentse testimise töövoogu. See tähendab, et me tõenäoliselt ei asu liidestuma päris töövoogu, vaid teeme (näiteks notebooki, mõne mugavalt modifitseeritava märgendustööriista või isegi exceli põhiselt) testimise töövoogu.

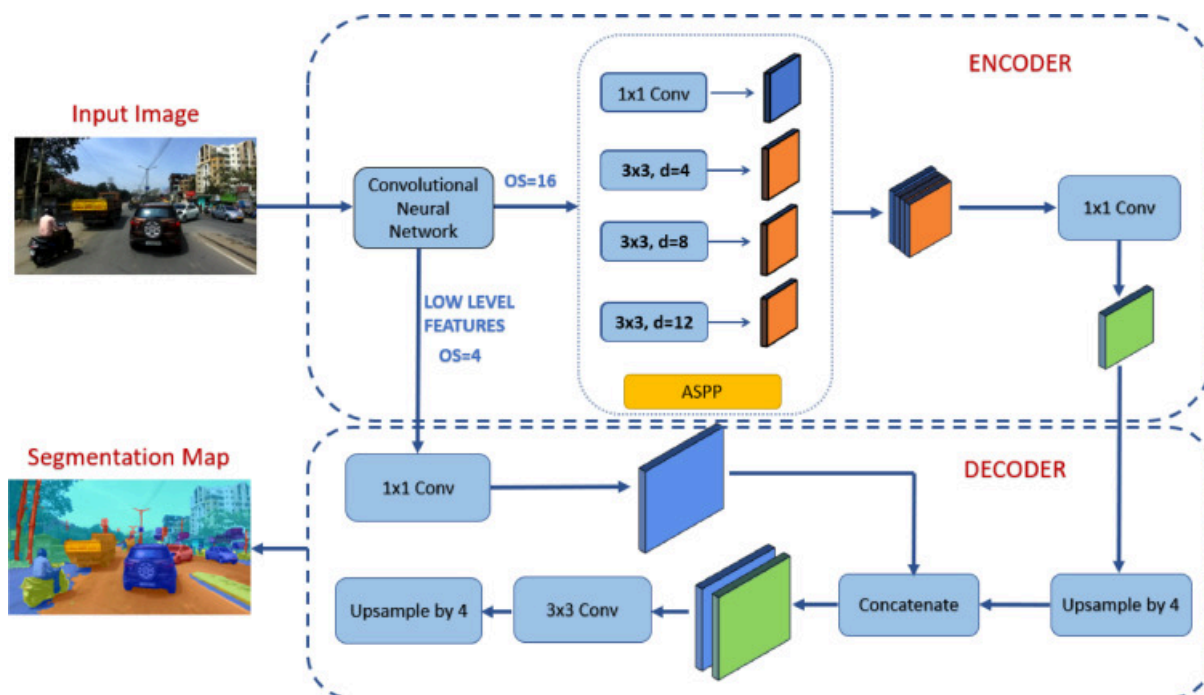
Ühine segmentatsioonimudel: hooned/teed/vooluveekogud

Sobivad arhitektuurid

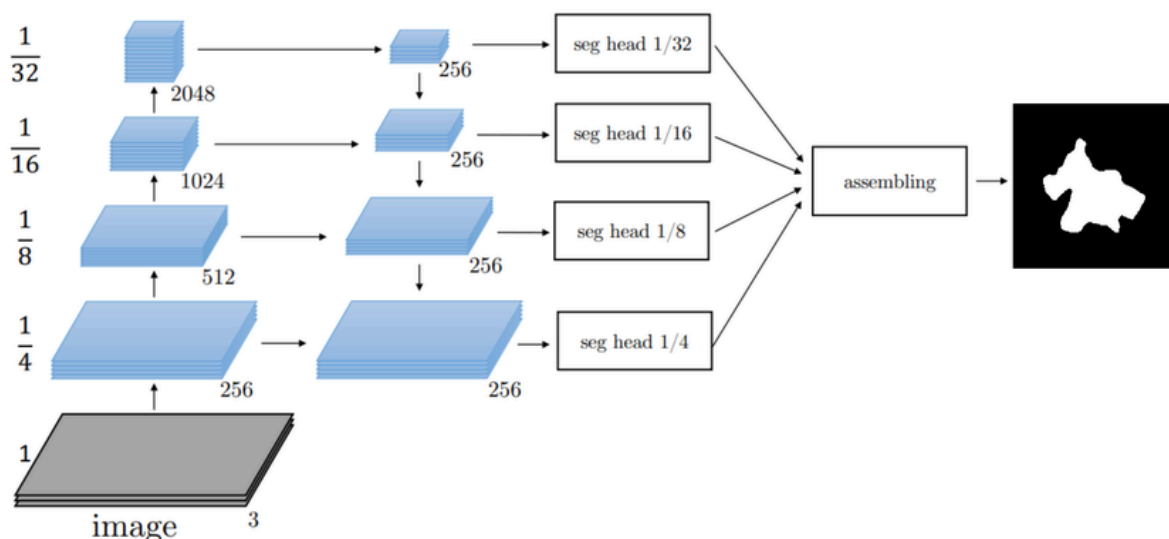
U-Net: Arhitektuur, mida iseloomustab ülekodeerimisühendusega (*skip connections*) kodeerija-dekodeerija struktuur. Tegemist on semantilise segmenteerimise standardiga, eriti muuhulgas kaugseire rakendustes. Kodeerija rada vähendab järk-järgult sisendpildi mõõtmeid, jäädvustades kõrgetasemelise kontekstiteabe. Dekodeerija rada suurendab seejärel neid tunnuseid, taastades järk-järgult ruumilise eraldusvõime. Ülekodeerimisühendused ühendavad otseselt kodeerijast pärit tunnuskaardid (*feature maps*) vastavate kihtidega dekodeerijas, võimaldades mudelil säilitada peeneid ruumilisi detaile, mis võivad allavalimise (*downsampling*) käigus kaduma minna. See muudab U-Net'i eriti tõhusaks erinevate suuruste ja kujudega objektide segmenteerimisel, mis on meid ootamas käesolevas projektis.



DeepLabv3+: Arhitektuur põhineb DeepLabi perekonna mudelitel ning kaasab tühimikkonvolutsioone (*atrous convolutions*) (tuntud ka kui laiendatud (*dilated*) konvolutsioonid), et tajuda mitme erineva skaala (näiteks puu ja mets) kontekstuaalset teavet ilma tunnускаardi eraldusvõimet vähendamata. Samuti rakendab see Tühimikkonvolutsioonidega Ruumilise Püramiidi Ühendamise (*Atrous Spatial Pyramid Pooling*, ASPP) moodulit, mis kasutab erineva laienduse (*dilation*) määraga tühimikkonvolutsioone, et eraldada tunnuseid mitmel skaalal. "+" versioon sisaldab dekodeerimoodulit, mis täpsustab segmenteerimise piire. DeepLabv3+ on näidanud tugevat sooritust erinevatel segmenteerimise näidisülesannetel (*benchmark*'idel) ning sobib hästi ortofotodel erineva suurused objektide käsitlemiseks.



Feature Pyramid Networks (FPN): FPN-e iseloomustab samuti mitme-skaalaliste tunnuste utiliseerimine. tunnuste pildil toimuvast. Nad kombineerivad madala resolutsiooniga, semantiliselt tugevaid tunnuseid närvivõrgu sügavamatest kihtidest kõrge resolutsiooniga, ruumiliselt täpsete tunnustega madalamatest kihtidest. See võimaldab mudelil tõhusalt segmenteerida objekte erinevatel skaaladel, kuna tal on juurdepääs nii peenetele detailidele kui ka kõrgetasemelisele kontekstuaalsele teabele mitmel resolutsioonil. FPN-id on eriti kasulikud väikeste objektide segmenteerimise parandamiseks, mis võib olla oluline üksikute majade või kitsaste teede tuvastamisel.



Erinevaid arhitektuure on veel palju, mis kas ehituvad eelnevalt kirjeldatule või on täiesti eriilmelised (näiteks SAM - Segment Anything Model). Hea ülevaade on näiteks [siin](#). Lõplik valitud arhitektuur ilmneb eksperimentaalse töö käigus lähtuvalt sellest, mis annab MaRu ülesande tarbeks parima tulemuse.

Lisaks, üks konkreetne kaardi segmenteerimise implementatsioon, mida kindlasti hea siin projektis aluseks võtta on Prantslaste loodud [FLAIR-1](#) - olemas nii mudelid kui alusandmed. Muuhulgas suudab arvesse võtta ka kõrgusandmeid seega igati hea kandidaat eksperimentide tegemiseks.

Mudeli treenimise nüansid

Kaofunktsioonid

Mitme klassiga semantilise segmenteerimise jaoks kasutatakse tavaliselt mitut kaofunktsiooni. Kategooriline ristentroopia (*Categorical Cross-Entropy*) on standardvalik mitme klassi klassifitseerimiseks, mõõtes erinevust klassidele (hooned, teed, vesi, taust) ennustatud tõenäosusjaotuse ja iga piksli tegeliku klassi märgenduse vahel. Dice'i kadu (*Dice Loss*) on eriti kasulik klasside ebaühtlase jaotuse (*class imbalance*) korral, kuna see mõõdab otseselt ennustatud ja tegelike segmenteerimismaskide kattuvust iga klassi jaoks, keskendudes ühisosa maksimeerimisele kogu ulatuses. Fokaalkadu (*Focal Loss*) on teine tõhus kaofunktsioon klasside ebaühtluse lahendamiseks, vähendades lihtsalt klassifitseeritavate pikslite mõju ja keskendudes rohkem raskesti klassifitseeritavatele näidetele, mis kuuluvad sageli vähemuse klassidesse. Praktikas võib kaofunktsioonide kombinatsioon, näiteks kategoorilise ristentroopia ja Dice'i kao kaalutud summa, viia

mõnikord parema tulemuseni, kasutades ära mõlema tugevusi. Sobiv kaofunktsioon valitakse taaskord eksperimentaalselt lähtudes ülesandest.

Klasside ebaühtlane jaotus

Klasside ebaühtlane jaotus, kus ühe või mitme klassi pikslite arv on teistest oluliselt väiksem, võib pärssida semantilise segmenteerimise mudeli kvaliteeti. Näiteks võib uuritaval alal voolavaid veekogusid olla vähem kui hoonestust või teid. Siinkohal kasutame mitmeid strateegiaid.

1. Kaalutud kaofunktsioonid omistavad vähemusklassidele suuremad kaalud, sundides mudelit pöörama rohkem tähelepanu nende klasside õigele klassifitseerimisele.
2. Vähemusklasside ülevahtimine (*upsampling*) hõlmab vähemusklasside treeningnäidiste arvu suurendamist, kas olemasolevate näidiste dubleerimise või andmete täiendamise tehnikate, nagu pööramise ja peegeldamise abil.
3. Enamusklasside alavalimine (*undersampling*) vähendab enamusklasside treeningnäidiste arvu, et luua tasakaalustatum klassijaotus, kuigi see võib põhjustada mõningate potentsiaalselt kasulike andmete kaotsiminekut.
4. Nagu varem mainitud, on fokaalkadu samuti tõhus tehnika klasside tasakaalutuse käsitlemiseks, keskendudes treenimisele raskesti klassifitseeritavatel näidetel. Klasside tasakaalutuse lahendamine on oluline tagamaks, et mudel õpib täpselt segmenteerima kõiki sihtklasse, sealhulgas neid, mis esinevad andmestikus harvemini.

Andmete augmenteerimine

Aitab parandada mudeli üldistusvõimet ja olemasolevast treeningandmete kogusest maksimumi välja võtta. Olemasolevaid pilte transformeeritakse, luues nendest uusi variatsioone, mis säilitavad originaali semantilise sisu, kuid esitavad seda erineval kujul. Ortofotode puhul võib selline andmete rikastamine hõlmata piltide pööramist erinevate nurkade all, peegeldamist, lõikamist, värvi muutmist, heleduse ja kontrasti reguleerimist ning müra lisamist. Näiteks võib sama hoonetega piirkonna pilt esineda treeningandmetes erinevate ilmastikutingimuste, valgustustasemete ja perspektiividega, aidates mudelil õppida tuvastama hooneid sõltumata nende konkreetsest esitusviisist. Eriti kasulik on see vähemusklasside puhul, suurendades nende klasside näidiste arvu ja parandades mudeli võimet neid objekte erinevates tingimustes ära tunda.

Mõõdikud

Põhiliseks mõõdikuks on klassipõhine F1-skoor. Tegemist on hea standardse mõõdikuga masinõppes üldiselt ning ka esile toodud hankija nõuetes, tõenäoliselt lähtudes hankija senistest kogemustes sobiliku mõõdiku valikul. F1-skoori definitsioon on ka

hankijapoolsetes dokumentides esitatud, seega siinkohal me ei asu seda lahti defineerima.

Treenitud mudeli hindamise strateegia

Standardne hindamise strateegia hõlmab märgendatud andmete jagamist kolme eraldi komplekti: treeningkomplekt, valideerimiskomplekt ja testikomplekt. Treeningkomplekti kasutatakse süvõppe mudeli treenimiseks. Valideerimiskomplekti kasutatakse treenimisprotsessi ajal mudeli hüperparameetrite häälestamiseks ja jõudluse jälgimiseks nägemata andmetel, aidates ka vältida ületreenimist (*overfitting*). Lõpuks kasutatakse testikomplekti, mis on treenimis- ja valideerimisandmetest täiesti sõltumatu, mudeli üldistusvõime lõplikuks, erapooletuks hindamiseks täiesti nägemata andmetel. Mudeli sooritust tuleks hinnata testikomplektil eraldi iga kolme sihtklassi jaoks, kasutades käesolevas projektis valitud mõõdikuid

Tihti on valikuks ka rist-valideerimine, kuid siinkohal on treenimine nõnda aja- ja ressursimahukas, et see poleks ajaliselt reaalne. Seda kompenseerime väga esindusliku testandmestikuga.

Reaalne kasutusjuhu testimine

Lähtuvalt andmestikust ja protseduurist mis sai tekitatud eelneva peatüki

Testandmestike ja testprotseduuride defineerimine raames, toimub siin testimine.

Tõenäoliselt leiab testimine aset ka iteratiivselt, ehk vähemalt ühe korra etapi keskel ja siis etapi lõpus.

Muudatuste mõõtmine

Muudatuste tuvastamine ja mõõtmine on oluline kahes aspektis. Esmalt, kuna sisuliselt on suur osa aastast-aastasse andmete kogumise ja andmehõive protsessi eesmärk tuvastada muutusi. Teisalt, andmehõive protsessi saab lihtsustada (kiirendada) eriti, kui keskenduda muudatuste märgendamisele. Ehk edukalt tuvastada, kus muudatust aset leidnud ei ole - neid kohti pole vaja lasta andmesisestajatel üle vaadata.

Piksli-põhise muudatuse mõõtmine

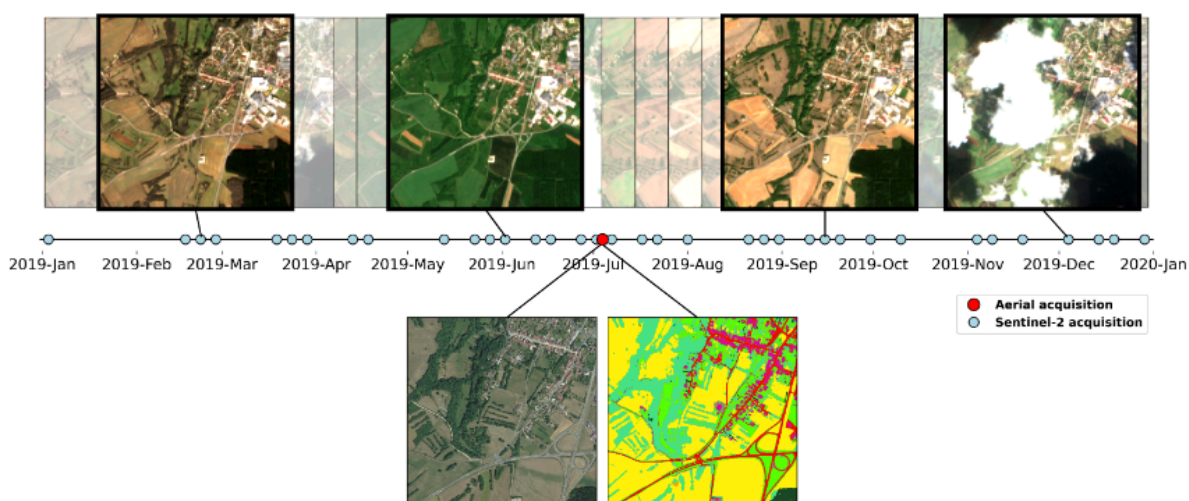
Muutuste tuvastamine toimub eri aastate andmeteset genereeritud semantiliste segmenteerimismaskide pikslipõhiste klassimärgendite võrdlemise teel. Iga piksli asukoha jaoks võrreldakse antud aasta maskis olevat ennustatud klassimärgendit järgnevate aastate maskide klassimärgenditega (või pikema ajaseeria ulatuses). Kui piksli klassimärgend muutub kahe ajapunkti vahel, näitab see maakatte muutust selles asukohas. Siinkohal tuleb ka eksperimenteerida erinevate silumise ja ühtlustamise algoritmidega, et pisimuudatused ei tekitaks liialt valepositiivseid. Samuti on huvipakkuv

mõte kasutada aastate lõikes teada-olevaid erinevusi treeningandmetes juba mudeli treenimise käigus, kuid selle mõttearenduse jätame juba projekti rüppe.

Satelliidipiltide kaasamine muudatuste mõõtmisel

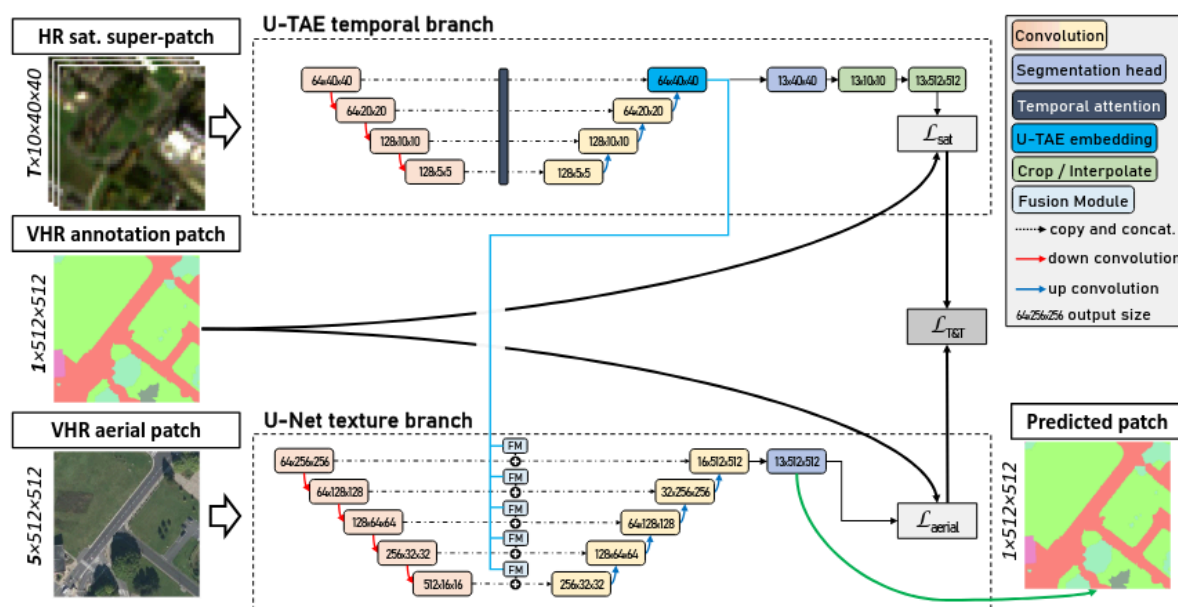
Suuremate muudatuste korral mõõtmiseksvõibolla võimalik kasutada satelliidipilte, tuginedes nende laekumise tihedusele võrreldes aerofotodega. Kirjeldatud lahendus ei pruugi olla sobilik hetkel saada olevate satelliidifotode (Sentinel 1 ja 2) resolutsiooni tõttu, aga lähenemine on huvipakkuv arvestades võimalusega tulevikus saada ligipääsu kõrgema resolutsiooniga satelliidifotodele.

Lahenduse tuumikosa moodustavad satelliidi fotod ning me plaanime neid kasutada stabiilse ja informeeritud ennustuse genereerimiseks pikkade aegridadena. Satelliidifotodele rakendatakse pilvemaski, et vähendada pilviste perioodide mõju ennustuskvaliteedile.



Selline lähenemine võimaldab meil jälgida muutusi nii lühikeses perspektiivis kui ka aasta lõikes sesoonselt ja kogu seda pikka satelliitseire aegrida veel omakorda ristvalideerida MaRU teiste kaugseire andmetega. Kõige tõenäolisem andmejaotus on, et me kasutame ajas sarnasuskahanevat tihedust ehk mida rohkem aega tagasi seda hõredamalt me võrdluspilte võtame, aga aasta tagasi samast perioodist võtame jällegi tihedamalt pilte, et oleks olemas hästi võrreldav referentsperiood. Loogika on siin see, et kui me näiteks kontrollime muudatuste veebruaris, siis referents perioodina on detsember-jaanuar ja aastatagune detsember-jaanuar oluliselt väärtuslikumad kui näiteks vahepealsed kuud. Ja vastupidi suvekuudega. Seega oleks ideaalne sisend vaatlusperiood ennustushetkest kuni 14 kuud tagasi, aga seda peiodi on võimalik ka pikemaks venitada nii palju kui andmete ajalugu on.

Sellise mudeli arhitektuuri koha pealt on tõenäoline, et kasutame modifitseeritud varianti [FLAIR-2](#) ülevaates kirjeldatud arhitektuurist:



Hoone korruste mudel

Kasutades eelnevalt arvutatud normaliseeritud digitaalset pinnamudelit (nDSM), keskendub mudeli arendus ALS ja ostofotode põhjal täpsete korruste arvu ennustusele, kust saab edasi jälgida ka nende muutust ajas. Esimese sammuna tuvastame hooned:

1. Juba märgistatud hoonete piirjoonte vektorkihte kasutatakse otse ALS-i ja vastavate aerofotode lõikamiseks teadaolevate hoonete kohalt
2. Märgistama hoonete jaoks saame kasutada projektis arendatud hoonete segmentatsioonimudelit.

Kui hooned on leitud, eraldatakse iga hoone kohta masinõppemudelisse sisestamiseks põhjalik tunnuste komplekt. Olulisimad on nDSM-ist tuletatud tunnused, mis kirjeldavad hoone vertikaalset mõõdet: maksimaalne kõrgus maapinnast, kõrguse protsentiilid (näiteks 90. või 95., et vähendada kõrvalekallete, nagu antennide mõju), keskmine kõrgus ja kõrguse standardhälve (mis viitab katuse iseloomule). Kõrgusandmeid täiendavad vastavast aerofoto lõigust saadud tunnused, mis annavad kontekstuaalset teavet, sealhulgas tekstuuriomadused (näiteks GLCM või LBP meetodeid), värviväärtuste (RGB) keskmine ja standardhälve ning erinevad muud klassikalise masinõppemise tunnused. Potentsiaalselt saaks kasutada ka kõrgetasemelised visuaalsed tunnused, mis on eraldatud eelnevalt treenitud konvolutsioonilise närvivõrgu (CNN), näiteks ResNet'i abil.

Arvestades nende eraldatud tunnuste struktureeritud, tabelikujulist olemust, keskendub esmane modelleerimisviis klassikalistele masinõppe algoritmidele. Meetodid nagu

gradientvõimendus (XGBoost, LightGBM) või juhumetsad (Random Forest) toimivad eeldatavasti hästi ning pakuvad selle ülesande jaoks sobivat tasakaalu täpsuse ja arenduse keerukuse vahel. Mudeleid treenitakse märgistatud andmestikuga, mis sisaldab eraldatud tunnuseid ja tegelikku korruste arvu. Kuigi need on tõhusad, võib esialgsete tulemuste parandamise vajadusel või ressursside olemasolul hinnata ka keerukamaid süvaõppe tehnikaid, nagu multimodaalsed fusioonvõrgud, mis suudavad ühiselt töödelda nii tooreid pildilõike kui ka LiDARi andmeid.

Lõpuks liigub süsteem muutuste tuvastamise ja hindamise juurde. Oluline samm on vastavate hoonete usaldusväärne vastavusse viimine järjestikustel aastatel, mis tavaliselt saavutatakse hoonete piirjoonte tsentroidide asukohtade võrdlemise ja püsivate unikaalsete ID-de määramisega. Sarnast ülesannet võib olla vaja lahendada ka meie üldise segmentatsioonimudeli juures. Seejärel arvutatakse iga hoone korruste muutus lihtsalt kui erinevus selle ennustatud korruste arvu vahel praegusel aastal ja eelmisel aastal. Hooned, mille muutus ei ole null, tuuakse esile.

Mudeli treenimiseks valitakse sobilik kaofunktsioon nagu keskmine absoluutviga (MAE) ja ruutkeskmine viga (RMSE). Lõplik hinnang tehakse hankija soovitusel F1 skooriga.

Mudeli arenduse etapp on iteratiivne, kasutades vigade analüüsi nõrkuste tuvastamiseks (nt probleemid teatud tüüpi hoonetega, andmete kvaliteedi varieeruvus) ja suunates täiustusi tunnuste eraldamises, segmenteerimises või mudeli valikus.

Hoone sissepääsu mudel

Selle ülesande lahendamiseks on kõige sobivam lähenemine objektituvastus, mille eesmärk on tuvastada hoonete sissepääsupunktide asukohad aerofotodelt täpsete punktidenä. Selleks plaanime kasutada süvaõppe objektituvastusmudeleid nagu näiteks Faster R-CNN, RetinaNet või EfficientDet, mis on näidanud häid tulemusi sarnaste ülesannete lahendamisel. Kuna eesmärk on ennustada punkti, mitte tüüpilist laiemat piiravat kasti, tuleb neid mudeleid veidi kohandada. Selleks on erinevaid võimalikke tehnikaid mis omakorda sõltub ka konkreetsest mudeli arhitektuurist. Näiteks Faster R-CNN puhul:

- mudel treenitud ennustama väga väikeseid piiravaid kaste, mille keskpunktid seejärel loetakse sissepääsupunktideks
- arhitektuuri saab kohandada sarnaselt võtmepunktide ennustamisele (nagu Mask R-CNN puhul kehahoiaku tuvastamisel)
- saab mudeli panna otse regresseerima sissepääsupunkti (x, y) koordinaate suhtes hoone enda piirava kastiga.

Esialgne lähenemine keskendub puhtalt aerofotode kasutamisele, kuna see on lihtsam ja kuna visuaalselt tundub täitsa reaalne ja lihtne seda teatava täpsusega teha. Intuiitiivselt

tundub, et LiDAR andmete lisamine võiks 3D-konteksti pakkudes aidata, eriti keerulisemate varikatuste või maastikudetailide tuvastamisel sissepääsude läheduses. Siiski, LiDAR andmete integreerimine lisab märkimisväärselt keerukust mudeli arhitektuuris. Seetõttu proovime LiDAR andmeid kaasata alles hilisemas faasis, kui ainult aerofotodel põhinev mudel ei anna piisavalt häid tulemusi ja projekti ajakava seda võimaldab.

Mudeli hindamisel ja praktilisel rakendamisel mõtleme lõpliku kasutusjuhu peale, kelleks on päästjad, kellel vaja õige külje alt majale läheneda. Selle asemel, et püüda ennustada täpset sissepääsupunkti, võiksime treenida mudelit hoopis tuvastama hoone õiget külge või fassaadi, kus peamine sissepääs asub. Selline lähenemine võib anda oluliselt robustsemaid ja usaldusväärsemaid tulemusi, olles samal ajal päästjate jaoks praktiliselt sama kasulik, juhatades nad kiiresti õigesse hoone küljele.



Projektiplaan ja ajakava

Projektiplaan kestused, maht ja järgnevus on ära toodud GANTT tabelis, mis on lisatud hanke failide hulka. GANTTi on märgitud ka puhveraeg, mis on viimased 3 kuud projekti lõpus. Tulp C-s on välja toodud kogu tegevuse töömaht andmeteadlasele ning sinise taustaga märgitud kastid (näiteks D4, E4, F4 "Andmete täpsustav kaardistamine" osas) tähistavad, millistele nädalatele töömaht jaotub. Järgnevalt kirjeldame ja põhjendame ära tegevuste sisu, eeldused (sh hankija roll), tulemid, läbiviijad. Tegevused on jaotatud **tegevusteks** ja **alategevusteks** - need on siin kirjeldatud, kuid nad tuginevad väga suuresti metoodika peatükis kirjeldatule.

Hankija soov on jaotada projekt 5'ks etapiks - iga mudeli kohta etapp. Meie soovitatud plaani järgi aga tuleks 3 mudelit. Kuna neist esimene on päris mahukas, siis teeme ettepaneku jagada projekti 4'ks **etapiks**. See asjaolu on kõige selgemini visuaalselt nähtav GANTT tabelis.

Meeskonnaliikmete rollid on põhjalikumalt lahti seletatud meeskonnaliikmete vormis.

Tegevuste all ei ole läbiviijana eraldi välja toodud **projektijuhti**, sest projektijuht on läbivalt kaasatud kõigis projekti tegevustes. Samuti ei ole välja toodud sidustööde tegijaid vaid ainult peamised rollid tegevuses, näiteks paljudes punktides osaleb **ruumiandmete ekspert** konsulteerivas rollis, osalevad ka hankija poolsed rollid, kuid seda tihti valideerivas või abistavas rollis, seega ei ole eraldi välja toodud. Samuti pole välja toodud korduvaid tehnilisi eeldusi, näiteks ligipääs ESTHub või LUMI infrale on välja toodud ainult esimese tegevuse all, milleks seda vaja on.

Tegevuste eest vastutamise loogika on kõigil tegevustel sama, seega toome selle siinkohal esile, mitte iga tegevuse juures eraldi:

1. Läbiviija (**andmeteadlane**) vastutab tegevuse edu eest
2. Projektijuht vastutab, et:
 - a. läbiviija panustab täies ulatuses
 - b. läbiviija töö ei oleks pärsitud projekti-üledest teguritest
 - c. vajadusel kaasab lisatööjõudu

1. Andmestike kogumine

1.1 Andmete täpsustav kaardistamine

Kirjeldus: Projekti alguses teostame olemasolevate andmeallikate põhjaliku kaardistamise, määratledes andmete kvaliteedi, kättesaadavuse ja sobivuse analüüsiks. Koostöös tellijaga määratleme, milliseid täiendavaid andmeallikaid oleks vaja kaasata parimate tulemuste saavutamiseks.

Eeldused: Tellija esindajate kättesaadavus konsultatsioonideks ja ligipääs algsetele andmeallikatele. Samuti eeldame, et tellija saab pakkuda valdkondlikku ekspertiisi andmete konteksti mõistmiseks.

Tulemid: Põhjalik ülevaade kõigist projektis kasutatavatest andmeallikatest, nende kvaliteedist ja piiratustest. Konkreetne kava, milliseid andmeid täpsemalt kasutame mudeli treenimiseks ja valideerimiseks.

1.2 Andmetorustike implementeerimine

Kirjeldus: Looime automatiseeritud andmetorustikud kõikide asjakohaste andmeallikate jaoks, tagades andmete järjepideva ja ühtlase töötlemise. Andmetorustikud disainitakse nii, et need oleksid taaskasutatavad kogu projekti vältel ja ühilduksid lõpliku lahendusega.

Eeldused: Ligipääs vajalikele süsteemidele (ESTHub vms) ja tellijapoolne tehniline tugi andmetele juurdepääsu küsimustes.

Tulemid: Töökindlad andmetorustikud, mis suudavad automaatselt hankida ja töödelda värskeid andmeid. Esimene versioon ajalooliste andmete väljavõttest analüüsiks.

1.3 Andmete puhastamine ja ühtlustamine

Kirjeldus: Teostame andmete kvaliteedikontrolli, tuvastades ja parandades puuduvaid väärtusi, ebakorrektsed andmeid ja muid probleeme, mis võivad mõjutada mudeli täpsust. Ühtlustame andmed nii, et need oleksid omavahel kooskõlas ja sobivas vormingus mudeli treenimiseks.

Eeldused: Ligipääs valdkonna ekspertidele, kes saavad aidata andmete anomaaliate tõlgendamisel ja valdkondlike otsuste tegemisel.

Tulemid: Puhastatud ja ühtlustatud andmekogum, mis on valmis masinõppe mudelite treenimiseks. Dokumentatsioon tehtud andmete puhastamise sammudest.

1.4 Testandmestike ja teststsenaariumite defineerimine

Kirjeldus: Määratleme koos tellijaga selged testjuhud ja -stsenaariumid, mis kajastavad reaalsel kasutuskonteksti. Need testid peavad katma erinevaid aastaaegu, piirkondi ja muid olulisi tegureid, et tagada mudeli üldistusvõime.

Eeldused: Tellija suudab selgelt väljendada oma vajadusi ja ootusi mudeli käitumisele erinevates olukordades.

Tulemid: Selgelt defineeritud testjuhtumid, hindamiskriteeriumid ja kvaliteedi lävendid, mis võimaldavad objektiivselt hinnata mudeli toimivust.

1.4 Jooksev koostöö hankija andmete ettevalmistaja(tega)

Kirjeldus: Teeme pidevat koostööd andmete ettevalmistaja(te)ga, et tagada treeningandmete kvaliteet ja ühtsus. Anname tagasisidet andmete kvaliteedi kohta ja teeme ettepanekuid protsessi parendamiseks.

Eeldused: Toimiv koostöö hankija andmete ettevalmistaja(tega), kes on valmis koostööks ja tagasiside vastuvõtmiseks. Selged juhised andmete märgenduspõhimõtete kohta.

Tulemid: Kvaliteetsed ja ühtlased andmete märgendused, mis toetavad mudeli edukat treenimist. Dokumenteeritud treeningandmete loomise protsess ja standardid.

1.5 Infrastruktuuri seadistamine

Kirjeldus: Seadistame vajaliku tehnilise infrastruktuuri, mis võimaldab tõhusalt treenida ja jooksutada masinõppe mudeleid. See hõlmab arvutusressursside, vajalike teekide ja keskkonnamuutujate seadistamist.

Eeldused: Juurdepääs piisavatele arvutusressurssidele (ESTHub, LUMI vms) ja vajalikele tarkvaralitsentsidele.

Tulemid: Töötav infrastruktuur, mis võimaldab kogu projekti vältel sujuvalt mudeleid treenida, testida ja jooksutada. Dokumentatsioon infrastruktuuri seadistamise kohta, mis võimaldab selle taasesitamist.

2. Testprotseduuride defineerimine

2.1 MaRu andmehõive töövoo testimise protseduuri defineerimine

Kirjeldus: Määratleme põhjaliku testimisprotseduuri Maa-ameti (MaRu) andmehõive töövoo jaoks. See hõlmab automatiseeritud ja manuaalsete testimisstsenaariumite loomist, mis simuleerivad reaalselt andmevoogu ja töötlemisprotsessi. Testid peavad kontrollima lahenduse suutlikkust tuvastada ja töödelda erinevaid geograafilisi objekte (hooned, teed, vooluveekogud) erineva kvaliteediga ortofotodelt.



Eeldused: Selge arusaam MaRu praegusest andmehõive töövoost ja nende konkreetsetest vajadustest. Ligipääs varasematele andmete töötlemise näidetele ja standarditele. MaRu esindaja kättesaadavus testimisprotseduuri valideerimiseks.

Tulemid: Dokumenteeritud testimisprotseduur, mis sisaldab erinevaid teststsenaariume, oodatavaid tulemusi ja kvaliteedi lävendeid. Automatiseeritud testskriptid, mis võimaldavad tulemuste kiiret ja objektiivset hindamist. Ühiselt kokku lepitud hindamismeetrika, mis vastab MaRu ootustele ja tööprotsessidele.

2.2 Ehitisregistri kasutusjuhu testimise protseduur defineerimine

Kirjeldus: Loomes spetsiifilise testimisprotseduuri Ehitisregistri kasutusjuhu jaoks, keskendudes hoonete tuvastamisele, korruste arvu määramisele ja nende muutuste jälgimisele ajas. Testid peavad kontrollima mudeli täpsust ja usaldusväärsust erinevates linnakeskkondades ning hoonete tüüpide puhul.

Eeldused: Põhjalik arusaam Ehitisregistri andmevajadustest ja andmekvaliteedi nõuetest. Juurdepääs olemasolevale Ehitisregistri andmestikule referentsina. Ehitisregistri esindaja osalemine testimisprotseduuri valideerimisel.

Tulemid: Detailne testimisplaan, mis sisaldab erinevaid hoonete tüüpe, asukohti ja konfiguratsioone. Selged kriteeriumid hoonete korruste arvu määramise täpsuse hindamiseks. Valideeritud testimisprotseduur, mis on kooskõlas Ehitisregistri tegelike vajaduste ja tööprotsessidega.

2.3 Päästeameti kasutusjuhu testimise protseduur defineerimine

Kirjeldus: Töötame välja testimisprotseduuri Päästeameti kasutusjuhu jaoks, keskendudes hoonete sissepääsude tuvastamisele ja täpsele lokaliseerimisele. Testprotseduur peab arvestama erinevate hoonetüüpide, ilmaolude ja ortofotode kvaliteediga, mis võivad mõjutada sissepääsude nähtavust.

Eeldused: Hea arusaam Päästeameti operatiivsetest vajadustest ja nõuetest hoonete sissepääsuinfo täpsusele. Ligipääs Päästeameti ekspertidele, kes saavad valideerida testimisprotseduuri vastavust tegelikele operatiivvajadustele. Näidisandmestik, mis sisaldab erinevat tüüpi hooneid koos valideeritud sissepääsuinfoga.

Tulemid: Põhjalik testimisprotseduur, mis katab erinevad hoonetüübid, keskkonnatingimused ja andmekvaliteedi varieeruvused. Selged kriteeriumid sissepääsude tuvastamise täpsuse hindamiseks, mis arvestavad Päästeameti operatiivseid vajadusi. Valideeritud ja dokumenteeritud testimisprotseduur, mis on kooskõlas Päästeameti tegelike tööprotsesside ja vajadustega.

3. Etapp 1 - Ühine segmentatsioonimudel - hooned/teed/vooluveekogud

3.1 Mudeli implementeerimine ja treenimine

Kirjeldus: Implementeerime ja treenime masinõppe mudeli, mis on võimeline tuvastama ja segmenteerima hooned, teed ja vooluveekogud aero- ja lidaripiltidel (potentsiaalselt kasutusel ka satelliidipildid). Valime parima mudeli arhitektuuri vastavalt eelnevalt tehtud andmeanalüüsile ja teststsenaariumide vajadustele. Treenimisprotsess hõlmab hüperparameetrite optimeerimist ja mitme treeningiteratsiooni läbiviimist.

Eeldused: Piisava mahuga kvaliteetsed ja märgendatud treeningandmed. Juurdepääs vajalikele arvutusressurssidele (arvutusklaas ESThub või LUMI) suurte mudelite treenimiseks. Selged mõõdikud mudeli soorituse hindamiseks.

Tulemid: Treenitud segmentatsioonimudel, mis suudab tuvastada ja segmenteerida hooned, teid ja vooluveekogusid erinevat tüüpi piltidelt. Mudeli konfiguratsiooni ja treenimisprotsessi põhjalik dokumentatsioon. Ülevaade mudeli sooritusest erinevate mõõdikute alusel.

3.2 Mudeli valideerimine ja täiustamine

Kirjeldus: Viime läbi põhjaliku mudeli valideerimise erinevate testandmestikega, mis kajastavad reaalselt kasutuskonteksti. Analüüsime mudeli sooritust erinevate objektiklasside lõikes. Identifitseerime probleemseid valdkonnad ja täiustame mudelit vastavalt, kasutades täiendavaid treeningandmeid või arhitektuurilisi kohandusi.

Eeldused: Esialgsed treenitud mudelid on olemas. Kvaliteetsed valideerimisandmed, mis ei kattu treeningandmetega. Tellija valdkonna ekspertide kättesaadavus mudeli väljundite kvalitatiivseks hindamiseks.

Tulemid: Valideeritud ja täiustatud segmentatsioonimudel. Detailne valideerimisraport, mis dokumenteerib mudeli sooritust erinevates tingimustes. Konkreetseid täiustusi, mis on tehtud esialgse mudeli parendamiseks.

3.3 Muudatuste tuvastuse meetodite arendamine

Kirjeldus: Arendame välja meetodid muudatuste tuvastamiseks ajas, võimaldades jälgida hoonete, teede ja veekogude muutusi erinevate ajaperioodide piltide alusel. See

hõlmab algoritmide loomist, mis võrdlevad erinevate ajahetkede segmentatsioone ja tuvastavad olulisi muutusi, eristades neid müra või hooajalistest muutustest.

Eeldused: Töötav segmentatsioonimudel on olemas. Juurdepääs mitmele ajaperioodile samast piirkonnast. Selged kriteeriumid, mis määratlevad olulise muutuse valdkonna kontekstis.

Tulemid: Toimiv muudatuste tuvastamise süsteem, mis suudab automaatselt identifitseerida ja klassifitseerida muutusi hoonestuses, teede võrgustikus ja voolaveekogu piirides. Algoritmi sooritus mõõdetuna täpsuse ja tundlikkuse näitajates. Dokumentatsioon, mis selgitab muudatuste tuvastamise metodoloogiat ja selle piiranguid.

3.4 MaRu andmehõive töövoos testimine

Kirjeldus: Testime kogu lahendust integreeritud MaRu andmehõive töövooga, simuleerides tegelikku kasutuskonteksti. Hindame lahenduse suutlikkust töötada reaalajas, käsitleda erinevaid andmesisendeid ja toota väljundeid, mis vastavad MaRu kvaliteedinõuetele. Jälgime kogu protsessi jõudlust ja stabiilsust pikema aja jooksul.

Eeldused: Valmis mudelid ja muudatuste tuvastamise süsteem on olemas. Juurdepääs MaRu andmehõive töövoole või selle simulatsioonile. MaRu esindajate kättesaadavus testimise tulemuste hindamiseks.

Tulemid: Kinnitatud lahendus, mis on edukalt testitud MaRu andmehõive töövoos. Dokumenteeritud testi tulemused, sealhulgas jõudluse mõõdikud ja võimalikud kitsaskohad. Soovitused optimeerimiseks ja edasisteks täiustusteks, mis on tuvastatud testimisprotsessi käigus.

4. Etapp 2 - Ühine segmentatsioonimudel - teed/vooluveekogud/hooned

4.1 Mudeli implementeerimine ja treenimine

Kirjeldus: Implementeerime ja treenime spetsialiseeritud segmentatsioonimudeli, mis on keskendatud hoonete, teede ja vooluveekogude tuvastamisele. Tuginedes eelmise etapi kogemustele, kohandame mudeli arhitektuuri ja treeningstrateegiat, et saavutada parem täpsus nende objektitüüpide puhul. Rakendame vajadusel täiendavaid tehnikaid nagu klasside kaalude tasakaalustamine või spetsiifilised andmerikastamise meetodid.

Eeldused: Kvaliteetsed märgendatud andmed hoonete, teede ja vooluveekogude kohta. Eelmisest etapist saadud kogemused ja teadmised, mis aitavad vältida potentsiaalseid probleeme. Piisavad arvutusressursid mudeli treenimiseks.

Tulemid: Optimeeritud segmentatsioonimudel. Dokumenteeritud treenimisprotsess, sealhulgas valitud parameetrid ja arhitektuurilised otsused.

4.2 Mudeli valideerimine ja täiustamine

Kirjeldus: Viime läbi põhjaliku valideerimisprotsessi. Tuvastame probleemseid stsenaariumid (näiteks varjud, sesoonsed muutused veekogudes) ja täiustame mudelit nende käsitlemiseks. Rakendame täiendavaid tehnikaid, mis parandavad segmentatsiooni kvaliteeti erinevates olukordades.

Eeldused: Esialgne treenitud mudel 4.1 punktist. Mitmekesine valideerimisandmestik, mis sisaldab erinevaid keskkonnatingimusi ja maastikutüüpe. MaRu ekspertide kättesaadavus valideerimistulemuste hindamiseks.

Tulemid: Täiustatud ja valideeritud segmentatsioonimudel. Detailne valideerimisraport, mis dokumenteerib mudeli sooritust erinevate stsenaariumide lõikes. Konkreetseid täiustused, mis on tehtud eelmise etapi kogemuste põhjal.

4.3 Muudatuste tuvastuse meetodite arendamine

Kirjeldus: Arendame edasi muudatuste tuvastamise meetodeid, mis on spetsiaalselt kohandatud hoonete, teede ja vooluveekogude dünaamika jälgimiseks. See hõlmab algoritmide loomist, mis suudavad eristada tegelikke muutusi (nagu tee-ehitus või voolusängi muutused) hooajalistest variatsioonidest või sensori erinevustest olenemata. Implementeerime meetodid muudatuste tuvastamiseks.

Eeldused: Eelmises etapis arendatud muudatuste tuvastamise algmeetodid. Aegridade andmed samadest piirkondadest, mis võimaldavad jälgida hoonete, teede ja vooluveekogude muutusi. Selged kriteeriumid, mis määratlevad erinevad muudatuste tüübid ja nende olulisuse.

Tulemid: Täiustatud muudatuste tuvastamise süsteem, mis on optimeeritud hoonete, teede ja vooluveekogude muutuste jälgimiseks. Muudatuste klassifitseerimise meetoodika, mis eristab muutusi. Dokumentatsioon, mis selgitab täiustatud süsteemi toimimist ja selle rakendamist praktikas.

4.4 MaRu andmehõive töövoo testimine

Kirjeldus: Testime spetsialiseeritud hoonete, teede ja vooluveekogude segmentatsioonimudeli ja muudatuste tuvastamise süsteemi integreeritust MaRu

andmehõive töövooga. Hindame süsteemi suutlikkust automaatselt tuvastada ja raporteerida olulisi muutusi hoonete, teede ja vooluveekogude piirides. Analüüsime süsteemi jõudlust ja täpsust tegelikes töötingimustes.

Eeldused: Täiustatud segmentatsioonimudel ja muudatuste tuvastamise süsteem on valmis. Defineeritud MaRu andmehõive töövoog, mis võimaldab lahenduse testimist. MaRu tiimi liikmete osalemine testimistulemuste hindamisel.

Tulemid: Valideeritud lahendus, mis on tehtud MaRu tegelike andmete töötlemisel. Dokumenteeritud testi tulemused, mis kinnitavad süsteemi suutlikkust vastata MaRu vajadustele. Konkreetseid näited tuvastatud muutustest ja nende klassifikatsioonist, mis demonstreerivad süsteemi praktilist väärtust.

5. Etapp 3 - Hoone kõrguse mudel

5.1 Mudeli implementeerimine ja treenimine

Kirjeldus: Implementeerime ja treenime spetsialiseeritud mudeli hoonete korruste arvu määramiseks, ühendades ortofotod ja LiDAR-põhised kõrgusandmed. Rakendame sobivat masinõppe algoritmi, mis suudab tuvastada hoonete kõrgust maapinnast ning konverteerida selle korruste arvuks. Mudel töötatakse välja nii, et see arvestaks erinevate hoonetüüpide arhitektuurilisi erinevusi.

Eeldused: Ligipääs kvaliteetsetele LiDAR andmetele ja ortofotodele. Märgeandmed, mis sisaldavad hoonete tegelikke korruste arve. Eelnevalt implementeeritud hoonete segmentatsioonimudel, mis võimaldab isoleerida hooned muust maastikust.

Tulemid: Treenitud mudel, mis suudab ennustada hoonete korruste arvu kõrgusandmete ja ortofotode põhjal. Dokumenteeritud mudeli arhitektuur ja treenimisprotsess. Esmane hinnang mudeli täpsusele kokkulepitud mõõdikute alusel.

5.2 Mudeli valideerimine ja täiustamine

Kirjeldus: Viime läbi põhjaliku valideerimisprotsessi, hinnates mudeli täpsust erinevate hoonetüüpide (korterimajad, eramajad, ärihooned, tööstushooned) ja arhitektuuriliste eripärade lõikes. Tuvastame ja analüüsime süstemaatilisi vigu ning täiustame mudelit nende lahendamiseks. Katsetame alternatiivseid algoritmilisi lähenemisi ja parameetreid optimaalse konfiguratsiooni leidmiseks.

Eeldused: Esialgne treenitud mudel punktist 5.1. Mitmekülgne valideerimisandmestik, mis sisaldab erinevaid hoonetüüpe ja ehitusstiile. Ehitisregistri andmed hoonete korruste arvu kontrollimiseks. Valdkonna ekspertide kättesaadavus tulemuste kvalitatiivse hindamise jaoks.

Tulemid: Täiustatud ja valideeritud mudel, mis demonstreerib kõrget täpsust hoonete korruste määramisel. Detailne valideerimisraport, mis dokumenteerib mudeli sooritust erinevate hoonetüüpide ja piirkondade lõikes. Selgelt määratletud täpsuspiirid ja piirangud erinevate stsenaariumide jaoks.

5.3 Muudatuste tuvastuse meetodite arendamine

Kirjeldus: Arendame spetsiifilised meetodid hoonete kõrguste ja korruste arvu muutuste tuvastamiseks ajas. Luuakse algoritmid, mis võrdlevad erinevatest ajaperioodidest pärit hoonete andmeid ning tuvastavad olulised muutused (näiteks juurdeehitused, lisakorrused). Implementeerime lahendused, mis suudavad eristada tegelikke struktuuralseid muutusi mõõtmistest või andmekvaliteedi variatsioonidest tingitud erinevustest.

Projekti käigus vaatame üle, kas siin on üldse vaja muudatusi mõõta. Hetkel tundub, et mitte, vaid vaja on lihtsalt andmestikus puuduvolevaid automaatselt märgendada.

Eeldused: Täpne hoonete korruste määramise mudel punktist 5.2. Mitmete ajaperioodide andmed samadest piirkondadest, mis võimaldavad jälgida ehitiste muutusi. Selged kriteeriumid, mis määratlevad olulise muutuse hoonete kõrguses või korruste arvus.

Tulemid: Toimiv muudatuste tuvastamise süsteem hoonete kõrguste ja korruste muutuste jälgimiseks. Algoritm, mis minimeerib valepositiivseid tulemusi. Dokumentatsioon, mis selgitab muudatuste tuvastamise metodoloogiat ja kriteeriumeid.

5.4 Ehitisregistri kasutusjuhu testimine

Kirjeldus: Testime väljatöötatud lahendust Ehitisregistri reaalset kasutusjuhtu simuleerivas keskkonnas. Hindame, kuidas mudel suudab toetada Ehitisregistri tööprotsesse, pakkudes automaatset korruste arvu tuvastamist. Analüüsime lahenduse täpsust, töökindlust ja efektiivsust võrreldes praeguse manuaalse protsessiga.

Eeldused: Töökorras mudel hoonete korruste määramiseks ja muutuste tuvastamiseks. Ligipääs Ehitisregistri andmetele mudeli valideerimiseks. Ehitisregistri esindajate osalemine testimisprotsessis, et hinnata lahenduse praktilisust ja kasutusmugavust.

Tulemid: Valideeritud lahendus, mis on edukalt testitud Ehitisregistri kasutusjuhuga. Dokumenteeritud testitulemused, mis näitavad süsteemi täpsust võrreldes Ehitisregistri

tegelike andmetega. Hinnang potentsiaalsele ajavõidule ja efektiivsuse tõusule, mida lahendus pakub võrreldes praeguse protsessiga. Soovitused lahenduse integreerimiseks Ehitisregistri tööprotsessidesse.

6. Etapp 4 - Hoone sissepääsu mudel

6.1 Mudeli implementeerimine ja treenimine

Kirjeldus: Implementeerime ja treenime objektituvastuse mudeli hoonete sissepääsude tuvastamiseks aerofotodelt. Rakendame kaasaegseid süvaõppe arhitektuure (nagu Faster R-CNN, RetinaNet või EfficientDet), mis on kohandatud punktobjektide tuvastamiseks. Mudel on treenitud tuvastama erinevate hoonetüüpide peamisi sissepääse, arvestades nii visuaalseid tunnuseid ortofotodelt kui ka täiendavat konteksti (näiteks sissepääsude tüüpiline paiknemine hoone fassaadil ja seos ümbritseva taristuga).

Eeldused: Piisav hulk märgendatud andmeid, kus on tähistatud hoonete sissepääsud. Toimiv hoonete segmentatsioonimudel eelmistest etappidest, mida saab kasutada huvi pakkuvate alade piiramiseks. Piisavad arvutusressursid konvolutsiooniliste närvivõrkude treenimiseks.

Tulemid: Treenitud mudel, mis suudab aerofotodelt tuvastada hoonete sissepääsude asukohti. Dokumenteeritud treenimisprotsess ja valitud mudeli arhitektuur. Esmane hinnang mudeli täpsusele ja usaldusväärsusele.

6.2 Mudeli valideerimine ja täiustamine

Kirjeldus: Viime läbi põhjaliku valideerimisprotsessi, testides mudeli võimekust tuvastada sissepääse erinevate hoonetüüpide, ilmastikutingimuste ja pildistamisnurdade puhul. Analüüsime mudeli vigu ja ebaselgeid juhtumeid, selgitades välja probleemsed stsenaariumid. Täiustame mudelit täiendavate treeningandmete või algoritmi kohandamisega, keskendudes eriti Päästeameti jaoks kriitiliste juhtumite täpsuse parandamisele.

Eeldused: Esialgne treenitud mudel punktist 6.1. Mitmekesine valideerimisandmestik, mis sisaldab erinevaid hoonetüüpe ja keskkondi. Päästeameti ekspertide kaasamine, et hinnata tuvastatud sissepääsude praktilisust operatiivtöö kontekstis.

Tulemid: Täiustatud ja valideeritud mudel, mis demonstreerib kõrget täpsust sissepääsude tuvastamisel. Detailne valideerimisraport, mis dokumenteerib mudeli

sooritust erinevates tingimustes. Selgitused mudeli piirangute ja usaldusväärsuse kohta erinevates stsenaariumides.

6.3 Muudatuste tuvastuse meetodite arendamine

Kirjeldus: Arendame meetodid sissepääsude muutuste tuvastamiseks ajas, võimaldades jälgida muudatusi hoonete ligipääsetavuses. Looime algoritmid, mis võrdlevad eri ajaperioodide sissepääsuandmeid ning tuvastavad olulisi muutusi (uued sissepääsud, suletud sissepääsud, muudetud ligipääsuteed). Implementeerime lahendused, mis aitavad hoida Päästeameti andmeid ajakohasena ja operatiivselt relevantsetena.

Projekti käigus vaatame üle, kas siin on üldse vaja muudatusi mõõta. Hetkel tundub, et mitte, vaid vaja on lihtsalt andmestikus puuduvolevaid automaatselt märgendada.

Eeldused: Täpne sissepääsude tuvastamise mudel punktist 6.2. Erinevate ajaperioodide andmed samadest piirkondadest, mis võimaldavad jälgida sissepääsude muutusi. Selged kriteeriumid, mis määratlevad olulise muutuse Päästeameti operatiivtöö kontekstis.

Tulemid: Toimiv muudatuste tuvastamise süsteem, mis suudab identifitseerida ja klassifitseerida sissepääsudega seotud muutusi. Automatiseeritud protsess, mis genereerib hoiatusi potentsiaalselt oluliste muutuste kohta. Dokumenteeritud metoodika, mis selgitab muudatuste tuvastamise algoritmi toimimist ja prioritseerimist.

6.4 Päästeameti kasutusjuhu testimine

Kirjeldus: Testime väljatöötatud lahendust Päästeameti operatiivtööd simuleerivas keskkonnas, keskendudes lahenduse praktilisele väärtusele kriisiolukordades. Hindame, kuidas mudel toetab Päästeameti tööprotsesse, pakkudes kriitilise tähtsusega informatsiooni hoonete ligipääsetavuse kohta. Analüüsime lahenduse töökindlust, täpsust ja kasutatavust operatiivtöö kontekstis.

Eeldused: Töökorras mudel sissepääsude tuvastamiseks ja muutuste jälgimiseks. Päästeameti esindajate aktiivne osalus testimisprotsessis. Testimisstsenaariumid, mis peegeldavad tegelikke operatiivolukordi.

Tulemid: Valideeritud lahendus, mis on edukalt testitud Päästeameti tegelikes töötingimustes. Dokumenteeritud testimistulemused, mis näitavad lahenduse praktilist väärtust. Hinnang lahenduse potentsiaalsele mõjule operatiivtöö efektiivsusele ja ohutusele. Soovitused lahenduse integreerimiseks Päästeameti infosüsteemidesse ja tööprotsessidesse.

7. Dokumentatsioon ja muud lõputegevused

7.1 Mudeli ja koodi dokumenteerimine

Kirjeldus: Koostame põhjaliku tehnilise dokumentatsiooni kõigi projekti käigus loodud mudelite ja koodi kohta vastavalt hanke nõuetele. Dokumenteerime iga mudeli lõppversiooni spetsifikatsiooni, arhitektuuri, treenimisprotsessi, kasutatud andmeid ja nende kombineerimise põhimõtteid. Koodi varustame põhjalike kommentaaridega, mis selgitavad funktsioonide eesmärki, sisendit ja väljundit. Dokumentatsioon sisaldab ka iga mudeli täpsushinnangut ja vastavust sihttasemetele. Kõik dokumendid koostatakse struktureeritult ja selges eesti keeles, mis võimaldab Maa- ja Ruumiametil mudeleid iseseisvalt edasi kasutada, muuta ja arendada.

Eeldused: Kõigi mudelite lõppversioonid on valmis ja töökindlad. Mudelite testimised on lõppenud ja vastavus sihttasemetele on kindlaks määratud.

Tulemid: Iga mudeli lõppversiooni detailne kirjeldusdokument, mis sisaldab andmeallikate kirjeldust, metoodika spetsifikatsiooni, täpsushinnangut ja kasutusjuhiseid. Kommenteeritud lähtekood, mis võimaldab mudelite treenimist ja jooksumist korrata. Mudelite töö tulemuste näited piltidena, mis illustreerivad süsteemi toimivust erinevates tingimustes. Dokumentatsioon, mis vastab täielikult hankes toodud nõuetele.

7.2 Lõpparuanne (vähemalt 4 nädalat enne projekti lõppu)

Kirjeldus: Koostame põhjaliku lõpparuande, mis koondab kogu projekti käigus tehtud töö, saavutatud tulemused ja tähelepanekud. Aruanne sisaldab detailset ülevaadet kõigist kasutatud andmeallikatest, nende analüüsimise ja kombineerimise meetoditest koos viidetega. Kirjeldame mudelite huvipakkuvate vaheversioonide tulemusi koos põhjendusega, miks need valiti. Hindame iga mudeli vastavust hanke tehnilises kirjelduses sätestatud sihttasemetele. Lisame soovitusel Maa- ja Ruumiametile tehnilise taristu arendamiseks ning võimalikeks jätkuprojektideks, mida käesoleva projekti skoobis polnud võimalik teostada.

Eeldused: Kõik projektis määratletud tegevused on lõpetatud või lõppfaasis. Iga mudeli lõppversiooni kirjeldusdokument on valmis. Mudelite vaheversioonide tulemused on dokumenteeritud ja analüüsitud. Kõik projekti jooksul kogetud tehnilised väljakutsed on kaardistatud ja nende lahendused dokumenteeritud.

Tulemid: Põhjalik lõpparuanne PDF formaadis, mis vastab hanke tehnilises kirjelduses toodud nõuetele. Aruanne sisaldab kõiki mudeli kirjeldusdokumente, vaheversioonide tulemusi, andmeallikate kirjeldusi, viiteid teadusartiklitele, projekti tegevuste ajagraafikut,

soovitusi tehnilise taristu ja jätkuarenduste osas ning projekti järeldusi ja hinnanguid. Aruanne on koostatud selges ja arusaadavas eesti keeles.

7.3 Muu dokumentatsioon ja lõputegevused

Kirjeldus: Lisaks põhilistele dokumentidele koostame ja anname üle kõik hanke tehnilises kirjelduses nõutud lisamaterjalid. Valmistame ette projekti tutvustava lühikokkuvõtte eesti ja inglise keeles. Koostame tulemusi tutvustavad esitlusslaidid ja esitleme neid. Viimistleme ja anname üle kõikide mudelite lõpp- ja huvipakkuvate vaheversioonide koodi. Lisaks teostame vajalikud lisamaterjalide koostamised, mis tagavad projekti tulemite lihtsa kasutusele võtmise ja jätkusuutlikkuse Maa- ja Ruumiametis.

Eeldused: Kõik projekti põhilised tulemid on valmis ja dokumenteeritud. Mudelid on testitud ja nende toimimine on kinnitatud nii tehnilise võimekuse kui ka praktilise kasutatavuse seisukohast Maa- ja Ruumiameti, Päästeameti ja Ehisregistri kasutusjuhtude raames. Kõigi loodud materjalide autoriõiguste küsimused on lahendatud vastavalt lepingule.

Tulemid: Projekti tutvustav kokkuvõte (1 A4) eesti ja inglise keeles PDF formaadis. Tulemeid tutvustavad esitlusslaidid eesti keeles PPTX või PDF formaadis. Kõikide mudelite lõppversioonide ja huvipakkuvate vaheversioonide kommenteeritud treening- ja jooksumiskood Pythoni keeles. Kõik materjalid vastavad hanke tehnilises kirjelduses peatükis 5.5 toodud nõuetele ja on üle antud kokkulepitud formaadis.

7.4 Projekti tulemusi tutvustav esitus

Kirjeldus: Valmistame ette ja viime läbi projekti tulemusi tutvustava esitluse juhtrühmale, töörühmale, laiendatud töörühmale ja teistele huvitatud osapooltele. Esitlusel tutvustame projekti eesmärke, metoodikat, saavutatud tulemusi ja nende potentsiaalset mõju Maa- ja Ruumiameti, Päästeameti ja Ehisregistri tööle. Demonstreerime reaalselt iga mudeli toimivust, näidates näiteid tuvastatud teedest, vooluveekogudest, hoonetest, hoonete korrustest ja sissepääsudest. Toome esile projekti käigus tekkinud väljakutsed ja nende lahendused ning anname soovitusi jätkutegevusteks.

Eeldused: Kõik projekti tulemid on valmis ja valideeritud. Esitlusmaterjalid on koostatud ja Maa- ja Ruumiametiga kooskõlastatud. Esitluseks vajalik tehniline keskkond on ette valmistatud, võimaldades mudelite reaalselt demonstreerimist. Kõik asjaosalised on esitlusele kutsutud.

Tulemid: Professionaalne ja informatiivne esitus, mis tutvustab projekti tulemusi selgelt ja arusaadavalt nii tehnilistele kui mittetehnilistele osapooltele. Esitlusmaterjalid, mis jäävad Maa- ja Ruumiametile kasutamiseks ka peale projekti lõppu. Arutelu tekkinud

küsimustele vastused ja täiendavad selgitused projekti tulemite kohta. Soovitused edasisteks sammudeks süvaõppe rakendamisel ruumiandmete hõives.

Riskianalüüs ja riskide maandamismeetmed

Projektiga kaasnevate võimalike projektiplaani edukat järgimist takistavate riskide kirjeldus, iga riski kohta selle esinemise tõenäosus ja tagajärg ning nende vältimise ja/või maandamise võimaluste kirjeldus ja lahendusettepanekud on välja toodud eraldi dokumendis "Riskianalüüs.xlsx".

Projekti eelarve

Projekti eelarve kujuneb vastavalt tunnihinnale 90 EUR/h. Projekti kestuseks on arvestatud 24 kuud, millest viimased 3 kuud on puhveraeg.

Projektijuhi ja ruumiandmete eksperdi keskmiseks (hõlmab ka kolme viimast puhverajaks mõeldud kuud) igakuiseks tundide mahuks hindame 20h. Seega saame projektijuhi ja ruumiandmete eksperdi summaarseks tundide kuluks $2 \cdot 24 \cdot 20h = 960h$. Valemis tähistab "2" kahte meeskonnaliiget ning "24" kuude arvu. Tunnihinnaga 90 EUR/h teeb see eelarve $960h \cdot 90 \text{ EUR/h} = 86400 \text{ EUR}$.

Andmeteadlase eeldatav tundide maht nädalates on välja loetav Ganttt-tabelist, kus on näidatud tegevused, millistel nädalatel need toimuvad ning milline on andmeteadlase töökoormus iga tegevuse all. Gantt-tabelist lähtuv andmeteadlase töö nädalate maht on 57, mis teeb $57 \cdot 40h/\text{nädalas} = 2280h$. Tunnihinnaga 90 EUR/h teeb see eelarve $2280h \cdot 90 \text{ EUR/h} = 205200 \text{ EUR}$.

Sellest lähtuvalt on projekti kogueelarve $86400 \text{ EUR} + 205200 \text{ EUR} = 291600 \text{ EUR}$.