KOKEILUKIIHDYTTÄMÖ 2024

LOPPURAPORTTI

Yleisten alueiden omaisuus-rekisteritietojen paikkansa-pitävyyden parantaminen (ID97)

Ritva Keko, Esko Laiho, Saara Rossi, Soile Aarnio, Lasse Kuisma ja (Virpi Vertainen) / KYMP

Hilla Tilhi ,Timo Kohtala / Digia Oy

Helsinki

YLREn tippa

HISSIPUHE

Tiedolla on erilaisia vaikutuksia. Tiedon laatu vaikuttaa suoraan päätöksentekoon, asiakastyytyväisyyteen ja operatiiviseen tehokkuuteen. Oikea tieto lisää luotettavuutta/uskottavuutta, vaikuttaa omaisuuden arvoon, mahdollistaa tuottavuuden kasvun, säästää kaupungin varoja, jne.

Yleisten alueiden rekisterin (YLRE) tiedon laadun parantaminen on tällä hetkellä satunnaista sekä reaktiivista, minkä takia tiedon laadun ongelmia korjataan adhoc -periaatteella. Projektin keskeisenä tavoitteena on rekisterissä olevan tieto-omaisuuden laadun parantaminen kehittämällä tiedon laadun parantamiseen määritelty automatisoitavissa oleva prosessi. Prosessin toteuttamiseksi tarvitsemme toimivia keinoja tarkastella rekisterin tietojen paikkansapitävyyttä ja löytää helposti korjaamista vaativat kohteet.

Projektissa on keskitytty rekisterin aluemaisten kohteiden virheiden löytämiseen kuvantunnistusta ja tekoälyä hyödyntäen. Tekniikalla tarkistusta olisi mahdollista tehdä säännöllisesti ja se mahdollistaisi rekisterin tiedon laadun tason selkeän nostamisen koskien pelkästään aluemaisia kohteita.

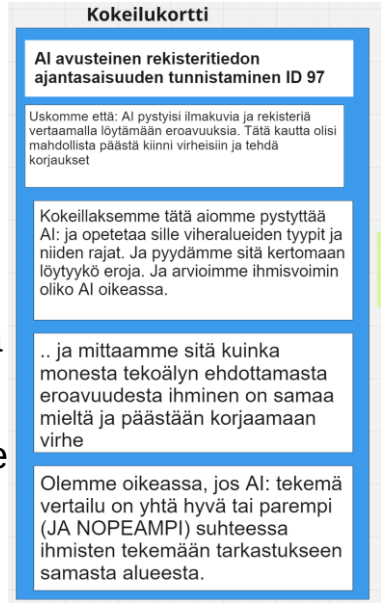
Yleisten alueiden tiedot toimivat Helsingissä monien toimintojen pohjana. Rekisterin sisältämän tiedon paikkansapitävyydellä ja oikeellisuuden parantamisella on vaikuttavuutta. Tiedon laadun parantaminen ei ole yksittäisen palvelun tehtävä, vaan koko organisaation yhteinen vastuu. Ymmärtämällä tiedon laadun merkityksen ja sitoutumalla sen parantamiseen, voimme saavuttaa merkittäviä tuloksia ja viedä organisaatiomme tiedolla johtamisen uudelle tasolle. Projektimme on osaltaan pieni askel eteenpäin tällä tiellä.

Sisällysluettelo

1. Tiivistelmä
2. Kokeilun tavoitteet
3. Kokeilun keskeiset opit
4. Kokeilun eteneminen
5. Kokeilun tuotokset
6. Opit kokeiltavan ratkaisun tai toimintatavan mahdollisuuksista
7. Opit asiakkaiden tai palvelun käyttäjien tarpeista
8. Opit ratkaisun kehittämisestä teknisesti
9. Opit kokeilemisesta yleensä
10. Opit kokeiluprojektin arjen pyörittämisestä
11. Kokeilun tekninen ympäristö
12. Kokeilun data
13. Jatkopäätökset ja -ideat

1. Tiivistelmä

- **Kokeilun tavoitteet**
  * Testata tekoälyn mahdollisuuksia tunnistaa omaisuustietorekisterin aluemaisten kohteiden ja ilmakuvan välillä poikkeavuuksia, jotka vaativat tarkempaa tarkastelua ja mahdollisesti tietojen päivitystä
  * Luoda käsitys ja tietopohjaa, toimiiko tekoäly rekisterin tietosisällön laadun tarkastelussa. Onko tekoälyyn perustuva toimintamalli toimiva, riittävän nopea ja tarkka?
- **Kokeilun opit**
  * Koulutettu koneoppimismalli soveltuu omaisuustietorekisterin laadunvarmistusmenetelmäksi
- **Suositukset jatkotoimenpiteiksi**
  * Kokeilu oli hyödyllinen, jatkokehittämistä ja menetelmän käyttöönottoa suositellaan.
  * Jatkokehittäminen vaatii projektoimisen
- **Kokeilualue**
  * Kokeilualueeksi valittiin Kumpulan kaupunginosan viheralueiden aluemaiset kohteet.



Helsinki

2. Kokeilun kuvaus

Yleisten alueiden rekisterin (YLRE) omaisuustiedoissa on virheitä. Rekisterin tiedot ovat tyypiltään alueita, viivoja tai pisteitä. Virheet heikentävät rekisterin luotettavuutta ja paikkansapitävyyttä. Rekisterin tietosisältö ei ole laadullisesti hyvällä tasolla. Kokeilussa testattiin uutta toimintamallia ja tekoälyn mahdollisuuksia rekisterin paikkansapitävyyden tarkastelussa.

Rekisterin virheiden löytäminen on sattumanvaraista ja perustunut manuaaliseen tarkasteluun, mitä ei ehditä tehdä systemaattisesti. Rekisterin tiedoissa on virheitä ja tietopuutteita. Eri tarkastelussa on havaittu mm., että kaakkoisen suurpiirin tiedoissa noin joka kymmenes kohde on virheellinen ja Kumpulan kokeilualueella noin joka neljäs kohde.

Kumpulan kokeilualueen rekisteritiedot tarkastettiin ennakkoon manuaalisesti ja tehtiin havainnot poikkeavuuksista rekisteritiedon ja ORTO-ilmakuvan välillä. Kokeilussa konenäölle opetettiin viheraluetyyppien pää- ja alaluokat sekä niiden ulkonäöt. Testausvaiheessa koulutettu konenäkö vertasi rekisterin tietoja ORTO-ilmakuviin.

Helsinki

3. Kokeilun keskeiset opit

Kokeilun onnistumista arvioitiin kolmella ennakkoon asetetulla mittarilla, joista kaikki toteutuivat hyvin tai erinomaisesti.

Määrä mittari	Aika mittari	Laatu mittari
Hypoteesi: Tekoälyavusteinen koneoppimismalli tunnistaa 80 % ne aluemaiset kohteet, joissa manuaalisessa tarkastelussa havaittiin poikkeamia rekisteritiedon ja ilmakuvan välillä.	**Onko koneoppimismalli riittävän tehokas** (=aikaa säästävä)? Manuaalisessa tarkastelussa kului kokeilualueen kaikkien viheralueiden aluemaisten kohteiden (1 168 kpl) läpikäyntiin reilut 2 vk työaikaa.	**Onnistuuko tekoäly arvioimaan laadullista onnistumista?** Ennustaako koneoppimismalli oikein, kuinka hyvin se on tunnistanut virheelliset ja tiedoiltaan korjaamista vaativat aluemaiset kohteet?
TOTEUMA: Malli oppi tunnistamaan n. 87 % silmävaraisesti tunnistetut poikkeavuudet ja virheelliseksi todetuista aluemaisista 78 kohteesta 76 kohdetta.	**TOTEUMA:** Koneoppimismalli tunnistaa 2 kohdetta sekunnissa. Aikaa em. kohteiden läpikäyntiin kuluu alle 10 min.	**TOTEUMA:** Konenäkömallin tulostiedostossa ilmoittama korkea varmuusprosentti lukuarvo (lähellä1,0) kertoo, että tunnistus on oikea. Kun lukuarvo x < 0,85 tunnistus ei ehkä ole oikein.

Helsinki

4.Kokeilun eteneminen

9 viikko palaveria
4 erillistä palaveria
(aloitus, data, pilvipalvelu, tekoälyn koulutus)

*Kokouskäytännöt
*Lähtömateriaalin läpikäynti
*Aikataulu
*Vakio kokouslokipohja
*Tuntiseurantataulukko
*Projektisuunnitelma (edellytykset, resurssit, riskit, tavoitteet – hyöty /tuotos, mittarit, viestintä)
*Pilvipalvelu/tekniset vaatimukset

*Pilvipalvelu käyttöön ja koulutus käyntiin vk 45 – 46 → ensimmäiset tulokset
*Lisädatan käyttöönotto (virheelliset kohteet /korjaus), vanhempi ORTO-kuva
*Koulutus tapahtui vk 45-46
*Mittareiden toteumat
*Tuotokset

KK	**Hankintavaihe:** TP, tarjous, tilaus, Aloituskokous 25.9./rutiinit Lähtömateriaali (Teams) (rekisteridata/aluemaiset kohteet, ORTO-ilmakuvat)				**Valmisteluvaihe:** Projekti & yhteistyö käyntiin: Fasiliteetit käyttöön, lähtödatan muokkaus/käsittely/karsinta, **pilvipalvelun odottelua**					**Tekoälyn koulutus:** Konenäkömallin koulutus, lisädatan käyttöönotto, tuotokset, tulosten ja projektin onnistumisen arviointi				**Projektin päättäminen:** Loppuraportin viimeistely Päätelmät Viestintä			
VK	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50		

Helsinki

5. Kokeilun tuotokset

- Tiedot kokeilualueen (Kumpulan alueen rekisterin aluemaiset kohteet) käsitellyistä kohteista (Container Azure Blob Storagessa)

- Koulutettu konenäkömalli Azure AI Vision -palvelussa (Huom. konenäkömalli ei ole suoraan siirrettävissä toiselle alustalle). Aineistoa voidaan käyttää suoraan uusien kohteiden alatyypin tunnistamiseen.

- Excel-tulostiedosto Kumpulan alueen kohteissa esiintyvistä virheistä ja niille tunnistetuista oikeista luokituksista tarkempaa vertailua varten

FID	Viheralue nr	Osan nr	Virhe rekisteris	Päätyyp	Alatyyppi	Tunnistettu alatyyppi	Varmuus	Tunnistet	Virhe havaittu	Materiaali	Huomiot 1
9325	1824	31	0	Muut alueet	Avokallio	Avokallio	0.999777138	1		Luonnontilainen	
8866	1816	12	0	Muut alueet	Avokallio	Avokallio	0.999712646	1		Luonnontilainen	kallio ulottuu oikealla pu
11094	1816	9	0	Muut alueet	Avokallio	Avokallio	0.999678135	1		Luonnontilainen	
8755	1816	13	0	Muut alueet	Avokallio	Avokallio	0.999675274	1		Luonnontilainen	
8755	1816	13	0	Muut alueet	Avokallio	Avokallio	0.999641776	1		Luonnontilainen	
10846	1824	45	0	Muut alueet	Avokallio	Avokallio	0.99955672	1		Luonnontilainen	
8976	1817	76	0	Muut alueet	Avokallio	Avokallio	0.99947983	1		Luonnontilainen	eri muoto, jatkuuko pide

Esimerkki osasta tulostiedostoa

Tulostiedoston varmuus% lukuarvo

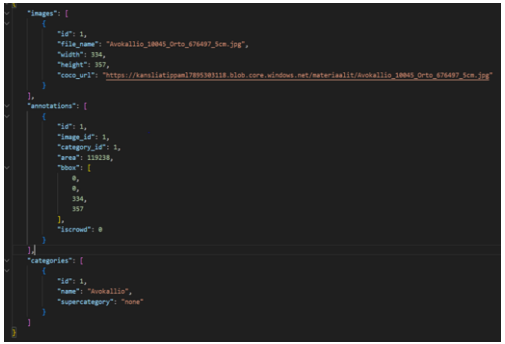
Helsinki

5. Kokeilun tuotokset

- **Python-algoritmi**
  - →ilmakuvien ja paikkatietojärjestelmän tietojen yhdistämiseksi ja optimoimiseksi hyvänlaatuista koulutusmateriaalin aikaan saamiseksi konenäölle
  - →koulutus- ja testi-datasettien jaotteluun
  - →tulostiedostojen muodostamiseen ja tulosten arviointiin tarvittavien tunnuslukujen laskemiseen

Esimerkki avokallio-tyyppisestä kohteesta

- Kuvien **annotointitiedostot** (COCO-tiedostot)
  - →voidaan käyttää myös muiden konenäkömallien kouluttamiseen

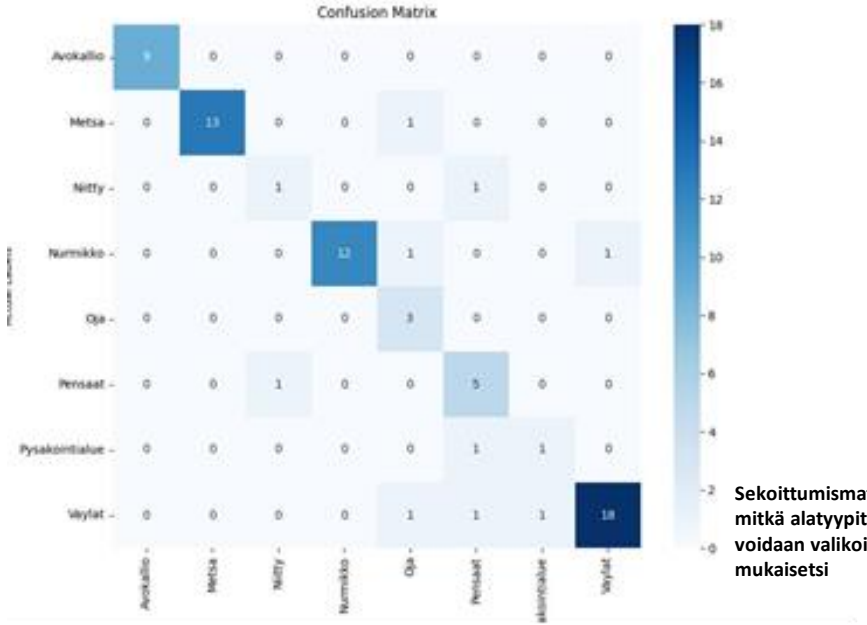


Esimerkki osasta COCO-tiedostoa

Helsinki

5. Kokeilun tuotokset

- Sekoittumismatriisi



Sekoittumismatriisin avulla voidaan tarkastella, mitkä alatyypit olivat mallille vaikeita, jolloin voidaan valikoida tulevaa koulutusdataa sen mukaisetsi

- Ehdotukset työmääräarvioineen toimenpiteistä mallin laadun parantamiseksi, laajentamiseksi ja tuotantokäytössä pyörivän sovelluksen rakentamiseksi

Helsinki

6. Opit kokeiltavan ratkaisun tai toimintatavan mahdollisuuksista

Projektissa saatiin selkeä näkemys koneoppimismallin mahdollisuuksista. Tekoälyn avulla on mahdollista kehittää toimiva omaisuustietorekisterin laadunvarmistamismenetelmä. Sen todettiin olevan:

- riittävän kyvykäs
- riittävän nopea
- riittävän luotettava

Käyttöönotto vaatii kehittämisvaiheen, jonka aikana mm. käyttöliittymä kehitetään ja koneoppimismalliin liittyvä sisältö (mitä asioita aluemaisten kohteiden lisäksi otetaan mukaan) määritellään.

KYMP toimialalta löytyy myös muita käyttötarkoituksia ja käyttäjäryhmiä, jotka voisivat hyötyä konenäkömallista yhdistettynä ilmakuvan ja karttatiedon poikkeavuuksien tunnistamisessa

Helsinki

7. Opit asiakkaiden tai palvelun käyttäjien tarpeista

- Projektin tavoitteena ei ole ollut tuottaa valmista loppukäyttäjiä eli rekisterin tietosisällöstä vastaavia henkilöitä palvelevaa ratkaisua esim. käyttöliittymää, vaan testata koneoppimismallia, soveltuisiko se rekisterin aluemaisissa kohteissa olevien poikkeavuuksien ja mahdollisten virheellisten kohteiden tunnistamiseen ja kannattaisiko menetelmästä jatkokehittää omaisuustieto-rekisterin sisällön laaduntarkastusmenetelmä, koska sellainen puuttuu.
- Mallia tulisi kokeilla myös viivamaisiin ja pistemäisiin kohteisiin ja selvittää, toimisiko konenäkömalli myös niissä olevien poikkeavuuksien ja puutteiden tunnistamisessa.
- Konenäkömalliin perustuva rekisterin sisällön laadun varmistaminen on menetelmänä nopea ja tehokas. Kun virheitä ei tarvitse etsiä, jää enemmän aikaa virheiden ja puutteiden korjaamiseen. Tästä hyötyvät kaikki karttapalveluaineiston käyttäjät.

Helsinki

8. Opit ratkaisun kehittämisestä teknisesti

A. Lähtöaineiston hyödynnettävyys

- Kokeilussa mukana ollut lähtöaineisto oli poikkeuksellisen laadukasta ja lähes kokonaan hyödynnettävissä mallin koulutukseen pienin kuvankäsittelyllisin muokkauksin.
- Aineisto oli etukäteen tarkistettua eli virheelliset kohteet oli tunnistetut, merkitty ja osalle virheellisistä kohteista kirjattu oikea alatyyppi. Manuaalisen tarkasteluun tulokset ja vihreelliseksi todettujen kohteiden oikea tieto olivat keskeisessä asemassa koneoppimismallin kouluttamista. Ilman kyseisiä tietoja kokeilu ei olisi onnistunut.
- Mikäli malli laajennetaan kattamaan lisää uusia pää- ja alatyyppeja aluemaisista kohteista tai muista kohteista esim. viivamaisista kohteista samankaltaista aineistoa tarvitaan lisää.
- Opittiin, että mallin kouluttamiseen riittää samasta alatyypista 20-50 kohdetta riittävän hyvään ja luotettavaan tunnistamiseen.
- Jos mallia halutaan laajentaa tunnistamaan uusia pää-/alatyyppejää tulee etsiä esimerkkejä nimenomaan kohteista joiden tyyppi on vielä aliedustettuna tai ei ole sisältynyt kokeilualueen aluemaisiin kohteisiin. Uusia isoja laajoja alueita, joissa on paljon samantyyppisiä kohteita, ei ole tarve uudestaan käydä läpi. Kokeilusta saatu aineisto on suoraan hyödynnettävissä mahdollisessa jatkoprojektissa

8. Opit ratkaisun kehittämisestä teknisesti

B. Teknisen ympäristön pystyttäminen

- **Rakennettaessa ympäristöä tämän kaltaiselle menetelmäkokeilulle, on huomioitava, että pilvipalvelun tekninen määrittely eli kaikki pilvipalvelun tarpeet eivät välttämättä ole lyötävissä lukkoon ennalta tai edes alkuvaiheessa. Tarpeet täydentyvät projektin aikana.**

- **Jos kokeilutoimintaa jatketaan niin, että pilvipalvelua tarvitsevat kokelut tehdään kaupungin omasa pilvipalveluympäristössä, tulee yhteistyön pilviresursseja hallinnoivan tahon kanssa olla mahdollisimman joustavaa. Näin vältyttäisiin tarpeettoman laajojen käyttöoikeuksien antamiselta (tietoturva) ja lopulta käyttämättä jäävien resurssien luomiselta (säästyvä aika).**

- **Eduksi olisi, että pilviympäristön ensimmäinen määrittely tapahtuisi mahdollisimman aikaisessa vaiheessa ennakoivasti.**

8. Opit ratkaisun kehittämisestä teknisesti

C. Tekoäly / koneoppiminen

Valittu konenäkömenetelmä (Azure AI Vision-palveluun omalla datalla koulutettava luokittelumalli) suoriutui tehtävästä jopa hieman odotuksia paremmin, kun kiinnitti seuraaviin asioihin huomiota jo koulutusdataa kootessa:

- Osa kohteista on hyvin pieniä tai asettuivat ilmakuvan reunalle, jolloin oli tarvetta joko venyttää kuva vaadittuun 50x50 pikseliä-kokoon tai hylätä kohde kokonaan opetusmateriaalin joukosta
- Osa alatyypeistä voi olla koneelle mahdoton tunnistaa. Esimerkiksi, kun mukana on hoitotapa, jota ei pysty kuvasta päättelemään. Kokeilussa samankaltaisia alatyyppejä yhdistettiin kokonaisuuksi. Yhdistäminen onnistui, koska liian erilaisten alatyyppien yhdistäminen olisi heikentänyt tunnistusta.
- Dataa tarvittaisiin eri vuodenajoilta esim. lehdettömältä ajalta, koska viheralueet näyttävät hyvinkin erilaisilta silloin. Kokeilussa käytettiin kahdelta eri vuodenajalta ja vuodelta peräisin olevia ilmakuvia.
- Mallia laajennettaessa tulee huomioida alatyyppien sisäinen variaatio ja esim. tarkoituksellisesti lisätä eri puu-/pensassukuja ja -lajeja sisältäviä kohteita.
- Ilmakuvissa on varjoisempia aluieita, joiden kohdalla kannattaa lisätä kirkkautta/kontrastia alatyypin tunnistamisen helpottamiseksi
- Malli ei vaadi yhden alatyypin kohteille minimimäärää. Käytäntö kuitenkin osoitti, että alatyypit, joista esimerkkejä oli vähemmän kuin 10, erottuivat heikommin. Yli 20 kuvan opetusmateriaalilla/alatyyppi tunnistus oli selvästi yli 90-prosenttista.

Helsinki

9. Opit kokeilemisesta yleensä

- **Kokeiluun liittyvät odotukset**
  Kokeilussa konkretisoitui, mitä kokeilun lopputuloksena on mahdollista saada aikaan → opittiin, että projektissa ei synny valmista heti käyttöön otettavaa menetelmää, mutta kokeilusta
  → saimme innostusta asian eteenpäin viemiseen
  → syntyi arvokasta tietoa tekoälyn/konenäkömallin mahdollisuuksista rekisterin laadullisessa tarkastelussa
  → löydettiin varteenotettavia uusia menetelmän käyttömahdollisuuksia
  → ymmärrys lisääntyi - kaikkea kehittämistä ei tarvitse tehdä kerralla vaan vaiheittain. Ensin keskitytään esim. rekisterissä suurimääräisiin kohteisiin, joiden oikeellisuudella on suuri vaikuttavuus toimintaan.

- **Mikä toimi hyvin, mitä jatkossa kannattaa tehdä samalla tavalla?**
  → Lähtöaineistona käytetty materiaali oli koneoppimismallin koulutuksessa toimivaa ja samantyyppisellä aineistolla kannattaa jatkaa
  → Hyödyntää mahdollisen laajemman lähtöaineiston muokkaamisessa projektissa käytettyjä menettelyjä ja saatua kokemusta

Helsinki

9. Opit kokeilemisesta yleensä

- **Mikä ei toiminut odotetusti? Mitä jatkossa kannattaa muuttaa tai tehdä toisin?**
  → Tekoälyä hyödyntävissä projekteissa pilvipalvelun pystyttäminen ei toiminut ja projektissa tarvittavaa olennaista työkalua jouduttiin odottamaan liian pitkään (koko lokakuu). Toteutusaika päättyi, eikä kaikkea saatavilla ollutta koulutusaineistoa päästy testaamaan (esim. viistoilmakuvia).
  Jos kokeiluja jatketaan kaupungin omasa pilvipalveluympäristössä, tulee yhteistyön pilviresursseja hallinnoivan tahon kanssa olla mahdollisimman joustavaa.

  → Koulutusmateriaaliin ORTO-ilmakuvien lisäksi muita kuvatyyppejä esim. viistoilmakuvia.
  → Tarvitaan eri vuodenaikana otettua ilmakuva-aineistoa (lehdettömään aikaan otettua) kohteiden tunnistamisessa sekä muuta saatavilla olevaa kuvamateriaalia (drone-kuvia).

- **Mitä opit muilta Kokeilukiihdyttämössä mukana olleilta tiimeiltä, kaupungin asiantuntijoilta tai valmentajilta?**
  → Muut projektit olivat kovin erilaisia vrt. oma, joten liittymäpintaa ei juuri ollut, mutta oli avartavaa seurata, kuinka projekteissa saatiin kehitettyä jopa käyttöön otettava työkalu.

Helsinki

10. Opit kokeiluprojektin arjen pyörittämisestä

- Työaikaa kokeiluun kului kaikkiaan 226 h syys-joulukuun aikana. Lähtöaineiston tuottamiseen (kokeilualueen läpikäyntiin manuaalisesti) kului työaikaa kesällä 2 vk, mikä ei ole laskelmassa mukana.

	SYYS	LOKA	MARRAS	JOULU	h YHT	tpv YHT
KOKEILIJAT	53	53	33	8	146	20
KUMPPANI	4	46	31	0	81	11
YHT:					**226**	**31**

Kokeilijalla projektissa 5 hengen työryhmä. Ajankäytöstä (146 h) projektipäällikön osuus 68 tuntia.

- Arjen pyörittämistä helpottivat mm. yhteistyönä kumppanin kanssa laadittu projektisuunnitelma (=yhteinen näkemys, mitä tehdään) sekä valmiiksi laaditut asialistapohja viikkokokouksille ja tuntiseurantapohja, viikkokokousaikojen ja muiden kokousaikojen kalenterointi koko projektiajalle.
- Kokeiluun varattu budjetti riitti esitettyjen tulosten aikaan saantiin. Lisäkoulutusmateriaalista ei ehditty kokeilla kaikkia (mm. viistoilmakuvat). Tulokset olisivat saattaneet vielä jopa parantua.
- Projektipäälliköllä tulee olla riittävästi aikaa projektin edistämiseksi. Lähtömateriaalin tuottamisessa todettiin, että toimiva tapa on keskittää tuottaminen yhdellä henkilölle tai työparille.

Helsinki

11. Kokeilun tekninen ympäristö

- Kokeilussa pyrittiin mahdollisimman kevyeen tekniseen ratkaisuun, jotta analyysille ja menetelmän hienosäädöille jäisi mahdollisimman paljon aikaa
- Kokeiluun liittyvä Python-koodi toteutettiin lokaalisti ja se on siirrettävissä pienin muokkauksin suoritettavaksi pilvipalvelun kontti-ratkaisuissa. Tätä ratkaisua käytettäisiin varsinaisessa sovelluksessa.
- Muokatut kuvat ja niihin liittyvät COCO-annotointitiedostot on siirretty containeriin Azure Blob Storageen (Azure AI Vision lukee kuvat vain tästä lähteestä)
- Konenäkö toteutettu Azure AI Vision -palvelussa. Etuna moniin vaihtoehtoisiin ratkaisuihin on koulutukseen tarvittavien kuvien pieni määrä. Mallia on esiopetettu hyvin laajasti jolloin muutaman esimerkkikohteen lisäys tuottaa jo oikean suuntaisia tuloksia. Heikkoutena palvelu ei tuota suoraan toiselle alustalle siirrettävää mallia
- Ulostulona tällä hetkellä lokaali Excel-tiedosto, lähinnä tulosten laadun arviointia varten. Tätä kohtaa on syytä pohtia tarkemmin lähestyttäessä käyttöönotettavaa sovellusta, esimerkiksi mahdolliset integraatiot muihin järjestelmiin.

12. Kokeilun data

Kokeilualueen rajaaminen/käytettävissä oleva data:

Kokeilu rajattiin tiukasti Kumpulan kaupunginosaan (rekisteritietoja päivitetty viimeksi 2010 luvulla) – selkeät maantieteelliset rajat isoissa kaduissa ja mukaan otettiin vain viheralueiden aluemaiset kohteet

+ Kompaktin kokoinen alue → manuaalinen tarkastelu oli mahdollista tehdä
+ Manuaalisessa tarkastelussa tunnistetut virheet toimivat arvokkaana koulutusaineistona.
+ Saatu hyvä tuntuma, mikä määrä koulutusdataa tarvitaan luotettavan tuloksen aikaan saamiseksi
+ Kokeiluun valittu sisältö eli yksi omaisuusrekisteritiedon kohdetyyppi, aluemaiset kohteet, oli riittävän laaja kokeilua ajatellen. Oli eduksi, että kokeilusta jätettiin pois viivamaiset ja pistemäiset kohteet.
- Koulutusdata jäi joidenkin kohteiden osalta riittämättömäksi ja jotkut aluemaiset kohteet pois kokonaan
- Kaikkea koulutusdataa ei ehditty testata esim. viistoilmakuvia

Lähtödatan muokkaus ja käsittely - näkemys toimivasta lähtödatasta koneoppimismallin kouluttamiseen syntyi

Kokeilija: Rekisteritiedot shapefile muodossa, riittämättömän (3-5/kohde) rekisterilähtöaineiston karsittiin pois, virheellisten kohteiden oikea pääluokka tai alaluokka lisättiin dataan, lisätietona jos aluemainen kohde ei ole KYMP vastuulla, ORTO-ilmakuvista eri muotojen hyödyntäminen (täyslehtisyys/lehdetön aika)

Kumppani: lähtödataa käsitelty niin, että konenäkömallin koulutus datalla mahdollista mm.

- Shapefile kuvien ja aluetyyppien, virheellisyystiedon ja aluetta rajaavien koordinaattien yhdistäminen
- Alatyyppien yhdistäminen (esim. eri niittytyypit)
- Kuvien muunto Azuren pikseli-vaatimusten mukaisiksi, kontrasteja/valoisuutta muokattu
- Virheellisten ja virheettömien kuvien jaottelu omiin kansioihin
- Jako opetus- ja testidataan (8 opetus- ja testidataryhmää)

13. Jatkopäätökset ja -ideat

- Projektin jatkamisesta ja menetelmän tuotantokäyttöön kehittämisestä ei ole päätetty, mutta omaisuustieto ja laatu tiimi tullee esittämään projektin jatkamista.
- Jatkotyö rahoitetaan yksikön kehittämistyöhön varatuista määrärahoista tai toimialan tekoälyn käytön edistämiseen varatuista määrärahoista, joiden käyttömahdollisuus selvitetään.
- Jatkoprojekti edellyttää, että jatkokehittäminen projektoidaan eli laaditaan tavoitteet, määritetään laajuus (mm. mukaan otettavan koulutusmateriaalin laajuus), nimetään henkilöt ja varmistetaan kehittämistyöhön määrärahat. Menetelmän käyttöönotto vaatii käyttösovelluksen kehittämisen.
- Tuotantokelpoisen ratkaisun rakentaminen edellyttää, että on selkeä näkemys käyttötavasta – käydäänkö omaisuustietorekisterin paikkansapitävyys läpi esim. vuosittain kertaluonteisena ajona, vai tarkastellaanko alueita muiden töiden ohella jatkuvaluonteisesti sekä selvittää mahdollisten integraatioiden tarve muihin järjestelmiin.

- Tekoälypohjaisen tarkastelumenetelmän etu on sen **nopeus / tehokkuus.**

  Kokeilualue käsitti n. 1 % rekisterissä olevista viheralueiden aluemaisista kohteista. Kumpulan alueella aluemaisia kohteita oli kaikkiaan 1 168 kpl. Tästä voidaan päätellä, että rekisterissä on kaikkiaan n. 116 800 viheralueiden aluemaista kohdetta.
  Koulutettu konenäkömalli käsittelee ja tunnistaa kohteet n. 2s/kohde. Tästä voidaan päätellä, että kaikkien viheralueiden kohteiden aluemaisten käsittelyyn kuluu aikaa 58 500 s → 973 min → 16 h →n. 2 htp.

  Kumpulan alueen viheralueiden aluemaisten kohteiden manuaaliseen läpikäyntiin kului työaikaa n. 2 vk.

Helsinki

13. Jatkopäätökset ja -ideat

Projektin aikana syntyneet ideat menetelmän muista käyttömahdollisuuksista:

Aluemaiset kohteet:

→ Työmaana olevien kohteiden tunnistaminen ja tietojen korjaaminen – digitointiavustajana käyttö
→ Tiemerkintöihin (ajoratamaalaukset/-merkinnät) liittyvien puutteiden tunnistaminen
→ Alueiden käyttö – tunnistaa/havaita vuokra-alueiden laajeneminen yli vuokra-alueen
→ Alueiden käyttö – havaita tonttien laajeneminen yli vuokra-alueen

Viivamaiset kohteet:

→ Rekisteristä puuttuvien tietojen täydentäminen esim. reunakivet (graniitti, betoni ja asvalttimakkara)
→ Hidastetöyssyjen tunnistaminen

Pistemäiset kohteet:

→ Tiettyjen haluttujen puiden tunnistaminen ilmakuvasta
→ Puistopuiden paikkojen saanti rekisteriin