

# RAKENNUSTEN KÄYTÖN ENNUSTAMINEN KONEOPPIMISELLA HYÖDYNTÄEN OLEMASSA OLEVAA MITTAUSDATAA – HANKKEEN LOPPUSELVITYS 1.6.2018-31.1.2019

## Tavoitteiden toteutuminen

Hankkeessa tavoite oli tehdä kustannustehokas menetelmä tilojen käytön ennustamiseksi. Hankkeessa testattiin olemassa olevan datan soveltuvuutta käytön ennustamiseksi, sekä ennustemallien tarkkuutta. Hanke jakautui vaiheisiin seuraavasti:

- State-of-art selvitys tilojen käytön estimointi ja –ennustemenetelmistä. Kartoitettiin koneoppimismenetelmiä estimoida tilojen käyttöä käyttäen olemassa olevaa mittausdataa sekä kartoitettiin soveltuvia aikasarjaennustusmalleja.
- Tutkimusdatan keruu.
- Potentiaalisten menetelmien testaus käytännössä ja soveltuvuuden arviointi.

Hanke toteutui budjetin mukaisesti ja suurimmilta osin suunnitelmien mukaisesti. Alustavasti toimivia menetelmiä rakennusten käytön estimointiin ja ennustamiseen löydettiin. Menetelmiä testattiin kerätyllä datalla ja ne vaikuttivat alustavasti toimivilta. Käyttökelpoista tutkimusdataa saatiin alkuperäisestä suunnitelmaa vähemmän ja sen tarkkuuteen liittyi epävarmuustekijöitä. Näin ollen menetelmien toimivuuteen liittyi epävarmuustekijöitä, jotka johtuvat kokeiluhankkeessa käytetyn datasetin laatuongelmista. Selvitetystä menetelmästä löydettiin lisäksi lukuisia kehitysmahdollisuuksia.

## Hankkeen tulokset

Hankkeen aikana saavutettiin alan state-of-the-art tietämys. Hankkeessa testattiin olemassa olevia estimointimalleja rakennusten käytön arviointiin sekä testattiin näistä muutamia aikasarjaennustemalleja. Tärkeimmät tulokset ja käytetyt menetelmät on esitetty liitteenä olevassa tulosraportissa. Tärkeimmät alustavat tulokset esiteltiin esittelytilaisuudessa. Esittelytilaisuuden kalvot julkaistaan tutkimusryhmän internet sivuilla, kun organisaation internetsivujen päivitysprojekti on valmis. Esityskalvot ovat tämän raportin liitteenä. Hankkeen tulokset eivät itsessään ylitä tieteellisen julkaisemisen kynnystä johtuen käytetyn datan epävarmuudesta, vaan vaativat jatkotutkimusta. Tuloksia julkaistaan tarkemmin tutkimusryhmän internetsivustolla, kun jatkotutkimukseen liittyvät tieteelliset artikkelit on julkaistu.

## Tapahtumat ja viestintä

Hanke on ollut esillä seuraavissa tapahtumissa

- 9.5.2018 Kira-digi kokeiluhankkeiden Kick-off
- 22.1.2019, tulosten esittely, Älykkään talotekniikan pilottikohde loppuseminaari
- 6.2.2019, hankkeen esittely, KNX Finland hallituksen kokous
- 7.2.2019, hankkeen esittely, Granlund Oy

Hankkeelle on perustettu omat internet-sivut. Esittelytilaisuuksien markkinoinnissa on hyödynnetty älyrakennusosalalla toimivia yhteistyökumppaneita sekä heidän verkostojaan, esim:

- KNX-Finland
- Aalto sustainability hub
- Aallon viestintä
- VTT Renovation Leap -hankkeen viestintä

## Tuloksien hyödynnettävyys, haasteet ja jatkotoimet

Alan toimijat saivat hankkeessa viimeisintä tietoa rakennusten käytön estimoinnista ja ennustamisesta käyttäen olemassa olevaa rakennusautomaatiosta saatavaa mittausdataa. Tuloksia on esitelty alan tilaisuuksissa sekä alan organisaatioille. Tilaisuuksissa hanke herätti mielenkiintoa alan toimijoiden keskuudessa.

Kokeillulla ennustusmenetelmällä on selkeää potentiaalia ja käyttökohteita tulevaisuudessa. Jatkotutkimusta myös tarvitaan, jotta estimointimenetelmien yleistettävyyttä rakennusten välillä pystytään parantaa, jotta dataa ennustemallien kehittämiseen saadaan kerättyä kustannustehokkaasti. Tällä hetkellä haasteita asettaa myös rakennusautomaatio datan saatavuus. Hankkeessa havaittiin merkittäviä haasteita datan keruuseen ja datan luotettavuuteen liittyen.

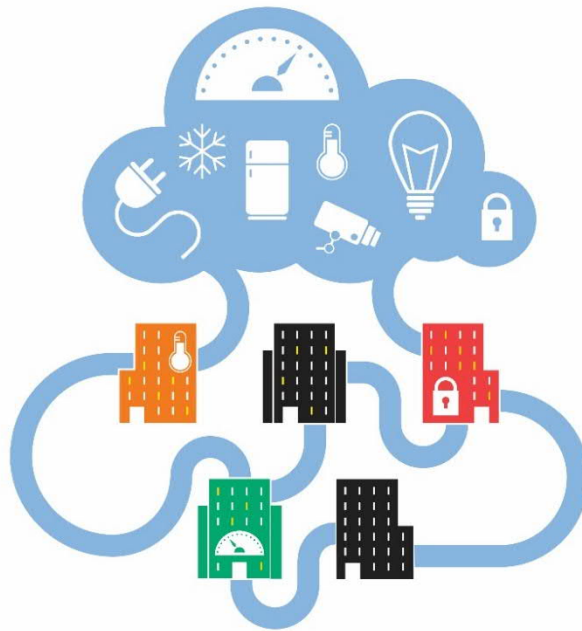
Hankkeessa saatavien tulosten perusteella tehdään jatkotutkimusta rakennusten käytön estimointimenetelmien yleistettävyyden parantamiseksi rakennusten välillä. Lisäksi ennustemallien luotettavuutta testataan suuremmalla määrällä tutkimuskohteita, jotta voidaan varmistua menetelmän toiminnasta tai kehittää parempia menetelmiä. Jatkotutkimuksista julkaistaan kaksi tieteellistä artikkelia myöhemmin kuluvan vuoden aikana.

Liitteet:

1. Rakennusten käytön ennustaminen koneoppimisella hyödyntäen olemassa olevaa mittausdataa: Kokeiluhankkeen tulosraportti
2. Esitystilaisuuden kalvot
3. Esitystilaisuuden ohjelma ja kutsu

LIITE 1

# Rakennusten käytön ennustaminen koneoppimisella hyödyntäen olemassa olevaa mittausdataa: Kokeiluhankkeen tulosraportti



Petri Jantunen, Tohtorikoulutettava

## Sisällysluettelo

1 Johdanto .....	3
2 Ihmismäärien mittaus, estimointi ja ennustaminen .....	3
3 Menetelmät .....	4
4 Tulokset .....	8
5 Yhteenveto.....	10
6 Lähteet.....	11

## 1 Johdanto

Suomalaisten toimistorakennusten käyttöasteet ja tilojen käyttötottumukset vaihtelevat huomattavasti. Lisäksi rakennuksen käyttö jakautuu usein epätasaisesti tilojen välillä. Käyttöastetietoa eikä sen ennustetta ole kuitenkaan yleensä saatavilla. Käytöllä on merkittävä vaikutus mm. rakennuksen käyttäjien viihtyvyyteen, sisäilmaolosuhteisiin, ylläpitokuluihin, energiankulutukseen ja laitteiden toimintaan. Keräämällä tietoa rakennusten ja niiden tilojen käytöstä sekä ennustamalla tulevaisuuden käyttöä, voidaan muun muassa ymmärtää paremmin käyttäjien tarpeita, tehostaa tilojen käyttöä, allokoida ylläpidon resursseja tehokkaammin, parantaa käyttäjien tyytyväisyyttä.

Kaikkein moderneimmissa rakennuksissa saattaa olla erillinen järjestelmä käyttäjien seuraamiseksi tai tilojen käytön mittaamiseksi. Nämä järjestelmät perustuvat usein esimerkiksi kameroihin ja konekääköön, käyttäjien mukanaan kantamiin tunnisteesiin tai matkapuhelimien laskemiseen. Nämä järjestelmät ovat kuitenkin harvinaisia ja usein varsin kalliita toteuttaa. Niihin usein liittyy myös käyttäjien yksityisyyden suojaan liittyviä ongelmia.

Useimmissa rakennuksissa on kuitenkin lukuisia antureita ja toimilaitteita, jotka mitaavat tilojen olosuhteita sekä vaikuttavat niihin. Useiden tutkimusten perusteella, näiden tuottaman datan avulla voidaan estimoida tilan käyttöä koneoppimista hyödyntäen. Tilojen käytön ennustamista on kuitenkin tutkittu vähän. Tilojen käytön arvioon perusteella voidaan ainakin teoriassa edelleen pyrkiä ennustamaan tulevaisuuden käyttöä käyttöhistorian ja muiden muuttujien avulla hyödyntäen koneoppimista ja tilastollisia menetelmiä, sillä tilojen käyttö rakennuksissa ei ole täysin satunnaista.

Hankkeessa testataan olemassa olevan rakennusautomaatiodatan hyödyntämistä tilojen käytön ennustamisessa sekä arvioidaan ennustemenetelmien tarkkuutta. Hankkeen tarkoitus on toimiva ratkaisu, eikä niinkään vertailla eri menetelmien tarkkuutta toisiinsa ja löytää absoluuttisesti parasta menetelmää kaikista mahdollisista.

## 2 Ihmismäärien mittaus, estimointi ja ennustaminen

Rakennusten ihmismäärien arviointiin on lukuisia eri menetelmiä. Osa menetelmistä perustuu läsnäolon tai ihmismäärien suoraan mittaamiseen, esimerkiksi läsnäolotunnistimien tai kameroiden avulla. Osa menetelmistä perustuu epäsuoraan arviointiin, esimerkiksi sisäilmaolosuhteiden muutoksista. Tässä tutkimuksessa keskityttiin ihmismäärien estimointiin epäsuorasti käyttäen yleensä rakennusautomaatiosta saatavilla olevaa dataa.

Epäsuoria mittausmenetelmät voivat perustua, esimerkiksi tilan hiilidioksidipitoisuus-, suhteellisen kosteus- ja/tai lämpötila-aikasarjojen analysointiin [1], käyttäjälaitteiden sähkönkulutuksen analysointiin [2]. Tarpeenmukaisesti ohjatuissa tiloissa ilmanvaihto ja lämmitysjärjestelmät vaikuttavat myös sisäilman olosuhteisiin, joten yhdistämällä tieto järjestelmien toiminnasta, voidaan malliin tuoda lisäinformaatiota, jonka perustella ainakin teoriassa on mahdollista päästä tarkempiin tuloksiin.

Estimontimallin laatimiseen on lukuisia eri menetelmiä. Tilan ihmismäärä ja läsnäolo ovat luonteeltaan diskreettejä muuttujia, joten tässä hankkeessa testattiin luokittelijoita, jotka päättelevät datan perusteella tilassa vallitsevan ihmismäärän tai läsnäolon. Luokitteluun on lukuisia eri koneoppimis- ja tilastollisia menetelmiä kuten neuroverkot, logistinen regressio, tukivektorikone (support vector machine) ja diskriminanttianalyysi. [3]

Aikasarjojen ennustamiseen on kehitetty useita eri menetelmiä. Yleisimpiä sovelluskohteita on esimerkiksi taloudellisten muuttujien ennustaminen. Niiden voidaan ajatella jakautuvan lineaarisiin, esimerkiksi autoregressiivinen malli, liukuvan keskiarvon malli tai niiden yhdistelmä, joko ulkoisilla muuttujilla tai ilman ja epälineaarisiin menetelmiin, esimerkiksi epälineaarinen autoregressiivinen malli ulkoisilla muuttujilla tai ilman. Näiden lisäksi on esimerkiksi neuroverkkoihin perustuvia malleja. Epälineaarinen autoregressio voidaan toteuttaa myös neuroverkoin. [4] ja [5]

### 3 Menetelmät

#### Datan keruu

Hankkeessa kerättiin dataa neljästä eri rakennuksesta. Eri rakennuksista saatavilla ollut data vaihteli niin ajallisessa ulottuvuudessa kuin datalähteiden osalta. Ensimmäisestä rakennuksesta saatavilla ollut data osoittautui epäluotettavaksi, sillä ihmismäärien keräämiseen liittyvä mittaustapa kattoi alle puolet rakennuksen käyttäjistä.

Toiseen rakennukseen asennettiin kamerat kolmeen huoneeseen kulkuaukon yläpuolelle, mitkä laskivat tilaan tulevia ja lähteviä ihmisiä. Lisäksi kerättiin saman aikaista dataa rakennuksen rakennusautomaatiojärjestelmästä. Kahdessa tilassa valaistusta ohjattiin läsnäoloantureiden perusteella, mitä tietoa ei kuitenkaan ollut rakennusautomaatiosta saatavilla. Valaistuksen muutokset aiheuttivat virheitä ihmismäärädataan, joten nämä kaksi tilaa tuli rajata hankkeen ulkopuolelle. Kolmannessa tilassa valaistus toimi kytkimen avulla ja se aiheutti niin ikään virheitä. Virheiden lukumäärä oli kuitenkin päivätasolla huomattavasti pienempi ja niitä pystyttiin korjaamaan jälkikäteen.

Kolmannen ja neljännen rakennuksen dataa oli saatavilla alle kolmen kuukauden ajalta ja ne päätettiin rajata riittämättöminä hankkeen ulkopuolelle. Kolmannessa rakennuksessa oli saatavilla ainoastaan tilojen läsnäolotieto. Neljännessä rakennuksessa oli käytettävissä läsnäolotiedon lisäksi myös LVIS järjestelmistä saatavat sisäilmanlaatu- sekä laitteiden toimintaa kuvaavat mittaukset.

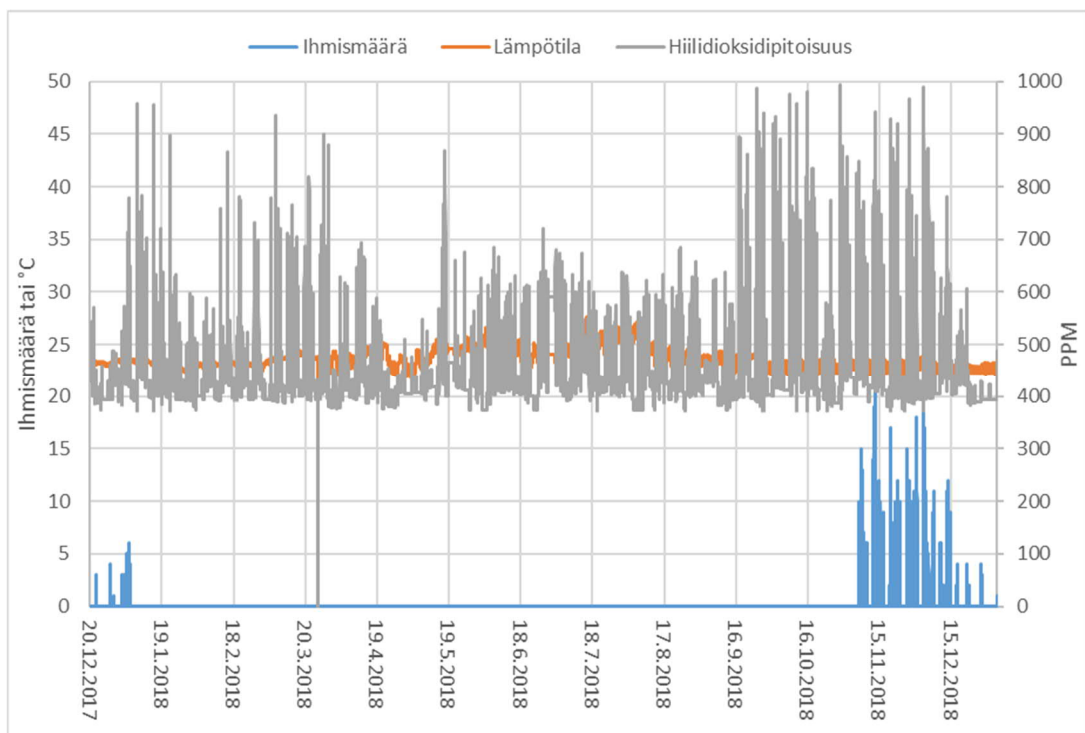
Toiseen rakennukseen asennettu ihmismääriä mittaava järjestelmä toteutettiin kameroiden avulla. [6] Järjestelmä seurasi kuvassa näkyviä ihmisiä. Järjestelmä laski ihmiset ja suunnat, joka ylitti kuvassa määritetyn alueen ja laski niiden virran ja suunnan asetetun virtuaalisen alueen yli. Ihmismäärät laskettiin näiden ihmisvirtojen kumulatiivisena summana. Yksityisyyden suojan varmistamiseksi, järjestelmä prosessoivat videokuvaa paikallisesti, ja tallensi ainoastaan ihmismäärä tapahtumatiedot. Tapahtumatiedot tallennettiin tutkimusryhmän palvelimelle.

Kamerajärjestelmän tarkkuus testattaessa vakiovalaistusolosuhteissa oli noin 95 %. Järjestelmän tarkkuudessa ei havaittu muutoksia, vaikka tilan valaistus oli kytketty pois päältä. Epätarkkuus kasvoi hetkellisesti, noin sekunnin ajaksi, kun valaistus kytkettiin päälle tai pois. Valaistuksen muutokset aiheuttivat häiriöitä hahmon tunnistukseen. Valaistuksen muuttuessa järjestelmä saattoi rekisteröidä yhden ihmisen

useana tai kirjasi monta tapahtumaa yhden sijasta. Valaistuksen muutos ei kuitenkaan pelkästään aiheuttanut tapahtumien kirjauksia, mikäli kuvassa ei ollut liikettä samanaikaisesti.

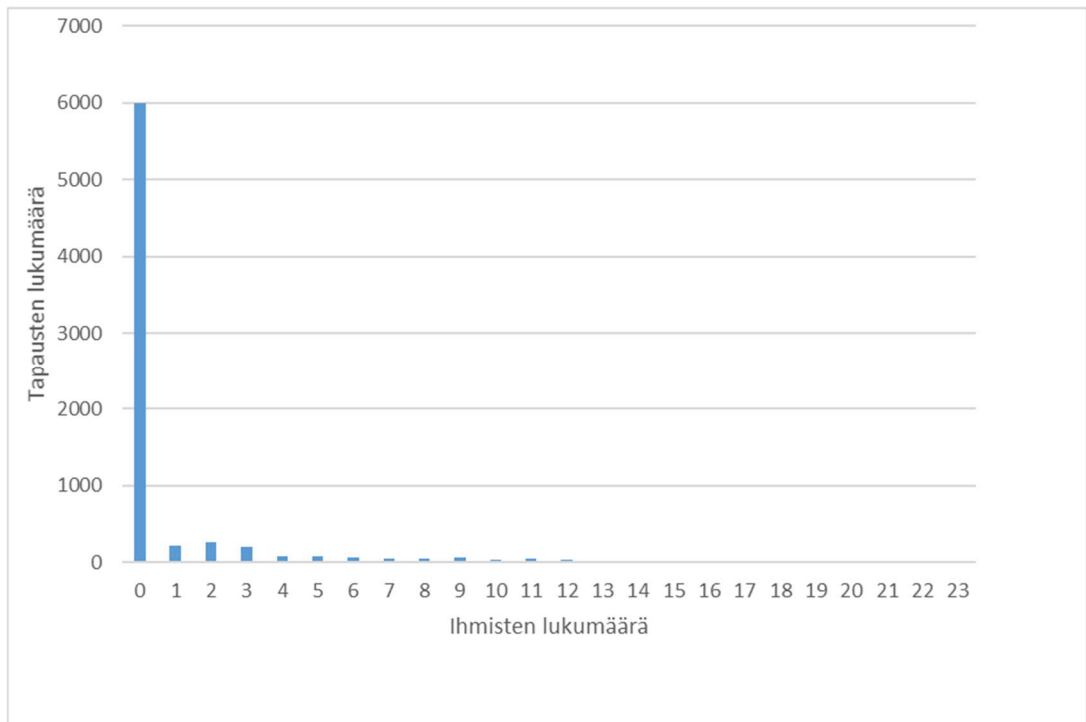
Kamerajärjestelmän virhetilanteet pystyttiin kartoittamaan etsimällä tiheästi esiintyneitä tapahtumia. Mikäli sekunnin sisällä tallennettiin useampi tapahtuma, ryhmiteltiin tapahtuma yhdeksi tapahtumaksi. Tämä perustui oletukseen, että todennäköisyys sille, että tavanomaisesta oviaukosta kulkee kaksi henkilö sekunnin aikana huomattavasti pienempi verrattuna virheen mahdollisuuteen. Tämän jälkeen jokaiselle ryhmitellylle tapahtumalle etsittiin nolasta poikkeava arvo, siten että ihmismäärä on päivän viimeisen tapahtuman jälkeen nolla ja että ihmismäärä on aina epänegatiivinen. Menetelmällä saatuihin ihmismääriin sisältyy korjauksen jälkeen virheen mahdollisuus, mutta virheen suuruus on näin huomattavasti pienempi ja perustuu parhaaseen arvioon. Ihmismäärädataa oli käytössä yhteensä kahden ja puolen kuukauden ajalta kahdessa eri mittausperiodissa. Ensimmäinen mittausperiodi oli 20.12.2017 – 5.1.2018. Toinen mittausperiodi ajoittui välille 7.11.2018 – 3.1.2019

Edellä kuvatun ihmismäärädatan lisäksi saatavissa oli rakennusautomaatiosta kyseisen tilan lämpötila ja hiilidioksidipitoisuus sekä rakennuksen kaikkien ilmanvaihtokoneiden poistoilman lämpötila sekä käyntiä kuvaava prosentuaalinen arvo. Edellä mainittua dataa saatiin käyttöön välillä 7.11.2018 – 3.1.2019. Rakennuksen ilmanvaihtokoneista ei ollut käytettävissä tarkempia tietoja, joten varmuutta tilaa palvelevasta ilmanvaihtokoneesta ei saatu. Rakennuksen kolmen ilmanvaihtokoneen datojen havaittiin korreloivan selkeästi enemmän tilan olosuhteiden kanssa. Näin ollen näiden kolmen ilmanvaihtokoneen mittaukset otettiin mukaan tarkasteluun ja loput jätettiin ulkopuolelle. Datalähteiden alkuperäinen näytteenottoväli vaihteli välillä 10...20 min ihmismääriä lukuun ottamatta, jossa ainoastaan tapahtumat kirjattiin. Kaikkien datalähteiden näytteenotto uudistettiin (resampling) keskiarvoistamalla neljään näytteeseen tunnissa. Ihmismäärät, lämpötila ja hiilidioksidipitoisuus on esitetty kuvassa 1.



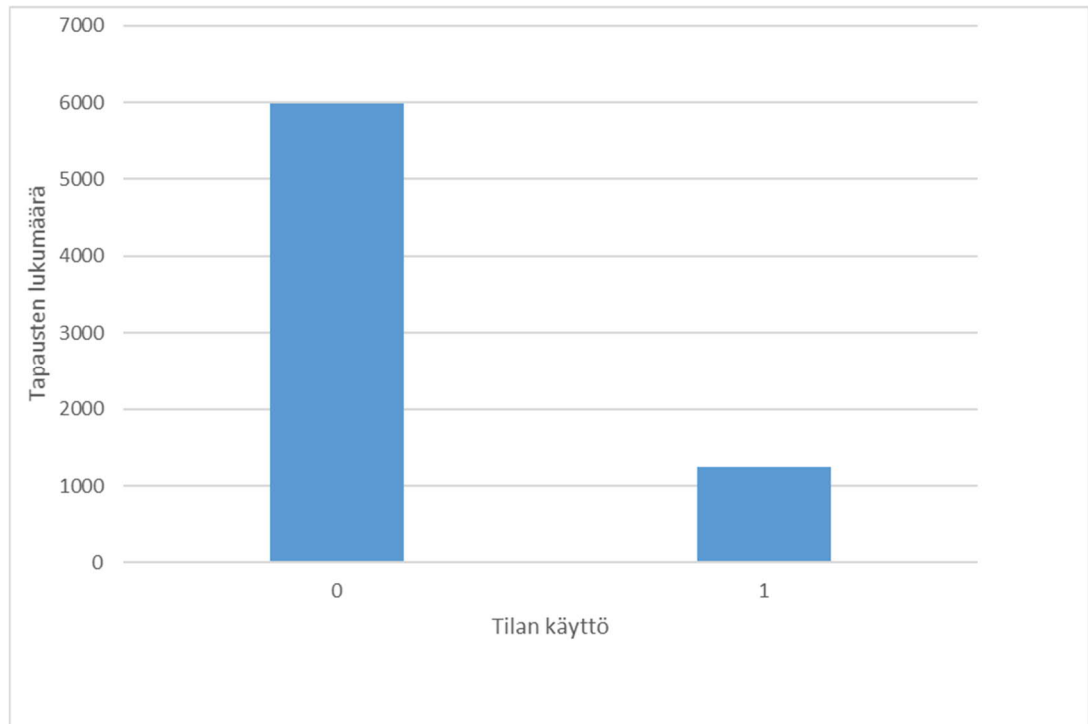
Kuva 1. Tilan ihmismäärät, lämpötila ja hiilidioksidipitoisuus datan keräysjaksolla.

Seurantajaksolla tilan korjatut ihmismäärät vaihtelivat välillä 0...23. Suurimmassa osassa mittauspisteitä tila ei ollut käytössä. Käytön aikana ihmismäärän vaihtelu oli suurta ja datapisteiden lukumäärä tietyllä ihmismäärällä oli hyvin pieni. Kuvassa 2 on esitetty ihmismäärien histogrammi datapisteittäin mittausajanjaksolla. Pienestä datapisteiden lukumäärästä johtuen ei ihmismäärien estimointimallin opettaminen onnistunut, siten että oltaisiin saatu tarkkoja tuloksia. Näin siitä huolimatta, vaikka opetuksessa annettiin suurempi painokerroin virheille, jotka tapahtuivat tilan ollessa käytössä. Tästä johtuen hankkeessa keskityttiin estimoimaan ainoastaan, että onko tila käytössä vai ei. Valitulla menetelmällä saatiin riittävä määrä opetusdataa molempiin luokkiin. Kuvassa 3 on esitetty läsnäolon histogrammi.



Kuva 2. Tilan käyttäjämäärien histogrammi. Y-akselilla datapisteiden lukumäärä. X-akselilla datapisteiden ihmismäärä.





Kuva 3 Tilan käytön histogrammi. Y-akselilla datapisteiden lukumäärä. X-akselilla tilan läsnäolo. Nolla edustaa tilannetta, jossa tila ei ole käytössä. Yksi edustaa tilannetta, jossa tila on käytössä.

Kerätty data jaettiin opetus, validointi ja testausdataan. Opetusdataa käytettiin neuroverkkomallin painokertoimien etsimisessä. Validointidataa eri neuroverkkoarkkitehtuurien vertailuun. Testidatan avulla arvioitiin mallin tarkkuutta. Testidata valittiin satunnaiselta kahdelta peräkkäiseltä viikolta mittausajanjaksolta. Validointidata poimittiin satunnaiselta viikolta. Loput käytettiin opettamiseen.

### Estimointimenetelmät

Testattaviksi estimointimenetelmiksi valittiin neuroverkot, sillä niiden on havaittu tutkimuksissa saavuttavan hyviä tuloksia käyttäjämäärien estimoinnissa. Lisäksi testattiin yleisesti luokitteluongelmissa käytettyä logistista regressiota.

Estimoinnissa käytettiin luokittelevaa neuroverkkoa. Logistista regressiolla ei saavutettu mielekkäitä tuloksia. Eri muuttujien polynomiyhdistelmiä ei testattu, sillä mallin parametrien määrä olisi kasvanut liian suureksi jo käytettäessä toisen asteen polynomeja. Datasta johtuen hankkeessa lopulta keskityttiin ainoastaan neuroverkkoestimointimalleihin, joka luokitteli tilan läsnäoloa.

Neuroverkko arkkitehtuuri valittiin opettamalla useita erilaisia arkkitehtuureja ja valitsemalla näistä neuroverkko, jonka sisäisen tarkkuuden (precision) ja herkkyden (recall) harmoninen keskiarvo oli suurin ( $F_1$  score).

Malleja testaattiin eri opetusdatalähteiden yhdistelmillä. Testatut yhdistelmät on esitetty taulukossa 1.

Taulukko 1. Taulukossa on esitetty opetuksessa käytetyt syötteet eri tapauksissa. CO<sub>2</sub> tarkoittaa hiilidioksidipitoisuutta, T lämpötilaa, RAU rakennusautomaatiodataa, d muuttujan edessä tarkoittaa aikaderivaatta, alaindeksi t:t-4 tarkoittaa aikasarjan pituutta (tässä tapauksessa nykyinen hetki ja neljä edellistä).

Tapaus	Syötteet	Tulosteet
1	CO <sub>2,t:t-4</sub> + dCO <sub>2,t:t-4</sub>	Läsnäolo tai ihmismäärä
2	T <sub>t:t-4</sub> + dT <sub>t:t-4</sub>	Läsnäolo tai ihmismäärä
3	CO <sub>2,t:t-4</sub> + dCO <sub>2,t:t-4</sub> + RAU <sub>t:t-4</sub>	Läsnäolo tai ihmismäärä
4	T <sub>t:t-4</sub> + dT <sub>t:t-4</sub> + RAU <sub>t:t-4</sub>	Läsnäolo tai ihmismäärä
5	CO <sub>2,t:t-4</sub> + dCO <sub>2,t:t-4</sub> + T <sub>t:t-4</sub> + dT <sub>t:t-4</sub> + RAU <sub>t:t-4</sub>	Läsnäolo tai ihmismäärä

### Ennustusmenetelmät

Ennustusmenetelmien opettamisessa ja testauksessa käytettiin tilan estimoitua läsnäoloa. Estimoitu läsnäolo tieto kattoi koko datankeruu aikavälin. Malleissa, joissa ulkoisia muuttujia käytettiin, ulkoisia muuttujina käytettiin kellonaikaa ja viikonpäivää.

Aikasarjaennustusmenetelmistä testattiin lineaarisia ja epälineaarisia menetelmiä ilman ulkoisia muuttujia ja niiden kanssa. Epälineaarisilla menetelmillä ei onnistuttu saamaan ennustuskykyisiä malleja. Epälineaarisista malleista parhaat tulokset saavutettiin epälineaarisella autoregressiivisellä mallilla ulkoisilla muuttujilla (NARX). Epälineaarisessa autoregressiivisessä mallissa käytettiin neuroverkkoja. Edellisten ajanhetkien lukumäärä oli 4 – 144, mikä vastasi 1-36h edellisiä arvoja. Optimaalinen aikaikkuna etsittiin kokeilemalla eri vaihtoehtoja. Toimiva neuroverkkoarkkitehtuuri etsittiin niin ikään kokeilemalla eri vaihtoehtoja.

## 4 Tulokset

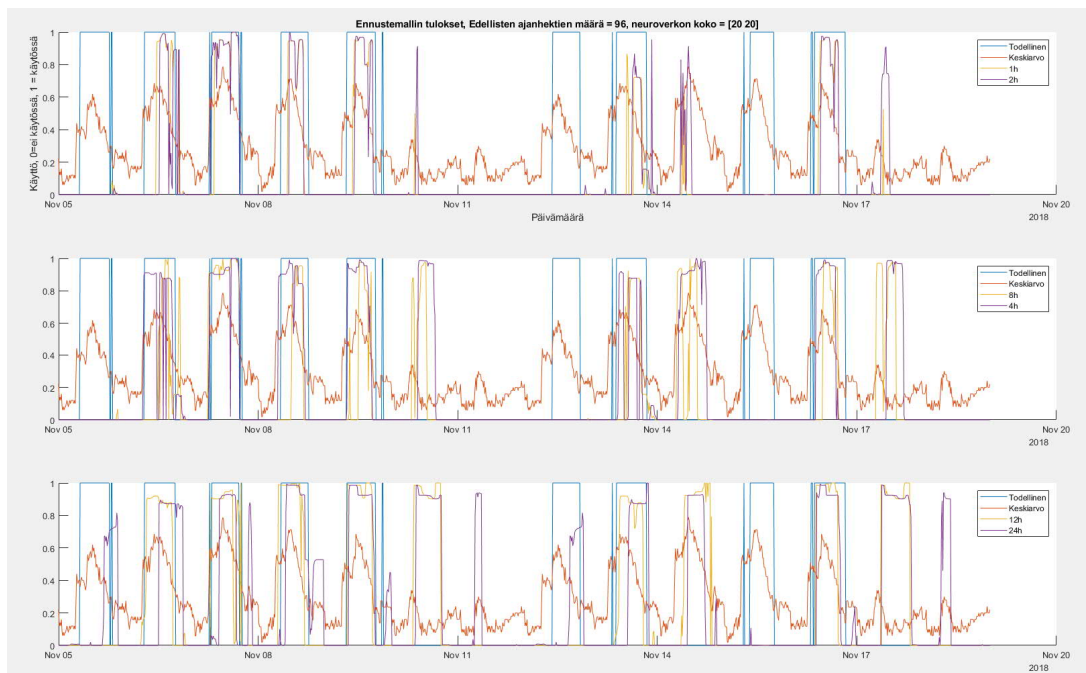
Paras estimointimallin tarkkuus saavutettiin, kun käytettiin taulukon 1 tapauksen 5 mukaisia syötteitä. Mallin syötteiden lukumäärä näin ollen oli 26. Paras tulos saavutettiin neuroverkkomallilla, jossa oli kaksi piilotettua tasoa (hidden layer), joissa molemmissa oli 26 neuronaa. Mallissa oli yksi ulostulo. Mallin antamat tulokset testidatalla on esitetty taulukossa 2. Mallin tarkkuus oli 90,4 %. Mallin sisäinen tarkkuus on 93,0 % ja herkkyys on 94,6 %.

Taulukko 2. Estimointimallin tulokset testidatalla.

Mallin ulostulo	Todellinen läsnäolo		
	0	1	
0 (ei käytössä)	347	26	93.0 %
	72.3 %	5.4 %	7.0 %
1 (käytössä)	20	87	81.3 %
	4.2 %	18.1 %	18.7 %
	94.6 %	77.0 %	90.4 %
	5.4 %	23.0 %	9.6 %

Ennustumalleissa parhaan tuloksen tuotti keskimäärin neuroverkko, jossa oli kaksi piilotettua neuronikerrosta, joissa molemmissa oli 20 neuronia. Piilotetuissa ja ulos-tulokerroksissa käytettiin softmax aktivointifunktiota. Parhaat tulokset saavutettiin edellisten ajanhetkien aikaikkunalla, joka huomioi tilan käytön 24 tunnin ajalta. Lisäksi malli sai ulkoisina syöteinä viikon päivän ja kellonajan numeerisessa muodossa. Tulosten havaittiin riippuvan huomattavasti neuroverkon rakenteesta. Lyhyillä ennustehorisonteilla yksinkertaiset mallit tuottivat parempia tuloksia, kun taas pitkillä ennustehorisonteilla monimutkaisemmilla malleilla saavutettiin parempia tuloksia.

Mallin tarkkuus on parempi lyhyillä kuin pitkillä ennustushorisonteilla. Ennustushorisontin pidentyessä mallin epätarkkuus kasvaa. Pidemmällä ennustushorisonteilla ennusteet lähestyvät viikonpäivän ja ajan mukaista keskiarvoa. Malli ei kykene ennustamaan testidatassa esiintynyttä päivää keskellä viikkoa, milloin tila ei ole käytössä. Tämä käyttämätön päivä myös vaikutti seuraavan päivän ennusteeseen, siten että malli piti epätodennäköisempänä tilan käyttöä tuolloin. Tämä johtuu todennäköisesti siitä, että tilan keskimääräinen käyttö arkipäivisin on hyvin saman kaltaista, eikä opetusdatassa esiintynyt vastaavia tilanteita. Lisäksi ulkoiset syötteet saattavat aiheuttaa mallin ylikorostamaan viikonpäivän ja kellonajan osuutta ennustuksessa. Toisaalta taas on epärealistista odottaa, että malli tuottaisi osaisi päivää etukäteen ennustaa, että tila ei ole käytössä, mikäli sen syöteenä ei ole muuta informaatiota, kuin edellisen 24 tunnin käyttö, kellonaika ja viikonpäivä. Tulokset on esitetty kuvassa 4.



Kuva 4. Tulokset validointi datalla. Ennustemallin tulokset (violetti ja keltainen käyrä). Ylimmällä rivillä on esitetty 1 ja 2 tunnin ennusteet. Keskimmäisellä rivillä on esitetty 4 ja 8 tunnin ennusteet. Alimmalla rivillä on esitetty 12 ja 24h ennusteet. Lisäksi jokaisessa kuvassa on esitetty tilan todellinen käyttö (sininen käyrä) ja tilan keskimääräinen käyttötieto viikonpäivän ja kellonajan mukaan (punainen viiva).

## 5 Yhteenveto

Testatulla menetelmillä saavutettiin alustavasti lupaavia tuloksia tilan läsnäolon estimoinnissa. Ennustuksen tarkkuus heikkenee, kun ennustuksen pituus kasvaa. Käytössä olleen datan luotettavuuden ja määrän johdosta käyttäjämääriä ei pystytty estimoimaan eikä ennustamaan luotettavasti. Lisäksi läsnäolon ennustamiseen ja estimointiin liittyviin tuloksiin tulee suhtautua varauksella ja tulokset ovat ainoastaan suuntaa antavia johtuen ongelmista datan luotettavuuden osalta. Poikkeavat tilanteet aiheuttivat suuria poikkeamia ennustusmallin ja todellisen tilan välillä. Mallista todennäköisesti puuttuu joitain tärkeitä muuttujia, esim. tenttiviikot, joiden lisääminen mallin syötteiksi voisi parantaa mallin toimintaa. Kattavammalla datalla voitaisiin todennäköisesti saavuttaa parempia tuloksia.

Jatkotutkimusta tarvitaan, jotta voidaan varmistua menetelmien toimivuudesta luotettavalla datalla ja, jotta voidaan optimoida mallien parametrit sekä lisätä malliin tärkeitä muuttujia. Optimoinnin kannalta datan luotettavuus on tärkeää, jotta voidaan mahdollisten ongelmien johtuvan valituista mallin parametreista eikä datassa olevista virheistä. Jatkotutkimusta tarvitaan menetelmien soveltamiseksi myös ihmismäärien ennustamiseen pelkän läsnäolon sijasta. Ihmismäärien ennustamiseen tulee kerätä dataa kattavampi data. Ennustemalleja vertailemalla voidaan mahdollisesti löytää tarkempia menetelmiä.

Käytettyjen estimointimallien ongelma on, että ne tarvitsevat opetusdataksi tilan läsnäolotiedon tai ihmismäärän. Mikäli rakennuksen tiloista tämä tieto on saatavilla, ei ihmismäärää tarvitse estimoida epäsuorasti ylipäättänsä. Jatkotutkimusta tarvitaan menetelmien löytämiseksi, mitkä eivät vaatisi tilakohtaista opetusdataa, vaan mallin opettamisessa voitaisiin käyttää esimerkiksi muista rakennuksista saatavaa dataa tai harvempaa ihmismäärä tietoa, jota voitaisiin kerätä esimerkiksi tilavaraussovelluksen tai kalenterien kautta.

## 6 Lähteet

- [1] A. Szczurek, M. Monika ja T. Pietrucha, "Occupancy determination based on time series of CO2 concentration, temperature and relative humidity," *Energy and Buildings*, osa/vuosik. 147, nro 15, pp. 142-154, 2017.
- [2] J. Zhao, B. Lasternas, K. P. Lam, R. Yun ja V. Loftness, "Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining," *Energy and Buildings*, pp. 341-355, 2014.
- [3] Z. Chen, C. Jiang ja L. Xie, "Building occupancy estimation and detection: A review," *Lihua*, pp. 260-270, 2018.
- [4] P. J. Brockwell ja R. A. Davis, *Introduction to Time Series and Forecasting*, Springer-Verlag, 1996.
- [5] S. Billings, *Nonlinear system identification : NARMAX methods in the time, frequency, and spatio-temporal domains*, John Wiley & Sons, Ltd, 2013.
- [6] Watershed Arts Trust Ltd., "GitHub repository," 1 6 2018. [Online]. Available: <https://github.com/WatershedArts/Footfall>.

## LIITE 2



# Rakennusten käytön ennustaminen koneoppimisella hyödyntäen olemassa olevaa mittausdataa

Petri Jantunen, Doctoral Candidate  
petri.jantunen@aalto.fi

## Esityksen sisältö

1. Tutkimuksen tausta
2. Kokeiluhanke
3. Tulokset



# Tutkimuksen tausta

## Käytön tuntemisen hyödyt

- Ylläpidon resurssien allokointi
- Tilatehokkuus
- Energiatehokkuus
- Ennakointi

## Tiedon hyödyntämisen vaikeudet

- Dataa ei ole
- Mittaaminen ongelma:
  - Yksityisyys
  - Kustannukset

# Kokeiluhanke



## Tavoitteet

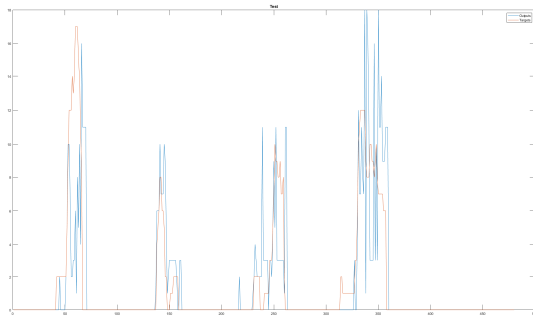
- **Testata olemassa olevan mittausdatan hyödyntämistä käytön ennustamisessa**
- **Testattiin olemassa olevia aikasarjaennustusmenetelmiä**
- **Arvioida ennustemallien tarkkuutta**

## Käytetty data

- **Ryhmätyötila**
- **Läsnäolomäärät**
  - Kolme kuukautta
- **Olosuhdetiedot**
  - CO2, noin vuosi
  - Lämpötila, noin vuosi
- **Ilmanvaihtokoneiden tiedot (usean koneen data)**
  - Poistoilman lämpötila, noin vuosi
  - Käynti, noin vuosi

# Tulokset

## Ihmismäärien arviointi



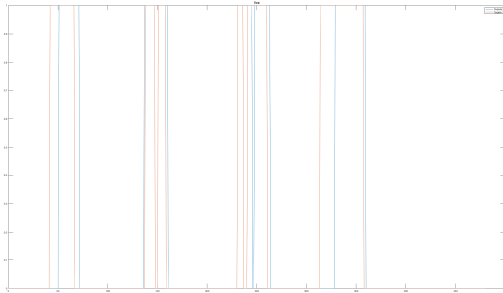
Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
340	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
320	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
300	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
280	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
260	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
240	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
220	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
180	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
160	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
140	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

# Läsnäolon arviointi,

Kaikki data hyödynnetty

**Precision = 0,93**

**Recall = 0,94**



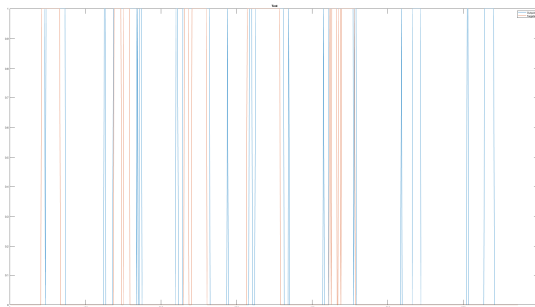
Mallin ulostulo	Todellinen läsnäolo		
	0	1	
0 (ei käytössä)	347 72.3 %	26 5.4 %	93.0 % 7.0 %
1 (käytössä)	20 4.2 %	87 18.1 %	81.3 % 18.7 %
	94.6 %	77.0 %	90.4 %
	5.4 %	23.0 %	9.6 %

# Läsnäolon arviointi,

Pelkkä lämpötila

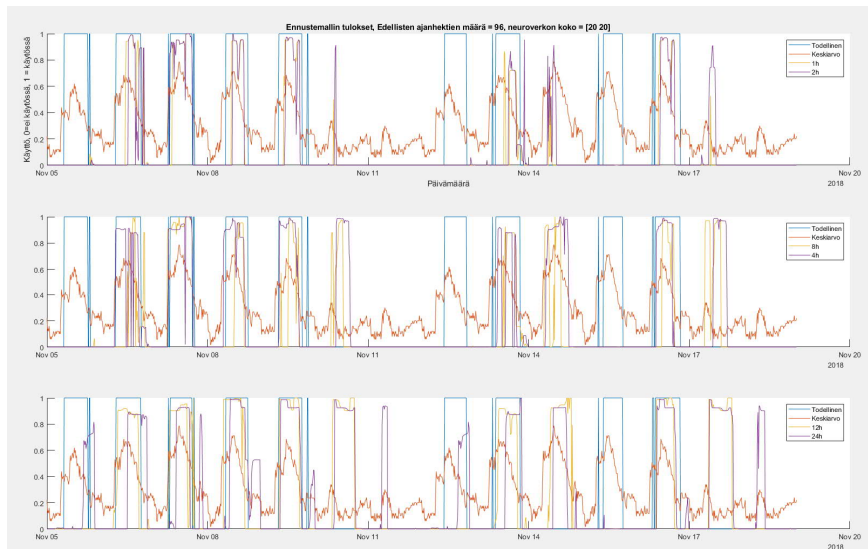
**Precision = 0,98**

**Recall = 0,87**



Tulokset validointidatalla			
Mallin ulostulo	Todellinen läsnäolo		
	0	1	
0 (ei)	448 66.7 %	9 1.3 %	98.0 % 2.0 %
1 (käytössä)	81 12.1 %	134 19.9 %	62.3 % 37.7 %
	84.7 %	93.7 %	86.6 %
	15.3 %	6.3 %	13.4 %

## Tuloksia, ennustaminen



## Tulosten arviointi

- **Luotettavien datojen määrä suhteellisen pieni**
  - Suurin osa käytettävistä datasta sisälsi epävarmuuksia liittyen ihmismääriin
  - Parempia tuloksia paremmalla datalla
- **Ihmismäärien perusteella luokiteltujen luokkien n oli pieni kuin ihmismäärä > 0**
- **Menetelmät vaikuttavat lupaavilta**
  - Kerätään lisää data
  - Pyritään parempaan tarkkuuteen
  - Ihmismäärien ennustaminen

## Yhteenveto

- **Käyttäjätieto on estimoitavissa olemassa olevasta mittausdatasta**
  - CO<sub>2</sub>
  - Lämpötila ja suhteellinen kosteus varauksin
  - Käyttäjälaitteiden sähkönkulutus (tutkimusten mukaan)
- **Ennustemallit**
  - Pitkällä ennustehorisontilla virhe kasvaa merkittävästi
  - Lyhyellä aikavälillä varsin tarkka
  - Tulee testata lisää menetelmiä ja enemmän dataa

## LIITE 3

Hei!

Älykkään talotekniikan pilottikohde-seminaarista 22.1. muutama asia. Nämä eivät ehkä kosketa teitä kaikkia, mutta pitäkää se mikä on hyvää

**1) Käykää itse ilmoittautumassa**, niin tiedän väkimäärän esim. lounasta ajatellen

<https://www.webropolsurveys.com/S/715ACFCF2AC63F5C.par>

Ilmoittautuneita on tällä hetkellä vajaat 30 kpl, ja tässä määrässä on vain pari hankkeen yritysten edustajaa ja tekin halunnette **ruokaa**.

**2)** Olen mitoitannut ohjelman niin, että jokaiselle **yrietykselle on 5 min aikaa** puhua ja tähän päälle tietysti esityksen vaihto. Yritetään pysyä tässä aikataulussa, niin meille ei tule kiire ja pääsemme lopettamaan ajoissa, jolloin viimeisetkin esitykset jaksetaan kuunnella.

5 min on haastava aika puhua, joten jos rutiinisi lyhyisiin esityksiin on vajavainen, niin **harjoittele pari kertaa** etukäteen työpöydän ääressä. Mene suoraan asiaan. Esityksissä tulisi fokuoioitua siihen, mitä oma tuote tarjoaa älykkääseen rakennukseen. 5 slideä on aika hyvä määrä. Vähemmälläkin pärjää.

Esitykset voi lähettää minulle **etukäteen**.

**3)** Tila on siis Aalto Design factory ja siellä "Stage" -niminen sali (<https://designfactory.aalto.fi/insiders/space-reservation/#stage>)

Paikkana se on vähän rouhea, mutta myös tavallista luentosalia inspiroivampi

Ständejä varten tilassa on pöytiä. Sähköäkin varmaan saa jos sitä joku tarvitsee, mutta ovat **jatkojohdot** olisi hyvä olla mukana.

Käytössä on videotykki ja äänentoistojärjestelmä

**4)** Paikalle voi saapua jo aamusta klo **9.30 alkavaan Ekotohokkaan rakentamisen sidosryhmätapaamiseen** (ohjelma ihan alinna). Sinne on ilmoittautuminen erikseen. Muuten aloitetaan lounaalla klo 12. Johtuen edellisestä tilaisuudesta olisi hyvä, jos esim. pöytiä ei siirreltäisi aamupäivän aikana. Yritän laittaa/laitattaa mahdollisimman paljon valmiiksi etukäteen.

**5) Tilaisuutta voi edelleen markkinoida** esim. ao. kutsulla. Ilmoittautumisen takaraja on 16.1. mutta se kyllä joustaa ja koskee lähinnä lounasta

Varmasti jotain unohdin, mutta kysykää niin täydennetään.

## ÄLYKKÄÄN TALOTEKNIIKAN PILOTTIKOHDE - loppuseminaari

---

### Arvoisa älykkään rakentamisen asiantuntija!

Kutsumme teidät Aalto-yliopiston Kira-Digi hankkeen  
"Älykkään talotekniikan pilottikohde" loppuseminaariin

ti 22.1. klo 12.30-15 (kevyt lounas klo 12...12.30)

Aalto Design Factory, Betonimiehenkuja 5 C, Espoo  
(<https://designfactory.aalto.fi/>)

Tilaisuudessa vuorottelevat Aalto-yliopiston hankkeen esittely  
sekä hankkeessa mukana olleiden yritysten puheenvuorot  
teemasta "mitä yrityksemme voi tarjota älykkäästi  
toteutettuihin rakennuksiin?"

Tilaisuudessa on esillä myös yritysten pieniä esittelypisteitä.

Ilmoittautuminen **16.1. mennessä** linkistä  
<https://www.webropolsurveys.com/S/715ACFCF2AC63F5C.par>

### Ohjelma

- 12.00 Kevyt lounas
- 12.30 Mitkä älyrakennukset,  
Professor of Practice **Jaakko Ketomäki**, Aalto-  
yliopisto
- Yrityspuheenvuorot:  
KNX Finland ry  
Granlund Oy  
Helvar Oy
- 13.00 Pilottikohteen esittely,  
Professor of Practice **Heikki Ihasalo**,  
Aalto-yliopisto
- Yrityspuheenvuorot:  
Connected Finland Oy  
Caverion Suomi Oy
- 14.00 Älykkään talotekniikan hyödyt,  
diplomityöntekijä **Laura Remes**  
Aalto-yliopisto
- Yrityspuheenvuorot:  
KT-interior Oy



**BECKHOFF**



Connected Finland

**SIEMENS**



MIRASYS

**Caverion**

**Schneider**  
Electric



Schneider Electric Finland Oy  
Tieto Oyj  
14.30 Rakennusten käytön ennustaminen  
koneoppimisella hyödyntäen olemassa olevaa  
mittausdataa,  
tohtorikoulutettava **Petri Jantunen**  
Aalto-yliopisto

Yrityspuheenvuorot:  
Fidelix Oy  
Siemens Oyj

Lisätietoja:  
Jaakko Ketomäki  
Professor of Practice, Aalto Yliopisto | p. 050 3000 119 |  
[jaakko.ketomaki@aalto.fi](mailto:jaakko.ketomaki@aalto.fi)

**P.S.**

Samana päivänä samassa paikassa klo 9.30-12 järjestetään  
Aalto yliopiston **Ekotehokkaan rakentamisen ja talotekniikan  
yhteistyötapaaminen**, jossa kerrotaan tutkimusalueiden  
ajankohtaiset kuulumiset ja esitellään meneillään olevia  
tutkimushankkeita. Sinne olette myös tervetulleita.

Ilmoittautumiset 15.1. mennessä [risto.kosonen@aalto.fi](mailto:risto.kosonen@aalto.fi) ja  
[seija.erander-luukkanen@aalto.fi](mailto:seija.erander-luukkanen@aalto.fi).

-----

**Subject:** EKOTEHOKKAAN RAKENTAMISEN JA TALOTEKNIIKAN YHTEISTYÖTAPAAMINEN 22.1.2019

Hyvät Yhteistyökumppanit,

Seuraava yhteistapaaminen pidetään  
Tiistaina **22.1.** klo **09:30-12:00**

Paikka on Design Factory Otaniemi (kts. liitetiedoston kartta 25 Betonimiehenkuja 5)

**Ohjelma:**

09:00-09:30 Kahvi

Eri laitoksien ajankohtaiset asiat ja valittujen tutkimusprojektien esittely:

09:30-10:15 Rakennustekniikan laitos

- Olli Seppänen (esittelee kurssinsa ja tutkimushankkeensa), 30 min
- Camilla Vornanen-Winqvist (Kuntotutkimusopas) ja Tuomas Alapieti/Raimo Mikkola (Sisäilmapoliisi), yhteensä 15 min

10:15-11:00 Konetekniikan laitos

- Yleiset kuulumiset, Risto Kosonen, 15 min
- Techno-economic performance of community sized solar heating system, Hassam ur Rehman, 15 min



- Indoor airflow characteristic under increased heat load condition, Sami Lestinen, 15 min

11:00-11:45 Sähkötekniikan ja automaatiotekniikan laitos

- Yleiset kuulumiset, Jaakko Ketomäki, 15 min
- Tilojen käyttöasteen reaaliaikainen seuranta, Heikki Ihasalo/Tuomas Juopperi, 15 min
- Sisäilmapoliisihanke (mittausjärjestelmien näkökulmasta), Heikki Ihasalo/Panu Harmo, 15 min

11:45-12:00 Loppukeskustelu

Vahvistatko osallistumisen ti 15.1. mennessä allekirjoittaneelle ja [seija.erander-luukkanen@aalto.fi](mailto:seija.erander-luukkanen@aalto.fi)

p.s

Kira-Digi -hankkeen "Älykkään talotekniikan pilottikohde" loppuseminaari pidetään Design factoryssä klo 12.30-15:00.

Lisätietoja ja ilmoittautumiset [jaakko.ketomaki@aalto.fi](mailto:jaakko.ketomaki@aalto.fi)

Tervetuloa, Risto

Jaakko Ketomäki

Professor of Practice, Aalto University | p. +358 50 3000 119 | [jaakko.ketomaki@aalto.fi](mailto:jaakko.ketomaki@aalto.fi)