



Sumeutetun aikasarjadataan visualisointi

Janne Saukkio

OPINNÄYTETYÖ
Toukokuu 2022

Tieto- ja viestintäteknikka
Ohjelmistotekniikka

TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Tieto- ja viestintäteknikka
Ohjelmistotekniikka

SAUKKIO, JANNE:
Sumeutetun aikasarjadataan visualisointi

Opinnäytetyö 32 sivua
Toukokuu 2022

Teollisuuden nykyaikaistumisen ja digitalisoitumisen myötä laitteiden lähettämä datamäärä on moninkertaistunut ja monipuolistunut. Suuria datamääriä on vaikea tulkita ja visualisoida, eikä erilaisten arvojen välisiä vuorovaikutuksia välttämättä pystytä tunnistamaan. Erilaisten arvojen vuorovaikutusten havainnointi mahdollistaa poikkeuksien tunnistamisen sekä saman tyyppisten laitteiden vertailun keskenään. Aiemmin tehdyn tutkimuksen pohjalta on todistettu, että erilaisten arvojen välistä suhdetta voidaan mallintaa sumean logiikan avulla tehokkaasti.

Opinnäytetyössä jatkettiin aiemman kehityksen pohjalta sumeaa mallintamiseen luotua työkalua, johon toteutettiin käyttöliittymä sumean mallin visualisointiin. Käyttöliittymää varten toteutettiin sovellusrajapinta, joka palauttaa sumeasta mallista lasketun luottamusjakauman. Jakauma ja sen avulla toteutettu visualisointi kuvaavat, kuinka yksittäinen laite vastaa sumean mallin muita saman tyyppisiä laitteita.

Opinnäytetyössä myös pohdittiin mahdollisia tulevaisuuden käyttömahdollisuuksia sumealle mallinnukselle ja mallin erilaisille visualisoinneille. Sumeaa mallia voidaan muun muassa käyttää laitteiden poikkeavuuksien seurantaan tai isomman laitejoukon käyttäytymisen vertailuun.

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Degree Programme in ICT Engineering
Software Engineering

SAUKKIO, JANNE:
Visualisation of Fuzzified Time-Series Data

Bachelor's thesis 32 pages
May 2022

Modernisation and digitalisation of heavy industry has resulted in an increased amount of data from machines. Large amounts of data can be hard to interpret and visualize effectively. Further problem is that relationships between data values cannot be seen in huge data sets or processed without a large amount of computing power. However, previous studies have proven that relationships between different values can be modelled efficiently with fuzzy logic.

In this thesis, a user interface was developed as part of an already existing web application for visualisation of a fuzzy model. A REST API was developed for the interface that calculates a confidence score distribution from the fuzzy model. The calculated distribution and the resulting visualisation of it describe how well a certain machine corresponds with other machines of same type in the fuzzy model.

Furthermore, possible future uses were discussed for fuzzy models and different kinds of visualisation methods of it. In future fuzzy models may be used for monitoring of anomalies in machines or for comparing usage of multiple machines against each other.

Key words: visualisation, charts, fuzzy logic, time-series data, fuzzification

SISÄLLYS

1	JOHDANTO	6
2	TARKOITUS JA TAVOITTEET	7
	2.1 Nykyinen tilanne	7
	2.2 Tavoitteet	7
3	TEKNOLOGIAT	9
	3.1 Sumea logiikka	9
	3.2 Sumea mallinnus	10
	3.3 Luottamuspisteet	13
	3.4 Timescale	14
	3.4.1 Continuous aggregation	14
	3.4.2 Retention policy	14
	3.5 Sovellusrajapinta	15
	3.6 Apache ECharts	15
4	TYÖN TOTEUTUS	16
	4.1 Sumeiden arvojen tiivistäminen	16
	4.2 Rajapinta	17
	4.3 Visualisointikäyttöliittymä	19
	4.4 Aikasarjakaaviot	22
5	POHDINTA	25
	5.1 Poikkeusilmoitukset	26
	5.2 Tutkakaavio	27
	5.3 Pylväskaavio	29
	5.4 Pistelämpökarttakaavio	30
	LÄHTEET	32

LYHENTEET JA TERMIT

Aikasarjadata	Aikasarjadata on aikavälein kerättyjen datapisteiden sarja, mikä antaa mahdollisuuden seurata ajan kuluessa datassa tapahtuvia muutoksia (Timescale 2020).
HTML5	Verkkosivuissa käytettävän HTML-merkintäkielen viides versio (engl. Hypertext Markup Language).
HTTP	Tiedonsiirtoprotokolla (engl. Hypertext Transfer Protocol), jota käytetään datan siirtoon verkkoselaimen ja palvelimen välillä.
IoT	Esineiden Internet (engl. Internet of Things).
JavaScript	Verkkoselainsovelluksissa käytettävä ohjelmointikieli.
JAX-RS	Jakarta EE:n määrittelemä ohje REST-rajapintojen luomiseen (engl. Jakarta RESTful Web Services).
KPI	Suorituskykymittari (engl. key performance indicator).
PostgreSQL	Avoimen lähdekoodin SQL-tietokanta.
Regatta Insights	Web-sovellus IoT-laitteiden datan visualisointiin, huoltokirjanpitoon ja vikailmoituksiin.
SQL	Tietokannan kyselykieli (engl. Structured Query Language) relaatiotietokantojen datan hakemista ja muokkaamista varten.

1 JOHDANTO

Teollisuuden nykyaikaistumisen ja digitalisoitumisen myötä yksittäiset laitteet voivat lähettää satoja erilaisia arvoja ja tietoja laitteen toiminnasta. Tämä väistämättä johtaa suureen määrään informaatiota, jota ei välttämättä pystytä tulkitsemaan tai näkemään erilaisten arvojen vuorovaikutusta toisiinsa.

Työn tarkoituksena on toteuttaa käyttöliittymä sumean mallin visualisointia varten osaksi toimeksiantajan (Remion Oy) Regatta Insights -sovellusta. Käyttöliittymän tarkoituksena on tarjota toimeksiantajan asiakkaille entistä laajempi ja reaaliaikainen näkyvyys omaan dataansa, jotta laitteiden lähettämien arvojen vuorovaikutusten ja laitteiden poikkeuksien havainnointi olisi mahdollista.

Aiemmat toimeksiantajan projektit datan analysointiin ovat olleet yksittäisiä katsauksia dataan ja niiden kehittäminen on ollut asiakaskohtaista. Yksittäiset analyysiprojektit eivät ole tarjonneet asiakkaille nopeaa ratkaisua reaaliaikaisen datan visualisointiin ja ymmärtämiseen.

2 TARKOITUS JA TAVOITTEET

2.1 Nykyinen tilanne

Toimeksiantaja tarjoaa asiakkailleen teollisen Internetin laitteiden datan keräystä, prosessointia ja analysointia. Ympäri maailmaa sijaitsevat laitteet lähettävät muutamana kymmenen sekunnin välein purskeina niiltä kerättyä dataa erilaisilla tiedonsiirtoprotokollilla. Laitteiden erilaisia arvoja luetaan mm. laitteiden ohjausjärjestelmistä ja sensoreista.

Jokaisen tyyppiselle laitteelle on määriteltä valmiiksi mitä ns. signaaleja ne lähettävät. Signaali on kuvaus jostain tietyntyyppisestä datasta ja se voi olla esimerkiksi moottorin lämpötila, laitteen sijainti tai akselin pyörimisnopeus. Kaikki signaaliarvot tallentuvat mittausajan ja laitetunnuksen kanssa tietokantaan, josta niitä voidaan lukea prosessoitavaksi ja visualisoitavaksi.

Nykyisellään signaalidatasta lasketaan erilaisia johdettuja signaaleja ja KPI-arvoja asiakkaan määrittelemällä tavalla. Signaaliarvoja voidaan myös tiivistää eri funktioilla pienemmiksi yhteenvedoiksi datan määrän vähentämiseksi. Dataa voidaan visualisoida toimeksiantajan web-käyttöliittymässä tai tarjota rajapintadatan käyttöä varten ulkoisissa järjestelmissä.

2.2 Tavoitteet

Työn pääasiallisena tavoitteena on toteuttaa sumean aikasarjadataan visualisointikäyttöliittymä osaksi jo olemassa olevaa Regatta Insights -sovellusta. Visualisointia varten toteutetaan sovellusrajapinta sumeiden arvojen laskemista varten, joka toimii osana toimeksiantajan muita rajapintoja sekä käyttäjänhallintaa. Rajapinnan pitää toimia tehokkaasti, jotta käyttöliittymä ja sen visualisoinnit ovat responsiivisia.

Toisena tavoitteena työssä on pohtia ja tutkia muita mahdollisia tapoja sumeiden arvojen ja niistä laskettujen luottamuspisteiden visualisointiin. Visualisointien pitää olla selkeitä ja ymmärrettäviä sekä olla toimeksiantajan asiakkaille tarpeellisia. Kaikissa työn visualisoinneissa on tavoitteena mahdollistaa poikkeuksien havainnointi ja parempi näkemys eri arvojen vuorovaikutuksiin.

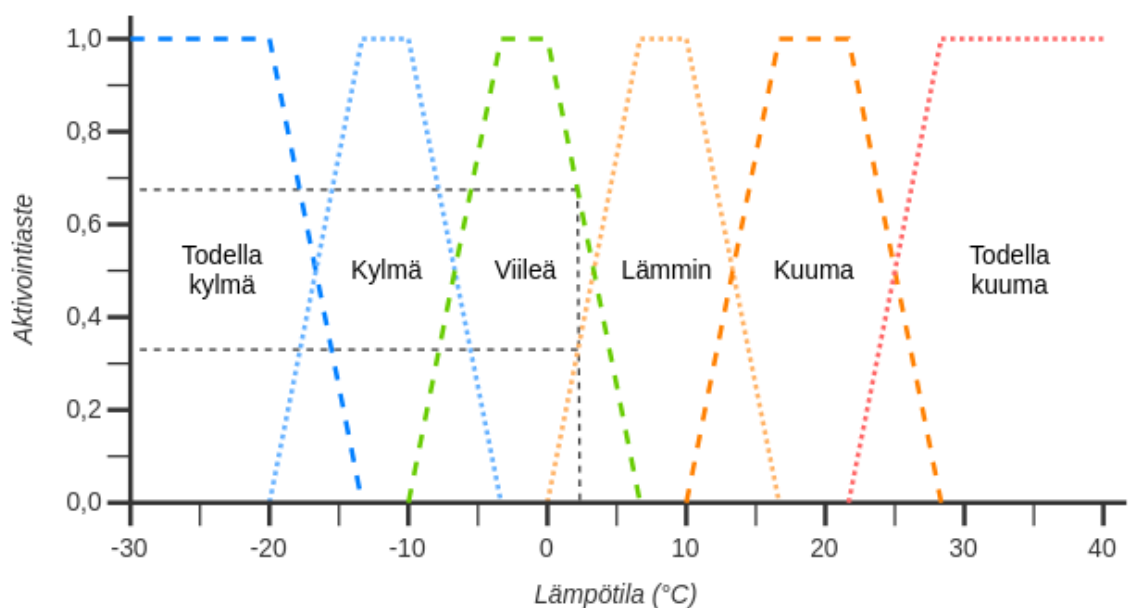
3 TEKNOLOGIAT

3.1 Sumea logiikka

Sumea logiikka on alun perin 1960-luvulla esitelty ajatus epätarkan datan hallintaan. Sumeassa logiikassa muuttujan arvo on välillä 0–1. Toisin kuin Boolean logiikassa, jossa arvo voi olla ainoastaan 0 tai 1 eli epätosi tai tosi, sumeassa logiikassa arvo voi olla samaan aikaan jossain määrin epätosi ja tosi (Wolf ym. 1996, 2.)

Lämpötila on yksi hyvä esimerkki muuttujasta, jota voidaan mallintaa sumealla logiikalla. Muuttujan eri arvoalueille voidaan asettaa kielellisiä termejä kuvaamaan muuttujan arvoa kielellisesti kuten ”todella kylmä”, ”kylmä”, ”viileä”, ”lämmin” ja niin edespäin.

Kuviossa 1 olevia alueita kutsutaan jäsenyyksifunktioiksi. Ne määrittävät aktivointiasteen ja kielellisen termin tietyille lämpötila-arvoille. Esimerkiksi lämpötila-arvo 2,5 °C on lämmin noin 0,32 aktivointiasteella ja viileä noin 0,68 aktivointiasteella. Muuttujan arvon kielellisen termin määrittämisprosessia kutsutaan sumeutukseksi (Wolf ym. 1996, 2).



KUVIO 1. Esimerkki lämpötilan kielellisesti mallintamisesta

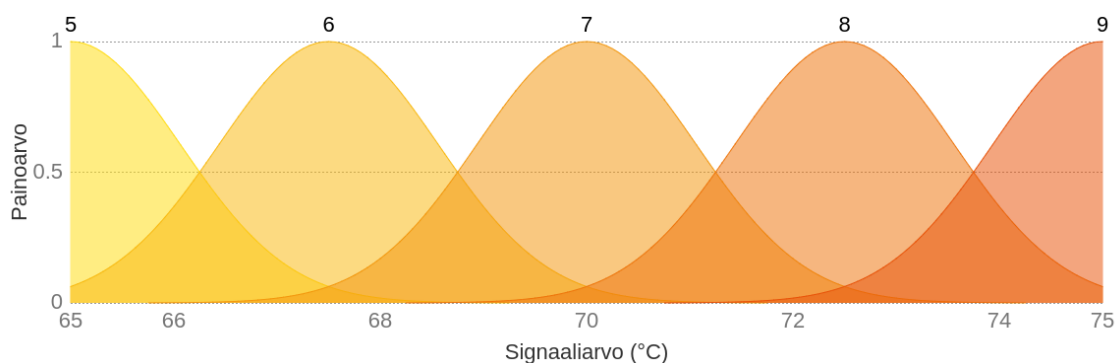
Jäsennysfunktiot voivat olla muodoltaan mitä tahansa ja ne voidaan sijoittaa niin, että jokaisella arvolla kaksi tai kolme jäsennysfunktiota aktivoituu yhtä aikaa (Keski-Heikkilä 2021, 5). Tärkeää kuitenkin on, että jokainen mahdollinen arvo on mallinnettu vähintään yhdellä jäsennysfunktiolla (Keski-Heikkilä 2021, 20).

3.2 Sumea mallinnus

Aikasarjadataan sumea mallinnusmenetelmä perustuu Vainio ym. (2008) ja Keski-Heikkilän (2021) tekemiin aiempiin tutkimuksiin, joissa on tutkittu muuttujien välisen vuorovaikutusten mallintamista sumean logiikan avulla. Menetelmän avulla voidaan suuria määriä dataa muuntaa sumeiksi arvoiksi tehokkaasti.

Jotta laitteiden lähettämää signaalidataa voidaan sumeuttaa, pitää luoda sumea malli. Sumeaan malliin tallentuu eri signaalien sumeiden kielellisten termien välinen riippuvuus toisistaan ajan suhteen, jossa monta arvoa on totta samanaikaisesti, mutta eri aktivointiasteilla. Malliin voi kuulua vain saman tyyppisiä laitteita, jotka lähettävät samoja signaaleja.

Jokaiselle mallissa käytettävälle signaalille pitää löytyä konsepti, joka on useasta kielellisestä termistä eli jäsennysfunktiosta muodostuva kuvaus signaalin arvolle. Konsepti määrittää miten signaalin arvot sumeutetaan ja mikä on signaalin oletettu arvoalue. Jäsennysfunktioiden määrä määrittää kuinka tarkasti signaalia mallinnetaan (Vainio ym. 2008, 45.). Konseptin jäsennysfunktiot asetellaan pääasiallisesti tasavälein siten, että ne leikkaavat toisiaan 0,5 aktivointiasteen kohdalla, kuten kuvion 2 esimerkissä. Jäsennysfunktioille voidaan asettaa kuvaavat nimet, mutta konseptin sisältäessä monia termejä on helpompi käyttää juoksevaa kokonaislukua.



KUVIO 2. Konseptin kielelliset termit eli jäsennysfunktiot

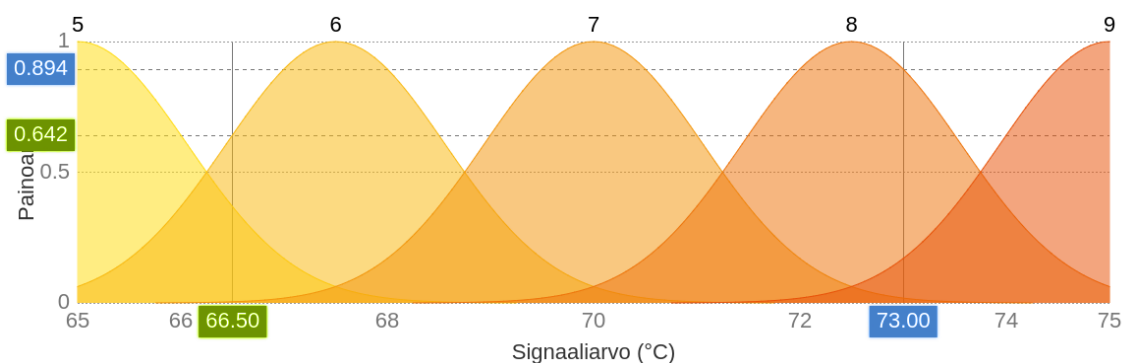
Kun sumea malli on määritelty, voidaan siihen asetetut signaalit sumeuttaa keskenään. Tässä sumeutusprosessissa lasketaan jokaiselle signaalin datapisteelle kaksi sumeaa arvoa valitsemalla jokaiselle signaaliarvolle kaksi korkeimman aktivointiasteen jäsennysfunktiota. Taulukosta 1 löytyy alla kuvattua prosessia esiteltäessä käytettävät esimerkki lämpötilasignaalien arvot, jossa kuvitteellinen laite on lähettänyt moottorin sekä öljyn lämpötila arvoja minuutin välein.

TAULUKKO 1. Esimerkkilaitteen sumeutettavat lämpötilasignaaliarvot

AIKA	MOOTTORIN	ÖLJYN
	LÄMPÖTILA (°C)	LÄMPÖTILA (°C)
12:00:00	66,5	73,0
12:01:00	63,2	71,1
12:02:00	69,7	74,8

Kuviossa 3 on mallinnettuna yllä olevien lämpötilasignaalien sumeutus käyttäen samaa konseptia. Kuvion jokaiselle jäsennysfunktiolle on asetettu yksilöllinen järjestysluku, jotta ne pystytään erottamaan toisistaan.

Sumeutuksessa moottorin lämpötilasignaalin arvolla 66,5 °C korkein aktivointiaste on jäsennysfunktiolla kuusi ja se on 0,642. Vastaavasti öljyn lämpötilasignaalin arvolla 73,0 °C korkein aktivointiaste on jäsennysfunktiolla kahdeksan ja se on 0,894. Samoilla arvoilla etsitään myös toiseksi korkeimman aktivointiasteen jäsennysfunktiot ja niiden saamat aktivointiasteet.



KUVIO 3. Kahden signaaliarvon korkeimman aktivointiasteen jäsennysfunktiot

Tästä saadaan yhden datapisteen kaksi korkeinta sumeaa arvoa, jotka esitetään taulukossa 2 signaalien jäsennysfunktioiden järjestysnumeroilla ja aktivointias-teilla. Sumeiden arvojen aktivointiasteet kertovat kuinka vahvasti muuttuja kuuluu jäsennysfunktioon (Keski-Heikkilä 2021, 23).

TAULUKKO 2. Yhden datapisteen kaksi korkeinta sumean termin arvoa

	MOOTTORIN LÄMPÖTILA	AKTIVOIN- TIASTE	ÖLJYN LÄMPÖTILA	AKTIVOIN- TIASTE
KORKEIN	6	0,642	8	0,894
2. KORKEIN	5	0,356	9	0,172

Mallin tuottamia sumeita arvoja voidaan tiivistää tallennettavien tietojen vähentämiseksi. Tiivistäessä sumeiden arvojen painoarvot muunnetaan yhteisen painon muotoon, jolloin jokaiselle riville saadaan yksi yhteinen painoarvo. Keski-Heikkilän (2021, 23) mukaan kaikista tarkin tapa painoarvojen yhdistämiseksi on laskea aktivointiasteista keskiarvo. Taulukon 2 sumeiden arvojen painoarvoista laske-taan keskiarvot alla olevalla kaavalla 1.

$$\frac{0,642 + 0,894}{2} = 0,768 \text{ ja } \frac{0,356 + 0,172}{2} = 0,264 \quad (1)$$

Kaikkien datapisteiden saadut tulokset tallennetaan alkuperäisen aikaleiman kanssa (taulukko 3). Jokainen rivi edustaa yhden datapisteen aikaleimaa ja akti-vointiastetasoa korkein (1) tai toiseksi korkein (2).

TAULUKKO 3. Laitteen lämpötilasignaalien sumeat arvot

AIKA	MOOTTORIN LÄMPÖTILA	ÖLJYN LÄMPÖTILA	AKTIVOINTI- TASO	PAINOARVO
12:00:00	6	8	1	0,768
12:00:00	5	9	2	0,264
12:01:00	4	7	1	0,695
12:01:00	3	8	2	0,340
12:02:00	7	9	1	0,978
12:02:00	6	8	2	0,114

3.3 Luottamuspisteet

Sumeaa mallia voidaan visualisoida luottamuspisteillä, jotka esittävät luottamusta siihen, että arvo on tietyllä arvovälillä. Grandperrin (2021) mukaan luottamuspisteitä voidaan kuvata mm. prosentuaalisella asteikoilla välillä 0–100 % tai kuvainnollisilla termeillä kuten ”matala”, ”keskitaso” ja ”korkea”. Edellä mainitut tavat soveltuvat sumeiden arvojen luottamuspisteiden esittämiseen, mutta tässä työssä käytetään prosentuaalista asteikkoa sen luottavuuden vuoksi.

Laskemalla sumeista arvoista luottamuspisteille erilaisia tasoja saadaan aikaiseksi luottamuspistejakauma, joka kuvaa luottamusta mallista valitulle selitettävän signaalin arvolle. Selitettävä signaali on mallista valittu tekijä, jonka käyttäytymistä halutaan verrata mallin muiden signaalien tuottamiin arvoihin. Mallin muita signaaleja käytetään selittävinä tekijöinä, joiden sumeista arvoista luottamuspistejakauma lasketaan. Selittävät signaalit nimensä mukaisesti selittävät miten selitettävän signaalin pitäisi käyttäytyä.

Luottamuspisteiden avulla siis esitetään mahdollisuus selitettävän signaalin arvon sijoittamiseen jollakin prosentuaalisen tason alueella. Esimerkkinä signaaliarvon sijaitessa >99 %:n luottamuspistetasoalueella voidaan sanoa, että signaaliarvo vastaa odotuksia toisin kuin taas arvon sijaitessa <33 %:n tason alueella vastaa se mallin odotuksia heikosti.

3.4 Timescale

Kaikki sovelluksessa käytetyt signaaliarvot tallennetaan Timescale-aikasarjatietokantaan. Timescale on avoimen lähdekoodin relaatiotietokanta, joka on kehitetty PostgreSQL-tietokannan päälle ja tukee täysin SQL-kyselykieltä, mahdollistaen aikasarjadataan ja perinteisen relaatiodataan tallentamisen samaan tietokantaan (Timescale n.d.). Koska Timescale on pohjimmiltaan relaatiotietokanta, voidaan samaa tietokantaa hyödyntää myös tavallisen relaatiodataan, kuten signaalien ja sumeiden mallien sekä muun metatiedon tallentamiseen.

3.4.1 Continuous aggregation

Timescalesta löytyy continuous aggregation -ominaisuus, joka automaattisesti taustalla ylläpitää datasta tiivistettyä tietokantanäkymää, kun tietokantaan lisätään uusia tai muokataan vanhoja rivejä. Tällä vältetään suurien tietorakenteiden ja taulujen lukemiselta turhaan uudestaan jokaisella kyselyllä, koska vain uudet tai muuttuneet datapisteet pitää tiivistää. (Timescale n.d.) Aikasarjadataan määrä kasvaa erittäin nopeasti ja isojen datamäärien hakemisesta voi tulla hidasta.

3.4.2 Retention policy

Aikasarjadataan kanssa tarkkojen arvojen tärkeys vähenee ajan kuluessa, jonka takia on usein haluttua poistaa vanhoja arvoja tietokannasta levytilan säästämiseksi. Timescale tarjoaa vanhojen arvojen poistoon retention policy -ominaisuuden, joka automaattisesti poistaa vanhoja arvoja sille asetetun aikavälin ja sääntöjen mukaisesti. (Timescale n.d.)

3.5 Sovellusrajapinta

Sovellusrajapinta eli REST-rajapinta (engl. representational state transfer) on tietty suunnitteluperiaatteet täyttävä rajapinta, jonka avulla eri sovellukset tai laitteet voivat yhdistää ja kommunikoida toistensa kanssa. Alun perin 2000-luvun alussa määritetyt periaatteet ovat luoneet yhtenäisen tavan rajapintojen toteutukseen. (IBM Cloud Education 2021.)

3.6 Apache ECharts

Apache ECharts on avoimen lähdekoodin selaimessa toimiva JavaScript-visuaalisointikirjasto, joka tukee monia erilaisia kaaviotyyppejä sekä toimintoja. HTML5-kanvaasiin pohjautuva grafiikkarenderöijä mahdollistaa EChartsin helpon laajennettavuuden ja tehokkaan suorituskyvyn (Li ym. 2018, 1).

4 TYÖN TOTEUTUS

Sumean mallin visualisointiin toteutettiin käyttöliittymä ja sitä varten REST-rajapinta raakadatan sekä sumeiden arvojen hakemista varten. Sumeita arvoja tiivistetään automaattisesti tietokannan sisällä paremman tehokkuuden saavuttamiseksi. Käyttöliittymä ja rajapinta toteutettiin osaksi jo aiemmin kehitettyä sumean mallinnuksen hallintatyökalua. Työkalulla pystytään luomaan ja muokkaamaan sumeita malleja sekä niiden konsepteja ja termejä web-käyttöliittymän kautta. Luodut mallit ja niiden tuottamat sumeat arvot tallentuvat tietokantaan, mutta niille ei toteutettu visualisointia tai jatkoprosessointia opinnäytetyötä edeltäneessä vaiheessa.

4.1 Sumeiden arvojen tiivistäminen

Sumeat arvot vievät noin kaksi kertaa enemmän rivejä tietokannassa raakadataan nähden, koska jokaista datapistettä kohden pyritään valitsemaan kaksi korkeinta sumeaa arvoa. Tarvittavien rivien vähentämiseksi sumeita arvoja voidaan tiivistää pidemmille aikaväleille yhdistämällä samanlaiset rivit toistensa kanssa ja summaamalla painoarvot yhteen. (Keski-Heikkilä 2021, 24, 55.) Ilman datan automaattista tiivistämistä sekä poistoa tietokantakyselyjen kesto kasvaisi koko ajan datan lisääntyessä.

Tietokantahakujen ja visualisoinnin nopeuttamista varten sumealle mallille luotiin kaksi asetusta: tiivistyksen aikaväli ja arvojen säilytysaikaväli. Molemmat asetukset asetetaan mallikohtaisesti sen luonnin yhteydessä.

Mallin tuottamat sumeat arvot tiivistetään Timescalen continuous aggregation -toiminnolla tietokantanäkymäksi, jossa sumeista arvoista luodaan halutulla aikavälillä pienempiä yhteenvetoja. Tiivistettyjä ja mallin sumeita arvoja poistetaan automaattisesti mallille asetetun arvojen säilytysaikavälin mukaan Timescalen retention policy -toiminnolla. Mallin arvojen säilytysaikaväli määrittää miltä ajalta mallin dataa säilytetään ja samalla kuinka pitkältä ajalta sumeita arvoja on käytettävissä visualisoinnissa.

4.2 Rajapinta

Sumean mallin tuottaman datan visualisointia varten toteutettiin REST-rajapinta, joka palauttaa annettujen parametrien mukaan mallista lasketut luottamuspisteet ja selitettävän signaalin raakadatan. Rajapinta toimii käyttöliittymän ja tietokannan välisenä yhteytenä, jolla tietoja saadaan siirrettyä molempiin suuntiin. Kuviossa 4 on esiteltyä esimerkki sumean mallin rajapinnan määrittämisestä käyttäen JAX-RS -merkintöjä.

```
@Path("/models")
public interface ModelResource {
    @GET
    @Path("/{modelId}/devices")
    List<Device> getModelDevices(
        @PathParam("modelId") UUID modelId,
        @QueryParam("offset") @DefaultValue("0") int offset,
        @QueryParam("limit") @DefaultValue("25") int limit
    );

    @GET
    @Path("/{modelId}/evaluate")
    EvaluationResult evaluateModel(
        @PathParam("modelId") UUID modelId,
        @BeanParam EvaluationParametersQuery parameters
    );
}
```

KUVIO 4. Esimerkki sumean mallin REST-rajapinnan määrittämisestä

Kuvion 4 laskentafunktiolle annetaan erilaisia laskentaan liittyviä arvoja ja tunnisteita `EvaluationParametersQuery`-luokkana (kuvio 5). Yhden luokan käyttäminen parametrejä määriteltäessä mahdollistaa niiden helpon muokkaamisen ja uudelleenkäyttämisen muissa rajapinnoissa.

```

public class EvaluationParametersQuery {
    @Parameter(description = "Tarkasteltavan laitteen tunniste")
    @QueryParam("deviceId")
    public UUID deviceId;

    @Parameter(description = "Mallista selitettävän signaalin nimi")
    @QueryParam("evaluatedSignal")
    public String evaluatedSignal;

    @Parameter(description = "Selitettävän signaalin selittävien signaalien nimet. Mikäli tyhjä kaikkia signaaleja käytetään.")
    @QueryParam("explainingSignals")
    public List<String> explainingSignals = new ArrayList<>();

    @Parameter(description = "Ulostulon tarkkuus ISO-8601 aikavälillä")
    @DefaultValue("PT5M")
    @QueryParam("plotResolution")
    public String plotResolution;
}

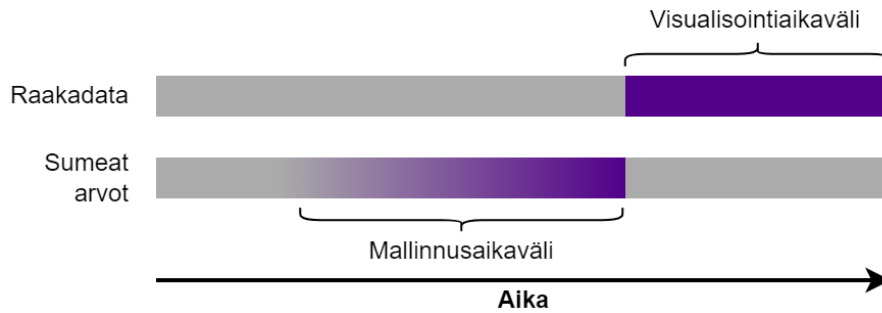
```

KUVIO 5. Esimerkki laskentarajapinnan hakuparametreista

Rajapinnalle annetaan hakuparametreina laitteen tunniste sekä mitä mallin signaaleista käytetään selitettävänä signaalina ja lopuista signaaleista mitkä ovat käytössä selittävinä signaaleina. Muita parametreja ovat mm. visualisoitava aikaväli ja aikavälin tarkkuus, jolla ulostuloa voidaan vielä tiivistää pienemmäksi sekä prosentuaaliset tasot, joilta luottamuspistejakauma halutaan laskea.

Malliin tallennettuja sumeita arvoja halutaan käyttää pidemmältä ajalta kuin visualisoitavalta aikaväliltä mallin tarkkuuden parantamiseksi. Useimmissa tapauksissa halutaan myös välttää visualisoitavan aikavälin arvojen vaikuttamista mallin laskentaan, koska valitun laitteen sumeat arvot löytyvät mallin arvoista. Tätä varten rajapintaan toteutettiin mallinnusaikaväliparametri, joka määrittää miltä ajalta sumeita arvoja käytetään laskennassa.

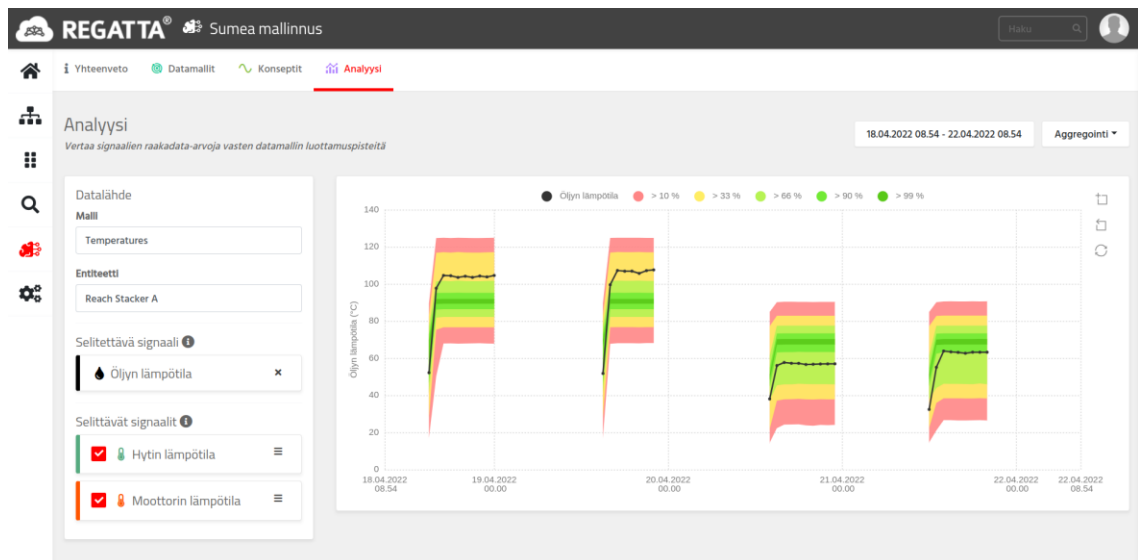
Kuviossa 6 on esimerkki, kuinka rajapinnan kahta edellä kuvattua aikaväliparametria käytetään luottamuspisteiden laskentaan. Laskettaessa mallista luottamuspisteitä käytetään sumeille arvoille puoliintumisaikaa, jolla mallin vanhemmat datat saavat pienemmän painoarvon ja näin ollen vaikuttavat vähemmän ulostuloon. Vanhempien arvojen painotuksen vähentäminen korostaa mallin uudempia sumeita arvoja, jotta malli kuvastaa paremmin laitteiden toimintaa lähempänä nykyistä ajanhetkeä.



KUVIO 6. Rajapinnan aikaväliparametrien käyttö

4.3 Visualisointikäyttöliittymä

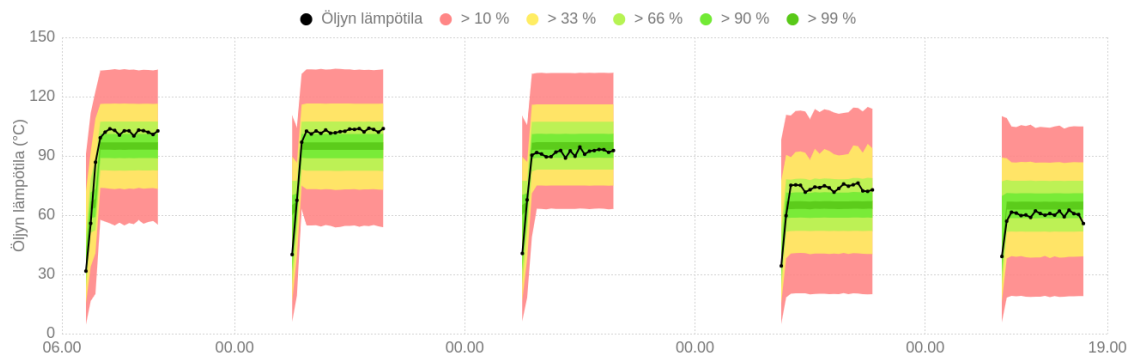
Sumean mallin visualisointikäyttöliittymä (kuva 1) toteutettiin osaksi jo olemassa olevaa Regatta Insights -käyttöliittymää uutena ominaisuutena. Käyttöliittymässä verrataan yhtä valittua laitetta vasten sumeaa mallia ja sen kaikkia saman tyyppiä laitteita. Laitteen signaaleista erikseen valitun selitettävän signaalin käyttäytymistä verrataan mallista laskettuun luottamuspistejakaumaan. Mallin muista signaaleista valitaan signaalit, joita halutaan käyttää selittävinä tekijöinä.



KUVA 1. Sumean mallin visualisointikäyttöliittymä

Kuvion 7 esimerkistä nähdään, kuinka mallista laskettu luottamuspistejakauma piirretään värillisinä alueina punaisesta vihreään. Alueiden päällä esitetään seli-

tettävä signaalin raakadata mustana viivana käyttäjän valitsemalta ajalta. Jakauman alueita lasketaan ja piirretään vain ajoilta, jolloin signaalista on saatavilla dataa, joka voi signaalista riippuen aiheuttaa kuvion mukaisia tyhjiä kohtia.



KUVIO 7. Sumean mallin luottamuspistejakauma ja selitettävän signaalin raakadata aikasarjakaaviolla

Käyttöliittymässä käytetystä ECharts-visualisointikirjastosta ei löydy valmista ratkaisua yllä olevassa kuviossa esiteltyjen alueiden piirtämiseen. Alueita varten työssä toteutettiin mukautettu kaaviosarjatyyppi, joka monikulmioita piirtämällä esittää luottamuspistejakauman eri tasot. Jokainen prosentuaalinen taso piirretään kaavioon omana sarjanaan yksilöllisellä värillä.

Rajapinta palauttaa jokaiselle luottamuspistejakauman tasolle ala- ja yläpisteet jokaista raakadatan aikaleimaa kohden. Pisteet muunnetaan kaaviolla näytettäväksi alueiksi kuvion 8 funktion avulla.

Funktiolle annettu lista sisältää yhden tason aikaleimat ja ala- sekä yläpisteet koko kaavion alueelta. Piirtämistä varten monikulmiot määritetään pisteinä alkaen vasemmasta alakulmasta ja päättyen vasempaan yläkulmaan vastapäivään. Jokaisen pisteen sekä aloitus- ja lopetuspisteen välille piirtyy viiva, joista lopulta muodostuu monikulmio.

Monikulmioita muodostetaan vähintään yksi per jakaumataso muuntamalla datapisteiden ECharts-kirjaston avulla kaavion sisäisiksi koordinaateiksi, jotta piste piirtyy oikeaan kohtaan kaaviossa. Mikäli datapisteiden aikaleimojen välissä esiintyy liian pitkä tauko, monikulmion muodostus katkaistaan ja aloitetaan uuden muodostaminen. Ilman aikaleimojen välien tarkistusta, alueiden välille piirtyisi

joissakin tapauksissa turhia tasaisia alueita, joista ei ole dataa, epäselkeyttämään kaaviota.

```
function renderArea(data: [number, number, number][[]]) {
  let y0: [number, number][[]] = [];
  let y1: [number, number][[]] = [];

  const polygons: [number, number][[]][[]] = [];

  // Käy läpi jokainen luottamuspistejakauma
  for (let i = 0; i <= data.length; i++) {
    // Tallenna monikulmio, kun käsitellään jakauman
    // viimeistä pistettä tai aikaleimoissa on liian pitkä tauko
    if (i >= data.length || shouldBreak(data[i], data[i - 1])) {
      polygons.push(y0.concat(y1));
      y0 = y1 = [];

      // Ei pisteitä jäljellä, hyppää ulos
      if (i >= data.length) {
        break;
      }
    }

    const [time, bottom, top] = data[i];

    // Muunna aika ja jakauman alareunan arvo kuvaajan
    // koordinaatiksi ja lisää se alareunan pistejoukkoon
    y0.push(coord([time, bottom]));

    // Muunna aika ja jakauman yläreunan arvo kuvaajan
    // koordinaatiksi ja lisää se yläreunan pistejoukon
    // ensimmäiseksi
    y1.unshift(coord([time, top]));
  }

  return polygons;
}
```

KUVIO 8. Esimerkki luottamuspistejakauman alueiden piirtämisfunktioista

4.4 Aikasarjakaaviot

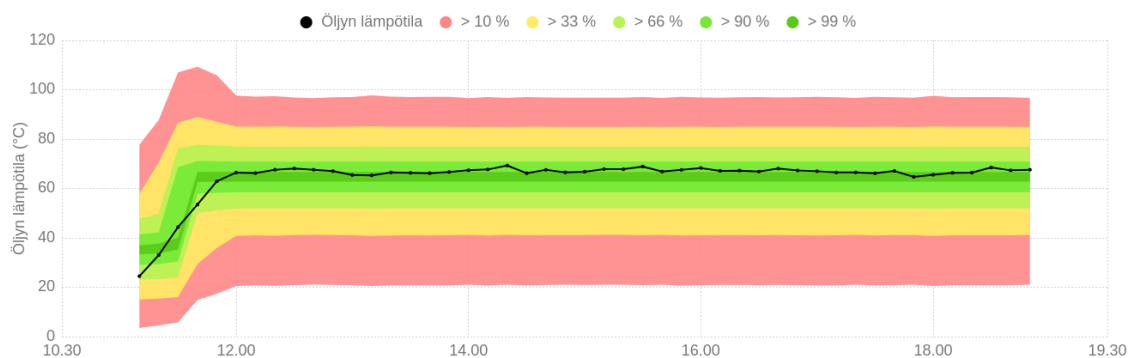
Työn painopisteenä oli toteuttaa käyttöliittymään visualisointi sumeasta mallista, jonka avulla voi havainnoida mallin laitteiden toimintaa. Seuraavissa visualisointiesimerkeissä ollaan mallinnettu useaa ympäri maailmaa toimivaa nostolaitetta. Laitteita käytetään noin kahdeksan tuntia päivässä yhtä jaksoisesti ja niiden lähettämistä signaaleista ollaan valittu: öljyn, hytin ja moottorin lämpötila. Lämpötila-arvot ovat tallennettu Celsius-asteina noin viiden minuutin tarkkuudella.

Kuviosta 9 nähdään kuinka viitenä kuviossa näkyvänä päivänä, laitteen käynnistyttyä on öljyn lämpötila kasvanut nopeasti ja tasoittunut noin tunnin käytön jälkeen. Laitteiden ollessa pois päältä ne eivät ole lähettäneet signaalidataa, jolloin niistä ei olla voitu myöskään laskea luottamuspisteitä.



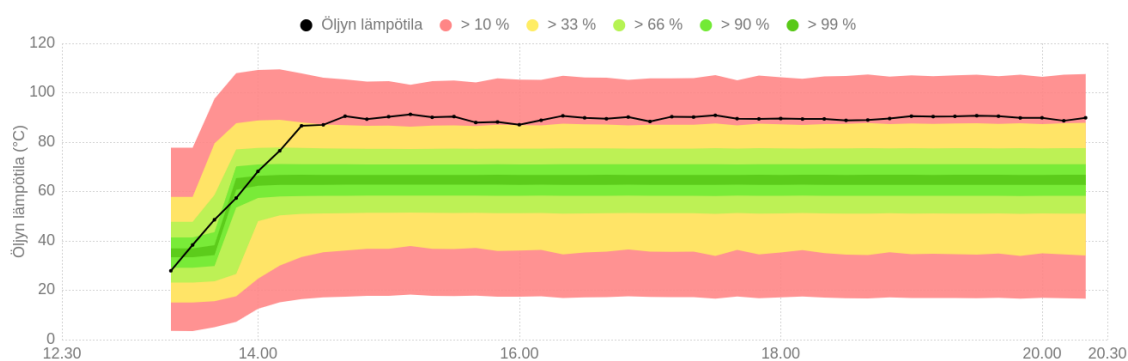
KUVIO 9. Öljyn lämpötila ja sumean mallin luottamuspistejakauma viiden päivän ajalta

Tarkastellessa laitteen yhtä yhtäjaksoista käyttökertaa huomataan kuviossa 10, että lämpötilan tasaannuttua mallin luottamuspistejakauma jakaantuu noin 20-100 °C alueelle. Kuviosta on myös huomattavissa, että laitteen öljyn lämpötila on vastannut mallin luottamuspistejakaumaa ollen pääasiassa >90 prosentin alueella.



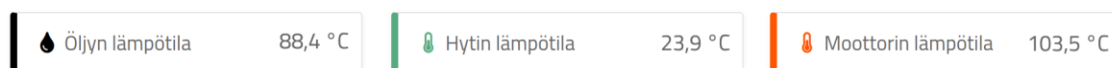
KUVIO 10. Laitteen normaali öljyn lämpötila verrattuna malliin luottamuspistejakaumaan

Kuvion 11 esimerkissä nähdään selvä poikkeus öljyn lämpötilassa. Lämpötila on kasvanut lähes oletetulla tavalla laitteen käynnistyttyä, mutta on noussut noin 30 °C astetta lämpimämmäksi kuin vastaavilla laitteilla. Selitettävän signaalin arvo sijaitsee luottamuspistejakauman kahden pienimmän tason raja-alueella. Tämä voitaisiin määrittää poikkeukseksi, koska signaaliarvo ei ole vastannut sumeaa mallia ja sitä kautta muita mallinnettuja laitteita.



KUVIO 11. Laitteen korkea öljyn lämpötila verrattuna mallin luottamuspistejakaumaan

Valitsemalla kaaviosta jonkin datapisteen näytetään kyseisen ajankohdan signaalien raakadata-arvot kaavion alapuolella (kuva 2). Arvoista voidaan päätellä mistä syystä poikkeus on mahdollisesti johtunut ja miksi signaali ei ole vastannut muiden laitteiden arvoja.



KUVA 2. Valitun ajankohdan signaalien raakadata-arvot

Visualisoinnissa käytettäviä selittäviä signaaleja on mahdollista poistaa käytöstä yksitellen. Rajaamalla pois mallin käytöstä eri signaaleja voidaan tutkia mitkä signaalit vaikuttavat toisiinsa. Poistettaessa selittävästä signaaleista moottorin lämpötila -signaalin, nähdään kuviossa 12 kuinka mallista ei enää saada laskettua tarkkaa luottamuspistejakaumaa. Tämä johtuu siitä, että jäljelle jäävä signaali, hytin lämpötila, ei pysty selittämään yksinään tai ollenkaan öljyn lämpötilaa.



KUVIO 12. Sumea malli ei kykene selittämään öljyn lämpötilaa vain hytin lämpötilan avulla

5 POHDINTA

Tässä työssä toteutettiin ensimmäisen vaiheen sumean mallin visualisointi. Käytölliittymästä saadun toimeksiantajan sisäisen positiivisen palautteen perusteella voidaan katsoa työn onnistuneen hyvin ja vastaavan odotuksia.

Työn merkitys on ja tulee olemaan toimeksiantajalle merkittävä sekä myynnin että tulevaisuuden jatkokehitystyön osalta. Toimeksiantajan asiakkaita kiinnostaa nopeampi ja helpompi tapa omien laitteidensa käyttäytymisen visualisointiin ja sitä kautta datan parempaan ymmärtämiseen. Sumeaa mallinnusta on hyvin vähän käytetty IoT-laitteiden tuottaman datan mallintamiseen eikä kaikki mahdollisuuksia ole vielä tiedossa.

Yhtenä isona tulevaisuuden kehityskohteena sumealle mallinnukselle on anomalioiden eli poikkeuksien tunnistaminen laitteiden signaalidatasta. Tunnistamalla erilaisia poikkeuksia ja niiden mahdollisia lähteitä voidaan ennustaa laitteen tarvitsevan esimerkiksi huoltoa tietyn ajan sisällä tai ilmoittaa, mikäli laitetta mahdollisesti käytetään väärin.

Kuten Keski-Heikkilä (2021, 32) työssään kuvaa voidaan poikkeuksien tunnistukseen käyttää mallista tehtyä ennustetta, vertaamalla mallin tuottamia painoarvoja signaalin raakadataan tai yhdistää manuaalisesti sumeiden sääntöjen pohjalta erilaisia sääntökokonaisuuksia. Yksinkertaisilla sääntökokonaisuuksilla saadaan helposti tunnistettua poikkeuksia. Esimerkkinä poikkeukseksi voidaan luokitella moottorin korkea lämpötila samaan aikaan, kun moottorin nopeus on matala.

Toisena kehityskohteena on kehittää mahdollisuus verrata saman tyyppisten laitteiden tai laitejoukkojen toimintaa keskenään sumean mallin avulla. Laitteiden valmistaja voisi haluta tietää kuinka eri laitteet vertautuvat toisiinsa ja miten niitä käytetään eri kulttuureissa ja ympäristöissä. Seuraavissa kappaleissa on pohdittu mahdollisia muita visualisointeja, joita voitaisiin käyttää työssä toteutetun aikasarjakaavion rinnalla.

5.1 Poikkeusilmoitukset

Työssä toteutettu aikasarjakaavio on hyvin teknisen tason toteutus eikä kuvaa laitteen tai laitejoukon tilaa selkeästi. Tulevaisuuden tavoitteena on antaa käyttäjälle mahdollisimman yksinkertainen ilmoitus toimivatko laitteet oletetulla tavalla. Tiedon välittäminen pitäisi tapahtua ns. yhdellä vilkaisulla, jolla käyttäjä näkee heti, jos jokin on pielessä.

Kuvassa 3 on esimerkki, miten laitteiden signaaleissa tapahtuneita poikkeuksia voitaisiin ilmaista selkeällä värillä sekä lyhyellä selitetekstillä. Punainen väri edustaa kriittistä tai akuuttia poikkeamaa, joka on tapahtunut lyhyen ajan sisällä. Keltaisella voidaan taas kuvata satunnaisesti tai harvemmin tapahtuneita poikkeamia, joihin saattaisi olla hyvä käyttäjän reagoida. Poikkeuksista voitaisiin myös lähettää ilmoitus käyttäjille esimerkiksi sähköpostilla ennakoivana huolto-toimenpiteenä.



Laite A:n **Moottorin lämpötila** on ollut poikkeuksellisesti *erittäin suuri* viimeisen kahden päivän ajan!



Laite A:n **Öljynpaine** on ollut satunnaisesti *korkea* viimeisen viikon aikana!



Laite B:n **Öljynpaine** on ollut usein *kohonnut* viimeisen 15 päivän aikana!

KUVA 3. Laitejoukon poikkeuksien tuottamia ilmoituksia

Tilanteissa, joissa laitteiden todetaan toimivan oikein eikä poikkeuksia olla havaittu voitaisiin esittää yksinkertainen OK-merkki sekä lyhyt seliteteksti (kuva 4).



Ei havaittuja poikkeuksia kohderyhmän laitteilla

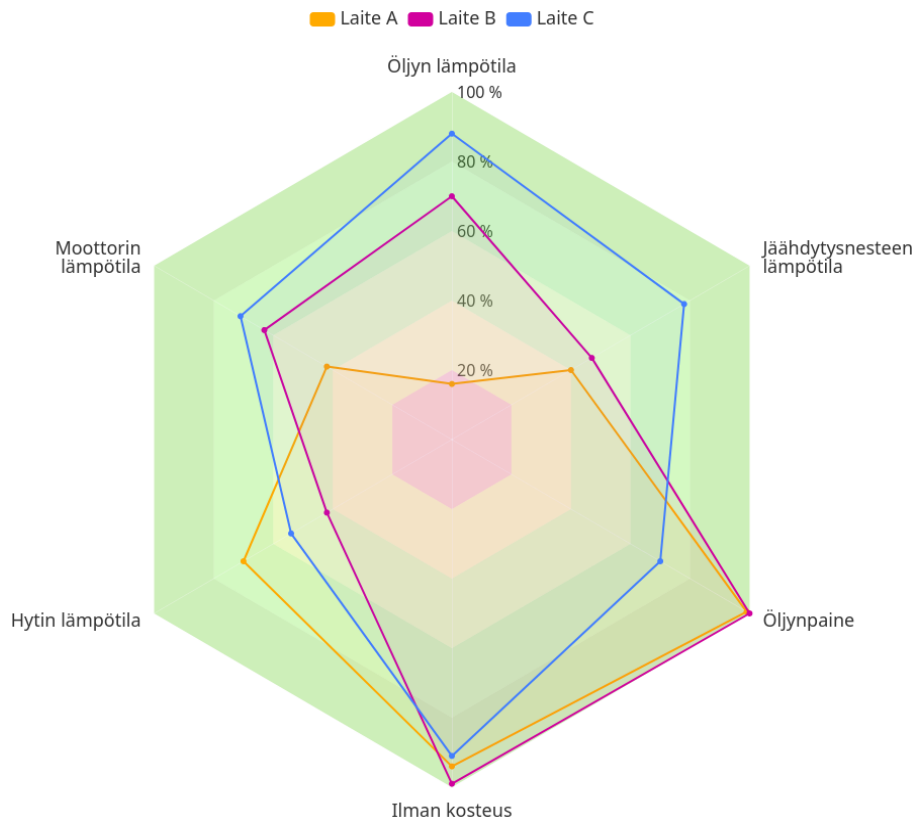
KUVA 4. Laitejoukolla ei ole havaittu poikkeamia -ilmoitus

5.2 Tutkakaavio

Edellä esitetyt ilmoitusmuotoiset kuvaajat eivät anna tarkkaa tietoa, vaan käyttäjälle pitää tarjota myös helppo mahdollisuus tutustua tarkemmin laitteen toimintaan tai mahdollisesti havaittuihin poikkeuksiin. Helposti luettavan yhteenvedon toteuttaminen saadaan aikaan poistamalla kaavioista aika-akseli. Aikaväliä kuitenkin käytetään mallin laskennassa sekä signaalien raakadatojen hakemisessa tietokannasta, vaikka sitä ei esitetä kaavioissa. Alla esitetyissä kaavioissa on oletettu, että visualisoitavan aikavälin hallinta tapahtuu kaavion ulkopuolella.

Laajempaa informaatiota varten voidaan käyttää laitteen tai laitejoukon signaalien visualisointiin niin sanottua tutkakaaviota. Tutkakaaviota käytetään yleisesti vertaillen muuttujia keskenään etsittäessä poikkeavuuksia muuttujien välillä. Tutkakaaviot ovat myös hyödyllisiä tehokkuuden tai suorituskyvyn esittämiseen, missä vertaillaan mitkä muuttujat ovat suoriutuneet hyvin ja mitkä taas huonosti. (The Data Visualisation Catalogue n.d.)

Kuviossa 13 taustalla on piirrettynä sumean mallin luottamuspistejakauman tasot samaan tapaan kuin kuvion 7 aikasarjakaaviossa värein punaisesta vihreään. Jokainen kaavion akseli edustaa yhtä signaalia ja sen prosentuaalista mahdollisuutta, jolla se on sijainnut jakaumatason alueella. Koska tutkakaaviossa signaalista voidaan esittää vain yksi arvo, on aikavälin jokaisen raakadatan datapisteen prosentuaalinen mahdollisuus verrattuna luottamuspisteisiin pitänyt tiivistää keskiarvolliseksi.

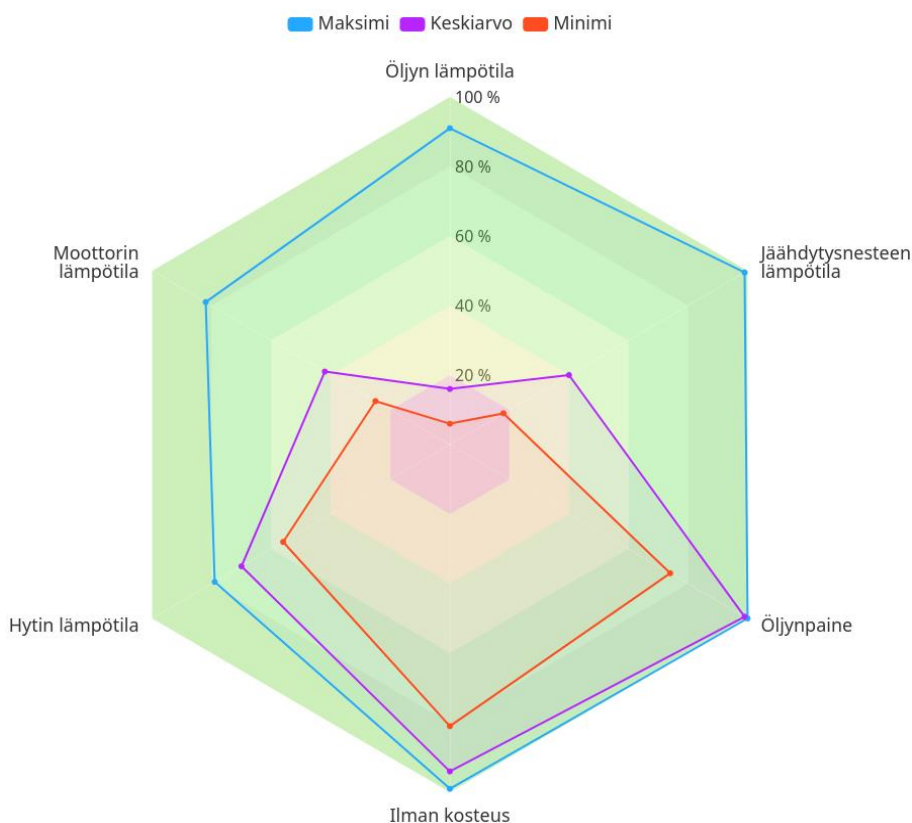


KUVIO 13. Laitejoukon sumean mallin signaalit verrattuna luottamuspistejakaumaan

Kuviossa on nähtävissä selvää eroavaisuutta laite A:n toiminnassa muihin mallin laitteisiin nähden. Laite A:n öljyn lämpötila vastaa huonosti mallista laskettua luottamuspistejakaumaa valitulla ajanjaksolla ja on näin ollen poikkeus muihin laitteisiin nähden. Myöskin laite A:n moottorin sekä jäähdytysnesteen lämpötila on vastannut visualisoitavalla ajanjaksolla muita laitteita heikommin sumeaa mallia. Kunnossapidonhenkilö tässä tapauksessa voisi käydä tarkistamassa työssä kehitetyltä aikasarjakaaviolta tarkemmin, miten laite A on käyttäytynyt eri ajan hetkinä.

Signaalin luotettavuuden esittäminen sumeaan malliin verrattuna keskiarvolla laskentatavalla saattaa tuottaa huonoja tuloksia signaalidatasta riippuen. Toisena vaihtoehtona kaaviossa voidaan myös esittää luottamusten minimiarvot, joka kertoo huonoimman mahdollisen tuloksen, minkä signaali on saavuttanut valitulla aikajaksolla.

Yhden laitteen signaalien tarkempia arvoja tarkastellessa voidaan tutkakaaviossa (kuvio 14) esittää signaaliarvojen luottamusten prosentuaaliset minimi-, maksimi- ja keskiarvot. Kaaviosta on helposti luettavissa signaalien käyttäytymisen arvoväli sekä luottamuspistealue mitä signaali on keskimäärin vastannut valitulla aikaväliltä.



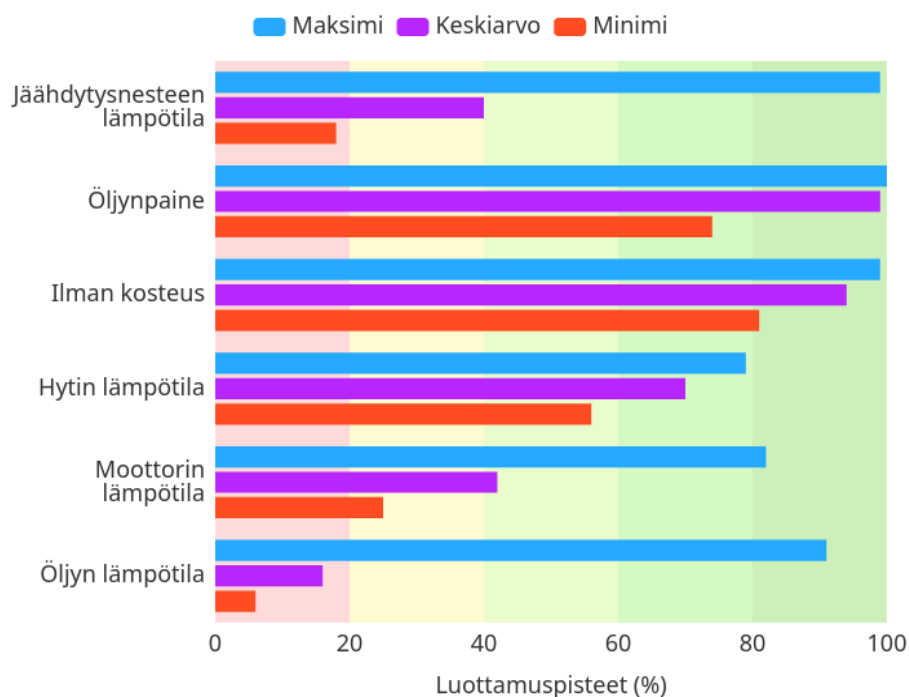
KUVIO 14. Yhden laitteen signaalien toteutumat verrattuna sumean mallin luottamuspistejakaumaan

Tutkakaaviosta nähdään helposti eri muuttujien väliset poikkeamat toisiinsa verrattuna. Kaavion ongelmaksi koituu helposti liian monta päällekkäistä pistesarjaa tai liian monta akselia, jonka jälkeen kaaviosta tulee heikosti luettava. Hyvänä tapana olisi näyttää korkeintaan viisi pistesarjaa ja kaavion koosta riippuen noin 10 akselia maksimissaan kerrallaan.

5.3 Pylväskaavio

Kuvion 14 tutkakaavion sijasta samat arvot voidaan näyttää myös perinteisemmällä tavalla pylväskaaviona (kuvio 15). Kaavioon piirretään jokaisen signaalin

prosentuaalinen mahdollisuus luottamuspisteiden toteutumiseen minimi-, maksimi- ja keskiarvoilla varten oma pylväänsä, jotka ryhmitellään signaalikohtaisesti. Pylväskaavion taustalle voidaan vaaka-akselille piirtää eri luottamuspistetasojen värit punaisesta vihreään kuten yllä olevissa tutkakaavioissa on toteutettu.



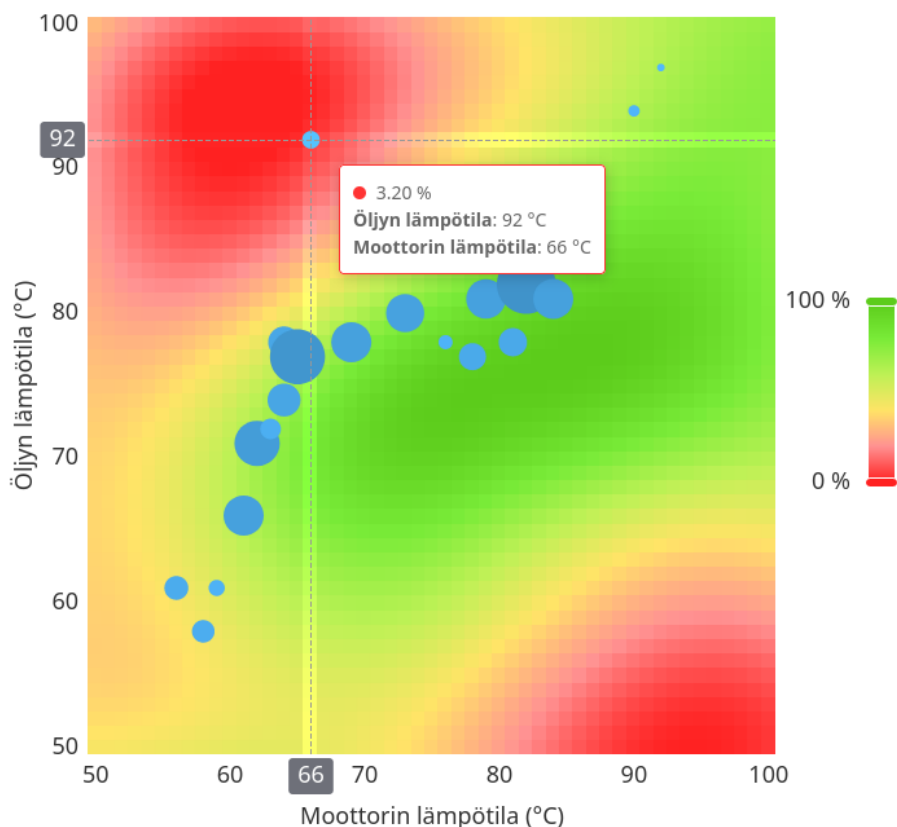
KUVIO 15. Laittejoukon sumean mallin signaalit verrattuna luottamuspistejakaumaan pylväskaaviona

Pylväskaavion hyötynä on sen parempi visualisointi ja hahmottamiskyky tilastollisesta näkökulmasta. Kuitenkin sumean mallin signaalien visualisoinnissa pylväskaavio ei ole riittävän selkeä, että siitä saisi ymmärryksen signaalien toiminnasta ensimmäisellä vilkaisulla. Kaaviossa on liikaa isoja elementtejä ja värejä, jotka häiritsevän sen lukemista eikä selkeää vertailua ole helppo tehdä.

5.4 Pistelämpökarttakaavio

Kolmas tapa mallin ja signaalien visualisointiin on yhdistää piste- ja lämpökarttakaavio (kuvio 16) toistensa kanssa, jossa pisteet ovat signaalien raaka-arvoja ja taustalla oleva värikartta mallin luottamuspistejakauma. Taustalle piirtyvä värikartta punaisesta vihreään kuvaa muiden laitteiden tuottamien arvojen oletusta samaan tapaan kuin työssä tehdyssä aikasarjakaaviossa.

Kuvion 16 esimerkissä pystyakselilla on öljyn lämpötila signaalin ja vaaka-akselilla moottorin lämpötila signaalin raakadata-arvot. Jokaista datapistettä voidaan myös kuvata eri kokoisella pisteellä, joka kuvastaa painoarvoa, kuinka usein signaalit ovat olleet kyseisissä arvoissa. Taustalla olevan värikartan ansiosta poikkeavuudet signaaleissa on helppo havaita.



KUVIO 16. Pistelämpökarttakaavio laitteen moottorin ja öljyn lämpötilan vertailuun

Lämpökartan hyödyntäminen sumean mallin visualisointiin on loogista, koska se kuvaa selkeästi aluetta, jolla muut laitteet ovat toimineet. Yllä esitetty lämpökartan ja pistekaavion yhdistelmä ei itsessään riitä poikkeuksen diagnosointiin vaan vaatisi esimerkiksi pistettä painamalla avautuvan lisätietoikkunan, jossa listataan ajankohdat, jolloin poikkeus on tapahtunut.

Kaavion ongelmana on siinä esitettävän tiedon suppeus, koska akseleille voidaan valita vain kaksi signaalia. Sumea malli voi sisältää useita kymmeniä eri muuttujia, joiden välisten vuorovaikutuksien ymmärtäminen olisi tärkeää.

LÄHTEET

Grandperrin, J. 2021. How to use confidence scores in machine learning models. Towards Data Science. Verkkosivu. Viitattu 12.4.2022. <https://towardsdatascience.com/how-to-use-confidence-scores-in-machine-learning-models-abe9773306fa>

IBM Cloud Education. 2021. What is a REST API? Verkkosivu. Viitattu 20.4.2022. <https://www.ibm.com/cloud/learn/rest-apis>

Keski-Heikkilä, T. 2021. Multivariate fuzzy modelling of time-series data. Kone-tekniikka. Aalto-yliopisto. Opinnäytetyö. Viitattu 16.3.2022. <http://urn.fi/URN:NBN:fi:aalto-202109059012>

Li, D., Mei, H., Shen, Y., Su, S., Zhang, W., Wang, J., Zu, M., Chen, W. 2018. ECharts: A declarative framework for rapid construction of web-based visualization. Visual Informatics 2 (2), 136–146. Viitattu 4.5.2022. <https://doi.org/10.1016/j.visinf.2018.04.011>

The Data Visualisation Catalogue. n.d. Radar Chart. Verkkosivu. Viitattu 25.4.2022. https://datavizcatalogue.com/methods/radar_chart.html

Timescale. 2020. What the heck is time-series data (and why do I need a timeseries database)? Verkkosivu. Viitattu 20.3.2022. <https://www.timescale.com/blog/what-theheck-is-time-series-data-and-why-do-i-need-a-time-series-databasedcf3b1b18563/>

Timescale. n.d. Timescale Docs. Verkkosivu. Viitattu 20.3.2022. <https://docs.timescale.com/timescaledb/latest/>

Vainio, A., Valtonen, M. & Vanhala, J. 2008. Proactive Fuzzy Control and Adaptation Methods for Smart Homes. IEEE Intelligent Systems 23 (2), 42–49. Viitattu 4.5.2022. <https://doi.org/10.1109/MIS.2008.33>

Wolf, T., Gutmann, B., Weber, H., Ferré-Borrull, J., Bosch, S., & Vallmitjana, S. 1996. Application of fuzzy-rule-based postprocessing to correlation methods in pattern recognition. Applied Optics 35 (35), 6955–6963. Viitattu 16.3.2022. <https://doi.org/10.1364/ao.35.006955>