

Poikkeavuuksien havaitseminen etäluettavien sähkómittareiden tuntisarjoista

Sami Lehtinen

Pro gradu –tutkielma



ITÄ-SUOMEN YLIOPISTO

Tietojenkäsittelytieteen laitos

Tietojenkäsittelytiede

Kesäkuu 2019

ITÄ-SUOMEN YLIOPISTO, Luonnontieteiden ja metsätieteiden tiedekunta,
Joensuu
Tietojenkäsittelytieteen laitos
Tietojenkäsittelytiede

Opiskelija, Sami Lehtinen: Poikkeavuuksien havaitseminen etäluettavien
sähkömittareiden tuntisarjoista
Pro gradu –tutkielma, 68 s., 1 liite (7 s.)
Pro gradu –tutkielman ohjaajat: FT Pasi Fränti ja Niklas Syrén
Kesäkuu 2019

Tiivistelmä:

Tutkielma käsittelee etäluettavien sähkömittareiden virhekytkentöjen havaitsemista mittareiden tuottaman mittausdatan perusteella. Väärin kytketty sähkömittari voi vääristää sähkömittarin mittaamaa sähkönkulutusta, jolloin asiakasta laskutetaan väärin perustein. Oikein toimivat sähkömittarit ovat pohja tasa-arvoiselle ja luotettavalle sähkönjakelualustalle. Tutkielman aineistona käytetään lähes 13 000 sähkömittarin yhden päivän aikana tuottamaa mittausdataa, jotka on kerätty vuoden 2018 lopulla. Tutkimusaineiston pohjalta toteutetaan testejä, joissa tarkoituksena on tutkia kuinka onnistuneesti eri poikkeavuuksien havaitsemistekniikat tunnistavat erityyppisiä virhekytkentöjä ja miten eri mittausdatan esitysmuodot sekä datan esiprosessointi vaikuttavat tunnistustarkkuuteen. Testauksen tulosten perusteella annetaan suosituksia, mitkä tekniikat ja millainen mittausdatan esitystapa toimivat parhaiten virheellisesti kytkettyjen sähkömittareiden tunnistamiseen.

Avainsanat: poikkeavuuksien havaitsemistekniikka, etäluettava sähkömittari, sähkö, sähköteho, virhekytkentä

UNIVERSITY OF EASTERN FINLAND, Faculty of Science and Forestry, Joensuu
School of Computing
Computer Science

Student, Sami Lehtinen: Detecting anomalies from smart meters' energy series
Master's Thesis, 68 p., 1 appendix (7 p.)
Supervisors of the Master's Thesis: PhD Pasi Fränti and Niklas Syrén
June 2019

Abstract:

The thesis is about detecting incorrectly installed electricity meters based on the measurement data. Incorrectly installed electricity meter can distort the measured electricity consumption and therefore customer can be incorrectly invoiced. Correctly working electricity meters are the base of equal and reliable electricity power transmission platform. The experiments in the thesis use energy series data collected from almost 13 000 electricity meters from one day in the end of 2018. Purpose of this thesis is to study how successfully different anomaly detection techniques recognize the data produced by incorrectly installed meters and to study how different data presentations affects to the results. Based on the test results, recommendations are made on which technologies and the type of data presentation are best suited for identifying incorrectly installed electricity meters.

Keywords: outlier detection, smart meter, electricity, electrical power, error-wiring

Lyhenneluettelo

AMI	Advanced Metering Infrastructure
A+ / A-	Electricity meter's energy register for active power
R+ / R-	Electricity meter's energy register for reactive power
1P	One-Phase electricity meter
3P3W	Three-Phase Three-Wire electricity meter
3P4W	Three-Phase Four-Wire electricity meter
I/O	Input / Output
PCA	Principal Component Analysis
OCSVM	One-Class Support Vector Machine
KNN	k-Nearest Neighbours
HBOS	Histogram-Based Outlier Score
OTS	One-Time Sampling
EE	Elliptic Envelope
LOF	Local Outlier Factor
CBLOF	Cluster-Based Local Outlier Factor
IF	Isolation Forest
MinA+	Search records with lowest A+ register proportion
MaxA-	Search records with highest A- register proportion
MinSum	Search records with lowest combined active power consumption
MMS	Min-Max Scaling
SS	Standard Scaling
RS	Robust Scaling

Sisällysluettelo

1. Johdanto	1
2. Etäluettavat sähkömittarit	2
2.1. Etäluettavat sähkömittarit yleisesti	2
2.2. Järjestelmän osat	3
2.3. Hyödyt ja haasteet.....	7
2.4. Sähkönmittaus.....	9
2.5. Mittarityypit	12
2.6. Vika-asennus.....	14
3. Poikkeavuuksien havaitseminen	19
3.1. Poikkeavuuksien havaitsemisen perusteet	19
3.2. Poikkeavuuksien havaitseminen älykkäässä sähkönmittauksessa	22
3.3. Tekniikoita	23
3.3.1. Principal Component Analysis	24
3.3.2. One-Class Support Vector Machine	24
3.3.3. k-Nearest Neighbors	25
3.3.4. Histogram-Based Outlier Score	25
3.3.5. One-Time Sampling.....	26
3.3.6. Elliptic Envelope.....	26
3.3.7. Local Outlier Factor	26
3.3.8. Cluster-Based Local Outlier Factor	27
3.3.9. Isolation Forest	27
3.4. Intuitiiviset sovelluslähtöiset tekniikat	28
3.5. Datan esiprosessointi	29
4. Virhekytkentöjen tunnistaminen energiasarjoista.....	31
4.1. Tutkimuskysymys	31
4.2. Tutkimussuunnitelma.....	31
4.3. Datalähteet	33
4.4. Datan esitysmuoto.....	34
4.5. Virheellisen datan simulointi	36
5. Tulokset	45
5.1. Kaikkien johtimien I/O kytketty ristiin.....	48
5.2. 3P4W-mittari, kaksi I/O kytketty ristiin	50
5.3. 3P4W-mittari, yksi I/O kytketty ristiin	52
5.4. 3P3W-mittari, yksi I/O kytketty ristiin	53
5.5. 3P3W-mittari, yksi I/O kytketty ristiin, epätasainen kuorma	55
5.6. Virheellisen datan määrän vaikutus tunnistustarkkuuteen.....	56
6. Pohdinta	57
7. Yhteenveto	63
Viitteet	65

Liite 1 69

1. Johdanto

Sähkön tuotanto, jakelu sekä kulutuksen mittaaminen ovat tärkeä osa maiden infrastruktuuria. Täsmällinen ja automatisoitu sähkönkulutuksen mittaaminen luo tasavertaisen ja luotettavan alustan sähkönjakelulle. Sähkön mittaaminen on hyvin huomaamaton operaatio, vaikka se on osa ihmisten jokapäiväistä arkea. Lähes jokaisessa asuinrakennuksessa on nykyään oma kiinteistön tai huoneiston sähkönkulutusta mittaava sähkömittari, joka mittaa sähkönkulutusta tarkasti ja luotettavasti laskutusta varten. Ennen sähköyhtiöt joutuivat käymään paikan päällä lukemassa näiltä mittareilta sekä kulutus että tuottolukemia. Hidas sähkönkulutuslukemien saaminen haittaa viallisten ja väärinkytettyjen mittareiden havaitsemista sekä vaikeuttaa sähkönkulutuksen seuraamista. Nykyään yhä useampaan mittariin on kiinnitetty etäluennan mahdollistava kommunikaatiomoduli, joka mahdollistaa sähkönkulutuslukemien lähettämisen sähköyhtiön luentajärjestelmään. Tämän avulla mittari voi lähettää jatkuvasti kulutuslukemia sekä tietoa mittarin toiminnasta sähköyhtiölle. Energiasarjojen jatkuva lähetys parantaa mittaustarkkuutta ja mahdollistaa väärin kytketyn sekä viallisen mittarin nopeamman tunnistamisen kuin myös sähkönkulutuksen piikkien ennakoimisen tai sähkön dynaamisen hinnoittelun. Etäluettavat sähkömittarit lisäävät myös kuluttajan oikeuksia tarkkojen sähkönkulutuksen historiatietojen ollessa saatavilla. [1][2]

Tämän pro gradu tutkielman tutkimuskysymykset käsittelevät sitä, kuinka hyvin poikkeavuuksien etsiminen sähkömittareiden tunneittain lähettämistä energiasarjoista sopii väärin asennettujen sähkömittareiden havaitsemiseen. Käytettävissä on laitteilta kerran tunnissa lähetetyt arvot A+, A-, R+ sekä R- -rekistereistä, jotka kertovat sähkömittarin mittaamasta *päätöhosta* sekä *loistehosta*. Nämä suureet ovat tärkeimmät sähkömittarin mittaamat arvot ja niiden avulla voidaan laskea kohteen sähkönkulutus sekä sähköntuotto, joiden perusteella kohdetta laskutetaan. Tutkielman kokeellisessa osiossa vertaillaan eri poikkeavuuksien etsintämenetelmien soveltuvuutta väärin asennettujen mittareiden havaitsemiseen.

2. Etäluettavat sähkömittarit

Tässä luvussa tehdään yleiskatsaus nykyaikaiseen sähkömittaukseen. Luku käsittelee älykästä sähkömittausta ja siihen liittyviä peruskäsitteitä. Luvussa käydään läpi mikä etäluettava sähkömittari on, mistä mittausrakenteet koostuvat sekä mitä hyötyjä ja haasteita etäluettava sähkömittaus tuo mukanaan. Luvussa käydään läpi myös sähkömittauksen perusteet, esitellään tavallisimmat sähkömittarityyppit sekä käydään läpi miten eri tavoin sähkömittari on mahdollista asentaa virheellisesti.

2.1. Etäluettavat sähkömittarit yleisesti

Ennen kuin sähkömittarit pystyivät lähettämään kulutuslukemia automaattisesti sähköyhtiöille, jouduttiin sähkömittarin lukemat käymään tarkastamassa paikanpäällä tai asiakkaat joutuivat itse ilmoittamaan sähkömittarinsa kulutuslukemat sähköyhtiölle. Tämä tarkoitti sitä, että sähköyhtiön oli palkattava työntekijöitä, joiden tehtävä oli käydä lukemassa sähköyhtiön sähkömittareiden kulutuslukemia ovelta ovelle tai kirjata asiakkaiden ilmoittamia kulutuslukemia. Tämä prosessi on kallis sekä aikaa vievä. Tämän lisäksi asiakkailta oli mahdollista saada vain pitkän aikavälin kulutuslukemia sen sijaan, että sähkönkulutusta olisi pystynyt seuraamaan reaaliajassa. Muun muassa näiden ongelmien ratkaisemiseksi kehitettiin etäluettava sähkömittari.

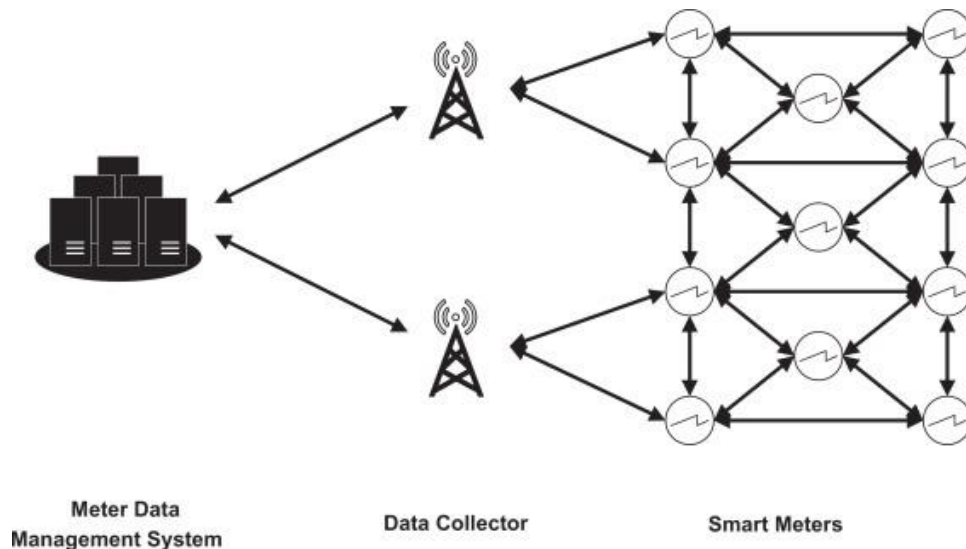
Etälukeminen tarkoittaa käytännössä sitä, että sähköyhtiön ei tarvitse enää käydä paikanpäällä lukemassa kulutuslukemia sähkömittareilta vaan tiedot lähetetään automaattisesti sähköyhtiön luentajärjestelmään tiedonsiirtoverkon avulla. Etäluettavat sähkömittarit ovat yleistymässä ympäri maailmaa ja vanhanaikaisia sähkömittareita ollaan päivittämässä kiihtyvällä vauhdilla etäluettaviin. Esimerkiksi Iso-Britanniassa on tavoitteena asentaa, jokaiseen kiinteistöön etäluettava sähkömittari vuoteen 2020 mennessä. [3] Suomessa puolestaan voimaan astui laki

vuonna 2013, jonka mukaan jakeluverkonhaltijan sähkökäyttöpaikoista vähintään 80 % tuli perustua tuntimittaukseen sekä mittauslaitteiston etäluentaan.¹ Vuonna 2014 Suomessa 98 % sähkömittareista oli etäluettavia. [4] Tulevaisuudessa yhä useampi maa on uusimassa sähkömittausinfrastruktuuriaan ja etäluettavat sähkömittarit tulevat yleistymään entisestään.

2.2. Järjestelmän osat

Advanced metering infrastructure (AMI) eli kehittynyt mittausinfrastruktuuri on keskeinen käsite nykyajan sähkökulutuksen mittauksessa. AMI tarkoittaa käsitteenä kehittyntä mittausinfrastruktuuria, joka tukee mittareiden etäluentamahdollisuutta. Kehittynyt mittausinfrastruktuuri voi viitata niin sähkön, kaasun, veden tai muun hyödykkeen kulutuksen mittaukseen, mutta tässä osiossa käsittelemme sitä nimenomaan sähkönmittauksen kontekstissa. AMI on integroitu järjestelmä, joka koostuu etäluettavista mittareista, kommunikaatioverkosta ja datan hallintajärjestelmistä. Se mahdollistaa kaksisuuntaisen kommunikoinnin mittarin sekä luentajärjestelmän välillä sekä automatisoidut mittausoiminnallisuudet. Kehittyneen mittausinfrastruktuurin avulla monet manuaalista työtä vaatineet operaatiot kuten sähkökulutuslukemien lukeminen mittarilta, mittarin lukeman vääristämisen, mittarin vioittumisen tai sähkökatkoksen havaitseminen sekä jännitteen mittaus voidaan automatisoida. [2][5]

¹ <https://www.finlex.fi/fi/laki/alkup/2009/20090066> (28.10.2018)



Kuva 1: Esimerkki kehittyneestä mittausinfrastruktuurista [6]

AMI sisältää tyypillisesti kolme komponenttia: etäluettavat sähkömittarit, tietoliikenneverkon sekä mittausdatan hallintajärjestelmän, kuten kuvassa 1 on havainnollistettu. Etäluettavat sähkömittarit ovat älykkään mittausympäristön ydin. Etäluettavan sähkömittarin tehtävä on mitata asiakkaan sähkönkulutusta tietyn intervallin kuten 5, 15, 30 tai 60 minuutin välein. Sähkönkulutuksen mittauksen lisäksi mittari voi mitata muun muassa jännitelukemia, jotka kertovat sähköverkon tilasta sekä pitää kirjaa siitä, mitkä mahdolliset lisäpalvelut asiakkaalla on käytössä. Etäluettavat sähkömittarit lähettävät nämä tiedot takaisin sähköyhtiön luentajärjestelmään, laskutusta, analysointia, prosessointia, sähköverkon kunnan tarkastusta ja dynaamista hinnoittelua varten.

Etäluettavien mittarien uudet toiminnallisuudet vaativat tietoliikenneverkon, joka mahdollistaa suurten datavirtojen siirtämisen mittareilta tarkasti ja luotettavasti. Tietoliikenneverkko yhdistää etäluettavat mittarit dataa vastaanottavaan pääjärjestelmään (*head-end system*), joka hallitsee kommunikaatiota mittareiden sekä muiden tietojärjestelmien välillä. Tietoliikenne mittareiden sekä pääjärjestelmän välillä voidaan toteuttaa niin langattomilla kuin langallisilla teknologioilla ja niitä voidaan yhdistää keskenään käyttötapauksesta riippuen. Useat sähköyhtiöt kehittävät omat tietoverkkoratkaisunsa ja valitsevat käytettävät teknologiat mitattavan alueen erityispiirteiden kuten asukastiheyden ja alueen pinnanmuotojen mukaan. Esimerkiksi kerrostalossa voi olla järkevää yhdistää mittarit langallisella

teknologialla, mittareiden ollessa hyvin lähekkäin, mutta haja-asutusalueella langaton kommunikointi on järkevämpi ratkaisu.

Langallisia tekniikoita, joita voidaan käyttää tiedonsiirtoon, ovat muun muassa valokuitukaapeli, sähköverkko (PLC), puhelinverkko tai digitaalinen tilaajayhteys (DSL). Langattomaan tiedonsiirtoon käytetään useimmiten radiotaajuutta (RF), jonka avulla dataa voidaan siirtää solmuverkoissa tai matkapuhelinverkoissa. Solmuverkko mahdollistaa topologian luomisen mittareiden kesken, jolloin vain osa verkon mittareista lähettää dataa pääjärjestelmään esimerkiksi matkapuhelinverkon kautta. Tällöin osa mittareista toimii orjina (*slave*), osa reitittiminä (*router*) ja osa isäntinä (*master*). Orjat lähettävät mittausdataa reitittimille tai isännille joista reitittiminä toimivat mittarit voivat lähettää mittausdataa edelleen isännille RF-verkon kautta. Isännät ovat mittareita, jotka ovat varustettu komponentilla, joka mahdollistaa internetyhteyden ja näin ollen yhteyden edelleen pääjärjestelmään. Tämä vähentää tiedonsiirtokustannuksia ja mukautuva solmuverkko auttaa palautumaan tilanteista, joissa tiedonsiirtoverkkoon tulee häiriö. [5] Oikean kommunikaatioteknologian valitsemiseen vaikuttaa yleisesti kaistanleveys, latensi, hinta, luotettavuus, kantama ja turvallisuusvaatimukset, jotka tulisi ottaa huomioon teknologioita valittaessa. Tietoverkkoratkaisuissa ei ole olemassa yhtä standardia vaan valinnat tulisi tehdä tarpeesta riippuen. Yleinen ja suosittu tapa on kuitenkin jakaa tietoverkko kahteen osaan siten, että ensimmäinen osa yhdistää sähkömittarit välipisteenä toimivaan solmukohtaan hitailla ja rajallisilla yhteyksillä kuten langattoman solmuverkon avulla ja toinen kerros yhdistää solmukohtat yhtiön pääjärjestelmään nopeilla, suuriin tiedonsiirtomääriin pystyvillä tekniikoilla. [5]

Pääjärjestelmän tehtävä kehittyneessä mittausinfrastruktuurissa on vastaanottaa jatkuvana virtana saapuvaa mittausdataa sekä lähettää operaatiokäskyjä. Pääjärjestelmän tärkein osa on *meter data management system* (MDMS), joka käsittelee ja tallentaa mittareilta saapuvaa mittausdataa ja tarjoaa rajapinnat muille dataa käyttäville järjestelmille, jotka eivät ole kykeneviä vastaanottamaan suuria määriä mittausdataa, kuten laskutusjärjestelmälle. Pääjärjestelmän tehtävä on myös mahdollistaa käskyjen lähettäminen mittareille kaksisuuntaisen kommunikoinnin mahdollistavan tietoverkon kautta. Näitä käskyjä voi olla esimerkiksi käsky

mittauslukemien lähetykselle, ohjelmistopäivityksen asennus, käsky jännitelukemien lähetykselle tai jokin muu AMI-operaatio. [2][5]

Itse AMI koostuu vain mittaukseen liittyvästä arkkitehtuuriratkaisuista ja pääjärjestelmästä. Pääjärjestelmän rinnalla voi toimia kuitenkin monia muita siihen integroituja järjestelmiä. Näitä järjestelmiä ovat *customer information system (CIS)*, *laskutusjärjestelmä*, *geographic information system (GIS)*, *outage management system (OMS)* sekä *distribution management system (DMS)*. [2][5] Laskutusjärjestelmä, web-portaali, CIS ja muut järjestelmät voivat hyödyntää esiprosessoitua mittausdataa MDMS:n kautta.

Laskutusjärjestelmä prosessoi mittarien lähettämää intervallidataa ja luo asiakkaiden laskut automaattisesti MDMS:n tarjoaman datan perusteella. CIS:n tehtävänä on hoitaa kaikki asiakassuhteisiin liittyvät operaatiot laskutushistoriasta, asiakaspalvelusta, uusien sähkösopimusten tekoon. GIS:n tehtävänä on hallita, analysoida ja esittää paikkatietoa kartalla. GIS voidaan integroida muiden järjestelmien kanssa, jolloin kartalla voidaan visualisoida esimerkiksi vikaantuneiksi ilmoitettuja mittareita, sähkökäyttöastetta tai asiakasdataa. GIS helpottaa vaikealukuisen datan tulkintaa visualisoimalla se ihmiselle helpommin ymmärrettävään muotoon. GIS:n avulla voidaan havaita esimerkiksi joitakin ongelma-alueita, joissa mittarit vioittuvat tavallista herkemmin tai mittarien välinen langaton kommunikaatio ei toimi oletetusti haastavien maan pinnanmuotojen vuoksi. Tällaisia tapauksia on hyvin vaikea havaita raakadatatista.

OMS:n tehtävänä on yksinkertaisesti tarkastella onko mittari toiminnassa vai ei. Tämän tiedon avulla saadaan helpommin selville, millä alueilla sähkökatkoksia on ilmennyt. Integroimalla OMS-järjestelmä GIS-järjestelmään, huoltotiimien lähettäminen vikakohteisiin on helpompaa, kun vikakohteiden sijainti on tarkasti tiedossa ja huoltotiimit voivat itse tarkastella vikakohteita kartalla. DMS on järjestelmä, joka monitoroi sekä hallitsee koko sähköverkon tilaa. Se toimii päätöksenteon tukena niille, jotka ylläpitävät sähköverkkoa ja ovat vastuussa toimitetun sähkön laadusta. DMS:n tarkoitus on parantaa sähkötoimituksen luotettavuutta vähentämällä sähkökatkoja, minimoimalla sähkökatkon pituudet sekä ylläpitämällä toimitettavalle sähkölle hyväksyttävää jännite- ja taajuustasoa. U.S.

Department of Energyn 2016 Advanced Metering Infrastructure and Customer Systems raportin [5] mukaan käytetyimmät AMI:in integroidut tietojärjestelmät olivat laskutusjärjestelmä ja CIS. Laskutusjärjestelmä oli käytössä 58 kyselyssä mukana olleessa AMI-projektissa 60:stä kun puolestaan CIS oli käytössä 53:ssa. OMS oli käytössä 37 projektissa kun taas DMS:n käyttö oli melko harvinaista. Se oli käytössä vain kuudessa kyselyyn osallistuneessa projektissa. [5]

2.3. Hyödyt ja haasteet

Vaikka etäluettavien sähkömittareiden asentaminen vanhojen tilalle on kallista, etäluettavien sähkömittareiden tuomat hyödyt ovat suuria. Suurin etäluettavien sähkömittareiden tuoma hyöty asiakkaille on täsmällinen kulutuksen mittaaminen. Näin ollen asiakkaan laskutus perustuu todelliseen sähkönkulutukseen ja arviolaskut sekä tasauslaskut eivät enää ole tarpeellisia ja sähkönkulutuksen laskutus muuttuu huomattavasti selkeämmäksi. Lisäksi kuluttajan oikeudet parantuvat, kun asiakaskohtainen sähkönkulutus ja mahdollinen sähköntuotanto on tallennettu tuntitasolla tai jopa useammin. Ongelmatilanteissa kulutuslukemat ovat tallessa ja asiakas voi itse tarkkailla sähkönkulutustaan sähköyhtiöiden online-palveluiden kautta. Asiaan perehtyneet asiakkaat pystyvät online-palveluista saadun tiedon perusteella saavuttamaan 5-15 % energiansäästöt muuttamalla heidän sähkönkulutustottumuksiaan. [1][2]

Sähkön hintaa on mahdollista muuttaa jopa tuntitasolla, kun sähköyhtiöt saavat reaaliaikaiset kulutuslukemat asiakkailta. Näin ollen asiakkaan on mahdollista valita sähkötuotteita, joiden hinnoittelu perustuu esimerkiksi sähkön tuntitason tukkuhintoihin sähköpörssissä. Tämä laajentaa sähkötuotteiden tarjontaa ja asiakas voi itse valita sopivimman sähkötuotteen itselleen hänen oman sähkönkäytön perusteella. Suurin hyöty etäluettavista sähkömittareista sähköyhtiöille on se, että sähkömittareiden kulutuslukemia ei tarvitse enää käydä lukemassa paikanpäällä, vaan kulutuslukemat lähetetään automaattisesti sähköyhtiölle. Tämä vähentää vilpin mahdollisuutta, kun kulutuslukemia lähetetään lyhyen ajanjakson välein sähköyhtiölle ja poikkeamat mittausdatassa on helppo havaita. [7] Sähköyhtiö saa myös nopeammin tiedon vikatilanteista sekä sähkönkulutuksen piikeistä, kun sähkömittarit lähettävät jatkuvasti tietoa sähköyhtiölle. Tämä nopeuttaa

vikatilanteiden selvittelyä ja mahdollistaa nopeamman reagoinnin sähkökulutuksen piikkeihin. [1][2][8] Reaaliaikainen kulutuksen mittausta avaa myös sähköyhtiöille mahdollisuuden kehittää uusia tuotteita sekä raportointipalveluita. Muun muassa GIS:n avulla voidaan luoda reaaliaikainen kuva sähköverkon tilasta, vikaantuneista laitteista sekä sähkökatkoista. Asiakkaille on mahdollista myydä dynaamiseen hinnoitteluun perustuvia sähkötuotteita ja tarjota sähkökäytön hallintaan ja monitorointiin soveltuvia online-palveluja. [2]

U.S. Department of Energy julkaisi syyskuussa 2016 Advanced Metering Infrastructure and Customer Systems raportin [5], joka käsittelee millaisia vaikutuksia älykkääseen energiamittaukseen tehdyillä investoinneilla on ollut Yhdysvalloissa. Raportin mukaan AMI on alentanut mittaus- ja laskutuskustannuksia huomattavasti. Sen ansiosta tarvitaan vähemmän työvoimaa, laskutus on tarkempaa, kiistat asiakkaiden kanssa ovat vähentyneet ja yleisesti toiminta on tehostunut.

Toiseksi löydökseksi raportti mainitsee asiakkaiden parantuneen kontrolloinnin sähkökulutuksen, siihen liittyvien kustannusten sekä laskutuksen osalta. Asiakkaat voivat hallita ja monitoroida sähkön käyttöönsä paremmin uusien työkalujen kuten web-portaalien ja uusien tekniikoiden avulla, jotka mahdollistavat käyttämään sähköä, kun se on halvimmillaan.

Kolmanneksi raportti mainitsee alentuneet käyttöomaisuusinvestoinnit, jotka johtuvat muun muassa alentuneista sähkökäyttöpiikeistä ja tehostuneesta ylläpidosta, kun sähköverkosta on mahdollista saada reaaliaikainen tilannekatsaus. Kun käytössä on kysyntäpohjainen hinnoittelu, jossa sähkön hinta määräytyy kysynnän ja tarjonnan mukaan, asiakkaat jakavat sähkökulutuksen tasaisemmin eikä suuria piikkejä sähkökäytössä pääse muodostumaan.

Viimeisenä löydöksenä raportti mainitsee alentuneet kustannukset sähkökatkoksista palautumisessa. AMI mahdollistaa reaaliaikaisen monitoroinnin, jolloin sähkökatkoksen sattuessa vika on helpompi paikantaa ja huoltomiehet voidaan lähettää nopeasti paikalle. [5]

Suurien hyötyjen lisäksi etäluettavat sähkömittarit tuovat muutamia uusia haasteita. Näitä ovat muun muassa tietoturvaongelmat, ihmisten yksityisyydensuoja sekä

mittausinfrastruktuurin uudistamiseen vaadittavien investointien suuruus. Koska useat etäluettavat sähkömittarit toimivat langattoman teknologian avulla, se altistaa sähkömittarit ja samalla koko kiinteistön virtalähteen kyberhyökkäyksille mahdollistaen teoreettisesti virtojen katkaisemisen kiinteistöltä. Teoriassa hyökkääjän päästessä sisään AMI-operaatioita lähettävään järjestelmään, hän voi lähettää sähköyhtiön omistamille sähkömittareille mitä tahansa AMI-operaatiokäskyjä. Näistä eniten vahinkoa aiheuttava on sähköjen katkaisemiseen tarkoitettu operaatio, joka katkaisee sähkön kiinteistöltä. Näin ollen järjestelmän tietoturva on oltava erittäin korkealla tasolla, jotta voidaan turvata sähkönjakeluun käytettävän infrastruktuurin toimivuus ja luotettavuus.

Myös mittauslukemien lähettämiseen tarvittavien turvallisten protokollien kehittäminen on ollut ongelmallista sähkömittareiden rajatun laskentatehon sekä mittarin pitkän eliniän vuoksi. Mittarin laskentatehon ollessa hyvin rajallinen salausprotokollan on oltava verrattain kevyt ja se voi olla murrettavissa suuren laskutehon omaavalla laitteistolla. [9] Heikentynyt yksityisyydensuoja on myös herättänyt huolta uudistuneessa sähkönkulutuksen mittauksessa. Jatkuvasti jopa viiden minuutin välein sähköyhtiölle lähetetyt sähkönkulutuslukemat avaavat sähköyhtiöille mahdollisuuden käyttäjien profiloinnille. Sähkönkäytön perusteella voidaan esimerkiksi ennustaa onko asukas kotonaan vai poissa tai onko asukas nukkumassa. Sähkönkäytön perusteella voidaan yrittää profiloida jopa mitä sähkölaitteita kiinteistön asukas käyttää. [10] Tämä tieto on monille tahoille arvokasta kuten velanperijöille ja mainostajille. Mitä harvemmin mittari tallentaa mittaustuloksia, sitä vaikeampi sähkönkäyttäjien profilointi on, mutta samalla se vaikeuttaa sähkönjakelijoita vastaamaan jatkuvasti muuttuvaan sähkönkysyntään ja vaikeuttaa sähköverkon tilan monitorointia.

2.4. Sähkönmittaus

Voimaloissa tuotettu sähkö siirretään sähkön käyttäjille sähköverkon avulla. Sähköverkossa sähkö kulkee ilmajohtoja tai maakaapeleita pitkin suurijännitteisenä *vaihtovirtana*, jonka avulla sähkövirta sekä johtojen vastuksen aiheuttama häviö saadaan pidettyä pienenä. Lähempänä kuluttajia, suurijännitteinen sähkö muutetaan muuntajassa turvallisemmaksi pienjännitteiseksi sähköksi. Tämä sähkö johdetaan

kuluttajien koteihin, jotta he saavat sähkölaitteiden tarvitsemaa sähkötehoa. Sähkömittari asennetaan kuluttajan kotitalouden ja pienjänniteverkon väliin. Näin sähkömittari pystyy mittaamaan kaiken sähköä, joka siirtyy sähköverkosta asiakkaan kiinteistöön tai kiinteistöstä sähköverkkoon.

Koska sähkönsiirto tapahtuu vaihtovirran avulla, tulee sähkömittarin kyetä mittaamaan vaihtovirran kahta tärkeintä tehoa jotka ovat *pätöteho* (P) sekä *loisteho* (Q). Näiden kahden tehon avulla pystytään laskemaan koko virtapiirin käyttämä *näennäisteho* (S). Pätöteho on sähkölaitteiden todellisuudessa käyttämää sähkötehoa. Pätöteho voidaan laskea sähkövirran voimakkuuden I , jännitteen U sekä niiden vaihe-eron kosinin tulona $P = U * I * \cos(\theta)$ ja sen yksikkönä käytetään wattia. Sähkövirta ilmaisee kuinka paljon sähkövarausta johtimen läpi kulkee aikayksikköä kohden ja sen yksikkönä käytetään ampeeria (A). Jännite puolestaan ilmaisee varauseroa kahden pisteen välillä eli sitä kuinka varautunut johdin on. Varauksen yksikkönä käytetään voltia (V). Sähkömittarin tärkein tehtävä on laskea mittarin läpi kulkeman pätötehon määrä, jonka avulla kotitalouksien laskutus tehdään. Useimmiten laskutus tehdään käytettyjen kilowattituntien (kWh) perusteella.

Loisteho on virtapiirissä olevaa värähtelevää tehoa, joka ei itsessään tee työtä. Se vain värähtelee edestakaisin sähkölaitteen sekä siirtoverkon välillä. Erilaiset sähkölaitteet kuluttavat tai synnyttävät eri määriä loistehoa. Esimerkiksi sähkömoottorit kuluttavat loistehoa. Vaikka loisteho ei tee työtä, on tärkeää, että sähköverkossa on sopiva määrä loistehoa. Loistehon vajavaisuus voi vähentää esimerkiksi sähköverkon sähkövirran määrää. Loisteho voidaan laskea sähkövirran, jännitteen ja niiden välisen vaihe-eron sinin tulona $Q = U * I * \sin(\theta)$ ja sen yksikkö on vari, jota käytetään erottamaan loisteho pätötehosta vaikkakin molemmat kuvaavat samaa sähköenergian suuretta. Loistehon kulutus on usein pienille sähkönkäyttäjille kuten omakoti- ja kerrostalotalouksille ilmaista, mutta suurille sähkönkäyttäjille loisteho on maksullista. Suuret teollisuuslaitokset voivat kuluttaa tai tuottaa suuria määriä loistehoa sähkönsiirtoverkkoon. Tämä aiheuttaa sähköyhtiölle kustannuksia, kun loistehoa tarvitsee siirtää sähköverkossa ja siksi loisteho on usein maksullista suurille teollisuuslaitoksille. [11]

Vaihtovirrassa sähkö ei virtaa vain yhteen suuntaan, vaan sähkövirran ja jännitteen määrät vaihtelevat sinimuotoisten signaalien mukaan. Vaihekulma kuvaa sitä kuinka paljon jännitteen sinimuotoinen signaali laahaa tai edistää sähkövirran sinimuotoista signaalia. Sinimuotoisten signaalien vaihtelua on kuvattu kuvassa 2. Mikäli signaalit kulkevat täsmälleen samaan aikaan, jolloin vaihekulma on 0, tarkoittaa se sitä, että kaikki piirissä oleva sähkövirta ja jännite saapuvat kohteeseen yhtä aikaa ja kaikki piirissä oleva sähköteho pystytään hyödyntämään. Tällöin piirissä on vain pätötehoa. Mikäli signaalien vaihekulma on 90 astetta tai -90 astetta, tarkoittaa se, että vaikka piirissä olisi riittävä määrä sähkövirtaa sekä jännitettä, saapuvat ne juuri väärään aikaan sähköä käyttävälle laitteelle, jolloin laite ei voi hyödyntää niitä. Tällöin piirissä on vain loistehoa. 180 asteen ero tarkoittaisi, että sähkö kulkee kohteesta lähteeseen. Vaihekulmasta voidaan siis päätellä sähkövirran suunta, sekä se kuinka varaava kohde on. [11]



Kuva 2: Resistiivinen kuorma (vasemmalla), induktiivinen kuorma (keskellä) ja kapasitiivinen kuorma (oikealla). [11]

Etäluettavilla sähkömittareilla on siis kyettävä mittaamaan sähköyhtiön asiakkaiden kulutettua tai esimerkiksi aurinkopaneeleilla tuotettua pätötehoa sekä sähköverkosta otetun tai tuotetun loistehon määrää. Sähkömittari tallentaa näiden kahden tehon kulutuksen sekä tuotannon jatkuvasti neljään eri rekisteriin; A+, A-, R+ sekä R- -rekisteriin.

- A+ Pätötehon kulutus
- A- Pätötehon tuotto
- R+ Kapasitiivinen loisteho
- R- Induktiivinen loistehon

Kaikki rekisterit ovat kumulatiivisia ja kasvavat sitä mukaan kun tuottoa tai kulutusta esiintyy. A+ -rekisteri kasvaa pätötehon (*active power*) kulutuksen mukaan kun

puolestaan A- -rekisteri kasvaa pätötehon tuotannon mukaan. R+ -rekisteri kasvaa pätötehon ohella tarvittun kapasitiivisen loistehon (*reactive power*) mukaan ja R- kasvaa tarvittun induktiivisen loistehon mukaan. Voidaan myös ajatella, että induktiivinen kuorma kuluttaa loistehoa ja kapasitiivinen kuorma tuottaa loistehoa. Etäluettava sähkömittari tallentaa jokaisen rekisterin hetkellisen arvon tietyin väliajoin kuten tunnin tai 15 minuutin välein. Asiakkaan laskutus tapahtuu näiden neljän eri rekisterin arvojen mukaan.

2.5. Mittarityypit

Kolme yleisintä sähkömittaria ovat *yksivaiheinen sähkömittari (1P)*, *kolmivaiheinen kolmijohdinmittari (3P3W)* sekä *kolmivaiheinen nelijohdinmittari (3P4W)*. Kolmivaihemittarit sisältävät kolme sisäistä johdinta L1, L2 sekä L3, joiden läpi kulkevan vaihtovirran jännitettä U , sähkövirtaa I sekä vaihekulmaa θ mitataan jatkuvasti. Niiden avulla mitataan pätötehon ja loistehon käyttöä, jotka tallennetaan mittarin kumulatiivisiin energiarekistereihin. [12] Kolmivaihevirta on siis yksi vaihtovirran eri muodoista, jossa kolmessa vierekkäisessä johtimessa kulkee sinimuotoinen, samantaajuinen ja yhtä suuri jännite. Jännitteiden sinimuotoisten signaalien välillä on 120 asteen eli 1/3 jakson vaihe ero, jolloin vaiheista yhteensä käytettävissä oleva teho on jatkuvasti yhtä suuri.

3P4W-mittarilla on lisäksi neljäs johdin, niin sanottu nollajohdin, joka toimii muille johtimille nollareferenssinä. Nollajohdin maadoitetaan, jonka avulla nollajohtimen sähkövirta sekä jännite saadaan hyvin lähelle nollaa. Johtimien jännitettä laskettaessa siis mitataan johtimen ja nollajohtimen jännitteen välistä eroa.

Yksinkertaisempi yksivaiheinen sähkömittari (1P) sisältää vain yhden sähkönsiirtoon käytettävän johtimen, jonka lisäksi se sisältää maadoitettavan nollajohtimen. Tällaisella mittarilla voidaan mitata vaihtovirran kulutusta, kun vaihtovirta siirretään vain yhtä johdinta pitkin.

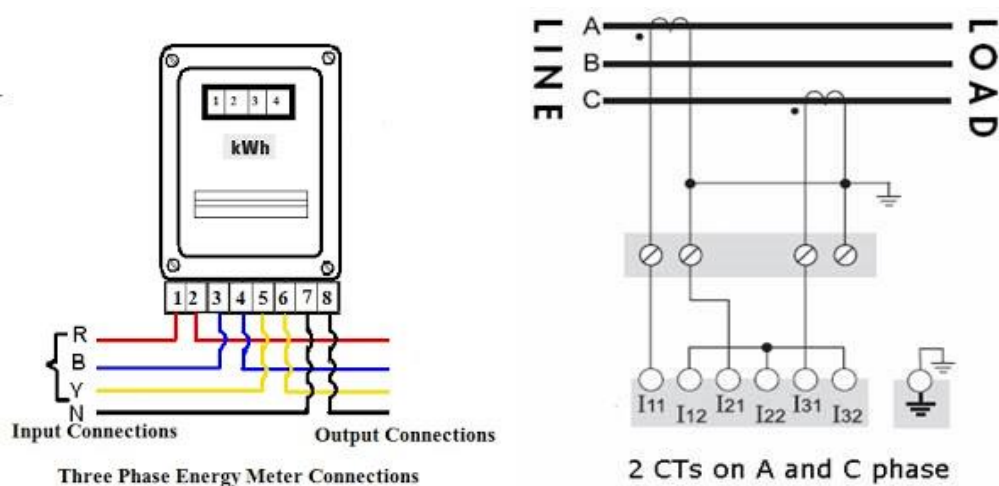
Suomessa käytetään yleisesti vaiheen ja nollan välisenä jännitteenä 230 volttia. 3P4W-mittari sisältää kolme erillistä wattimittaria, jotka laskevat jokaiselle vaiheelle kulutuksen erikseen nollajohdinta hyödyntäen. Muun muassa Norjassa, jossa

sähkömittarin maadoittaminen kivisen maaperän takia ei onnistu sekä osassa Keski-Euroopan maissa on käytössä kolmivaihejärjestelmiä, jossa 230 voltia saadaan kahden vaihejohtimen väliltä eikä vaiheen ja nollajohtimen väliltä. Näissä järjestelmissä pistorasioiden molemmissa pistokkeissa on hengenvaarallinen noin 133 voltin jännite-ero maahan verrattuna. Tällaisen kolmivaihejärjestelmän sähkönkulutusta mitataan 3P3W-mittarilla. Tällaisissa mittareissa riittää, että käytetään vain kahta wattimittaria, jolloin mittaus tehdään vain kahden vaiheen väliltä. Tämän vuoksi kyseisillä mittareilla ei voida laskea vaihekohtaista kulutusta, vaan ainoastaan koko sähkömittarin läpi kulkema sähköteho kahden mittauksen summana. [13] 3P3W-mittarissa on myös mahdollista käyttää kolmea wattimittaria, jolloin jokaisen johtimen kuormaa mitataan erikseen.

Kolmivaiheisissa sähkömittareissa on teoriassa mahdollista, että yhdessä johtimessa on kulutusta kun toisessa on tuotantoa. Eri mittarityypit käyttäytyvät kyseisessä tilanteessa eri tavoin. 3P4W-mittarilla on tällaisessa tilanteessa mahdollista mitata jokaisen johtimen kulutus ja tuotanto erikseen, jolloin molemmat A+ sekä A- -rekisterit voivat kasvaa samanaikaisesti, sillä jokaisen vaiheen kulutusta sekä tuotantoa mitataan erikseen omilla watti-mittareilla. 3P4W-mittari on myös mahdollista konfiguroida *netottavaksi*, jolloin mittari summaa pätötehon mittaustulokset yhteen, jolloin vain yksi A-rekisteri kasvaa kerralla riippuen onko kulutus suurempaa kuin tuotanto vai tuotanto suurempaa kuin kulutus mitattavan aikavälin aikana.

3P3W-mittareiden tapauksessa vain toinen A-rekisteri voi kasvaa kerralla, sillä mittarin sisältämän kahden wattimittarin mittaustuloksen summa ei kerro vaihekohtaisesta kulutuksesta tai tuotosta, vaan se kertoo koko mittarin läpi kulkeneesta tehosta. R-rekisterit ovat puolestaan aina summautuvia, jolloin vain toinen rekisterin arvo voi kasvaa kerralla. Tämä johtuu siitä, että samassa virtapiirissä syntyessään induktiivinen ja kapasitiivinen loisteho kumoavat toisensa, joten piiristä voidaan mitata vain toista loistehoa kerrallaan.

Yksivaiheisessa sähkömittarissa vain yksi A-rekisteri sekä yksi R-rekisteri voivat kasvaa samanaikaisesti, sillä mittari sisältää vain yhden wattimittarin, jonka arvoja mittari tallentaa.



Kuva 3: 3P4W-mittarin (vasemmalla)² ja 3P3W-mittarin (oikealla)³ kytkennät

2.6. Vika-asennus

Sähkömittareissa jokaiselle johtimelle on omat sisääntulo sekä ulostuloliitännät, jotka tulee kytkeä oikein mittaria asennettaessa. Käytännön mahdollisuudet väärinkytkenälle ovat johtimien sisääntulojen sekä ulostulojen kytkeminen ristiin tai väärään järjestykseen. [14] Kuvassa 3 esitetyn 3P4W-mittarin tapauksessa tämä tarkoittaa yhden, kahden tai kolmen johtimen sisääntulon sekä ulostulon kytkemistä ristiin, johtimien kytkemistä väärään järjestykseen tai nollajohtimen sisääntulon ja ulostulon kytkemistä ristiin. 3P3W-mittareiden tapauksessa väärinkytkeä voi tapahtua samalla tavoin kuin 3P4W-mittarin väärinkytkeä, lukuun ottamatta nollajohtimen sisääntulon ja ulostulon kytkemistä ristiin, sillä 3P3W-mittareissa ei ole kyseistä johdinta. Yksivaihemittarissa on mahdollista kytkeä, joko vaihejännitejohtimen tai nollajohtimen sisääntulo sekä ulostulo ristiin. [14][15]

Sähkömittarin väärinasennus muuttaa mittarin mittaustuloksia. 3P4W-mittareissa yhden L1, L2 tai L3 -johtimen sisääntulon ja ulostuloon kytkeminen ristiin aiheuttaa sähkövirran suunnan vaihtumisen yhdessä kolmesta vaiheesta. Tämä tarkoittaa netottavalla konfiguraatiolla asennettujen mittareiden tapauksessa molempien A-

² <https://electricalstandards.blogspot.com/2017/04/energy-meter-connections-single-phase.html>

³ <https://www.schneider-electric.com.sg/en/faqs/FA295682/>

rekisterien arvojen pienenemistä kolmannekseen siitä, mitä oikein asennettu mittari mittaisi, mikäli kaikkien vaiheiden kuormat ovat balanssissa. Myös R-rekisterien arvot putoavat noin kolmannekseen siitä, mitä oikein asennettu mittari tuottaisi. Tämä johtuu siitä, että mittari olettaa yhdessä johtimessa kulkevan sähkövirran olevan tuotantoa kulutuksen sijaan, sillä mittarin sisällä sähkövirta kulkee väärään suuntaan kyseisessä johtimessa.

Mikäli kaksi johdinta on asennettu ristiin 3P4W-mittarissa, kaksi kolmesta vaiheesta mittaakin kulutusta tuotantona, sähkövirran suunnan kääntyessä niissä. Tämä tarkoittaa, että yksi kolmesta vaiheesta mittaa kulutusta, kun kaksi muuta vaihetta mittaavat tuotantoa. Tämä tarkoittaa, että arvot, jotka kuuluisivat A+ -rekisteriin, lisääntyvätkin A- -rekisteriin siten, että lisääntyvä arvo on kolmannes alkuperäisestä A+ -rekisterin arvosta. Samoin käy myös R-rekistereille.

Mikäli kaikkien kolmen johtimen sisään- ja ulostulo ovat asennettu ristiin, kaikissa vaiheissa virran suunta kääntyy, jolloin mittari mittaa kulutuksen sijaan tuotantoa. Tämä aiheuttaa sen, että arvot, jotka kuuluisivat A+ -rekisteriin, kääntyvätkin A- -rekisteriin. Myös R+ -rekisteriin kuuluvat arvot kääntyvät R- -rekisteriin ja R- -rekisterin arvot R+ -rekisteriin sähkövirran suunnan vaihtuessa.

Nollajohtimen sisään- ja ulostulon kytkeminen ristiin ei vaikuta mittaustulokseen, sillä sen jännite on joka tapauksessa nolla. Myöskään vaihejännitteiden kytkeminen väärään järjestykseen ei vaikuta mittaustulokseen ja näin ollen ei ole varsinainen kytkentävirhe mittauksen osalta, vaikkakin väärä vaihejärjestys voi vaikuttaa esimerkiksi sähkömoottorien pyörimissuuntaan. Ei-netottavien 3P4W-mittareiden tapauksessa A-rekisterien arvot kertyvät samanaikaisesti A+ sekä A- -rekistereihin riippuen siitä montako johdinta on asennettu ristiin. [14][15]

3P3W-mittareiden tapauksessa vaihejännitteiden kytkeminen ristiin aiheuttaa hieman erilaista poikkeamaa mittaustulokseen verrattuna nelijohdinmittareihin. Mikäli L2-johdin toimii kaksi wattimittaria sisältävässä 3P3W-mittarissa referenssinä L1:lle sekä L3:lle, L2-johtimen sisääntulon ja ulostulon kytkeminen ristiin ei aiheuta muutosta itse mittaustuloksessa. Tämä johtuu siitä, että L2-johtimen sähkövirtaa ei käytetä laskettaessa mittarin läpi kulkemaa sähkötehoa vaan laskentaan käytetään

vain L1 ja L3 -johtimien sähkövirtaa sekä L1 ja L3 -johtimen jännite-eroa L2-johtimeen nähden. [11]

Tilanteessa, jossa 3P3W-mittarissa joko L1 tai L3 on kytketty ristiin, A-rekistereihin kuuluvat arvot valuvat R-rekisterien ja R-rekisterien arvot valuvat A-rekistereihin. Se tallentuuko A-rekisterien arvot R+ vai R- -rekisteriin, riippuu siitä, kumpi johdin on kytketty ristiin. Testauksen perusteella balansoidulla kuormalla, jossa jännitteet sekä vaihekulmat olivat samat, L1 tai L3 käännettäessä A+ -rekisterin suhteellinen osuus R-rekistereiden suhteelliseen osuuteen verrattuna pieneni, kun vaihekulmat vaihtelivat 45 ja -45 asteen välillä. Tarkemmat tulokset eri vaihekulmilla voi nähdä liitteestä. [Liite 1]

Tilanteessa, jossa L1 sekä L3 sisääntulo ja ulostulo on kytketty ristiin, mittarin kahdessa mitattavassa vaiheessa virran suunta kääntyy, jolloin A+ -rekisteriin kuuluvat arvot kääntyvätkin A- -rekisteriin sekä R+ -rekisteriin kuuluvat arvot kääntyvät R- -rekisteriin ja toisin päin. [Liite 1]

3P3W-mittareissa vaiheiden kytkeminen väärään järjestykseen ei vaikuta itse mittaustuloksiin millään tavoin, joten se ei ole 3P3W-mittareidenkaan tapauksessa varsinaisesti kytkentävirhe mittaustuloksiin osalta. Kaikissa edellä mainituissa tapauksissa oletetaan, että vaiheiden kuormat ovat balanssissa keskenään. Käytännössä vaiheiden kuormat eivät ole täysin identtisiä ja joskus vaiheiden välillä saattaa olla suuriakin eroja jännitteen ja sähkövirran määrässä sekä vaihekulmissa, mikä aiheuttaa vaihtelua siinä, millaista mittaustuloksiin virheellisesti asennettu mittari tuottaa. [16] Mikäli oletetaan, että johtimien jännite, sähkövirta ja vaihekulma voivat olla mitä vain, väärinkytkeä voi teoriassa muuttaa mittarin mittaustulokset näennäisesti satunnaisiksi. [14][15][17]

1P-mittarissa ainoan johtimen sisääntulon sekä ulostulon kytkeminen ristiin saa sähkövirran kulkemaan 180-astetta väärään suuntaan, jolloin kaikki A+ -rekisteriin kuuluva teho kääntyykin A- -rekisteriin. Nollajohtimen sisääntulon ja ulostulon kytkeminen ristiin ei muuta mittaustuloksia.

Taulukko 1: Esimerkki virhekytkennän vaikutuksesta mittaustulokseen

Virhekytkentä	A+	A-	R+	R-
Oikein kytketty	30	0	24	0
Kaikkien johtimien I/O asennettu ristiin	0	30	0	24
3P4W, 2 I/O kytketty ristiin	10	20	8	16
3P4W, 1 I/O kytketty ristiin	20	10	16	8
Netottava 3P4W, 2 I/O kytketty ristiin	0	10	0	8
Netottava 3P4W, 1 I/O kytketty	10	0	8	0
3P3W, L1 I/O käännetty	14,5	0	0	18
3P3W, L3 I/O käännetty	0	14,5	18	0
3P3W, 1 I/O käännetty, epätasainen kuorma	10	0	0	25

Tästä päätellen voidaan olettaa, että tavallista suuremmat A- arvot sekä A+ ja R - rekisterien epätavallinen suhde voivat indikoida väärin asennetusta sähkömittarista. Virhekytkennän vaikutusta mittaustulokseen on myös havainnollistettu taulukossa 1, josta tämä voidaan havaita. Tulee kuitenkin ottaa huomioon, että osassa kohteista saattaa olla esimerkiksi aurinkopaneeleista johtuvaa tuotantoa, jolloin A- kasvu voi tapahtua. Netottavassa konfiguraatiossa olevassa mittarissa tuotannon pitää kuitenkin olla suurempaa kuin kulutus, jotta niiden summa olisi negatiivinen ja näin ollen A-rekisteri kasvaisi.

Mittari voi myös vikaantua jostain syystä. Viallinen sähkömittari voi lopettaa mittaamisen, jolloin sen lähettämät kumulatiiviset rekisteriarvot eivät enää kasva. Mikäli sähkömittarin lähettämän rekisterin arvo pienenee, se on varma merkki mittarin vikaantumisesta, sillä näin ei pitäisi pystyä käymään. Lisäksi, jos mittarin lähettämä rekisterin arvo kasvaa erittäin suuren summan, on sekin merkki mittarin vikaantumisesta. Viallisten mittareiden havaitseminen on huomattavasti helpompaa

verrattuna väärin asennettujen etäluettavien sähkömittareiden havaitsemiseen, sillä se voidaan tehdä luentajärjestelmään saapuvien rekisteriarvojen validoinnilla, jossa varmistetaan, että uudet saapuvat arvot ovat järkeviä suhteessa edellisiin arvoihin. Mikäli, mittari tuottaa vain nolla-arvoja kaikkiin rekistereihin, ei mittarin viallisuudesta voida olla varmoja, sillä samankaltaista mittausdataa voi tuottaa kohde, jossa ei ole lainkaan kulutusta tai tuotantoa. Tämän lisäksi myös sähkömittarin mittaamat epätavallisen suuret jännite ja sähkövirta lukemat ovat merkki siitä, että mittarissa on mahdollisesti jotain vikaa tai sitä ei ole asennettu oikein. Sallitut lukemat ovat kuteinkin mittarityypistä riippuvaisia.

3. Poikkeavuuksien havaitseminen

Tässä luvussa käsitellään poikkeavuuksien havaitsemistekniikoita tiedonlouhinnan välineenä. Poikkeavuuksien havaitsemistekniikat ovat tärkeä työkalu havaitessa poikkeavia datapisteitä datamassan joukosta. Poikkeavat datapisteet indikoivat usein ongelmatilannetta tai muuta epätavallista tapahtumaa, joka voi vaatia käyttäjän tarkempaa huomiota. Poikkeavuuksien havaitsemistekniikoita on mahdollista käyttää, niin lääketieteessä, pankkialalla kuin sähkömittauksessa etsimään poikkeavia tapahtumia.

3.1. Poikkeavuuksien havaitsemisen perusteet

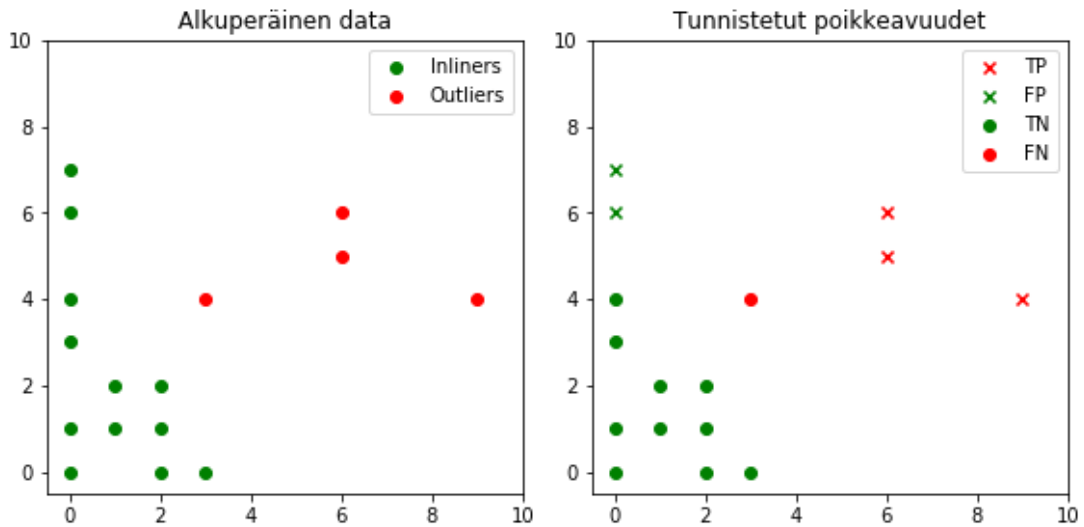
Tiedonlouhinta on joukko työkaluja, joiden tarkoitus on löytää merkittävää informaatiota datamassasta. Yksi tiedonlouhinnassa käytettävistä menetelmistä on poikkeavuuksien havaitsemistekniikat, joita kutsutaan myös anomaly detection tekniikoiksi. Niiden tarkoitus on identifioida tyypillisestä datasta poikkeavia datapisteitä, tapahtumia tai kaavoja. [18] Tyypillisesti tavanomaisesta datasta huomattavasti poikkeavat datapisteet kertovat jonkinlaisesta ongelmasta. Poikkeavuuksien havaitsemistekniikoita voidaan käyttää useissa eri aiheilyhteiksissä, etsimään tavanomaisesta poikkeavia datapisteitä kuten havaitsemaan pankkikortin väärinkäytön, sähkömittarin lukeman vääristämisen, yrityksen tietojärjestelmään tunkeutumisen tai syövän havaitsemisen röntgenkuvasta. Se millainen datapiste luokitellaan poikkeavaksi, ja millaisia datapisteitä halutaan tunnistaa datamassasta, riippuu aiheilyhteystestä. Poikkeavat havainnot jaetaan usein kolmeen eri luokkaan, jotka ovat *globaali poikkeama*, *kontekstuaalinen poikkeama* sekä *kollektiivinen poikkeama*. [19] Globaalilla poikkeamalla tarkoitetaan yhtä datainstanssia, jonka arvo on muihin datapisteisiin verrattuna poikkeuksellinen. Esimerkiksi tavallista suurempi osto luottokortilla voidaan luokitella globaaliksi poikkeamaksi. Kontekstuaalinen poikkeama on poikkeava datapiste tietyssä kontekstissa. Esimerkiksi loma-aikoina suuremmat ja tiheimmät luottokortilla tehdyt ostot eivät välttämättä ole poikkeamia, mutta arkena samanlainen kulutus voidaan luokitella poikkeavuudeksi. Kollektiiviset poikkeavuudet ovat joukko poikkeavia datapisteitä, jotka voidaan yhdessä luokitella poikkeavuudeksi. Esimerkiksi useat tiheästi tehdyt

tilisiirrot ulkomaille voidaan luokitella poikkeavuudeksi, vaikka tilin käyttäjällä olisi tapana tehdä satunnaisesti tilisiirtoja ulkomaille. Tilisiirrot yhdessä luovat poikkeavuuden, vaikka yksittäiset tapahtumat eivät vaikuttaisi tavallisesta poikkeavilta.

Poikkeavuuksien havaitsemismenetelmät voidaan jakaa kolmeen eri kategoriaan. Kategoriat ovat *ohjatut* (supervised anomaly detection), *puoliohjatut* (semisupervised anomaly detection) sekä *ohjaamattomat* poikkeavuuksien havaitsemismenetelmät (unsupervised anomaly detection). [20] Ohjatut tekniikat olettavat, että saatavilla on koulutusdataa, jossa jokainen datapiste on merkitty joko normaaliksi tai poikkeavuudeksi. Tyypillinen lähestymistapa ohjatuissa poikkeavuuksien havaitsemismenetelmissä on rakentaa ennustava malli tavalliselle vs. poikkeavalle datalle. Uusia ennennäkemättömiä dataa voidaan verrata ennustavaan malliin, joka luokittelee mihin luokkaan kyseinen datapiste kuuluu. Puoliohjatut tekniikat olettavat, että saatavilla on koulutusdataa, jossa vain normaaleiksi luokiteltavat datapisteet ovat merkitty. Tyypillinen lähestymistapa on rakentaa malli kuvaamaan vain normaaliksi luokiteltavaa dataa ja käyttää kyseistä mallia tunnistamaan poikkeavuudet testidatan joukosta. Koska puoliohjatuissa poikkeavuuksien havaitsemistekniikoissa ei tarvita poikkeavuuksiksi merkittyjä datapisteitä, se on huomattavasti laajemmin sovellettava tekniikka ohjattuihin tekniikkoihin verrattuna. Viimeinen kategoria on ohjaamattomat poikkeavuuksien havaitsemismenetelmät. Kyseiset menetelmät eivät vaadi ollenkaan koulutusdataa, jonka vuoksi ne ovat kaikista laajimmin sovellettavia tekniikoita. Ohjaamattomat menetelmät perustuvat oletukseen, että normaaliksi luokiteltavia datapisteitä esiintyy datassa toistuvammin verrattuna poikkeavuuksiksi luokiteltaviin, joita oletetaan esiintyvän harvemmin. Mikäli oletus ei päde testattavassa kontekstissa, kyseiset tekniikat eivät välttämättä tunnista poikkeavuuksia tehokkaasti. [20] Tässä työssä keskitytään näihin ohjaamattomiin menetelmiin.

Kun poikkeavuuksien havaitsemistekniikka luokittelee datapisteitä, voidaan luokittelun oikeellisuus jaotella neljään eri kategoriaan, jotka ovat *true positive* (TP), *false positive* (FP), *true negative* (TN) sekä *false negative* (FN). Mikäli tekniikka luokittelee datapisteen oikein poikkeavuudeksi, luokittelu määritellään TP:ksi ja jos datapiste luokitellaan oikein tavalliseksi, luokittelu määritellään TN:ksi. Jos

tekniikka luokittelee tavalliseksi laskettavan datapisteen poikkeavuudeksi, kyseessä on FP ja mikäli tekniikka luokittelee poikkeavuudeksi laskettavan datapisteen tavalliseksi, kyseessä on FN. Luokittelun oikeellisuutta on kuvattu kuvassa 4.



Kuva 4: Esimerkki luokittelun oikeellisuudesta

Poikkeavuuksien havaitsemistekniikan tunnistustarkkuutta kuvataan tavallisesti kahdella eri suureella, jotka ovat tarkkuus eli *precision* sekä herkkyys eli *recall*. Tarkkuus voidaan laskea kaavalla

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

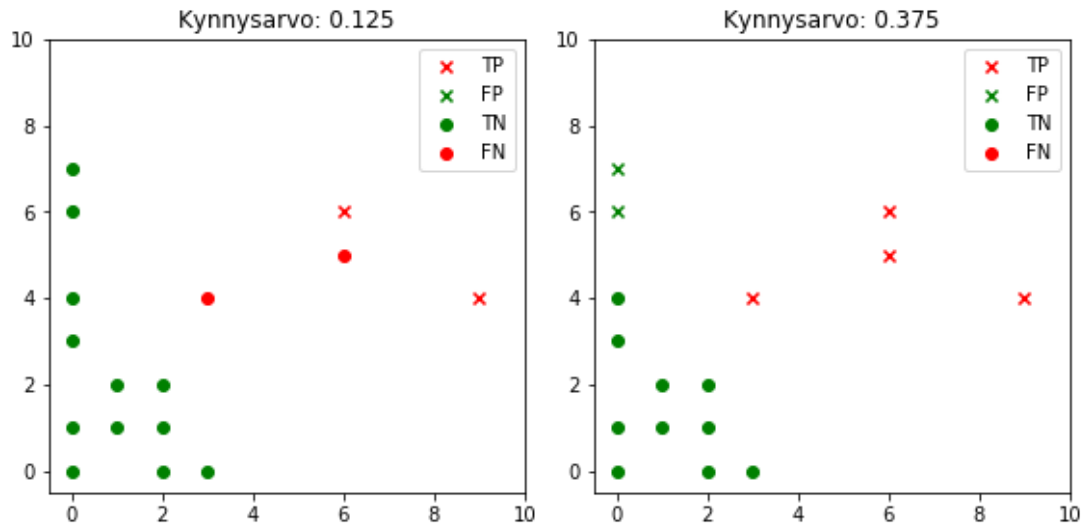
Tarkkuus kuvaa kuinka monta prosenttia tekniikan poikkeavuuksiksi luokitelluista datapisteistä on oikeasti poikkeavuuksia. Herkkyys voidaan laskea kaavalla

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

ja se kuvaa puolestaan kuinka monta prosenttia kaikista poikkeavuuksiksi laskettavista datapisteistä tekniikka tunnisti oikein poikkeavuuksiksi.

Poikkeavuuksien havaitsemistekniikat koostuvat kahdesta vaiheesta. Ensin datapisteet pisteytetään ja järjestetään sen mukaan kuinka vahvasti tekniikka luokittelee datapisteen poikkeavuudeksi. Toisessa vaiheessa tekniikka luokittelee tietyn osan suurimman pisteytyksen saaneista datapisteistä poikkeavuuksiksi kynnysarvon eli tekniikalle asetetun *contamination raten* mukaan. Tekniikoiden

poikkeavuuksiksi luokiteltavien datapisteiden prosentuaalista osuutta voidaan vaihdella muuttamalla tekniikan kynnyksisarvoa. Kynnyksisarvoa nostamalla tekniikka luokittelee suuremman osan datapisteistä poikkeavuuksiksi ja kynnyksisarvoa laskemalla pienemmän osan.



Kuva 5: Esimerkki kynnyksisarvon vaikutuksesta luokitteluun

Kuvassa 5 ensimmäisessä kaaviossa tarkkuudeksi voidaan laskea 1,00 ja herkkyydeksi 0,50 ja toisessa kaaviossa tarkkuus on 0,67, mutta herkkyyys on 1,00. Kynnyksisarvon kohottaminen siis lisää väärin hälytysten määrää, mutta yhä useampi poikkeava datapiste tunnistetaan poikkeavuudeksi. Kynnyksisarvon laskeminen puolestaan lisää tarkkuutta vähentämällä vääriä hälytyksiä, mutta siinä tapauksessa yhä useampi poikkeavuudeksi luokiteltavista datapisteistä jää tunnistamatta.

3.2. Poikkeavuuksien havaitseminen älykkäässä sähkömittauksessa

Etäluettavat sähkömittarit, kiihtyvä mittausinfrastruktuurien kehitys ja niiden tuottamat suuret datamassat mahdollistavat uusia tapoja tutkia ja analysoida mittausdataa ja eristää siitä uutta hyödyllistä informaatiota. Poikkeavuuksien havaitsemismenetelmiä on aiemmin sovellettu AMI:n yhteydessä muutamiin eri tarkoituksiin. Poikkeavuuksien havaitsemismenetelmiä on hyödynnetty muun muassa tunnistamaan eri syistä johtuvaa sähköhävikkiä, kuten sähkövarkauksia, virheellisiä mittarilukemia ja vikaantuneita mittareita. [21] Etenkin sähkövarkauden

tunnistamista poikkeavuuksien havaitsemismenetelmillä on tutkittu paljon ja niistä on saatu melko hyviä tuloksia. [21][22][23] Esimerkiksi eräässä tutkimuksessa pystyttiin tunnistamaan peukaloituja sähkömittareita mittareiden tuntisarjojen perusteella jopa 94 % tarkkuudella virheellisten tunnistusten osuuden ollessa 11 %. [22] Poikkeavuuksien havaitsemismenetelmiä on sovellettu AMI-kontekstissa useissa eri tutkimuksissa ja tulokset ovat olleet lupaavia.

Kun poikkeavuuksia yritetään havaita energiarekisterien tuntisarjojen perusteella, tunnistus tehdään usein ohjaamattomilla tai puoli ohjatuilla tekniikoilla. Tämä johtuu siitä että, koulutusdatan saaminen, jossa mittausdatat on valmiiksi luokiteltu normaaleiksi ja poikkeaviksi, on erittäin vaikeaa. Jotta tiedetään varmasti, onko mittarin tuottama mittausdata normaalia vai poikkeavaksi luokiteltavaa, pitäisi mittari tarkastaa fyysisesti, jotta tiedetään onko sitä peukaloitu tai toimiiko mittari kuten pitääkin. Puoli ohjattu malli on mahdollista rakentaa käyttämällä dataa, jonka oletetaan olevan normaalia, mutta tällöin ei kuitenkaan ole varmuutta, että datan seassa ei olisi joitakin poikkeavuuksiksi luokiteltavia mittausdatoja. Ohjaamattomat tekniikat ovat kaikista helpoin tapa soveltaa poikkeavuuksien havaitsemismenetelmiä, kun tunnistus tehdään sähkömittareiden tuntisarjojen perusteella. Tällöin koulutusdatalle ei ole tarvetta ja tunnistus voidaan suorittaa käyttämällä luokittelematonta dataa.

3.3. Tekniikoita

Ohjaamattomat poikkeavuuksien havaitsemismenetelmät luokittelevat datapisteet poikkeamiksi tai tavallisiksi puhtaasti datapisteiden ominaisuuksien perusteella, kun koulutusdataa ei ole saatavilla. Ohjaamattomien poikkeavuuksien havaitsemismenetelmien toiminta voi perustua useisiin erilaisiin tekijöihin kuten etäisyyteen, tiheyteen tai tilastolliseen todennäköisyyteen ja ne ovat laajasti sovellettavissa eri sovellusalueille. Ohessa on listaus muutamista ohjaamattomista ja puoli ohjatuista tekniikoista ja niiden toiminnasta.

3.3.1. Principal Component Analysis

Tavallisesti *principal component analysis* (PCA) -tekniikkaa eli pääkomponenttianalyysia käytetään vähentämään moniulotteisen datan ulottuvuuksia. Tekniikassa etsitään moniulotteisesta datasta ne komponentit, jotka kuvaavat dataa kaikista parhaiten, jolloin mahdollisimman vähän merkittävää informaatiota katoaa. Tämä tarkoittaa, että datasta etsitään ne avaruuden pinnat, joille projisoituna data tuottaa suurimman varianssin. Pääkomponenttianalyysia voidaan käyttää myös poikkeavuuksien havaitsemistekniikkana. Mikäli otoskoko on suuri, voidaan olettaa, että $\sum_{i=1}^q \frac{y_i^2}{\lambda_i}$, $q \leq p$, noudattaa khiin neliö (χ^2) jakaumaa, jossa p kuvaa ulottuvuuksien määrää, q vapausastetta, y_i datapisteen pääkomponentteja ja λ ominaisarvoja (eigenvalues). Datapiste luokitellaan poikkeavuudeksi, mikäli

$$\sum_{i=1}^q \frac{y_i^2}{\lambda_i} > \chi_q^2(\alpha) \quad (3)$$

jossa $\chi_q^2(\alpha)$ on ylempi α prosenttipiste khiin neliö -jakaumassa, vapausasteen ollessa q . [24]

3.3.2. One-Class Support Vector Machine

One-class support vector machine (OCSVM) on Schölkopf et al. [25] esittelemä tekniikka datapisteiden luokitteluun. Tekniikka luokittelee datapisteet normaaleiksi tai poikkeavuuksiksi luomalla hypertason datapisteiden välille, joka erottaa normaaliksi luokiteltavat datapisteet poikkeavuuksiksi luokiteltavista. Poikkeavuuksiksi luokiteltavat datapisteet pisteytetään ja järjestää vertaamalla kuinka kaukana ne sijaitsevat luotuun hypertasoon verrattuna. Pisteytys voidaan tehdä pisteytysfunktiolla

$$q_{SVM}(x) := \rho - (w \cdot \Phi(x)) \quad (4)$$

jossa $\Phi(x)$ kuvaa mapping funktiota ja jossa painokerroin w ja offset ρ optimoidaan ratkaisemalla yhtälö

$$\min_{\xi \in \mathbb{R}^n, \rho \in \mathbb{R}} \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i - \rho \quad (5)$$

joka noudattaa $(w \cdot \Phi(x_i)) \geq \rho - \xi_i, \xi_i \geq 0, v \in (0,1)$. [25][26]

3.3.3. k-Nearest Neighbors

k-nearest neighbors (KNN) on etäisyyteen perustuva lähestymistapa. Se perustuu oletukseen, että normaaliksi luokiteltava data esiintyy tiheämmissä ryppäissä poikkeavuuksiksi luokiteltaviin datoihin verrattuna. KNN-tekniikkaa voidaan käyttää poikkeuksien havaitsemistekniikkana käyttämällä klusterin k :nneksi lähimmän datapisteen etäisyyttä arviona todennäköisyydelle, että kyseinen piste on poikkeavuus. Formaalisti datapisteiden pisteytys voidaan tehdä määrittelemällä pisteytysfunktio

$$q_{kthNN}(x) := d^k(x; X) \quad (6)$$

jossa $d^k(x; X)$ on etäisyys x :n ja sen k -lähimmän naapurin välillä. [26][27]

3.3.4. Histogram-Based Outlier Score

Histogram-based outlier score (HBOS) on tiheyteen perustuva lähestymistapa. HBOS perustuu oletukseen, että tiheämmin esiintyvät datapisteet esittävät normaaliksi luokiteltavaa dataa kun taas harvemmin esiintyvät datapisteet esittävät poikkeavaksi luokiteltavaa dataa. HBOS luo jokaisesta eri piirteestä oman histogrammin, joko laskemalla eri kategoristen muuttujien määrän tai numeeristen muuttujien tapauksessa määrittelemällä ensin arvoalueet, joihin piirteet tullaan luokittelemaan ja sen jälkeen laskemalla kuinka monta instanssia kuhunkin luokkaan kuuluu. Näin jokaiselle piirteelle saadaan muodostettua omat histogrammit, joiden korkeudet esittävät esiintymien määrää kyseisessä luokassa tai arvovälissä. HBOS-tekniikka mahdollistaa myös arvoalueiden dynaamisen määrittelyn. Kun jokaiselle piirteelle on saatu luotua histogrammi, histogrammit vielä skaalataan siten, että suurin sallittu histogrammin arvo on 1, jotta jokainen piirre on samanarvoinen keskenään. Tämän jälkeen voidaan laskea HBOS jokaiselle datapisteelle x kaavalla

$$q_{HBOS}(x) = \sum_{i=0}^d \log\left(\frac{1}{hist_i(x)}\right) \quad (7)$$

jossa d kuvaa eri piirteiden lukumäärää. Mitä harvinaisempi datapiste on, sitä suuremman HBOS-arvon se saa. [28]

3.3.5. One-Time Sampling

One-time sampling (OTS) on etäisyyteen perustuva tekniikka, jossa käytetään hyväksi satunnaisotosta. Tekniikassa muodostetaan satunnaisesti ja itsenäisesti muodostettu osajoukko $S(X) \subset X$ vain kerran ja tämän jälkeen määritellään pisteytysfunktion

$$q_{Sp}(x) := \min_{x' \in S(X)} d(x, x') \quad (8)$$

avulla pisteet jokaiselle datapisteelle $x \in X$. Pisteytys tehdään siis mittaamalla datapisteen etäisyys satunnaisesti valitun osajoukon lähimpään datapisteeseen. Tekniikan suurin hyöty on sen nopeus. Sen aikavaativuus on vain $O(nms)$, jossa m kuvaa eri piirteiden lukumäärää ja s kuvaa satunnaisotoksen kokoa. [26]

3.3.6. Elliptic Envelope

Elliptic envelope (EE) on minimum covariance determinant (MCD) -algoritmiin perustuva tekniikka. MCD on tilastolliseen todennäköisyyteen perustuva lähestymistapa. Se perustuu oletukseen, että tavallisiksi luokiteltujen datapisteiden jakauma noudattaa jotain tiettyä todennäköisyysjakaumaa kuten normaalijakaumaa. MCD olettaa, että tavallisiksi luokitellut datapisteet esiintyvät suuremman todennäköisyyden alueella, lähellä jakauman odotusarvoa. Poikkeamiksi oletetaan datapisteet, jotka esiintyvät jakauman reunoilla, jossa todennäköisyys kyseisen datapisteen esiintymiselle on hyvin pieni. [29]

3.3.7. Local Outlier Factor

Local outlier factor (LOF) on tiheyteen perustuva lähestymistapa, joka analysoi datapisteen paikallista poikkeamaa pisteen naapureihin verrattuna. Tekniikka laskee jokaiselle pisteelle paikallisen tiheyden etsimällä ensin k -lähintä naapuria, joiden

etäisyyksien avulla tehdään arvio tiheydestä. Vertailemalla datapisteen paikallista tiheyttä sen naapureihin voidaan identifioida datapisteet, joiden paikallinen tiheys on huomattavasti alhaisempi sen naapureihin verrattuna. Formaalisti datapisteen saavutettavuustiheys voidaan kuvata kaavalla

$$p(x) := |N^k(x)|^{-1} \left(\sum_{x' \in N^k(x)} \max\{d^k(x', X), d(x, x')\} \right)^{-1} \quad (9)$$

jossa $N^k(x)$ kuvaa x :n k -lähintä naapuria. Datapiste voidaan pisteyttää laskemalla sille LOF, joka määritellään x :n paikallisen saavutettavuustiheyden ja x :n k -lähimpien naapureiden paikallisten saavutettavuustiheyksien keskiarvon suhteena:

$$q_{LOF}(x) := \left(|N^k(x)|^{-1} \sum_{x' \in N^k(x)} p(x') \right) p(x)^{-1} \quad (10)$$

Tekniikan aikavaativuus on $O(n^2m)$, jossa m kuvaa datan ulottuvuuksien lukumäärää. [26][30]

3.3.8. Cluster-Based Local Outlier Factor

Cluster-based local outlier factor (CBLOF) on myös tiheyteen perustuva lähestymistapa, joka osaa ottaa huomioon myös klusterin koon, johon se kuuluu. Tämän avulla on mahdollista tunnistaa tiheä klusteri poikkeavuudeksi, jos se poikkeaa suuresti tavallisesta datasta. Tekniikka jaottelee ensin datan klustereihin ja erottelee suuret klusterit pienistä. Sen jälkeen tekniikka laskee jokaiselle datapisteelle CBLOF arvon, joka kuvaa datapisteen klusterin keskustan etäisyyttä lähimmän suuren klusterin keskustaan. Näin ollen tekniikka pystyy tunnistamaan tiheän klusterin poikkeavuudeksi, mikäli klusterin etäisyys tavalliseen dataan on suuri. [31]

3.3.9. Isolation Forest

Isolation forest (IF) -tekniikan ideana on eristää datapisteet, ja pisteyttää datapiste sen mukaan kauanko sen eristäminen kestää. Jos datapisteen eristäminen kestää kauan ja useita kierroksia se tarkoittaa, että datapiste ei ole luultavasti tavallisesta

poikkeava, sillä sen eristäminen muista datapisteistä on hankalaa. Datapisteen eristäminen tapahtuu seuraavalla menetelmällä.⁴ [32]

- 1) Valitaan datapiste, joka halutaan eristää.
- 2) Jokaiselle piirteelle asetetaan raja-arvot, jotka ovat kyseisen piirteen pienin ja suurin arvo, joiden välillä eristäminen tehdään.
- 3) Valitaan piirre sattumanvaraisesti.
- 4) Valitaan satunnaisluku kyseisen piirteen raja-arvojen väliltä:
 - a) Jos datapisteen piirteen arvo on suurempi kuin satunnaisluku, vaihdetaan satunnaisluku piirteen raja-arvon uudeksi alarajaksi.
 - b) Jos datapisteen piirteen arvo on pienempi kuin satunnaisluku, vaihdetaan satunnaisluku piirteen raja-arvon uudeksi ylärajaksi.
- 5) Toistetaan kohtia 3-4, kunnes tarkasteltu datapiste on ainut piste rajojen sisällä.
- 6) Lasketaan kuinka monesti kohdat 3-4 täytyi toistaa. Tätä lukua kutsutaan eristysluvuksi.

3.4. Intuiitiiviset sovelluslähtöiset tekniikat

Mikäli on tiedossa, minkä tyylinen data voi viitata poikkeuksesta tai jostakin ongelmasta, on mahdollista luoda myös intuitiivisia, sovelluslähtöisiä poikkeavuuksien havaitsemistekniikoita. Luvussa 3.3 esitellyt tekniikat ovat sovellettavissa useisiin eri sovelluksiin ja niitä ei ole kehitetty mihinkään erityistarkoitukseen.

Sähkömittauksessa on tiedossa, millaista mittausdataa väärin kytketty sähkömittari voi tuottaa. Tällöin on mahdollista käyttää myös intuitiivisia tekniikoita etsimään kyseisten mittareiden tuottamaa poikkeuksellista mittausdataa. On tiedossa, että väärinkytkenät näkyvät etenkin tavallista suurempana A- arvona tai A+, A-, R+ ja R- -rekistereihin kertyneen tehon epätavallisena suhteena, etenkin 3P3W-mittarin

⁴ <https://quantdare.com/isolation-forest-algorithm/>

väärinkytkenän tapauksessa. Tässä työssä määrittelemme kolme intuitiivista tekniikkaa, joilla pyritään löytämään poikkeuksellista mittausdataa.

- MinSum
- MinA+
- MaxA-

Datan esitysmuodon tapauksessa, jossa mittausdata esitetään pätötehon tuntikulutuksena, etsitään vektoreita, jossa pätötehon yhteiskulutus on ollut kaikista pienintä (*MinSum*). Tämä johtuu siitä, että väärinasennetuissa etäluettavissa sähkömittareissa A- -rekisteri on taipuvaisempi kasvamaan ja pätötehon kulutus on usein pienempää oikein asennettuun mittariin verrattuna. Esitysmuodon tapauksessa, jossa mittausdata esitetään rekisterien suhteellisina osuuksina, etsitään mittausdataa, jossa A-, R+ sekä R- -rekisterien summa on mahdollisimman suuri ja A+ -rekisteri mahdollisimman pieni (*MinA+*). Etenkin 3P3W-mittarin väärinasennus muuttaa A ja R -rekisterien suhdetta ja saattaa osassa tapauksista kasvattaa muita rekistereitä A+ -rekisterin sijaan. Kolmantena intuitiivisena tekniikkana käytetään prosentuaalisesti suurimpien A- -rekisterien etsintää (*MaxA-*).

Esimerkiksi mikäli seuraavista energiarekisterien prosentuaalisista osuuksista, jossa [A+, A-, R+, R-], tekniikat etsisivät yhden poikkeavan instanssin, MinA+ tekniikka luokittelisi c:n poikkeavuudeksi sen A+ ollessa kaikista pienin ja MaxA- luokittelisi a:n poikkeavuudeksi sen A- ollessa kaikista suurin.

- a) [0.4, 0.3, 0.2, 0.1]
- b) [0.3, 0.2, 0.4, 0.1]
- c) [0.1, 0.2, 0.3, 0.4]

3.5. Datan esiprosessointi

Datan esiprosessointi on tärkeä osa poikkeavuuksien tunnistamisessa. Alkuperäinen data saattaa sisältää esimerkiksi erittäin suuria poikkeamia, jonka vuoksi pienemmät poikkeamat jäävät havaitsematta tai alkuperäisessä datassa voi olla kohinaa, joka heikentää joidenkin poikkeavuuksien havaitsemismenetelmien toimivuutta. Siksi on tyypillistä, että käytettävälle datalle tehdään muunnoksia ennen kuin siitä yritetään

havaita poikkeavuuksia. Datan esiprosessointi voi olla esimerkiksi arvojen skaalaamista uudelle välille tai arvojen skaalaaminen normaalijakaumalle.

Yleisiä datan esiprosessointitekniikoita ovat muun muassa *min-max skaalaus*, *normalisointi*, *robustinen skaalaus* sekä *pääkomponenttianalyysi*. Min-max skaalaus muuttaa datan skaalaamalla sen [0, 1] välille. Kaava, jolla muunnos tehdään, on

$$x' = \frac{x_i - X_{min}}{X_{max} - X_{min}} \quad (11)$$

Normalisointi muuttaa datan jakauman vastaamaan normaalijakaumaa $X \sim N(0,1)$. Normalisoinnin avulla datapisteet saadaan keskitettyä 0:n ympärille ja keskihajonnaksi saadaan vakioitua 1. Kaava, jolla muunnos tehdään, on

$$x' = \frac{x_i - \mu}{\sigma} \quad (12)$$

jossa μ on vektoreiden keskiarvo ja σ vektoreiden keskihajonta.

Robustinen skaalaus sietää paremmin suuria poikkeamia kuin esimerkiksi normalisointi, jossa yksi datamassasta selvästi poikkeava arvo voi haitata normalisoitujen arvojen tasaista jakautumista. Kuten min-max skaalaus, robustinen skaalaus muuttaa datan skaalaamalla vektorit tietylle välille, mutta robustinen skaalaus käyttää min ja max arvojen sijaan alakvartiilia Q_1 ja yläkvartiilia Q_3 . Alakvartiili on datapiste, jota pienempiä arvoja on aineistossa 25 % ja yläkvartiili on arvo, jota suurempia arvoja on aineistossa 25 %. Kaava, jolla muutos tehdään, on

$$x' = \frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \quad (13)$$

Pääkomponenttianalyysin (PCA) avulla on mahdollista vähentää datan moniulotteisuutta. PCA eristää moniulotteisesta datasta keskeisimmät komponentit ja esittää ne käyttäen vähemmän ulottuvuuksia kuin alkuperäisessä datassa siten, että mahdollisimman vähän merkittävää dataa häviää. [33]

4. Virhekytkentöjen tunnistaminen energiasarjoista

Tässä luvussa tutkitaan kuinka hyvin eri poikkeavuuksien havaitsemistekniikat toimivat virheellisesti asennettujen sähkömittareiden havaitsemiseen mittareiden lähettämien energiarekisterit sisältävien tuntisarjojen perusteella. Testauksessa vertaillaan, miten eri poikkeavuuksien havaitsemistekniikat, eri datan esiprosessointitekniikat sekä virheellisen mittausdatan suhteellinen määrä vaikuttavat virhekytkentöjen havaitsemiseen.

4.1. Tutkimuskysymys

Tutkielman tutkimuskysymys on kuinka hyvin eri poikkeavuuksien havaitsemistekniikat sopivat netottavien sähkömittareiden virhekytkentöjen tunnistamiseen yhden päivän A+, A-, R+ sekä R- -energiarekisterien tuntisarjojen perusteella. Tutkittavat virhekytkennät ovat:

1. Kaikkien johtimien sisään ja ulostulo on kytketty ristiin.
2. Netottava 3P4W-mittari, jossa kahden johtimen sisään ja ulostulo on kytketty ristiin.
3. Netottava 3P4W-mittari, jossa yhden johtimen sisään ja ulostulo on kytketty ristiin.
4. 3P3W-mittari, jossa yhden johtimen sisään ja ulostulo on kytketty ristiin.
5. 3P3W-mittari, jossa yhden johtimen sisään ja ulostulo on kytketty ristiin ja jossa johtimien kuormat ovat erisuuruiset.

4.2. Tutkimussuunnitelma

Tavoitteena on löytää tutkimuskysymykseen vastaus luomalla oikean mittausdatan sekaan virhekytkentää simuloivia mittausdatoja ja yrittämällä tunnistaa niitä oikean mittausdatan joukosta. Tutkittavia virhekytkentöjä on viisi erilaista.

Testaus koostuu kolmesta eri vaiheesta:

1. Mittausdatan muuttaminen sopivaan esitysmuotoon

2. Virheellisen mittausdatan simuloiminen eri virhekytkentöjen mukaisesti
3. Virhekytkentää simuloivien mittausdatojen etsiminen
 - Eri poikkeavuuksien havaitsemistekniikat
 - Eri datan esiprosessointi
 - Eri virheellisen mittausdatan määrä

Ensimmäisessä vaiheessa valitaan mittausdatalle sopiva esitysmuoto. Prosessoimattomat energiarekisteriarvot eivät itsessään kerro paljoakaan, vaan mittareiden tuottamista rekisteriarvoista täytyy erotella tarkoituksenmukaiset piirteet, kuten eri rekisterien väliset suhteelliset osuudet tai tallennettujen rekisteriarvojen muutos tunneittain.

Toisessa vaiheessa muutamme osan alkuperäisestä mittausdatasta vastaamaan viittä erityyppistä virheellisesti asennetun mittarin tuottamaa mittausdataa, jolla simuloidaan virheellisesti kytkettyä sähkömittaria. Virheellinen mittausdata sekoitetaan alkuperäisen mittausdatan joukkoon.

Kolmannessa vaiheessa alkuperäiseen mittausdataan sekoitettua virheellistä mittausdataa yritetään etsiä eri poikkeavuuksien havaitsemistekniikoiden avulla. Samalla tutkitaan kuinka eri datan esiprosessointimenetelmät sekä vaihteleva virheellisen mittausdatan määrä vaikuttavat poikkeavuuksien havaitsemiseen. Virheellisen mittausdatan määränä käytetään 20 ja 100 kappaletta eli 0,158 % ja 0,790 %. Testauksen jälkeen tulokset analysoidaan ja tehdään suosituksia, mitkä menetelmä toimivat parhaiten virheellisen mittausdatan ja täten virheellisesti asennettujen mittarien tunnistamiseksi.

Poikkeavuuksien havaitsemistekniikoiden vertailu suoritetaan testaamalla jokaista viittä eri virhekytkentätapausta erikseen. Jokaisen virhekytkentätapauksen testaus suoritetaan käyttäen kahta eri mittausdatan esitysmuotoa: pätötehon tuntikulutusta sekä energiarekisterien prosentuaalisia osuuksia. Vertailussa on mukana PCA [24], OCSVM [25], k-NN [27], HBOS [28], one-time sampling [26], elliptic envelope [29], LOF [30], CBLOF [31] ja isolation forest [32] tekniikat. Pätötehon tuntikulutuksen tapauksessa vertailussa on mukana myös intuitiivinen MinSum-tekniikka sekä energiarekisterien suhteellisia osuuksia käytettäessä mukana on MinA+ sekä MaxA- -tekniikat. Muita vartenotettavia mutta tässä työssä tapahtuvan

testauksen ulkopuolelle jääviä menetelmiä ovat muun muassa mean-shift outlier detection (MOD) [34], medoid-shift outlier detection (DOD) [34], outlier detection using indegree number (ODIN) [35] sekä reverse unreachability (NC) [36]. Datan esiprosessointitekniikoista vertailussa mukana on pääkomponenttianalyysi, min-max skaalaus, normalisointi sekä robustinen skaalaus.

4.3. Datalähteet

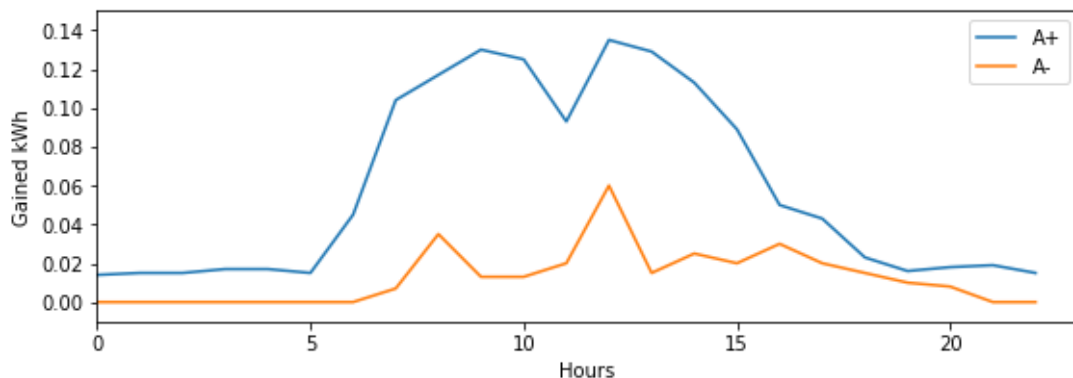
Etäluettavat sähkömittarit mittaavat myös monia muita arvoja pätötehon sekä loistehon lisäksi, mutta niitä ei välttämättä lähetetä automaattisesti luentajärjestelmään, vaan laite lähettää ne vain pyydettyäessä. Tämän vuoksi väärin asennettujen laitteiden havaitsemiseen käytetään vain A+, A-, R+ sekä R- -rekisterien arvoja. Muun datan, kuten vaihekohtaisten vaihekulmien tai sähkövirran määrän jatkuva kysely laitteilta voi rasittaa mittausdatan lähetykseen käytettävää tiedonsiirtoverkkoa. Tämän vuoksi tutkielmassa käytetään vain suoraan luentajärjestelmästä saatavia energiarekisterien tuntisarjoja, jotka ovat aina saatavilla kaikille mittareille. Käytössä oleva data, jonka pohjalta testit tehdään, on aitoa mittausdataa, joka sisältää yhteensä 12 663 kappaletta 1P, 3P3W sekä 3P4W-mittarien A+, A-, R+ sekä R- -rekisterien tuntisarjat yhdeltä päivältä.

Mittaustulokset ovat mitattu 23.10.2018. Energiarekisterien lisäksi data sisältää mittarin uniikin laitekoodin, jonka avulla identifiointi voidaan tehdä sekä mittausajankohdan. Analysoitavaan dataan otetaan mukaan vain laitteet, joilta on saatavissa jokaisen energiarekisterin lukema jokaiselta tunnilta. Alkuperäisessä datassa joiltakin mittareilta puuttuu joitakin yksittäisiä mittauksia. Nämä on siivottu pois datamassasta. Tämän lisäksi datamassasta on poistettu sellaiset laitteet, joiden kumulatiivisen A+ -rekisterin arvo on hyvin pieni, alle 10 kWh, jolloin sähkömittarin mittaamassa kohteessa ei ole ollut lähes lainkaan kulutusta. Tämä toimenpide tehdään siksi, että hyvin pienissä rekisteriarvoissa saattaa suhteessa olla hyvin paljon kohinaa, jolloin ne helposti näyttävät väärin asennetulta, vaikka mittari toimisi oikein.

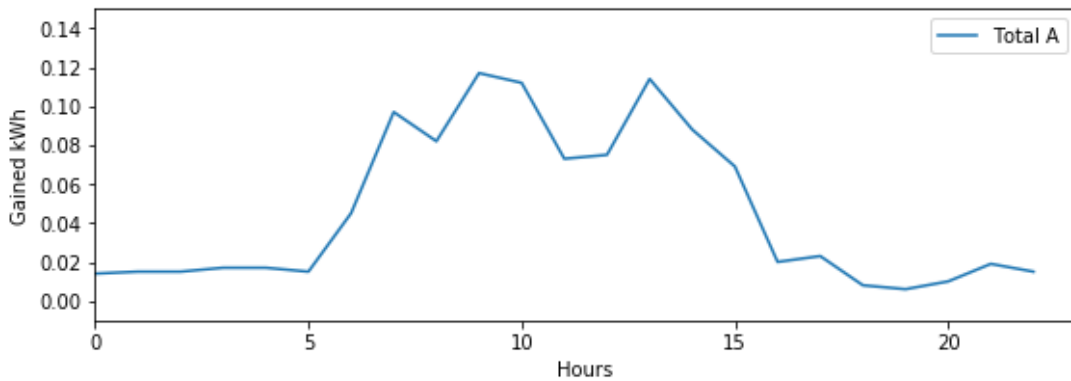
4.4. Datan esitysmuoto

Testauksen pohjana toimii noin 12 663 sähkömittarin mittausdata yhdeltä päivältä. Mittausdata koostuu tunneittain tallennetuista hetkellisistä energiarekisterien lukemista. Tämä tarkoittaa, että päivän aikana kumulatiivisista A+, A-, R+ sekä R- -rekistereistä tallennetaan hetkellinen arvo 24 kertaa. Voidaan siis ajatella, että jokaisen mittarin jokainen rekisteri luo 24-ulotteisen vektorin, joka koostuu 24 eri rekisteriarvosta. Jokainen mittari tuottaa siis yhteensä 96 uutta rekisterilukemaa päivän aikana.

Poikkeavuuksien havaitsemistekniikoiden vertailussa tullaan käyttämään kahta eri mittausdatan esitystapaa: pätötehon tuntikulutusta sekä energiarekisterien prosentuaalisia osuuksia. Tiedetään, että virheellisesti asennetun netottavan mittarin A- -rekisterin suuret arvot ovat usein viittaus virheellisestä kytkennästä. Tämän vuoksi ensimmäisenä esitystapana datasta käytetään käyrää, joka koostuu A- rekisterien tuntimuutosten erotuksesta eli yksi piste kuvaa kuinka monta kWh pätötehoa on yhteensä kulutettu tunnin aikana. Näin mittausdata saadaan myös vastaamaan netottavaksi asetettua sähkömittaria. Näin ollen jokaiselle laitteelle luodaan oma 23 pisteestä koostuva tuntisarjakäyrävektori, joka kuvaa pätötehon yhdistettyä kulutusta ja tuotantoa eri tunteina kuten kuvassa 7 on havainnollistettu.

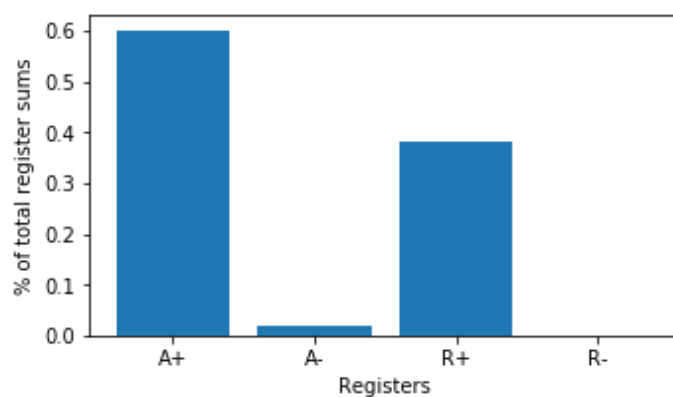


Kuva 6: Esimerkki pätötehon kulutuksesta ja tuotannosta tunneittain



Kuva 7: Esimerkki pätötehon yhdistetystä käytöstä

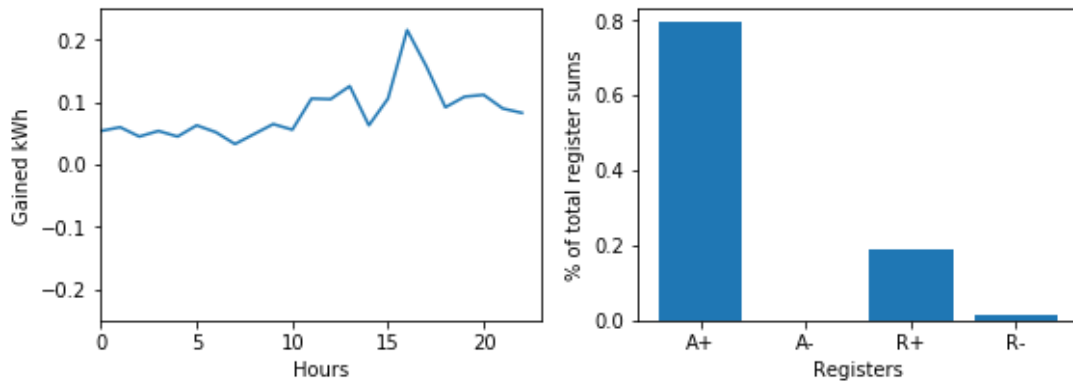
Toinen virhekytkentää indikoiva merkki etenkin 3P3W-mittareilla on A-rekisterien sekä R-rekisterien epätavallinen suhde. A sekä R -rekisterien suhde on usein tavallisesta poikkeava kun yhden tai kahden johtimen sisääntulo sekä ulostulo on sähkömittarissa kytketty ristiin. Tämän vuoksi toiseksi datan esitysmuodoksi valikoitui rekisterien prosentuaaliset osuudet. Kumuloituvat rekisterit on jaettu kaikkien rekisterien summalla. Tästä syntyy neliulotteinen vektori, joka kuvaa rekisterien välisiä suhteellisia osuuksia ja joiden yhteenlaskettu summa on yksi. Esimerkiksi mikäli mittarin kumulatiivisten rekisterien lukemat olisivat [300, 10, 190, 0] niin uudeksi vektoriksi saataisiin [0.60, 0.02, 0.38, 0.00], jossa luvut kuvaavat rekisterien prosentiosuutta kaikesta mittarin läpi kulkemasta sähkötehosta, kuten kuvassa 8 on havainnollistettu.



Kuva 8: Esimerkki rekisterien suhteellisista osuuksista

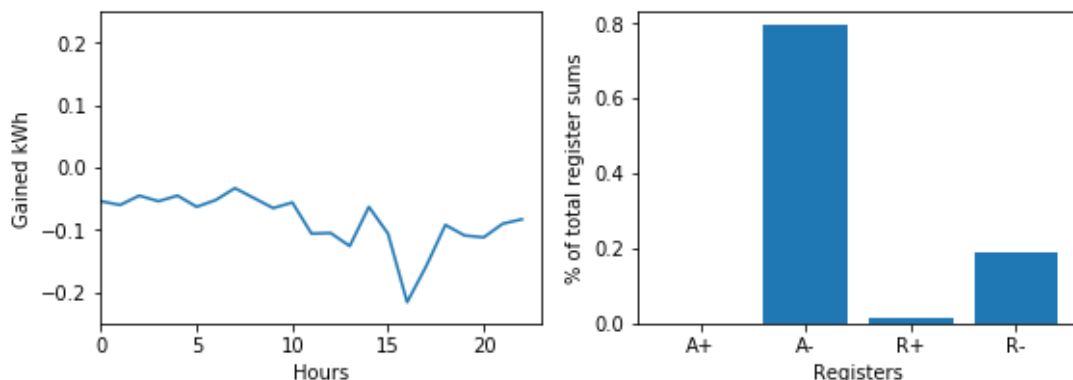
4.5. Virheellisen datan simulointi

Jotta pystymme vertaamaan poikkeavuuksien havaitsemistekniikoiden kykyä tunnistaa virhekytkentöjä, täytyy mittausdatan sekaan luoda virheellistä dataa, joka simuloi väärin kytketyn sähkömittarin tuottamaa mittausdataa. Virhekytkentää simuloiva mittausdata luodaan muuttamalla osa alkuperäisestä mittausdatasta vastaamaan virhekytkennän tuottamaa mittausdataa. Eri virhekytkennät voivat muuttaa mittausdataa viidellä eri tavalla. Kuvassa 9 on esimerkit tavallisesta mittausdatasta.



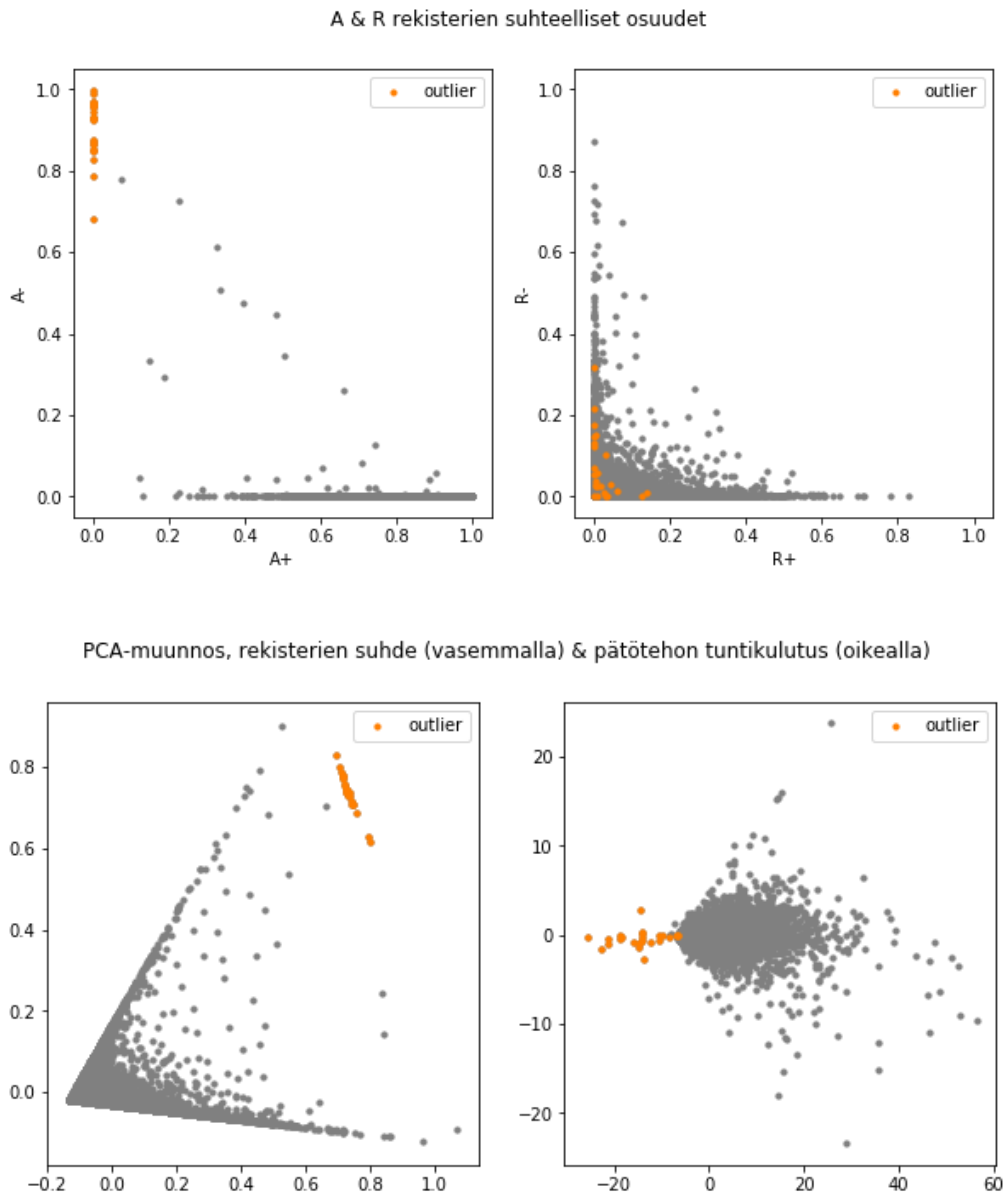
Kuva 9: Esimerkki alkuperäisestä mittausdatasta

Mikäli mittareissa jokaisen johtimen sisääntulo ja ulostulo on kytketty ristiin, sähkövirran suunta kääntyy eli mitattu kulutus muuttuu tuotannoksi. Tilannetta voidaan esittää kääntämällä A+ ja A- sekä R+ ja R- rekisterit keskenään alkuperäiseen mittausdataan verrattuna. Virhekytkennän vaikutusta mittausdataan on kuvattu kuvassa 10.



Kuva 10: Mittausdata kaikkien johtimien I/O ollessa käännettynä

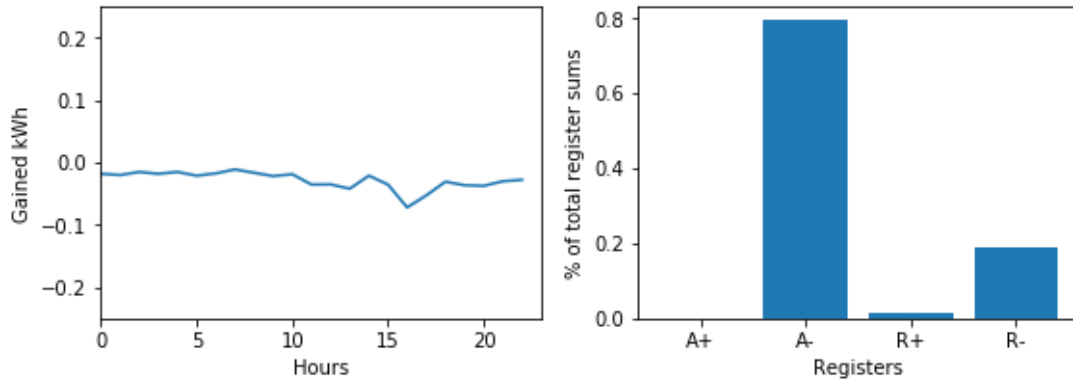
Kuvassa 11 on kuvattu kuinka kyseisen virhekytkennän tuottama mittaustieto sijoittuu tavalliseen mittaustietoon nähden, kun tavallisen mittaustietoon sekaan on simuloitu 25 virheellistä mittaustietoa.



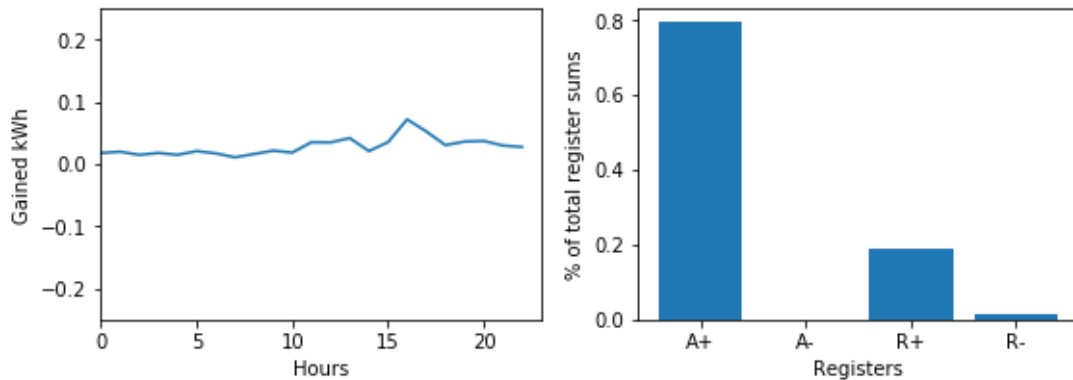
Kuva 11: Esimerkki datan jaottumisesta kaikkien johtimien I/O ollessa käännettynä

Mikäli puolestaan netottavassa 3P4W-mittarissa kahden johtimen sisääntulo ja ulostulo on kytketty ristiin, mittari tuottaa noin $-1/3$ alkuperäisistä A sekä R arvoista. Jos yhden johtimen sisääntulo sekä ulostulo ovat kytketty ristiin, mittari tuottaa noin $1/3$ alkuperäisistä A sekä R arvoista olettaen, että kaikkien johtimien kuormat ovat

yhtä suuria eli johtimissa kulkee samansuuruinen jännite ja virta sekä vaihekulmat ovat kaikkien vaiheiden johtimissa samat. Kyseisten virhekytkentöjen vaikutusta mittausdataan on kuvattu kuvissa 12 sekä 13.



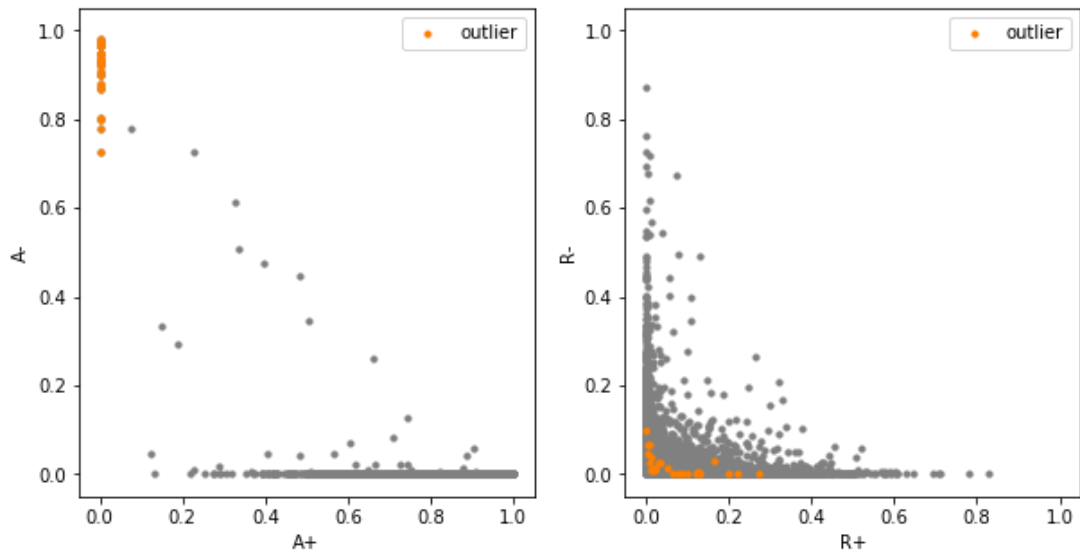
Kuva 12: Mittausdata 3P4W-mittarin kahden johtimen väärinkytkenässä



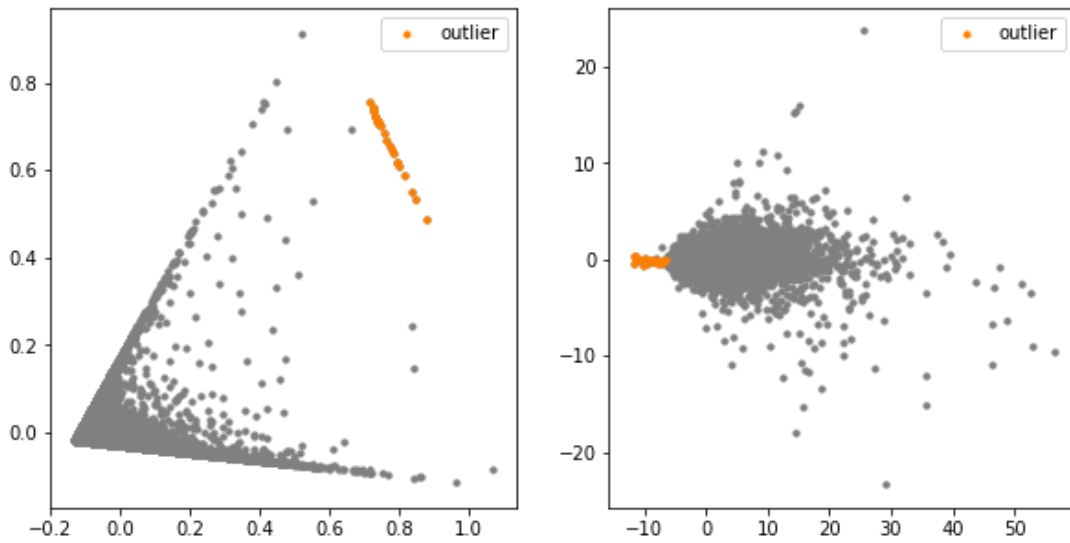
Kuva 13: Mittausdata 3P4W-mittarin yhden johtimen väärinkytkenässä

Kuvassa 14 sekä 15 on kuvattu kuinka kyseisten virhekytkentöjen tuottama mittausdata sijoittuu tavalliseen mittausdataan nähden, kun tavallisen mittausdatan sekaan on simuloitu 25 virheellistä mittausdataa.

A & R rekisterien suhteelliset osuudet

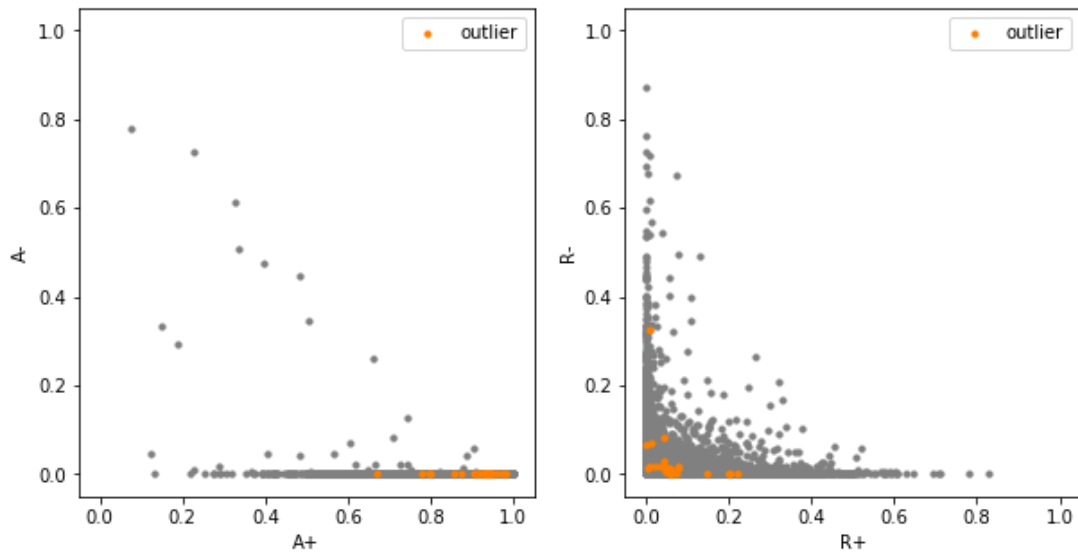


PCA-muunnos, rekisterien suhde (vasemmalla) & päätöksen tuntikulutus (oikealla)

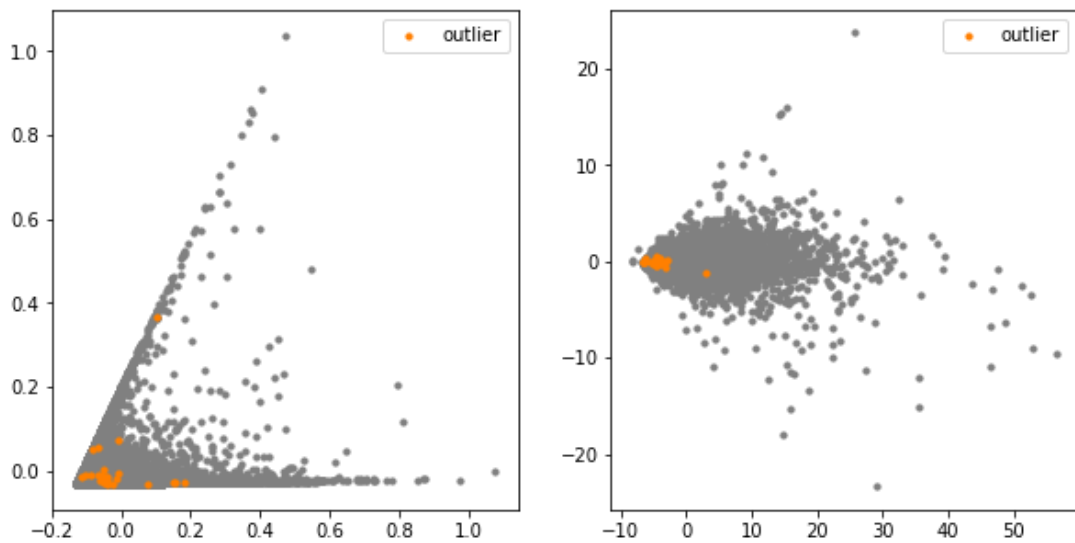


Kuva 14: Esimerkki datan jaottumisesta 3P4W-mittarin kahden johtimen väärinkytkenässä

A & R rekisterien suhteelliset osuudet



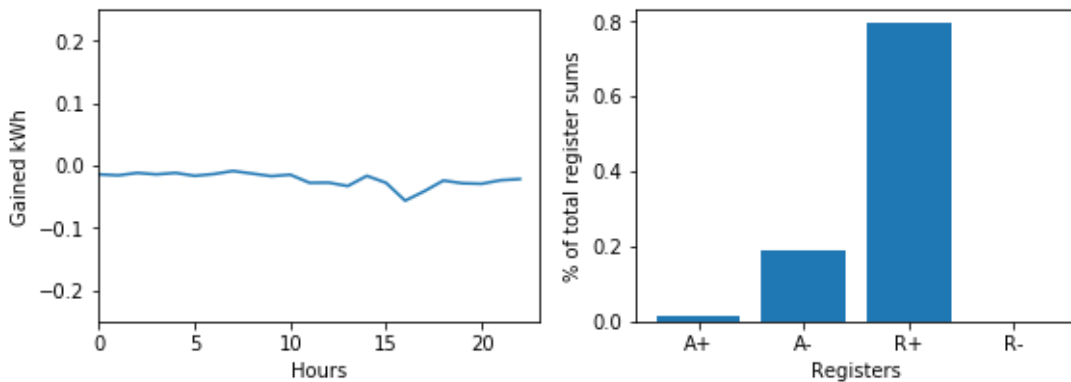
PCA-muunnos, rekisterien suhde (vasemmalla) & pätötehon tuntikulutus (oikealla)



Kuva 15: Esimerkki datan jaottumisesta 3P4W-mittarin yhden johtimen väärinkytkenässä

Neljäs tapaus on 3P3W-mittarin yhden johtimen sisääntulon ja ulostulon asentaminen ristiin. Tämä aiheuttaa sen, että mittarissa, jossa on kulutusta eli A+ -rekisterin arvo kasvaa jatkuvasti ja vaihekulmat ovat -90 sekä 90 välillä, energiarekisterien suhteelliset osuudet muuttuvat oikein kytkettyyn mittariin verrattuna. Se, miten A sekä R-rekisterien suhteelliset osuudet muuttuvat, riippuu siitä, kumpi mittaava vaihejohdin on kytketty väärin ja mitkä ovat vaiheiden

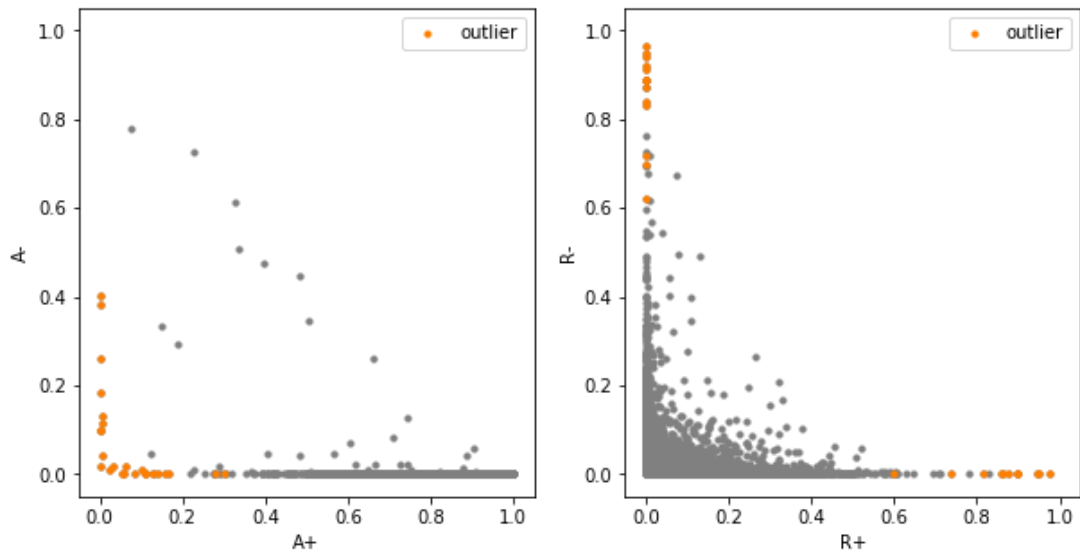
vaihekulmat. L1 ollessa kytketty väärin, vaihekulmasta riippuen, mittarissa kasvaa joko A+ tai A- -rekisteri sekä R- -rekisteri. L3 ollessa kytketty ristiin A+ tai A- kasvaa vaihekulmasta riippuen sekä R+ -rekisteri kasvaa. Rekisterien väliset suhteet riippuvat myös vaihekulmista. [Liite 1] Esimerkiksi tilanne, jossa oikein kytketty mittari mittaisi lukemat [39262, 0, 6955, 0] vaihekulman ollessa 10, L1-johtimen virheellisen kytkennän jälkeen sama mittari mittaisi lukemat [4018, 0, 0, 22641]. Mikäli L3-johdin olisi kytketty väärin, mittari mittaisi lukemat [0, 4018, 22641, 0]. Virhekytkentätaulukosta voidaan havaita, että rekisterien suhteelliset osuudet eivät muutu virhekytkennöissä, vaan kokonaistehon määrä muuttuu sekä rekisterien paikat vaihtuvat alkuperäiseen tilanteeseen nähden. Kun data esitetään rekisterien suhteellisina osuuksina, virheellistä mittausdataa voidaan yksinkertaisesti simuloida vaihtamalla rekisterien paikkoja liitteenä olevan Excel-taulukon mukaisesti, sillä rekisterien suhteelliset osuudet eivät muutu virhekytkennässä. Kun data esitetään pätehtöön tuntikulutusikäyränä, väärin kytkettyä sähkömittaria voidaan simuloida pienentämällä tai suurentamalla tuntikulutusta liitteenä olevan Excel-taulukon mukaisesti. Virhekytkennän vaikutusta mittausdataan on kuvattu kuvassa 16.



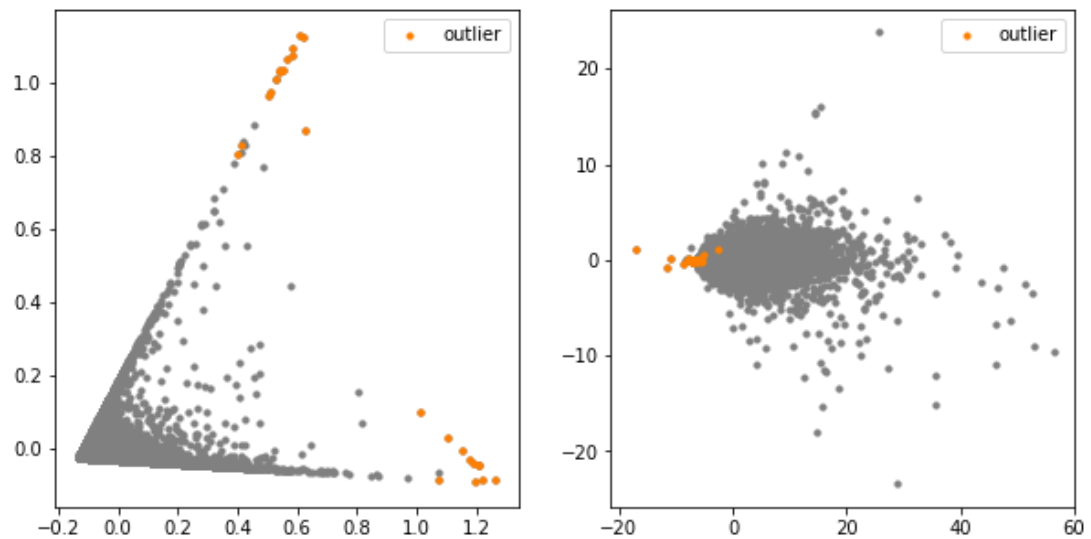
Kuva 16: Mittausdata 3P3W-mittarin L3-johtimen väärinkytkennässä

Kuvassa 17 on kuvattu kuinka kyseisen virhekytkennän tuottama mittausdata sijoittuu tavalliseen mittausdataan nähden, kun tavallisen mittausdatan sekaan on simuloitu 25 virheellistä mittausdataa.

A & R rekisterien suhteelliset osuudet



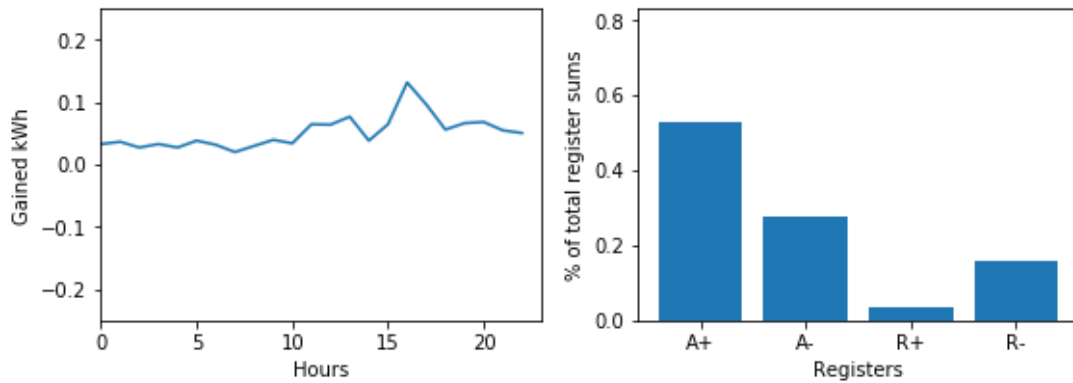
PCA-muunnos, rekisterien suhde (vasemmalla) & pätötehon tuntikulutus (oikealla)



Kuva 17: Esimerkki datan jaottumisesta 3P3W-mittarin yhden johtimen väärinkytkenässä

Mikäli vaiheiden kuormat eivät ole balanssissa ja vaiheiden vaihekulmat sekä sähkövirta voivat vaihdella johtimien kesken, ei A ja R -rekisterien muutosta virhekytkennässä voi ennakoida, jos vaihekohtaiset vaihekulmat eivät ole tiedossa. Tässä tapauksessa virhekytkentä voi tuottaa eri mittaustuloksia riippuen siitä kuinka paljon sähkövirtaa kussakin johtimessa kulkee ja mitkä ovat eri vaiheiden vaihekulmat. Näin ollen tapaus, jossa johtimien kuormat eivät ole balanssissa, voi

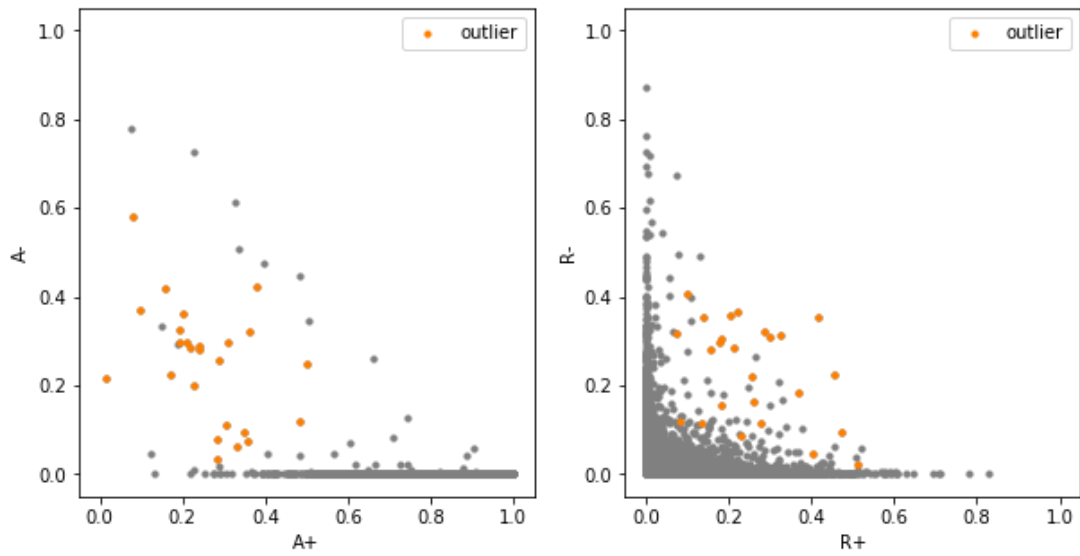
tuottaa näennäisesti satunnaista mittausdataa ja jonka takia liitteenä olevaa Excel-taulukkoa [Liite 1] ei voida soveltaa tähän tapaukseen. Tätä tapausta simuloidaan luomalla satunnaisia rekisterien suhteita siten, että A+, A-, R+ ja R- -rekisterien summa on 1, kun data esitetään rekisterien suhteellisina osuuksina. Kun data esitetään pätötehon tuntikulutusena, virheellistä mittausdataa simuloidaan kertomalla pätötehon tuntikulutus satunnaisella luvulla -2 ja 2 välillä.



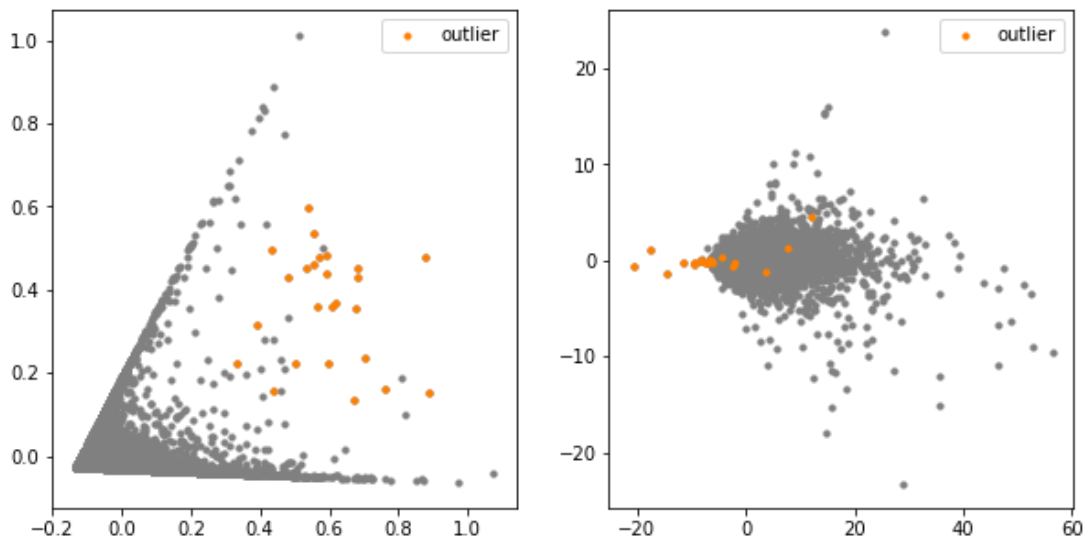
Kuva 18: Mittausdata 3P3W-mittarin yhden johtimen virhekytkennässä epäbalanssilla kuormalla

Kuvassa 19 on kuvattu kuinka kyseisen virhekytkennän tuottama mittausdata sijoittuu tavalliseen mittausdataan nähden, kun tavallisen mittausdatan sekaan on simuloitu 25 virheellistä mittausdataa.

A & R rekisterien suhteelliset osuudet



PCA-muunnos, rekisterien suhde (vasemmalla) & päätötehon tuntikulutus (oikealla)



Kuva 19: Esimerkki datan jaottumisesta 3P3W-mittarin yhden johtimen virhekytkennässä epäbalanssilla kuormalla

5. Tulokset

Tässä osiossa esitetään tulokset, jotka saatiin vertailemalla eri poikkeavuuksien havaitsemistekniikoiden tehokkuutta sähkömittarien väärinkytentöjen havaitsemisessa. Testissä vaihdeltiin virheellisen mittausdatan määrää 20 ja 100 välillä, jolloin outlierien prosentuaalinen osuus vaihtelee välillä 0,158 % ja 0,790 %. Toistoja testissä tehtiin 10 kertaa, joiden keskiarvot ovat esitettyinä tulostaulukoissa.

Tulostaulukkojen sarakkeet kuvaavat eri poikkeavuuksien havaitsemistekniikoita ja sarakkeet kuvaavat eri datan esiprosessointitekniikoita. Tulokset kertovat kunkin poikkeavuuksien havaitsemistekniikan ja datan esiprosessointitekniikan kombinaation saavuttaman tarkkuuden sekä herkkyyden. Havaitsemistekniikoiden etsimä virheellisen mittausdatan määrä eli contamination rate on asetettu vastaamaan mittausdatan seassa olevan virheellisen mittausdatan määrää, jolloin *Precision* = *Recall*. Väärien hälytysten määrä on siis yhtä suuri kuin tunnistamatta jätettyjen poikkeavuuksien määrä.

Testaukset tehtiin Jupyter Notebook -alustalla, johon oli asennettuna Python 3.6 ohjelmointikieli. Kaikki tekniikat ovat valmiiksi implementoituina ja vapaasti käytettävissä sklearn⁵, PyOD⁶ sekä Kenchi⁷ Python -kirjastoissa. Kaikissa käytetyissä tekniikoissa on mahdollista määrittää contamination rate eli kuinka suuren osan datasta tekniikka luokittelee poikkeavuuksiksi. Tämä helpottaa tekniikoiden vertailua, kun kaikki tekniikat etsivät saman määrän poikkeavia mittausdatoja. Etsittyä poikkeavuuksien määrää lukuun ottamatta kaikissa tekniikoissa käytetään pakettien oletusasetuksia, jotka on listattu seuraavaksi.

- Principal Component Analysis (PCA)

```
pyod.models.pca.PCA(n_components=None,  
n_selected_components=None, contamination=c, copy=True,
```

⁵ <https://scikit-learn.org/stable/>

⁶ <https://pyod.readthedocs.io/>

⁷ <https://github.com/HazureChi/kenchi>

```
whiten=False, svd_solver='auto', tol=0.0,
iterated_power='auto', random_state=None, weighted=True,
standardization=True)
```

- **One-Class Support Vector Machine (OCSVM)**

```
pyod.models.ocsvm.OCSVM(kernel='rbf', degree=3, gamma='auto',
coef0=0.0, tol=0.001, nu=0.5, shrinking=True, cache_size=200,
verbose=False, max_iter=-1, contamination=c)
```

- **k-Nearest Neighbors (KNN)**

```
pyod.models.knn.KNN(contamination=c, n_neighbors=5,
method='largest', radius=1.0, algorithm='auto', leaf_size=30,
metric='minkowski', p=2, metric_params=None, n_jobs=1,
**kwargs)
```

- **Histogram-Based Outlier Score (HBOS)**

```
pyod.models.hbos.HBOS(n_bins=10, alpha=0.1, tol=0.5,
contamination=c)
```

- **One-Time Sampling (OTS)**

```
kenchi.outlier_detection.OneTimeSampling(contamination=c,
metric='euclidean', novelty=false, n_subsamples=20,
random_state=20, metric_params=None)
```

- **Elliptic Envelope (EE)**

```
sklearn.covariance.EllipticEnvelope(store_precision=True,
assume_centered=False, support_fraction=None, contamination=c,
random_state=None)
```

- **Local Outlier Factor (LOF)**

```
pyod.models.lof.LOF(n_neighbors=20, algorithm='auto',
leaf_size=30, metric='minkowski', p=2, metric_params=None,
contamination=c, n_jobs=1)
```

- **Cluster-Based Local Outlier Factor (CBLOF)**

```
pyod.models.cblof.CBLOF(n_clusters=8, contamination=c,
clustering_estimator=None, alpha=0.9, beta=5,
use_weights=False, check_estimator=False, random_state=None,
n_jobs=1)
```

- **Isolation Forest (IF)**

```
pyod.models.iforest.IForest(n_estimators=100,
max_samples='auto', contamination=c, max_features=1.0,
bootstrap=False, n_jobs=1, behaviour='old', random_state=None,
verbose=0)
```

Käytetyt datan esiprosessointitekniikat ovat vapaasti saatavilla sklearn Python-kirjastossa.

- **2-ulotteinen pääkomponenttianalyysi (PCA)**⁸
`sklearn.decomposition.PCA(n_components=2, copy=True, whiten=False, svd_solver='auto', tol=0.0, iterated_power='auto', random_state=None)`
- **Normalisointi (SS)**⁹
`sklearn.preprocessing.StandardScaler(copy=True, with_mean=True, with_std=True)`
- **Min-max skaalaus (MMS)**¹⁰
`sklearn.preprocessing.MinMaxScaler(feature_range=(0, 1), copy=True)`
- **Robustinen skaalaus (RS)**¹¹
`sklearn.preprocessing.RobustScaler(with_centering=True, with_scaling=True, quantile_range=(25.0, 75.0), copy=True)`

⁸ <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

⁹ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

¹⁰ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

¹¹ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>

5.1. Kaikkien johtimien I/O kytketty ristiin

Taulukko 2: Tunnistusprosentti käyttäen pätehtehon tuntikulutusta, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.5	4.0	4.0	13.5	3.5	0.0	88.5	0.0	0.0	86.0
PCA	0.5	12.0	11.5	2.5	2.5	0.5	88.0	0.5	0.0	86.0
SS	0.5	4.0	4.0	13.5	3.0	0.0	88.5	0.5	0.0	86.0
MMS	0.5	0.5	4.0	18.5	1.5	0.0	88.5	0.0	0.0	86.0
RS	0.5	4.0	4.0	13.5	2.0	0.0	88.0	0.5	0.0	86.0

Taulukko 3: Tunnistusprosentti käyttäen pätehtehon tuntikulutusta, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	5.2	5.3	2.9	21.2	9.8	1.2	22.4	2.7	3.3	95.0
PCA	4.7	14.2	13.7	1.6	17.6	1.4	23.3	2.6	4.1	93.3
SS	5.2	8.9	2.9	21.3	7.6	1.2	22.3	4.3	4.9	95.0
MMS	5.2	6.6	2.7	28.7	10.7	1.2	22.1	3.0	3.3	94.9
RS	5.2	8.7	2.8	21.9	7.8	1.2	22.3	4.2	5.7	95.0

Taulukko 4: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	96.5	99.5	14.5	40.0	97.5	94.5	0.0	99.0	50.5	100.0	94.5
PCA	71.5	85.5	9.0	37.0	72.5	73.0	0.0	87.0	94.5	63.5	62.0
SS	96.5	23.5	19.0	34.5	95.5	94.5	0.0	95.5	44.5	98.0	94.5
MMS	96.5	95.0	15.0	40.0	86.5	94.5	0.0	94.0	47.0	95.0	94.5
RS	96.5	20.0	17.0	41.0	14.0	94.5	1.0	1.0	49.5	0.5	94.5

Taulukko 5: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	93.0	98.6	17.2	42.3	80.2	97.8	8.6	98.9	94.7	100.0	97.9
PCA	97.9	99.1	14.1	9.5	78.8	98.9	9.7	99.0	98.7	98.0	97.9
SS	93.0	40.2	18.7	43.3	94.5	97.8	8.1	92.9	92.5	91.6	97.9
MMS	93.0	98.0	17.3	42.6	70.4	97.8	9.1	98.6	92.5	99.0	97.9
RS	93.0	29.8	16.0	43.3	59.2	97.8	5.3	41.9	94.8	6.0	97.9

Tulostaulukoista 2-5 voimme havaita, että tapaukset, joissa kaikkien johtimien sisään- ja ulostulo ovat kytketty ristiin, saadaan paremmin havaittua käyttämällä energiarekisterien suhteellisia osuuksia pätehtehon tuntikulutukseen verrattuna. Pätehtehon tuntikulutusta käyttämällä LOF-tekniikan tunnistustarkkuus oli 88,5 %, kun virheellisen mittausdatan määrä oli 20 ja MinSum-tekniikan tunnistustarkkuus oli jopa 95 %, kun virheellisen mittausdatan määrä oli 100. Suurin osa tekniikoista kuitenkin epäonnistuu tunnistamaan virheelliset mittausdatat, kun käytössä on pätehtehon tuntikulutus. Vaikka osa tekniikoista toimi myös käyttämällä pätehtehon

tuntikulutusta, parhaat tulokset saatiin käyttämällä rekisterien suhteellisia osuuksia. Kaikista parhaiten virhekytkentää simuloivat mittausdatat tunnisti MinA+ -tekniikka, joka tunnisti kaikki virheelliset mittausdatat kun mittausdatan esitystapana oli rekisterien prosentuaaliset osuudet ilman muunnosta. OCSVM sekä CBLOF tunnistivat virhekytkentää simuloivat mittausdatat lähes yhtä hyvin kuin MinA+ -tekniikka ja ne saavuttivat noin 99 % tunnistustarkkuuden virheellisen mittausdatan määrästä riippumatta.

5.2. 3P4W-mittari, kaksi I/O kytketty ristiin

Taulukko 6: Tunnistusprosentti käyttäen päätöshon tuntikulutusta, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	2.5	0.0	0.0	87.5	0.0	0.0	84.5
PCA	0.0	1.0	0.5	0.0	0.0	0.0	88.0	0.0	0.0	82.0
SS	0.0	0.0	0.0	2.5	0.0	0.0	87.5	0.0	0.0	84.5
MMS	0.0	0.0	0.0	2.5	0.0	0.0	87.5	0.0	0.0	85.0
RS	0.0	0.0	0.0	2.5	0.0	0.0	87.5	0.0	0.0	84.5

Taulukko 7: Tunnistusprosentti käyttäen päätöshon tuntikulutusta, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	2.7	0.0	0.0	19.6	0.0	0.0	93.6
PCA	0.1	3.0	3.3	0.4	1.3	0.1	21.2	0.0	0.1	91.3
SS	0.0	0.0	0.0	2.7	0.0	0.0	19.4	0.0	0.0	93.6
MMS	0.0	0.0	0.0	7.3	0.1	0.0	19.3	0.0	0.0	93.6
RS	0.0	0.0	0.0	2.8	0.1	0.0	19.2	0.0	0.0	93.6

Taulukko 8: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	96.5	99.0	10.5	43.0	96.5	95.5	0.0	97.0	28.0	100.0	94.5
PCA	59.5	83.5	5.5	70.5	62.5	59.5	0.0	85.5	91.0	46.0	46.5
SS	96.5	18.0	15.0	18.5	96.0	95.5	1.0	96.5	47.5	96.5	94.0
MMS	96.5	94.5	10.5	46.0	84.5	95.5	0.0	95.5	32.0	95.0	94.0
RS	96.5	8.5	8.5	46.0	5.0	95.5	0.0	0.5	38.0	0.5	94.0

Taulukko 9: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	91.6	98.3	16.4	42.2	79.6	97.6	9.6	98.3	90.3	99.8	97.4
PCA	96.9	99.0	14.9	31.5	98.8	98.2	6.7	99.0	97.9	96.5	95.9
SS	91.6	36.8	17.3	43.7	84.1	97.6	10.2	91.7	88.5	91.0	97.4
MMS	91.6	98.0	16.5	42.9	88.6	97.6	9.5	97.7	87.6	99.0	97.4
RS	91.6	17.4	13.1	43.7	40.7	97.6	13.9	48.5	91.0	0.8	97.4

Tulostaulukosta 6-9 voidaan havaita, että virhekytkentä, jossa kolmivaiheisessa nelijohdinmittarissa kaksi johdinta on kytketty ristiin, on mahdollista havaita melko hyvin eri poikkeavuuksien havaitsemistekniikoilla. Parhaiten virhekytkennät tunnisti MinA+ -tekniikka kun mittausdata oli esitetty energiarekisterien suhteellina osuuksina ilman muunnosta. MinA+ -tekniikka tunnisti kaikki virhekytkentää simuloivat mittausdatat kun alkuperäisen datan sekaan oli sekoitettu 20 virhekytkentää simuloivaa mittausdataa. 100 simuloitun mittausdatan tapauksessa tekniikka tunnisti 99,8 % simuloitusta mittausdatasta. OCSVM sekä CBLOF

tunnistivat lähes yhtä monta virhekytkentää kuin MinA+ tekniikka, kun data esitettiin rekisterien prosentuaalisina osuuksina. Käytettäessä datan esitysmuotona päätötehon tuntikulutusta LOF-tekniikka toimi melko hyvin saavuttaen jopa 88 % tunnistustarkkuuden, kun datan seassa oli 20 virhekytkentää. MinSum-tekniikka saavutti 93,6 % tunnistustarkkuuden kun datan seassa oli 100 virheellistä mittausdataa. Yleisesti ottaen käyttämällä datan esitysmuotona rekisterien suhteellisia osuuksia virheelliset mittausdatat oli helpompi tunnistaa ja tunnistustarkkuudet olivat suurempia, vaikka muutama tekniikka tunnisti virhekytkentöjä myös päätötehon tuntikulutuksesta.

5.3. 3P4W-mittari, yksi I/O kytketty ristiin

Taulukko 10: Tunnistusprosentti käyttäen päätöksen tuntikulutusta, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCA	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
SS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MMS	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
RS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Taulukko 11: Tunnistusprosentti käyttäen päätöksen tuntikulutusta, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.7
PCA	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.7
SS	0.0	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.9
MMS	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.7
RS	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.4

Taulukko 12: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	0.5	0.0	0.5	0.0	0.5	0.5	0.5	0.0	0.0	0.0	0.0
PCA	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0
SS	0.5	0.5	0.5	0.0	0.5	0.5	0.5	0.5	0.0	0.5	0.0
MMS	0.5	0.0	0.5	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0
RS	0.5	0.0	0.0	0.5	0.0	0.5	0.5	0.0	0.0	0.0	0.0

Taulukko 13: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	0.1	0.4	0.5	0.2	0.2	0.4	0.4	0.3	0.2	0.4	0.6
PCA	0.1	0.3	0.5	0.3	0.2	0.3	0.4	0.3	0.2	0.2	0.6
SS	0.1	0.1	0.6	0.4	0.2	0.4	0.4	0.1	0.0	0.1	0.6
MMS	0.1	0.4	0.5	0.2	0.2	0.4	0.5	0.4	0.2	0.4	0.6
RS	0.1	0.2	0.4	0.5	0.3	0.3	0.6	0.2	0.1	0.2	0.6

Tulostaulukoista 10-13 voidaan havaita, että virhekytkentää, jossa kolmivaiheisen nelijohdinmittarin yhden johtimen sisään- ja ulostulo on kytketty ristiin, on hyvin vaikea havaita energiasarjojen perusteella, kun sähkömittari on konfiguroitu netottavaksi. Parhaana tekniikkana LOF kykeni löytämään keskimäärin vain 1.7 % virheellisestä datasta käyttäen min-max skaalattua päätöksen tuntikulutusta.

5.4. 3P3W-mittari, yksi I/O kytketty ristiin

Taulukko 14: Tunnistusprosentti käyttäen päätötehon tuntikulutusta, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	0.0	0.0	0.0	48.5	0.0	0.0	48.0
PCA	0.0	0.5	0.5	0.0	0.0	0.0	49.0	0.0	0.0	45.0
SS	0.0	0.0	0.0	0.0	0.0	0.0	48.5	0.0	0.0	48.0
MMS	0.0	0.0	0.0	0.0	0.0	0.0	48.5	0.0	0.0	48.0
RS	0.0	0.0	0.0	0.0	0.0	0.0	48.5	0.0	0.0	48.0

Taulukko 15: Tunnistusprosentti käyttäen päätötehon tuntikulutusta, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	0.0	0.0	0.0	0.0	0.0	25.3	0.0	0.0	48.6
PCA	0.0	1.1	1.5	0.0	0.3	0.0	32.0	0.0	0.0	48.3
SS	0.0	0.0	0.0	0.0	0.0	0.0	25.3	0.0	0.0	48.2
MMS	0.0	0.0	0.0	0.0	0.0	0.0	25.4	0.0	0.0	48.1
RS	0.0	0.0	0.0	0.0	0.0	0.0	25.3	0.0	0.0	48.1

Taulukko 16: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	50.0	90.5	17.5	26.5	73.5	31.0	1.5	91.0	71.0	90.5	32.0
PCA	57.5	93.5	14.0	52.0	84.0	52.5	32.0	93.5	52.5	52.5	48.5
SS	50.0	31.0	27.0	26.5	48.5	31.0	5.5	49.5	67.0	50.0	32.0
MMS	50.0	90.0	21.0	26.5	70.0	31.0	0.5	86.0	70.0	84.0	32.0
RS	50.0	22.5	12.5	26.5	53.0	31.5	9.0	52.5	64.0	52.5	32.0

Taulukko 17: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	88.5	96.7	23.4	38.3	79.9	72.6	3.5	96.3	90.5	95.6	70.5
PCA	92.7	97.1	18.2	49.9	73.0	71.1	3.9	96.7	91.5	47.8	49.3
SS	88.5	66.0	39.9	33.6	81.6	72.6	7.0	86.6	91.3	86.2.0	70.5
MMS	88.5	96.0	25.1	38.3	87.2	72.6	3.5	96.5	91.0	93.7	70.5
RS	88.5	41.8	20.8	25.6	58.7	72.7	4.2	60.1	89.8	47.8	70.5

Tulostaulukosta 14-17 voidaan havaita, että kyseinen virhekytkentä on helpompi havaita esittämällä mittausdata energiarekisterien suhteellisina osuuksina verrattuna päätötehon tuntikulutukseen. Parhaiten virhekytkennät havaitsivat CBLOF ja OCSVM –tekniikat, kun data oli muutettu kaksiulotteiseksi PCA:n avulla. CBLOF tunnisti keskimäärin 93,5 % virhekytkennöistä, kun virhekytkentää simuloivien mittausdatojen määrä oli 20 ja 96,7 % kun simuloituja mittausdatoja oli 100 kappaletta. OCSVM tunnisti keskimäärin 93,5 % virhekytkentää simuloivista mittausdatoista kun simuloituja mittausdatoja oli alkuperäisen mittausdatan seassa 20

kappaletta ja 97,1 %, kun simuloitun mittausdatan määrä oli 100 kappaletta. MinA+ -tekniikka suoriutui vain hieman huonommin OCSVM sekä CBLOF -tekniikoihin verrattuna. Parhaiten MinA+ -tekniikka tunnisti virhekytkennät käyttämällä alkuperäistä dataa ilman muunnoksia. MinA+ tunnisti 90,5 % simuloituista virhekytkennöistä kun simuloitun datan määrä oli 20 kappaletta ja 95,6 % simuloitun datan määrän ollessa 100.

Kyseisessä virhekytkentätapauksessa voidaan huomata, että useiden tekniikoiden tunnistustarkkuus nousee huomattavasti, kun virhekytkentää simuloivan mittausdatan suhteellinen osuus kasvaa ja kun data esitetään rekisterien suhteellisina osuuksina. Esimerkiksi Isolation Forest -tekniikka kykeni tunnistamaan parhaimmillaan keskimäärin 71,0 % simuloitusta mittausdatan määrän ollessa 20, mutta määrän ollessa 100 tekniikka kykeni tunnistamaan parhaimmillaan 91,5 % virhekytkennöistä. PCA -tekniikan tulos nousi jopa 35,2 % ja Elliptic Envelope -tekniikan tulos nousi myös huomattavasti, kun virhekytkentöjä simuloivien mittausdatojen määrä nousi kahdestakymmenestä sataan.

5.5. 3P3W-mittari, yksi I/O kytketty ristiin, epätasainen kuorma

Taulukko 18: Tunnistusprosentti käyttäen päätöksen tuntikulutusta, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	0.0	1.5	1.5	6.5	0.5	0.0	46.0	0.0	0.0	46.5
PCA	0.5	4.5	8.0	0.5	0.5	0.5	46.0	0.0	0.0	44.5
SS	0.0	1.0	1.5	6.5	0.5	0.0	46.0	0.5	0.0	46.5
MMS	0.0	0.0	1.5	7.0	0.0	0.0	46.0	0.0	0.0	46.5
RS	0.0	1.0	1.5	6.5	1.0	0.0	46.0	0.0	0.0	46.5

Taulukko 19: Tunnistusprosentti käyttäen päätöksen tuntikulutusta, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinSum
Orig.	2.8	5.4	3.3	7.9	4.2	0.8	28.1	1.7	0.1	48.8
PCA	2.7	10.6	10.4	1.8	6.2	1.1	34.5	1.5	1.4	42.9
SS	2.8	5.1	3.6	8.1	4.0	0.8	28.0	1.2	0.1	48.8
MMS	2.8	3.5	3.4	10.7	4.1	0.8	28.0	1.4	0.1	48.8
RS	2.8	4.7	3.6	8.3	4.7	0.8	28.1	1.7	0.1	48.6

Taulukko 20: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 20

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	48.5	30.0	48.5	85.0	33.5	52.0	13.0	31.5	73.5	54.0	50.5
PCA	22.5	30.5	56.5	59.0	23.0	21.0	4.0	30.5	48.5	14.0	14.5
SS	48.5	48.0	50.0	72.5	50.5	52.0	15.5	47.5	68.5	57.0	50.5
MMS	48.5	25.0	49.5	85.0	35.5	52.0	13.5	30.5	72.0	57.5	50.5
RS	48.5	61.0	58.5	81.0	15.5	52.0	7.0	9.5	71.5	11.5	50.5

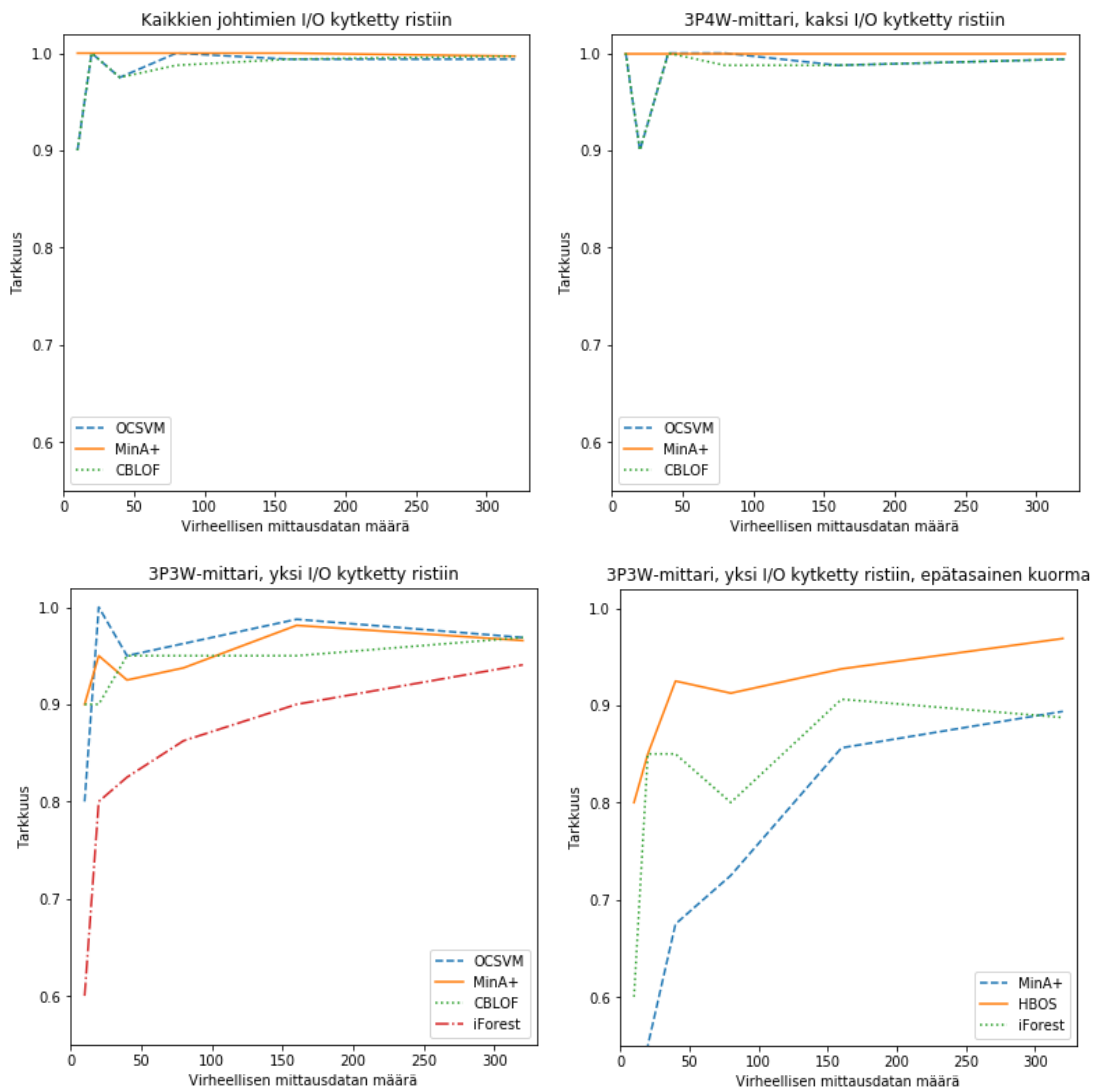
Taulukko 21: Tunnistusprosentti käyttäen rekisterien suhteita, virhedatan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Orig.	74.0	67.3	76.1	93.7	62.2	87.9	4.0	69.7	89.1	80.3	88.1
PCA	62.2	65.9	58.3	74.1	52.2	67.7	2.0	68.3	67.9	46.4	45.4
SS	74.0	81.2	80.0	90.3	73.5	87.9	6.1	74.1	88.4	84.0	88.1
MMS	74.0	64.9	77.2	93.6	61.7	87.9	5.0	67.3	89.0	80.6	88.1
RS	74.0	70.3	67.8	88.4	56.0	87.9	6.9	52.0	86.9	48.3	88.1

Tulostaulukosta 18-21 voidaan nähdä, että tässäkin virhekytkentätapauksessa väärin kytketyn mittarin tuottamaa mittausdataa simuloivat datapisteet tunnistetaan paremmin esittämällä data rekisterien suhteellisina osuuksina. Se kuinka hyvin eri poikkeavuuksien havaitsemistekniikat tunnistavat virheellisesti kytketyn mittarin tuottamaa mittausdataa vaihtelee suuresti eri tekniikoiden ja virheellisen mittausdatan suhteellisen määrän kesken. Parhaiten virhekytkennät tunnisti HBOS-tekniikka, joka tunnisti parhaimmillaan 85,0 % simuloidusta mittausdatasta virhedatan määrän ollessa 20 ja 93,7 % virheellisen mittausdatan määrän ollessa 100.

5.6. Virheellisen datan määrän vaikutus tunnistustarkkuuteen

Tässä luvussa on tutkittu tarkemmin, miten virheellisen mittausdatan määrä vaikuttaa tekniikoiden tunnistustarkkuuteen eri virhekytkentätapauksissa lukuun ottamatta 3P4W-mittarin yhden johtimen virhekytkentää. Kuvassa 20 on testattu suurimmat tunnistustarkkuudet saavuttaneita tekniikoita käyttäen mittausdatan esitysmuotona rekisterien suhteellisia osuuksia ilman esiprosessointia. Kuvasta voidaan huomata, että virhekytkentätapauksessa, jossa kaikkien johtimien sisään- ja ulostulo on kytketty ristiin ja 3P4W-mittarin kahden johtimen virhekytkennässä, virheellisen datan määrällä ei ole juurikaan vaikutusta tunnistustarkkuuteen. 3P3W-mittarin virhekytkennöissä virheellisen datan määrän kasvu näyttäisi lisäävän tekniikoiden tunnistustarkkuutta.



Kuva 20: Parhaiden tekniikoiden tunnistustarkkuudet eri virhekytkentätapauksissa

6. Pohdinta

Kaikkiaan tulostaulukoista voidaan nähdä, että esitettäessä mittausdata päätötehon tuntikulutuksena, testattujen poikkeavuuksien havaitsemistekniikoiden on hyvin vaikea tunnistaa virhekytkentöjä simuloivia mittausdatoja oikean mittausdatan joukosta. Jokaisessa virhekytkentätapauksessa paras tunnistusprosentti saatiin käyttämällä mittausdatan esitystapana energiarekisterien suhteellisia osuuksia päätötehon tuntikulutuksen sijaan. Energiarekisterien suhteellisia osuuksia käyttäessä poikkeavuuksien havaitsemistekniikat vaativat myös vähemmän laskentatehoa vektoreiden ollessa neliulotteisia 23-ulotteisten päätötehon tuntikulutusvektoreiden sijaan. Lisäksi energiarekisterien suhteelliset osuudet saadaan muodostettua, vaikka mittarilta tulisi päivän aikana vain yksi mittaustulos sähköyhtiön luentajärjestelmään, kun taas päätötehon tuntikulutusvektorin muodostukseen vaaditaan sähkömittarilta jokainen päivän aikana lähetetty mittaustulos.

Virhekytkentätapauksessa, jossa jokaisen johtimen sisään ja ulostulo on kytketty ristiin, virhekytkentää simuloivat mittausdatat tunnisti parhaiten MinA+ -tekniikka, joka tunnisti keskimäärin 100 % virheellisestä mittausdatasta kun virheellisen mittausdatan määrä oli 20 ja 100, ja mittausdatan esitysmuotona käytettiin energiarekisterien suhteellisia osuuksia. OCSVM ja CBLOF tunnistivat lähes kaikki virhekytkentätapaukset käyttämällä mittausdatan esitystapana energiarekisterien suhteellista osuutta, mutta ne eivät tunnistaneet yhtä suurta osuutta kuin MinA+ -tekniikka. Tämä virhekytkentätapaus on melko triviaali tunnistaa, sillä kyseisessä virhekytkennässä sähkön kulutus mitataankin tuotantona. Vain harvalla oikein kytketyllä sähkömittarilla A- -rekisteri on suuri, jolloin kyseiset virhekytkennät erottuvat helposti oikein kytkettyjen laitteiden joukosta. MaxA- -tekniikka löysi 94,5 % ja 97,9 % kun käytössä oli energiarekisterien suhteelliset osuudet. Testien perusteella kuitenkin MinA+ -tekniikka toimi kaikista parhaiten tässä virhekytkentätapauksessa.

Virhekytkennässä, jossa 3P4W-mittarin kahden johtimen sisään ja ulostulo on kytketty ristiin, parhaiten virhekytkentää simuloivat mittausdatat tunnisti jälleen MinA+ -tekniikka käyttämällä mittausdatan esitysmuotona energiarekisterien

suhteellisia osuuksia. MinA+ -tekniikka tunnisti keskimäärin 100 % virheellisestä mittaustulosta, kun virheellistä mittaustulosta oli 20 kappaletta ja 99,8 % kun virheellistä mittaustulosta oli 100 kappaletta. Moni muu tekniikka, kuten PCA, OCSVM, EE, CBLOF sekä MaxA- tunnistivat yli 95 % kun data oli esitettyä energiarekisterien suhteellisina osuuksina, mutta yksikään niistä ei tunnistanut yhtä suurta osuutta kuin MinA+ -tekniikka. Kyseinen virhekytkentätapaus on melko triviaalia havaita, sillä väärin kytketty mittari muuttaa kaikkien rekisterien arvon $-1/3$ alkuperäisestä, jolloin pätötehon kulutus näyttää mittarille pieneltä tuotannolta. Tämä virhekytkentä tapaus on kuitenkin mahdollista sekoittaa tapaukseen, jossa esimerkiksi kesämökille on asennettu aurinkopaneeleita, jolloin kyseinen kohde voi tuottaa hyvin samantapaista mittaustulosta kuin kyseinen väärinkytkentä. Rekisterien prosentuaaliset osuudet vastaavat käytännössä virhekytkentätilannetta, jossa kaikkien johtimien sisääntulo ja ulostulo on kytketty ristiin, joten näiden kahden virhekytkentätapausten tulokset ovat hyvin samankaltaisia.

Virhekytkentätapauksessa, jossa netottavaksi konfiguroidun 3P4W-mittarin yhden johtimen sisään ja ulostulo on kytketty ristiin, yksikään testattu poikkeuksien havaitsemistekniikka ei kyennyt tunnistamaan virhekytkentää simuloivia mittaustulosta luotettavasti. Parhaimmillaankin tunnistustarkkuus jäi alle kahden prosentin. Tämä on sinänsä ymmärrettävää, sillä kyseisessä virhekytkentätapauksessa kaikki mittarin mittaustulokset pienenevät kolmanneksella, jolloin virhekytkentä on erittäin helppo sekoittaa oikein kytkettyyn mittariin, jossa kulutus on pientä. Rekisterien suhteelliset osuudet pysyvät myös täysin samana oikein asennettuun mittariin verrattuna. Tätä virhekytkentää onkin mahdotonta havaita vain energiarekisterien tai pätötehon tuntikulutuskäyrän perusteella. Kyseisen virhekytkennän havaitsemiseen tarvittaisiin tarkempaa tietoa vaihekohtaisesta jännitteestä, virrasta sekä vaihekulmasta. Mikäli 3P4W-mittari ei ole konfiguroitu netottavaksi, MinA+ sekä MaxA- -tekniikat voisivat tunnistaa kyseisellä tavalla väärin kytkettyjä sähkömittareita, sillä ei-netottavan 3P4W-mittarin molemmat A-rekisterit voivat kasvaa yhtä aikaa, jolloin yhden johtimen aiheuttama tuotanto ei jäisi piiloon vaan tallentuisi A- -rekisteriin.

3P3W-mittarin virhekytkentätapaus, jossa L1 tai L3 -johtimen sisään ja ulostulo on kytketty ristiin, tuotti mielenkiintoisia tuloksia. Parhaiten virhekytkennät tunnisti

CBLOF-tekniikka, joka tunnisti keskimäärin 93,5 % virheellisestä mittausdatasta, kun virheellistä mittausdataa oli 20 kappaletta ja 96,7 %, kun virheellistä mittausdataa oli 100 kappaletta. Tulokset saatiin kun mittausdata esitettiin energiarekisterien suhteellisina osuuksina ja jolle oli tehty PCA-muunnos, joka muutti datan kaksiulotteiseksi. MinA+ -tekniikka tunnisti lähes yhtä suuren osan virhekytkennöistä, 90,5 % ja 95,6 %, kun mittausdata esitettiin energiarekisterien suhteellisina osuuksina ilman muunnosta. Se, onko ero tilastollisesti merkittävä, tulisi vielä testata, mutta vaikuttaa siltä, että CBLOF-tekniikka tunnistaisi kyseisiä virhekytkentöjä hieman MinA+ -tekniikkaa paremmin.

Viimeinen virhekytkentätapaus oli 3P3W-mittarin L1 tai L3 -johtimen sisään ja ulostulon kytkeminen ristiin mittarissa, jonka johtimien kuormat eivät olleet balanssissa keskenään. Kyseisen virhekytkentätapauksen tunnisti kaikista parhaiten HBOS-tekniikka, joka tunnisti 85,0 % ja 93,7 % virheellisen mittausdatan määrän ollessa 20 ja 100. Toinen tekniikka, joka tunnisti kyseisen virhekytkennän hyvin, oli isolation forest, joka tunnisti 73,5 % ja 89,1 %, kun mittausdata esitettiin energiarekisterien suhteellisina osuuksina. Molempien tekniikoiden tapauksessa voidaan huomata, että tunnistuksen tarkkuus parani kun virhedatan suhteellinen osuus kasvoi. Tätä voi kuitenkin osaksi selittää se, että oikeaksi merkittävän mittausdatan joukossa saattaa olla valmiiksi virheellisesti kytkettyjä sähkömittareita, joita ei ole havaittu aiemmin.

Aiemmat virhekytkentätapaukset hyvin tunnistanut MinA+ -tekniikka ei tunnistanut 3P3W-mittarin virhekytkentää yhtä hyvin, kun johtimien kuormat eivät olleet balanssissa. Se tunnisti vain 57,0 % ja 84,0 % virheellisestä datasta, kun käytössä oli rekisterien suhteelliset osuudet, jotka oli normalisoitu. MaxA- -tekniikka tunnisti kyseisen virhekytkennän lähes yhtä hyvin kuin MinA+ -tekniikka. MaxA- -tekniikka tunnisti 50,5 % ja 88,1 % virheellisestä datasta kun data esitettiin energiarekisterien suhteellisina osuuksina.

Yleisesti 3P3W-mittarin virhekytkennässä tapaukset, jossa johtimien kuormat ovat suhteellisen balansoitu, ovat yleisempiä kuin tilanne, jossa johtimien kuormat ovat erittäin epäbalanssissa. Tämän vuoksi 3P3W-mittarin virhekytkennän tunnistamisessa tulokset, jotka saatiin testaamalla virhekytkentää balansoidulla

kuormalla, ovat oleellisemmat verrattuna tuloksiin, jotka saatiin simuloimalla epäbalansoitua kuormaa. Mikäli poikkeavuuksien havaitsemistekniikoita käytettäisiin oikeasti etsimään virheellisesti kytkettyjä 3P3W-mittareita, on tärkeää, että sama tekniikka löytää virheellisesti kytketyt mittarit, riippumatta siitä, ovatko johtimien kuormat balanssissa vai ei. Tulostaulukoista voidaan nähdä, että CBLOF-tekniikka, joka tunnisti suuren osan virhekytkennöistä balansoidulla kuormalla, ei tunnistanut virhekytkentää kovinkaan luotettavasti, kun johtimien kuormat eivät olleet balanssissa. Samoin kävi OCSVM ja HBOS-tekniikoille. Tekniikat tunnistivat toisessa virhekytkentätapauksessa suuren osan virhekytkennöistä, mutta eivät kyenneet yhtä hyvään tunnistusprosenttiin toisessa tapauksessa. Isolation forest -tekniikka tunnisti molemmissa tapauksissa virhekytkennät verrattain hyvin ja se löysi 71,0 % ja 90,5 % virheellisistä mittausdatoista tasaisella kuormalla virheellisen datan määrän ollessa 20 sekä 100 ja käytetyn mittausdatan esitysmuotona käytettäessä energiarekisterien suhteellisia osuuksia. Epätasaisella kuormalla tekniikka tunnisti 73,5 % ja 89,1 %. Keskimääräinen tunnistusprosentti tekniikalla oli neljässä eri testatussa 3P3W-mittarin virhekytkentää simuloivassa testauksessa täten yhteensä 81,0 %. MinA+ -tekniikka tunnisti myös verrattain luotettavasti kaikkia erityyppisiä virhekytkentätapauksia 3P3W-mittarin tapauksessa ja tasaisen kuorman tapauksessa se tunnistikin virheellisen mittausdatan isolation forest -tekniikkaa paremmin.

Paras yleinen tekniikka virhekytkentöjen etsimiseen näyttäisi testauksen perusteella olevan MinA+ -tekniikka, jossa yksinkertaisesti etsitään mittareita, joissa A+ rekisterin suhteellinen osuus on kaikista pienin. Se tunnisti lähes kaikki virhekytkentää simuloivat mittausdatat, kun kyseessä olivat tapaus, jossa kaikkien vaiheiden johtimien sisään ja ulostulot olivat kytketty ristiin ja tapaus, jossa netottavan 3P4W-mittarin kahden johtimen sisään- ja ulostulo olivat kytketty ristiin ja mittausdata esitettiin energiarekisterien suhteellisina osuuksina. 3P3W-mittarin virhekytkennän tapauksissa osa muista tekniikoista tunnisti suuremman osan virhekytkentää simuloivista mittausdatoista, mutta vain isolation forest -tekniikka tunnisti lähes yhtä suuren osan virheellisestä mittausdatasta. 3P3W-mittarin virhekytkennän tapauksessa MinA+ -tekniikan tunnistustarkkuus oli keskimäärin 80,1 % neljässä eri testitilanteessa, kun isolation forest -tekniikan keskimääräinen tunnistustarkkuus oli 81,0 %. Todellisuudessa kuitenkin tasaisen kuorman virhekytkentätapauksen tunnistaminen on oleellisempaa ja MinA+ -tekniikka toimi

huomattavasti isolation forest -tekniikkaa paremmin kyseisessä tilanteessa. Vaikka keskimääräinen tunnistustarkkuus oli isolation forest -tekniikalla suurempi 3P3W-mittarin virhekytkennän tapauksessa, kyseinen tekniikka ei toiminut muissa virhekytkentätapauksissa yhtä hyvin kuin MinA+, jolloin sitä voidaan soveltaa vain 3P3W-mittareiden virhekytkentöjen havaitsemiseen. MinA+ puolestaan sopii kaikkien virhekytkentätapausten tunnistamiseen vaikeasti havaittavaa netottavan 3P4W-mittarin yhden johtimen väärinkytkentää lukuun ottamatta.

Datan esiprosessointitekniikoilla ei saavutettu merkittävästi parempia tuloksia alkuperäiseen dataan verrattuna. Muutamissa tapauksissa pääkomponenttianalyysillä sekä normalisoinnilla saavutettiin korkeampi tarkkuus, mutta erot olivat hyvin pieniä ja esiprosessointitekniikoiden tuoma hyöty vaihteli suuresti eri virhekytkentätapauksissa. Kokonaisuudessaan käytetyt datan esiprosessointitekniikat eivät lisänneet merkittävästi testattujen poikkeavuuksien havaitsemistekniikoiden tarkkuutta.

Testauksessa käytettyä menetelmää voidaan hyödyntää oikeassa ympäristössä havaitsemaan N epäilyttävintä mittaria, mutta oikeassa ympäristössä ei ole tietoa kuinka monta mittaria on asennettu virheellisesti. Asettamalla havaitsemistekniikoiden etsimien poikkeavuuksien määrän, käyttäjä voi hallita kuinka monta epäilyttävintä mittaria hän haluaa etsiä mittausdatan joukosta. Saatavat tulokset kannattaa vielä tutkia käsin, jotta saadaan varmuus mittarin asennusvirheestä. Poikkeavuuksien havaitsemistekniikat eivät siis itsessään anna varmuutta sille, onko mittari asennettu väärin vai ei. MinA+ ja MaxA- -tekniikoiden tapauksessa on kuitenkin mahdollista asettaa algoritmeille ennalta asetetut rajat, jonka mukaan tunnistus tehdään eikä vain etsiä N kappaletta epäilyttävimpiä mittareita. Asettamalla MinA+ -tekniikalle kynnykseksi esimerkiksi 0,1 tekniikka olisi merkannut epäilyttäväksi kaikki mittarit, joiden A+ -rekisterin suhteellinen osuus on alle 10 % mittarin läpi kulkeneesta kokonaistehosta.

Olisi myös mahdollista rakentaa esimääritellyt rajat, jossa tietyn rajan sisällä olevat tulokset merkataan eri vakavuusasteilla, kuten esimerkiksi asettamalla tulokselle 0 - 0,05 vakavuusasteen kriittinen, 0,05 - 0,10 vakavuusasteen suuri ja niin edelleen. Tämä mahdollistaa myös automaation, kun kaikki mittarit merkataan epäilyttäväksi,

jossa A+ -rekisterin suhteellinen osuus on alle määritellyn rajan. Tämä estää myös sen, että tilanteessa, jossa kaikki sähköyhtiön mittarit toimivat oikein, N kappaletta epäilyttävintä mittaria merkataan viallisiksi, vaikka kaikkien mittareiden A+ -rekisterien suhteelliset osuudet ovat täysin kelvollisia.

MinA+ -tekniikan tunnistustarkkuutta on myös mahdollista lisätä määrittelemällä energiarekistereille rajat, jonka alle jääviä mittareita ei oteta mukaan analysointiin. Jos mittari on mitannut vain hyvin vähän kulutusta, on mahdollista, että rekisterien suhteelliset osuudet ovat tavallisesta poikkeavat. Esimerkiksi tilanne, jossa mittarin rekisterien arvot ovat $[0.5, 0, 0, 2]$ on täysin mahdollinen, eikä siinä ole mitään epäilyttävää, mutta MinA+ -tekniikka luultavasti merkitsisi kyseisen mittaustuloksen epäilyttäväksi. Poistamalla analysoinnista kohteet, jossa kulutus on ollut hyvin pientä, estetään MinA+ -tekniikkaa tunnistamaan valideja mittaustuloksia virhekytkennäksi.

Poikkeuksien havaitsemistekniikat ovat vertailukelpoisia vain keskenään ja tulostaulukoissa olevat tunnistustarkkuudet ovat vain suuntaa antavia tuloksia. On mahdollista, että käyttämällä toista datasettiä tulokset olisivat voineet olla hieman erit. Ei myöskään ole varmuutta onko käytetyn mittaustulosten seassa virheellisesti asennettuja mittareita, jolloin vaikka poikkeuksien havaitsemistekniikka olisi luokitellut kyseisen mittarin tuottaman mittaustuloksen poikkeavuudeksi, se olisi tulkittu virheelliseksi luokitteluksi. Testauksen tarkoituksena oli vertailla eri poikkeuksien havaitsemistekniikoita ja datan esiprosessointimenetelmiä keskenään, jolloin vain tulosten eroilla on merkitystä absoluuttisen tunnistusprosentin sijaan.

Tutkimuksen pohjalta olisi mahdollista tehdä paljon jatkotutkimusta. Kokeellisessa osiossa testattiin vain pieni osa kaikista saatavilla olevista poikkeavuuksien havaitsemistekniikoista. Tekniikoiden vertailua voitaisiin laajentaa moniin muihin tekniikoihin, kuten tilastolliseen todennäköisyyteen perustuviin angle-based outlier detection sekä stochastic outlier selection -tekniikoihin tai neuroverkkoon perustuvaan auto encoderiin. Myös tässä tutkimuksessa käytetyistä menestyvistä poikkeavuuksien havaitsemistekniikoista voisi tehdä tarkempaa tutkimusta sekä tutkia miten muut mittaustulosten esitystavat vaikuttavat tekniikoiden tunnistustarkkuuteen.

7. Yhteenveto

Tutkielmassa tutkittiin, miten eri virheellisesti kytketyt 1P, 3P3W ja netottavat 3P4W-mittarit olisi mahdollista havaita etäluettavien sähkömittareiden lähettämien energiarekisterien arvojen perusteella. Tutkielmassa käsiteltiin viittä eri virhekytkentätapausta, jotka vaikuttavat sähkömittarin mittaustuloksiin ja näin ollen vaikuttavat siihen, miten asiakasta laskutetaan. Kyseiset virhekytkennät olivat tapaukset, jossa mittarin kaikkien johtimien sisään ja ulostulo on kytketty ristiin, netottava 3P4W-mittari, jossa kahden tai yhden johtimen sisään- ja ulostulo on kytketty ristiin sekä 3P3W-mittari, jossa L1 tai L3 -johtimen sisään- ja ulostulo on kytketty ristiin ja jossa johtimien kuormat ovat balanssissa tai epäbalanssissa.

Tutkielmassa vertailtiin, mikä poikkeavuuksien havaitsemistekniikka kykenee tunnistamaan suurimman osan virhekytkentää simuloivista mittausdatoista. Taulukkoon 22 on kerätty tekniikoiden tunnistustarkkuudet eri virhekytkentätilanteissa kun tunnistus on tehty käyttäen rekisterien suhteellisia osuuksia ilman esiprosessointia. MinA+ -tekniikka, jossa etsitään pienimpiä A+ -rekisterien suhteellisia osuuksia, osoittautui kaikista toimivimmaksi tekniikaksi. Sen avulla tunnistettiin kohtalaisen luotettavasti virhekytkentää simuloivat mittausdatat kaikissa muissa tapauksissa, paitsi netottavan 3P4W-mittarin yhden johtimen sisään- ja ulostulon ristiin kytkennässä. Testauksen perusteella näyttäisi siltä, että parhaiten virhekytkentöjen tunnistamiseen energiarekisterien perusteella toimii intuitiivinen tapa, etsiä pieniä A+ -rekisterin suhteellisia osuuksia sen sijaan, että tehtävään sovellettaisiin ohjaamattomia poikkeavuuksien havaitsemistekniikoita. Ainoa poikkeus oli 3P3W-mittarin yhden johtimen virhekytkentä epäbalanssilla kuormalla, jossa isolation forest sekä HBOS -tekniikat tunnistivat suuremman osan virhekytkentää simuloivista mittausdatoista. Tilanne, jossa kuorma on täysin epäbalanssissa, on kuitenkin melko harvinainen ja 3P3W-mittarit ovat käytössä vain muutamassa maassa. Tämän takia kyseinen virhekytkentätilanne on melko harvinainen, jonka takia pienten A+ -rekisterien suhteellisten osuuksien etsintä on luotettavampi tapa tunnistaa 3P3W-mittarin virhekytkentä.

Taulukko 22: Yhteenveto tekniikoiden tarkkuudesta, rekisterien suhde, virheellisen datan määrä 100

	PCA	OCSVM	KNN	HBOS	OTS	EE	LOF	CBLOF	IF	MinA+	MaxA-
Jokainen I/O	93.0	98.6	17.2	42.3	80.2	97.8	8.6	98.9	94.7	100.0	97.9
3P4W, 2 I/O	91.6	98.3	16.4	42.2	79.6	97.6	9.6	98.3	90.3	99.8	97.4
3P4W, 1 I/O	0.1	0.4	0.5	0.2	0.2	0.4	0.4	0.3	0.2	0.4	0.6
3P3W, 1 I/O	88.5	96.7	23.4	38.3	79.9	72.6	3.5	96.3	90.5	95.6	70.5
3P3W, 1 I/O, epätasainen	74.0	67.3	76.1	93.7	62.2	87.9	4.0	69.7	89.1	80.3	88.1

Energiarekisterien perusteella on mahdollista havaita vain osa virhekytkennöistä ja kaikkien mahdollisten virhekytkentöjen havaitsemiseen tarvittaisiin paljon enemmän informaatiota, joita ei automaattisesti lähetetä sähköyhtiölle, kuten vaihekohtaiset jännite-, virta- ja vaihekulmalukemat, joiden perusteella voitaisiin tehdä tarkempaa analyysia mittarin kytkennästä. [14] Energiarekisterit analysoimalla voidaan kuitenkin tunnistaa osa virhekytkentätapauksista ja automatisoimalla prosessi, voidaan tunnistaa epäilyttäviä mittareita ja saada osa väärin kytketyistä mittareista korjattua.

Viitteet

- [1] S. Darby, 2010, Smart Metering: What Potential for Householder Engagement, *Building Research and Information*, 38(5), s. 442-457
- [2] G. R. Barai, S. Krishnan & B. Venkatesh, 2015, Smart Metering and Functionalities of Smart Meters in Smart Grid - a Review, *2015 IEEE Electrical Power and Energy Conference (EPEC)*, London, s. 138-145
- [3] L. Zhou, F. Xu & Y. Ma, 2010, Impact of Smart Metering on Energy Efficiency, *2010 International Conference on Machine Learning and Cybernetics*, Qingdao, s. 3213-3218
- [4] N. Uribe-Pérez, L. Hernández, D. Vega & I. Angulo, 2016, State of the Art and Trends Review of Smart Metering in Electricity Grids, *Applied Sciences*, 6(3), s. 68-92
- [5] U.S. Department of Energy, 2016, Advanced Metering Infrastructure and Customer Systems: Results from the Smart Grid Investment Grant Program
- [6] A. Hansen, J. Staggs & S. Shenoï, 2017, Security Analysis of an Advanced Metering Infrastructure, *International Journal of Critical Infrastructure Protection*, 18(C), s. 3-19
- [7] P. Kukuča & I. Chrapčiak, 2016, From Smart Metering to Smart Grid, *Measurement Science Review*, 16(3), s. 142-148
- [8] K. C. Budka, J. G. Deshpande, T. L. Doumi, M. Madden, & T. Mew, 2010, Communication Network Architecture and Design Principles for Smart Grids, *Bell Labs Technical Journal*, 15(2), s. 205-227
- [9] A. A. Cárdenas et al., 2014, A Framework for Evaluating Intrusion Detection Architectures in Advanced Metering Infrastructures, *IEEE Transactions on Smart Grid*, 5(2), s. 906-915

- [10] M. Jin, R. Jia, Z. Kang, I. Konstantakopoulos, & C. Spanos, 2014, PresenceSense: Zero-training Algorithm for Individual Presence Detection Based on Power Monitoring, *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, Memphis, s.1-10
- [11] S. Herman, *Alternating Current Fundamentals*, 8th ed., s. 23-65, 242-276
- [12] S. Arote, G. N. Mulay & A. Khaparde, 2016, Design and Implementation of Smart Three Phase Energy Meter, *2016 International Conference on Smart Grid and Clean Energy Technologies*, Chengdu, s. 44-49
- [13] S. Kusui & M. Kogane, 1999, Apparent Power Transducer for Three-Phase Three-Wire System, *Electronics Letters*, 35(1), s. 5-7
- [14] J. Li, W. Xue, L. Gao, B. Zhang & M. Zhou, 2018, A New Method to Analyze the Error-wiring of Electric Energy Metering Device, *IOP Conference Series: Materials Science and Engineering*, 381(1), s. 1-6
- [15] Q. Zhiming, F. Wang, F. Wenli, 2015, Study in Step by Step Electric Energy Meter Connection Detection Method, *Indonesian Journal of Electrical Engineering*, 14(1), s. 90-96
- [16] R. Pelikan, 2004, Three-Phase Electricity, *National Driller*, 25(10)
- [17] T. He, S. Li, S. Wu, J. Peng & X. Guo, 2018, Research on Frequency Adaptability of Doubly-fed Induction Generator under Asynchronous Power Grid, *IOP Conference Series: Earth and Environmental Science*, 186(5) s. 1-8
- [18] A. Zimek & E. Schubert, 2017, Outlier Detection, *Encyclopedia of Database Systems*, s. 1–5
- [19] J. Han, M. Kamber & J. Pei, *Data Mining: Concepts and Techniques*, 3rd Edition, Chapter 12
- [20] V. Chandola, A. Banerjee & V. Kumar, 2009, Anomaly Detection: A Survey. *ACM Computing Surveys*, 41(3), Article 15

- [21] E. Villar-Rodriguez, J. Del Ser, I. Oregi, M. Nekane & B. Gil-Lopez, 2017 Detection of Non-Technical Losses in Smart Meter Data Based on Load Curve Profiling and Time Series Analysis, *Energy*, 137(1), s. 118-128
- [22] P. Jokar, N. Arianpoo & V. Leung, 2016, Electricity Theft Detection in AMI Using Customers' Consumption Patterns, *IEEE Transactions on Smart Grid*, 7(1), s. 216-226
- [23] S. Sahoo, D. Nikovski, T. Muso & K. Tsuru, 2015, Electricity Theft Detection Using Smart Meter Data, *2015 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT 2015)*, Washington, s. 1-5
- [24] M-L. Shyu, S-C. Chen, K. Sarinnapakorn & L. W. Chang, 2003, A Novel Anomaly Detection Scheme Based on Principal Component Classifier. *ICDM Foundation and New Direction of Data Mining workshop*, Florida, s.172-179
- [25] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola & R. Williamson, 2001, Estimating the Support of a High-Dimensional Distribution, *Neural computation*, 13(7), s. 1443–1471
- [26] M. Sugiyama & K. Borgwardt, 2013, Rapid Distance-Based Outlier Detection via Sampling, *Advances in Neural Information Processing Systems 26*, Nevada, s. 467-475
- [27] S. Ramaswamy, R. Rastogi & K. Shim, 2000, Efficient Algorithms for Mining Outliers from Large Data Sets. *ACM Sigmod Record*, 29(2), s. 427–438
- [28] M. Goldstein & A. Dengel, 2012, Histogram-based Outlier Score (HBOS): A Fast Unsupervised Anomaly Detection Algorithm, *Poster and Demo Track of the 35th German Conference on Artificial Intelligence*, Saarbrücken, s. 59-63
- [29] P. Rousseeuw & K. Driessen, 1999, A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, 41(3), s. 212-223
- [30] M. Breunig, H-P. Kriegel, R. Ng & J. Sander, 2000, LOF: Identifying Density-Based Local Outliers, *ACM Sigmod Record*, 29(2), s. 93–104

- [31] Z. He, X. Xu & S. Deng, 2003, Discovering Cluster-Based Local Outliers, *Pattern Recognition Letters*, 24(9-10), s. 1641–1650
- [32] F. Liu, K. Ting & Z-H Zhou, 2008, Isolation Forest. *IEEE International Conference on Data Mining*, Pisa, s. 413-422
- [33] L. Shalabi, S. Ziyad & B. Al-Kasasbeh, 2006, Data Mining: A Preprocessing Engine, *Journal of Computer Science*, 2(9), s. 735-739
- [34] J. Yang, S. Rahardja & P. Fränti, 2018, Mean-Shift Outlier Detection, *International Conference on Fuzzy Systems and Data Mining (FSDM)*, Bangkok, s. 208-215
- [35] V. Hautamäki, I. Kärkkäinen, & P. Fränti, 2004, Outlier Detection Using k-Nearest Neighbor Graph, *International Conference on Pattern Recognition (ICPR)*, Cambridge, s. 430-433
- [36] X. Li, J. Lv, and Z. Yi, 2018, An Efficient Representation-Based Method for Boundary Point and Outlier Detection, *IEEE Transactions on Neural Networks and Learning Systems*, 29 (1), s. 51-62

Liite 1

Esimerkki 3P3W-mittarin väärinkytkenän vaikutuksesta energiarekisterikertymään balansoidulla kuormalla käyttäen 230V ja 1A.

Oikein kytketyn mittarin tuottama teho

A+	A-	R+	R-	Vaihekulma
39859	0	0	41	0
39720	0	3511	0	5
39262	0	6955	0	10
38521	0	10273	0	15
37464	0	13700	0	20
36138	0	17006	0	25
34524	0	19910	0	30
32653	0	22962	0	35
30537	0	25660	0	40
28195	0	28220	0	45
25624	0	30540	0	50
22875	0	32675	0	55
19942	0	34534	0	60
16852	0	36106	0	65
13642	0	37473	0	70
10323	0	38494	0	75
6913	0	39258	0	80
3486	0	39708	0	85
0	0	39879	0	90
0	3476	39713	0	95
0	6916	39252	0	100
0	10315	38506	0	105
0	13639	37453	0	110
0	16847	36120	0	115
0	19930	34504	0	120
0	22868	32677	0	125
0	25625	30501	0	130
0	28196	28225	0	135
0	30543	25553	0	140
0	32663	22879	0	145
0	34529	19208	0	150
0	36136	16718	0	155
0	37470	13552	0	160
0	38511	10354	0	165
0	39272	6916	0	170
0	39718	3419	0	175
0	39871	0	25	180
0	39712	0	3475	185

0	39262	0	6921	190
0	38517	0	10306	195
0	37471	0	13730	200
0	36139	0	16986	205
0	34531	0	19932	210
0	32672	0	22737	215
0	30552	0	25627	220
0	28195	0	28245	225
0	25626	0	30509	230
0	22875	0	32690	235
0	19943	0	34526	240
0	16853	0	36128	245
0	13634	0	37496	250
0	10321	0	38504	255
0	6923	0	39279	260
0	3479	0	39727	265
0	0	0	39878	270
3473	0	0	39708	275
6928	0	0	39267	280
10325	0	0	38530	285
13631	0	0	37453	290
16850	0	0	36161	295
19929	0	0	34563	300
22877	0	0	32700	305
25628	0	0	30560	310
28198	0	0	28169	315
30554	0	0	25553	320
32658	0	0	22761	325
34531	0	0	19937	330
36138	0	0	17026	335
37468	0	0	13677	340
38518	0	0	10392	345
39268	0	0	6975	350
39730	0	0	3482	355

L1-johdin kytketty ristiin

A+	A-	R+	R-	Vaihekulma
30	0	0	22998	0
2042	0	0	22904	5
4018	0	0	22641	10
5952	0	0	22248	15
7896	0	0	21588	20
9726	0	0	20889	25
11477	0	0	19924	30
13219	0	0	18954	35

14805	0	0	17662	40
16243	0	0	16201	45
17625	0	0	14771	50
18871	0	0	13218	55
19925	0	0	11500	60
20826	0	0	9691	65
21586	0	0	7845	70
22214	0	0	5940	75
22668	0	0	3978	80
22899	0	0	1993	85
23021	0	0	0	90
22892	0	2007	0	95
22667	0	3999	0	100
22235	0	5971	0	105
21622	0	7885	0	110
20864	0	9744	0	115
19883	0	11484	0	120
18837	0	13198	0	125
17634	0	14826	0	130
16253	0	16268	0	135
14794	0	17682	0	140
13151	0	18673	0	145
11498	0	19929	0	150
9737	0	20705	0	155
7838	0	21668	0	160
5921	0	22246	0	165
3964	0	22659	0	170
1990	0	22901	0	175
0	0	22993	0	180
0	1994	22920	0	185
0	3981	22667	0	190
0	5979	22213	0	195
0	7885	21606	0	200
0	9709	20976	0	205
0	11518	19768	0	210
0	13199	18844	0	215
0	14800	17607	0	220
0	16301	16312	0	225
0	17657	14816	0	230
0	18881	13210	0	235
0	19899	11460	0	240
0	20886	9728	0	245
0	21616	7861	0	250
0	22240	5951	0	255
0	22653	3982	0	260
0	22890	1984	0	265
0	23012	0	0	270
0	22930	0	2026	275

0	22626	0	4001	280
0	22249	0	5977	285
0	21617	0	7882	290
0	20878	0	9756	295
0	19946	0	11539	300
0	18843	0	13209	305
0	17624	0	14802	310
0	16243	0	16248	315
0	14755	0	17557	320
0	13194	0	18877	325
0	11522	0	19944	330
0	9730	0	20828	335
0	7838	0	21671	340
0	5925	0	22260	345
0	3960	0	22683	350
0	1987	0	22936	355

L3-johdin kytketty ristiin

A+	A-	R+	R-	Vaihekulma
0	0	23001	0	0
0	2000	22917	0	5
0	4023	22627	0	10
0	5980	22201	0	15
0	7898	21549	0	20
0	9701	20974	0	25
0	11518	19610	0	30
0	13220	18906	0	35
0	14765	17540	0	40
0	16283	16287	0	45
0	17620	14764	0	50
0	18871	13205	0	55
0	19930	11495	0	60
0	20823	9689	0	65
0	21598	7838	0	70
0	22220	5944	0	75
0	22664	3984	0	80
0	22923	1993	0	85
0	23028	0	0	90
0	22884	0	2019	95
0	22665	0	4017	100
0	22238	0	5970	105
0	21580	0	7872	110
0	20840	0	9736	115
0	19893	0	11502	120
0	18809	0	13179	125

0	17619	0	14811	130
0	16226	0	16215	135
0	14756	0	17584	140
0	13200	0	18921	145
0	11486	0	19933	150
0	9696	0	20899	155
0	7855	0	21630	160
0	5932	0	22246	165
0	3964	0	22680	170
0	1993	0	22910	175
0	0	0	23012	180
2007	0	0	22912	185
3993	0	0	22654	190
5981	0	0	22203	195
7897	0	0	21590	200
9746	0	0	20735	205
11490	0	0	19927	210
13183	0	0	18717	215
14786	0	0	17577	220
16279	0	0	16284	225
17635	0	0	14789	230
18838	0	0	13180	235
19906	0	0	11489	240
20848	0	0	9707	245
21593	0	0	7852	250
22216	0	0	5951	255
22636	0	0	3986	260
22898	0	0	1986	265
22980	0	0	0	270
22945	0	2009	0	275
22629	0	3993	0	280
22240	0	5970	0	285
21601	0	7859	0	290
20856	0	9739	0	295
19926	0	11514	0	300
18815	0	13171	0	305
17628	0	14815	0	310
16262	0	16273	0	315
14804	0	17701	0	320
13189	0	18821	0	325
11472	0	19925	0	330
9743	0	20686	0	335
7879	0	21574	0	340
5921	0	22263	0	345
4011	0	22625	0	350
1978	0	22928	0	355

L1 & L3 -johtimet käännetty

A+	A-	R+	R-	Vaihekulma
0	39889	32	0	0
0	39728	0	3519	5
0	39273	0	6984	10
0	38532	0	10376	15
0	37485	0	13629	20
0	36143	0	16709	25
0	34541	0	19931	30
0	32676	0	22740	35
0	30550	0	25596	40
0	28194	0	28230	45
0	25637	0	30528	50
0	22878	0	32689	55
0	19946	0	34524	60
0	16856	0	36144	65
0	13632	0	37500	70
0	10318	0	38517	75
0	6922	0	39283	80
0	3473	0	39745	85
0	0	0	39897	90
3483	0	0	39740	95
6931	0	0	39264	100
10331	0	0	38503	105
13645	0	0	37488	110
16871	0	0	36142	115
19946	0	0	34523	120
22879	0	0	32650	125
25642	0	0	30582	130
28203	0	0	28148	135
30559	0	0	25701	140
32674	0	0	23001	145
34547	0	0	19940	150
36152	0	0	17005	155
37485	0	0	13582	160
38531	0	0	10321	165
39283	0	0	6871	170
39739	0	0	3489	175
39887	0	0	48	180
39735	0	3443	0	185
39287	0	6975	0	190
38512	0	10333	0	195
37478	0	13729	0	200
36150	0	16797	0	205
34543	0	19922	0	210
32672	0	22979	0	215
30559	0	25642	0	220

28198	0	28253	0	225
25627	0	30578	0	230
22869	0	32690	0	235
19932	0	34526	0	240
16841	0	36132	0	245
13635	0	37504	0	250
10323	0	38526	0	255
6910	0	39283	0	260
3469	0	39718	0	265
0	0	39864	0	270
0	3487	39717	0	275
0	6932	39254	0	280
0	10337	38518	0	285
0	13650	37467	0	290
0	16855	36151	0	295
0	19951	34508	0	300
0	22882	32690	0	305
0	25637	30533	0	310
0	28215	28214	0	315
0	30560	25578	0	320
0	32676	23001	0	325
0	34549	19941	0	330
0	36151	17023	0	335
0	37487	13590	0	340
0	38527	10323	0	345
0	39285	6881	0	350
0	39725	3445	0	355