

---

# Data-analytiikan hyödyntäminen ennakoivassa kunnossapidossa

---

Pro gradu -tutkielma  
Turun yliopisto  
Tietotekniikan laitos  
Data-analytiikka  
2022  
Janne Iltanen

TURUN YLIOPISTO  
Tietotekniikan laitos

JANNE ILTANEN: Data-analytiikan hyödyntäminen ennakoivassa kunnossapidossa

Pro gradu -tutkielma, 55 s.

Data-analytiikka

Huhtikuu 2022

---

Suunnittelemattomat tuotantokatkokset ovat ongelmallisia yrityksille ja niitä koitetaan välttää viimeiseen asti. Tuotantokatkosten välttämiseksi yrityksissä onkin siirrytty ongelmia korjaavasta kunnossapidosta kohti ehkäisevää tai peräti ennakoivaa kunnossapitoa. Erilaisten data-analytiikka menetelmien hyödyntämistä ennakoivassa kunnossapidossa on tutkittu laajasti ja menetelmien kehittyminen on tehnyt niistä helpommin hyödynnettäviä.

Tässä tutkielmassa tutustutaan data-analytiikan ja tekoälymetelmien hyödyntämiseen ennakoivassa kunnossapidossa, sekä toteutetaan valikoiduille Finnsementin Paraisten sementtitehtaan koneille tärinämallit ydinestimointia hyödyntäen. Tutkielmassa käytettiin pääasiallisina lähteinä aihealueeseen liittyvää kirjallisuutta sekä joitakin internetistä löytyviä artikkeleita ja raportteja.

Asiasanat: Data-analytiikka, Ennakoiva kunnossapito, Koneoppiminen, Ydinestimointi

UNIVERSITY OF TURKU  
Department of Computing

JANNE ILTANEN: Data-analytiikan hyödyntäminen ennakoivassa kunnossapidossa

Master of Science Thesis, 55 p.

Data analytics

April 2022

---

Unplanned interruptions in production are problematic for companies. In order to avoid production interruptions, companies have moved from corrective maintenance to preventive or even predictive maintenance. The utilization of various data analytic methods in predictive maintenance has been extensively studied and the development of the methods has made them easier to utilize.

In this thesis, the utilization of data analytics and artificial intelligence methods in preventive maintenance are introduced. There is also implementation of vibration models for selected machines at Finnsementti's Parainen cement plant using kernel density estimation. The main sources used in the thesis were literature related to the topic, as well as some articles and reports available on the Internet.

Keywords: Data analytics, Predictive maintenance, Machine learning, Kernel density estimation

# Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Kunnonvalvonnan teoria</b>	<b>3</b>
2.1	Kunnossapito . . . . .	3
2.2	Kunnonvalvonnan menetelmät . . . . .	4
2.2.1	Tärinäanalyysi . . . . .	4
2.2.2	Voiteluaineanalyysi . . . . .	5
2.2.3	Suorituskykyanalyysi . . . . .	6
2.2.4	Lämpökuvaus . . . . .	6
2.2.5	Akustiset mittaukset . . . . .	6
2.3	Yleisimmät vikatilanteet . . . . .	7
2.3.1	Laakereiden viat . . . . .	7
2.3.2	Staattori- ja roottoriviat . . . . .	7
2.3.3	Roottoritangon rikkoutuminen ja päätyrenkaan viat . . . . .	8
2.3.4	Epäkeskisyyteen liittyvät viat . . . . .	9
2.4	Tekoälymenetelmät kunnonvalvonnassa . . . . .	9
2.4.1	Geneettiset algoritmit . . . . .	11
2.4.2	Tukivektorikoneet . . . . .	11
2.4.3	Sumean logiikan järjestelmät . . . . .	12
2.4.4	Neuroverkot . . . . .	14

2.4.5	Satunnaismetsät . . . . .	17
2.4.6	K:n keskiarvon ryhmittely . . . . .	19
2.4.7	Pääkomponenttianalyysi . . . . .	21
2.4.8	Fourier-analyysi . . . . .	21
<b>3</b>	<b>Sementtitehdas toimintaympäristönä</b>	<b>23</b>
3.1	Järjestelmän kuvaus . . . . .	23
3.2	Datan esittely . . . . .	24
3.3	Käytetyt laitteet ja ohjelmistot . . . . .	25
3.3.1	Fyysiset laitteet . . . . .	26
3.3.2	Pilvikomponentit . . . . .	26
3.4	Sementin valmistus . . . . .	27
3.5	Kunnossapidon nykytila . . . . .	28
3.6	Tehtaalle asennetut sensorit . . . . .	29
<b>4</b>	<b>Koneiden normaalitilan mallintaminen</b>	<b>32</b>
4.1	Cross-industry standard process for datamining . . . . .	32
4.1.1	Liiketoiminnan ymmärtäminen . . . . .	32
4.1.2	Datan ymmärtäminen . . . . .	34
4.1.3	Datan valmistelu . . . . .	37
4.1.4	Mallintaminen . . . . .	37
4.1.5	Tulosten arviointi . . . . .	38
4.1.6	Käyttöönotto . . . . .	38
4.2	Mallintaminen . . . . .	38
4.2.1	Parametriset todennäköisyysjakaumat . . . . .	39
4.2.2	Ydinestimointi . . . . .	44
4.2.3	Kahden näytteen Kolmogorov-Smirnovin-testi . . . . .	51
<b>5</b>	<b>Yhteenveto</b>	<b>53</b>



# Kuvat

2.1	Eri tärinälähteiden yhteisvaikutus vaikuttaa lopulta useisiin erilaisiin tärinä vasteisiin. [5]	5
2.2	Geneettinen algoritmi [8].	12
2.3	Sumean logiikan toimintaperiaate [12].	13
2.4	MLP-neuroverkko. Syöttökerros (engl. input layer) syöttää analysoitavan datan piilotetulle kerrokselle (engl. hidden layer), joka muodostaa taas tuloskerroksen (engl. output layer). [14]	15
2.5	Graaffinen esitys autoenkoodausta hyödyntäneen tutkimuksen neuroverkkojen rakenteesta [18].	17
2.6	Yksinkertaistettu esimerkki satunnaismetsien toiminnasta. [19]	18
2.7	Esimerkki moniulotteisesta keskiarvoryhmittelystä, mustat tähdet kuvaavat ehjiä laakereita, punaisella ovat tapaukset joissa laakerin sisäkehä on vioittunut ja vihreällä ulkokehästä vioittuneet. Sinisellä merkatuissa taas kehien välissä olevissa kuulissa on ollut vikaa. XYZ-akselit kuvaavat taas tutkimuksessa mitattuja ominaisuuksia. [21]	20
2.8	Pääkomponenttianalyysin toiminta [22].	22
3.1	Nykyisen järjestelmän kuvaus [28].	24
3.2	RuuviTag avattuna.	25
3.3	Sementin valmistus. [33]	27
4.1	Yhteenvedo CRISP-DM:n vaiheista ja tuloksista. [37]	33

4.2	Laatikkokuvaaja raakamyly 7:n hammaskehän voitelupumpun väri- nästä. . . . .	35
4.3	Histogrammi raakamyly seitsemän ketjuelevaattori ykkösen tärinästä.	35
4.4	Histogrammi sementtiuunin ruuvikuljettimen sähkösuodattimen tärinästä. . . . .	36
4.5	Fitter-kirjaston tulokset raakamyly 7:n erotuspiirin separaattori 1:n tärinälle. . . . .	40
4.6	Fitter-kirjaston tulokset sementtiuunin vaakaruuvikuljetin kakkosen tärinälle. . . . .	41
4.7	Ydinestimoinnissa käytettyjä ydinfunktioita [53]. . . . .	45
4.8	Ydinestimointi uunin sähkösuodattimen poikittaisesta ruuvikuljettimesta, sekä analyysit estimoinnin onnistumisesta. . . . .	47
4.9	Ydinestimointi sementtiuunin syötön homogenisoinnin 3. vaakaruuvikuljettimesta, sekä analyysit estimoinnin onnistumisesta. . . . .	48



# Taulukot

2.1	Tekoälymenetelmien vertailu . . . . .	10
3.1	Näyte kerätystä datasta . . . . .	25
3.2	Myllyihin asennetut sensorit . . . . .	30
3.3	Sementtiuuniin asennetut sensorit . . . . .	31
4.1	Ydinestimoinnit myllyjen antureista . . . . .	49
4.2	Ydinestimoinnit uunin antureista . . . . .	50

# 1 Johdanto

Tämän työn tarkoituksena on selvittää, miten Finnsementin Paraisten sementitehtaalle vuonna 2018 rakennettua valikoitujen laitteiden värinänseurantajärjestelmän keräämää dataa voitaisiin hyödyntää ennakoivassa kunnossapidossa. Ennakoidulla kunnossapidolla tarkoitetaan toimia, joissa koneiden kuntoa valvotaan erilaisia menetelmin ja koneille tehtävät huoltotoimenpiteet toteutetaan kunnonvalvonnasta saatujen tietojen perusteella.

Työ koostuu viidestä luvusta. Luku 2 käsittelee kunnonvalvonnan menetelmiä, yleisimpiä vikatilanteita ja kunnonvalvonnassa hyödynnettyjä tekoälymenetelmiä. Luku 3 esittelee värinädatan keräämisen käytetyn laitteiston ja siihen liittyvät sovellukset. Kappaleessa 3.4 käydään läpi sementin valmistus pääpiirteittäin ja kappale 3.6 listaa tehtaalle asennetut 44 sensoria niiden asennuspaikkoineen.

Luku 4 on kokeellinen osuus, jossa alaluvussa 4.1 käydään läpi projektin eri vaiheet Cross-industry standard process for datamining -mallin mukaisesti. Prosessin eri vaiheissa perehdytään liiketoiminnan tarpeisiin, datan rakenteeseen ja jalostamiseen analytiikkaa varten. Lisäksi luvussa tehdään katsaus projektin tulosten arviointiin ja itse lopulliseen käyttöönottoon. Varsinainen datan mallintaminen tapahtuu alaluvussa 4.2. Lopullinen mallintaminen toteutettiin ydinestimoinnilla. Lopulta luvussa 5 on yhteenveto tehdystä työstä ja sen tuloksista.

Finnsementti Oy on yli 100 vuotta toiminut suomalainen sementinvalmistaja. Finnsementin historian voidaan katsoa alkaneen vuonna 1914, kun Paraisten Kalk-

---

kivuori Osakeyhtiö aloitti sementin valmistuksen. Viisi vuotta myöhemmin myös Lohjan Kalkkitehdas aloitti sementin valmistuksen ja vuonna 1992 yritysten sementtituotannot yhdistyivät. Yrityksellä on sementtitehtaat Paraisilla ja Lappeenrannassa sekä kuonajauhetehtas Raahessa. Vuodesta 1999 yritys on ollut osa CRH:ta, joka on yksi maailman suurimmista rakennusteollisuudenkonserneista [1]. Finnsementin liikevaihto vuonna 2020 oli 132 miljoonaa euroa ja se työllistää noin 230 työntekijää [2].

## 2 Kunnonvalvonnan teoria

Alusta alkaen erilaisiin sähkökoneisiin on laitettu erilaisia suojauksia esimerkiksi ylivirran, ylijännitteen ja maasulun varalta. Laitteiden tullessa monimutkaisemmiksi on ratkaisuja etsitty myös vikadiagnostiikan ja kunnonvalvonnan puolelta, joten ongelmatilanteiden havainnoinnista jo alkuvaiheessa onkin tullut erittäin tärkeää, jotta välttyttäisiin suunnittelemattomilta tuotantokatkoksilta. [3]

### 2.1 Kunnossapito

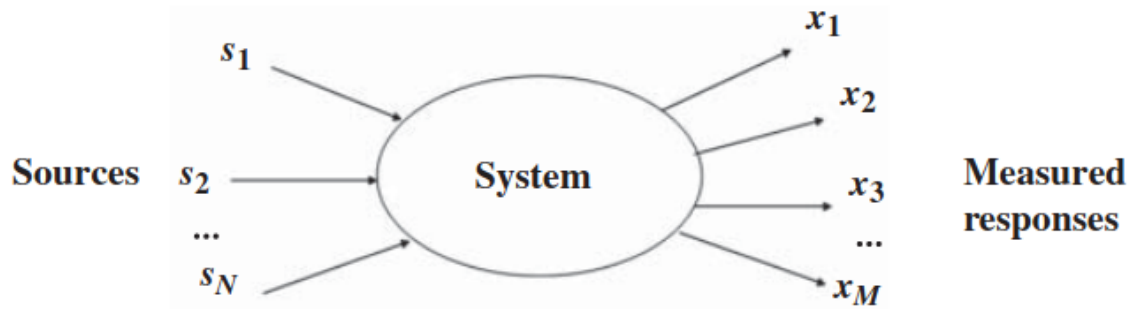
Kunnossapito voidaan jakaa kolmeen eri kategoriaan: korjaavaan, ennakoivaan ja ehkäisevään kunnossapitoon. Korjaavassa kunnossapidossa (engl. corrective maintenance, run-to-failure) kunnostustoimia tapahtuu vasta, kun koneet lakkaavat toimimasta ja tuotanto pysähtyy. Ehkäisevässä kunnossapidossa (engl. preventive maintenance) koneita huolletaan ennakkoon määritellyn aikataulun mukaisesti, mikä suhteellisen hyvin auttaa välttämään turhia pysähdyksiä tuotannossa, mutta joskus joudutaan kuitenkin tekemään aikataulun ulkopuolisia korjaavia toimenpiteitä. Ennakoivalla kunnossapidolla (engl. predictive maintenance) tarkoitetaan menetelmiä, jossa koneiden rikkoutuminen pyritään ennakoimaan ja korjaavat toimenpiteet tehdään tämän ennakkoinnin perusteella. Varhainen havaitseminen voidaan toteuttaa tavalla, joihin tutustutaan tarkemmin kappaleessa 2.2. [4]

## 2.2 Kunnonvalvonnan menetelmät

Oleellinen osa ennakoivaa kunnossapitoa on kunnonvalvonta. Koneiden kunnonvalvonnalla (engl. condition monitoring) tarkoitetaan mallia, jossa koneiden sen hetkistä tilaa seurataan ja pyritään sen avulla ennustamaan tilan kehitystä. Käytännössä tämä vaatii tiedonhankintaa koneen sisällä tapahtuvasta toiminnasta ulkoisten lähteiden avulla koneiden käytön aikana. Kaksi yleisintä tapaa tiedon hankintaan ovat värinän seuranta ja voiteluaineiden analysointi. [5]

### 2.2.1 Värinäanalyysi

Kaikilla täysin kunnossa olevilla koneilla on omanlaisensa värinämalli, joka muuttuu, kun laitteissa alkaa kehittyä vikoja. Muutoksia värinässä on mahdollista havaita jo ennen laitteiden lopullista hajoamista. Kuva 2.1 esittää, miten koneesta mitattu värinä on aina yhdistelmä useita eri värinän lähteitä (kuvassa  $s_1-s_N$ ). Koneen rakenteen (kuvassa system) läpi kulkiessa värinöiden yhteisvaikutus muodostaa erilaisia mitattavia värinävasteita (kuvassa  $x_1-x_M$ ). Haasteena analyysia tehdessä onkin selvittää, johtuuko ongelma jostain koneen osasta vai värinää kuljettavasta rakenteesta. Useimmiten viat kuitenkin löytyvät jostain koneen komponentista. Yleistykseenä voidaan sanoa, että normaalitilassa vakionopeudella käyvä kone värisee aina samalla perustaaajuudella. Se onkin yleisin tapa seurata koneiden kuntoa, sillä sen avulla voidaan havaita tapahtuvia muutoksia välittömästi, jos seuranta on jatkuvaa, mutta tämän lisäksi menetelmä soveltuu myös satunnaisempaan käyttöön. Voiteluaineanalyysiin verrattuna värinäanalyysi on myös tarkempi, sillä se osoittaa viallisen komponentin todennäköisemmin, esimerkiksi voiteluaineanalyysin tuloksia heikentää se, että useissa laakereissa on samanlainen metallien koostumus. Lisäksi värinäanalyysiin voidaan yhdistää kulumisanalyysia. Signaalinkäsittelytekniikoiden kehittyminen on johtanut siihen, että jo pienempiä muutoksia voidaan havaita ja erottaa kohinasta. [5]



Kuva 2.1: Eri tärinälähteiden yhteisvaikutus vaikuttaa lopulta useisiin erilaisiin tärinävästeisiin. [5]

### 2.2.2 Voiteluaineanalyysi

Voiteluaineanalyysi voidaan jakaa kolmeen kategoriaan: siruilmaisimiin, spektografisiin öljyanalyysimenetelmiin sekä kulumishiukkasanalyysiin. Siruilmaisimilla tarkoitetaan suodattimia ja magneettitulppia, jotka on suunniteltu keräämään lohjonneet sirut sekä muut epäpuhtaudet voiteluainejärjestelmän kierrosta. Vaihtoehtoisesti näytteitä voidaan tutkia myös voitelujärjestelmässä olevan ikkunan kautta. Havaituista epäpuhtauksista tarkastellaan esimerkiksi niiden määrää, laatua, muotoa sekä kokoa. Voiteluaineanalyysin heikkoutena voidaan pitää viivettä näytteenoton ja tulosten saamisen välillä. [5]

Spektografisessa öljyanalyysissä voiteluaineesta otetaan säännöllisesti näytteitä, joille tehdään kemiallisia-analyysejä. Analyyseissä voiteluaineen seasta havaitut tietyn tyyppiset materiaalit voivat kertoa tarkemmin, mitkä osat ovat kuluneet tai näytteistä voidaan havaita muita epäpuhtauksia, kuten esimerkiksi jäähtyksen puolelta vuotanutta vettä. [5]

Kulumishiukkasanalyysissä eli ferrografiassa tarkkaillaan voiteluaineessa olevia kulumishiukkasia. Menetelmä perustuu voiteluaineen seassa olevien kulumishiukkasten määrässä ja kokojakaumassa tapahtuviin muutoksiin, jotka korreloivat koneen kunnan kanssa: kun hiukkasten koko ja esiintyvyys kasvaa, viittaa se koneen voimak-

kaaseen kulumiseen. Yksittäisiä hiukkasia tarkastelemalla taas saadaan tietoa, mistä kuluneesta komponentista hiukkaset ovat lähtöisin ja onko niiden syntymekanismi ollut esimerkiksi lastuaminen tai hankautuminen. [6]

### 2.2.3 Suorituskykyanalyysi

Tietyille laitteille suorituskykyanalyysi on tehokas tapa seurata sen toimintakuntoa. Esimerkiksi imupaineen muutokset voivat kertoa suodattimien tukkeutumisesta tai jos kompressorin tilavuushyötysuhde laskee, voi venttiileissä olla vuotoja.. Myös virtauspolkuanalyysia käytetään, jos laitteessa on erilaisia antureita, joiden tuottamasta datasta voidaan laskea erilaisia hyötysuhteita ja verrata normaalitilaan. [5]

### 2.2.4 Lämpökuvaus

Myös lämpökuvausta voidaan hyödyntää kunnonvalvonnassa. Nykyaikaiset mittalaitteet voivat huomata jopa pieniä lämpötilojen muutoksia. Esimerkiksi laakerien tapauksessa lämpötilat alkavat nousta niiden elinkaaren loppuvaiheessa, kun vierintäelementit alkavat liukumaan. Laakerien häiriötiloista saadaankin tieto nopeammin värinä- ja akustisten mittausten avulla. Lämpökuvausta hyödynnetään myös sähkökytkentöjen kuumien pisteiden tai tulenkestävien rakenteiden vuorausten vuotojen havaitsemiseen. [5]

### 2.2.5 Akustiset mittaukset

Akustisella päästöllä (engl. acoustic emission) tarkoitetaan ääntä, joka syntyy kahden väliaineen (esimerkiksi laakerin vierintäelementin ja liukupinnan) vuorovaikutuksesta suhteellisessa liikkeessä. Tämä ääni muuttuu osien kuluessa tai niiden viikaantumisen yhteydessä. Alun perin akustisen mittauksen menetelmät kehitettiin esimerkiksi kiinteiden rakenteiden testaukseen, ilman että niille aiheutuu vaurioita. Myöhemmin menetelmän käyttöä on laajennettu myös pyöriviin koneisiin, kuten

laakereihin, pumppuihin ja vaihdelaatikoihin. Menetelmän etu värinäanalyysiin verrattuna on sen suurempi herkkyys, jolloin häiriöitä voidaan havaita aikaisemmin, mutta haasteena akustisten mittausten hyödyntämisessä on kuitenkin datan vaikea käsittely, tulkinta ja luokittelu. [7]

## 2.3 Yleisimmät vikatilanteet

Sähkömoottorien vikaantumisten pääasialliset syyt voidaan jakaa muutamaankin eri kategoriaan, jotka kattavat valtaosan kaikista vikaantumisista. Alla on käyty läpi nämä yleisimmät vikatyypit ja niihin johtavat syyt.

### 2.3.1 Laakereiden viat

Pääosa erilaisista sähkömoottoreista käyttää erilaisia kuula- tai liukulaakereita ja lähes 40–50 % kaikista sähkömoottorien vioista liittyy laakereihin. Laakerilla tarkoitetaan kahden elementin välistä osaa, joka ohjaa niiden liikettä suhteessa toisiinsa. Laakerit ovat alttiita väsymisvaurioille, vaikka toimintaolosuhteet olisivat hyvät ja laakeriin kohdistuva kuorma olisi tasapainossa. Muita tyypillisiä syitä laakereiden vikoihin ovat kontaminaatiot, vääränlainen voitelu tai laakerien virheellinen asentaminen. Laakereiden häiriöt saattavat kasvattaa laitteen värinää sekä aiheuttaa meteliä. [3]

### 2.3.2 Staattori- ja roottoriviat

Toinen yleinen ongelmakohta sähkömoottoreissa liittyy staattorien ja roottorien vikoihin, jotka aiheuttavat noin 30–40 % kaikista sähkömoottorien vioista. Roottorista saatetaan käyttää myös nimityksiä pyöräjä ja ankkuri. Tämän kategorian ongelmat johtuvat usein eristyksen ongelmista eristyksessä, jonka syitä useimmiten ovat:

- Staattorin sydämen tai käämin korkealämpötila



- Ytimen löysä laminointi
- Päätykäämien löysä jäykistys
- Öljystä, kosteudesta tai liasta johtuvat kontaminaatiot
- Oikosulku
- Käynnistysjännitykset
- Sähköpurkaukset
- Vuodot jäähdytysjärjestelmässä

[3]

### 2.3.3 Roottoritangon rikkoutuminen ja päätyrenkaan viat

Kahta edellä mainittua selvästi harvinaisempi ongelmatyyppi, noin 5–10 prosentissa tapauksista, on roottoritangon tai päätyrenkaiden vaurioitumisesta johtuva vika. Roottorit voivat olla joko valettuja kokonaisuuksia tai valmistettu erillisistä osista. Valetut roottorit ovat lujatekoisempia kuin osista kootut, mutta rikkoutumisen jälkeen valettu roottori pitää vaihtaa kokonaan uuteen, sillä niiden korjaaminen on käytännössä mahdotonta. [3]

Roottoritangon ja päätyrenkaiden rikkoutumiselle on useita mahdollisia syitä:

- Liasta lämmöstä ja epätasapainosta aiheutuva rasitus
  - Pääasiassa valmistetut roottorit
- Sähkömagneettinen kuormitus
  - Epätasainen magneettinen veto
  - Sähkömagneettinen kohina ja värinä

- Valmistusvirheestä johtuvat jäännösjäännitykset
- Dynaamiset jännitykset, joita aiheuttavat akselin vääntömomentti, keskipoisvoimat ja sykliset jännitykset
- Ympäristöstä johtuva rasitus, kuten kontaminaatiot tai roottorin kuluminen
- Mekaaninen rasitus
  - Löysä laminointi
  - Osien väsyminen
  - Laakeriongelmat

[3]

#### 2.3.4 Epäkeskisyyteen liittyvät viat

Epäkeskisyydellä tarkoitetaan tilaa, jossa ilmarako roottorin ja staattorin välissä on epätasainen. Epäkeskisyyden kasvaminen johtaa epätasaiseen magneettiseen vetoon, joka saattaa aiheuttaa staattorin ja roottorin välistä hankausta. [3]

## 2.4 Tekoälymenetelmät kunnonvalvonnassa

Kunnonvalvonnan ennakoinnissa voidaan hyödyntää useita erilaisia tekoälymenetelmiä, kuten geneettisiä algoritmeja, tukivektorikoneita (engl. support vector machine, SVM), sumean logiikan järjestelmiä (engl. fuzzy logic system, FLS), satunnaismetsiä (engl. random forest) ja keinotekoisia neuroverkkoja (engl. artificial neural network, ANN). [8][4]

Taulukossa 2.1 on lyhyesti vertailtu eri tekoälymenetelmiä.

Taulukko 2.1: Tekoälymenetelmien vertailu

Menetelmä	Vahvuudet	Heikkoudet
Satunnaismetsät	Yksinkertainen rakenne Ei tarvetta datan esikäsittelylle	Iso mahdollisuus ylikouluttamiselle Pitkä koulutusaika
Tukivektorkoneet	Toiminta korkealla ulottovuudella, Toimii epälineaarisisa prosesseissa, Tietojen määrittelyä ei tarvita	Ei kykene turhan tiedon suodattamiseen, Monimutkainen hallinta, Riski päällekkäisyyteen
Pääkomponenttianalyysi	Ei päällekkäisyyttä Pieni koulutusaika Datan visualisointi	Mahdollisuus piirteiden häviämiseen, joka heikentää tarkkuutta
Geneettiset algoritmit	Mukautuva algoritmi, Nopea prosessointi	Riski päällekkäisyyteen
Neuroverkot	Nopea oppiminen, Ei altis kohinalle	Ylikouluttamisen riski, Mustan laatikon menetelmä, Yli- ja alisovittamisen vaikea tasapainotus
Sumean logiikan järjestelmät	Yksinkertainen rakenne Mukautuva algoritmi Helppo toistettavuus	Vaatii suuren datamäärän Epätarkka data johtaa heikkoihin tuloksiin
K:n keskiarvon ryhmittely	Helppo implementointi Soveltuu eri muotoisten tietojoukkojen luokitteluun	Riippuvuus lähtöarvoista Haastava skaalaus, jos käytössä useita ulottuvuuksia
Fourier-analyysi	Vain vähän kadotettua informaatiota Helppo vertailu tutkimusten välillä	Kompromissi aika/taajuus-resoluution välillä Aikajakson pituus kriittinen lopputuloksen kannalta

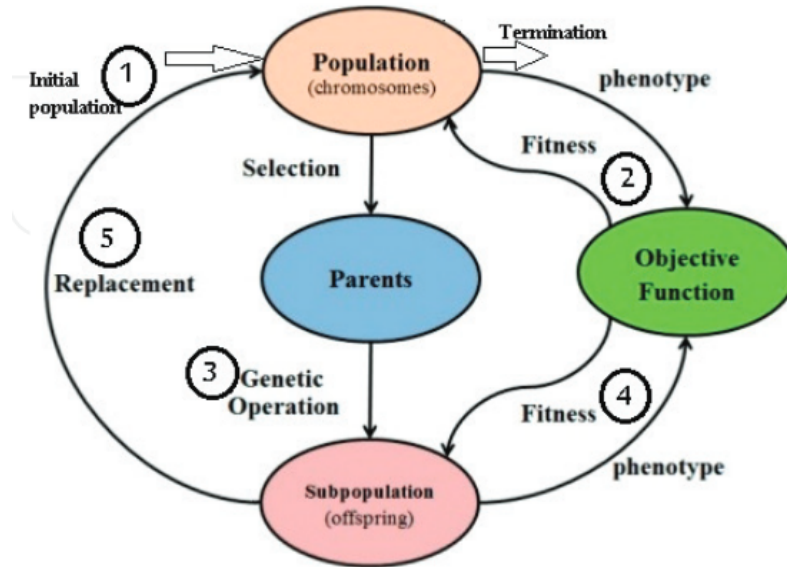
### 2.4.1 Geneettiset algoritmit

Geneettiset algoritmit ovat John Hollandin 1970-luvulla kehittämä tekoälymenetelmä, joka perustuu luonnolliseen biologista evoluutiota jäljittelevään valintaprosessiin. Algoritmi muokkaa jatkuvasti yksittäisten ratkaisujen populaatiota ja valitsee jokaisella kierroksella satunnaisesti yksilöitä nykyisestä populaatiosta ja käyttää niitä seuraavan sukupolven luomiseen; jokaisella kierroksella malli kehittyy kohti optimaalista ratkaisua. Kuvassa 2.2 on esitetty geneettisen algoritmin kierto. Geneettisissä algoritmissa populaation yksittäistä näytettä kutsutaan kromosomiksi, joista jokainen saattaa olla vastaus ongelman ratkaisemiseksi. Objektiviiset funktiot arvioivat yksittäisten kromosomien soveltuvuutta. Parhaiten soveltuvista luodaan jälkeläisiä geneettisten risteytys- ja mutaatio-operaatioiden perusteella. Syntyneitä jälkeläisiä voidaan hyödyntää päätöksenteossa tai ne voidaan käyttää algoritmin seuraavalla kierroksella. [8]

Aikaisemmissa tutkimuksissa geneettisiä algoritmeja on hyödynnetty esimerkiksi tutkittaessa kuulalaakereiden kuntoa, joissa tietoa laakereiden toiminnasta kerättiin kolmella kiihtyvyyssensorilla ja yhdellä mikrofoniolla. Testit tehtiin kahdeksalle vialliselle ja yhdelle ehjälle laakerille, testeissä kierrosluku oli tasainen 1400 rpm ja vääntö 14730 newtonia. Tutkimuksessa geneettiset algoritmit löysivät datasta merkitsevät ominaisuudet, jotka sitten syötettiin neuroverkolle. Tällä menetelmällä saatiin varsin hyviä tuloksia ja vikaantuneiden laakereiden luokittelu onnistui huomattavan hyvällä tarkkuudella. [9]

### 2.4.2 Tukivektorikoneet

Tukivektorikoneet ovat laajasti käytetty kokoelma ohjattuja oppimismenetelmiä datan luokittelua varten tehtäviin. Eräs tukivektorikoneiden tunnuspiirre on niiden hyvä tarkkuus löytää luokkien raja-arvot ja erottaa ne toisistaan. Alun perin tukivektorikoneet kehitettiin binääriseen luokitteluun, mutta myöhemmin niitä on laa-



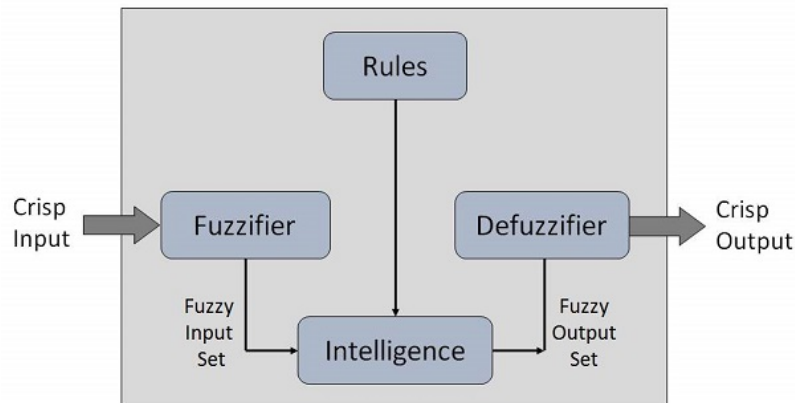
Kuva 2.2: Geneettinen algoritmi [8].

jennettu luokittelemaan myös moniluokkaisia muuttujia. [4] [8]

Tukivektorikoneita on hyödynnetty esimerkiksi tutkimuksessa, jossa diagnosoi-  
ttiin vikoja alhaisella nopeudella toimivissa laakereissa. Tutkimuksissa laakereista  
kerättiin tietoa sekä niiden tärinästä että akustisista päästöistä. Tutkimus toteu-  
tettiin laboratorio-olosuhteissa SKF NF307-laakereille, joita oli rikottu tarkoituksel-  
la ja viat jaettiin kuuteen eri luokkaan. Tutkimuksessa tukivektorikoneen käyttöä  
verrattiin relevanssivektorikoneella (engl. relevance vector machine, RVM) tehtyyn  
analytiikkaan, jossa RVM todettiin kuitenkin tarkemmaksi menetelmäksi. [10]

### 2.4.3 Sumean logiikan järjestelmät

Sumean logiikan järjestelmäksi kutsutaan matemaatikko Lotfi Zadehin vuonna 1965  
kehittämää moniarvoista logiikka, joka mahdollistaa tavallisesti binääristen arvojen  
esittämisen desimaalimuotoisina väliarvoina[8]. Sumea logiikka voi hyödyntää jo-  
ko valvottua tai valvomatonta oppimista tai niiden yhdistelmää. Hybridimallissa  
prosessi on viisivaiheinen. Ensimmäisessä vaiheessa koulutusdatan avulla löydetään



Kuva 2.3: Sumean logiikan toimintaperiaate [12].

jäsenfunktioiden keskuksset ja leveydet itseorganisoitujen klustereiden avulla. Seuraavaksi etsitään kilpailullisen oppimisen avulla sumean logiikan säännöt ja näille säännöille tehdään lopulta eliminointi ja niiden yhdistely. Lopuksi etsitään optimaaliset jäsenyysfunktioit virheiden vastavirta-algoritmin avulla. [11]

Sumean logiikan toimintaperiaate on esitetty kuvassa 2.3. Järjestelmässä selkeä syöte sumennetaan ja tämän sumennetun tiedon käsittelyyn sovelletaan erilaisia sääntöjä. Syötetyn tiedon käsittelyn jälkeen se syötetään seuraavaan vaiheeseen, jossa tiedosta tehdään taas selkeässä muodossa oleva syöte. [12]

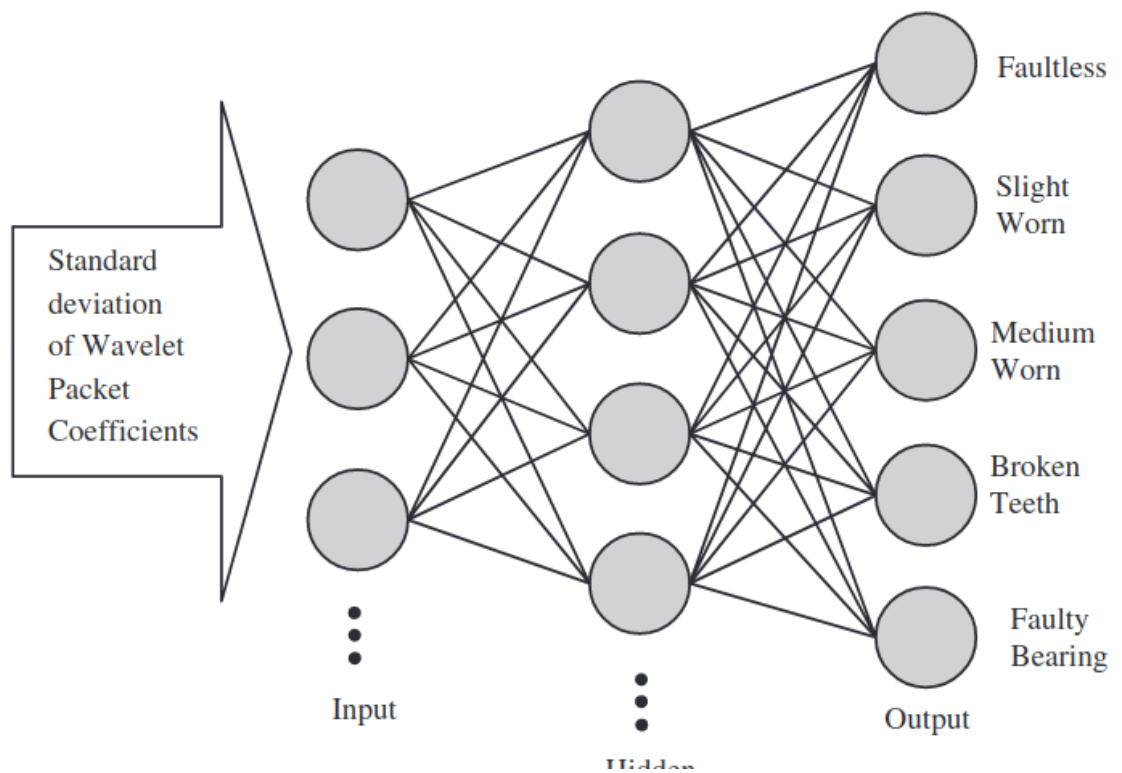
Sumean logiikan järjestelmiä on käytetty analysoimaan yhteyttä akustisten päästöjen ja työkalujen kulumisen välillä. Tutkimuksessa mitattiin titaanimetallikomposiitin leikkaamisesta syntyvää ääntä. Kerättyä dataa käsiteltiin sumealla suodatuksella (engl. fuzzy filtering). Saatujen tulosten mukaan sumean logiikan hyödyntäminen paransi tuloksia perinteiseen neliöllisen keskiarvon avulla tehtävään akustiseen analyysiin verrattuna ja metodin käyttö myös mahdollisti seurannan automatisoinnin. [13]

#### 2.4.4 Neuroverkot

Neuroverkot ovat tekoälymenetelmä, jossa tiedon prosessointi noudattaa biologisen hermojärjestelmän, käytännössä aivojen, tapaa prosessoida tietoa [8]. Menetelmässä prosessointiyksiköt, joita kutsutaan solmuiksi tai neuroneiksi, on yhdistetty toisiinsa viestintäkanavilla ja solmut käyttävät ainoastaan aikaisemmalta solmultaan saatua tietoa päätöksenteossa. Neuroverkkojen etuna on päätöksenteon perustuminen historialliseen dataan, jolloin ratkaisun tekemiseen ei tarvita tarkempaa asiantuntijatietoa. Lisäksi neuroverkkoa voidaan hyödyntää reaaliaikaisissa sovelluksissa ilman, että sovellukseen tarvitsee tehdä muutoksia jokaisen päivityksen yhteydessä. Neuroverkoilla on myös huonoja puolia, kuten se, että verkosto saattaa tehdä johtopäätöksiä, jotka sotivat sovelluksen vahvistamia sääntöjä ja teorioita vastaan. Myös verkon kehittäminen on aikaa vievää ja lopulta neuroverkojen varsinaista päätöksentekoprosessia ei tiedetä, vaan kyseessä on niin kutsuttu mustan laatikon menetelmä. Lisäksi neuroverkon kouluttamiseen tarvitaan erittäin suuri määrä dataa. [4]

#### Syväoppiminen

Neuroverkkojen yhteydessä aletaan puhumaan syväoppimisesta, kun piilotettuja kerroksia on enemmän kuin yksi. Syväoppimisen idea on se, että alemman tason piilotetut kerrokset laskevat tuloksia, joita myöhemmän vaiheen piilotetut kerrokset käyttävät syötteinään. Syvemmät verkot mahdollistavat, että solmut käyttävät pienempiä painotuksia ja yksiköitä päätöksentekoon. Tämä mahdollistaa suuremman tarkkuuden, kun yhden solmun päätelmille ei anneta niin suurta painoarvoa. [15]



Kuva 2.4: MLP-neuroverkko. Syöttökerros (engl. input layer) syöttää analysoitavan datan piilotetulle kerrokselle (engl. hidden layer), joka muodostaa taas tuloskerroksen (engl. output layer). [14]



### Monikerroksinen perseptroniverkko

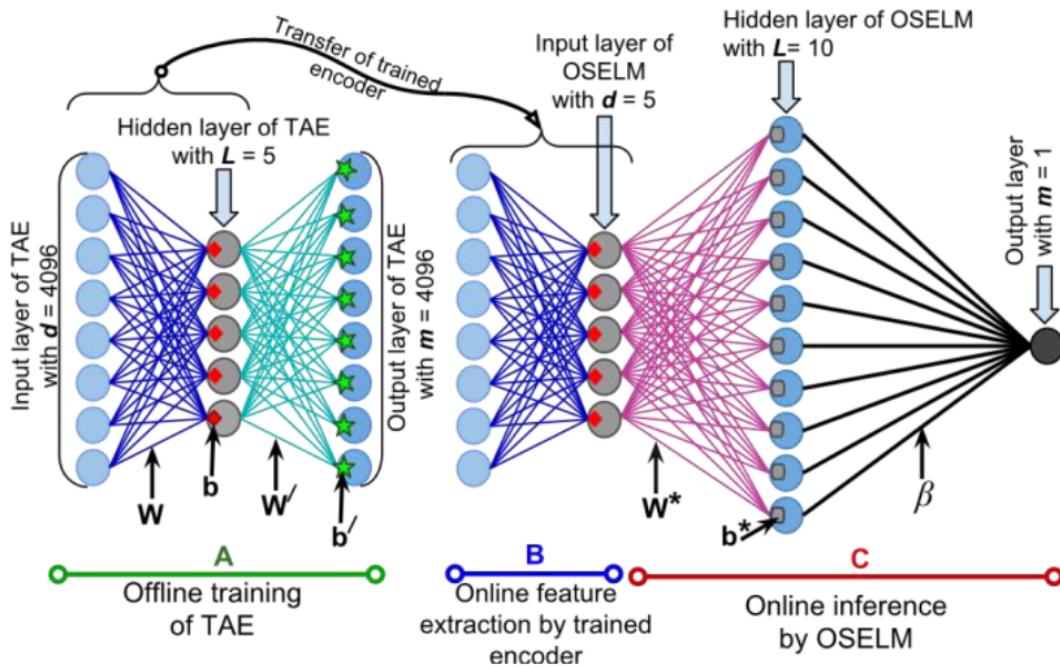
Kuva 2.4 esittää yhden piilotetun kerroksen monikerroksista perseptroniverkkoa (engl. multilayer perceptron). Piilotettuja kerroksia voi olla enemmänkin kuin yksi. Jokainen solmu voi hoitaa sekä summausta että aktivointia yhtä aikaa. [16]

MLP-neuroverkkojen käyttöä vaihdelaatikoiden laakereiden ja hammaspyörien kunnan seurannassa on tutkittu laboratorio-olosuhteissa. Tutkimus toteutettiin niin, että sähkömoottorit pyörittivät 1420 kierroksen minuuttinopeudella neljävaihteista moottoripyörän vaihdelaatikkoa. Järjestelmästä kerättiin tietoa kierroslukumittarilla sekä kiihtyvyyssanturilla, jolla seurattiin tärinää. Signaalin esikäsittelyssä signaalit synkronointiin aallokepaketeille (engl. wavelet packet) laskettiin keskihajonta. Aallockeanalyysia käytetään kohinan vähentämiseen ja signaalien hajottamiseen ominaisuuskomponentteihin. Syötteenä MPL-neuroverkolle käytettiin aallokepakettien kerrointen keskihajontaa. Tutkimuksessa onnistuttiin tunnistamaan hammaspyörien ja laakereiden häiriöt täydellisellä sadan prosentin tarkkuudella. Kuvassa 2.4 on tutkimuksessa käytetty neuroverkon rakenne, syöttökerroksessa solmuja oli 16, piilotetussa kerroksessa 20 ja tuloskerroksessa viisi, yksi jokaisella tilaluokitukselle. [14]

### Autoenkoodaja

Autoenkoodaja (engl. autoencoder) on neuroverkkotyyppi, joka käyttää takaisinsäyksen (engl. backpropagation) algoritmia annotoimattoman datan ominaisuuksien oppimiseen. Autoenkoodajaan toiminnassa on kaksi vaihetta: ensimmäisessä syötettävä data muutetaan pieniulotteiseen esitysavaruuuteen, mistä kerätään mahdollisimman paljon koulutukseen käytettäviä ominaisuuksia. Toisessa vaiheessa kerätyt ominaisuudet taas syötetään takaisin syöttöavaruuuteen. [17]

Aikaisemmissa tutkimuksissa autoenkoodausta on käytetty Cincinnatiin yliopiston aikasemmin luoman datan analysointiin. Data koostui kolmesta tietojoukosta, jossa jokaisessa oli tärinäsignaali neljältä eri laakerilta. Dataa esikäsiteltiin poista-

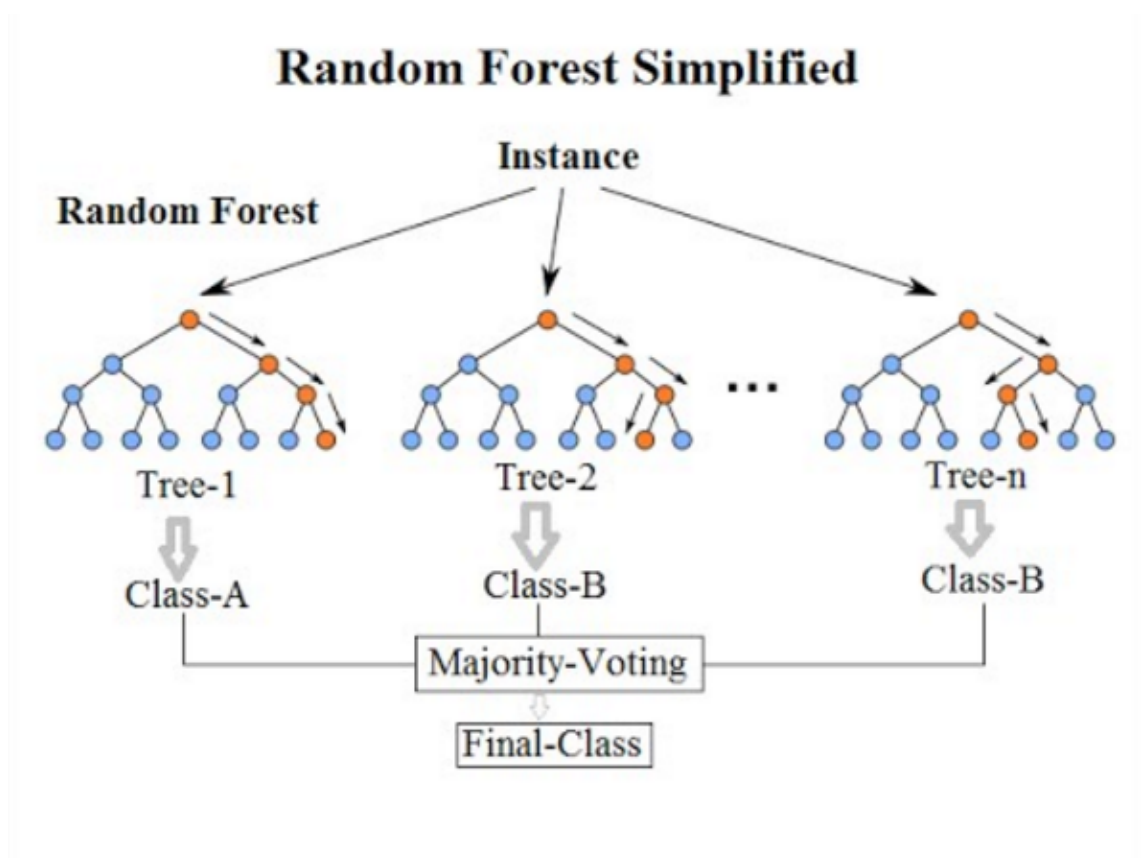


Kuva 2.5: Graaffinen esitys autoenkoodausta hyödyntäneen tutkimuksen neuroverkkojen rakenteesta [18].

malla niistä kohinaa. Kuvassa 2.5 on esitetty tutkimuksessa käytetty rakenne. Aluksi autoenkoodausta hyödyntävä neuroverkko koulutettiin erikseen ja siitä saatua mallia hyödynnettiin seuraavan vaiheen syötteenä neuroverkolle. Tutkimuksessa tultiin lopputulokseen, että autoenkooderin käyttö soveltuu laakereiden reaaliaikaiseen kunnonvalvontaan hyvin. [18]

### 2.4.5 Satunnaismetsät

Nimensä mukaisesti satunnaismetsissä luodaan joukko (ns. metsä) päätöspuita, joiden ennustamien lopputulosten enemmistöstä luodaan lopullinen algoritmin lopputulema. Kuvassa 2.6 esitetään yksinkertaistettu esimerkki satunnaismetsien toiminnasta, kolme erillistä päätöspuita tekee luokittelun itsenäisesti ja lopulta lopulliseen luokitteluun käytetään sitä tulosta, jonka enemmistö päätöspuista on tuottanut. Sa-



Kuva 2.6: Yksinkertaistettu esimerkki satunnaismetsien toiminnasta. [19]

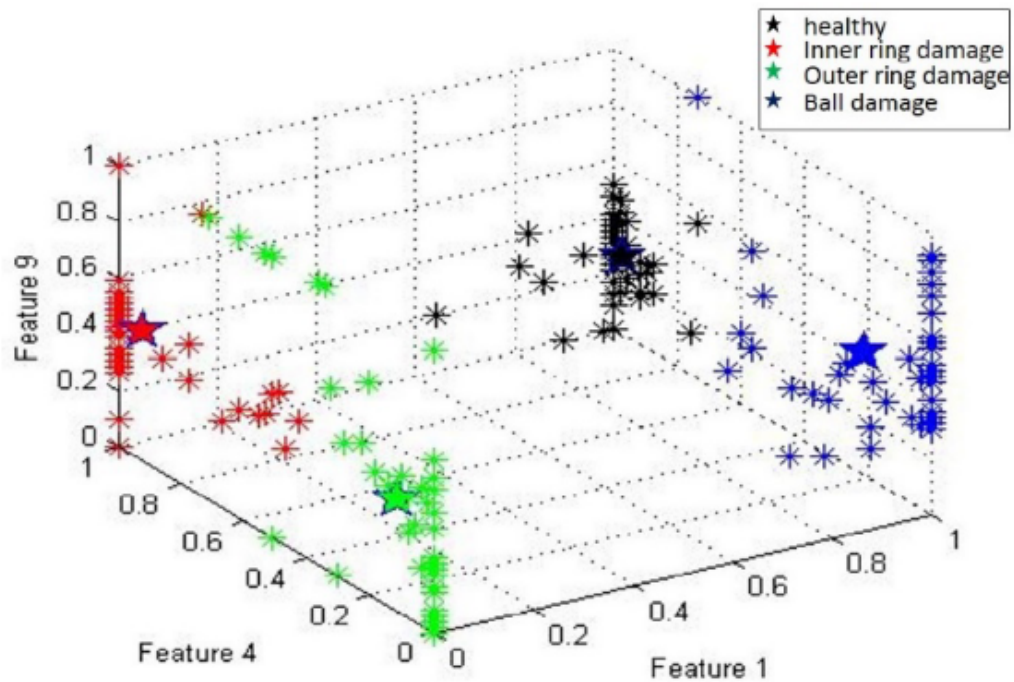
tunnaismetsissä käytetään kahta erilaista algoritmia ominaisuuksien valintamenetelyyn: ominaisuuksien valintaan sädehakua ja suodatusmenetelmänä Kolmogorov-Smirnovin testiin perustuvaa menetelmää. Kolmogorov-Smirnovin testiin tutustutaan tarkemmin kappaleessa 4.2.3. Menetelmän suosio perustuu sen kykyyn käsitellä suuren joukon muuttujia ennusteen tekemiseen. Menetelmän heikkoutena on sen monimutkaisuus, joka johtaa monia muita koneoppimisalgoritmeja suurempaan aikavaatimukseen. [4]

Satunnaismetsien käyttöä kunnonvalvonnassa on tutkittu laboratorio-olosuhteissa yhdessä geneettisten algoritmien kanssa. Tutkimuksessa rakennettiin ympäristö, jossa mitattiin hammaspyörän toimintaa seitsemässä eri tilanteessa normaalista ehjästä hammaspyörästä eriasteisilla vioilla varustettuihin hammaspyöriin. Hammaspyöriä pyöritettiin tasaisella 1650 kierroksen nopeudella sen yhdistäessä moottoria ja vaihdelaatikkaa. Seurannassa kerättiin tärinän raakadataa. Myöhemmin raakadatalle tehtiin esimerkiksi Fourier-muunnos ja laskettiin neliöllinen keskiarvo, sekä käytettiin muita statistisen analyysin klassisia menetelmiä. Satunnaismetsillä tehtävään analyysiin käytettävät datan ominaisuudet etsittiin geneettisillä algoritmeilla. Itse satunnaismetsäanalyysi tehtiin 800 syvää puuta sisältävän metsän avulla. Tutkimuksessa menetelmä todettiin toimivaksi ja sen uskottiin olevan sovellettavissa myös todelliseen tehdasympäristöön. [20]

Fourier-analyysi on esitelty tarkemmin kappaleessa 2.4.8.

### 2.4.6 K:n keskiarvon ryhmittely

K:n keskiarvon ryhmittely (engl. K-means clustering) on ohjaamattoman oppimisen luokittelualgoritmi, jossa kohteen luokittelu määritellään sitä lähimpänä olevien naapureiden luokittelun perusteella. Menetelmän nimessä oleva K-kirjain voidaan korvata numerolla, joka kertoo kuinka monta lähintä naapuria käytetään päätöksenteon apuna. Algoritmi käyttää ominaisuuksien samankaltaisuuden tulkintaan euklidista



Kuva 2.7: Esimerkki moniulotteisesta keskiarvoryhmittelmästä, mustat tähdet kuvaavat ehjiä laakereita, punaisella ovat tapaukset joissa laakerin sisäkehä on vioittunut ja vihreällä ulkokehästä vioittuneet. Sinisellä merkatuissa taas kehien välissä olevissa kuulissa on ollut vikaa. XYZ-akselit kuvaavat taas tutkimuksessa mitattuja ominaisuuksia. [21]

etäisyyttä. Kuvassa 2.7 on esimerkki moniulotteisesta eli useata eri ominaisuuksia hyödyntävästä ryhmittelmästä. [4]

Keskiarvoryhmittely on käytetty yhdessä geneettisten algoritmien kanssa laakereiden vikadiagnostiikkaan. Ennen keskiarvoryhmittelyä hyödynnettiin geneettisiä algoritmeja, jottei ryhmittely jäisi jumiin paikalliseen optimiin, vaan sille haluttiin antaa parempi lähtökohta kuin ennakkoon käsittelemätöntä dataa käytettäessä. Tutkimus tehtiin laboratorio-olosuhteissa ja se osoitti menetelmän toimivaksi. [21]

### 2.4.7 Pääkomponenttianalyysi

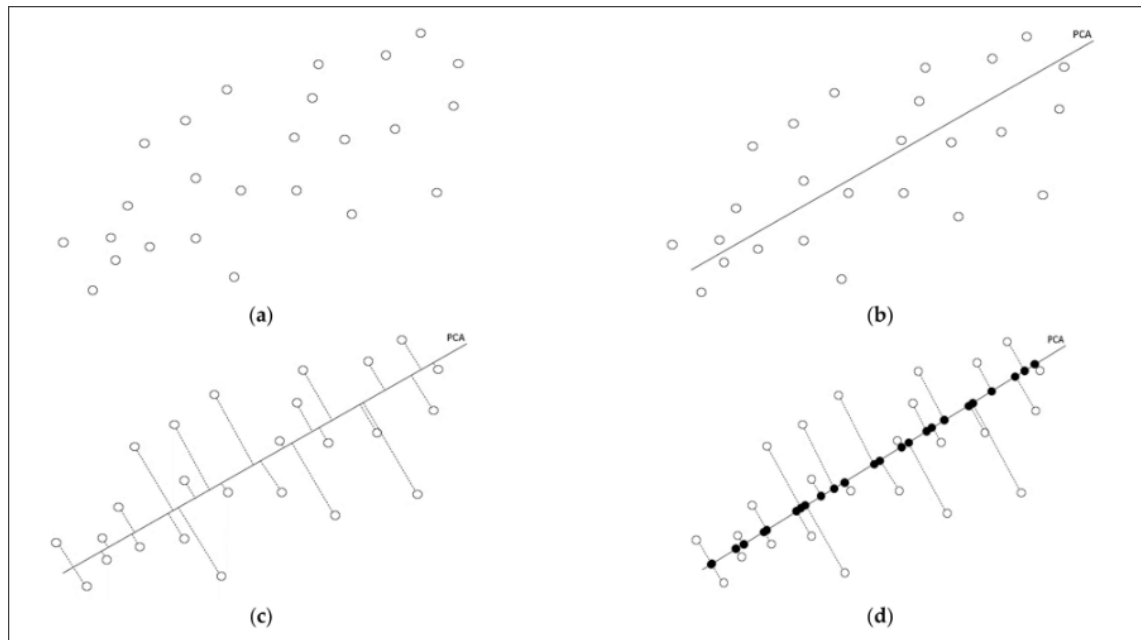
Pääkomponenttianalyysi (engl. principal component analysis, PCA) on valvomattoman oppimisen menetelmä, jossa pyritään löytämään datan keskeisimmät komponentit hukkaamatta merkittävää määrää dataa. Kuvassa 2.8 on esitetty pääkomponenttianalyysin toiminta pääpiirteittäin: Kohdassa (a) on alkuperäiset datapisteet esitettynä, (b) esittää optimaalisen PCA-vektorin valintaa. Kohdassa (c) alkuperäisten datapisteet suhteutetaan PCA-vektoriin ja vaiheessa (d) uusi tietojoukko määritellään suhteessa vektoriin. Menetelmän etuna on, että se tarkastelee jokaista datan pistettä riippumattomana komponenttina eikä etsi korrelaatioita niiden välillä. Tästä johtuen menetelmä voi merkittävästi säästää kouluttamiseen ja prosessointiin vaadittavaa aikaa. Heikkoutena on taas se, että analyysin aikana dataa saattaa kaikesta huolimatta hävitä, joka saattaa johtaa tulosten tarkkuuden heikentymiseen. [22]

Aikaisemmin pääkomponenttianalyysia on käytetty suurten riippusiltojen käytönaikaisessa kunnonvalvonnassa. Järjestelmä seuraa sillan värinävasteita reaaliaikaisesti ja tekee tarvittaessa hälytykset niiden perusteella. Yhteen siltaan on asennettu 12 sensoria, joista jokaisen värinää seurataan paikasta riippuen 1–3 akselilla. Järjestelmässä on luotu malli siltojen normaalitilasta, mistä poikkeaminen katsotaan häiriöksi. Siltojen liikennevirta on hyvin samanlainen päivästä toiseen. [23]

### 2.4.8 Fourier-analyysi

Klassisessa Fourier-analyysissä ideana on ilmaista signaalit sinimuotoisina komponentteina. Koneiden värinän analysoinnissa käytetään pääasiallisesti jaksollisia signaaleja. Muutamia poikkeuksia lukuunottamatta käytännössä kaikki signaalit on mahdollista hajottaa Fourier-analyysiin avulla. [5]

Fourier-analyysillä on kuitenkin heikkoutensa, jos esitettävät signaalit on hyvin lokalisoitu ajan suhteen. Näissä tapauksissa signaalikuvion havaitseminen ja tunnis-



Kuva 2.8: Pääkomponenttianalyysin toiminta [22].

taminen on vaikeaa. Tähän ratkaisuksi on kehitytty muuntaminen aallokepaketeiksi (engl. wavelet packet transform, WPT). Aallokepakettien käyttö mahdollistaa vaihtoehdoisen menetelmän löytää tärinästä sen aika-taajuus-informaatio. Menetelmässä vain vähän informaatiota esittävät ominaisuuskomponentit voidaan hylätä tilastollisia menetelmiä hyödyntävien valintakriteerien avulla. Supistettua piirrevektoria voidaan käyttää syötteenä esimerkiksi neuroverkkoon perustuvalla luokittelijalle, joka laskee merkittävästi verkon kouluttamiseen tarvittavaa aikaa ja parantaa neuroverkon yleistyskykyä. [24]

Kappaleessa 2.4.5 on esitelty aikaisempi tutkimus, jossa satunnaismetsien yhteydessä on käytetty Fourier-analyysia ja aallokepaketteja.

# 3 Sementtitehdas toimintaympäristönä

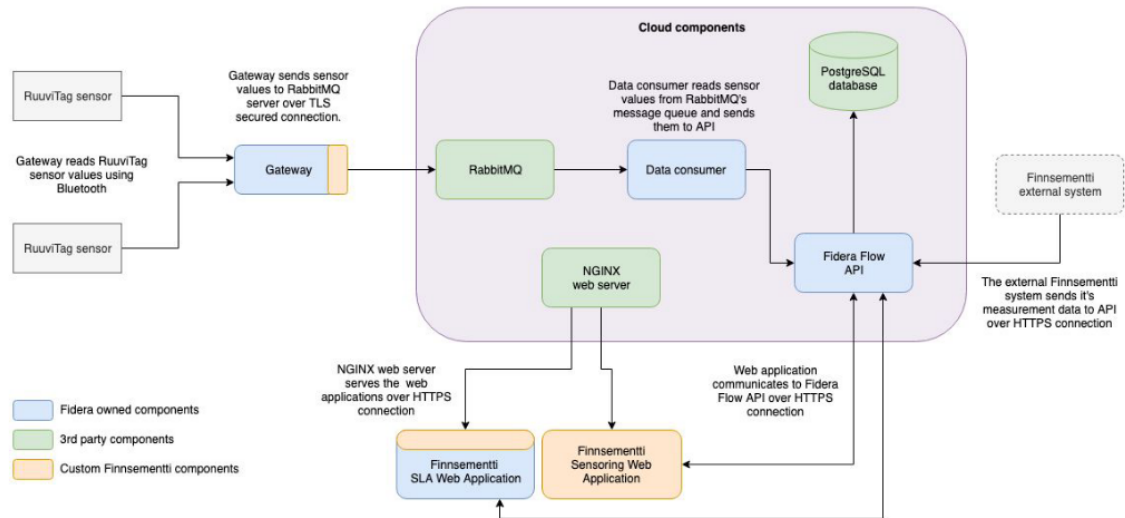
Vuonna 2018 Finnsementti osallistui yhtenä kilpailukohteena DNA:n järjestämään #iotnextlevel -kilpailuun, jonka voittajaksi selviytyi Fidera Oy. Kilpailun jälkeen Finnsementin Paraisten sementtitehtaalle asennettiin Fideran kehittämä Security Platform -järjestelmä [25][26].

## 3.1 Järjestelmän kuvaus

Kuvassa 3.1 esitetyssä toteutuksessa tehtaan koneisiin on kiinnitetty RuuviTag-sensoreita, jotka mittaavat värähtelyn ja lämpötilan 30 minuutin välein. Lisäksi mitaustuloksiin liitetään kirjauskerran yksilöivä järjestysnumero. Värähtelyarvo saadaan ottamalla näytteitä 400 Hz:n taajuudella noin 2 sekunnin ajan ja laskemalla niistä kiihtyvyyden neliöllinen keskiarvo. Sensori lähettää tuotettua dataa BLE-teknologialla 10 sekunnin välein, mutta data pysyy muuttumattoman seuraavaan mittaukseen asti. BLE tulee sanoista Bluetooth Low Energy, joka on vähävirtainen lyhyiden välimatkojen langaton tiedonsiirtoteknologia [27]. Datan lähettämällä useammin kuin mittauksia tehdään pyritään siihen, että yhdyskäytävä varmasti saa vastaanotettua tiedot huonommissakin olosuhteissa. [28]

Sensoreiden lähettämä data luetaan Fidera Flow gateway -yhdyskäytävälaitteilla. Laite erottaa eri sensorit toisistaan BLE MAC-osoitteiden avulla. MAC-osoite (engl.





Kuva 3.1: Nykyisen järjestelmän kuvaus [28].

Media Access Control) on laitteen yksilöivä uniikki 48-bittinen osoite, joka on määritetty valmistajan toimesta kaikille verkkoon liitettäville laitteille [29]. Yhdyskäytävä välittää saadut tiedot eteenpäin TLS-salattuna RabbitMQ-viestinvälitysohjelmistolle, josta Fidera Flow API -rajapinta tallentaa tiedot PostgreSQL-tietokantaan. TLS (engl. Transport Layer Security) on itse viestintäprotokollasta riippumaton teknologia kahden välisen viestinnän salaamiseen [30]. Mikäli yhdyskäytävä on päällä, mutta internetyhteys ei toimi, mittaukset tallennetaan välimuistiin, josta ne lähetetään yhteyden palautumisen jälkeen. [28]

## 3.2 Datan esittely

Analysoinnissa käytetty data sisältää sensoreiden keräämiä värinäarvoja yhteensä 1 685 891 kappaletta aikaväliltä 9.8.2019-22.1.2021. Taulukossa 3.1 on esitettyä viisi uusinta merkintää datasta. Mitatun värinäarvon neliöllisen keskiarvon lisäksi datassa on merkinnän aikaleima, koneen yksilöivä techIdent -tunnus, jolla laite tunnetaan muissa järjestelmissä, toiminnallinen alue (bigArea) sekä tarkempi sijainti



Kuva 3.2: RuuviTag avattuna.

(areaname).

Taulukko 3.1: Näyte kerätystä datasta

timestamp	rms	machinename	areaname	techIdent	bigArea
2021-01-22 04:58:44.944	109	KAUKOLÄMMÖN KIERTOPUMPPU	DISTRICT HEATING	24606PUM2	KILN
2021-01-22 04:58:52.703	28	RUUVIKULJETIN (PITKITTÄINEN)	ELECTRICAL FILTER	24796STR	KILN
2021-01-22 04:59:35.841	123	VAAKARUUVIKULJETIN 2	KILN FEED HOMOGENIZING	23241STR	KILN
2021-01-22 04:59:43.707	941	ILMARÄNNI SYKLONIEN VÄLISSÄ	RAWMILL R7 SEPARATING CIRCUIT	22336LUR	R7
2021-01-22 04:59:44.233	53	KAUKOLÄMMÖN KIERTOPUMPPU	DISTRICT HEATING	24610PUM	KILN

### 3.3 Käytetyt laitteet ja ohjelmistot

Toteutus voidaan jakaa kahteen osaan: tehtaalla oleviin fyysisiin laitteisiin sekä pilvikomponentteihin.

### 3.3.1 Fyysiset laitteet

#### RuuviTag

RuuviTag on suomalaisen Ruuvi Innovations Oy:n kehittämä Bluetooth Low Energy -sensori, jolla on mahdollista mitata lämpötilaa, ilmankosteutta, ilmanpainetta sekä liikettä. Sensorin laiteohjelmisto, elektroniikka ja siihen liittyvät mobiilisovellukset ovat kaikki avointa lähdekoodia. [31] Kuvassa 3.2 on RuuviTag avattuna ja siinä näkyvät paristo, laitteen käyttöpainikkeet sekä mittauksiin käytetyt komponentit.

#### Fidera Flow Gateway

Fidera Flow Gateway on 4G-modeemin sisältävä BLE-sensoreita lukeva yhdyskäytävälaite. [28]

### 3.3.2 Pilvikomponentit

#### RabbitMQ

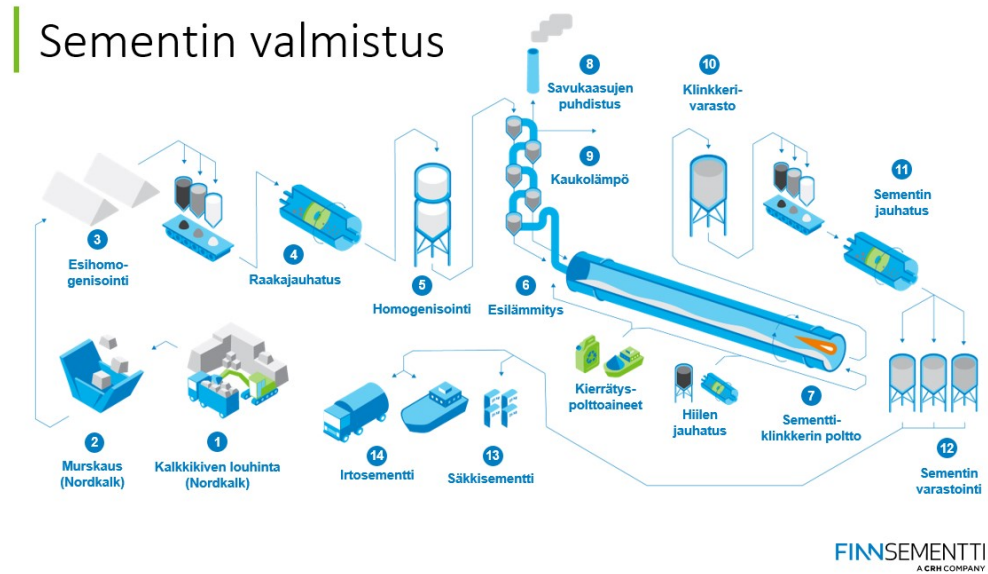
RabbitMQ on avoimen lähdekoodin viestinvälityspalvelin, joka tukee useita viestinvälitysprotokollia, kuten AMQP, STOMP ja MQTT. Ohjelmisto toteuttaa viestien välittämisen, jonojärjestyksen sekä reitittämisen yhdyskäytävän ja pilvikomponenttien välillä. [28]

#### PostgreSQL

PostgreSQL on vuodesta 1986 kehitteillä ollut avoimen lähdekoodin tietokannan hallintajärjestelmä, joka perustuu relaatiomalliin. [32]

#### Fidera Flow API ja Data Consumer

Fidera Flow API on NodeJS:llä toteutettu HTTP-rajapinta, joka tarjoaa käyttöliittymän tietokantaan sekä hoitaa datan tallentamisen SQL-tietokantaan. Data consu-



Kuva 3.3: Sementin valmistus. [33]

mer muuttaa yhdyskäytävien välittämän datan rajapinnalle sopivaan muotoon. Rajapinta mahdollistaa myös muiden järjestelmien tuottaman datan hyödyntämisen ja tallentamisen tietokantaan. [28]

### 3.4 Sementin valmistus

Kuvassa 3.3 on esitetty sementin valmistusprosessi pääpiirteittäin. Sementin valmistus alkaa sen pääraaka-aineen eli kalkkikiven louhinnasta ja murskauksesta. Murskattu kiviaines esihomogenisoidaan, jonka jälkeen kalkkikivi ja raaka-aineet jauheetaan raakajauhemylyssä hienoksi jauheeksi. Homogenisoinnilla tarkoitetaan prosessia, jossa seoksen sisältämät ainekset yritetään saada jakautumaan mahdollisimman tasaisesti. Valmiille raakajauheelle tehdään vielä lopullinen homogenisointi, jonka jälkeen aine esilämmitetään ja lopulta syötetään kiertouuniin, jossa lämpötila nousee jopa 1400 asteeseen. Sementtiuunia lämmitetään erilaisilla kierrätyspolttoaineilla,

kivihiilillä sekä petrokoksilla, ja uunin hukkalämpöä voidaan hyödyntää esimerkiksi kaukolämmössä. Korkea lämpötila saa aikaan kemiallisen reaktion, jossa kalkki-, pii-, alumiini- sekä rautayhdisteet muuttuvat kalsiumyhdisteiksi sintraantuen sementtiklinkkeriksi. Saatu klinkkeri varastoidaan ja lopulta jauhetaan kuulamylyillä yhdessä seosaineiden ja kipsin kanssa valmiiksi sementiksi. Erilaisia sementtejä saadaan säätelämällä klinkkerin koostumusta, käytettyjen seosaineiden suhteita sekä sementin hienoutta säätelämällä. Lopulta valmis sementti kuljetetaan asiakkaille, joko sellaisenaan tai säkkeihin pakattuna. [33] [34]

### 3.5 Kunnossapidon nykytila

Tällä hetkellä Finnsementin kunnossapitoa toteutetaan pääasiallisesti ehkäisevän kunnossapidon periaatteella. Määräaikaistarkastusten lisäksi ennakkohuoltotarkastajat suorittavat kunnonvalvontaa koneille ennalta laadittujen aikataulujen mukaisesti yhdessä kunnossapitoinsinöörin kanssa. Mikäli tarkastusten yhteydessä havaitaan asioita, jotka vaativat toimenpiteitä, tehdään vikailmoitus ja arvioidaan vaadittujen toimenpiteiden kiireellisyys. Mikäli korjauksen toteuttaminen vaatii tuotannon tai sen osan keskeyttämistä, mutta ei ole akuutti, huolto suoritetaan suunniteltujen tuotantokatkosten tai välitöntä korjausta vaativien vikojen huollon yhteydessä, jotta tuotantokatkoksia olisi mahdollisimman vähän. Töiden suunnittelusta vastaa kunnossapidon työnjohtaja. Pääasiassa kunnossapitotöistä vastaavat Finnsementin oman päiväryhmän kunnossapitomekaanikot, tarvittaessa töissä käytetään myös aliurakoitsijoita. [35]

### 3.6 Tehtaalle asennetut sensorit

Finnsementin Paraisten sementtitehtaalle on asennettu yhteensä 44 sensoria, jotka on sijoitettu kolmelle eri toiminnalliselle alueelle. Sementtiuuniin alueella olevat sensorit on listattu taulukossa 3.3, kuvan 3.3 kohdat 6-9), myllyihin kiinnitetyt sensorit on listattu taulukossa 3.2, seurannassa olevia myllyjä ovat raakamylly 7 (kuvassa 3.3 kohta 4) ja hiilimylly 6 (kuvassa 3.3 ennen kohtaa seitsemän).

Taulukko 3.2: Myllyihin asennetut sensorit

Koneen nimi	Toiminnallinen alue	Tarkempi sijainti
Keskipakopuhallin	Hiilimyly 6	Hiilimyly 6
Ketjusyötin	Hiilimyly 6	Hiilimyly 6
Pölysuodatin	Hiilimyly 6	Hiilimyly 6
Ruuvikuljetin Hiilisykloonasta	Hiilimyly 6	Hiilimyly 6
Ruuvipumppu	Hiilimyly 6	Hiilimyly 6
Ilmaränni Syklonien Välissä	Raakamyly 7	Erotuspiiri
Ketjuelevaattori 1	Raakamyly 7	Erotuspiiri
Ketjuelevaattori 2	Raakamyly 7	Erotuspiiri
Separaattori 1	Raakamyly 7	Erotuspiiri
Separaattori 2	Raakamyly 7	Erotuspiiri
Siilon Elevaattori 3 (raakajauho)	Raakamyly 7	Erotuspiiri
Tarkastusseula Pyörivä	Raakamyly 7	Erotuspiiri
Kaksiportainen Hammasvaihde	Raakamyly 7	Myllypiiri
Kiertoöljypumppu	Raakamyly 7	Myllypiiri
Kiertoöljypumppu 2,2 Kw Laakeri Sisään	Raakamyly 7	Myllypiiri
Kiertoöljypumppu 2,2 Kw Laakeri Ulos	Raakamyly 7	Myllypiiri
Voitelupumppu Hammaskehälle	Raakamyly 7	Myllypiiri

Taulukko 3.3: Sementtiuuniin asennetut sensorit

Koneen nimi	Toiminnallinen alue	Tarkempi sijainti
Jauhevaaka Nr.2 A-Puoli	Sementtiuuni	A-Puoli
Sulkusyötin 1. A-Puoli	Sementtiuuni	A-Puoli
Sulkusyötin A-Puoli	Sementtiuuni	A-Puoli
Jauhevaaka Nr.2 B-Puoli	Sementtiuuni	B-Puoli
Sulkusyötin 2. B-Puoli	Sementtiuuni	B-Puoli
Sulkusyötin, B-Puoli	Sementtiuuni	B-Puoli
Ruuvikuljetin Jäähdytystorni	Sementtiuuni	Jäähdytystorni
Kaukolämmön Kiertopumppu	Sementtiuuni	Kaukolämpö
Annosteluruuvi	Sementtiuuni	Kierrätyspolttoaineen syöttö
Sulkusyötin	Sementtiuuni	Kierrätyspolttoaineen syöttö
Ruuvikuljetin (pitkittäinen)	Sementtiuuni	Sähkösuodatin
Ruuvikuljetin (poikittainen)	Sementtiuuni	Sähkösuodatin
Ruuvikuljetin (yli Menevä)	Sementtiuuni	Sähkösuodatin
Ruuvikuljetin, Sähkösuodatin	Sementtiuuni	Sähkösuodatin
Sähkösuodattimen Puhallin	Sementtiuuni	Sähkösuodatin
Unin Elevaattori	Sementtiuuni	Syötön homogenisointi
Vaakaruuvikuljetin	Sementtiuuni	Syötön homogenisointi
Vaakaruuvikuljetin 2	Sementtiuuni	Syötön homogenisointi
Vaakaruuvikuljetin 3	Sementtiuuni	Syötön homogenisointi
Varaelevaattori	Sementtiuuni	Syötön homogenisointi



# 4 Koneiden normaalitilan mallintaminen

Työn kokeellinen osuus toteutettiin noudattamalla Cross-industry standard process for data mining eli CRISP-DM -menetelmää. Lopulta käytettävissä olevaa tärinädataa päätettiin hyödyntää mallintamalla sensoreiden tuottamasta datasta koneiden normaalitila, jonka poikkeamat katsotaan vikatilanteiksi.

## 4.1 Cross-industry standard process for datamining

CRISP-DM sisältää data-analytiikkaprojektin tyypilliset vaiheet ja niihin liittyvät tehtävät kuvauksineen; eri vaiheita on yhteensä kuusi. [36]

Prosessin vaiheita ei ole määritelty tiukasti lineaariseen järjestykseen vaan on tyypillistä, että eri vaiheiden välillä liikutaan joustavasti tarpeen mukaan edestakaisin. Lisäksi vaiheita voidaan poistaa tai lisätä, mikäli tarve niin vaatii. [36] Kuvassa 4.1 on yleiskatsaus prosessin vaiheisiin ja vaiheiden tavoitteisiin.

### 4.1.1 Liiketoiminnan ymmärtäminen

Mallin ensimmäinen vaihe on liiketoiminnan sekä sen tarpeiden ja tavoitteiden ymmärtäminen. Liiketoiminnan ymmärtämisen ensimmäinen vaihe on määritellä ongelma, joka tiedon louhinnalla halutaan ratkaista, sekä määritellä kaikki siihen liitty-

<b>Business Understanding</b>	<b>Data Understanding</b>	<b>Data Preparation</b>	<b>Modeling</b>	<b>Evaluation</b>	<b>Deployment</b>
<p><b>Determine Business Objectives</b> Background Business Objectives Business Success Criteria</p> <p><b>Assess Situation</b> Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</p> <p><b>Determine Data Mining Goals</b> Data Mining Goals Data Mining Success Criteria</p> <p><b>Produce Project Plan</b> Project Plan Initial Assessment of Tools and Techniques</p>	<p><b>Collect Initial Data</b> Initial Data Collection Report</p> <p><b>Describe Data</b> Data Description Report</p> <p><b>Explore Data</b> Data Exploration Report</p> <p><b>Verify Data Quality</b> Data Quality Report</p>	<p><i>Data Set</i> Data Set Description</p> <p><b>Select Data</b> Rationale for Inclusion/ Exclusion</p> <p><b>Clean Data</b> Data Cleaning Report</p> <p><b>Construct Data</b> Derived Attributes Generated Records</p> <p><b>Integrate Data</b> Merged Data</p> <p><b>Format Data</b> Reformatted Data</p>	<p><b>Select Modeling Technique</b> Modeling Technique Modeling Assumptions</p> <p><b>Generate Test Design</b> Test Design</p> <p><b>Build Model</b> Parameter Settings Models Model Description</p> <p><b>Assess Model</b> Model Assessment Revised Parameter Settings</p>	<p><b>Evaluate Results</b> Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</p> <p><b>Review Process</b> Review of Process</p> <p><b>Determine Next Steps</b> List of Possible Actions Decision</p>	<p><b>Plan Deployment</b> Deployment Plan</p> <p><b>Plan Monitoring and Maintenance</b> Monitoring and Maintenance Plan</p> <p><b>Produce Final Report</b> Final Report Final Presentation</p> <p><b>Review Project</b> Experience Documentation</p>

Kuva 4.1: Yhteenvedo CRISP-DM:n vaiheista ja tuloksista. [37]

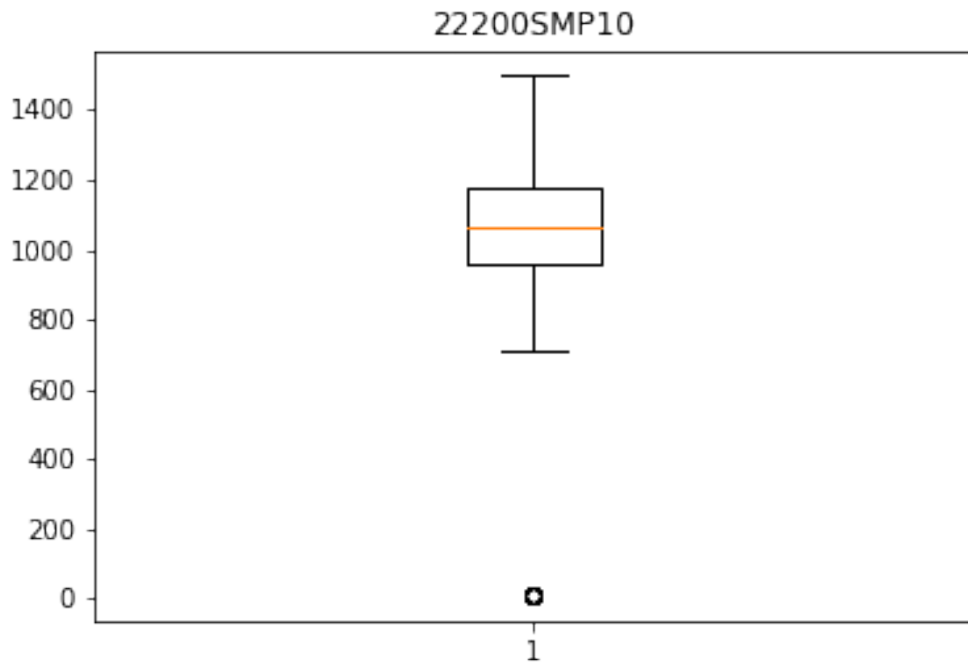
vät liiketoimintakysymykset ja sen asettamat vaatimukset mahdollisimman tarkasti. Lisäksi tulee määritellä ongelman ratkaisemisesta saatava hyöty. [36] [37]

Tämän työn tavoitteena on kehittää mallit, joissa hyödynnetään värinän seurantajärjestelmän keräämää dataa. Mallien avulla pyritään havaitsemaan laitteissa olevia ongelmia ennakkoon, jotta huollot voitaisiin tehdä suunnitellusti eikä tuotantoon synny katkoksia. Suunnittelemattomat tuotantokatkokset kasvattavat kustannuksia sekä johtavat menetettyyn työaikaan.

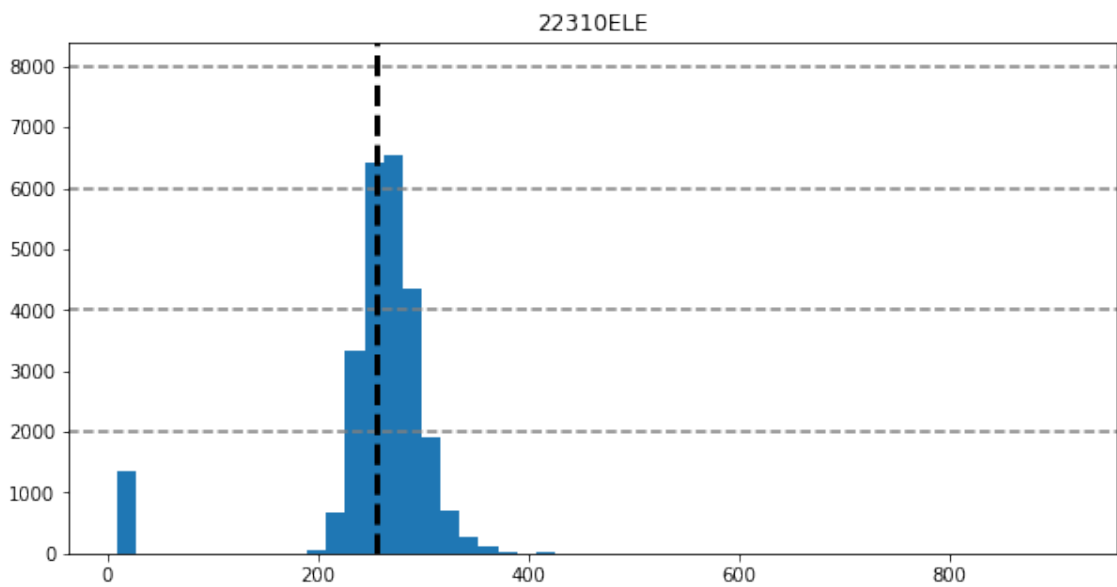
### 4.1.2 Datan ymmärtäminen

Tässä vaiheessa luodaan tarkempi katsaus käytettävissä olevaan dataan, jotta vältetään ongelmia datan valmistelussa. Mikäli olemassa oleva data ei tyydytä tarpeita, dataa voidaan hankkia lisää joko ostamalla tarpeisiin sopivaa dataa kolmannelta osapuolelta tai kerätä kokonaan uutta dataa. Datan keräämisen jälkeen luodaan katsaus dataan, sen määrään, muuttujien tyypeihin sekä siihen, että eri lähteistä olevassa datassa ei ole ristiriitoja tiedon esittämisessä. Monesti on hyödyllistä käyttää myös perinteisiä tilastotieteen työkaluja dataan tutustumisessa. [36] [37] Esimerkiksi kuvassa 4.2 on laatikkokuvaaja raakamyly 7:n hammaskehän voitelupumpun tärinästä. Kuvaajasta nähdään, että värinän neliöllisen keskiarvon mediaani on 1067 (oranssi viiva). 25. persentiili on 958 (laatikon alareuna), 75. persentiili on 1174 (laatikon yläreuna). Vaihteluväli on 1497 (ylin viiva) ja 720 (alin viiva). Lisäksi nollan lähellä on poikkeava havainto, joka on kuvattu ympyrällä. Koska data on harvoin täydellistä, osa datan ymmärtämisestä on myös sen laadun varmistaminen. Tietoa saattaa puuttua tai siinä voi olla mittavirheitä tai tiedon syöttämisessä sattuneita virheitä. Lisäksi muuttujien tai sarakkeiden nimissä saattaa olla epä johdonmukaisuuksia tai tiedon muotoilu on muuttunut rivien välillä. [36] [37]

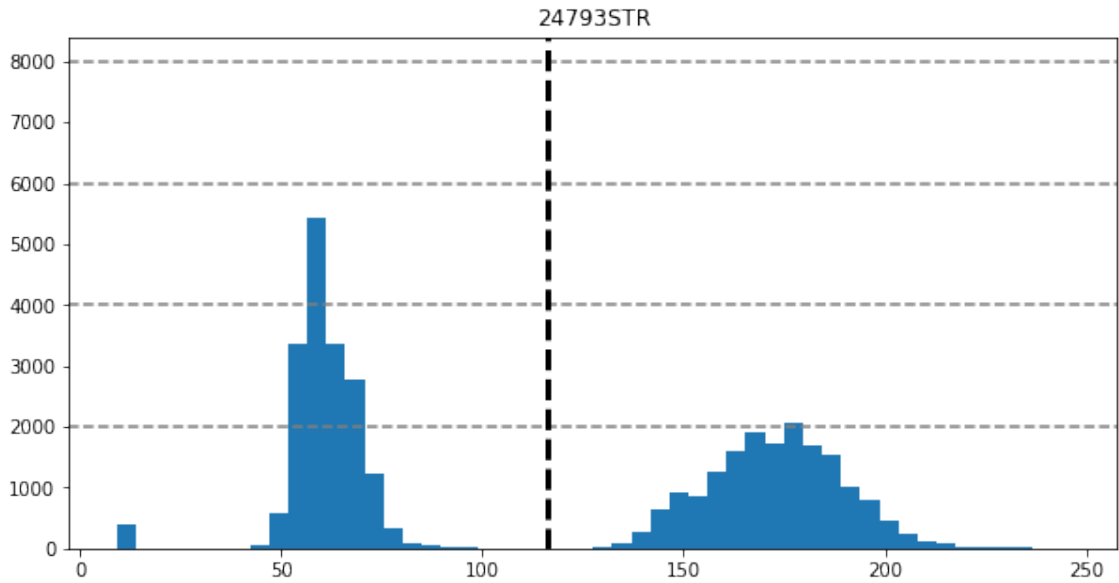
Histogrammi on pylväskuvio, joka esittää datassa olevien arvojoukkojen suhteita toisiinsa [38]. Kuvan 4.3 histogrammi esittää raakamyly 7:n ketjuelevaattorin nu-



Kuva 4.2: Laatikkokuvaaja raakamyly 7:n hammaskehän voitelupumpun värinästä.



Kuva 4.3: Histogrammi raakamyly seitsemän ketjuelevaattori ykkösen tärinästä.



Kuva 4.4: Histogrammi sementtiuunin ruuvikuljettimen sähkösuodattimen tärinästä.

mero yksi tärinää ja histogrammissa 4.4 on esitetty sementtiuunin ruuvikuljettimen sähkösuodattimelta kerätyt tärinäarvot. Mustalla pystyviivalla on esitetty datassa esiintyvien tärinäarvojen keskiarvo. Huomionarvoista on, että kuvaajien x-akselit eivät ole samalla leveydellä, eivätkä myöskään pylväiden leveydet ole samat ja muutenkin koneista kerätyt tärinäarvot ovat jakaumaltaan hyvin erilaiset. Ketjuelevaattorin tärinän neliöllinen keskiarvo on välillä 200–400, mutta sähkösuodattimen tapauksessa tärinää on kahdella eri alueella välillä 50–80 ja 140–220 keskiarvon jäädessä näiden kahden alueen väliin, jossa tärinää itsessään ei pääsääntöisesti havaita.

Analysointiin käytettävä oleva data sisältää 44 eri sensorilta kerättyjä värinäarvoja yhteensä 1 685 891 kappaletta. Data on esitelty tarkemmin kappaleessa 3.2. Datan laadussa oli joitakin ongelmia, esimerkiksi puolen tunnin väli tiedontallennuksessa aiheutti haasteita, sekä datapisteen aikaleima tulee kellonajasta, jolloin tieto tallennettiin tietokantaan eikä siitä, milloin se kerättiin yhdyskäytävän väliin. Tämä aiheutti sen, että vaikka kolme datapistettä kerättiin puolen tunnin

välein, niiden aikaleimat olivat muutaman sekunnin sisällä toisistaan silloin, kun yhdyskäytävän yhteys tietokantaan palautui. Lisäksi koneiden luonteenomainen käyttö hankaloittaa aikasarja-analytiikkaa, sillä koneet voivat olla poissa päältä tai eri käyttöasteilla tuotannon tarpeista riippuen.

### 4.1.3 Datan valmistelu

Datan valmistelussa rakennetaan varsinaiset tietojoukot, joita analysoinnissa käytetään. Tietojoukot rakennetaan yhdistelemällä eri lähteistä kerättyä dataa analyysiin soveltuviksi tauluiksi. Tässä vaiheessa dataa myös siistitään esimerkiksi korvaamalla tai poistamalla tyhjiä tai puuttuvia arvoja. [37] [36]

Valmisteluvaiheessa yhdistin tärinädatan tietoon toiminnallisten alueiden toiminnasta ja poistin sellaiset värähtelyarvot sellaisilta ajanjaksoilta, jolloin toiminnallinen alue ei ollut lainkaan käytössä. Osana datan valmistelua mallintamista varten oli datan jakaminen koulutus- ja testidataan. Aluksi jaoin sensoreiden datat koulutus- ja testidatoihin laittamalla ensimmäiset 67 % koulutusdataksi ja loput 33 % testidataan. Mallintamisvaiheessa tämä kuitenkin osoittautui huonoksi ratkaisuksi ja paremmat tulokset saavutettiin, kun kaksi peräkkäistä päivää laitettiin koulutusdataan ja kolmas testidataksi. Tätä toistettiin kunnes koko data oli käyty läpi.

### 4.1.4 Mallintaminen

Mallintamisvaiheessa dataan käytetään erilaisia mallinnustekniikoita ja yritetään löytää optimaalisin tekniikka ja malli datan esittämiseen sekä analytiikan tekemiseen. Datan valmistelun ja mallintamisen suhde on läheinen, usein mallintaessa datasta löytyy uusia ongelmia, joita voidaan korjata palaamalla valmisteluvaiheeseen. Lisäksi mallintamisessa valitaan metodi, jolla mallin toimivuutta voidaan arvioida. [37]

Mallien toimivuuden todentamiseen käytin pääasiassa Kolmogorov-Smirnovin testiä. Testi perustuu empiirisen ja hypoteettisen kumulatiivisen jakaumafunktion eroihin [39]. Mallintaminen käytännössä käydään läpi kappaleessa 4.2 ja Kolmogorov-Smirnovin testiin kappaleessa 4.2.3.

#### 4.1.5 Tulosten arviointi

Tulosten arviointiin päästään, kun tiedon louhinta on tehty ja todettu, että mallit ovat teknisesti oikein ja ne täyttävät projektille asetetut tavoitteet. Tulosten arvioinnissa ei perehdytä enää mallin toimivuuteen vaan siihen, miten hyvin tulokset vastaavat liiketoiminnan asettamiin tavoitteisiin ja kysymyksiin. [36]

#### 4.1.6 Käyttöönotto

Viimeisenä vaiheena on mallin jalkauttaminen tuotantoon. Projektista riippuen käyttöönotto voi yksinkertaisimmillaan olla pelkkä loppuraportin tuottaminen, mutta monimutkaisemmissa tapauksissa voidaan vaatia toistuvan tiedonlouhintaprosessin implementointia. Useissa tapauksissa varsinainen implementointi ei jää data-analyytikon tehtäväksi vaan siitä on vastuussa tehtävänannon antanut organisaatio. [37]

## 4.2 Mallintaminen

Mallintamisessa käytössä ei ollut dataa, jolla koneilta kerätyt värinäarvot olisi ollut mahdollista luokitella, milloin koneet ovat toimineet normaalisti ja milloin eivät. Tästä syystä päädyttiin käyttämään menetelmiä, joilla mallinnetaan koneiden normaalien toimintatilojen värinäarvojen jakauma. Jos tästä normaalitilasta poiketaan tilanne tulkitaan, epäilyksi vikatilanteeksi ja kunnossapito voi tämän perusteella ryhtyä selvittämään oikeita toimenpiteitä. Normaalitilanteen mallintamista kokeil-

tiin erilaisilla menetelmillä, kuten parametrisillä ja parametrittömillä tiheysfunktion estimointimenetelmillä.

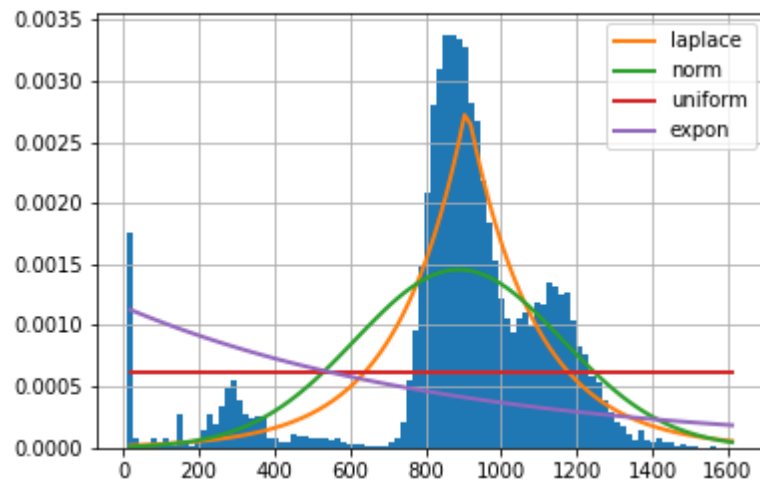
### 4.2.1 Parametriset todennäköisyysjakaumat

Sopivien jakaumien löytämiseen käytettiin Fitter-nimistä Python-kirjastoa, jonka avulla on mahdollista etsiä datan sopivuutta kahdeksallakymmenellä erilaiselle Sci-pyn tukemalle jakaumalle, joille kirjasto selvittää myös parhaimmat mahdolliset parametrit. Lisäksi kirjasto myös pisteyttää mallien soveltuvuuden neljällä eri menetelmällä: neliösummavirhe (engl. Sum square error, SSE), Akiaken tietokriteeri (engl. Akaike information criterion, AIC), Bayesialainen tietokriteeri (engl. Bayesian information criterion, BIC) sekä Kullback–Leibler-divergenssi. Fitter myös tulostaa datan histogrammin ja näyttää graafisen vertailun parhaiten soveltuviin jakaumiin.

[40]

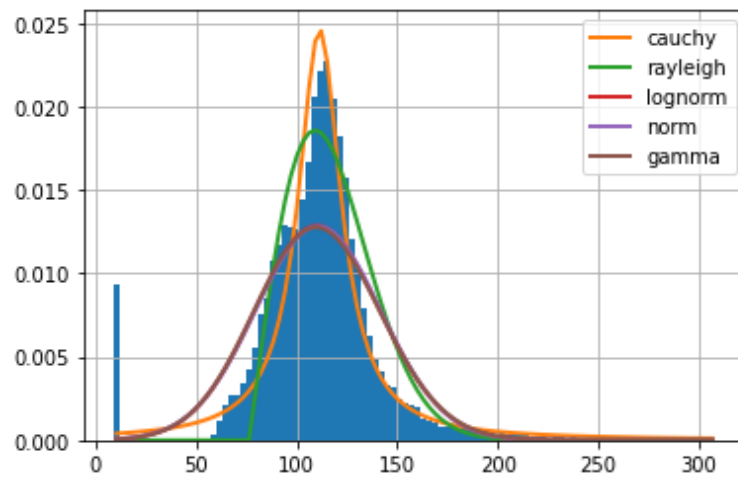


	sumsquare_error	aic	bic	kl_div
<b>laplace</b>	0.000020	1637.035843	-2.621472e+05	inf
<b>norm</b>	0.000043	1604.976786	-2.524737e+05	inf
<b>uniform</b>	0.000081	1480.797892	-2.441781e+05	inf
<b>expon</b>	0.000099	1542.904138	-2.415888e+05	inf
<b>alpha</b>	inf	inf	inf	inf



Kuva 4.5: Fitter-kirjaston tulokset raakamylyy 7:n erotuspiirin separaattori 1:n tärinäälle.

	sumsquare_error	aic	bic	kl_div
<b>cauchy</b>	0.000267	1396.854273	-674378.081091	inf
<b>rayleigh</b>	0.000379	inf	-661744.720705	inf
<b>lognorm</b>	0.000642	1854.385449	-642769.871152	inf
<b>norm</b>	0.000645	1883.178831	-642609.325871	inf
<b>gamma</b>	0.000654	1815.290085	-642098.610196	inf



Kuva 4.6: Fitter-kirjaston tulokset sementtiuunin vaakaruuvikuljetin kakkosen tärinäälle.

Kuvassa 4.5 on Fitter-kirjaston tuottama analyysi raakamyly seitsemän erotuspiirin separaattori 1:n tärinädatasta. Fitter on valikoinut parhaiten soveltuviksi jakaumiksi Laplacen, normaalijakauman, tasajakauman (engl. continuous uniform distribution), eksponenttijakauman (engl. exponential distribution), myös muita jakaumia koitettiin mutta ne onnistuivat kuvaamaan dataa vielä heikommin. Tuloksista voidaan todeta, ettei edes mikään "parhaiten" sopivista jakaumista kuvaa dataa riittäväällä tarkkuudella, eikä Kullback-Leibler-divergenssin avulla ole edes onnistuttu pisteyttämään yhtäkään vertailuun valikoituneista jakaumista. Neliösummavirheen perusteella parhaiten soveltuva jakauma olisi laplace-jakauma, mutta kuvajasta voidaan todeta, ettei sekään kuvaa todellista dataa onnistuneesti.

Kuva 4.6 esittää sementtiuunin vaakaruuvikuljettimen numero kaksi tärinästä tehty analyysi. Silmämääräisesti löydetty parametriset jakaumat muistuttavat edellistä hieman paremmin todellista dataa, mutta käytetyt testit osoittavat, ettei siinä kuitenkaan onnistuta kovin hyvin. Aikaisemmin mainitun normaalijakauman lisäksi parhaiten soveltuvia jakaumia olivat testin mukaan Cauchy, Rayleigh, gamma sekä logaritminormaalijakauma.

$$f(x, \theta, s) = \frac{1}{2s} e^{-|x-\theta|/s}, \quad (1)$$

Kaava 1 on laplace-jakauman tiheysfunktion kaava, jossa sijainnin parametri  $\theta = (-\infty, \infty)$  ja skaalan parametri  $s > 0$ . Sekä  $-\infty < x < \infty$ . [41].

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2)$$

Kaava 2 on normaalijakauman tiheysfunktion kaava. Kaavan vakioista  $\mu$  on keskiarvo, odotusarvo sekä mediaani,  $\sigma^2$  on varianssi. Kaavassa  $|x| < \infty$ ,  $\mu < \infty$  ja varianssin tulee olla nollaa suurempi. [42]

$$f(x) = \lambda e^{-\lambda x}, \text{ jos } x \geq 0, \quad (3)$$

$$f(x) = 0, \text{ jos } x < 0, \quad (4)$$

Kaavat 3 ja 4 ovat eksponenttijakauman tiheysfunktioita, kaavaa 3 käytetään jos  $x \geq 0$ . Kaavassa  $\lambda$  on odotusarvon käänteisluku. Kaavaa neljä käytetään tilanteessa jossa  $x < 0$ , jolloin koko jakauma on nolla. [43]

$$f(x|a, b) = \frac{1}{b-a}, \quad (5)$$

Tasajakauman tiheysfunktion kaava on kohdassa 5. Edellä mainituista jakaumista poiketen jakauma on nimensä mukaisesti tasainen kahden rajaparametrin välillä. Parametri  $a$  ( $-\infty < a < b$ ) on alempi päätepiste ja ylempi on  $b$  ( $a < b < \infty$ ). Muissa tapauksissa jakauma on nolla. [44]

$$f(x, \alpha, \lambda) = (1 - e^{-(\lambda x)^2})^\alpha, \quad (6)$$

Kaava 6 on Rayleigh jakauman tiheysfunktio. Kaavan vakioista  $\alpha$  on jakauman muoto ja  $\lambda$  on sen skaala.  $X$ :n tulee olla nollaa suurempi. [45]

$$f(x, s) = \frac{1}{sx\sqrt{2\pi}} e\left(-\frac{\log^2(x)}{2s^2}\right), \quad (7)$$

Logaritminormaalijakauman tiheysfunktio on kaava 7, jossa  $s$  on muodon parametri [46].

$$f(x, \alpha) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \quad (8)$$

Gammajakauman tiheysfunktion kaavassa (kaava 8),  $\alpha$  on muodon parametri ja  $\Gamma(\alpha)$  on gammafunktio [47].

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad (9)$$

Kaava 9 on Cauchyn jakauman tiheysfunktio [48].

$$f(x, \alpha) = \frac{1}{x^2 \phi(\alpha) \sqrt{2\pi}} * \exp\left(-\frac{1}{2}\left(\alpha - \frac{1}{x}\right)^2\right), \quad (10)$$

Kaava 10 on Alpha jakauman tiheysfunktio, missä  $x > 0$  ja muodon parametri  $\alpha > 0$ .  $\phi$  on normaali kertymäfunktio. [49]

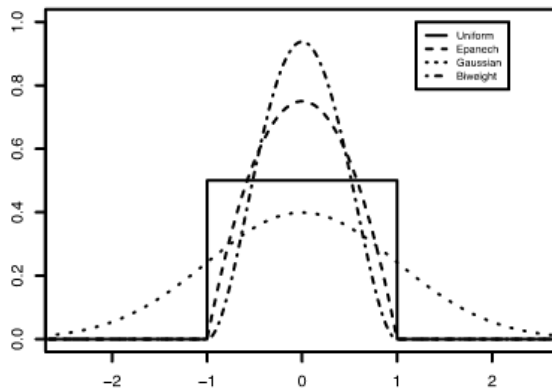
### 4.2.2 Ydinestimointi

Ydinestimointi (engl. kernel density estimation, KDE) on parametritön tiheysfunktio, toisin kuin histogrammi ydinestimointi luo tasaisemman mallin tiheysfunktioista (engl. probability density function, PDF) [50]. Murray Rosenblat esitti ensimmäisen kerran ydinestimoinnin konseptin vuonna 1956.

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K \frac{x-x_i}{h}, \quad (11)$$

Kaavassa 11 on esitetty ydinestimointi, jossa funktio  $K$  on ydinfunktio ja  $h$  on ytimen leveys.  $X$ :llä esitetään otoksia jakaumasta, joiden tiheyttä ei tunneta. Ytimen leveyden tulee olla suurempi kuin nolla [51].  $N$  tarkoittaa otannan kokoa [52]. Ydinestimointi voidaan toteuttaa useilla erilaisilla ydinfunktioilla, joista muutamia on esitelty kuvassa 4.7 [53].

Ytimen leveyden valinta on oleellinen osa estimoinnin onnistumista, ja joidenkin tutkimusten mukaan sen valinta on jopa tärkeämpää kuin käytetyn ytimen. Pienet ytimen leveydet kasvattavat luodun mallin varianssia, mikä saa sen näyttämään heiluilta, mutta liika ytimen leveys taas saattaa tehdä siitä liian tasaisen eikä se silloin esitä mallin pohjana käytetyn datan todellisia piirteitä. Kuten taulukoista 4.1 ja 4.2 nähdään, sopivat ytimen leveydet ovat todella yksilöllisiä. [54]

**Figure 3.3** Kernel weight functions

Kernel weight	$K(x)$
Uniform	$\frac{1}{2}I( x  \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$
Epanechnikov	$\frac{3}{4}(1-x^2)I( x  \leq 1)$
Biweight	$\frac{15}{16}(1-x^2)^2I( x  \leq 1)$
Triweight	$\frac{35}{32}(1-x^2)^3I( x  \leq 1)$

**Table 3.1** : Kernel weight functions.

Kuva 4.7: Ydineestimoinnissa käytettyjä ydinfunktioita [53].

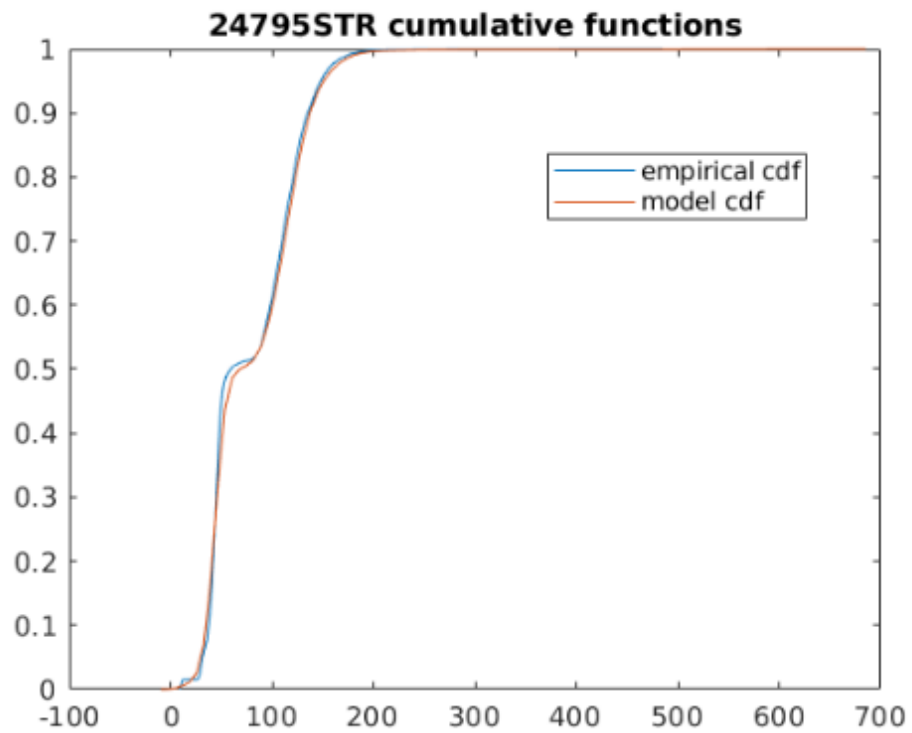
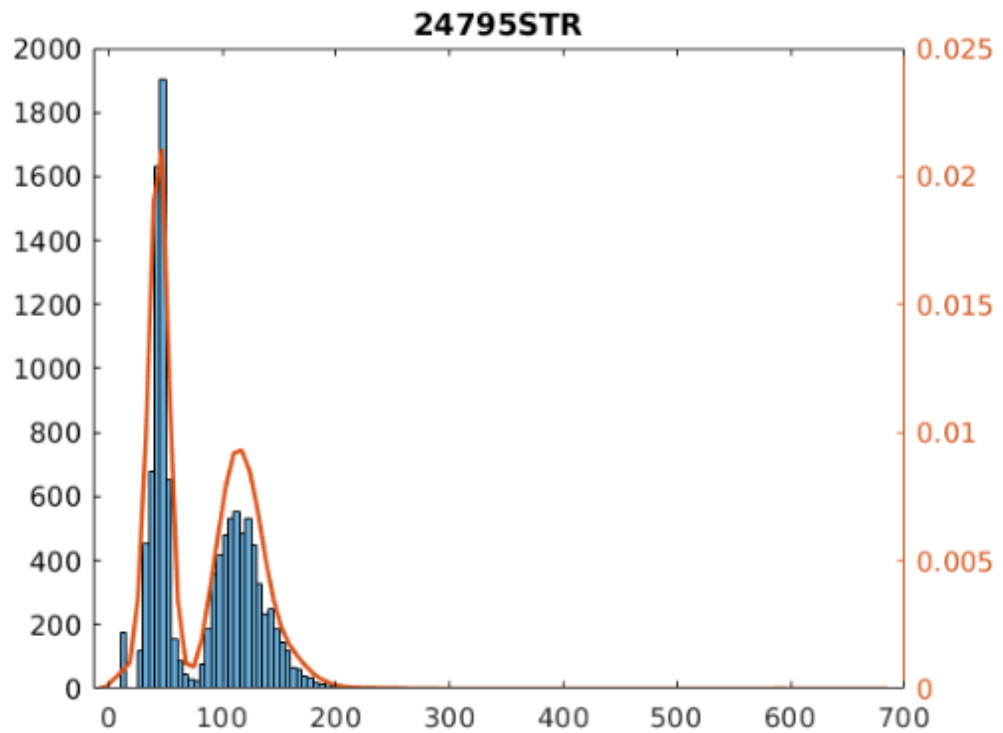
### Ydineestimoinnin toteuttaminen

Toteutin ydineestimoinnin Matlabin `ksdensity`-funktioilla. Oletuksena Matlab käyttää ytimenä normaaliytimen tasoitusfunktioita (engl. normal kernel smoothing) ja laskee itsenäisesti ytimelle sopivimman leveyden [55]. Kuvassa 4.8 on sementtiuunin sähkösuodattimen poikittaisen ruuvikuljettimen ydineestimointi sekä siihen liittyvät kumulatiiviset funktiot. Ylemmän kuvaajan oranssi viiva on ydineestimoinnilla luotu malli ja sinisellä oleva histogrammi esittää testidataa. Silmämääräisesti voidaan todeta, että luotu malli ja testidata vastaavat varsin hyvin toisiaan. Alemmassa kuvaajassa on mallin toimivuuden arviointiin käytetyt kumulatiiviset funktiot, sininen viiva esittää testidataa ja oranssi luotua mallia. Tästäkin voidaan todeta, että silmämääräisesti malli muistuttaa todellista dataa. Lopuksi kuvaajien alla on kahden näytteen Kolmogorov-Smirnov-testin tulos, jonka mukaan molemmat datat tulevat samasta jakaumasta. Voidaan todeta, että tälle ruuvikuljettimelle onnistuttiin luomaan sen tärinää kuvastava malli ydineestimoinnin avulla.

Kuvassa 4.9 on sementtiuunin syötön homogenisoinnin 3. vaakaruuviikuljettimen ydineestimointi. Silmämääräisesti voidaan todeta, että luotu malli muistuttaa etäi-

---

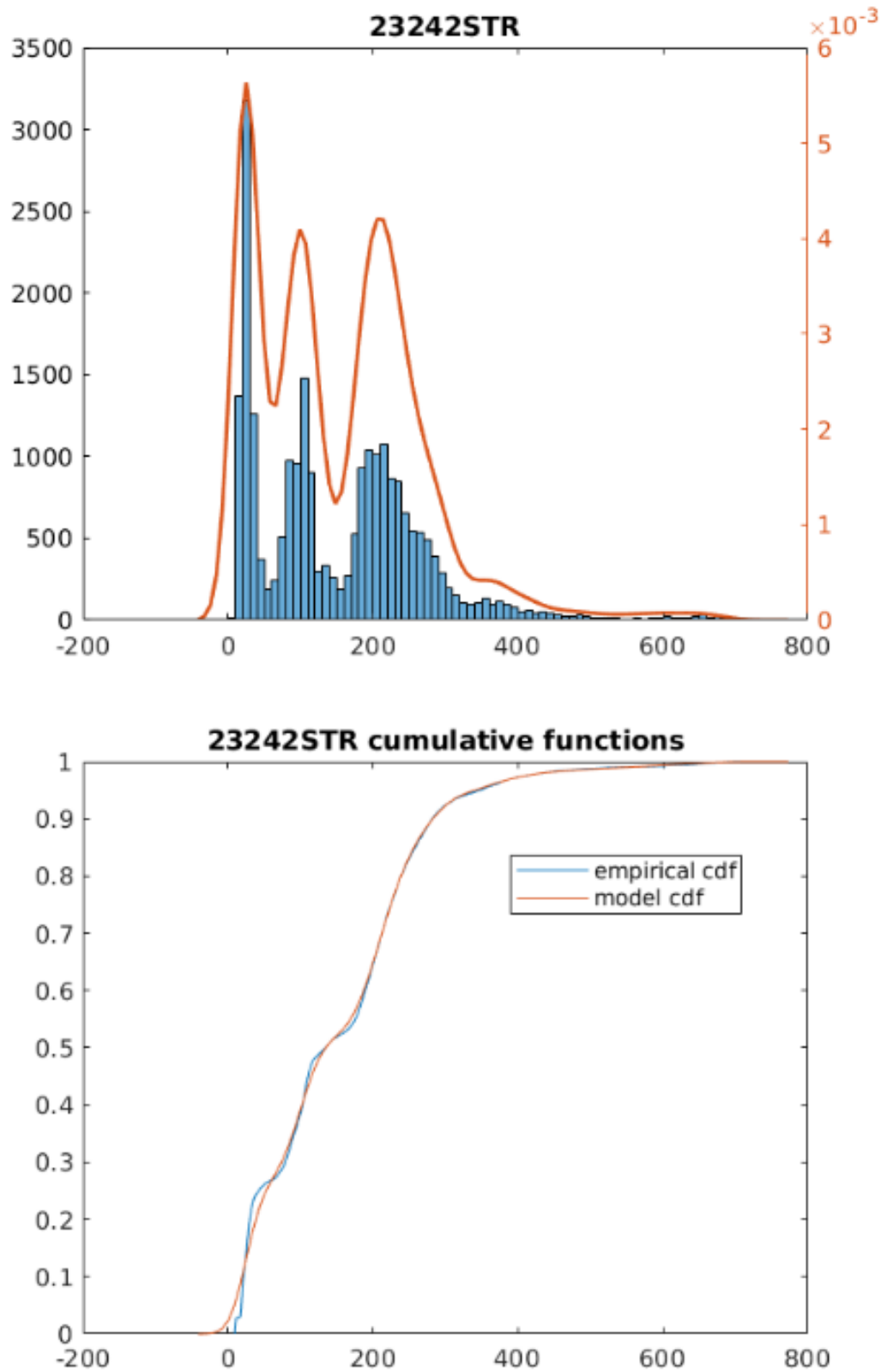
sesti todellista testidataa, muttei onnistu siinä kovin hyvin. Myös kumulatiivisista funktioista nähdään, ettei malli onnistu esittämään testidataa kovin hyvin, vaikka jossain määrin mukaileekin sitä. Lopuksi Kolmogorov-Smirnovin-testi vahvistaa, etteivät malli ja testidata noudata samaa jakaumaa. Taulukoissa 4.1 ja 4.2 on esiteltyä anturikohtaisesti, noudattavatko ydinestimointi ja testidata samaa jakaumaa, sekä käytetty ytimen leveys.



24795STR result for kolmogorov-smirnov is true

Kuva 4.8: Ydinestimointi uunin sähkösuodattimen poikittaisesta ruuvikuljettimesta, sekä analyysit estimoinnin onnistumisesta.





23242STR result for kolmogorov-smirnov is false

Kuva 4.9: Ydinestimointi sementtiuunin syötön homogenisoinnin 3. vaakaruuvikujettimesta, sekä analyysit estimoinnin onnistumisesta.

Taulukko 4.1: Ydinestimoinnit myllyjen antureista

TechIdent	Alue	Koneen nimi	Ytimen leveys	Sama jakaumaa?
28021MAT	Hiilimyly 6	Ketjussyötin	8.06258999428144	Ei
28230FIL	Hiilimyly 6	Pölysuodatin	32.3021772441528	Kyllä
28045FLÄ	Hiilimyly 6	Pölysuodatin	13.1741262559479	Kyllä
28041STR	Hiilimyly 6	Ruuvikuljetin hiilisykloonasta	12.7067673309978	Kyllä
28043STR	Hiilimyly 6	Ruuvikuljetin hiilisykloonasta	20.9375762847038	Kyllä
28062PUM	Hiilimyly 6	Ruuvipumppu	5.84488886450492	Kyllä
22336LUR	Raakamyly 7	Ilmarämi syklonien välissä	17.1061345039861	Kyllä
22201VÄX1	Raakamyly 7	Kaksiportainen hammasvaihde	3.98863868248857	Kyllä
22310ELE	Raakamyly 7	Ketjuelevaattori 1	4.00133931152311	Kyllä
22320ELE	Raakamyly 7	Ketjuelevaattori 2	5.94267767568423	Kyllä
22200OLP6	Raakamyly 7	Kiertoöljypumppu	23.9737574011115	Kyllä
22200OLP7	Raakamyly 7	Kiertoöljypumppu	12.4355118910958	Kyllä
22200OLP3	Raakamyly 7	Kiertoöljypumppu (laakeri sisään)	9.85798011646346	Kyllä
22200OLP4	Raakamyly 7	Kiertoöljypumppu (laakeri ulos)	10.7078030885322	Kyllä
22312SEP	Raakamyly 7	Separattori 1	23.9976131263544	Kyllä
22322SEP	Raakamyly 7	Separattori 2	28.8201065870837	Kyllä
22410ELE	Raakamyly 7	Siilon elevaattori 3 (raakajauho)	3.23250283126732	Kyllä
22412ROT	Raakamyly 7	Tarkastusseula pyörivä	18.0047117892711	Kyllä
22200SMP10	Raakamyly 7	Voitelupumppu hammaskehälle	55.2139791914163	Kyllä

Taulukko 4.2: Ydinestimoinnit uunin antureista

TechIdent	Alue	Koneen nimi	Ytimen leveys	Sama jakaumaa?
28619STR	Uuni	Annosteluruuvi	32.4786007907248	Kyllä
23230ELE	Uuni	Elevaattori	9.1790123183153	Kyllä
24103VÅG	Uuni	Jauhevaaka 2	7.31308093487639	Kyllä
24203VÅG	Uuni	Jauhevaaka 2 (A-puoli)	3.34440673698413	Kyllä
24608PUM2	Uuni	Kaukolämmön kiertopumppu	37.2499738609628	Kyllä
24608PUM1	Uuni	Kaukolämmön kiertopumppu	6.74449620535238	Kyllä
24606PUM1	Uuni	Kaukolämmön kiertopumppu	4.32218167340987	Kyllä
24606PUM2	Uuni	Kaukolämmön kiertopumppu	20.7205188041077	Ei
24610PUM	Uuni	Kaukolämmön kiertopumppu	21.6322863340837	Kyllä
24796STR	Uuni	Ruuvikuljetin (pitkittäinen)	8.80683838186236	Ei
24795STR	Uuni	Ruuvikuljetin (poikittainen)	6.49703469433849	Kyllä
24797STR	Uuni	Ruuvikuljetin (yli menevä)	5.62056487159011	Kyllä
24751STR	Uuni	Ruuvikuljetin jäähdytystorni	17.3639346663972	Kyllä
24791STR	Uuni	Ruuvikuljetin, sähkösuodatin	9.00957587894713	Kyllä
24793STR	Uuni	Ruuvikuljetin, sähkösuodatin	10.8903326335306	Kyllä
28620SLU	Uuni	Sulkusyötin	12.4849271768369	Kyllä
24204SLU	Uuni	Sulkusyötin (A-puoli)	8.90256537445625	Kyllä
24104SLU	Uuni	Sulkusyötin (B-puoli)	23.218336198096	Kyllä
24202SLU	Uuni	Sulkusyötin 1 (A-puoli)	2.50792795535158	Kyllä
24102SLU	Uuni	Sulkusyötin 2 (B-puoli)	2.72139025995237	Kyllä
24802FLÄ	Uuni	Sähkösuodattimen puhallin	19.4609135152673	Kyllä
23240STR	Uuni	Vaakaruuvikuljetin 1	8.5470269063274	Kyllä
23241STR	Uuni	Vaakaruuvikuljetin 2	2.91970153984205	Ei
23242STR	Uuni	Vaakaruuvikuljetin 3	16.5179507084244	Ei
23220ELE	Uuni	Varaelevaattori	0.417072658819841	Kyllä

### 4.2.3 Kahden näytteen Kolmogorov-Smirnovin-testi

Luotujen mallien oikeellisuuden todentamiseen käytettiin kahden näytteen Kolmogorov-Smirnovin-testiä; testillä todennetaan noudattavatko kaksi toisistaan riippumatonta näytettä samaa jakaumaa. Testissä molemmille joukoille lasketaan kumulatiivinen jakaumafunktio käyttäen samaa intervallia. Muodostettuja jakaumafunktioita verrataan keskenään vähentämällä jokaisesta intervallista yhden askelfunktion toisesta. Testi keskittyy näistä laskutoimituksista saatuihin suurimpiin poikkeamiin. [56]

$$D_{m,n} = \max[S_m(X) - S_n(X)], \quad (12)$$

Kaava 12 on yksihäntäisen Kolmogorov-Smirnovin-testin kaava, jossa vähennetään kahden näytteen kumulatiivisten jakaumien välit kussakin intervallien mukaan jaetussa pisteessä, jossa  $S_m(X)$  on m-kokoisen näytteen kumulatiivinen jakaumafunktio ja  $S_n(X)$  on vastaavasti toisen n-kokoisen näytteen kumulatiivinen jakaumafunktio. Lisäksi jakaumafunktioista on määritelty, että  $S_m(X) = \frac{K}{m}$  ja vastaavasti myös  $S_n(X) = \frac{K}{n}$ . K on datan määrä, joka on pienempi tai yhtä suuri kuin X. Kun  $D_{m,n}$  on ratkaistu määritetään suurin näistä eroista. Mikäli suurin ero on suurempi kuin valittu tilastollinen merkitsevyystaso, todetaan, että näytteet eivät noudata samaa jakaumaa. [56]

$$D_{m,n} = \max|S_m(X) - S_n(X)|, \quad (13)$$

Kaksihäntäinen Kolmogorov-Smirnovin-testi on kaava 13, joka eroaa yksihäntäisestä siten, että siinä tarkastellaan suurinta absoluuttista eroa jakaumien saman kohdan pisteiden välillä riippumatta niiden suunnasta. Vaikka ero laskutavassa ei ole iso, on käytetyllä menetelmällä kuitenkin vaikutusta; yksihäntäisen testin tulos tarkoittaa, että yksi satunnaisesti valituista näytteistä on suurempi kuin toisen näytteen vastaavan kohdan arvo, mutta kaksihäntäinen testi taas tarkoittaa näytteiden tulevan selvästi kahdesta eri populaatioista. Matlab tukee testin molempia versioita,

mutta oletuksena se käyttää kaksihäntäistä (engl. two-tailed[11] tai two-sided test) versiota [57]. [56]

Tilastollisen merkitsevyyden tason valitseminen on merkittävä osa testin hyödyntämistä [56]. Matlab käyttää oletuksena kahden näytteen Kolmogorov-Smirnovin testissä 5 % merkitsevyyden eroa, joten sitä käytettiin myös tässä työssä [57].

## 5 Yhteenveto

Tämän pro gradu -työn tavoitteena oli selvittää data-analytiikan hyödyntämistä enakoivassa kunnossapidossa ja sitä, miten siinä voitaisiin hyödyntää Finnsementin Paraisten sementtitehtaalla olevaa järjestelmää, joka kerää dataa valikoitujen tehtaalla olevien laitteiden aiheuttamasta tärinästä. Tutkimus osoitti, että kerätystä datasta on mahdollista rakentaa malli, joka esittää koneiden aiheuttamaa tärinää niiden luonteenomaisen käytön aikana. Poikkeamat näistä malleista voidaan tulkitta vikatilanteiksi, joiden aiheuttajaa on syytä lähteä selvittämään. Normaalitilan mallintamiseen päädyttiin menetelmänä, sillä käytettävissä ei ollut dataa siitä, milloin seuratut koneet ovat olleet jossakin vikatilassa ja millaista tärinä olisi tällaisissä tilanteissa ollut.

Kaikille käytössä olleille sensoreille ei saatu rakennettua luotettavaa mallia koulutusjoukosta, vaan jouduttiin toteamaan testijoukon noudattavan erilaista jakaumaa. Tämä saattaa johtua esimerkiksi huonolaatuisesta datasta tai siitä, että seurantajakson aikana kyseisissä koneissa on ollut piileviä vikoja, joita on mahdollisesti jopa korjattu seurantajakson aikana. Vastaavasti myös niille sensoreille, joille onnistuttiin luomaan mallit, saattaa olla käynyt niin, että koneissa on ollut piileviä vikoja koko seurantajakson ajan eivätkä niistä luodut mallit täten välttämättä vastaa niiden normaalia toimintatilaa.

Hyvin toteutettu ja mahdollisimman aikaisessa vaiheessa tehty kunnossapito on yritykselle arvokasta. Varhaisella puuttumisella vikatilanteisiin mahdollisimman ai-

kaisin ehkäistään vikojen pahentuminen – sekä parhaassa tapauksessa onnistutaan välttämään suunnittelemattomat tuotantokatkokset. Rahallisen menetyksen lisäksi suunnittelemattomat tuotantokatkokset johtavat myös menetettyyn työaikaan ja saattavat vaikuttaa myös tuotteen laatuun ja suuremmissa ongelmissa myös toimitusvarmuuteen.

Aikaisemmissa tutkimuksissa data-analytiikkaa ja tekoälymenetelmiä on käytetty laajasti kunnonvalvonnan apuna, mutta monet tutkimuksista on toteutettu vakioituissa laboratorio-olosuhteissa, joten niiden hyödyntäminen sellaisenaan reaali-maailmassa saattaa olla haasteellista, sillä todellisissa käyttöpaikoissa koneen osiin vaikuttaa myös sen lähistöllä tapahtuva muu toiminta. Tutkimusten tulokset ovat kuitenkin lupaavia ja monet tekoälymenetelmät sekä niiden yhdistelmät tuntuvat soveltuvan ennakoivaan kunnossapitoon. Tämän työn onnistuminen on kuitenkin osoitus, että menetelmät ovat myös siirrettävissä reaali-maailmaan.

Vaikka projektia voikin pitää onnistuneena, löytyi datan keräämiseen käytetystä järjestelmästä myös kehityskohteita. Kerätystä datasta tallennettiin vain neliöllinen keskiarvo, vaikka joitakin data-analytiikkamenetelmiä käytettäessä saattaisi olla hyödyllistä käyttää myös tärinän raaka-arvoja eri akseleilta. Lisäksi tietojen tallennuksessa on ollut käytössä pitkänä pidettävä puolen tunnin intervalli, vaikkakin puolen tunnin aikana koneiden käynti saattaa muuttua moneen kertaan merkittävästi. Tämä saattoi vaikuttaa datan laatuun ja useissa muissa tutkimuksissa tietojen tallentamisessa onkin käytetty huomattavasti lyhyempiä välejä. Luonnollisesti tällä on myös omat haittapuolensa, esimerkiksi lisääntyneen virrankulutuksen muodossa. Näiden muutosten toteuttaminen avaisi mielenkiintoisia jatkojalostusmahdollisuuksia nyt tehdylle työlle. Tutkielmaa tehdessä havaittiin käytössä olleesta datasta myös, että tiedon tallentamisen yhteydessä tallennettu aikaleima ei välttämättä vastannut todellista ajankohtaa, jolloin tieto oli sensorilta todellisuudessa tallennettu, vaan kertoi ajankohdan, jolloin tieto oli siirretty yhdyskäytävälaitteelta tietokan-

taan, ja tässä saattoi olla huomattavakin ero tallennetun tiedon ja todellisuuden välillä.

Jatkossa olisi mielenkiintoista tutkia muiden järjestelmien keräämän datan hyödyntämistä kunnonvalvonnassa ja selvittää sen avaamia mahdollisuuksia. Muista järjestelmistä voisi saada esimerkiksi muita hyödynnettäviä vasteita nyt kerätyn tärinädatan lisäksi tai mahdollisesti nykyistä dataa voisi annotoida vikaantumista kertovien raporttien avulla. Lisäksi tehtaalle voisi asentaa myös muun tyyppisiä antureita, esimerkiksi akustisten päästöjen seuraaminen oli käytössä monessa aikaisemmassa tutkimuksessa. Suuremman datamäärän hyödyntäminen tutkimuksessa saattaisi avata mahdollisuuksia käyttää myös muita data-analytiikkamenetelmiä kuin nyt käytettyä ydinestimointia.



# Lähdeluettelo

- [1] *Finnsementin historia*, <https://finnsementti.fi/yritys/historia/>, tarkastettu 22.11.2021, 2021.
- [2] *Lukuja*, <https://finnsementti.fi/yritys/lukuja/>, tarkastettu 22.11.2021, 2021.
- [3] S. Nandi, H. A. Toliyat ja X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review", *IEEE transactions on energy conversion*, vol. 20, nro 4, s. 719–729, 2005.
- [4] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto ja S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance", *Computers Industrial Engineering*, vol. 137, s. 106 024, 2019, ISSN: 0360-8352. DOI: <https://doi.org/10.1016/j.cie.2019.106024>. url: <https://www.sciencedirect.com/science/article/pii/S0360835219304838>.
- [5] R. B. Randall, *Vibration-based condition monitoring: industrial, automotive and aerospace applications*. John Wiley & Sons, 2021.
- [6] A. R., *Kunnossapito menestystekijänä*. Opetushallitus, 2010.
- [7] D. Mba ja R. Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: bearings, pumps, gearboxes, engines, and rotating structures", *Shock and Vibration Digest*, vol. 38, nro 1, s. 3–18, 2006.

- [8] Y. H. Ali, "Artificial Intelligence Application in Machine Condition Monitoring and Fault Diagnosis", teoksessa *Artificial Intelligence*, M. A. Aceves-Fernandez, toim., Rijeka: IntechOpen, 2018, luku 14. DOI: 10.5772/intechopen.74932. url: <https://doi.org/10.5772/intechopen.74932>.
- [9] A. Saxena ja A. Saad, "Genetic algorithms for artificial neural net-based condition monitoring system design for rotating mechanical systems", teoksessa *Applied Soft Computing Technologies: The Challenge of Complexity*, Springer, 2006, s. 135–149.
- [10] A. Widodo, B.-S. Yang, E. Y. Kim, A. C. Tan ja J. Mathew, "Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine", *Nondestructive Testing and Evaluation*, vol. 24, nro 4, s. 313–328, 2009.
- [11] C.-T. Lin, C. S. G. Lee et al., "Neural-network-based fuzzy logic control and decision system", *IEEE Transactions on computers*, vol. 40, nro 12, s. 1320–1336, 1991.
- [12] *Artificial Intelligence - Fuzzy Logic Systems*. url: [https://www.tutorialspoint.com/artificial\\_intelligence/artificial\\_intelligence\\_fuzzy\\_logic\\_systems.htm](https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_fuzzy_logic_systems.htm).
- [13] Q. Ren, L. Baron ja M. Balazinski, "Type-2 fuzzy modeling for acoustic emission signal in precision manufacturing", *Modelling and Simulation in Engineering*, vol. 2011, 2011.
- [14] J. Rafiee, F. Arvani, A. Harifi ja M. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network", *Mechanical systems and signal processing*, vol. 21, nro 4, s. 1746–1754, 2007.
- [15] N. Kriegeskorte ja T. Golan, "Neural network models and deep learning", *Current Biology*, vol. 29, nro 7, R231–R236, 2019.

- [16] A. A. Heidari, H. Faris, I. Aljarah ja S. Mirjalili, "An efficient hybrid multilayer perceptron neural network with grasshopper optimization", *Soft Computing*, vol. 23, nro 17, s. 7941–7958, 2019.
- [17] M. Bhuvaneshwari, E. G. M. Kanaga, J. Anitha, K. Raimond ja S. T. George, "Chapter 7 - A comprehensive review on deep learning techniques for a BCI-based communication system", teoksessa *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics*, P. N, S. Kautish ja S.-L. Peng, toim., Academic Press, 2021, s. 131–157, ISBN: 978-0-12-821633-0. DOI: <https://doi.org/10.1016/B978-0-12-821633-0.00013-1>. url: <https://www.sciencedirect.com/science/article/pii/B9780128216330000131>.
- [18] M. Roy, S. K. Bose, B. Kar, P. K. Gopalakrishnan ja A. Basu, "A stacked autoencoder neural network based automated feature extraction method for anomaly detection in on-line condition monitoring", teoksessa *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2018, s. 1501–1507.
- [19] V. Jagannath, *Diagram of random decision forest*. maaliskuu 2017. url: <https://community.tibco.com/wiki/random-forest-template-tibco-spotfire>.
- [20] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés ja C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest", *Mechanical Systems and Signal Processing*, vol. 70, s. 87–103, 2016.
- [21] M. M. Ettefagh, M. Ghaemi ja M. Y. Asr, "Bearing fault diagnosis using hybrid genetic algorithm K-means clustering", teoksessa *2014 IEEE international symposium on innovations in intelligent systems and applications (INISTA) proceedings*, IEEE, 2014, s. 84–89.

- [22] K. Kudelina, B. Asad, T. Vaimann, A. Rassölkin, A. Kallaste ja H. V. Khang, ”Methods of Condition Monitoring and Fault Detection for Electrical Machines”, *Energies*, vol. 14, nro 22, s. 7459, 2021.
- [23] Z. Nie, E. Guo, J. Li, H. Hao, H. Ma ja H. Jiang, ”Bridge condition monitoring using fixed moving principal component analysis”, *Structural Control and Health Monitoring*, vol. 27, nro 6, e2535, 2020.
- [24] G. G. Yen ja K.-C. Lin, ”Wavelet packet feature extraction for vibration monitoring”, *IEEE transactions on industrial electronics*, vol. 47, nro 3, s. 650–667, 2000.
- [25] *Koneiden käyttöikä pitenee vuosikymmenillä IoT-seurannan ansiosta*, <https://www.fidera.fi/kayttoika-pitenee-vuosikymmenilla-iot-seurannan-ansiosta/>, tarkastettu 23.11.2021, 2020.
- [26] *#iotnextlevel - Kilpailun esittely*, <https://www.dna.fi/yrityksille/iotnextlevel>, tarkastettu 23.11.2021, 2018.
- [27] C. Gomez, J. Oller ja J. Paradells, ”Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology”, *Sensors*, vol. 12, nro 9, s. 11 734–11 753, 2012.
- [28] *Finnsementti condition monitoring system overview*, Sisäinen dokumentaatio, 2020.
- [29] T. A. CONTAIN, ”Real-time travel time estimates using media access control address matching”, *ITE journal*, 2008.
- [30] E. Rescorla ja T. Dierks, ”The transport layer security (TLS) protocol version 1.3”, 2018.
- [31] *RuuviTag tuotesivu*, <https://ruuvi.com/fi/ruuvitag/>, tarkastettu 23.11.2021, 2021.
- [32] *About*, <https://www.postgresql.org/about/>, tarkastettu 23.11.2021, 2021.

- [33] *Yritysesitys*, tarkastettu 30.11.2021, 2021.
- [34] *Sementin valmistus*, <https://finnsementti.fi/palvelut/tietoa-sementista/valmistus/>, tarkastettu 29.11.2021, 2021.
- [35] M. Öhman, ”Finnsementti Sensoring: kunnossapidon kehitysprojekti”, 2020.
- [36] IBM, ”IBM SPSS Modeler CRISP-DM Guide”, 1994, 2016.
- [37] R. Wirth ja J. Hipp, ”CRISP-DM: Towards a standard process model for data mining”, teoksessa *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Springer-Verlag London, UK, vol. 1, 2000, s. 29–39.
- [38] *2.4.1 Pylväskuvio*. url: [https://tilastokoulu.stat.fi/verkkokoulu\\_v2.xql?course\\_id=tkoulu\\_tilaj&lesson\\_id=2&subject\\_id=19&page\\_type=sisalto](https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=2&subject_id=19&page_type=sisalto).
- [39] F. J. Massey Jr, ”The Kolmogorov-Smirnov test for goodness of fit”, *Journal of the American statistical Association*, vol. 46, nro 253, s. 68–78, 1951.
- [40] R. Raoniari, *Finding the Best Distribution that Fits Your Data using Python’s Fitter Library*, kesäkuu 2021. url: <https://medium.com/the-researchers-guide/finding-the-best-distribution-that-fits-your-data-using-pythons-fitter-library-319a5a0972e9>.
- [41] S. Kotz, T. Kozubowski ja K. Podgórski, *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*, 183. Springer Science & Business Media, 2001.
- [42] J. K. Patel ja C. B. Read, *Handbook of the normal distribution*. CRC Press, 1996, vol. 150.
- [43] O. Ibe, *Fundamentals of applied probability and random processes*. Academic Press, 2014.

- 
- [44] *Uniform Distribution (Continuous)*. url: <https://se.mathworks.com/help/stats/uniform-distribution-continuous.html>.
- [45] D. Kundu ja M. Z. Raqab, "Generalized Rayleigh distribution: different methods of estimations", *Computational statistics & data analysis*, vol. 49, nro 1, s. 187–200, 2005.
- [46] *scipy.stats.lognorm*. url: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.lognorm.html>.
- [47] *scipy.stats.gamma*. url: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gamma.html>.
- [48] *scipy.stats.cauchy*. url: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.cauchy.html>.
- [49] *scipy.stats.alpha*. url: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.alpha.html>.
- [50] S. Węglarczyk, "Kernel density estimation and its application", teoksessa *ITM Web of Conferences*, EDP Sciences, vol. 23, 2018.
- [51] T. Ledl, "Kernel density estimation: theory and application in discriminant analysis", *Austrian journal of statistics*, vol. 33, nro 3, s. 267–279, 2004.
- [52] V. A. Epanechnikov, "Non-Parametric Estimation of a Multivariate Probability Density", *Theory of Probability & Its Applications*, vol. 14, nro 1, s. 153–158, 1969. DOI: 10.1137/1114019. eprint: <https://doi.org/10.1137/1114019>. url: <https://doi.org/10.1137/1114019>.
- [53] A. Z. Zambom ja D. Ronaldo, "A review of kernel density estimation with applications to econometrics", *International Econometric Review*, vol. 5, nro 1, s. 20–42, 2013.
- [54] B. A. Turlach, "Bandwidth selection in kernel density estimation: A review", teoksessa *CORE and Institut de Statistique*, Citeseer, 1993.

- 
- [55] *Fit Kernel Distribution Using ksdensity*. url: <https://se.mathworks.com/help/stats/generate-a-kernel-probability-density-estimate-using-ksdensity.html>.
- [56] P.-C. Lin, B. Wu ja J. Watada, "Kolmogorov-Smirnov two sample test with continuous fuzzy data", teoksessa *Integrated Uncertainty Management and Applications*, Springer, 2010, s. 175–186.
- [57] *kstest2*. url: <https://se.mathworks.com/help/stats/kstest2.html>.