

# **Kaukolämpötuotannon optimointi, raportointi ja visualisointi SaaS-palveluna**

**Case: Kemin Energia ja Vesi Oy**

LAB-ammattikorkeakoulu

Insinööri (ylempi AMK), Digitaaliset ratkaisut

2021

Jarmo Erola

Antti Lehikoinen

## Tiivistelmä

Tekijä(t) Erola, Jarmo Lehikoinen, Antti	Julkaisun laji Opinnäytetyö, YAMK Sivumäärä 94	Valmistumisaika 2021
Työn nimi <b>Kaukolämpötuotannon optimointi, raportointi ja visualisointi SaaS-palveluna</b> Case: Kemin Energia ja Vesi Oy		
Tutkinto Insinööri (ylempi AMK), Digitaaliset ratkaisut		
Toimeksiantajan nimi, titteli ja organisaatio Jani Peurasaari, kaukolämpöjohtaja, Kemin Energia ja Vesi Oy		
Tiivistelmä <p>Kaukolämpötuotanto on yleistynyt viimeisen kolmen vuosikymmenen aikana Suomessa merkittävästi. Samaan aikaan päästötavoitteet ovat jatkuvasti kiristyneet, minkä myötä kaukolämpötuotannolta edellytetään entistä ekologisempia sekä energia- ja kustannustehokkaampia ratkaisuja. Matemaattinen optimointi ja koneoppiminen tarjoavat työkaluja tuotannon optimointiin, jonka tuloksia voidaan tuoda päätöksenteon tueksi datan visualisointimenetelmillä.</p> <p>Työn lähtökohtana oli aiemmin asiakkaalle toteutetusta raportointijärjestelmästä nousevat jatkokehitystarpeet. Tarkoituksena oli luoda kaukolämpötuotannon optimointimalli, joka tuottaisi suosituksia, miten tuotantoa voitaisiin ajaa taloudellisemmin ja ekologisemmin. Tuotannon optimoinnin tulokset saatettiin käytännön työn tueksi raportointi- ja visualisointiratkaisuilla.</p> <p>Opinnäytetyön tutkimuksellinen osuus toteutettiin kirjallisuuskatsauksena ja käytännön toteutuksessa hyödynnettiin suunnittelutieteen käytäntöjä. Työ tehtiin asiakaslähteisesti ja kehityksen suuntaviivoja määriteltiin yhdessä asiakkaan kanssa, käytännön työstä nousseiden tarpeiden pohjalta. Yhteisen suunnittelun pohjalta pyrittiin hahmottamaan, minkälaisista tiedoista ja visualisoinneista olisi mahdollisimman suuri hyöty asiakkaan käytännön työn helpottamiseksi ja sitä mukaa kaukolämpötuotannon tehostamiseksi.</p> <p>Optimointiratkaisujen osalta havaittiin, että sekä matemaattiset menetelmät että koneoppimISRatkaisut ovat käyttökelpoisia kaukolämpötuotannon optimoinnin kannalta. Tuotanto- ja optimointidatan visualisoinnin todettiin tuovan tukea kaukolämpötuotannon päätöksentekoon ja tuotannon ennakointiin.</p>		
Asiasanat kaukolämpö, optimointi, koneoppiminen, raportointi, visualisointi		

## Abstract

Author(s)	Type of Publication	Published
Erola, Jarmo	Master's Thesis	2021
Lehikoinen, Antti	Number of Pages	
	94	
Title of Publication		
<b>District heating optimization, reporting and visualization delivered as a service</b>		
Case: Kemin Energia ja Vesi Oy		
Name of Degree		
Master of Engineering, Digital Solutions		
Name, title and organization of the client		
Jani Peurasaari, District Heating Director, Kemin Energia ja Vesi Oy		
Abstract		
<p>District heating has become significantly more general in Finland over the past three decades. At the same time emission targets have been steadily tightened which means that more environmentally friendly, energy- and cost-effective solutions are required for district heating production. Mathematical optimization and machine learning provide tools for optimizing production and generate results which can support decision-making via different data visualization methods.</p> <p>The starting point of the thesis was the need for further development concerning a previously implemented reporting system for the customer. The aim was to create an optimization model for district heating production that would produce recommendations on how to run production more economically and ecologically. The outcome of the optimization was then visualized and added to the existing reporting solutions.</p> <p>The research part of the thesis was carried out as a literature review and the practical implementation utilized the practices of design science. Research was carried out using customer-oriented methods, and the development guidelines were defined together with the customer based on their needs. With careful planning, the aim was to outline what kind of information and visualizations would benefit the most to facilitate the customer's needs and consequently increase the efficiency of district heating production.</p> <p>Mathematical methods and machine learning solutions were found to be useful for optimizing district heating production. Visualization of production and optimization data was found to provide support for district heating production decision-making and production forecasting.</p>		
Keywords		
district heating, optimization, machine learning, reporting, visualization		

## Sisällys

1	Johdanto.....	1
1.1	Työn tausta ja tavoitteet.....	1
1.2	Työn toteutus, tutkimuskysymykset ja rajaukset .....	1
2	Kaukolämmön tuotanto.....	3
2.1	Yleistä.....	3
2.2	Kaukolämmöntuotannon kannattavuus .....	4
2.2.1	Polttoaineen hinta ja päästökauppa .....	5
2.2.2	Kannattavuuden laskenta .....	6
3	Kemin Energia ja Vesi Oy .....	9
3.1	Yhtiön historiaa.....	9
3.2	Yhtiön toiminta.....	9
3.3	Kemin Energia ja Vesi Oy:n kaukolämmön tuotanto .....	10
3.4	Automaatiojärjestelmä ja raportointi.....	11
4	Matemaattinen optimointi.....	14
4.1	Johdanto matemaattiseen optimointiin.....	14
4.1.1	Kohdefunktio.....	14
4.1.2	Päätösmuuttujat.....	16
4.1.3	Rajoitteet .....	16
4.1.4	Mallintaminen .....	18
4.1.5	Konveksisuus .....	19
4.2	Geneettiset algoritmit.....	20
4.3	Gradienttilaskumenetelmä .....	24
4.4	Pienimmän neliösumman menetelmä .....	25
5	Koneoppiminen.....	28
5.1	Johdanto koneoppimiseen .....	28
5.2	Ohjattu ja ohjaamaton oppiminen .....	28
5.3	Yhdistelmäoppiminen .....	30
5.3.1	Bootstrap ja bagging.....	30
5.3.2	Boosting .....	30
5.4	Koneoppimisen vaiheet .....	31
5.4.1	Datan kerääminen ja käsittely.....	31
5.4.2	Muuttujien valinta.....	32
5.4.3	Mallin luonti .....	33
5.4.4	Validointi.....	34

5.5	Regressiomenetelmät.....	35
6	Kaukolämmön tuotannon kannattavuuden optimointi.....	37
6.1	Johdanto optimointitehtävään.....	37
6.2	Konsepti ongelmanratkaisusta.....	38
6.3	Optimaalisen tuotantotavan selvittäminen lineaarisella optimoinnilla.....	39
6.4	Optimaalisen tuotantotavan selvittäminen geneettisten algoritmien avulla.....	41
6.5	Koneoppimisen soveltaminen optimaalisen tuotantotavan selvittämisessä.....	48
6.5.1	Data ja sen käsittely.....	48
6.5.2	Mallin luonti polynomiregressiolla.....	51
6.5.3	Mallin luonti satunnaismetsäregressiolla.....	52
6.5.4	Mallin luonti gradienttitehostamisregressiolla.....	53
7	Raportointi- ja visualisointityökalut.....	55
7.1	Microsoft Power BI.....	55
7.2	Muut visualisointityökalut.....	56
8	Tuotanto- ja optimointidatan käsittely visualisointia varten.....	57
8.1	Datan puhdistaminen ja muotoilu.....	57
8.2	Tiedon väliset suhteet ja linkitys.....	59
9	Datan visualisointimenetelmät ja visualisoinnin tavoitteet.....	61
9.1	Datan visualisointi.....	61
9.2	Visualisointitavat ja värien käyttö.....	62
9.3	Yleiset visualisoinnin tavoitteet.....	64
9.4	Tuotantodatan visualisoinnin tavoitteet.....	64
9.5	Optimointi- ja kannattavuusdatan visualisoinnin tavoitteet.....	65
10	Tuotantodatan visualisointi.....	66
10.1	Reaaliaikainen tuotantonäkymä.....	66
10.2	Optimointidatan visualisointi.....	69
10.2.1	Tulojen ja menojen visualisointi.....	70
10.2.2	Kannattavuuden visualisointi.....	71
10.2.3	Lämmöntuotannon vertailu optimaaliseen tuotantoratkaisuun.....	72
10.2.4	Tuotannon taloudellisuus suhteessa optimaaliseen tuotantotapaan.....	75
10.3	Taulukkopohjainen raportointinäkymä.....	76
10.4	Raporttinäkymien käyttö mobiililaitteilla.....	78
11	Johtopäätökset ja pohdintaa.....	80
11.1	Käytetyn tutkimusmenetelmän soveltuvuus.....	80
11.2	Optimaalisen kaukolämmön tuotantokombinaation selvittäminen.....	80

11.3	Koneoppimisen menetelmien soveltuminen optimaalisen kaukolämmön tuotantokombinaation selvittämisessä .....	81
11.4	Tiedon reaaliaikaisella visualisoinnilla saavutetut hyödyt .....	83
11.5	Haasteet .....	83
11.6	Jatkokehitys .....	85
11.6.1	Kannattavuuden laskenta ja optimointi .....	85
11.6.2	Optimointi tekoälyn menetelmillä .....	85
11.6.3	Matemaattisen optimoinnin ja koneoppimisen yhdistäminen .....	86
11.6.4	Optimointi- ja tuotantodatan visualisointi .....	86
11.6.5	Ennustedatan ja päästölaskentojen yhdistäminen .....	87
11.6.6	Tuotannon tarpeen ennustamisen jatkokehitys .....	87
11.6.7	Polttoaine-erien riittävyyden ennustaminen .....	88
11.6.8	Ennakoiva laitteiden kunnan seuranta .....	88
12	Yhteenveto .....	90
	Lähteet .....	91

# 1 Johdanto

## 1.1 Työn tausta ja tavoitteet

Tässä LAB-ammattikorkeakoulun ylemmän ammattikorkeakoulun opinnäytetyössä käsitellään kaukolämpötuotannon optimointia matemaattisen optimoinnin ja koneoppimisen avulla sekä selvitetään tapoja visualisoida tuotantodataa kaukolämpötuotannon päivittäisen päätöksenteon avuksi. Kaukolämpötuotannon optimoinnilla on suuri merkitys kaukolämpötuotannon ympäristöystävällisyyteen ja taloudelliseen kannattavuuteen ja visualisointi puolestaan tukee optimoinnin tulosten saattamista käytäntöön. Työssä luodaan pohjaa kokonaistaloudellisen mallin kehittämiseksi, joka käsittäisi kaukolämpötuotannon tuotantotavat, tuotannon tarpeen ennusteen sekä eri tuotantomuotojen kustannukset ja tuotot. Kokonaistaloudellisen mallin avulla voitaisiin havainnollistaa eri tekijöiden vaikutuksia toisiinsa ja siten tuottaa tietoa tuotannon ohjaamisen avuksi.

Kaukolämpötuotannon optimoinnin ja visualisoinnin lisäksi työssä käsitellään matemaattisia optimointimenetelmiä sekä koneoppimisen menetelmien soveltuvuutta optimointitehtävien ratkaisemiseen. Tänä päivänä datalla johtaminen korostuu, kun dataa hyödynnetään prosessien ja liiketoiminnan kehittämisessä sekä päätöksenteossa. Koneoppimisen soveltuvuudesta optimointitehtävissä on vain vähän aikaisempaa tutkimusta, mutta samanaikaisesti koneoppimisen monipuolinen hyödyntäminen yritysten toiminnoissa lisääntyy. Näin ollen on perusteltua selvittää, soveltuvatko koneoppimisen menetelmät myös optimointitehtävien ratkaisemiseen.

Lähtökohtana opinnäytetyössä on aiemmin Kemin Energia ja Vesi Oy:lle toteutettu raportointijärjestelmä ja sen jatkokehitystarpeet. Opinnäytetyö on tehty yhteistyössä asiakkaan kanssa, joka osallistui jatkokehitystarpeiden määrittelyyn käytännön työstä nousseiden tarpeiden kautta. Yhteisen suunnittelun pohjalta määritettiin mistä tiedoista asiakkaalle olisi mahdollisimman suuri hyöty käytännön työn helpottamiseksi ja sitä mukaa kaukolämpötuotannon tehostamiseksi.

## 1.2 Työn toteutus, tutkimuskysymykset ja rajaukset

Tutkimus toteutetaan kirjallisuuskatsauksena sekä suunnittelutieteen (Design science) periaatteita noudattaen. Kirjallisuuskatsausta tehdään keskeisimmistä aiheista, kuten matemaattisesta optimoinnista, koneoppimisesta sekä visualisoinnista. Kirjallisuuskatsauksen avulla varmistetaan riittävä tietoperusta suunnittelutieteen mukaisesti toteutettaville ensimmäisille karkeille ratkaisuehdotuksille.

Suunnittelutiede on iteratiivinen prosessi, jonka avulla pyritään luomaan uusia ja innovatiivisia artefakteja. Artefaktit voivat olla muun muassa algoritmeja tai menetelmiä, jotka ratkaisevat olemassa olevan ongelman. Sen periaatteisiin kuuluu ongelman syvällinen ymmärtäminen ja sen pohjalta tutkimuskysymysten asettaminen. Suunnittelutieteelle on ominaista tehdä karkea suunnitelma ongelman ratkaisusta ja kehittää sitä suunnitteluvaiheen aikana. Kehitetyn ratkaisun tuloksia arvioidaan ja tarvittaessa sitä jatkokehitetään parempien tulosten saamiseksi. Arvioinnin pohjalta ratkaisua voidaan edelleen kehittää ja tuottaa entistä parempia tuloksia. Lopullisia tuloksia arvioidaan tavoitteisiin nähden ja esitellään johtopäätökset. (Männistö 2018.) Tässä työssä optimointimalli ja visualisointi ovat artefakteja, jotka toteutetaan iteratiivisesti suunnittelutieteen periaatteita noudattaen.

### Tutkimuskysymykset

1. Miten matemaattisen optimoinnin menetelmillä voidaan selvittää kaukolämmön tuotannon kannattavin tuotantokombinaatio?
2. Miten koneoppimisen menetelmät soveltuvat kaukolämmön tuotannon optimointiin?
3. Millä tavoilla kaukolämpötuotannon mittaus- ja ennustedataa sekä talousdataa yhdistämällä ja visualisoimalla voidaan tuottaa lisätietoa tuotannon päätöksenteon tueksi?

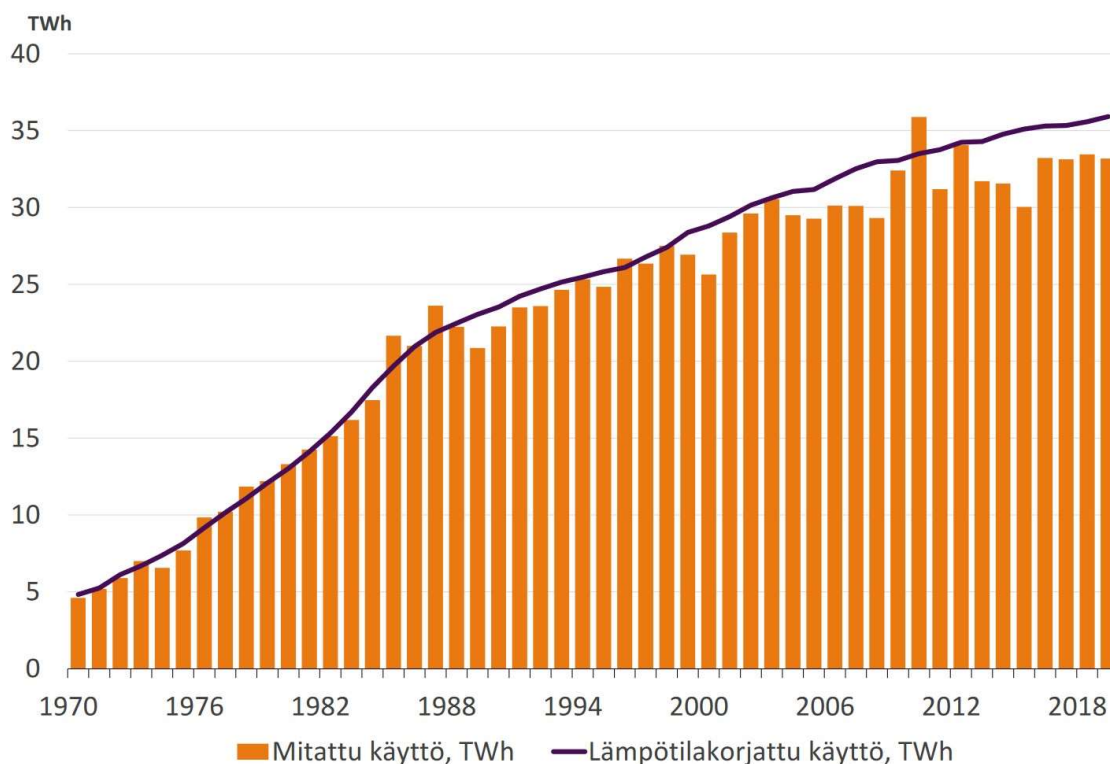
Tämän työn käytännön osuudesta on matemaattisen optimoinnin ja koneoppimisen osalta rajattu pois Kemin Energia ja Vesi Oy:n kiinteän polttoaineen KPA2-laitoksen käsittely, sillä sen toiminnasta ei ollut dataa saatavilla vielä työn kirjoitusvaiheessa. Koneoppimisen osalta käsitellään vain regressiomenetelmiä, jotka liittyvät oleellisesti tämän työn tutkimusongelmiin. Optimointiratkaisuja ja niiden tulosten visualisointia käsiteltäessä ei tietosuojasyistä tuoda esille laskentojen taustalla olevia talouslukuja.



## 2 Kaukolämmön tuotanto

### 2.1 Yleistä

Kaukolämmöllä tarkoitetaan erillisissä lämpölaitoksissa tuotettua lämpöenergiaa tai sähkön ja lämmön yhteistuotantolaitoksissa talteen otettua sähkötuotannon turbiinien hukkalämpöenergiaa. Kaukolämpöenergia jaellaan asiakkaille kaukolämpöverkostoa pitkin, jossa energian siirtämisen väliaineena toimii kuuma vesi (Mäkelä & Tuunanen 2015, 11). Vuoden 2019 lopussa kaukolämpöä jaeltiin Suomessa yhteensä 174 kunnan alueella ja kaukolämpöverkostojen yhteispituus oli 15430 kilometriä (Energiateollisuus 2020a, 2–3). Kaukolämpö on yleisin lämmitysmuoto kaupungeissa ja taajamissa (Energiateollisuus 2021). Yhteensä kaukolämmöllä lämmitetyissä kiinteistöissä asui noin 2,9 miljoonaa suomalaista vuoden 2019 lopussa. Kaukolämmön asiakkaiden määrä sekä kaukolämpöverkon pituus ovat kasvaneet tasaisesti viimeisen viiden vuosikymmenen aikana. (Energiateollisuus 2020a, 5–6.) Kuvio 1 esittää kaukolämpötuotannon kasvua Suomessa vuosien 1970 ja 2019 väliseltä ajanjaksolta, jonka aikana tuotanto on lähes 7-kertaistunut.



Kuvio 1. Kaukolämmön tuotannon kasvu Suomessa (Energiateollisuus 2020a)

Kaukolämpötuotannon polttoaineina käytetään paikkakunta- ja tuotantolaitoskohtaisesti puuta tai jotakin muuta biomassaa, kivihiiltä, maakaasua, turvetta, öljyä tai poltettavaksi sopivaa yhdyskuntajätettä. Käytettävän polttoaineen valintaan vaikuttavia tekijöitä ovat polttoaineen saatavuus ja toimintavarmuus, sekä kokonaistaloudellisuus ja vaikutukset ympäristöön (Energiateollisuus 2020b). Kaukolämmön tuotannossa on yleisesti pyritty lisäämään kotimaisten uusiutuvien polttoaineiden käyttöä ja vuonna 2019 uusiutuvilla polttoaineilla tuotetun kaukolämmön määrä ylitti ensimmäistä kertaa fossiilisilla polttoaineilla tuotetun määrän (Tilastokeskus 2019). Vuonna 2019 kaukolämpöä tuotettiin polttoaineilla yhteensä noin 33 terawattituntia, josta uusiutuvien osuus oli 15,3 TWh ja fossiilisilla polttoaineilla tuotettu osuus 13,2 TWh. Vielä vuonna 2019 ei-fossiiliseksi luokitellun turpeen osuus tuotannosta oli 5,7 TWh ja muilla energianlähteillä tuotettiin 4,0 TWh. Kokonaistuotannosta noin 10 prosenttia tuotettiin lämmön talteenotolla hukkalämmöstä, hyödyntäen savukaasupesureita ja lämpöpumppuja (Energiateollisuus 2020a, 3). Polttoaineiden valinnalla voidaan vaikuttaa hiilidioksidipäästöihin ja sitä kautta ilmastonmuutokseen. Energiantuotantoa optimoimalla pystytään tuottamaan tarvittava määrä energiaa välttäen turhan energian muostamista.

Vertailtaessa muihin lämmöntuotantomenetelmiin kaukolämmityksen suurimmiksi eduiksi voidaan lukea energiatehokkuus, ympäristöystävällisyys, kokonaistaloudellisuus ja toimintavarmuus. Erityisesti yhdistetyn sähkön- ja lämmöntuotannon energiatehokkuus sekä biopolttoaineiden osuuden kasvu lisäävät kaukolämmön mielekkyyttä lämpöenergian lähteenä (Mäkelä & Tuunanen 2015, 12). Kaukolämmön eduksi voidaan lukea myös se, että polttoaineena voidaan joustavasti käyttää erilaisia, lähialueilta helposti saatavilla olevia materiaaleja, jolloin lämmöntuotanto ei ole riippuvainen kauempaa kuljetettavista, tuotantomuodosta riippuen harvinaisistakin polttoaineista. Lisäksi kaukolämmön käyttöpaikoille asennettavia laitteistojen huolto- ja korjaustarve on kohtalaisen vähäinen (Motiva 2019).

## 2.2 Kaukolämmöntuotannon kannattavuus

Energiantuotannon kannattavuuteen vaikuttaa moni asia ja esimerkiksi jo energiantuotannon kustannustenkin tarkastelu on monimutkainen ongelma. Kustannusten yksityiskohtainen tarkastelu edellyttäisi voimalaitoksen hyötysuhteeseen vaikuttavien tekijöiden teknistä selvittämistä. Myös kausivaihtelut vuodenaikojen mukaan vaikuttavat kaukolämmön kysyntään ja näin kaukolämmöstä saataviin tuloihin: Kaukolämpöä käytetään talvikaudella huomattavasti enemmän kuin kesäkaudella ja näin valtaosa vuotuisesta kokonaiskulutuksesta kohdistuu vain muutamaan kuukauteen. Myös polttoaineiden päästöoikeuksien hinta vaihtelee kausittain. (Neittaanmäki & Schroderus 2010, 9.) Seuraavissa alaluvuissa käsitellään

tarkemmin kaukolämmön kannattavuuteen vaikuttavia tekijöitä, kuten polttoaineen hinnan muodostumista ja kannattavuuteen vaikuttavia muita kustannuksia.

### 2.2.1 Polttoaineen hinta ja päästökauppa

Polttoainekustannukset ovat kaukolämmöntuotannon kustannuksista merkittävimpiä (Neittaanmäki & Schroderus 2010, 4). Fossiilisia polttoaineita käyttäville energiantuotantolaitoksille muodostuu lisäkustannuksia, jotka pakottavat vähäpäästöisempien energiantuotantomuotojen käyttämiseen. EU:n asettamia päästötavoitteita seurataan päästökaupan avulla, jonka tarkoituksena on vaikuttaa kasvihuonepäästöjen muodostumiseen EU:n alueella. Energiantuotantolaitoksilla on oltava voimassa oleva päästölupa, jonka puitteissa laitoksilla on velvoite raportoida päästöt viranomaisille. Laitokset hankkivat fossiilisilla polttoaineilla aiheutettujen päästöjen kattamiseen päästöoikeuksia, joita tarvitaan jokaista hiilidioksiditonnia kohden. Päästöoikeuksia laitokset voivat saada ilmaiseksi, mutta niitä voi myös hankkia huutokaupasta. Päästöoikeuksilla voi lisäksi tehdä kauppaa EU:n alueella toimivissa pörssiissä, joissa esimerkiksi ylimääräiset päästöoikeudet voi myydä. Päästöoikeuksien huutokaupattava määrä vähenee vuosittain, jonka myötä päästöt vähenevät ja päästöoikeudet kallistuvat. Näin ollen vähäpäästöinen energiantuotanto muodostuu kannattavamaksi kuin esimerkiksi kivihiiilen tai turpeen käyttö. (Itkonen 2020.)

Polttoaineen hinta muodostuu tuottajahinnasta sekä kuljetus- ja päästöoikeusmaksusta. (Neittaanmäki & Schroderus 2010, 19). Polttoainemuodoista päästökaupan, eli päästöoikeusmaksujen, piirissä ovat öljy, kivihiihi, maakaasu ja energiaturve (Makkonen 2014, 25–27). Päästökaupasta vapautettuja polttoaineita ovat metsäenergia, tuulivoima, vesivoima ja ydinvoima (Makkonen 2014, 27–29). Polttoaineen hiilidioksidipäästöt vaikuttavat päästökaupan kautta polttoaineen hintaan ja kannattavuuteen. Eniten hiilidioksidipäästöjä aiheuttavat turve (382 KgCO<sub>2</sub>/MWh) ja kivihiihi (341 KgCO<sub>2</sub>/MWh). Raskas polttoöljy aiheuttaa hiilidioksidipäästöjä 279 KgCO<sub>2</sub>/MWh, kevyt polttoöljy 267 KgCO<sub>2</sub>/MWh ja maakaasu 202 KgCO<sub>2</sub>/MWh. Uusiutuvien polttoaineiden päästökerroin on puolestaan nolla. (Motiva 2020, 29–30.)

Polttoaineen hintaan vaikuttaa lisäksi mahdolliset tuet (Neittaanmäki & Schroderus 2010, 19). Esimerkiksi uusiutuvaan energiaan liittyviin investointeihin on saatavilla useita erilaisia tukia ja rahoitusta. Myös turve voidaan mainita esimerkkinä polttoaineesta, joka on saanut tukea erityisen alhaisesta verokohtelusta verrattuna muihin fossiilisiin polttoaineisiin. Turpeen käyttöä halutaan kuitenkin vähentää ja turpeen hintaan tultaneen säätelämään jatkossa ns. lattiahinnan avulla: Jos päästöoikeuden hinta laskee, turpeen verotus nousee, jotta polttoaineen hinta säilyy halutulla minimitasolla. (Eskonen 2020.)

Vuoden 2021 aikana päästöoikeuden hinta on noussut merkittävästi, kuten kuvio 2 havainnollistaa. Samaan aikaan myös turpeen energiavero on noussut jyrkästi. Päästöjen hinnan äkillinen nousu on aiheuttanut lämmityskustannusten kasvua kaukolämpöasiakkaille, sillä lähes 40 % kaukolämpöenergiasta on tuotettu kivihiilellä, öljyllä, maakaasulla tai turpeella. Tämän seurauksena energiayhtiöt ovat joutuneet nopeuttamaan muun muassa turpeen käytön lopettamista ja siirtymistä muiden polttoaineiden käyttöön. (Hartikainen 2021.)

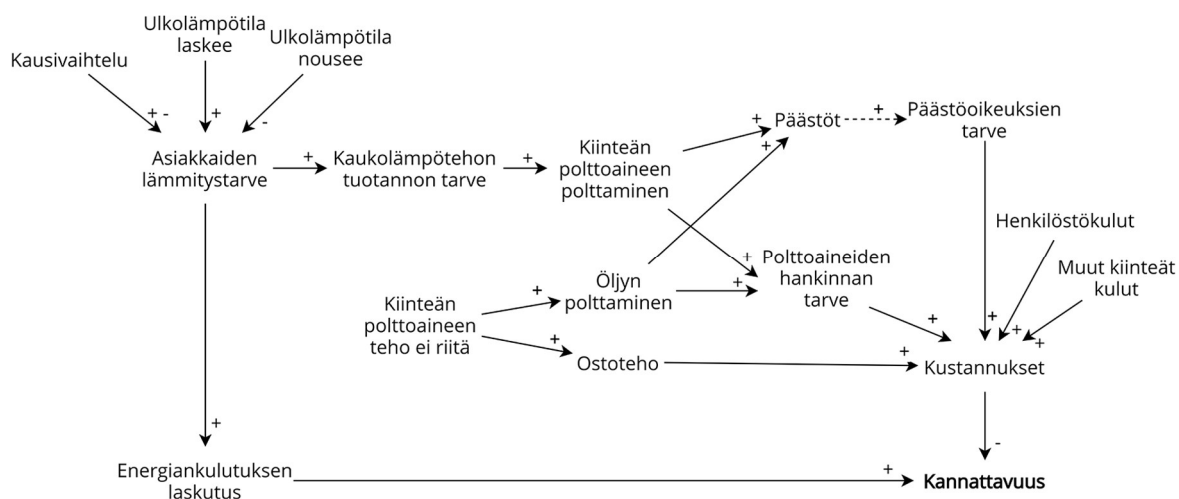


Kuvio 2. Hiilidioksiditonin hinnan kehitys (Helsingin Sanomat 2021)

### 2.2.2 Kannattavuuden laskenta

Ongelman monimutkaisuudesta johtuen, energian tuotannon kannattavuutta arvioitaessa malleja ja oletuksia on yksinkertaistettava. Esimerkiksi Neittaanmäki ja Schroderus (2010) ovat huomioineet pelkistetyssä energiantuotanto- ja lämpölaitosten oleellisina kustannustekijöinä käytettävän polttoaineen hinnan, käyttö- ja pääomakustannukset sekä omistajien tuotto-odotukset. Käyttökustannuksilla tarkoitetaan voimalaitosyksikön käyttö-, huolto-, materiaali- ja jäte- sekä muita vastaavia kuluja. Käyttö- ja pääomakulut vaihtelevat 20–40 prosentin välillä riippuen voimalan käyttöasteesta. Polttoainekustannukset ovat kuitenkin kaikista kuluista merkittävimpiä. Lisäksi polttoainekustannusten hinta voi vaihdella ja olla vaikeasti ennakoitavissa. (Neittaanmäki & Schroderus 2010, 2–4). Käyttö- ja pääomakulut ovat kiinteitä kuluja (Neittaanmäki & Schroderus 2010, 12). Kannattavuuslaskennassa voidaan huomioida polttoainekulujen ja kiinteiden kulujen lisäksi omistajien tuotto-odotus ja kaukolämmön tuotannon hävikki (Neittaanmäki & Schroderus 2010, 6). Energiantuotannon merkittävin tuottotekijä on lämpöenergian myyntihinta, jolla on iso merkitys kannattavuuteen.

Kuviossa 3 on havainnollistettu kannattavuuteen vaikuttavia tekijöitä. Kuvio esittää kausali-teettia, jossa jokaisella nimikkeellä on joko kasvattava tai vähentävä vaikutus muihin nimikkeisiin. Kaukolämmön tuotanto perustuu lämmitystarpeeseen, johon vaikuttavat ulkolämpötila sekä kausivaihtelu. Muun muassa vuoden ja vuorokauden ajat sekä lomat ja pyhäpäivät muodostavat kausivaihtelun. Kasvava lämmityksen tarve aiheuttaa lämpöenergian tuottamista joko kiinteällä polttoaineella, öljyllä tai ostetulla teholla. Näistä öljy ja mahdollinen fossiilinen kiinteä polttoaine aiheuttavat päästöjä. Päästöt eivät suoraan aiheuta päästöoikeuksien tarvetta, sillä päästöt on arvioitava etukäteen ja hankittava tarvittava määrä päästöoikeuksia. Muita kustannustekijöitä ovat henkilöstökulut ja muut kiinteät kulut. Kannattavuus muodostuu asiakkaiden lämpöenergian käytön laskutuksen ja kaikkien muodostuvien kustannusten perusteella.



Kuvio 3. Kaukolämmön tuotannon kannattavuuteen vaikuttavia tekijöitä

Yrityksen kannattavuutta kuvataan yleensä erilaisilla vakiintuneilla tunnusluvuilla, kuten myyntikateprosentilla, käyttökateprosentilla tai liikevoittoprosentilla (engl. EBIT, Earnings Before Interests and Taxes). Myyntikateprosentti tarkoittaa myyntikatteen suhteellista osuutta liikevaihdosta. Toisin sanoen se kertoo, kuinka suuri osuus myyntituotosta jää jäljelle, kun niistä vähennetään myydyin palvelun hankintakustannukset eli muuttuvat kulut. (Visma 2019.) Myyntikatteen ja myyntikateprosentin laskukaava on esitetty kaavassa 2.1. Kaukolämmön tuotannossa liikevaihdosta vähennettäviä muuttuvia kuluja ovat suurimaksi osaksi polttoaineen hinta.

$$\text{Myyntikate} = \text{Liikevaihto} - \text{Muuttuvat kulut} \quad (2.1a)$$

$$\text{Myyntikate}\% = \frac{\text{Myyntikate}}{\text{Liikevaihto}} * 100 \quad (2.1b)$$

(Visma 2019; Visma a). Käyttökateprosentti kuvaa käyttökateen suhteellista osuutta liikevaihdosta. Käyttökate lasketaan lisäämällä yrityksen liikevoittoon poistot ja arvonalentumiset ja se kertoo näin ollen, paljonko yritykselle jää katetta liikevaihdosta kaikkien toimintakulujen vähentämisen jälkeen. Käyttökateprosentti tarkoittaa käyttökateen suhteellista osuutta liikevaihdosta. (Visma 2019.) Käyttökateen ja käyttökateprosentin laskukaava on esitetty kaavassa 2.2. Myyntikatteeseen verrattuna käyttökatteesta vähennetään siis muuttuvien kulujen lisäksi myös kiinteät kulut. Kaukolämmön tuotannossa kaikki käyttö- ja pääomakulut ovat kiinteitä kuluja.

$$\text{Käyttökate} = \text{Liikevaihto} - (\text{Kiinteät kulut} + \text{Muuttuvat kulut}) \quad (2.2a)$$

$$\text{Käyttökate}\% = \frac{\text{Käyttökate}}{\text{Liikevaihto}} * 100 \quad (2.2b)$$

(Visma 2019; Visma b).

### 3 Kemin Energia ja Vesi Oy

#### 3.1 Yhtiön historiaa

Kemin Energia ja Vesi Oy on osa Kemin kaupungin omistamaa Kemin Energia -konsernia. Yritys toimii Kemin kaupungin alueella ja sen toimialana ovat sähkön siirto, kaukolämmön tuotanto ja myynti sekä vesihuolto. Kaukolämpötoiminta Kemin alueella käynnistettiin alun perin vuonna 1975 silloisen Kemin kaupungin sähkölaitoksen toimesta, jonka nimi muutettiin sittemmin Kemin kaupungin energialaitokseksi. Vuonna 1999 kaupungin harjoittama energialaitostoiminta yhtiöitettiin Kemin Energia Oy -nimisen osakeyhtiön alaisuuteen, joka aloitti toimintansa vuoden 2000 alussa. Vuonna 2016 Kemin Energia Oy osti Kemin kaupungilta Kemin Vesi Oy:n osakekannan, minkä myötä syntyneen konsernin nimeksi tuli Kemin Energia ja Vesi Oy. Vesiyhtiö sulautettiin tytäryhtiösulautumisena emoyhtiöönnsä vuoden 2020 alussa. Kokonaisuudessaan konsernin tuottamat palvelut kattavat Kemissä puhtaan veden, jätevesien johtamisen ja käsittelyn, sähkön toimittamisen, lämmön tuotannon ja myynnin. (Kemin Energia ja Vesi Oy 2021a.)

#### 3.2 Yhtiön toiminta

Kemin Energia ja Vesi Oy:n kaukolämpötoiminta on keskittynyt Kemin kaupungin keskustaan ja sitä ympäröiviin kaupunginosiin. Vuonna 2020 Kemin Energia ja Vesi Oy:n kaukolämpö- ja vesihuoltotoiminta työllisti Kemin alueella 63 henkilöä ja koko konsernin liikevaihto oli noin 21,1 miljoonaa euroa, josta lämpöliiketoiminnan osuus oli noin 9,1 miljoonaa euroa. Tällä hetkellä kaukolämpö on Kemin kaupungin suosituin lämmitysmuoto ja se kattaa noin puolet Kemin asuintaloista. Kaukolämpöverkosto on yhtenäinen ja sen pituus oli vuoden 2020 lopussa yhteensä noin 58 kilometriä. Verkostoon on kytketty yhteensä 444 kiinteistöä, joista suurin osa on kerros- tai rivitaloja sekä julkisia rakennuksia. Kaukolämmön piirissä oli vuoden 2020 lopussa yli puolet Kemin noin 20500 asukkaasta ja yhtiön toimittamalla kaukolämmöllä lämmitetään yhteensä noin 3,7 miljoonaa rakennuskuutiometriä. (Kemin Energia ja Vesi Oy 2021c.)

Konserni on panostanut vahvasti energian tuottamiseen hiilineutraalisti ja vuonna 2020 yhtiön tuottamasta kaukolämmöstä 97,6 prosenttia tuotetaan lähialueilta hankituilla kotimaisilla polttoaineilla. Saman vuoden aikana tuotetusta kaukolämmöstä noin 52,6 prosenttia tuotettiin hiilineutraalisti uusiutuvia polttoaineita käyttäen ja alkuvuodesta 2021 avatun uuden lämpövoimalayksikön myötä uusiutuvien polttoaineiden käytön osuus kasvaa entisestään (Kemin Energia ja Vesi Oy, 2021c). Myös osa kaukolämpötuotannon sähköntarpeesta tuotetaan uusiutuvilla energiamuodoilla, kuten vuonna 2015 hankitulla 30 kilowatin

aurinkovoimalalla. Yhtiö pyrkii täysin hiilineutraaliksi vuoteen 2024 mennessä. Yhtiöllä on ISO 9001:2015-standardin mukainen laatusertifikaatti. (Kemin Energia ja Vesi Oy, 2021b)

### 3.3 Kemin Energia ja Vesi Oy:n kaukolämmön tuotanto

Kemin Energia ja Vesi Oy:n pääasiallinen tapa tuottaa kaukolämpöä on kiinteän polttoaineen, kuten turpeen ja hakkeen polttaminen lämpölaitoksissa. Vuonna 2006 käyttöön otettu KPA1 -laitos tuottaa kaukolämpöä maksimissaan 32 megawatin teholla, josta 7,5 megawattia saadaan savukaasujen pesu- ja lämmön talteenottolaitteistosta (Kemin Energia ja Vesi Oy, 2018). KPA1-laitoksen tuottaman kaukolämmön pääasiallinen säätötehon lähde on paikallinen Metsä Fibre Oy:n teollisuuslaitos, josta Kemin Energia ja Vesi Oy ostaa tuotannon hukkalämpöä. Ostettavissa oleva lämpötehon kokonaismäärä riippuu sekä vuodenajasta että käynnissä olevasta teollisuustuotannosta ja sen saatavuus vaihtelee runsaasti. Teoreettinen maksimiteho ostettavalle teholle on noin 25–30 megawattia, mutta keskimäärin ostettavan säätötehon tarve vaihtelee 0–5 megawatin välillä. Lämpimään vuodenaikaan ostoteholla voidaan kattaa suhteessa suurempi osuus säätötehosta, koska hukkalämpöä on tuolloin saatavilla enemmän ja teollisuuslaitoksen oma lämmitystarve on selvästi pienempi kuin talvikaudella. Kaikkein kylmimpinä ajanjaksoina teollisuuslaitoksen tuottama lämpöenergia menee laitoksen omaan käyttöön, eikä ostotehoa ole tällöin saatavilla.

Laitoksille saapuvien kiinteän polttoaineen kuormien energiansisällön selvittämiseksi kaikki kuormat punnitaan ennen varastointia ja niistä otetaan näytteitä. Kuormanäytteiden perusteella määritellään kunkin polttoainekuorman kosteus hyödyntämällä uunikuivatusmenetelmää, jossa näytteet punnitaan tarkkuusvaa'alla ennen uunissa kuivattamista ja kuivattamisen jälkeen. Uunikuivatusmenetelmä tulee suorittaa standardin SFS-EN ISO 18134–2:2015 tai sitä vastaavaksi todennetun menetelmän mukaisesti. Polttoaineiden kosteuden mittaus asianmukaisella kuivatus- ja mittausmenetelmällä on myös edellytyksenä päästökauppaan mukaan pääsemiselle. (Anttila & Rönkkönen 2017, 43–45.)

Tuotannon omavaraisuuden kasvattamiseksi ja hiilineutraalin tuotantokapasiteetin lisäämiseksi Kemin Energia ja Vesi Oy otti vuoden 2021 ensimmäisellä neljänneksellä käyttöön uuden KPA2-laitoksen, joka käyttää polttoaineena ainoastaan puupohjaisia polttoaineita, kuten haketta, pilkettä ja sahanpurua. Uuden laitoksen maksimiteho on 22 megawattia, josta 4 megawattia saadaan lämmön talteenotosta. Tätä opinnäytetyötä kirjoitettaessa uusi laitos on ollut vasta koekäytössä, mutta jatkossa laitos tulee muuttamaan kaukolämpötuotannon ajamista siten, että lämpimään vuodenaikaan KPA2-laitos vastaa koko tuotannon-tarpeesta. Sään kylmetessä käynnistetään myös KPA1-laitos, jota tullaan ajamaan laitoksen minimiteholla KPA2:n rinnalla siten, että vasta KPA2:n maksimitehon saavutettaessa KPA1:n tehoa aletaan nostaa. Tällä tavoin toimittaessa saavutetaan hyöty puupolttoaineita



käyttävän KPA2-laitoksen ympäristöystävällisyydestä, mutta samalla KPA1:sta on saatavilla nopeasti lisätehoa, mikäli lämmöntarve nousee. Kylmään vuodenaikaan on mielekästä ajaa molempia laitoksia samaan aikaan, koska laitoksen käynnistäminen vaatii tilanteesta riippuen yhdestä kolmeen vuorokautta. Oletettavasti uudet tuotantojärjestelyt tulevat vaikuttamaan vähentävästi ostotehon tarpeeseen, joskin myös ostotehon tarjonta tulee muuttumaan vuonna 2023 avattavaksi suunnitellun Kemin biotuotetehtaan myötä (Metsä Fibre 2021).

Suuremmissa kulutushuipuissa, kiinteän polttoaineen laitosten seisokkien aikana tai tilanteissa, joissa ostotehoa ei ole saatavilla, voidaan kaukolämpöä tarvittaessa tuottaa myös kevytöljyä polttavissa lämpökeskuksissa. Kemin Energia ja Vesi Oy:llä on yhteensä viisi öljylaitosta, joista kaksi on kiinteitä laitoksia ja kolme siirreltävässä tarpeen mukaisesti. Yhteen laskettuna öljylaitosten teoreettinen maksimiteho on noin 71 megawattia, mutta käytännössä niiden tuottama huipputeho kaukolämpöverkkoon voi suurimmillaan olla noin 60 megawattia (Kemin Energia ja Vesi Oy, 2018). KPA1-laitoksen öljyä käyttävät tukipolttimet huomioiden öljytehoa on saatavilla yhteensä noin 70 megawattia. Päästökaupasta ja kustannussyistä johtuen kaukolämmön tuotantoa öljyllä pyritään välttämään viimeiseen asti ja esimerkiksi vuoden 2020 kokonaisenergiämäärästä vain 2,4 prosenttia tuotettiin öljylaitosten avulla (Kemin Energia ja Vesi Oy 2021c.)

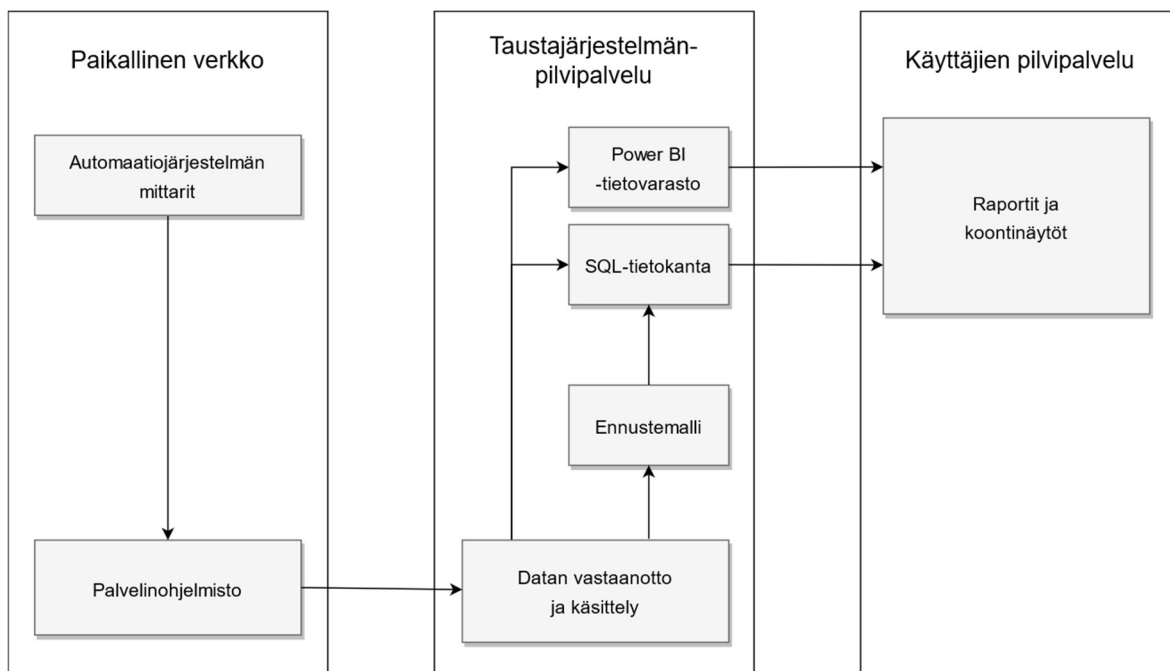
Energiateollisuus ry:n (2020b) julkaiseman tilaston mukaan Kemin Energia ja Vesi Oy oli kaukolämmön käyttömäärien perusteella Suomen 36. suurin kaukolämmön tuottaja vuonna 2019. Yhtiö tuotti vuonna 2019 kaukolämpötehoa 178,2 gigawattituntia, josta lämmön talteenoton osuus oli 23,7 GWh ja ostotehon 42,8 GWh. Kaukolämpöverkon lämpöhäviöt ja mittauserot huomioiden käytetyn kaukolämmön yhteismäärä oli 162,0 GWh.

### 3.4 Automaatiojärjestelmä ja raportointi

Kemin Energia ja Vesi Oy:n kaukolämpötuotannon automaatiojärjestelmä perustuu Siemensin mittaus- ja ohjausjärjestelmään. Edeltävän toimeksiannon puitteissa Kemin Energia ja Vesi Oy:lle on toteutettu reaaliaikainen energiantuotannon seurantajärjestelmä, joka kommunikoi olemassa olevan automaatiojärjestelmän kanssa. Kokonaisuudessaan lämpölaitokset ja kaukolämpöverkosto sisältävät yhteensä yli 22000 erilaista mittauspistettä, joista noin 500 mittauspistettä ovat raportoinnin kannalta olennaisia. Kaikki mittausdata kootaan keskitetysti valvomoon, jossa ajetaan Siemensin WinCC-palvelinohjelmistoa. Valvomosta data siirtyy MQTT-protokollaa hyödyntäen Microsoftin Azure-pilvipalveluun, jossa data vastaanotetaan mikropalvelua hyödyntäen. Mikropalvelu tilaa halutut mittaukset WinCC-palvelimelta. Datan vastaanottamisen lisäksi mikropalvelu määrittelee mittaustiedoille tallennussijainnin konfiguraatitiedoston perusteella, johon on määritelty, mihin

laitoskokonaisuuteen mittauspisteet kuuluvat ja onko kyseistä dataa tarkoitus esittää reaaliaikaisilla koontinäytöillä vai laajemmissa raportointinäkymissä.

Kuvio 4 havainnollistaa automaatiojärjestelmän tuottaman datan siirtoa pilvipalveluun tallennettavaksi. Kaikki mittauksiedot tallennetaan minuuttitasolla Azure SQL Server -tietokantapalvelimelle, minkä lisäksi osa tiedoista viedään Microsoft Power BI:n omaan tietovarastoon reaaliaikaisia näkymiä varten.



Kuvio 4. Tiedonsiirto automaatiojärjestelmästä pilvipohjaiseen raportointiin

Jotta raportointinäkymien käyttö olisi mahdollisimman sujuvaa ja nopeaa, SQL-palvelimelle tallennettavasta datasta luodaan ajoitettuja SQL-proseduureja hyödyntäen tunti-, päivä- ja kuukausikohtaisia koosterivejä. Niitä käyttämällä raportointinäkymistä saadaan huomattavasti selkeämpiä ja suorituskyvyn kannalta kevyempiä verrattuna pelkän minuuttidatan käyttöön, koska kerralla näytettävää dataa on määrällisesti vähemmän. Koosterivien käyttö mahdollistaa myös esimerkiksi pidemmän aikavälin keskiarvojen ja muiden laskennallisten koosteiden sekä trendien seuraamisen minuuttidataa yksinkertaisemmin.

Tallennuksen yhteydessä lämpökeskuksista saatavien mittausten ja polttoaineiden kulustietojen perusteella luodaan laskennallista dataa lämmöntuotannon päästöistä. Tuotettua dataa hyödynnetään eri raporttinäkymissä sekä viranomaisia varten tuotettavassa, kuukausittain automaattisesti muodostettavassa päästöraportissa. Päästömittausten laskennoissa

käytetään Tilastokeskuksen vuosittain julkaisemassa polttoaineluokituksessa määriteltyjä päästökertoimia (Tilastokeskus, 2021), jotka vaihtelevat polttoainetyypeittäin. Polttoaineiden energiansisällön mittauksen tapaan, myös päästölaskentojen tuottaminen on edellytyksenä päästökauppaan mukaan pääsemiselle. Osa päästölaskennoista, kuten esimerkiksi öljylaitosten tuottamat päästöt, perustuvat suoraan kulutettuun polttoainemäärään. Vastavasti puuta polttamalla tuotetun kaukolämpötehon päästölaskennat perustuvat kuukausitasolla mitattuun energiamäärään gigawattitunteina, koska puupolttoaineet eivät ole täysin tasalaatuisia energiansisällöltään ja lisäksi polttoainekuormia saatetaan sekoittaa keskenään, jolloin eri kuormien osuuksien selvittäminen ilman tarkkaa punnitusta on vaikeaa.

## 4 Matemaattinen optimointi

### 4.1 Johdanto matemaattiseen optimointiin

Optimoinnissa on kyse pyrkimyksestä löytää paras ratkaisu, joka täyttää ongelmalle asetetun tavoitteen. Optimointi on tärkeä työkalu fyysisten järjestelmien tarkastelussa ja se toimii usein osana ongelmanratkaisua sekä päätöksentekoa (Nocedal & Wright 2006, 2). Päätöksentekoon liittyy usein rajallisten resurssien tehokas käyttö minimoiden samalla erilaisten riskien vaikutusta. Optimointia voidaan käytännössä soveltaa kaikkialla, kuten tietojenkäsittelyssä, organisaatioiden strategian kehittämisessä, taloustieteissä ja terveydenhuollossa (Simon 2013, 11). Tuotteen hinnoittelu on yksi esimerkki optimoinnin kohteesta, johon vaikuttavia tekijöitä ovat mm. kilpailu, markkina-asema ja tuotantokustannukset. Optimointi on myös tärkeä osa koneoppimista, jossa laadukkaimman mallin tuottamisessa tavoitellaan virheiden minimointia.

Optimointitehtävän ratkaisu vaatii ongelman mallintamisen, joka voi mahdollisesti olla optimointitehtävän tärkein vaihe. Ensimmäiseksi on asetettava tavoite, jota halutaan optimoida. Tehtävä pitää sisällään erilaisia muuttujia, joiden arvot pyritään tavoitteen mukaisesti selvittämään optimoinnin avulla. Optimointitehtävä sisältää usein myös rajoitteita, jotka rajaavat muuttujien mahdollisia arvoja. Mallinnettu ongelma voidaan ratkaista käyttäen erilaisia optimointialgoritmeja, jotka tyypillisesti ratkaisevat tietynlaisia optimointiongelmia. Suoritetun optimoinnin jälkeen on tärkeää varmistaa, että oikeanlainen ratkaisu on löytynyt. Optimoinnin tulos voi myös antaa vihjeitä siitä, miten optimointitehtävän mallintamista voisi kehittää. (Nocedal & Wright 2006, 2.)

Matemaattisen optimoinnin keskeisiä käsitteitä ovat muun muassa kohdefunktio, päätös-  
muuttujat sekä rajoitteet. Kohdefunktio kuvaa optimoinnin tavoitetta, johon pyritään löytämällä päätös-  
muuttujien sopivat arvot huomioiden vallitsevat rajoitteet. Tärkeitä käsitteitä ovat myös paikallinen minimi ja globaali minimi, jotka ovat funktion ääriarvoja ja kuvastavat optimointitehtävän mahdollisten ratkaisujen määrää. Edellä mainitut käsitteet kuvataan tarkemmin alaluvuissa.

#### 4.1.1 Kohdefunktio

Optimointitehtävän ratkaisemiseksi tarvitaan tavoite, jonka suorituskykyä voidaan mitata. Tällainen tavoite voi olla esimerkiksi tuoton maksimointi tai kustannusten minimointi, jolloin voidaan puhua maksimointi- tai minimointitehtävästä. Kohdefunktioista käytetään minimointitehtävän tapauksessa termiä kustannusfunktio (engl. cost function) ja maksimointiongelman tapauksessa termiä sopivuusfunktio (engl. fitness function) (Simon 2013, 13).

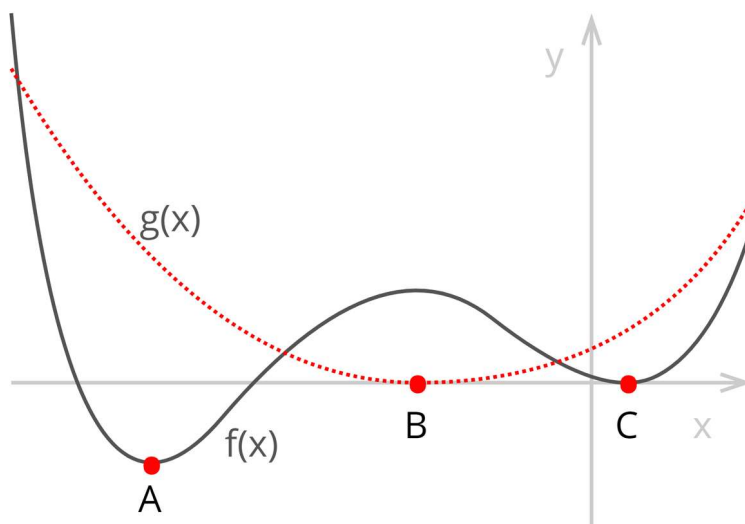
Kaavassa 4.1a on esitetty minimointitehtävä, jossa optimoidaan yhtä kohdefunktiota. Optimointitehtävä voi sisältää myös useamman kohdefunktion, jolloin tavoitellaan usean kohdefunktion minimointia tai maksimointia samanaikaisesti (Simon 2013, 16). Tällöin puhutaan monitavoiteoptimoinnista, jonka mukainen matemaattinen muoto on esitetty kaavassa 4.1b. Kaavojen 4.1a ja 4.1b funktiot  $f$  ovat samanlaiset.

$$\min_x f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1. \quad (4.1a)$$

$$\min_x [f(x) \text{ ja } g(x)], \quad \text{jossa } f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1 \quad (4.1b)$$

$$\text{ja } g(x) = 2(x + 1)^2$$

(Simon 2013, 16). Kaavassa 4.1b on kuvattu minimointitehtävä, joka sisältää kaksi minimoitavaa funktiota;  $f(x)$  ja  $g(x)$ . Nämä funktiot on havainnollistettu kuviossa 5, johon on myös merkitty niiden minimipisteet. Kaikkein pienintä minimipistettä kutsutaan globaaliksi minimiksi ja muita minimipisteitä paikallisiksi minimeiksi. Kohdefunktiolla  $g(x)$  on yksi minimipisteessä B. Kohdefunktiolla  $f(x)$  on kaksi minimipistettä, joista piste C on paikallinen minimi ja piste A globaali minimi. Tällöin kaavan 4.1a minimointitehtävässä ratkaisu on pisteessä A. Kaavan 4.1b minimointitehtävässä ei ole selkeää, mikä olisi ongelman ratkaisun kannalta paras  $x$ :n arvo, sillä kohdefunktiot ovat ristiriidassa keskenään (Simon 2013, 17). Useamman minimin sisältävän ongelman tapauksessa globaalin minimin löytäminen voi olla usein vaikeaa, jolloin sen ratkaisemiseksi voidaan soveltaa muun muassa evoluutioalgoritmeja (Simon 2013, 19–20).



Kuvio 5. Kohdefunktiot ja niiden minimipisteet (Simon 2013, 17)

#### 4.1.2 Päätösmuuttujat

Optimointitehtävät sisältävät erilaisia päätösmuuttujia, joiden parhaita arvoja optimoinnilla etsitään. Optimoinnilla pyritään löytämään sellaiset päätösmuuttujien arvot, jotka maksimoivat tai minimoivat kohdefunktion. Päätösmuuttujat kuvaavat optimointitehtävään liittyviä ominaisuuksia, joiden tilaa tai määrää voidaan mitata joko numeerisena tai kategorisena. Tällaisia ominaisuuksia voivat olla esimerkiksi lämpötila tai auton merkki. Lämpötila on esimerkki jatkuvasta muuttujasta, sillä sen mahdolliset arvot voidaan mitata äärettömän tarkasti. Auton merkkejä on rajallinen määrä, joten sitä käsitellään diskreettinä muuttujana. (Nocedal & Wright 2006, 2, 5.)

#### 4.1.3 Rajoitteet

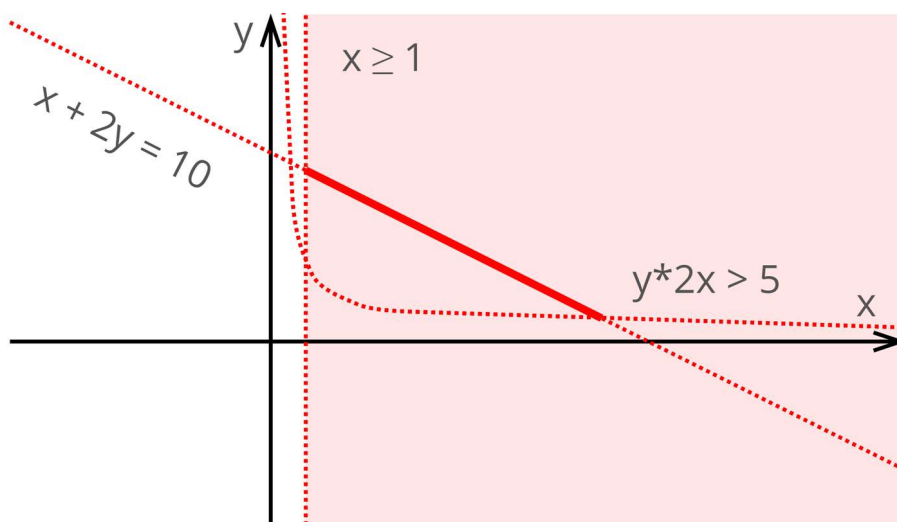
Optimointitehtävät sisältävät lähes aina rajoitteita, jotka rajaavat mahdollisten ratkaisujen määrää. Optimointitehtävää, joka sisältää rajoitteita, kutsutaan rajoitetuksi optimointitehtäväksi. Rajoitteet määritellään yhtälö- tai epäyhtälörajoitteina, jotka voivat olla lineaarisia tai epälineaarisia. Yhtälörajoitteet ovat yhtälöitä, jotka määräävät muuttujien sallitun arvojoukon. Epäyhtälörajoitteilla puolestaan voidaan vertailla suuruusjärjestystä, kuten esimerkiksi minimi- tai maksimiarvoa (Nocedal & Wright 2006, 6.) Kaavassa 4.2a on esitetty esimerkki yhtälörajoitteesta, jossa yhtälön vasemmalla puolella olevan kahden muuttujan laskutoimituksen on oltava 10. Kaavassa 4.2b on esitetty esimerkki epäyhtälörajoitteesta, jossa  $x$ -muuttujan arvo on oltava suurempi tai yhtä suuri kuin 1. Kaavassa 4.2c on epälineaarinen epäyhtälörajoite.

$$x + 2y = 10 \quad (4.2a)$$

$$x \geq 1 \quad (4.2b)$$

$$y * 2x > 5 \quad (4.2c)$$

Kuviossa 6 on havainnollistettu kaavojen 4.2a ja 4.2b yhtälöiden mukaiset rajoitteet katkoviivoilla erotettuina suorina. Kaavan 4.2b yhtälön ollessa epälineaarinen se esitetään käyränä. Epäyhtälörajoitteiden tapauksessa käyvät arvot ovat sellaisia, jolloin vertailun tulos on tosi. Näin ollen esimerkiksi kaavan 4.2b mukaisen yhtälön tapauksessa muuttuja  $x$  voi olla arvoltaan vähintään 1, joten sitä suuremmat arvot ovat käypiä arvoja. Tämä alue on esitetty tummalla alueella kuvaajan oikeassa reunassa. Vastaavasti kaavan 4.2c mukaisen epälineaarisen yhtälön rajaama alue on esitetty niin ikään tummalla alueella. Kaavan 4.2a yhtälörajoite ei rajaa tiettyä aluetta, mutta määrää käypien arvojen kuuluvan tälle suoralle. Kaikki rajoitteet huomioiden käyvät arvot sijaitsevat kuvaajan keskellä olevalla korostetulla janalla.



Kuvio 6. Erilaisia yhtälö- ja epäyhtälörajoitteita

#### 4.1.4 Mallintaminen

Optimointitehtävän ratkaiseminen vaatii aluksi ongelman mallintamisen, joka auttaa ymmärtämään ongelmaa sekä sen ympäristöä. Matemaattisilla malleilla voidaan kuvata todellisen maailman ongelmia matemaattisin kaavoin ja ilmaisin. Tämän jälkeen malli ratkaistaan käyttämällä sopivaa optimointialgoritmia ja lopuksi analysoidaan ratkaisun oikeellisuus. Optimointitehtävä voidaan esittää standardimuotoisena matemaattisena kaavana, joka on esitetty kaavassa 4.3.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{ehdoin} \quad \begin{aligned} c_i(x) &= 0, & i \in a, \\ c_i(x) &\geq 0, & i \in b, \end{aligned} \quad (4.3)$$

*jossa*

$f(x)$  = kohdefunktio

$x$  = muuttujavektori

$c_i$  = rajoitefunktiot

$a$  = yhtälörajoitteidenjoukko

$b$  = epäyhtälörajoitteidenjoukko

Kaavan 4.3 mukaisessa standardissa esitettävässä muuttujavektori  $x$  käsittää kaikki tehtävän päätösmuuttujat ja rajoitefunktiot kaikki tehtävään liittyvät rajoitteet. Standardimuotoinen esitystapa ei kuitenkaan ole välttämättä käytännöllisin muoto mallintaa optimointitehtäviä, joten päätösmuuttujat voidaan tarvittaessa nimetä kuvaavin nimin. (Nocedal & Wright 2006, 2–3.)

Optimointitehtävät voidaan esittää myös epästandardissa muodossa, joka on usein helpommin luettavissa. Optimointitehtävä voidaan esittää esimerkiksi kaavan 4.4 mukaisesti. Kaavassa 4.4a kohdefunktio esitetään yksinkertaisessa muodossa ja kaavassa 4.4b listataan kaikki yhtälö- ja epäyhtälörajoitteet, joita optimointitehtävään kuuluu.

$$\min_x f(x) = x^4 + 5x^3 + 4x^2 - 4x + 1 \quad (4.4a)$$

$$\text{ehdoin} \quad (4.4b)$$

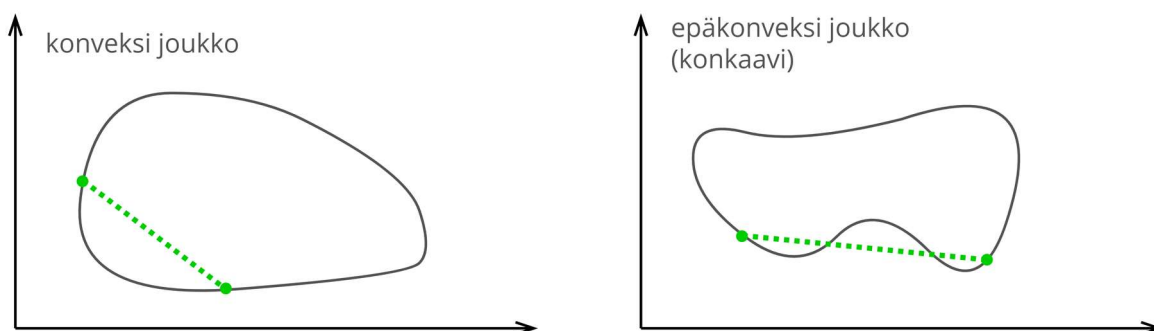
$$x \geq 1$$

$$x \leq 5.5$$



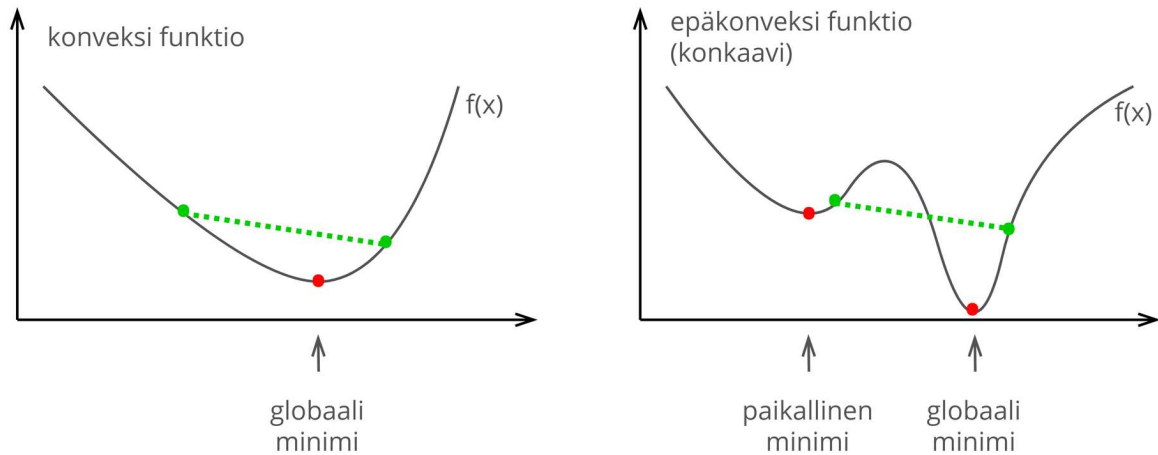
#### 4.1.5 Konveksisuus

Konveksisuus on keskeinen osa optimointia, sillä sen avulla voidaan päätellä, onko optimointitehtävällä yksi vai useampi minimikohta. Minimien määrä kuvaa mahdollisten ratkaisujen määrää, joista kaikki eivät välttämättä ole optimaalisia ratkaisuja. Konvekksi tarkoittaa alaspäin kuperaa, jonka vastakohta on konkaavi, eli ylöspäin kupera. Konveksisuudella viitataan sekä joukkoihin että funktioihin. Konvekksi joukko on sellainen, jossa joukon minkä tahansa kahden pisteen välillä oleva jana pysyy koko ajan rajatun alueen sisällä. Kuviossa 7 on esitetty kaksi joukkoa, joiden kuvaajiin on havainnollistettu kahden satunnaisen pisteen välillä oleva jana. Vasemmanpuoleisessa kuvaajassa jana pysyy rajatun alueen sisällä, jolloin kyseessä on konvekksi joukko. Oikeanpuoleisen kuvaajan tapauksessa jana ylittää rajatun alueen, joten kyseessä on epäkonvekksi, eli konkaavi joukko. (Nocedal & Wright 2006, 8.)



Kuvio 7. Konvekksi ja epäkonvekksi joukko

Funktio on konvekksi, jos janan pisteiden arvot ovat suurempia tai yhtä suuria kuin funktion arvot (Nocedal & Wright 2006, 8). Tätä on havainnollistettu kuviossa 8, jossa on esitetty kaksi funktiota. Funktioiden kuvaajiin on piirretty kahden satunnaisen pisteen välillä oleva jana. Konveksissa funktiossa janan pisteiden arvot eivät alita funktion arvoja missään vaiheessa. Oikeanpuoleisessa kuvaajassa on kyse epäkonveksista funktiosta, jossa osa janan pisteiden arvoista alittaa funktion arvot. (Nocedal & Wright 2006, 8.)



Kuvio 8. Konvekssi ja epäkonvekssi funktio

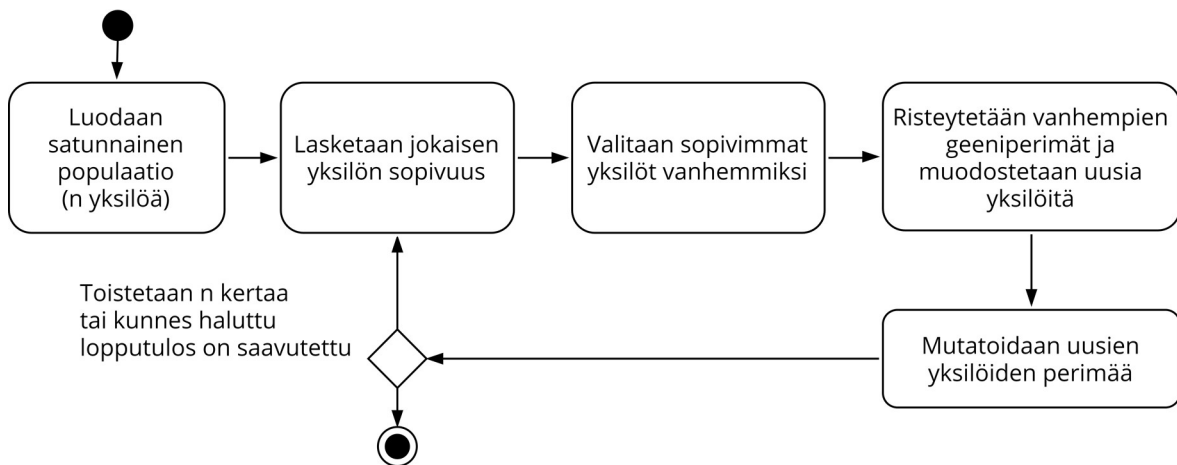
Kuviossa 8 on myös havainnollistettu konveksin ja epäkonveksin funktion minimikohtia. Konveksin optimointitehtävän tapauksessa sen lokali minimi on samalla sen globaali minimi, kun epäkonveksilla funktiolla voi olla useampia minimikohtia. Lineaarinen optimointitehtävä on tyypillinen esimerkki konveksista tehtävästä, jolla voi olla vain yksi globaali minimi. Toisaalta lineaarinen funktio on samaan aikaan sekä konvekssi että konkaavi. (Nocedal & Wright 2006, 7–8.)

#### 4.2 Geneettiset algoritmit

Geneettiset algoritmit ovat evoluutioalgoritmeihin kuuluvia algoritmeja, joiden avulla pyritään löytämään paras ratkaisu haastaviin optimointiongelmiin. Parhaimman ratkaisun löytäminen etsimällä suuresta määrästä vaihtoehtoja voi olla lähes mahdotonta. Evoluutioalgoritmeja voidaan soveltaa käytännössä kaikenlaisten optimointiongelmiin ratkaisemiseksi, minkä myötä ne ovat paljon tutkittuja ja käytössä monissa eri sovelluskohteissa. Evoluutioalgoritmeja on sovellettu muun muassa robottien optimaalisten liikeratojen muodostamiseen, neuroverkkojen opettamiseen sekä lääketieteellisten diagnoosien muodostamiseen. (Simon 2013, 12.)

Geneettiset algoritmit ovat tunnetuimpia ja käytetyimpiä evoluutioalgoritmeja. Niiden toimintaan on otettu vaikutteita todellisesta maailmasta ja niillä simuloidaan luonnonvalintaa, jonka keskeisimmät käsitteet ovat periytyminen, mutaatio sekä kyky lisääntyä. Luonnonvalinnan perusominaisuuksina voidaan pitää lisääntymiskykyistä populaatiota, siihen muodostuvaa variaatiota sekä yksilöiden rajallista elinikää. Geneettiset algoritmit simuloivat kaikkia edellä mainittuja perusominaisuuksia ratkaisun löytämiseksi. Kuviossa 9 havainnollistetussa prosessikaaviossa on kuvattu algoritmin toimintaa, jossa ensimmäisessä vaiheessa

luodaan satunnainen populaatio. Algoritmi laskee populaation yksilöiden sopivuuden ja valitsee näistä sopivimmat uusien yksilöiden vanhemmiksi. Näiden yksilöiden geneettinen informaatio risteytetään, jonka jälkeen uuden yksilön perimään luodaan variaatiota mutatoimalla sen informaatiota. Tämän pohjalta syntyy uusi yksilö uuteen populaatioon, joka muodostaa seuraavan sukupolven sisältäen entistä parempia yksilöitä. Geneettinen algoritmi on iteratiivinen prosessi, joka pysähtyy sukupolvien maksimimäärän kohdalla tai riittävän hyvän ratkaisun löytyessä. (Simon 2013, 35.)



Kuvio 9. Geneettisen algoritmin prosessi (Simon 2013, 55–56)

Geneettisillä algoritmeilla ratkaistavaa optimointitehtävää varten tulee määrittellä kohdefunktio, jonka perusteella sopivat yksilöt voidaan valita (Simon 2013, 44). Tehtävän luonteesta riippuen se voidaan vapaasti määrittellä joko maksimointi- tai minimointitehtävänä. Maksimointitehtävän tapauksessa kohdefunktiosta käytetään nimitystä sopivuusfunktio. Geneettisillä algoritmeilla voidaan esimerkiksi etsiä haluttu sana, jonka algoritmi itsenäisesti muodostaa annetun sopivuusfunktion avulla. Kaavassa 4.5 on esitetty esimerkki sopivuusfunktioista, joka palauttaa prosenttiarvon siitä kuinka suuren osan oikeita kirjaimia yksilön perimä sisältää etsittävään sanaan nähden.

$$\max f(a, b) = a/b, \quad (4.5)$$

*jossa*

*a = yksilön sisältämien oikeiden kirjainten lukumäärä*  
*b = halutun sanan kirjainten lukumäärä.*

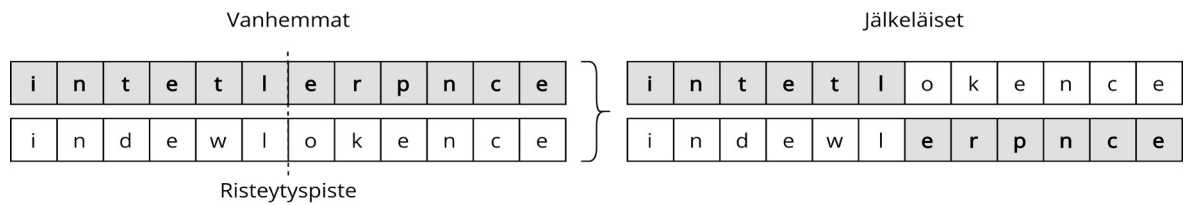
Esimerkissä etsitään sanaa "intelligence", jonka kirjaimia etsitään satunnaisesta merkkijonosta "yuyqkoigence". Etsittävä sana sisältää 12 merkkiä ja satunnainen merkkijono etsittävään sanaan nähden 6 kappaletta oikeita kirjaimia. Näin ollen yksilön sopivuus on kaavan 4.5 sopivuusfunktiolla 0.5, jota verrataan populaation muiden yksilöiden sopivuuteen. Prosenttiarvo kuvastaa suoraan yksilön todennäköisyyttä tulla valituksi muodostamaan jälkeläisiä. Kuviossa 10 on havainnollistettu populaation viisi yksilöä ja niiden lasketut sopivuudet. Jokainen rivi kuvastaa yksilöä, joka koostuu yksittäisistä kirjaimista. Genetiikan termein yksittäistä kirjainta kutsutaan tässä esimerkissä alleeliksi, yksittäistä yksilön ominaisuutta geeniksi ja kaikkien geenien kokonaisuutta kromosomiksi (Simon 2013, 45). Tässä esimerkissä on kuvattu vain yksi geeni, joka muodostaa yksittäisen sanan. Kromosomiin voidaan sisällyttää useita geenejä esimerkiksi koodaamalla, jossa yksittäinen geeni vastaa tiettyä ominaisuutta (Simon 2013, 44).

i	n	t	e	t	l	e	r	p	n	c	e	83%
i	n	d	e	w	l	o	k	e	n	c	e	66%
y	u	y	q	k	o	i	g	e	n	c	e	50%
a	i	n	e	w	q	l	k	k	m	v	e	42%
j	t	o	l	a	c	x	q	h	n	c	e	33%

Kuvio 10. Populaation yksilöt ja niiden sopivuudet

Kuvion 10 populaatio lienee muutaman sukupolven tulos, sillä monien yksilöiden perimässä on havaittavissa oikeanlaisia kirjaimia oikeilla paikoilla. Laskettujen sopivuusprosenttien perusteella populaatiosta valitaan yksilöt ja risteytetään ne jälkeläisten muodostamiseen. Valinta voidaan suorittaa arpomalla, jossa kunkin yksilön sopivuusprosentti vastaa valituksi tulemisen todennäköisyyttä (Simon 2013, 47). Näin ollen populaation keskimmaisella yksilöllä olisi 50 prosentin mahdollisuus tulla valituksi. Kuviossa 11 on havainnollistettu kaksi valittua yksilöä vanhemmiksi. Vanhempien perimää risteytetään määrittämällä risteytyspiste, jonka johdosta kummastakin vanhemmasta siirtyy perimää jälkeläisille (Simon 2013, 46). Esimerkin tapauksessa risteytyspiste sijaitsee sanan keskikohdassa, jolloin ensimmäinen jälkeläinen saa kuvion 11 mukaisesti alkuosan ensimmäiseltä vanhemmalta ja

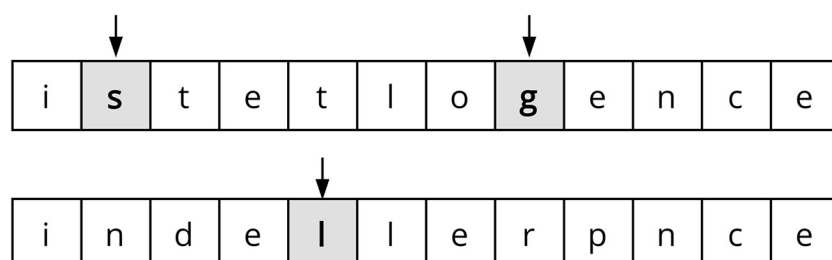
loppuosan toiselta vanhemmalta. Geneettisten algoritmien tapauksessa vanhempia voi olla useampia ja puolestaan jälkeläisiäkin voi muodostaa useita. Koko populaation läpikäynnin jälkeen vanhemmat kuolevat ja jälkeläiset muodostavat uuden populaation (Simon 2013, 46).



Kuvio 11. Kahden vanhemman risteytys ja jälkeläisten muodostus (Simon 2013, 46)

Jälkeläisten muodostamisen jälkeen uuteen populaatioon tuodaan variaatiota mutatoimalla jälkeläisten perimää. Mutaation avulla algoritmi löytää uusia ratkaisuja ja pystyy jatkamaan kehittymistä. Mutaatiolle asetetaan mutaatioaste, jolla ilmaistaan kuinka todennäköisesti yksittäinen alleeli voi muuttua. Liian pieni mutaatioaste ei luo riittävästi variaatiota, kun taas liian suuri mutaatioaste saattaa hidastaa kehittymistä. (Simon 2013, 49.)

Kuviossa 12 on havainnollistettu aiemmassa vaiheessa muodostetut jälkeläiset, joihin kumpaankin on kohdistettu mutaatiota. Ensimmäisestä jälkeläisestä huomataan, että ensimmäinen mutatoitunut kirjain muuttui vääräksi kirjaimeksi, mutta puolestaan toinen mutatoitunut kirjain muuttui oikeaksi. Näin ollen mutaatio ei itsessään vaikuttanut yksilön sopivuuteen, mutta seuraavan sukupolven entistä sopivimpien yksilöiden muodostamiseen sillä on voinut olla suuri merkitys. Geneettinen algoritmi toistaa sopivien yksilöiden valintaa, jälkeläisten muodostamista ja niiden mutaatiota, kunnes mahdollinen ratkaisu on löytynyt tai sukupolvien maksimimäärä on saavutettu.

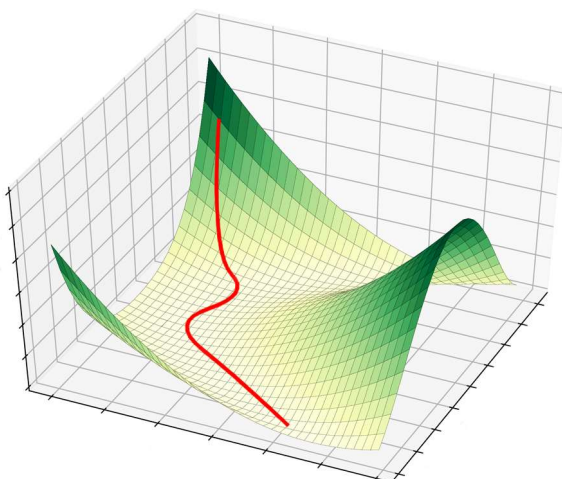


Kuvio 12. Yksilöihin kohdistuva mutaatio

### 4.3 Gradienttilaskumenetelmä

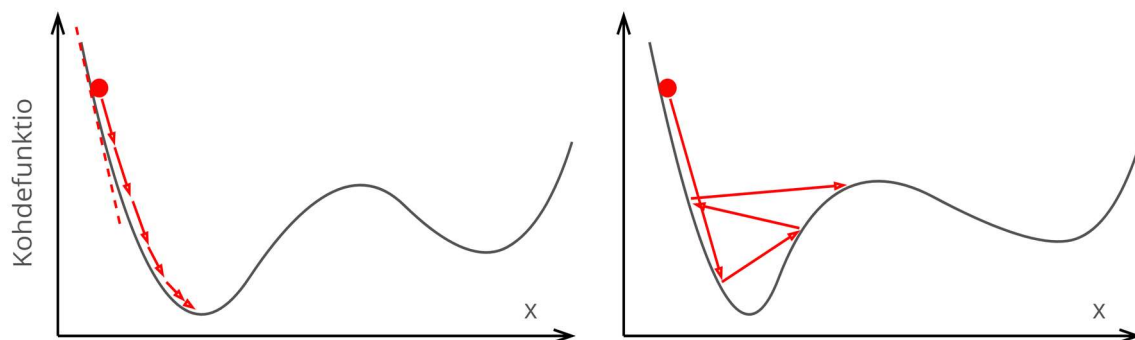
Gradienttilaskumenetelmä (engl. gradient descent) lienee maailman käytetyin optimointimenetelmä, jota hyödynnetään laajalti esimerkiksi koneoppimisessa ja neuroverkoissa. Gradienttilaskumenetelmän tavoitteena on löytää optimointitehtävän muuttujien optimaalisimmat arvot, jotka minimoivat tai maksimoivat kohdefunktion. Menetelmä perustuu derivointiin, jonka avulla voidaan selvittää kohdefunktion tietyssä pisteessä oleva kulmakerroin, eli gradientti. Minimointitehtävän tapauksessa käytetään negatiivista gradienttia, joka osoittaa jyrkimmän laskevan suunnan, toisin sanoen mihin suuntaan kohdefunktio pienenee eniten. Vastaavasti maksimointitehtävässä käytetään positiivista gradienttia, eli mihin suuntaan kohdefunktio kasvaa eniten. Menetelmän käyttö siis vaatii, että kohdefunktio on differentoituva. (Witten ym. 2011, 236–237.)

Gradienttilaskumenetelmää verrataan usein vuorelta laskeutumiseen siten, että pyritään laskeutumaan aina jyrkintä kohtaa alaspäin. Vuori kuvaa tässä tapauksessa kohdefunktion globaalia tai lokaalia maksimia. Kohdefunktion määrittämä pinta voi sisältää useita vuoria ja kuiluja riippuen optimointitehtävän monimutkaisuudesta. Gradienttilaskumenetelmällä voidaan kuitenkin löytää vain lähin minimi, joten mikäli optimointitehtävä on konvekksi, voidaan löytyneen minimin olettaa olevan globaali minimi (Witten ym. 2011, 237). Kuviossa 13 on havainnollistettu gradienttilaskumenetelmällä etenemistä kolmiulotteisessa tehtävässä, jossa pystyakseli kuvaa kohdefunktion arvoa suhteessa kahteen vaakakakseleilla oleviin muuttujiin.



Kuvio 13. Gradienttilaskumenetelmällä laskeudutaan iteratiivisesti "vuoren" rinnettä alas

Globaalin minimin etsiminen gradienttilaskumenetelmää käyttäen aloitetaan satunnaisesta pisteestä, josta edetään jyrkimmän laskun suuntaan tietyn verran. Tämä on gradienttilaskumenetelmän yksi parametri, jota kutsutaan oppimisasteeksi (engl. learning rate). Oppimisasteen toimintaa on havainnollistettu kuviossa 14, jonka mukaisessa tilanteessa kohdefunktiolla on kaksi minimipistettä. Algoritmi etenee negatiivisen gradientin suuntaan oppimisasteen verran lähestyen koko ajan globaalia minimiä. Kuviossa 14 on myös havainnollistettu tilanne, jossa oppimisaste on liian suuri. Tällöin algoritmi etenee liian suurin askelin ja voi kadottaa minimin. Globaalin minimin kadottaminen tarkoittaa tällöin optimaalisen ratkaisun kadottamista. Liian pieni oppimisaste puolestaan tekee algoritmin suorittamisesta hitaan. (Géron 2017, 111–112.)



Kuvio 14. Gradienttilaskumenetelmän toiminta erilaisella oppimisasteella

#### 4.4 Pienimmän neliösumman menetelmä

Pienimmän neliösumman menetelmä (PNS-menetelmä) on matemaattisen optimoinnin menetelmä, jota käytetään regressioanalyysissä. Regressioanalyysi on mahdollisesti käytetyin tilastollinen menetelmä, jonka avulla tutkitaan ja mallinnetaan muuttujien välisiä suhteita (Montgomery ym. 2012, 1). Regressiomalli toteutetaan sovittamalla yhden tai useamman asteen polynomi muuttujien havaintoarvojen välille. Toisin sanoen pyritään selittämään muuttujan vaikutus toiseen muuttujaan matemaattisen mallin avulla. PNS-menetelmän avulla pyritään löytämään regressiomallin optimaaliset parametrit, joilla mallin virhe on mahdollisimman pieni. Lineaarinen regressiomalli voidaan esittää kaavan 4.6 mukaisesti.

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad (4.6)$$

*jossa*

$y$  = selitettävä muuttuja

$\beta_0$  = vakiotermi

$\beta_1$  = kulmakerroin

$x$  = selittävä muuttuja

$\varepsilon$  = virhe.

Vakiotermiä ja kulmakerrointa kutsutaan usein regressiokertoimiksi. Vakiotermi osoittaa mikä on selitettävän muuttujan arvo silloin, kun selittävän muuttujan arvo on 0. Tätä kutsutaan myös leikkauspisteeksi. Kulmakerroin puolestaan osoittaa kuinka suuren muutoksen selittävän muuttujan muutos aiheuttaa selitettävään muuttujaan. Suora ei voi kulkea jokaisen havaintoarvon kautta, joten regressiomallin kaavassa on mukana satunnainen muuttuja, joka osoittaa mallin virheen. Yksittäisen havaintoarvon ja estimaatin erotusta kutsutaan jäännöstermiksi, eli residuaaliksi. (Montgomery ym. 2012, 12–13, 15.)

Pienimmän neliösumman menetelmällä voidaan selvittää entuudestaan tuntemattomat regressiokertoimet minimoimalla virhe. PNS-menetelmässä virhe lasketaan residuaalien neliöiden summana. Residuaalit voivat olla joko positiivisia tai negatiivisia riippuen jääkö ne estimaatin ala- vai yläpuolelle. Neliöjuuren avulla mallia ikään kuin rangaistaan enemmän mitä suurempi residuaali on. PNS-menetelmä voidaan esittää kaavan 4.7 avulla.

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2, \quad (4.7)$$

*jossa*

$y$  = selitettävä muuttuja

$\beta_0$  = vakiotermi

$\beta_1$  = kulmakerroin

$x$  = selittävä muuttuja.

Optimaalisten regressiokerrointen löytyessä residuaalien summa on aina nolla. Tällöin myös residuaalien keskiarvo on nolla. Mikäli regressiotehtävä sisältää kaksi keskenään vahvasti korreloivaa selittävää muuttujaa, voi kyseessä olla multikollineaarisuusongelma, joka voi aiheuttaa regressiomallin vääristymistä. (Montgomery ym. 2012, 13, 20, 289.) Kuviossa 15 on havainnollistettu regressiotehtävä, jossa havaintoarvojen välille on sovitettu



suora  $y = \beta_0 + \beta_1 x + \varepsilon$ . Suoran ja havaintoarvojen välillä olevat katkoviivat esittävät residuaaleja.



Kuvio 15. Sovitettu regressiosuora ja residuaalit

## 5 Koneoppiminen

### 5.1 Johdanto koneoppimiseen

Koneoppiminen on tietojenkäsittelytieteen osa-alue, jossa koneen suoritusta tietyssä tehtävässä kehitetään kokemuksen avulla. Koneoppimisen tavoitteena on tuottaa malli, jolla pystytään tekemään ennusteita. Malli on matemaattinen esitys todellisen maailman ilmiöstä. Mallin muodostamista varten koneen on löydettävä datasta riippuvuuksia ja säännönmukaisuuksia, joista se voi oppia itsenäisesti ilman ihmisen tuottamaa ohjelmakoodia. Koneoppimisella on vahva yhteys tilastotieteeseen sekä matemaattiseen optimointiin, sillä koneoppiminen hyödyntää monia matemaattisia menetelmiä näiltä alueilta. Witten, Frank ja Hall (2011) kuvailevat karkealla tasolla tilastotieteen ja koneoppimisen eroja: tilastotieteessä keskitytään enemmän hypoteesien testaamiseen, kun taas koneoppimisella keskitytään mallintamiseen. Optimoinnissa pyritään löytämään optimaalinen ratkaisu ongelmaan, mutta sen löytäminen voi olla laskennallisesti haastavaa tai liian vaikea mallinnettavaksi. Koneoppimisella voidaan mallintaa monimutkaisia ongelmia ja löytää dataa parhaiten kuvaileva malli, jossa virheen määrä on minimoitu erilaisten optimointimenetelmien avulla. Koneoppimisen käytön eduiksi voidaan luokitella trendien havaitseminen suuressakin datamäärässä. Koneoppimisen mallien tarkkuus voi olla myös paljon parempi kuin ihmisen. Koneoppimisen algoritmit voivat käsitellä suuren määrän dataa ja kartuttaa loputtomasti kokemusta verrattuna ihmisäivöihin. (Géron 2017, 10, 6.)

Koneoppimisessa data on keskeisessä roolissa. Datan avulla konetta voidaan opettaa ja tuottaa dataa kuvaileva koneoppimisen malli. Koneoppimisen hyödyntäminen vaatii datan ymmärrystä, jotta pystytään toteuttamaan oikeanlainen malli ja tuottamaan datalla arvoa. Datan laatutekijät vaikuttavat oleellisesti koneoppimisella saataviin tuloksiin, joten tämä on tärkeä vaihe ennen varsinaisten mallien luontia. Opetusvaiheessa voidaan testata useampia eri algoritmeja, jotka soveltuvat todennäköisimmin ratkaistavaan ongelmaan. Opetusvaiheessa osa datasta käytetään mallin opettamiseen, jonka jälkeen toisella osalla testataan mallin kyky kuvailla opetettua dataa. Mallin suorituskykyä voidaan mitata erilaisilla tunnusluvuilla, jotka kuvaavat mallissa olevan virheen määrää suhteessa dataan. (Géron 2017, 16.)

### 5.2 Ohjattu ja ohjaamaton oppiminen

Koneoppiminen voidaan jakaa eri kategorioihin muun muassa sen perusteella, sisältyykö opettamiseen ihmisen suorittamaa valvontaa. Tällaisiin menetelmiin lukeutuvat ohjatut, ohjaamattomat, puoliohjatut sekä vahvistusoppimisen menetelmät. Oikean menetelmän valinta riippuu käytettävissä olevasta datasta sekä ratkaistavan ongelman luonteesta.

Ohjatun oppimisen menetelmät ovat eniten käytettyjä menetelmiä, joissa algoritmille annetaan syötteenä selittäviä muuttujia sekä niitä vastaava selitettävä muuttuja. Ohjatussa opettamisessa käytettävästä datasta voidaan käyttää nimitystä luokiteltu data, eli se sisältää selitettävän muuttujan. Koneen tehtävänä on mallintaa muuttujien väliset suhteet sekä ennustaa haluttua lopputulosta syötettyjen arvojen perusteella. Ennustamista voidaan kutsua myös estimoinniksi, jolloin mallin tuottamaa tulosta kutsutaan estimaatiksi. Ohjatun oppimisen menetelmät voidaan jakaa regressio- ja luokittelumenetelmiin. Menetelmien erona on vastauksen muodostaminen jatkuvista arvoista tai rajatusta joukosta vaihtoehtoja, eli luokittelu. Luokittelutehtävän tapauksessa selittävänä muuttujana voisi olla esimerkiksi sähköpostin sisältö ja selitettävänä muuttujana tieto onko kyseessä roskaposti vai ei. (Géron 2017, 8.)

Ohjaamattomassa oppimisessä data on luokittelematonta, eli se sisältää ainoastaan selittäviä muuttujia. Ohjaamattomassa oppimisessä koneen tehtävänä on luoda malli perustuen datassa olevien havaintojen samanlaisuuteen. Mallin toimivuuden arviointi on hankalaa, sillä sen tuloksia ei voida verrata ohjatun oppimisen tavoin haluttuihin lopputuloksiin. Ohjaamatonta oppimista voidaan muun muassa hyödyntää luomalla uutta dataa olemassa olevan pohjalta, esimerkiksi generoimalla ihmisten kasvokuvia käyttäen datana todellisten ihmisten kasvokuvia. Tällaisessa käyttökohteessa sen soveltamisesta puhutaan generatiivisena oppimisena. (Géron 2017, 10, 411.)

Edellä mainittujen menetelmien hybridi on puoliohjattu oppiminen (engl. semisupervised learning), jossa data on osittain luokiteltua ja luokittelematonta. Käyttökohteena voi olla esimerkiksi suuren datamäärän käsittely, jossa sen läpikäynti ihmisen toimesta on työlästä tai mahdotonta. Menetelmässä luokitellun datan pohjalta voidaan luoda malli, jolla voidaan ennustaa luokittelemattoman datan todennäköisimmät selitettävät muuttujat. Lopputuloksena data on luokiteltua, jonka käyttämiseen voidaan hyödyntää ohjatun oppimisen menetelmiä. Toinen tyypillinen puoliohjatun oppimisen lähestymistapa on rinnakkaisoppiminen (engl. co-training), jossa luodaan iteratiivisesti useampia malleja datan eri osista ja käytetään niitä luokittelemattoman datan luokitteluun. Datat luokittelemiseksi mallien parhaimmat tapaukset yhdistetään luokiteltuun dataan ja käytetään seuraavan iteraation opetusdatana. (Witten ym. 2011, 294–298.)

Vahvistusoppiminen (engl. reinforcement learning) sisältää oppimisagentin, jolle annetaan joko positiivista tai negatiivista palautetta sen toiminnan perusteella. Tässä menetelmässä malli oppii halutunlaista toimintatapaa, jonka pohjalta se kykenee toimimaan itsenäisesti erilaisissa ympäristöissä. Vahvistusoppimisen käyttökohteeksi voi olla esimerkiksi robotti, joka

itsenäisesti havainnoi ympäristöään ja pyrkii suorittamaan sille asetetut tehtävät oikealla tavalla. (Géron 2017, 13–14.)

### 5.3 Yhdistelmäoppiminen

Yhdistelmäoppimisen (eng. ensemble learning) menetelmillä luodaan useita yksinkertaisia malleja ja yhdistetään niitä muodostamaan yksittäinen entistä parempi ennustemalli (Hastie ym. 2013, 605). Yhdistelmäoppimisen menetelmillä pystytään usein tuottamaan yksittäistä algoritmia parempi ennustetarkkuus (Géron 2017, 74). Tässä luvussa käsitellään yleisimpiä yhdistelmäoppimisen menetelmiä, kuten bootstrap-otantamenetelmään perustuvaa aggregointia sekä boosting-menetelmää.

#### 5.3.1 Bootstrap ja bagging

Bootstrap on tilastollinen otantamenetelmä, jossa aineistosta luodaan useita satunnaisia otoksia. Menetelmän avulla voidaan arvioida aineiston tilastollista tarkkuutta estimoimalla tietyn tunnusluvun jakaumaa tai soveltaa sitä ennustevirheen arviointiin. Jokaisesta otoksesta lasketaan halutut tunnusluvut, kuten esimerkiksi keskihajonnat, joista muodostetaan yksittäinen bootstrap-estimaatti. (Hastie ym. 2013, 249.)

Bootstrap aggregointi eli bagging on bootstrapiin perustuva menetelmä, jossa ensimmäinen vaihe koostuu useiden mallien opettamisesta bootstrap-otannalla valituilla satunnaisilla otoksilla. Toisessa vaiheessa mallit yhdistetään joko äänestämällä tai mallien estimaattien keskiarvon avulla. Esimerkiksi regressiopuun tapauksessa bootstrap-aggregoitu estimaatti on mallien ennusteiden keskiarvo, mikä vähentää sen keskihajontaa johtaen parempaan ennusteeseen. (Hastie ym. 2013, 283.)

#### 5.3.2 Boosting

Boosting on yksi tehokkaimmista oppimismenetelmistä, jota voidaan soveltaa regressio- ja luokitteluongelmiin. Boosting-menetelmässä luodaan joukko heikkoja malleja, joiden tulokset yhdistetään yhdeksi tehokkaaksi malliksi. Boosting on bagging-menetelmän variaatio, jossa yksittäiset mallit luodaan ja opetetaan peräkkäin tietyllä otoksella. Boostingissa mallin virheelliset estimaatit kootaan otokseksi ja käytetään seuraavan mallin opettamisessa. Prosessia voidaan toistaa useita kertoja, minkä johdosta mallin kokonaistarkkuus paranee. (Hastie ym. 2013, 337.)

## 5.4 Koneoppimisen vaiheet

Koneoppimisen soveltaminen edellyttää tiettyjä vaiheita. Ensimmäinen vaihe koostuu datan keräämisestä ja sen arvioinnista, jossa tutkitaan datan käyttökelpoisuutta. Seuraavassa vaiheessa varmistetaan datan laatu ja sitä käsitellään tarpeen mukaan. Datan avulla koneoppimisen malli tuotetaan ja sen suorituskykyä arvioidaan erilaisten tunnuslukujen avulla. Käytännössä datan käsittelyä, muuttujien valintaa, mallin opettamista ja sen testaamista suoritetaan iteratiivisesti, kunnes riittävän hyvä malli saadaan tuotettua. Lopuksi malli sijoitetaan käyttökohteeseensa ennusteiden tuottamiseksi. (Géron 2017, 5.)

### 5.4.1 Datan kerääminen ja käsittely

Koneoppimisen hyödyntäminen edellyttää ohjatun ja ohjaamattoman opettamisen tapauksessa dataa, jonka pohjalta malli voidaan opettaa. Dataa voidaan kerätä monia erilaisia ilmiöitä mittaamalla ja tallentamalla käsiteltävään muotoon. Nykyään dataa voidaan kerätä hyvinkin monipuolisesti, sillä digitaalisia sensoreita on kaikkialla. Dataa arvioitaessa on kiinnostettava huomiota siihen kuinka paljon hyödyllistä informaatiota se sisältää. Joissakin tapauksissa oleellista tietoa voi puuttua, sen luotettavuudesta tai laadusta ei voida olla varmoja tai datassa ilmeneviä ilmiöitä ei pystytä selittämään (Montgomery ym. 2012, 7).

Datan koostaminen käytännössä voi olla haastavaa, sillä dataa voi olla monessa erilaisessa muodossa. Sitä voi olla esimerkiksi saatavilla eri ajankohdilta ja erilaisella tarkkuudella, kuten tunnin tai päivän tarkkuudella. Dataa käsittelemällä voidaan varmistaa sen laatu ennen koneoppivien mallien toteutusta. Dataa voidaan käsitellä muun muassa muokkaamalla puuttuvia ja poikkeavia arvoja, mutta myös muuntamalla dataa erilaiseen, helpommin käsiteltävään muotoon. Oleellista on myös pohtia millä tarkkuudella dataa on käsiteltävä, jotta siitä olisi eniten hyötyä. Datan käsittely voi olla työläs ja aikaa vievä prosessi, mutta se on tärkeä vaihe koneoppimisen prosessissa. Näin ollen voidaan todeta, että koneoppimisen malli on yhtä hyvä kuin sen opettamiseen käytetty data on.

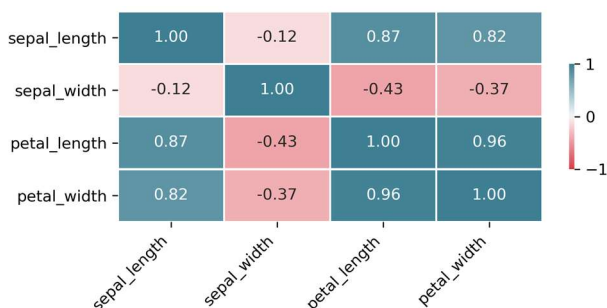
Datassa voi olla puuttuvia arvoja, joiden syynä voi olla esimerkiksi häiriö tiedonkeruujärjestelmässä tai datan konversiosta muodostuvat häiriöt. Tyhjiä arvoja sisältävät havainnot eivät välttämättä ole käyttökelpoisia, jolloin tällaiset havainnot tyypillisesti siivotaan datasta pois. Aikasarjadatan tapauksessa havainnot ovat peräkkäisiä, jolloin tyhjä arvot voidaan tilanteesta riippuen täyttää joko edellisellä tai seuraavalla tunnetulla arvolla.

Datassa esiintyy tyhjiä arvojen lisäksi poikkeamia, jotka ovat muista havainnoista oleellisesti poikkeavia havaintoja. Niillä voi olla suuri vaikutus mallin sovittamisessa, jonka seurauksena malli voi antaa vääristyneitä tuloksia. Poikkeamien luokittelu vaatii datan

ymmärrystä, sillä ne voivat olla todellisia havaintovirheitä, mutta ne voivat olla myös täysin valideja havaintoja. Dataa voidaan analysoida ja visualisoida monipuolisesti ymmärryksen muodostamiseksi. Yksittäisten muuttujien osalta on tunnettava niiden mahdollinen arvoalue, jonka avulla poikkeavat havainnot voidaan helposti luokitella. Data voi sisältää myös erilaisia ilmiöitä, jotka vaativat asiantuntijan tuen niiden tunnistamiseksi ja selittämiseksi. Poikkeamina voidaan ajatella sellaiset havainnot, joiden poistaminen aiheuttaa oleellisia muutoksia analyysin tuloksissa. Vastaavasti mikäli merkittävää muutosta analyysin tuloksissa ei havaita, voidaan olettaa havainnon olevan normaali. (Montgomery ym. 2012, 152–153.)

#### 5.4.2 Muuttujien valinta

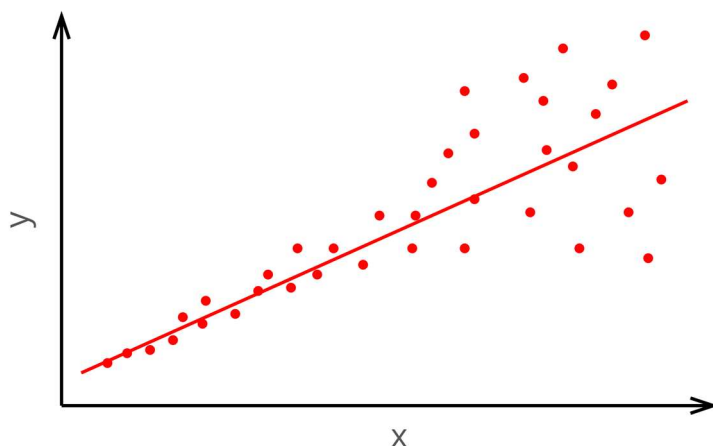
Muuttujien valinta on tärkeä vaihe, sillä vain ongelmanratkaisun kannalta oleellisimpien muuttujien hyödyntäminen voi tuoda toivottuja lopputuloksia koneoppimisessa. Datassa olevia muuttujia voi olla suuri määrä, mutta kaikki eivät välttämättä ole tarpeellisia. Epäoleelliset muuttujat voivat pahimmillaan vaikuttaa negatiivisesti koneoppivan mallin tarkkuuteen. Muuttujia voidaan tarvittaessa myös yhdistää tai jalostaa uusiksi muuttujiksi, jotka tuottavat dataan uutta informaatiota. Muuttujat voidaan valita ongelman luonteesta riippuen esimerkiksi asiantuntijuuden perusteella, kun ongelman toimintaympäristö ja data on tuttua. Muuttujat voivat myös sisältää yllättäviä suhteita muihin muuttujiin, joita ei välttämättä löydetä ilman perusteellisempaa tarkastelua. Muuttujien valintaa voidaan tehdä muun muassa tilastollisella testaamisella, jonka avulla voidaan selvittää millä selittävillä muuttujilla on suurin vaikutus selitettävään muuttujaan. Muuttujien välistä korrelaatiota voidaan visualisoida myös kuvion 16 mukaisella lämpökarttakaaviona, jossa korrelointi on osoitettu arvovälillä -1–1. Negatiiviset kertoimet osoittavat negatiivisen korrelaation ja positiiviset kertoimet puolestaan positiivisen korrelaation. Korrelaation avulla voidaan päätellä millä muuttujilla on eniten selityskykyä selitettävän muuttujan vaihteluun. (Géron 2017, 25–26.)



Kuvio 16. Muuttujien väliset korrelaatiokertoimet lämpökarttakaaviona

Datassa voi esiintyä multikollineaarisuutta, jossa selittävät muuttujat korreloivat vahvasti keskenään. Tämä voi johtaa mallin vääristymiseen, joka esimerkiksi regressioanalyysissä ilmenee regressiokertoimien suurena varianssina. Tämän johdosta mallin ennusteet ovat epävarmoja, mikä tekee mallista heikon. Multikollineaarisuusongelmaa voidaan käsitellä muun muassa keräämällä lisää dataa tai käyttämällä regularisointimenetelmiä, kuten ridge- tai lassoregressiota. Ongelma voidaan mahdollisesti kiertää valitsemalla vain sellaiset muuttujat, joiden välillä ei ole vahvaa korrelaatiota. (Montgomery ym. 2012, 285, 288–289.)

Datassa voi esiintyä myös heteroskedastisuutta, jossa residuaalien hajonta merkittävästi muuttuu selittävän muuttujan arvojen suhteen. Kuviossa 17 on havainnollistettu heteroskedastinen aineisto, jonka perusteella regressiosuora on sovitettu. Residuaalit ovat pienillä  $x$ :n arvoilla pieniä ja ne kasvavat lähes samassa suhteessa  $x$ :n kasvaessa. Näin ollen heteroskedastisella regressiosuoralla ei voida luotettavasti selittää  $y$ :n vaihtelua  $x$ :n avulla. (Mattila 2003.) Heteroskedastisuutta saattaa aiheuttaa poikkeavat arvot, joiden poistamisella heteroskedastisuutta voidaan pienentää.



Kuvio 17. Heteroskedastinen aineisto (Mattila 2003)

#### 5.4.3 Mallin luonti

Koneoppivan mallin luonnissa on kyse opettamisesta, jossa data pyritään yleistämään mallin avulla. Ennen koneoppimisen mallin luontia on valittava ongelmaan parhaiten soveltuva algoritmi. Algoritmeja on monia erilaisia ohjatun ja ohjaamattoman oppimisen ongelmiin. Ohjatun oppimisen algoritmeina voidaan käyttää esimerkiksi lineaarista regressiota, päätöspuita tai neuroverkkoja. Käytettävissä olevasta datasta luodaan otos, jota käytetään mallin opettamiseen. Tyypillisesti opetusdataksi valitaan 80 % käytettävissä olevasta datasta ja loput 20 % muodostavaa tällöin testidatan. Testidatan avulla mallin toimintaa voidaan

arvioida mittaamalla todellisten ja estimoitujen arvojen välistä eroa, eli yleistysvirhettä. (Géron 2017, 29.)

Käytännössä mallin opettaminen tarkoittaa funktion sovittamista kuvailemaan opetusdataa mahdollisimman hyvin. Opetusvaiheen aikana algoritmin luonteesta riippuen voidaan käyttää optimointimenetelmänä esimerkiksi pienimmän neliösumman menetelmää tai gradienttilaskumenetelmää. Monet algoritmit sisältävät useita parametreja, joilla mallin opetusta voidaan hienosäätää entistä tarkemmaksi. Näitä parametreja kutsutaan hyperparametreiksi. (Géron 2017, 28, 111.)

#### 5.4.4 Validointi

Mallin suorituskykyä arvioidaan vertailemalla kuinka tarkkoja ennusteita se tuottaa. Ennusteita verrataan testidatassa oleviin selitettäviin muuttujiin, joiden välistä virhettä arvioidaan useilla eri tunnusluvuilla. Tunnuksluvut paljastavat mallista tietynlaisia piirteitä, jotka voivat vaikuttaa mallin käyttökelpoisuuteen.

Regressiomallin luotettavuutta arvioidaan selitysasteella  $R^2$ , jonka avulla voidaan osoittaa kuinka suuren osuuden selittävät muuttujat pystyvät selittämään selitettävän muuttujan vaihtelusta. Selitysaste ilmaistaan lukuna väliltä 0–1, jossa 1 tarkoittaa, että se selittää selitettävän muuttujan vaihtelun täysin. Selitysaste kuitenkin kasvaa, kun malliin lisätään uusia selittäviä muuttujia, vaikka ne eivät todellisuudessa parantaisi mallin selityskykyä. Selitysaste ei huomioi hajontaa eikä näin ollen myöskään ennustevirhettä, minkä takia sen tulokinnassa on oltava huolellinen. Korjatun selitysasteen avulla voidaan vertailla kahden erillisen mallin luotettavuutta, sillä se huomioi mallissa käytettyjen selittävien muuttujien lukumäärän sekä datan otoskoon. Korjattu selitysaste voi tuloksellaan myös rangaista, mikäli malliin on lisätty sellaisia muuttujia, jotka eivät tuo lisäarvoa. (Montgomery ym. 2012, 36, 87–88.)

Keskineliövirhe (engl. Mean Squared Error, MSE) on yksi käytetyimmistä regressiomenetelmien tunnusluvuista. Se osoittaa mallin havaintoarvojen ja estimaatin välisen keskineliövirheen. Keskineliövirheen neliö (engl. Root Mean Squared Error, RMSE) kuvaa residuaalien keskihajontaa ja keskimääräinen absoluuttinen virhe (engl. Mean Absolute Error, MAE) osoittaa absoluuttisen etäisyyden havaintoarvojen ja estimaatin välillä. MAE huomioi mallin virheen suuruutta, minkä seurauksena esimerkiksi datassa mahdollisesti olevat poikkeavat arvot korostuvat. (Witten ym. 2011, 180–181.)

Mallin ennustevirhettä voidaan testata joko yksittäin tai käyttämällä ristiinvalidointia, jossa testataan useampaa mallia käyttäen erilaisia otoksia datasta. Ristiinvalidoinnissa datasta tehdään tietty määrä samankokoisia otoksia, joita käytetään ristiin siten, että kukin malli



opetetaan ja validoidaan eri otoksilla. Tyypillisenä otoksien lukumääränä voidaan pitää kymmentä otosta. Ristiinvalidoinnin avulla voidaan hyödyntää dataa tehokkaammin valitsemalla sellainen malli, jonka ennustustarkkuus on paras. (Hastie ym. 2013, 283.)

Opetusalgoritmin ja sen parametrien valinnan seurauksena mallissa voi esiintyä ali- tai ylisovittumista. Alisovittunut malli ei kykene kuvailemaan dataa riittävän hyvin, kun taas ylisovittunut kuvailee sitä liian hyvin (Géron 2017, 26, 28). Tämän seurauksena ali- tai ylisovittuneella mallilla ei voida tehdä kovinkaan luotettavia ennusteita entuudestaan tuntemattomalla datalla (Hastie ym. 2013, 364). Ali- ja ylisovittumista voidaan estää regularisoinnilla, jonka tavoitteena on pienentää mallin yleistysvirhettä. Regularisoinnissa mallin virhefunktioon lisätään sen monimutkaisuudesta riippuva termi. Ridge- ja lasso-regressiomenetelmät ovat yleisimpiä käytettyjä regularisointimenetelmiä, jotka pyrkivät estämään mallin parametrien kasvamisen liian suureksi ja näin pienentämään ylisovittumisen riskiä. Ridge-regressiossa regressiokertoimilla pyritään minimoimaan residuaalien neliöiden summan. Lassoregressiossa puolestaan pyritään rajoittamaan regressiokertoimien itseisarvojen summaa. (Hastie ym. 2013, 61–63, 68.)

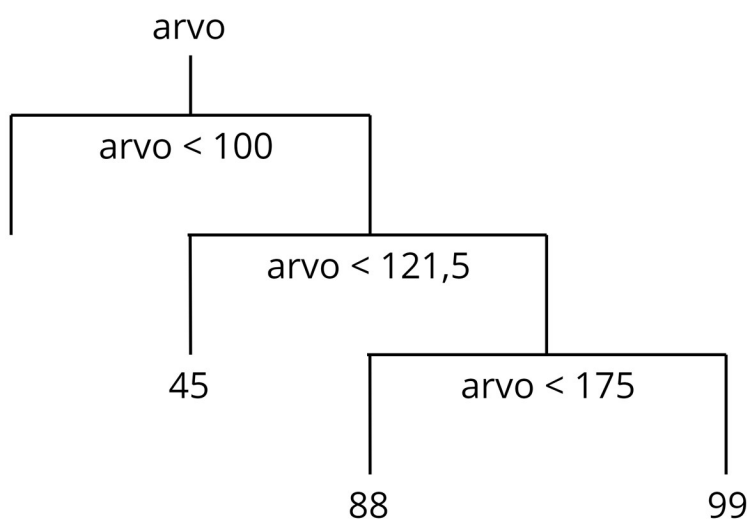
## 5.5 Regressiomenetelmät

Regressiomenetelmät ovat tilastollisia menetelmiä, joilla tutkitaan muuttujien välisiä riippuvuuksia. Regressiomenetelmiä käytetään silloin, kun muuttujat ovat jatkuvia muuttujia. Menetelmillä voidaan selvittää kuinka suuri vaikutus yhdellä tai useammalla muuttujalla on selitettävään muuttujaan rakentamalla tilastollinen malli. Mallien avulla voidaan todeta, mikälainen riippuvuus muuttujien välillä on ja kuinka voimakas se on. Regressiomenetelmiä hyödynnetään muun muassa ennustemallien toteuttamisessa. Ne sisältävät tietynlaisen regressioyhtälön sekä kertoimia, jotka yhdessä muodostavat ennustamisessa käytettävän funktion, eli regressiomallin. Regressiomalli osoittaa lähes poikkeuksetta vain likiarvoisen muuttujien välisen suhteen, sillä regressiolla pyritään yleistämään sille syötetyt havainnot (Géron 2017, 24). Regressiomenetelmiä voidaan soveltaa sekä lineaaristen että epälineaaristen mallien tuottamisessa. (Montgomery ym. 2012, 1–3.) Vaikka regressiomalli osoittaa muuttujien välisen riippuvuuden, se ei kuitenkaan osoita kausaliteettia, jossa toinen muuttuja aiheuttaisi toisen muuttujan muutoksen (Montgomery ym. 2012, 5). Esimerkiksi kahden keskenään korreloivan muuttujan tapauksessa ei voida päätellä kumpi aiheuttaa muutoksen toiseen muuttujaan. Regressiomenetelmistä voi kuitenkin olla apua syy-seuraus-suhteiden löytämisessä.

Yleisin regressiomenetelmä on lineaarinen regressio, jossa muuttujien välinen riippuvuus esitetään regressiosuoran avulla. Regressiosuora sovitetaan havaintopisteisiin optimoimalla suoran kulmakerroin ja leikkauspiste, jotka määräävät sen sijainnin. Lineaarinen

regressio on luonteeltaan lineaarinen, mutta mikäli regressioyhtälö sisältää muuttujien korrattamista toiseen tai useampaan potenssiin, on kyseessä epälineaarinen regressio (Géron 2017, 122). Yksinkertaisin epälineaarinen regressiomenetelmä on polynomiregressio, jonka muoto on eksponentista riippuen kaari tai monimuotoisempi. Logistinen regressio on luonteeltaan binäärinen ja sen avulla voidaan selvittää todennäköisyys kahden tapahtuman välillä (Géron 2017, 134). Käyttökohteena voi olla esimerkiksi luokittelu vastausvaihtoehtojen “kyllä” ja “ei” välillä.

Regressiota voidaan toteuttaa myös puumenetelmillä, joita tyypillisesti käytetään luokittelumalleissa. Regressiopuissa toteutetaan myös eräänlaista luokittelua vertailemalla, mihin arvoalueeseen syötettävä arvo sijoittuu. Regressiopuu-algoritmi pyrkii datan perusteella valitsemaan vertailussa käytettävät arvot. Kuvion 18 mukaisella puumaisella esitystavalla tällainen informaatio on selkeästi havainnollistettavissa. Kuviossa regressiopuulle syötetään arvo, jota vertaillaan puun haarakohdissa ja palautetaan sitä vastaava estimaatti. Eräs suosittu regressiopuun variaatio on satunnaismetsä (engl. random forest), joka perustuu bagging-menetelmään. Siinä malli muodostetaan luomalla useita regressiopuita datasta tehtävillä bootstrap-otannoilla ja tuottamalla niiden keskiarvon perusteella yksi malli. (Hastie ym. 2013, 307, 587.)



Kuvio 18. Regressiopuu (Montgomery ym. 2012, 525)

## 6 Kaukolämmön tuotannon kannattavuuden optimointi

### 6.1 Johdanto optimointitehtävään

Kaukolämmön tuotannossa optimoidaan polttoaineiden tehokasta käyttöä tarvittavan tehon tuottamiseksi. Kaukolämmön tuotannosta vastaavien operaattoreiden tekemä optimointi perustuu suurilta osin heidän omaan kokemukseensa, mitä karttuu mitä enemmän laitoksen operointi tulee tutummaksi. Operaattorin kokemuksen sekä laitoksen mittaustietojen analysoinnin avulla algoritmit voivat oppia ymmärtämään monimutkaisia suhteita ja niiden vaikutusta kaukolämmön tuotantoon.

Ratkaistava ongelma on kaukolämmön tuotannon kannattavuuden optimointi. Optimoinnin kohteena on kaukolämmön tuotantoon liittyvät tuotantomuodot, joiden paras kombinaatio pyritään selvittämään. Tuotantomuodoille on määritetty yhden tuotettavan megawatin kustannukset, joiden perusteella kannattavuutta voidaan mitata. Kiinteän polttoaineen kattilan lämmöntalteenottoteho vaihtelee polttoaineen laadusta riippuen 10–20 %. Tässä optimointitehtävässä oletetaan käytettävän laadukasta polttoainetta, jolloin lämmöntalteenottoteho on 20 % kattilan tehosta. Ongelmaan liittyvät rajoitteet ovat pääosin tuotannolliset maksimit. Ostotehon osalta ei kuitenkaan ole käytettävissä yksiselitteistä maksimitehoa, sillä sitä ei entuudestaan tunneta. Ostotehoa tarjoava Metsä Fibre Oy voi käyttää tuottamansa tehon kokonaan tai osittain itse. Kovimmilla pakkasasteilla ostotehon saatavuus todennäköisesti katkeaa, joten esimerkkinä voidaan olettaa ostotehoa olevan saatavilla -20 °C saakka. Muiden kulujen oletetaan jakautuvan tasaisesti koko vuodelle. Tulojen osalta on syytä huomata, että todellisessa ympäristössä tulot muodostuvat asiakkaan käyttämän energian perusteella. Näin ollen tuloja voi muodostua vain kysynnän mukaisesta määrästä tehoa eikä optimoinnissa mahdollisesti ilmenevää ylitehoa voi laskuttaa keneltäkään.

Kaavassa 6.1 on esitetty optimointitehtävä matemaattisessa muodossa. Kaavassa 6.1a lasketaan tuotettu kokonaisteho, jossa muodostuu kiinteän polttoaineen tehon, ostotehon sekä öljytehon summasta. Kaavassa 6.1b lasketaan tuotannon perusteella tulot. Kaavassa 6.1c lasketaan kulut, joihin vaikuttaa päätösmuuttujien KPA, MF ja OIL mukaiset kertoimet ja kunkin tuotantomuodon kustannus megawattituntia kohden. Kuluihin lasketaan mukaan myös muut kiinteät kulut. Päästöoikeuksien hinta on laskettu mukaan kiinteisiin kuluihin, sillä ne maksetaan erikseen niillä käytävän kaupankäynnin perusteella. Tehon tarpeen ennuste on saatavilla tunnin tarkkuudella, joten muut kulut on myös huomioitava tunnin tarkkuudella. Tämän aikasidonaisuuden takia optimointitehtävällä selvitetään kunkin tuotantomuodon energiamäärää tunnissa, jonka yksikkö on megawattitunti (MWh). Kaavassa 6.1d on esitetty kohdefunktio, jossa lasketaan tuloista ja kuluista muodostuva

kannattavuusprosentti. Laskettava kannattavuusprosentti vastaa suurin piirtein yrityksen käyttökateprosenttia. Kaavassa 6.1e listataan kaikki tunnetut rajoitteet. Mikäli ostotehon maksimiteho tunnettaisiin tai optimoinnissa haluttaisiin rajoittaa öljytehon käyttämistä, voitaisiin ne lisätä rajoitteiden joukkoon.

$$tuotanto = KPA + MF + OIL \quad (6.1a)$$

$$tulot = tuotanto * I_{EUR\_per\_MWh} \quad (6.1b)$$

$$kulut = KPA * C_{KPA} + MF * C_{MF} + OIL * C_{OIL} + C_{muut} \quad (6.1c)$$

$$\max f(P) = \frac{tulot - kulut}{tulot} \quad (6.1d)$$

$$ehdoin \quad (6.1e)$$

$$KPA \geq 0$$

$$OIL \geq 0$$

$$MF \geq 0$$

$$KPA \leq KPA_{max}$$

$$KPA + OIL + MF = D$$

$$T \leq -20 \rightarrow MF = 0,$$

jossa

$C_{KPA}$  = kiinteän polttoaineen tehon kustannus €/MWh

$C_{MF}$  = ostotehon kustannus €/MWh

$C_{OIL}$  = öljytehon kustannus €/MWh

$C_{muut}$  = kiinteät kulut €/h

$D$  = kysyntä MWh

$I_{EUR\_per\_MWh}$  = tulot €/MWh

$KPA$  = kiinteän polttoaineen teho MWh

$MF$  = ostoteho MWh

$OIL$  = öljyteho MWh

$T$  = ulkolämpötila °C

## 6.2 Konsepti ongelmanratkaisusta

Optimointitehtävää voidaan lähteä ratkaisemaan joko matemaattisilla tai koneoppimisen menetelmillä. Näiden lähestymistapojen erona on, että matemaattiselle optimoinnille kerrotaan entuudestaan tunnetut säännöt, kun taas koneoppimisella säännöt pyritään löytämään käytettävissä olevasta datasta. Tässä työssä hyödynnetään Kemian Energia ja Vesi Oy:llä käytössä olevaa kaukolämmön tuotannon ennustemallia, joka tuottaa ennustetta kokonaistehon tarpeesta säännusteen pohjalta. Kokonaistehon ennuste antaa pohjan optimoinnille,

jolla pyritään selvittämään optimaalinen tuotantokombinaatio maksimoiden kannattavuus tietyllä ajanhetkellä.

Työn kehitysvaiheessa vertaillaan muutamaa erilaista lähestymistapaa ongelman ratkaisemiseksi. Ensimmäinen lähestymistapa on lineaarinen optimointi, sillä kaavan 6.1 kohdefunktio ja rajoitteet ovat lineaarisia. Lineaarisella optimoinnilla pystytään selvittämään matemaattisesti optimaalinen tuotantotapojen kombinaatio kysynnän suhteen. Ongelman ratkaisemiseksi testataan myös geneettistä algoritmia, jonka tuloksia vertaillaan lineaarisen optimoinnin tuloksiin. Geneettiselle algoritmilta esitellään myös kaavan 6.1 mukainen kohdefunktio ja vallitsevat rajoitteet. Kolmas lähestymistapa on koneoppimisen menetelmät, joissa hyödynnetään kaukolämmön tuotannosta kerättyä mittaustietoa. Koneoppimisen mallien toteutuksessa hyödynnetään erilaisia regressiomenetelmiä, joista vertaillaan yksinkertaista polynomiregressiota sekä yhdistelmäoppimisen menetelmiin perustuvaa satunnaismetsäregressiota ja gradienttitehostamisregressiota. Teknisinä menetelminä sovelletaan Python-ohjelmointikieltä ja erilaisia kirjastoja, joilla dataa voidaan käsitellä ja käyttää koneoppivien mallien toteutukseen.

### 6.3 Optimaalisen tuotantotavan selvittäminen lineaarisella optimoinnilla

Optimaalinen tuotantotapa kuvaa sellaista tuotantotapojen kombinaatiota, joka tuottaa suurimman kannattavuuden. Optimaalinen kombinaatio voidaan selvittää lineaarisella optimoinnilla, jonka edellytyksenä kohdefunktion ja rajoitteiden tulee olla lineaariset. Lineaarinen optimointitehtävä vaatii, ettei yksittäistä muuttujaa ei ole korotettu toiseen tai korkeampaan potenssiin eikä niitä ole kerrottu toisillaan. Optimointitehtävän kohdefunktio, päätös-  
muuttujat sekä rajoitteet on kuvattu kaavassa 6.1, joka täyttää lineaarisen optimointitehtävän edellytykset.

Kohdefunktiossa kannattavuutta maksimoidaan vähentämällä tuloista polttoaineiden muodostamat kustannukset sekä muut toiminnan kustannukset. Matemaattisessa optimoinnissa algoritmi pyrkii löytämään sellaiset KPA-, OIL- ja MF-muuttujien arvot, jotka tuottavat kohdefunktion suurimman arvon. Ongelman rajoitteista on jätetty pois öljytehon maksimiarvo, jonka laskennallinen maksimiteho on noin 70 MWh. Öljytehon maksimi on mahdolliseen kysyntään nähden suuri eikä se näin ollen ole ongelman ratkaisun kannalta oleellinen. Myös ostotehon maksimin ollessa tuntematon, ei sillä ole rajoitetta. Rajoitteissa on huomioitu, että tuotantomuotojen tehot on oltava positiivisia lukuja sekä niiden kokonaistehon on vastattava tarvittavaa tehoa. Lineaarisen optimointitehtävän tapauksessa kyseessä on konvekssi tapaus, joten löydetty ratkaisu on aina globaali maksimi sekä näin myös paras ratkaisu.

Lineaarista optimointia toteutetaan PuLP-nimisellä avoimen lähdekoodin Python-kirjastolla, jonka avulla optimointitehtävät voidaan kuvata matemaattisina malleina. Mallien kuvaamat ongelmat voidaan ratkaista erilaisilla kirjaston tarjoamilla ratkaisualgoritmeilla. Kuvassa 1 on esitetty Python-koodina ongelman kuvaaminen PuLP-kirjaston avulla sekä tehtävän ratkaisu.

```

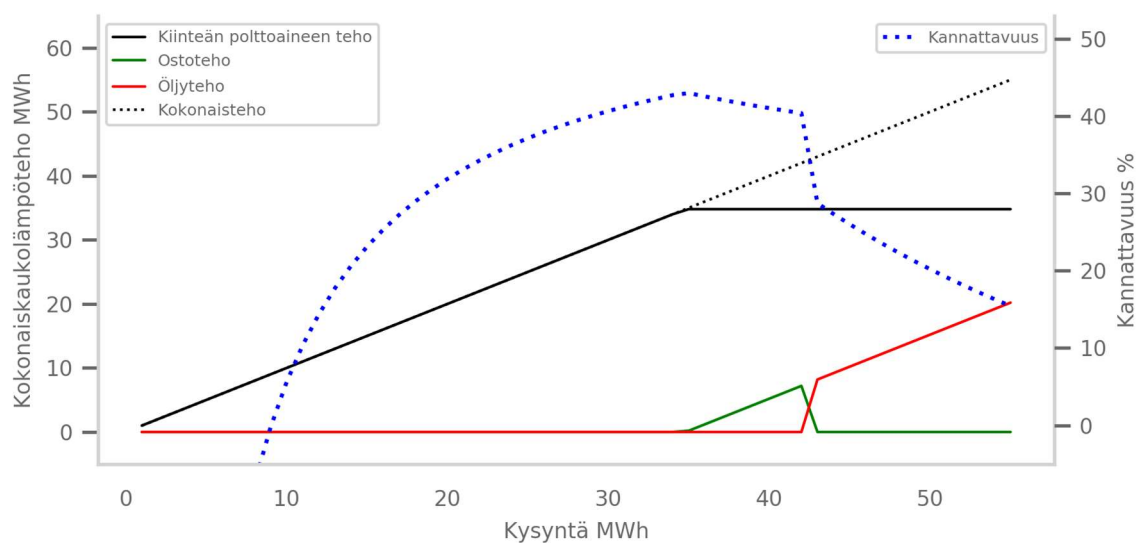
1 model = LpProblem(
2     name='Maksimoi_kaukolämmön_tuotannon_kannattavuus',
3     sense=LpMaximize
4 )
5
6 # Problem variables
7 KPA = LpVariable('KPA', lowBound=0)
8 OIL = LpVariable('OIL', lowBound=0)
9 MF  = LpVariable('MF', lowBound=0)
10
11 # Objective Function
12 model += income - cost_OTHER - cost_KPA * KPA - cost_OIL * OIL - cost_MF * MF
13
14 # Constraints
15 model += KPA <= max_KPA
16 model += KPA >= 0
17 model += OIL >= 0
18 model += MF >= 0
19 model += KPA + OIL + MF == demand
20
21 if MF_not_available:
22     model += MF = 0
23
24 # Solve
25 status = model.solve()

```

Kuva 1. Optimointitehtävän ratkaisu PuLP-kirjaston avulla

Koodin ensimmäisessä vaiheessa luodaan lineaarinen optimointitehtävä käyttäen *LpProblem*-metodia, jossa tehtävälle voidaan antaa nimi ja määrittää sen tyyppi, tässä tapauksessa kyseessä maksimointitehtävä. Seuraavassa vaiheessa määritellään kaikki tehtävään liittyvät muuttujat, joita ovat kolme tuotantomuotoa. Kohdefunktio ilmaistaan kuten matemaattinen funktio, jossa lasketaan jokaisen tuotantomuodon kokonaiskustannus kertomalla muuttuja yksikköhinnalla ja vähentäen kaikki kulut tuloista. Kohdefunktion jälkeen määritellään kaikki tehtävään liittyvät rajoitteet, jotka on listattu pitkälti samoin kuten matemaattisessa kuvauksessa. Rajoitteisiin voidaan vaikuttaa erilaisten ehtojen täyttymisen osalta, kuten ostotehon tapauksessa on määritetty. Mikäli ostotehoa ei ole saatavilla, asetetaan ostotehon maksimitehoksi 0. Lopuksi tehtävä ratkaistaan kutsumalla sen *solve*-metodia.

Kuviossa 19 on havainnollistettu lineaarisen optimoinnin tulokset kuvaajalla, joka osoittaa missä suhteessa tarvittava teho tulisi muodostaa, jotta kannattavuus olisi maksimoitu. Tulokset on koostettu suorittamalla lineaarista optimointia kysynnän funktiona arvoilla 1–52. Kiinteän polttoaineen ollessa edullisin tuotantomuoto on sitä kannattavinta käyttää ensisijaisena tuotantomuotona. Kun kiinteän polttoaineen maksimiteho saavutetaan, käytetään toiseksi edullisimpana tuotantomuotona ostotehoa. Ostotehon saatavuus voi loppua esimerkiksi, kun ulkolämpötila laskee alle -20 asteen, joka vastaa laskennallisesti noin 43 megawattia. Tämän jälkeen tehoa on tuotettava öljyllä, joka on tehomuodoista kallein vaihtoehto ja aiheuttaa merkittävän notkahduksen kannattavuudessa. Kuviosta voidaan myös päätellä, että kaukolämmön tuottaminen ei olisi alle 10 megawatin kohdalla kannattavaa. Tähän on mahdollisesti syynä muiden kulujen jakautuminen tasaisesti koko vuodelle, jolloin kesäkuukausien pieni tehon tuotanto hieman vääristää kannattavuutta. On myös huomioitava, että kaikkien optimoinnissa käytettyjen parametrien ollessa suuripiirteisiä arvioita on optimoinnin tulos myös suuripiirteinen.



Kuvio 19. Lineaarisella optimoinnilla selvitettyt optimaaliset tuotantokombinaatiot kysynnän suhteen

#### 6.4 Optimaalisen tuotantotavan selvittäminen geneettisten algoritmien avulla

Geneettisillä algoritmeilla voidaan ratkaista useita erilaisia optimointiongelmia simuloimalla luonnonvalintaa, periytymistä sekä mutaatiota. Tätä prosessia toistetaan uusien sukupolvien muodostamiseksi, jotka sisältävät optimointitehtävän kannalta entistä parempia

yksilöitä. Optimointitehtävässä kyse on kannattavuuden maksimoinnista, jonka noudattaa kaavan 6.1 optimointitehtävän matemaattista mallia. Optimointitehtävään etsitään ratkaisua käyttämällä Python-ohjelmointikieltä geneettisten algoritmien toteuttamiseen sekä NumPy-kirjastoa, jonka avulla voidaan helposti suorittaa laskentaa taulukoilla ja matriiseilla.

Algoritmin tavoitteena on itsenäisesti löytää paras tuotantomuotojen kombinaatio, joka maksimoi kohdefunktion. Tunnetuista rajoitteista on jätetty pois ehto, joka määrää kaikkien tuotantomuotojen summan olevan yhtä suuri kuin kysyntä. Geneettisten algoritmien tapauksessa tämän saavuttaminen olisi hankalaa, sillä mutaatio jonkin päätösmuuttujan arvossa aiheuttaisi muutoksen muihinkin päätösmuuttujiin, jotta niiden summa pysyisi aina samana. Päätösmuuttujien minimi- ja maksimiarvojen käyttäminen on oleellista, sillä ilman niitä tulokset voivat vääristyä ja tuottaa mahdottomia tuloksia. Kaavassa 6.2 on esitetty osa kaavan 6.1e rajoitteista, joita käytetään geneettisen algoritmin soveltamisessa.

$$\begin{aligned}
 & \textit{ehdot} && (6.2) \\
 & KPA \geq 0 \\
 & OIL \geq 0 \\
 & MF \geq 0 \\
 & KPA \leq KPA_{max} \\
 & T \leq -20 \rightarrow MF = 0,
 \end{aligned}$$

$$\begin{aligned}
 & \textit{jossa} \\
 & KPA = \textit{kiinteän polttoaineen teho MWh} \\
 & MF = \textit{ostoteho MWh} \\
 & OIL = \textit{öljyteho MWh} \\
 & T = \textit{ulkolämpötila } ^\circ C
 \end{aligned}$$

Ratkaisun löytämiseksi on ensin luotava populaatio. Populaation koolla on merkitystä muun muassa ratkaisun löytymiseen vähemmällä sukupolvien määrällä, sillä suurempi populaatio sisältää todennäköisemmin ratkaisun kannalta sopivia piirteitä kuin pienempi populaatio. Toisaalta suuremman populaation käsittely vaatii enemmän laskentatehoa. Optimointitehtävä käsittää kolme päätösmuuttujaa, jotka on määritettävä jokaiseen populaation yksilöön. Kuvassa 2 on esitetty osa generoidusta populaatiosta, jossa jokainen yksilö on esitetty omana rivinään. Kukin päätösmuuttuja kuvaa yksittäistä ominaisuutta, eli geeniä, jotka määritetään positiivisina satunnaislukuina väliltä 0–20. Yksilön ensimmäinen alkio kuvaa kiinteän polttoaineen tehoa, toinen ostotehoa ja kolmas öljytehoa.



```

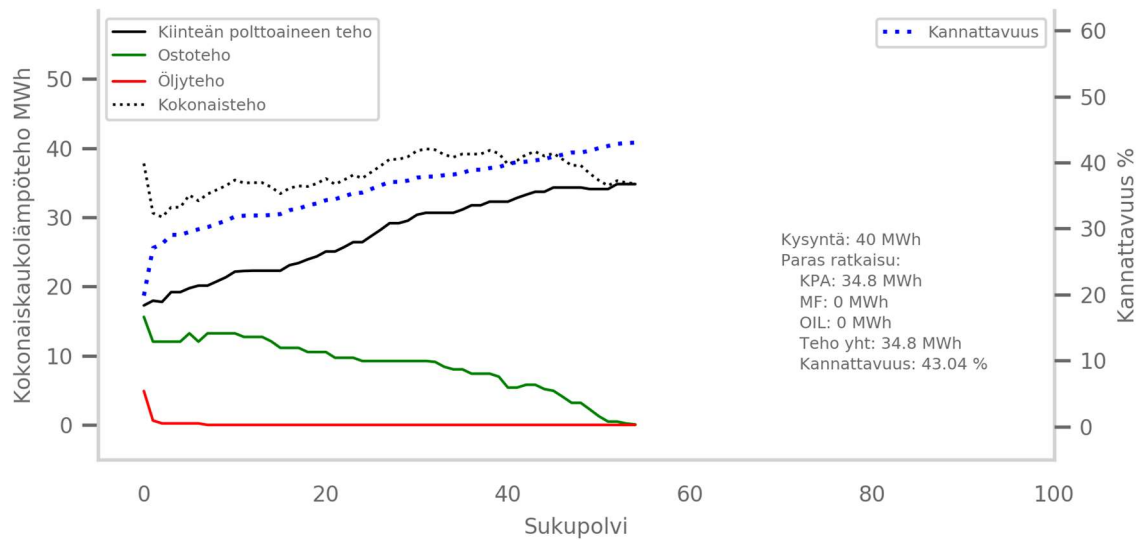
[[10.56659382  1.87144286 15.84894649]
 [18.62068089 17.67391934 18.41790061]
 [ 9.71496732  1.50942413 10.11693985]
 ...
 [17.14109914 13.56206897  2.77324024]
 [15.66290656 11.31317949 14.46634334]
 [ 2.5811073  3.49807242  6.75954844]]

```

Kuva 2. Generoitu populaatio

Geneettisten algoritmien prosessin ensimmäisessä vaiheessa populaation yksilöiden sopivuudet lasketaan kaavan 6.2 sopivuusfunktiolla. Tämän tuloksena jokaiselle yksilölle lasketaan kannattavuusprosentti niiden sisältämien arvojen perusteella. Seuraavassa vaiheessa populaatiosta valitaan kaksi yksilöä vanhemmiksi. Valinta tehdään yksinkertaisesti suoraan sopivuuden perusteella ilman, että valinta tehtäisiin satunnaisesti sopivuutta painottaen. Ensimmäisen vanhemman kohdalla sen sopivuutta muokataan siten, ettei se voi tulla uudestaan valituksi. Valittujen vanhempien perimä risteytetään sekoittamalla niiden geenejä risteytyspisteen perusteella. Risteytyspiste määrätään satunnaisesti siten, että toinen vanhemmista luovuttaa yhden alkion ja toinen kaksi alkia. Perimät risteyttämällä luodaan kaksi uutta alkia, joihin kohdistetaan mutaatiota. Mutaatiota kohdistetaan satunnaiseen alkioon jokaisen yksilön kohdalla, jolloin mutaatioaste on 100 %. Mutaatiossa alkion arvoa muokataan satunnaisluvulla, joka on väliltä  $-1-1$ . Mutaatiossa on huomioitava optimointitehtävälle asetetut rajoitteet, jolloin alkion uuden arvon ollessa negatiivinen, on se asetettava arvoksi 0. Kiinteän polttoaineen teho on myös rajoitettava sen maksimitehon perusteella.

Kuviossa 20 on esitetty ensimmäisen testin tulokset, jossa on kuvattu kunkin tuotantomuodon teho kehittyminen jokaisen sukupolven kohdalla. Kuvaajasta huomataan, kuinka algoritmi on todennut kiinteän polttoaineen vaikuttavan positiivisesti yksilöiden sopivuuteen ja kasvattanut sen osuutta. Algoritmi karsii myös kalliimpien tuotantomuotojen osuutta.



Kuvio 20. Geneettisellä algoritmilla optimoidut tuotantomuodot

Algoritmilla löydetty ratkaisu ei kuitenkaan ole tyydyttävä, sillä sen perusteella tehoa tuotetaisiin liian vähän kysyntään nähden. Syynä tähän on todennäköisesti se, ettei algoritmi saa riittävästi vihjeitä tavoitellusta tilasta, jota pyritään sopivuusfunktiossa laskemaan. Toisin sanoen kalliimpien tuotantomuotojen kustannukset saattavat heikentää kannattavuutta sen verran, ettei se kannusta algoritmia kasvattamaan niiden osuutta. Tähän ratkaisuna on tarkasteltava sopivuusfunktioita ja lisättävä siihen virhetermi, joka pienenee mitä lähempänä tuotantomuotojen yhteenlaskettu teho on kysyntää. Näin algoritmi kannustetaan tuottamaan sellaisen ratkaisun, joka täyttää kysynnän. Kaavassa 6.3a on esitetty virhetermin laskukaava ja kaavassa 6.3b virhetermin huomioiminen sopivuusfunktiossa. Virhetermi muodostetaan laskemalla itseisarvo kysynnän ja tuotetun kokonaistehon erotuksesta ja kertomalla se megawattitunnin hinnalla. Seuraavassa on esitetty optimointitehtävä kokonaisuudessaan, jossa on yhdistetty kaavat 6.1, 6.2 ja 6.3.

$$tuotanto = KPA + MF + OIL \quad (6.1a)$$

$$tulot = tuotanto * I_{EUR\_per\_MWh} \quad (6.1b)$$

$$kulut = KPA * C_{KPA} + MF * C_{MF} + OIL * C_{OIL} + C_{muut} \quad (6.1c)$$

$$virhe = |D - tuotanto| * I_{EUR\_per\_MWh} \quad (6.3a)$$

$$\max f(P) = \frac{tulot - kulut - virhe}{tulot} \quad (6.3b)$$

$$ehdoin \quad (6.2)$$

$$KPA \geq 0$$

$$OIL \geq 0$$

$$MF \geq 0$$

$$KPA \leq KPA_{max}$$

$$T \leq -20 \rightarrow MF = 0,$$

jossa

$C_{KPA}$  = kiinteän polttoaineen tehon kustannus €/MWh

$C_{MF}$  = ostotehon kustannus €/MWh

$C_{OIL}$  = öljytehon kustannus €/MWh

$C_{muut}$  = kiinteät kulut €/h

$D$  = kysyntä MWh

$I_{EUR\_per\_MWh}$  = tulot €/MWh

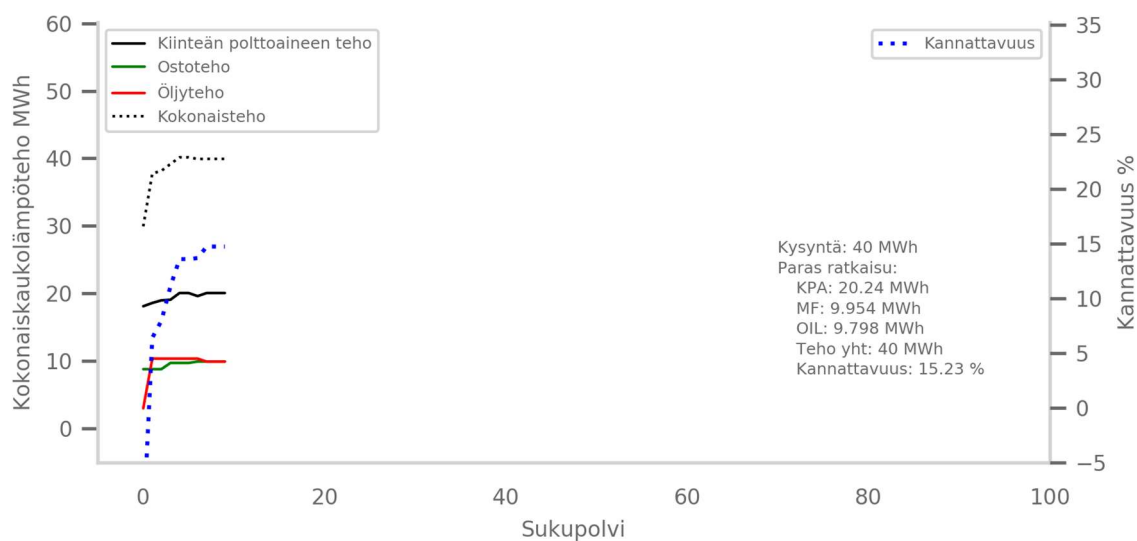
$KPA$  = kiinteän polttoaineen teho MWh

$MF$  = ostoteho MWh

$OIL$  = öljyteho MWh

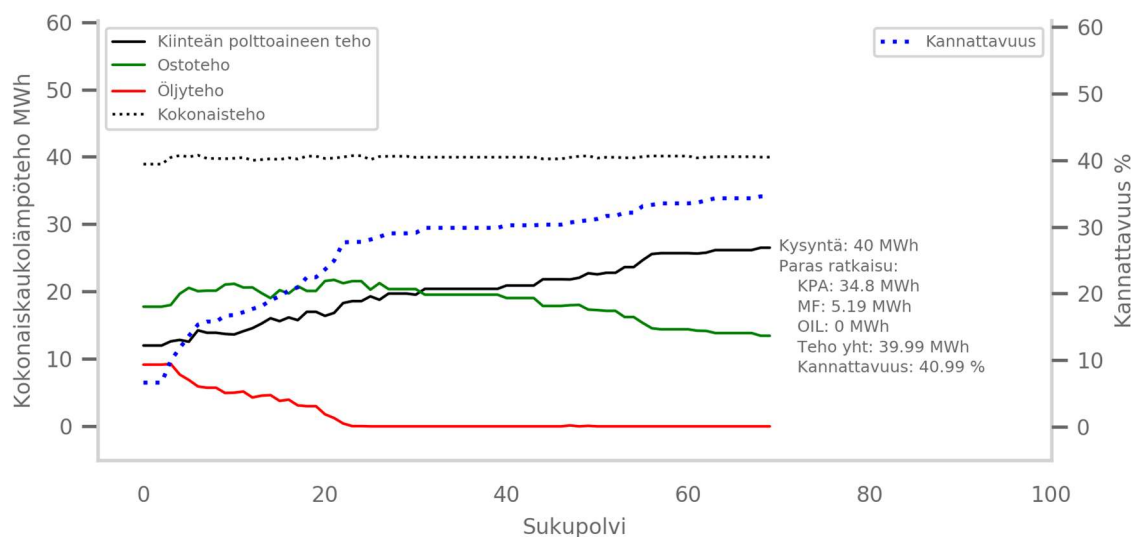
$T$  = ulkolämpötila °C

Kuviossa 21 on havainnollistettu uuden testin tulokset, kun sopivuusfunktiossa on huomioitu virhetermi. Tällä kertaa algoritmi löytää keinon kasvattaa kannattavuutta ja tavoitella samalla kysyntää vastaavaa kokonaistehoa. Algoritmi kuitenkin pysähtyy aikaisessa vaiheessa ja tyytyy kysynnän täyttävään ratkaisuun ilman minkäänlaista optimointia. Tässä kohti syynä voi olla liian pieni variaatio, sillä populaatioon ei kehity sellaisia yksilöitä, jotka voisivat viedä kehitystä optimaalisempaan suuntaan. Vaikka populaation koon kasvattaisi huomattavasti suuremmaksi, ei tällä kuitenkaan ole vaikutusta lopputulokseen.



Kuvio 21. Optimointi huomioiden virhe tavoitetilään nähden

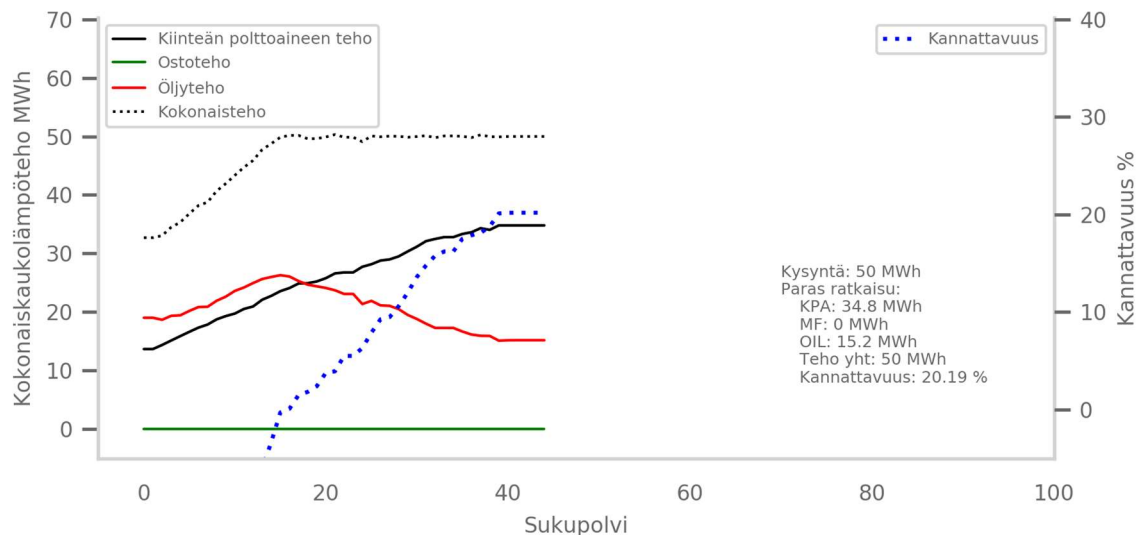
Variaation lisäämiseksi testataan luoda mutaatiota yksilön jokaiseen alkioon, joiden arvot voivat muuttua -1 ja 1 välillä. Kuviossa 22 on havainnollistettu tällainen tapaus, jossa tapahtuu kehitystä jatkuvan mutaation ansiosta. Sen lisäksi että algoritmi pyrkii täyttämään kysynnän, se myös kykenee löytämään edullisimman tuotantokombinaation.



Kuvio 22. Optimaalinen tulos aktiivisella mutaatiolla

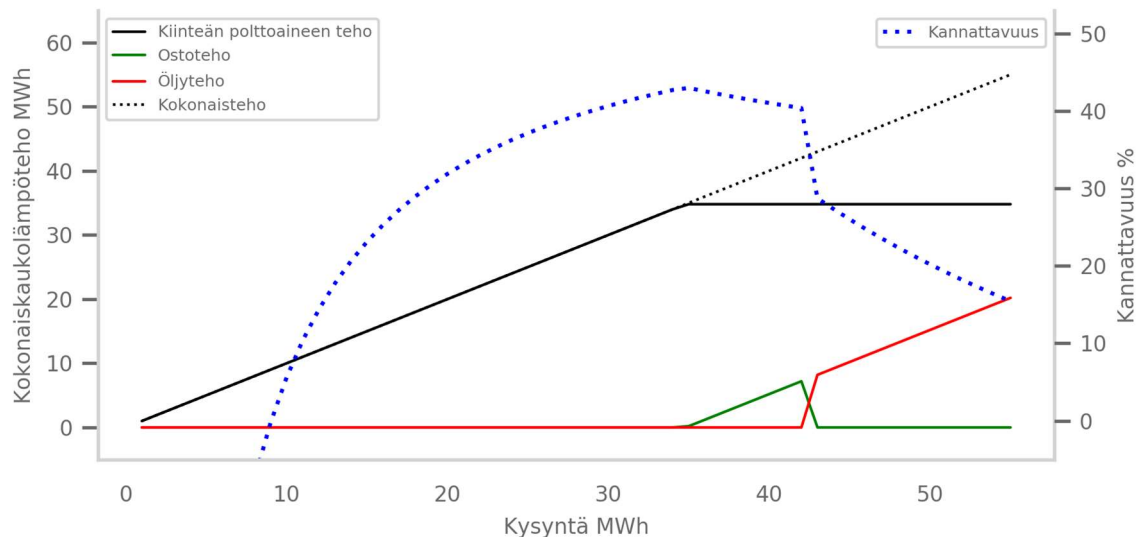
Testin tulos on lupaava, joten on syytä testata vielä muita tilanteita, joissa myös öljyteholle pitäisi olla tarvetta. Öljytehon tarve kasvaa sellaisessa tilanteessa, jossa ostotehoa ei ole saatavilla. Optimointitehtävässä ostotehon oletetaan olevan saatavilla  $-20\text{ }^{\circ}\text{C}$  saakka, jonka jälkeen sen maksimiarvo on käytännössä 0. Tällaisen tilanteen simulointi tarkoittaa geneettisten algoritmien tapauksessa sitä, että ostoteho olisi asetettava populaatioon aina arvoksi 0. Toisin sanoen optimointia toteutettaisiin vain kahdella muuttujalla.

50 MWh kysynnällä oletus on, että öljytehoa on käytettävä. Kuviossa 23 on esitetty tilanne, jossa algoritmi löytää optimaalisen ratkaisun käyttäen kiinteän polttoaineen tehoa sekä öljytehoa. Kuviosta huomataan, kuinka algoritmi tavoittelee alussa kysynnän tasoa kasvattamalla kummankin tuotantomuodon osuutta. Tämän jälkeen algoritmi optimoi pienin askelin tehomuotojen välistä osuutta kannattavuutta parantaen ja löytäen optimaalisen ratkaisun.



Kuvio 23. Optimaalinen ratkaisu kiinteän polttoaineen tehon ja öljytehon välillä

Kuviossa 24 on esitetty kooste algoritmin avulla löydettävistä ratkaisuista kysynnän suhteen. Tulokset on koostettu suorittamalla geneettisen algoritmin prosessia kysynnän funktiona arvoilla 1–55. Kuviosta huomataan, kuinka osto- ja öljyteho vaikuttavat kannattavuuteen, kun kiinteän polttoaineen teho on saavuttanut maksimin.



Kuvio 24. Geneettisellä algoritmilla selvitettyt optimaaliset tuotantokombinaatiot kysynnän suhteen

## 6.5 Koneoppimisen soveltaminen optimaalisen tuotantotavan selvittämisessä

Koneoppimisen avulla hyödynnetään kaukolämmön tuotannosta kerättyä mittaustietoa ja muodostetaan sitä kuvaileva ennustemalli. Ennustemalli paljastaa datassa olevien muuttujien väliset suhteet. Ennen koneoppimisen mallien luontia tutkitaan käytettävissä olevaa aineistoa ja varmistetaan sen laatu. Malleja luodaan käyttäen kolmea erilaista regressioalgoritmia: polynomiregressio (engl. polynomial regression), satunnaismetsäregressio (engl. random forest regression) sekä tehostetun gradientin regressio (engl. gradient boosting regression). Algoritmit on valittu pääosin siten, että voidaan vertailla yksinkertaisempien algoritmien suoriutumista monimutkaisempiin algoritmeihin, jotka hyödyntävät yhdistelmäoppimisen menetelmiä.

### 6.5.1 Data ja sen käsittely

Tutkimuksessa käytössä oleva data koostuu kaukolämmön tuotannon ja paikallisen sääaseman mittaustiedoista. Mittaustietoja on saatavilla minuutin tarkkuudella, mutta niitä on myös koostettu tunnin ja päivän tarkkuudelle käyttökohteesta riippuen. Koneoppivan mallin opettamiseksi riittää tunnin tarkkuudella oleva data. Mallin opettamiseen käytettävä otos muodostetaan vuoden 2019 lopusta lähtien saatavilla olevista mittaustiedoista, jota on noin 16 kuukauden ajalta. Tiedonkeruujärjestelmässä mittaustietojen pohjalta on automaattisesti laskettu erilaisia muuttujia, jotta tietojen käsittely on helpompaa. Tällaisia muuttujia ovat muun muassa kokonaiskaukolämpöteho, joka sisältää tuotantomuotojen yhteenlasketun

tehon vähennettynä mahdollisella omakäyttöenergian osuudella. Kiinteän polttoaineen tehossa on laskettu mukaan savukaasuista kerätty lämpöenergia, eli lämmön talteenotto. Myös kaikkien öljykattiloiden yhteenlaskettu teho on saatavilla yksittäisenä muuttujana. Muuttujien valinta toteutetaan siis käyttämällä näitä valmiiksi laskettuja muuttujia, joilla jokaisella on vaikutusta kaukolämpöenergian tuotannossa. Taulukossa 1 on listattu kaikki muodostetussa datassa olevat muuttujat. *Paiva\_aika*-muuttujalla ei itsessään ole roolia tässä optimointitehtävässä, mutta siitä voitaisiin johtaa uusia muuttujia, kuten kuukausi ja tunti. Nämä toimisivat mallissa omina parametreinaan ja niiden avulla voitaisiin estimoida tuotantokombinaatiot päivän aikana muuttuvan kysynnän suhteen.

<b>Muuttujan nimi</b>	<b>Yksikkö</b>	<b>Kuvaus</b>
paiva_aika	-	Havainnon päivämäärä ja kellonaika ISO8601-formaatissa
Kokonais_KL_teho	MWh	Yhteenlaskettu kokonaisteho
KPA_teho	MWh	Kiinteän polttoaineen teho
MF_teho	MWh	Ostoteho
OIL_teho	MWh	Öljyteho

Taulukko 1. Käytettävissä olevan datan muuttujat

Datan laatu on varmistettava ennen sen käyttöä koneoppivan mallin opettamisessa. Datasta poistetaan tyhjät ja tilastollisesti poikkeavat arvot. Koneoppimisen mallit halutaan opettaa datalla, joka kuvastaa ideaalista kaukolämmön tuotantoa. Tästä syystä datasta pyritään suodattamaan epänormaalit tapahtumat, kuten esimerkiksi kiinteän polttoaineen kattilan epätavallinen toiminta, joka aiheuttaa osto- ja öljytehon käyttämistä. Tällainen tilanne voi olla esimerkiksi huolto. Myös datankeruujärjestelmässä voi ilmetä ongelmia, minkä johdosta joissakin mittaustiedoissa saattaa esiintyä virheitä.

Datan analysointia ja käsittelyä toteutetaan Python-ohjelmointikielellä ja Pandas-kirjastolla, jonka avulla suurenkin datamäärän käsittely ja analysointi on helppoa. Pandas-kirjastosta löytyvä DataFrame soveltuu taulukkomaisen datan käsittelyyn, jossa data voidaan tuoda esimerkiksi CSV-tiedostona tai suoraan tietokannasta noutaen. Datan pohjalta luodun DataFrame:n avulla kaikki tyhjät arvot voidaan joko poistaa tai täyttää edellisellä tai seuraavalla

arvolla. Tyhjän arvon poistaminen tarkoittaa käytännössä koko havainnon poistamista, joten myös käyttökelpoista dataa saattaa kadota poistamisen yhteydessä. Tässä tapauksessa tyhjiä arvoja sisältävien rivien osuus koko datassa oli alle 0,1 %, joten näiden poistaminen ei aiheuta merkittävää informaation katoamista.

Tilastollisesti poikkeavat arvot voidaan löytää normaalijakauman avulla, jossa poikkeaviksi luokitellaan normaalijakauman päissä sijaitsevat havainnot. Tällä ei kuitenkaan voida suodattaa kaikkia datassa olevia poikkeustilanteita, sillä ne eivät välttämättä erotu normaalijakaumassa. Datan käsittelemiseksi on rakennettu algoritmi, jonka avulla tutkitaan kuinka eri tuotantomuodot käyttäytyvät ja luokitella havainnot tietyssä tilanteessa poikkeavaksi toiminnaksi. Poikkeavaa toimintaa on muun muassa kiinteän polttoaineen tehon nopea laskeminen sekä sen nollateho. Poikkeavaksi toiminnaksi luokitellut havainnot voidaan poistaa samalla tavalla kuin tyhjiä arvoja sisältävät havainnot. Tyhjien arvojen ja poikkeavien tilanteiden poistaminen datasta tuottaa koneoppivan mallin opettamiseen soveltuvan otoksen. Otoksen muuttujien tunnuslukuja on esitetty taulukossa 2. Taulukossa prosenttiluvut osoittavat ylä- ja alakvartiileja sekä mediaania.

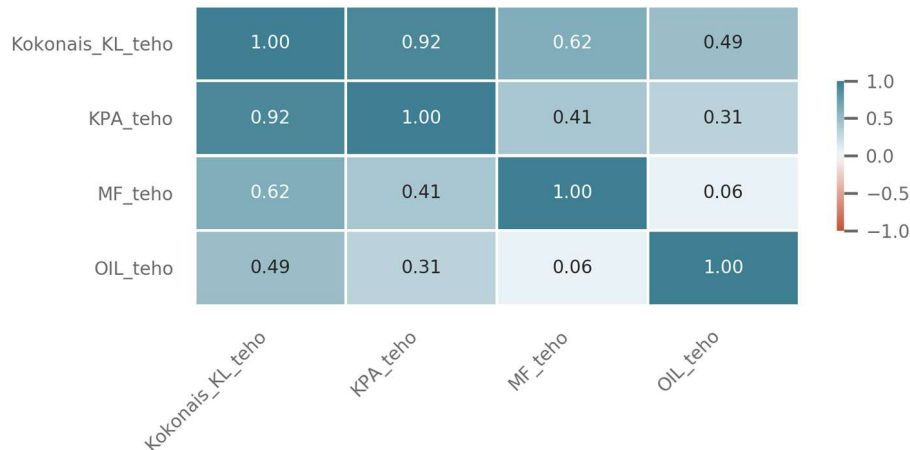
	<b>Kokonais_KL_teho</b>	<b>KPA_teho</b>	<b>MF_teho</b>	<b>OIL_teho</b>
<b>lukumäärä</b>	10220	10220	10220	10220
<b>keskiarvo</b>	20,8148	15,4814	2,1105	0,7027
<b>keskihajonta</b>	10,6639	8,4658	2,7897	2,6211
<b>minimi</b>	2,0687	2,0687	0	0
<b>25 %</b>	12,085	9,7740	0	0
<b>50 %</b>	22,1883	17,6302	1,0253	0
<b>75 %</b>	27,9894	22,5612	3,2791	0
<b>maksimi</b>	52,3916	36,3533	19,4654	30,2290

Taulukko 2. Käsitellyn datan muuttujien tunnuslukuja

Tunnuslukuja tarkastelemalla voidaan todeta, että kaukolämpöenergian tuottamiseksi pyritään käyttämään aina kiinteän polttoaineen tehoa ja ostotehoa toiseksi eniten. Öljytehoa puolestaan käytetään harvoin, sillä sen käyttö jakautuu yläkvartiilipisteen (75 %) yläpuolelle. Muuttujien välisiä suhteita voidaan tarkastella vielä kuvion 25 mukaisella



korrelaatiokertoimia esittäväällä lämpökarttakaaviolla. Korrelaatiokertoimia tarkastelemalla huomataan suuri korrelaatio kokonaiskaukolämpötehon ja kiinteän polttoaineen tehon välillä. Näin ollen voidaan todeta, että pelkästään kiinteän polttoaineen teholla voidaan selittää 92 prosenttisesti kokonaiskaukolämpötehon muodostuminen. Korrelaatiokertoimista huomataan myös, ettei muuttujien välillä ole negatiivista korrelaatiota, eli jokainen tuotantomuoto kasvattaa kokonaistehon määrää.



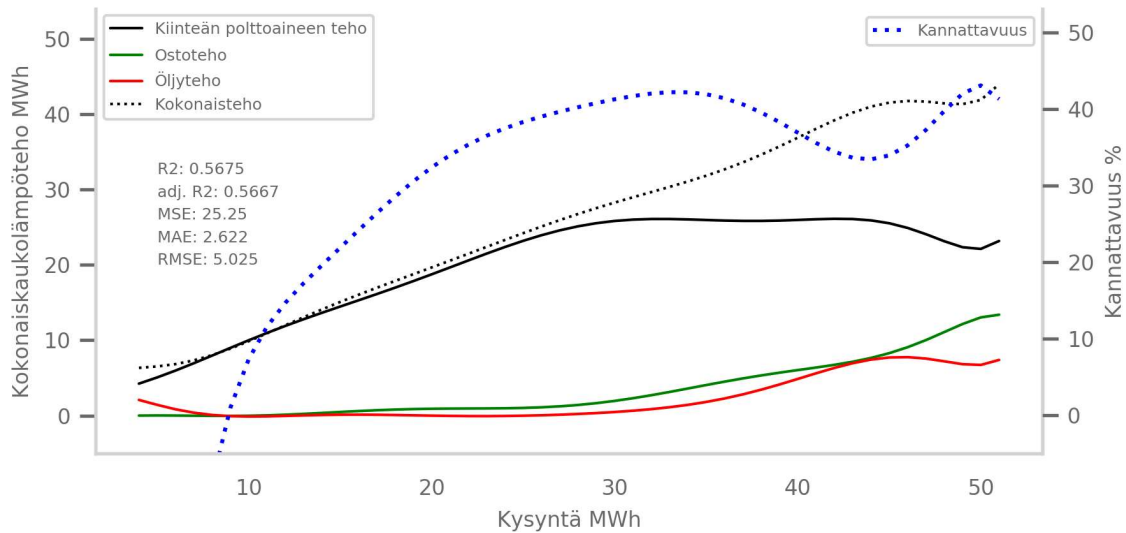
Kuvio 25. Muuttujien väliset korrelaatiokertoimet

### 6.5.2 Mallin luonti polynomiregressiolla

Polynomiregressiolla voidaan sovittaa suoraa monimutkaisempi muoto havaintoarvojen suhteen. Scikit-Learn-kirjasto sisältää PolynomialFeatures-luokan, jolle voidaan määrittää polynomifunktion asteluku. Asteluvun avulla malli voidaan sovittaa esimerkiksi käyttäen toisen tai kolmannen asteen yhtälöä. Tehtyjen testien perusteella sopivaksi asteluvuksi löytyi arvo 8, jonka mukaisella mallilla saavutettiin paras tarkkuus. Polynomiregressiomallin selitysasteeksi saatiin 57 %. Jokainen mallin muodostamisessa käytettävä muuttuja on merkitsevä, joten selitystasetta voidaan pitää luotettavan tunnuslukuna selittämään muuttujien välisiä suhteita. Sitä voidaan pitää myös vertailukelpoisena muihin malleihin nähden, sillä kaikki mallit opetetaan samalla opetusdatalla. Selitystasteen ollessa tarkkuuden suhteellinen mittari, on syytä tarkastella myös muita tunnuslukuja. Esimerkiksi keskineliövirheen neliö (RMSE) huomioi myös virheen keskihajonnan, joten se on myös vertailukelpoinen tunnusluku.

Kuviossa 26 on havainnollistettu polynomiregression avulla estimoitu tehon tuotanto ja kannattavuus kysynnän suhteen. Algoritmin luonteen mukaisesti kuvion viivat ovat muodoltaan pyöreitä. Kuvioista huomataan, että kokonaisteho ei täytä kysyntää suuremmilla tehoilla.

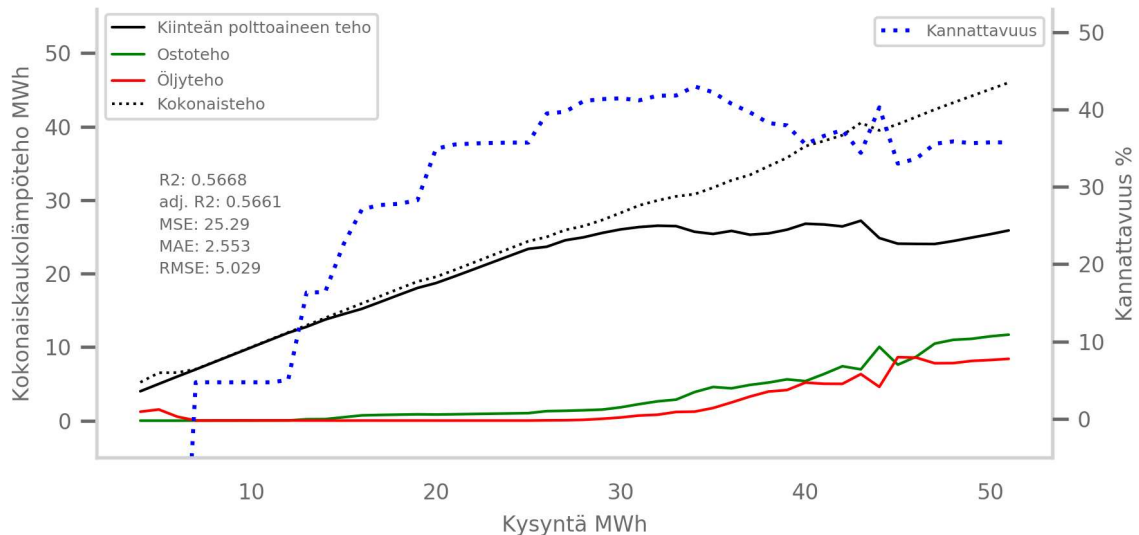
Tähän on syynä mallin tekemä yleistys, jossa se osoittaa vain keskimääräisen tehontuotannon havaintoarvojen välillä.



Kuvio 26. Polynomiregressiolla luotu malli ja tuotantokombinaatiot kysynnän suhteen

### 6.5.3 Mallin luonti satunnaismetsäregressiolla

Satunnaismetsäregressio on regressiotehtävissä käytettävä yhdistelmäoppimisen algoritmi, jossa luodaan useita päätöspuita satunnaisella opetusdatan otoksella. Opetusalgoritmillemme voidaan antaa parametreina generoitavien puiden lukumäärä sekä puiden maksimisyvyys. Sopivien parametrien löytäminen parhaimman mallin tuottamiseksi vaatii iteratiivista hakua, jota voidaan toteuttaa Scikit-Learn-kirjaston GridSearchCV-luokalla. Haku suoritetaan määritetyille opetusalgoritmin parametreille, jonka perusteella malli luodaan useita kertoja. Haun perusteella paras malli saatiin muodostettua, kun puiden lukumääräksi asetetaan 120 ja niiden maksimisyvyudeksi 8. Kuviossa 27 on esitetty satunnaismetsäregressiolla tuotetun mallin tuotantokombinaatiot ja kannattavuus kysynnän suhteen. Malli tuottaa hieman kulmikkaan lopputuloksen johtuen algoritmin puumaisesta rakenteesta.

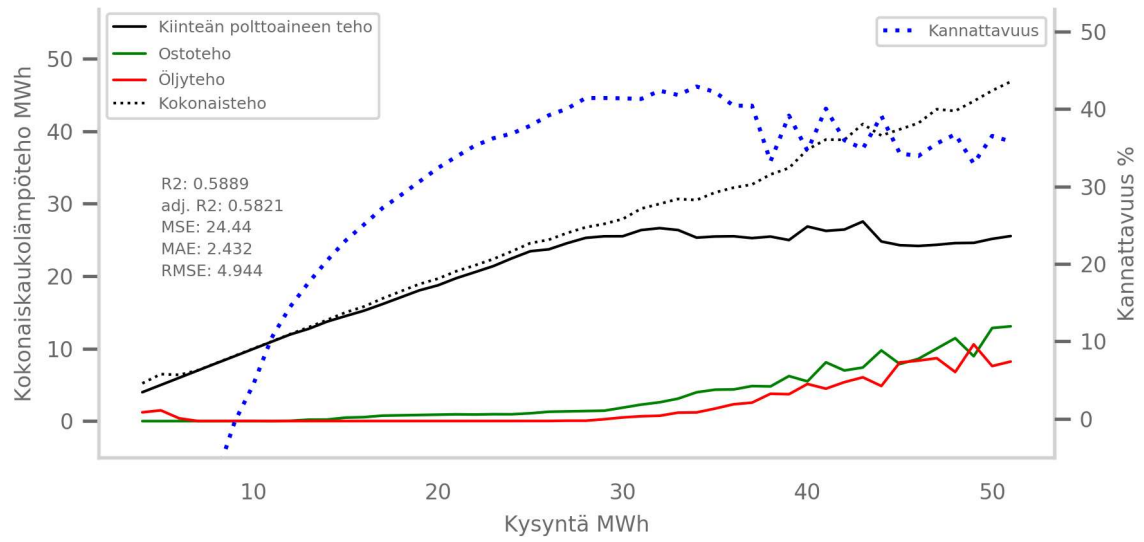


Kuvio 27. Satunnaismetsäregressiolla luotu malli ja tuotantokombinaatiot kysynnän suhteen

#### 6.5.4 Mallin luonti gradienttitehostamisregressiolla

Gradienttitehostamisregressio (engl. Gradient Boosting Regression) on yksi boosting-menetelmään perustuva yhdistelmäoppimisen algoritmi. Algoritmi tuottaa useita yksinkertaisia päätöspuita ja yhdistää niiden parhaimmat puolet yhdeksi tehokkaaksi malliksi. Algoritmi valitsee opettamiseen käytettävät muuttujat satunnaisesti jokaiselle generoitavalle päätöspuulle ja käyttää edellisen iteraation virheellisiä estimaatteja uuden iteraation opetusdatana (Géron 2017, 195). Scikit-Learn-kirjastosta löytyvälle GradientBoostingRegressor-luokalle voidaan määrittää muun muassa generoitavien puiden lukumäärä, niiden syvyys ja oppimisaste. Optimaalisten hyperparametrien arvojen löytäminen vaatii satunnaismetsäregressioon tavoin hakua. Määritettävien hyperparametrien määrä on suuri, joten optimaalisten parametrien löytäminen voi olla hankalaa ja syytä rajata vain muutama parametriin. Paras malli saavutettiin kuitenkin opetusalgoritmin parametrien oletusasetuksilla, jossa generoitavien puiden lukumäärä on 100, maksimisyvyys 3 ja oppimisaste 0,1.

Kuviossa 28 on havainnollistettu gradienttitehostamisregressiolla toteutetun mallin tuotantokombinaatiot ja kannattavuus kysynnän suhteen. Kuviosta voidaan päätellä, että malli tuottaa vakaan tuloksen, kun kysyntä on suurin piirtein alle 35 MW. Datassa on havaittavissa hajontaa suurimmilla tehoilla, jolloin myös osto- ja öljytehon osuus on kasvussa. Tunnuslukujen perusteella malli on tarkkuudeltaan tähän mennessä paras, sillä selitysaste on aavistuksen parempi ja mallin virhe muihin malleihin nähden pienempi.



Kuvio 28. Gradienttitehostamisregressiolla luotu malli ja tuotantokombinaatiot kysynnän suhteen

## 7 Raportointi- ja visualisointityökalut

Vuonna 2019 Kemin Energia ja Vesi Oy:n toteuttaman automaatiojärjestelmän uudistamisen yhteydessä kaukolämpötuotannon raportointitarpeita varten luotiin uusi, Microsoft Power BI –palvelua hyödyntävä raportointikonaisuus. Uudella raportointijärjestelmällä korvattiin aiempi, työlääksi ja hitaaksi koettu Excel-pohjainen raportointi. Power BI:n valintaa raportointityökaluksi puolsi myös se, että yhtiöllä oli käytössään muita Microsoft 365 –palveluita, joihin Power BI -palvelu oli helposti liitettävissä ja jonka myötä palvelu oli suoraan käytettävissä samalla Azure AD -kirjautumisella kuin muutkin käytössä olevat palvelut. Tästä syystä myös kaukolämpötuotannon optimointimallin tulosten visualisointi päätettiin kehittää samalle alustalle, jotta kaikki raportit ja visualisoinnit olisivat helposti tarkasteltavissa samasta palvelusta. Optimointimallin visualisoinnin ohella raportointia ja koontinäyttöjä laajennettiin kattamaan myös uuden KPA2-laitoksen mittauksien sekä polttoainetoimistusten raportoinnit.

### 7.1 Microsoft Power BI

Microsoftin dokumentaation mukaan Power BI on ”kokoelma ohjelmistopalveluja, sovelluksia ja yhdistimiä, jotka yhdessä muuntavat toisiinsa liittymättömistä lähteistä peräisin olevan tiedon johdonmukaisiksi, visuaalisesti vaikuttaviksi ja vuorovaikutteisiksi näkemyksiksi”. Power BI mahdollistaa datan noutamisen useista erilaisista lähteistä, kuten esimerkiksi tietokannoista, rajapinnoilta, tiedostoista tai suoraan verkkosivujen lähdekoodista. Noudetut tiedot kootaan Power BI:ssa tietomalliin, jossa data säilötään taulukkomuotoisena. Varsinaiset raporttinäkymät luodaan Power BI Desktop –sovelluksessa ja julkaistaan tämän jälkeen Power BI Service -verkkopalveluun, joka on verkkoselaimella käytettävä Saas-palvelu raporttien tarkasteluun, muokkaamiseen ja jakamiseen valituille käyttäjärhyhmille. Power BI Service mahdollistaa myös raporttien luomisen, mutta sillä ei ole mahdollista muokata datamallin tietoja tai datamuunnoksia Power BI Desktop –sovelluksen tapaan. Lisäksi Power BI Service mahdollistaa koontinäyttöjen luomisen, joihin voidaan upottaa sekä yksittäisiä visualisointeja että useammasta visualisoinnista koostuvia kokonaisuuksia olemassa olevista raporteista, tai luoda suoraan reaaliaikaisesta datalähteestä tietonsa saavia visualisointeja. (Microsoft 2021.)

Kemin Energian ja Veden tapauksessa hyödynnetään sekä Power BI Desktop –sovelluksella luotuja raportteja, että Power BI Service –verkkopalveluun luotuja reaaliaikaisia koontinäyttöjä. Pääosin taulukkopohjaiset raporttinäkymät mahdollistavat datan tarkastelun käyttäjän valintojen mukaisella tarkkuudella valitulta aikajaksolta, kuten esimerkiksi tuotantotietojen ja tuotannosta aiheutuneiden päästöjen tarkastelun tietyltä kuukaudesta.

Tietolähteenä taulukkopohjaisissa raporttinäkymissä toimii Azure SQL –tietokantapalvelin. Raportointinäkymien avulla tuotetaan automaattisesti myös viranomaisraportoinnin vaatimukset täyttävät päästöraportit. Taulukkopohjaisten raporttinäkymien ohella kokonaisuus sisältää myös reaaliaikaisia koontinäyttöjä, joissa esitetään tuotannon keskeisimpiä avainlukuja. Koontinäyttöjen data haetaan suoraan Power BI:hin määritetystä reaaliaikaisesta tietojoukosta, johon uutta mittaus- ja laskentatietoa saapuu minuutin väliajoin. Siinä missä SQL-tietokannan tauluihin tallennetaan useita satoja erilaisia mittauksia, reaaliaikänäyttöjen tietovarastoon tallennetaan vain sellaiset mittaustiedot, joita on tarve seurata reaaliaikaisesti.

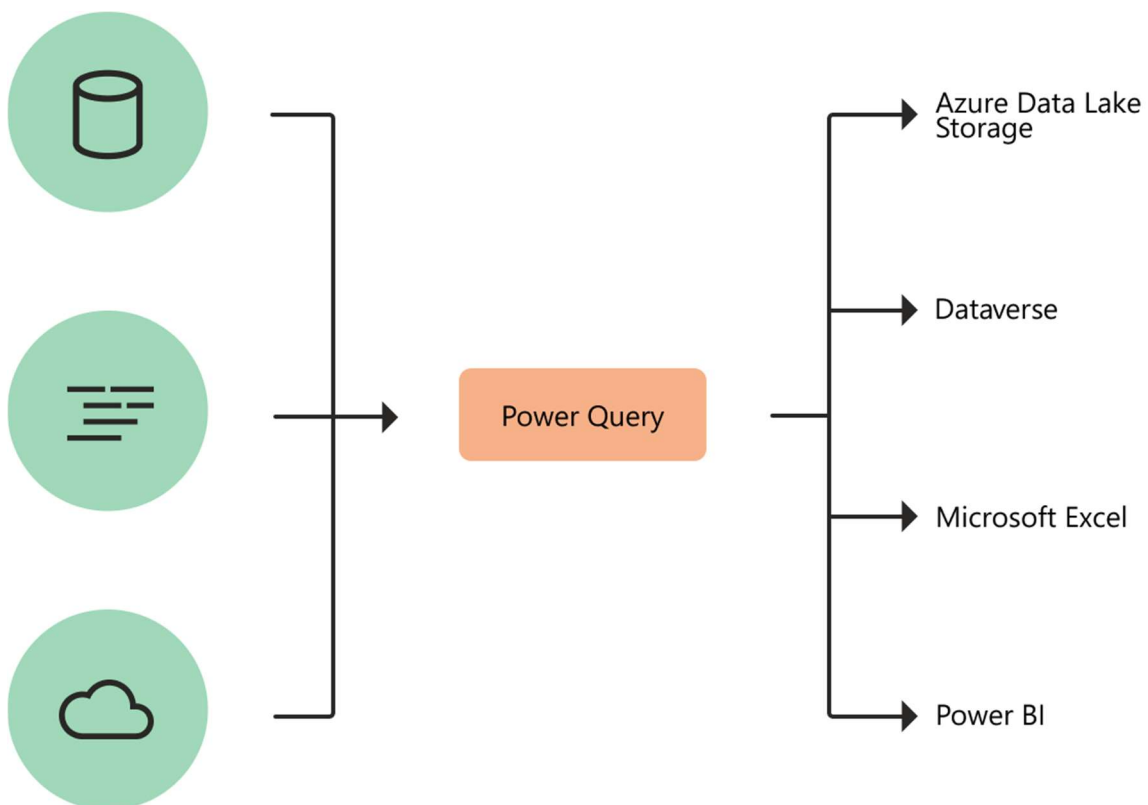
## 7.2 Muut visualisointityökalut

Lähtökohtaisesti Power BI:n oletettiin kattavan projektin raportointiin liittyvät erilaiset visualisointitarpeet, mutta projektissa päätettiin jättää mahdollisuus myös R-ohjelmointikielellä toteutettuja visualisointeja varten, mikäli Power BI:n oletuksena sisältämillä visualisoinneilla ei päästäisi haluttuun lopputulokseen. Power BI tukee suoraan R-visualisointien upottamista raportteihin, jonka myötä lopputuloksesta tulisi toteutustavasta riippumatta asiakkaan käyttökokemuksen kannalta yhtenäinen ja saumaton.

## 8 Tuotanto- ja optimointidatan käsittely visualisointia varten

### 8.1 Datan puhdistaminen ja muotoilu

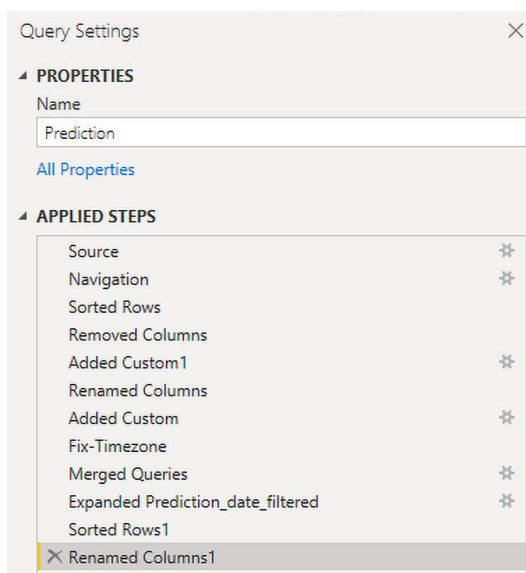
Datan visualisoinnin edellytyksenä on, että käytettävä data on riittävän laadukasta, eli se ei sisällä rakenteellisia ristiriitoja, puutteita tai epätarkkuuksia (Elgabry 2019). Lisäksi data tulee muotoilla valitun visualisointityökalun vaatimalla tavalla. Power BI:ssa datan muotoiluun käytetään Microsoftin Power Query -moduulia, joka mahdollistaa useiden eri datalähteiden yhdistämisen sekä datan käsittelemisen ja muotoilun. Power Query on käytettävissä useissa Microsoftin tuotteissa Excelin kaltaisista työpöytäsovelluksista aina Azure-pohjaisiin pilvipalveluihin (Microsoft 2020a).



Kuvio 29. Microsoft Power Query (Microsoft 2020a)

Tietojen käsittelyä varten luodaan Power Queryn omaa M-ohjelmointikieltä hyödyntäen sarja työvaiheita, joissa dataan kohdistetaan haluttuja toimenpiteitä, joita ovat esimerkiksi eri datalähteiden aikaleimakenttien yhtenäistäminen aikajanaavertailuja varten tai puutteellisen datan paikkaaminen joko laskennallisesti tai muiden sarakkeiden tietoja hyödyntäen. Power Queryn avulla dataan voidaan myös kohdistaa erilaisia matemaattisia toimituksia, kuten esimerkiksi normaalijakauman tai keskiarvon laskentoja.

Mikäli joltakin aikajaksolta ei ole saatavissa dataa, puuttuvien tietojen asiayhteys määrittää sen, millä tavalla ne käsitellään. Sellaisten mittausarvojen kohdalla, joita ei ole mahdollista muodostaa jälkikäteen, arvot voidaan jättää joko tyhjäksi tai nolaksi riippuen siitä, olisiko mahdollinen nolla-arvo lopputuloksen kannalta merkitsevä, tai että aiheuttaisiko nolla-arvo vääristymää esimerkiksi keskiarvoja laskettaessa. Joissakin tapauksissa voidaan käyttää arvojen täyttötoimintoa, jolloin puuttuviksi arvoiksi valitaan samasta sarakkeesta joko seuraava tai edellinen, tyhjistä poikkeava arvo. Kuvassa 3 on nähtävissä tyypillinen Power Query -työjono, jossa määritetään erilaisia datan siivoamiseen ja yhdistämiseen liittyviä työvaiheita. Power BI:n visualisointityökalujen tapaan myös datan muunnosvaiheessa voidaan hyödyntää R-ohjelmointikielellä toteutettuja skriptejä datan puhdistamiseen, muotoiluun tai muuhun käsittelyyn.



Kuva 3. Lista Power Query -työvaiheista

Datan muunnosvaiheessa Power BI pyrkii tunnistamaan datan tietotyyppin ja esitysmuodon automaattisesti. Mikäli tietolähde tarjoaa rakenteellista dataa, kuten esimerkiksi tietokannasta tietoja noudettaessa, Power BI määrittää datatyyppit suoraan tietokannan taulurakenteen perusteella. Jos datalähteenä käytetään tekstipohjaista tietoa, joka ei sisällä toistuvaa rakennetta, pyrkii Power BI päättämään sarakkeiden tietotyyppit niiden sisällön perusteella. Tarvittaessa käyttäjällä on mahdollisuus yliajaa automaattisesti asetetut tyyppit. Tietotyyppin muutos saattaa olla tarpeellista esimerkiksi tilanteissa, joissa tekstipohjaisesta tiedosta etsitään tietyssä muodossa olevaa päivämäärää (Microsoft 2020b).

Kemin Energia ja Vesi Oy:n käyttämä automaatiojärjestelmä sisältää useita erilaisia fyysisiä mittarilaitteita ja niiden myötä ison määrän erilaisia mitattavia suureita, joiden vuoksi dataan

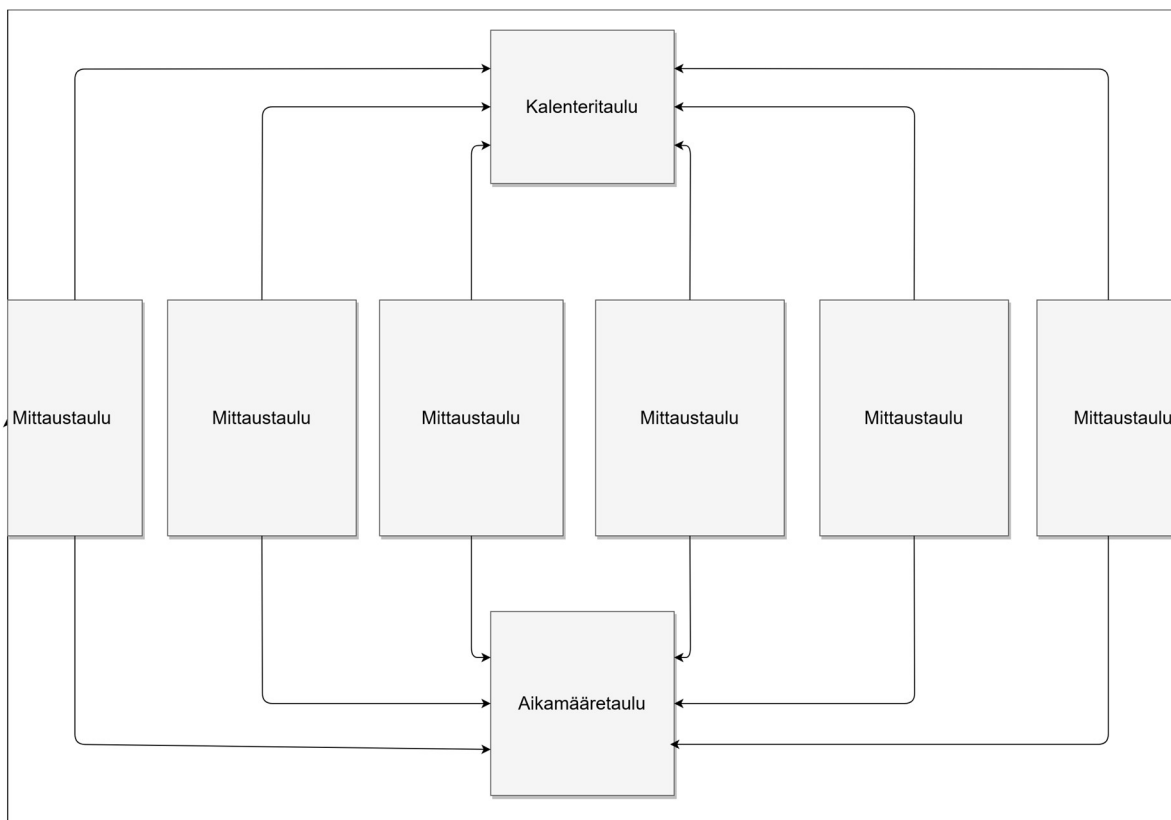


joudutaan tekemään useita yksikkömuunnoksia sekä erilaisia laskentoja. Tyypillisimpiä yksikkömuunnoksia ovat esimerkiksi jouleina mitatun lämpöenergian muuttaminen wattitunneiksi sekä kulutuslaskennat, joissa verrataan mittauksen uusinta arvoa edelliseen arvoon ja lasketaan näiden välinen erotus. Kaukolämpötuotannon päästölaskentojen osalta kunkin tuotantomuodon päästöt lasketaan erikseen sen perusteella, minkälaisia polttoainekohtaisia kertoimia laskenta vaatii.

## 8.2 Tiedon väliset suhteet ja linkitys

Power BI:n raporttinäkymissä visualisoitavaan dataan tulee määritellä tietojen välisiä suhteita, joiden avulla dataa voidaan esittää samanaikaisesti useasta eri taulusta Power BI:n tietomallin sisällä. Mikäli tallennettu tieto on aikaan sidonnaista ja se sisältää aikaleiman, on eri taulujen sisältämiä tietoja mielekästä linkittää aikaleimojen perusteella. Aikaleimojen linkityksen avulla eri taulujen tietoja voidaan esittää ajallisesti rinnakkain sekä suodattaa tietoja esimerkiksi datatyyppin tai aikarajauksen mukaisesti.

Kemin Energia ja Vesi Oy:n raportoinnin tapauksessa tietomallin rakenne on melko yksinkertainen, koska käytännössä lähes kaikki tietomallin taulut sisältävät vain mittaustietoja, joilla ei ole suoraa yhteyttä toisiinsa, mutta jotka voidaan linkittää toisiinsa mittaustietojen sisältämien aikaleimojen perusteella. Kuvassa 4 on nähtävissä osa taulujen välisen linkityksen toteutuksesta. Linkitystä varten Power BI:n tietomalliin luotiin kalenteritaulu, joka linkityy kunkin mittaustaulun aikaleimakenttään. Linkityksen avulla raportteihin voitiin luoda päivämäärävalitsin, joka mahdollistaa tietojen näyttämisen tietyltä aikaväliltä. Lisäksi tietomalliin luotiin aikamääretaulu, jonka perusteella käyttäjä voi valita, tarkastellaanko dataa tunti-, päivä- vai kuukausitasolla.



Kuva 4. Power BI:n tietomallissa datataulut ovat linkitetty sekä kalenteritauluun että aikamääretauluun.

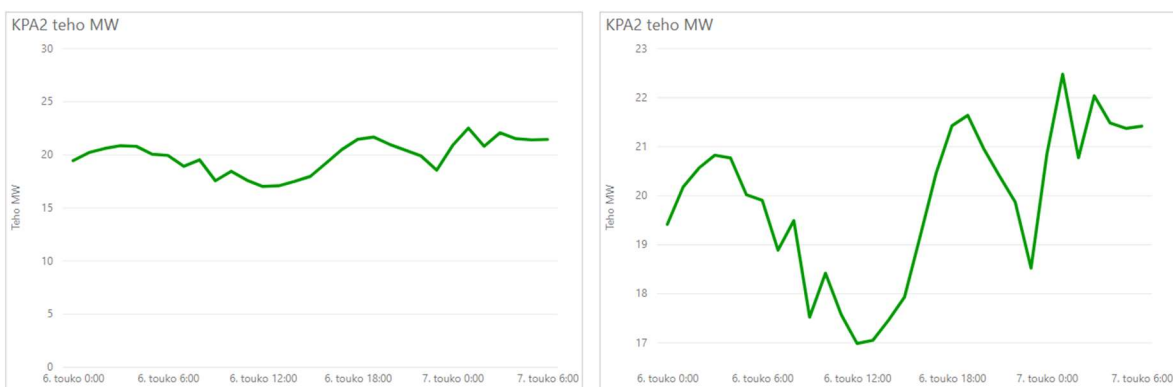
Raportointijärjestelmän reaaliaikaisille koantinäyttöille tuotava data visualisoidaan lähes sellaisenaan, kuin se saapuu Power BI:n reaaliaikaiseen tietojoukkoon. Kaukolämpötuotannon eri tuotantomuotojen tehoihin sekä säätietoihin liittyvistä mittausarvoista näytettäväksi valitaan viimeisin saapunut arvo, joka pyöristetään mitattavan suureen kannalta sopivaan tarkkuuteen. Osalle reaaliaikanäyttöillä näytettävistä arvoista suoritetaan joitakin laskentoja, kuten esimerkiksi vanhemman KPA1-laitoksen kohdalla, jossa mittauksista lasketaan osuudet hiilineutraalin ja fossiilisten polttoaineiden välillä, joiden avulla visualisoidaan tuotannon ekologisuutta.

## 9 Datan visualisointimenetelmät ja visualisoinnin tavoitteet

### 9.1 Datan visualisointi

Datan visualisoinnilla tarkoitetaan yleisesti visuaalisten keinojen hyödyntämistä siinä, että esimerkiksi tietokantaan tallennettu numeerinen data saadaan muutettua helposti ymmärrettävään ja mahdollisimman yksiselitteisesti tulkittavaan visuaaliseen muotoon (Ware 2021, 3). Robert Kosara toteaa julkaisussaan (Kosara 2007, 2), että datan visualisoinnille ei ole olemassa yksittäistä, yleisesti hyväksyttyä määritelmää ja että merkitys vaihtelee paljon asiayhteydestä ja toimialasta riippuen. Kosara ehdottaa, että visualisoinnin määritelmän täyttämiseksi sen tulee toteuttaa kolme ehtoa; visualisoinnin tulee perustua ei-visuaaliseen dataan, lopputuloksena tulee syntyä nähtävissä oleva kuva ja että tuotetusta kuvasta tulee olla selkeäsi luettavissa ja ymmärrettävissä sen sisältämä informaatio.

Siinä missä pelkkien lukujen tarkastelu sinänsä antaa faktatietoa datan taustalla olevista asioista ja ilmiöistä, on kokonais kuvan tai trendien hahmottaminen pelkästä numeerisesta datasta haastavaa. Visuaaliseen muotoon muunnetusta datasta voidaan helpommin tulkita datan sisältämät suhteet, muutosten suunnat tai datan sisältämät poikkeamat. Sopivan visualisointitavan valinnalla voidaan suuriakin tietomääriä esittää pienessä tilassa ja helposti ymmärrettävällä tavalla. Toisaalta valitsemalla epäolennaista dataa tai vääränlaisia visualisointitapoja voidaan visualisoinnilla joko vahingossa tai tarkoitushakuisesti tuottaa myös harhaanjohtavia mielikuvia, joissa olennainen data on hävitetty epäolennaisen sekaan; kuvioista 30 on nähtävissä, miten kuvaajien numeeristen asteikoiden aloituskohdan valinnalla voidaan merkittävästi vaikuttaa siihen, vaikuttaako kuvaajasta tulkittava trendi tai muutos merkittävältä vai ei.



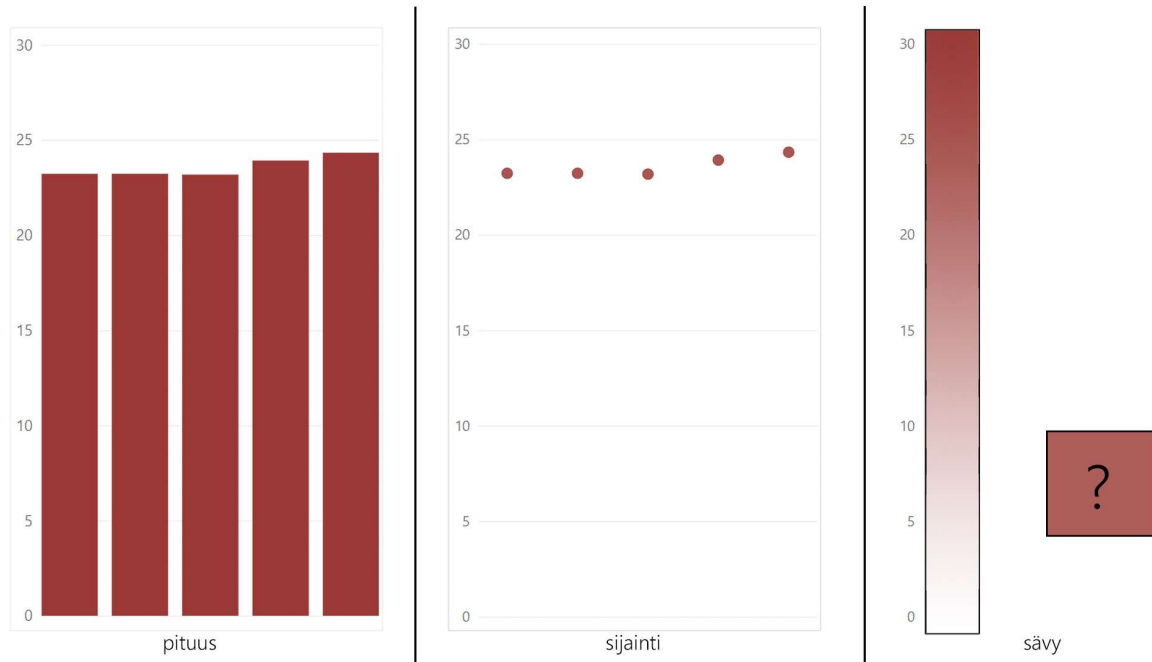
Kuvio 30. Skaalan valinnalla voidaan merkittävästi vaikuttaa kuvaajan tarjoamaan vaikutelmaan, vaikka taustalla oleva data olisikin täysin identtistä

Dataa visualisoitaessa tulisi aina pyrkiä valitsemaan selkein saatavilla oleva ja datan muotoon sopiva esitystapa. Esitettävän datan tulee sisältää vain kokonaisuuden kannalta olennaista tietoa, jotta voitaisiin välttää mahdolliset väärät tulkinnat. Lähtökohtaisesti datan visualisointi tulisi toteuttaa siten, että käyttäjän huomio saadaan mahdollisimman nopeasti kiinnittymään kaikkein tärkeimpiin tietoihin (Ware 2021, 14). Visualisoinnin tulisi sisältää mahdollisimman paljon saatavilla olevaa olennaista tietoa, kuitenkin tekemättä kokonaisuudesta liian sekavaa. Hyvin toteutetulla visualisoinnilla voidaan antaa käyttäjälle nopeasti tulkittava yleiskuva, mutta samalla tarjota myös mahdollisuus tutkia tietoja tarkemmin. (Koponen & Hildén 2019, 43)

Erilaisten visualisointitekniikoiden käytöllä itsessään ei ole minkäänlaista itseisarvoa, vaan niiden käyttäminen saa merkityksensä visualisoinneilla tuotettujen tulosten perusteella. Mikäli visualisointi ei tuota minkäänlaista lisäarvoa datan tulkittavuuteen tai jos tieto voidaan esittää selkeämmin pelkällä tekstillä, tulisi visualisoinnin käyttöä välttää (Ware 2021, 28).

## 9.2 Visualisointitavat ja värien käyttö

Dataa visualisoitaessa tulisi valita tarkoituksenmukainen visualisointitapa sen mukaan, mitä datasta halutaan tuoda esille. Koposen ja Hildénin (2019, 62) mukaan tapoja, joilla tieto muutetaan visuaaliseen muotoon, voidaan kutsua visuaalisiksi muuttujiksi. Mikäli ollaan esittämässä määrällistä dataa, jossa datan sisältämä numeerinen arvo on olennaisin tieto, tulisi käyttää visualisointityyppejä, joissa merkittävyys esitetään joko dataa kuvaavan muodon pituutta tai sijaintia käyttäen, kuten esimerkiksi pylväs-, sironta- tai viivakaaviossa. Jos taas halutaan esittää tiedon sisältämien arvojen järjestystä, kaikkein havainnollisimpia tapoja ovat määrällisen tiedon tapaan tiedon sijainti, mutta myös järjestyksen esittäminen yksittäisen värin kirkkauskaalan avulla, jossa värin kirkkausasteista käy ilmi tiedon sisältämien arvojen järjestys. Tiedon ollessa kategorisoitavissa erilaisiin luokkiin, voidaan eri luokat esittää joko sijaintia, muotoja tai eri värejä hyödyntäen. Kuviosta 31 on nähtävissä, miten sekä pylväs- että sirontakaaviosta on helpohkosti pääteltävissä kuvaajien esittämät arvot, mutta numeroarvon päättelyminen väriliu'usta on huomattavan vaikeaa ja epätarkkaa.



Kuvio 31. Numeerisen arvon esitys erilaisilla visualisointitavoilla

Käytettäessä värejä datan luokitteluun, tulisi huomioida visualisoinnin tulkitsijan hahmotuskyvyn rajallisuus sen suhteen, montako eri väriä keskivertoihminen pystyy hahmottamaan kerralla. Eri kirkkaussävyjen käytettävissä oleva määrä on vahvasti sidonnainen siihen, mistä väristä on kyse; esimerkiksi värikylläisestä keltaisesta on helpohkosti erotettavissa jopa kymmenen eri kirkkausastetta, kun taas värikylläisyydeltään vastaavasta violetin sävystä on vaikeaa erottaa yli kolmea eri kirkkausasetta. Mikäli tietoa visualisoidaan kategorisoidusti ja kategorioiden erottamiseen toisistaan halutaan käyttää eri värejä, tulisi käytettävien värien määrä rajata maksimissaan kymmeneen erilaiseen sävyyn, koska sitä suuremmalla värimäärällä eroteltua dataa on jo vaikea hahmottaa samasta kaaviosta (Ware 2021, 124). Vastaavasti jos valittujen värisävyjen määrä on pieni, mutta eri kirkkausasteiden määrä suurehko, käytettävissä olevat värisävyt alkavat olla jo niin lähellä toisiaan, että niitä on vaikea erottaa toisistaan. Suurempien määrien kategorisoinnissa olisi syytä käyttää muita visuaalisia tapoja erotella kategoriat, kuten esimerkiksi visuaalisten muotojen tai tekstilyhenteiden käyttö (Koponen & Hildén 2019, 74). Värien käytön suhteen tulisi huomioida myös värien erottuvuus toisistaan tapauksissa, joissa visualisointien hyödyntäjillä on värinäön heikkouksia. Miehistä jopa 8 % ja naisista 0,5 % kärsii tyypillisimmistä värinäön heikkouksista, kuten puna-viher- tai viher-punaheikkoudesta (Duodecim Terveyskirjasto 2020).

Visualisoinneissa käytettäviä värejä valitessa tulee huomioida, että niihin saattaa liittyä toimialaan tai toimintaympäristöön liittyviä oletuksia, joilla voi olla merkittäviäkin vaikutuksia visualisointien tulkintaan. Siinä missä esimerkiksi laadunvalvonnan ammattilaiselle vihreä väri tarkoittaa hyväksyttyä, lääkärille sama väri tarkoittaa tartunnan saanutta tapausta.

Alakohtaisten miellelyhtymien lisäksi myös kulttuurilla saattaa olla vaikutusta värien tulkin-tatapaan ja datan visualisointia tehtäessä on syytä olla tietoinen, mikäli joihinkin väreihin liittyy esimerkiksi erityisen vahva poliittinen lataus (Koponen & Hildén 2019, 73). Toisaalta myös itse visualisoitava data asettaa rajoja tulkintojen tekemiselle, ja kulttuurisidonnaisuuksien kanssa suurimmat riskit tulevatkin todennäköisesti tilanteissa, joissa visualisoidaan esimerkiksi mielipidemittauksien dataa. Sen sijaan teollista mittausdataa visualisoitaessa on turvallisempaa olettaa, että kuvaajissa käytetyistä väreistä ei aleta etsiä piilomerkityksiä.

### 9.3 Yleiset visualisoinnin tavoitteet

Tuotanto-, ennuste- ja optimointidatan visualisoinnin päätavoite on tukea Kemin Energia ja Vesi Oy:n kaukolämpötuotannon ohjaamista tiedolla johtamisen keinoin. Päätöksenteon tu-eksi toteutettavien visualisointien tarkoituksena on tuottaa selkeä tilannekuva koko tuotan-totilanteesta, jonka avulla voidaan seurata reaaliaikaisesti tuotannon avainlukuja ja rea-goida muutoksiin ennakoivasti. Reaaliaikaisen tiedon ohella päätöksentekoa pyritään hel-pottamaan visualisoimalla tuotannon tarpeen ennuste. Tilannekuvan ohella erilaiset histo-riadatan visualisoinnit tarjoavat mahdollisuuden tarkastella jo tapahtuneita tuotannon tilan-teita ja sitä mukaa pyrkiä kehittämään tuotannon ohjausta tehokkaampaan ja ekologisem-paan suuntaan.

### 9.4 Tuotantodatan visualisoinnin tavoitteet

Tärkein tavoite tuotantodatan visualisoinnin osalta on esittää tuotannon kannalta olennai-simmat mittaukset ja laskennat mahdollisimman yksiselitteisesti ja selkeästi siten, että nii-den perusteella voidaan suorittaa kaukolämpötuotannon ajoon liittyviä päätöksiä ja toimen-piteitä. Tuotannon tilanteen seurannan kannalta keskeisimmät mittarit ovat eri lämmöntuo-tantotapojen reaaliaikaiset tehomittarit, sekä sääennusteeseen perustuva, koneoppimisen avulla luotu ennuste tuotannon tarpeesta. Ennusteen tuottaman datan avulla pyritään pää-semään koko kaukolämpötuotannon ekologisuuden kannalta kaikkein keskeisimpään ta-voitteeseen, eli öljyn käytön minimointiin. Ideaalitulanteessa ennusteen avulla voitaisiin ha-vaita tulevan säätilan muutoksen aiheuttama äkillinen nousu tuotannon tarpeessa, jolloin kiinteällä polttoaineella tuotetun lämpötehon ennakoidulla nostolla voitaisiin välttää öljyä polttoaineenaan käyttävien laitosten käynnistäminen. Ennakoivalla öljyn käytön välttämi-sellä on saavutettavissa myös merkittäviä taloudellisia säästöjä, koska öljy kuuluu uusiutu-mattomuutensa myötä päästökaupan piiriin ja sen käytöstä aiheutuu merkittäviä lisäkustan-nuksia.

## 9.5 Optimointi- ja kannattavuusdatan visualisoinnin tavoitteet

Talousdatan visualisoinnin tärkeimmäksi tavoitteeksi asetettiin asiakkaan ideoima yksinkertainen kannattavuusmittari, josta olisi nähtävissä reaaliaikainen tai korkeintaan muutaman tunnin viipeellä päivittyvä tilanne kaukolämmön tuotannon aiheuttamista kuluista suhteessa saavutettuun taloudelliseen tuottoon. Yksittäinen ja selkeä talousmittari olisi helposti tulkittava ja yksiselitteinen kuvaus siitä, miten sen hetkinen tuotantotapa vaikuttaa kokonaistaloudellisesti. Yksittäisen talousmittarin lisäksi voittoa tai tappiota esittävä data tulisi voida esittää myös pidemmältä aikajaksolta, jolloin pystyttäisiin tekemään vertailuja erilaisten tuotannon tilanteiden välillä ja siten havainnollistamaan esimerkiksi säätilan muutosten vaikutusta tuotannon tarpeeseen ja sitä mukaa kannattavuuteen. Lisäksi kannattavuusdatasta olisi tarkasteltavissa laitosten huoltokatkosten vaikutus taloudellisesti, etenkin jos vikatilanne tai suunniteltu huolto on syystä tai toisesta pitkittynyt. Vastaavasti kannattavuuden kuvaajasta olisi selkeästi nähtävissä öljylaitosten käytön negatiivinen vaikutus kannattavuuteen.

Muita tavoitteita optimointidatan visualisoinnin suhteen on pyrkiä luomaan erilaisia näkymiä ja kuvaajia, joista on tarkasteltavissa todellisen tuotantotilanteen ja optimaalisen tilanteen eroavaisuudet. Yhdistämällä tuotantoon liittyvää talousdataa sekä tuotannon poikkeavuutta kuvaavaa dataa, voidaan esittää minkälaisia taloudellisia vaikutuksia tuotannon optimoinnilla voisi olla, sekä esittää rinnakkain ideaalitalanteen taloudellinen tuotto ja toteutuneesta tuotannosta saavutettu tuotto. Myös talousdataan tulisi voida liittää kaukolämmön tuotantotarpeen ennuste, jonka avulla voitaisiin esittää arvioita kannattavuuden muutoksista tulevina päivinä. Lisäksi talous- ja ennustedataa visualisoimalla voitaisiin helposti havainnollistaa se, millä tavalla mikäkin tuotantotapa vaikuttaa kannattavuuteen lähitulevaisuudessa.

## 10 Tuotantodatan visualisointi

Eri lämmöntuotantomuotojen teholukemien visualisoinnissa päädyttiin käyttämään jo aiemmin käytössä ollutta, asiakkaan kanssa yhdessä valittua väripalettia, jossa eri tuotantomuodoille oli valittu tuotantomuotoa kuvaava väri. Uusiutuvaa, kiinteätä polttoainetta käyttävien laitosten tehomittaukset visualisoidaan laitoskohtaisilla vihreän eri sävyillä ja vastaavasti laitosten lämmön talteenotto visualisoidaan turkoosin eri sävyillä. Asiakkaan toiveesta öljytehon visualisointiin haluttiin saada mahdollisimman tehokkaasti huomiota herättävä väri, joten sen osalta päädyttiin kirkkaan punaiseen sävyyn. Lähinnä säätötehona käytettävää ostotehoa päätettiin visualisoida sinisellä sävyllä. Tuotannon tarpeen ennusteen väriksi valittiin muista käytetyistä värisävyistä riittävästi erottuva oranssin sävy. Visualisointien väripalettiin oli jo aiemmin valittu lähtökohtaisesti mahdollisimman helposti toisistaan erottuvat sävyt ja uudistusten myötä sävyvalintoja säädettiin vielä siten, että mittarien värit erottuvat toisistaan myös tapauksissa, joissa katsojalla on mahdollisia värinäön heikkouksia (Koponen & Hildén 2019, 74).

### 10.1 Reaaliaikainen tuotantonäkymä

Reaaliaikaisen tuotantonäkymän päätavoitteena oli jo vuoden 2019 uudistuksessa koota kaikki olennainen tieto yhdelle koontinäytölle siten, että käyttäjän ei tarvitsisi vaihtaa näkymää eri raporttien välillä, vaan että keskeisimmät tiedot ovat nähtävillä jatkuvasti. Perusteena tietojen kokoamiselle nähtäväksi yhdellä kertaa oli myös se, että koontinäyttöjä saatetaan käyttää tiloissa, joissa näyttöä ohjaavaa tietokonetta ei ole tarkoitus käyttää mihinkään muuhun kuin koontinäytön esittämiseen, eikä käytössä ole välttämättä mitään tietokoneen hallintalaitteita.

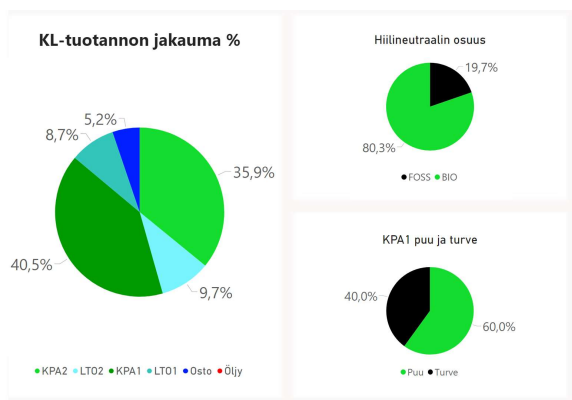
Koontinäyttöjen suunnittelussa pyrittiin siihen, että esitettävät tiedot ovat tarkoituksenmukaisia ja vastaavat käyttäjien toiveisiin datan esittämisen suhteen. Sisällöt on pyritty jaottelamaan loogisiin kokonaisuuksiin ja värien käyttö on rajoitettu itse kuvaajien sisältöön, jolloin graafinen ulkoasu pysyy yksinkertaisena, eikä vie huomiota olennaiselta sisällöltä. Myös eri visualisointien vaatima tila näytöltä on pyritty pitämään minimissään, samalla kuitenkin korostaen olennaisimpia tietoja antamalla niille enemmän näkyvyyttä. (Yin ym. 2014, 51.)





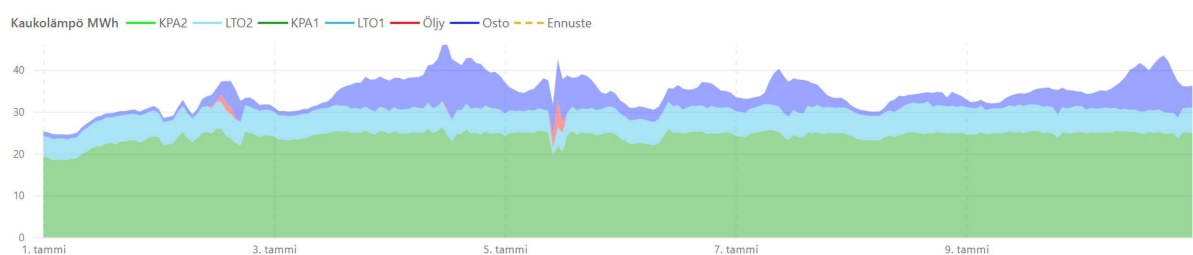
Kuva 5. Kuvakaappaus reaaliaikaisesta tuotantonäkymästä maaliskuulta 2021, jolloin KPA2-laitos otettiin käyttöön

Kuvassa 5 näkyvä tuotantonäkymä koostuu sekä reaaliaikaisista mittareista että historiatietoa näyttävistä kuvaajista. Näytön alalaitaan on koottu mittareita reaaliaikaisista tuotantolukemista, joista on nähtävissä kokonaistuotannon ja ulkolämpötilan ohella kunkin eri tuotantomuodon tuottama lämpöteho. Mittarityyppisten visualisointien skaalat ovat suhteutettu kunkin tuotantotavan maksimitehoon megawatteina. Näin ollen mittarien esittämä teho, eli mittarin värillisen osan koko, ei ole suoraan verrattavissa eri mittarien kesken, vaan jokaisesta mittarista käy ilmi kunkin tuotetun tehon suhde sen potentiaaliseen maksimitehoon. Jokaisessa tehomittarissa hetkellinen teho esitetään myös isokokoisena numerona mittarin keskellä, jolloin käsityksen eri tuotantomääristä saa helposti nopealla vilkaisulla. Lisäksi näytön alalaidassa on kaksi ympyrädiagrammia, joista käy ilmi kiinteän polttoaineen laitosten tehon ja lämmön talteenoton osuudet. Kuva 6 sisältää näkymän oikeaan laitaan koottuja sektoridiagrammeja, joista isoimmasta on nähtävissä eri lämmöntuotantomuotojen osuudet tuotetusta kokonaistehosta, lämmön talteenotto mukaan lukien. Lisäksi kuvaajista käy ilmi KPA1-laitoksen puun ja turpeen käytön välinen suhde, sekä tuotannon hiilineutraaliutta kuvaava diagrammi uusiutuvan bioenergian ja fossiilisen energian osuuksien välillä.

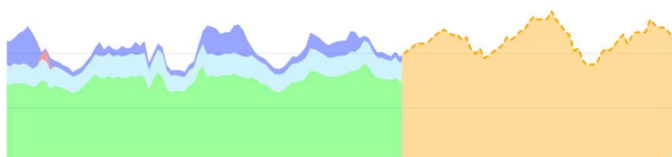


Kuva 6. Kuvakaappaus tuotantotapojen jakaumasta tuotantomuodoittain sekä tuotannon ekologisuutta kuvaavat diagrammit

Reaaliaikaisten mittarien ohella näytöllä ylimpänä oleva pinottu aluekaavio esittää historia-tietoa eri tuotantomuotojen tuottamasta kaukolämpötehosta. Kuvassa 7 näkyvässä kaaviossa näytettävät tiedot päivittyvät kolmen tunnin välein ja kaavio sisältää päivämäärärajaustyökalun halutun aikajakson tarkastelua varten. Tarkimmillaan dataa voidaan tarkastella kaaviosta tuntitasolla. Mikäli tarkasteluun valittu aikajakso käsittää nykyhetken, mitaustiedon jatkeena näytetään kuvassa 8 esitetty ennuste tuotannon tarpeesta. Ennuste ei ota kantaa siihen, millä tuotantotavalla tarvittava teho tulisi tuottaa, koska ennustetta muodostettaessa ei ole saatavilla tarkkoja tietoja esimerkiksi ostotehon saatavilla olevasta määrästä tai eri laitosten tilapäisistä tuotannon keskeytyksistä.

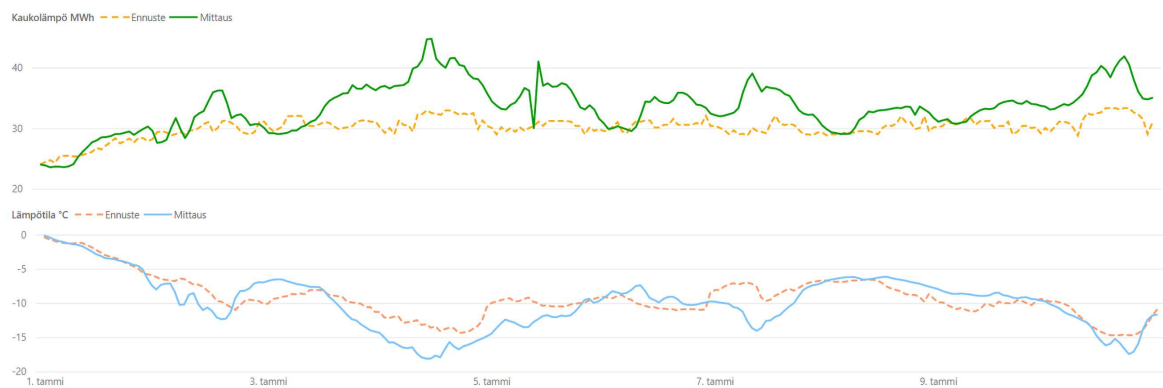


Kuva 7. Aluekaavio lämpöenergian tuotannosta tuotantomuodoittain



Kuva 8. Tuotantojakauma ja tuotannon tarpeen ennuste

Kuvassa 9 on nähtävillä koontinäytön tuotantojakauman alapuolella olevat kaksi viivakaaviota, joista ylempi kuvaa tuotantoennusteen osuvuutta suhteessa tuotettuun tehoon. Alempi kuvaaja esittää lämpötilaennustetta ja mitattua ulkolämpötilaa. Alun perin kunkin mittaus- ja ennustekaavion kohdalla oli tarkoitus näyttää rinnakkain kaksi erillistä viivakaaviota, joista toinen esittää mittaus- ja ennustearvot nykyhetkeen saakka ja toinen tulevan ennusteen, mutta lopulta mittaus- ja ennustedatat päädyttiin yhdistämään yksittäisiin viivakuvaajiin, koska Power BI:n reaaliaikanaäytöillä eri visualisointien skaalojen synkronoiminen todettiin haastavaksi tai senhetkisillä menetelmillä jopa mahdottomaksi. Erillisten kaavioiden käyttö olisi mahdollistanut esimerkiksi aikajakson historiatiedon tutkimisen samanaikaisesti ennusteen näyttämisen ohella, joka olisi mahdollistanut enemmän erilaisia vertailuja.



Kuva 9. Ruutukaappaus ennusteen osuvuutta kuvaavista viivadiagrammeista

Kuvassa 5 koontinäytön oikeassa yläkulmassa oleva pylväsdiagrammi esittää kuluneen viikon aikana saapuneiden polttoaineiden ja muiden tuotannossa tarvittavien materiaalien kuormatoimitusten jakauman. Kuvaajasta käy ilmi myös lähtevien kuormien, eli kaukolämpötuotannon tapauksessa tuhkakuormien määrä.

Kaikissa tuotantonäkymän kuvaajissa on käytössä kuvaajien päälle aukeavat työkaluvihjeet, jotka aukeavat, kun hiiren kursori viedään kuvaajan päälle. Työkaluvihjeistä on nähtävissä kunkin visualisoinnin esitystavasta ja visualisoitavasta datasta riippuen olennaisimmat avainluvut.

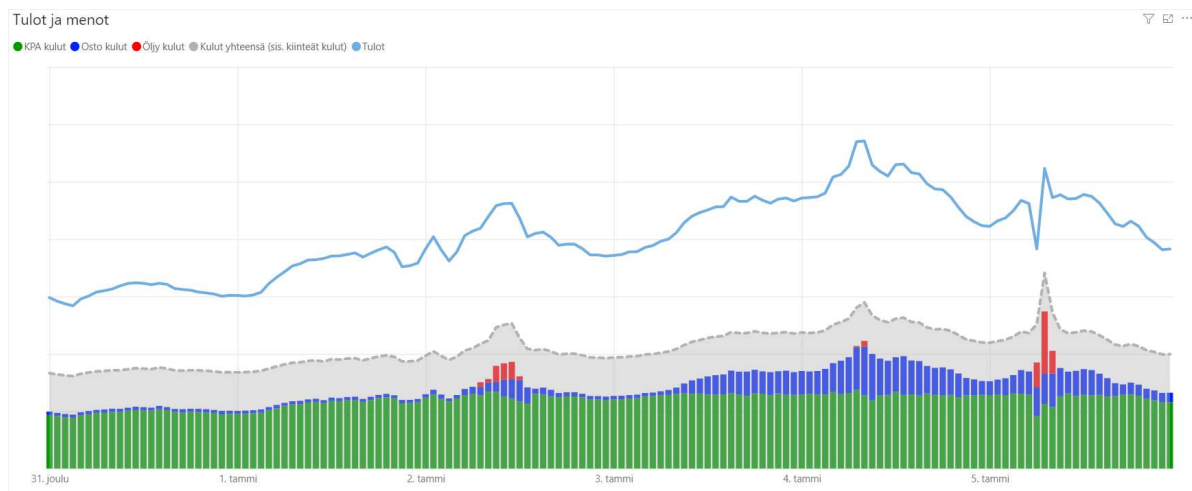
## 10.2 Optimointidatan visualisointi

Kaukolämpötuotannon optimointimallin osalta oli alun perin tarkoitus käyttää joko reaaliaikaista tai lähes reaaliaikaista, todellista kulutusdataa ja yhdistää siihen ajallisesti vastaavaa talousdataa. Haluttuja tietoa ei kuitenkaan ollut saatavilla tämän työn aikataulun puitteissa, joten optimointi jouduttiin tekemään huomattavasti suunniteltua pienemmällä ja

pelkistetymmällä datalla. Tästä johtuen optimointimallilla tuotetut johtopäätökset eivät välttämättä ole sellaisenaan täysin reaali maailmaa vastaavia, eikä esimerkiksi visualisointien tulkitseminen ole todellisessa ympäristössä aivan niin yksiselitteistä, kuin mitä se on käytössä olleella, rajoitetummalla datalla.

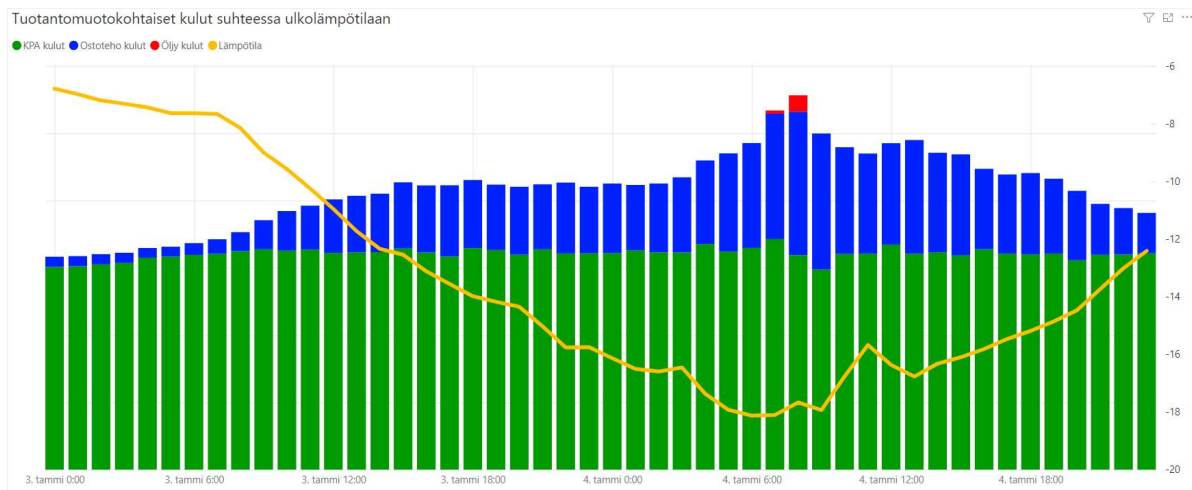
### 10.2.1 Tulojen ja menojen visualisointi

Optimoinnin tulosten tulkitsemisen kannalta on olennaista tietää lähtötilanne, johon optimoinnin tuottamia lukemia verrataan. Tästä johtuen on syytä visualisoida tuotannon kannalta olennaisimmat talousluvut, kuten kaukolämpötoiminnan tuotot, tuotantotapakohtaiset kulut sekä kiinteät kulut, kuten laitosten ylläpitokulut ja henkilöstön palkkakulut. Kuviossa 32 tulojen ja menojen visualisoinnin esitystapana toimii yhdistetty palkki- ja viivadiagrammi, josta käy ilmi kaukolämpötoiminnan tuottojen ja kulujen välinen suhde, sekä tuotantotapakohtaiset kulut selkeästi eriteltynä. Eri tuotantomuotojen visualisoinnissa käytetään samankaltaista väriskaalaa kuin koontinäyttöjen kuvaajissa. Todellisessa käyttöympäristössä kuvaajassa esitetään myös talouslukuja kuvaaja y-akselilla.



Kuvio 32. Eri tuotantomuotojen kulut suhteessa tuloihin

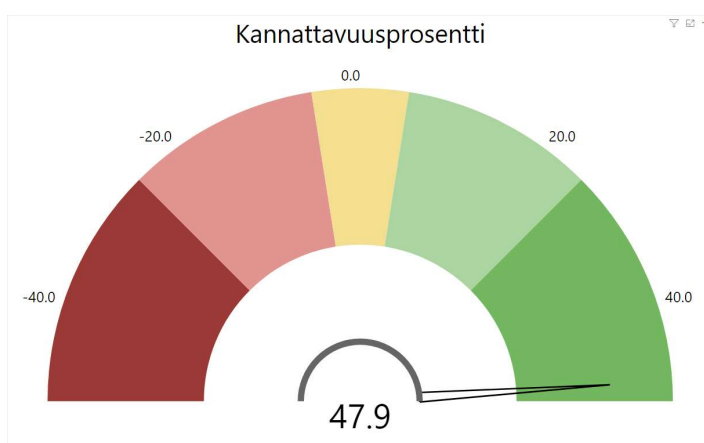
Tuotannon kulujakauma on mielekästä esittää myös suhteessa ulkolämpötilaan, koska lämpötila on tuotannon tarpeen kannalta merkittävin vaikuttava tekijä. Esitystapana on tulojen ja menojen tapaan palkki- ja viivadiagrammi, jossa eri tuotantomuotojen kulut visualisoidaan palkkeina ja ulkolämpötila viivana. Kuviossa 33 käy selkeästi ilmi lämpötilan vaikutus tuotantokuluihin, etenkin kun lämpötila kiristyy lähelle 20:ta pakkasastetta.



Kuvio 33. Tuotannon kulut suhteessa ulkolämpötilaan

### 10.2.2 Kannattavuuden visualisointi

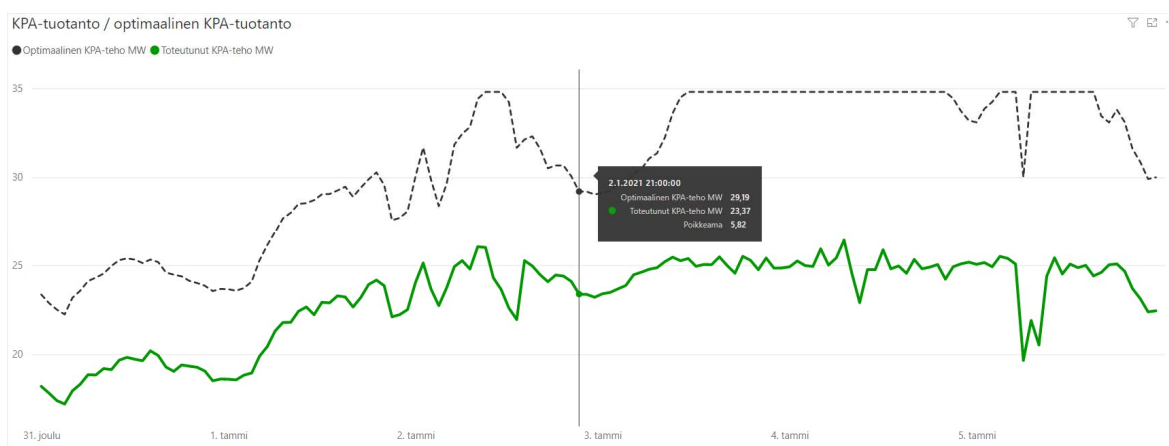
Optimointimallin avulla laskettu tuotannon hetkellinen kannattavuusprosentti voidaan esittää reaaliaikanaäytöllä tuotantodatan ohella. Esitystavaksi soveltuu yksinkertainen, kuviossa 34 esillä oleva mittarinäyttö, josta on nähtävillä hetkellinen kannattavuusprosentti. Itse lukeman lisäksi kannattavuutta esitetään mittarin asteikkoon valituilla väreillä siten, että negatiivisen kannattavuuden punaisista sävyistä siirrytään nollatason keltaisen sävyn kautta vihreään. Kannattavuuden väriskaala on jaettu viiteen eriväriseen osioon sen sijaan, että käytettäisiin liukuvaa väriskaalaa, koska erityisesti numerodatan tarkka tulkitseminen liukuvasta väriskaalasta on tutkitusti haastavaa (Koponen & Hildén 2019, 79).



Kuvio 34. Hetkellisen kannattavuuden visualisointi

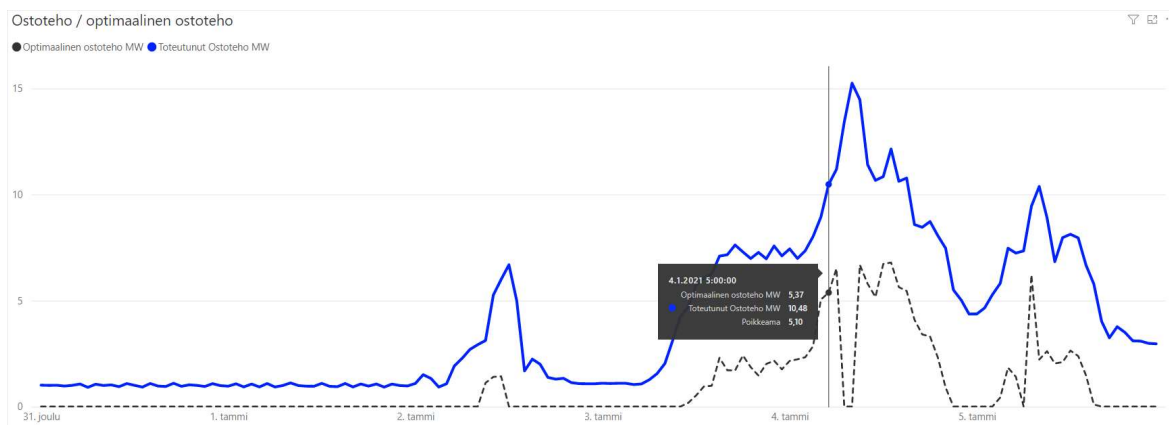
### 10.2.3 Lämmöntuotannon vertailu optimaaliseen tuotantoratkaisuun

Optimoitumallin tuottamaa dataa voidaan hyödyntää myös vertailemalla tiettyä ajanjaksona tapahtunutta kaukolämmön tuotantoa laskennallisesti optimaaliseen tuotantotapaan. Etenkin talvikaudella, jolloin lämmöntuotannon tarve on suurehko, mutta ostotehoa ei välttämättä ole runsaasti saatavilla, lähes poikkeuksetta optimaalisin tuotantotapa on kiinteän polttoaineen käyttö. Tämän myötä olennaisinta olisi visualisoida KPA-laitosten tuotanto. Vertailu voidaan toteuttaa kuvion 35 esittämällä tavalla viivadiagrammilla, jossa esitetään kuvaajaan valitun tuotantotavan lämpöenergian määrä ja sen laskennallisesti optimaalinen määrä omina kuvaajinaan. Absoluuttisten tuotantomäärien ohella voidaan esittää myös toteutuneen tuotannon ja optimituotannon välisen poikkeaman suuruus työkaluvihjeessä.



Kuvio 35. Kiinteän polttoaineen laitoksen lämpötuotannon vertailu laskennallisesti optimaaliseen tuotantoon

Kuviossa 36 on esitelty ostotehon optimaalinen määrä samaan tapaan kuin kiinteän polttoaineen tapauksessa. Kiinteän polttoaineen kuvaajan tapaan poikkeama esitetään kaavioiden päällä työkaluvihjeessä. Vertailtaessa toteutunutta ostotehoa sen optimaaliseen määrään tulee huomioida, että ostotehon saatavilla oleva määrä vaihtelee ja käytännössä täysin optimaalisen tilanteen saavuttaminen voi olla äärimmäisen haastavaa.



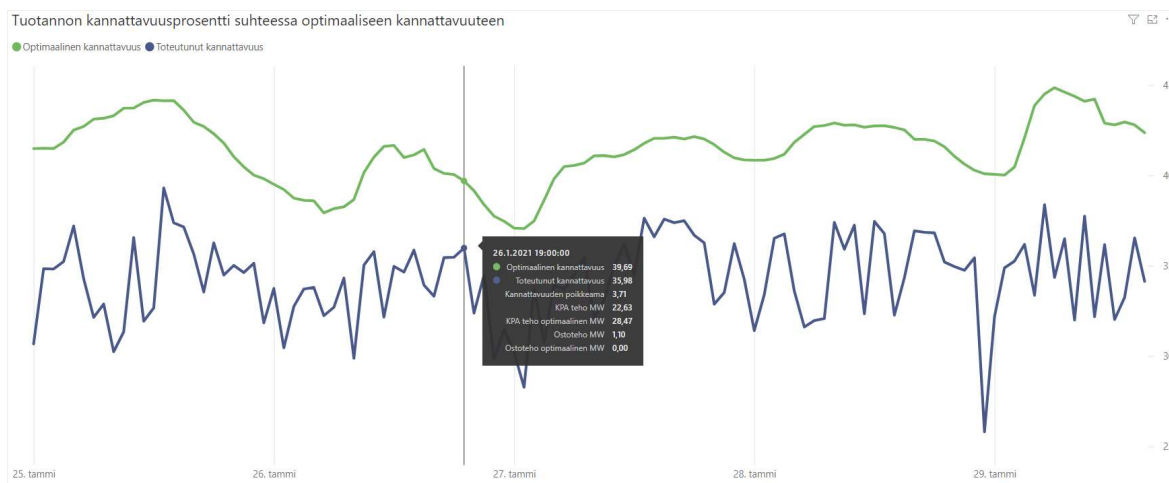
Kuvio 36. Ostetun lämpötehon määrä suhteessa optimaaliseen ostotehoon

Kiinteän polttoaineen tuotannosta ja ostotehosta poiketen öljytehon osalta poikkeamakuvaajaa ei ole mielekästä esittää, koska öljytehon käyttö on käytännössä joka tilanteessa viimeinen toivottava vaihtoehto. Käytännössä öljyllä tuotetun kaukolämmön optimimäärä olisi lähes joka tapauksessa nollassa vain aivan kovimpia pakkassäitä tai muun tuotannon keskeytyksiä lukuun ottamatta, jolloin öljyn käyttö olisi pakollista lämmöntarpeen tyydyttämiseksi.

Optimaalisen tuotannon vertailua voidaan tehdä sekä tulevaan että jo tapahtuneeseen tuotantoon. Vertailemalla optimituotantoa tulevien vuorokausien tuotantoennusteeseen voidaan esittää laskennallisesti optimaalinen tapa tulevien vuorokausien osalta ja siten helpottaa päätöksentekoa tuotantotavasta. Vastaavasti historiadataan verrattaessa voidaan visuaalisesti läpikäydä jälkikäteen tilanteita, joissa optimaalisen ja toteutuneen tuotannon välillä on ollut havaittavissa merkittäviä eroja. Tilanteita analysoimalla ja optimointidataa hyödyntämällä voidaan siten saada parempaa näkemystä siitä, millä tavalla vastaavanlaisissa tilanteissa olisi ollut mahdollista toimia taloudellisemmin ja ekologisemmin.

Tuotannon ja optimaalisen tuotannon vertailun ohella voidaan visualisoida myös pelkästään kannattavuuden poikkeama optimaalisesta. Kuvio 37 esittää kannattavuutta ja sen poikkeamaa optimaalisesta viivadiagrammina visualisointuna. Tällä tavoin kuvaaja pysyy selkeämpänä ja siitä on nopealla vilkaisulla nähtävissä tärkein tieto. Toisaalta, sijoittamalla

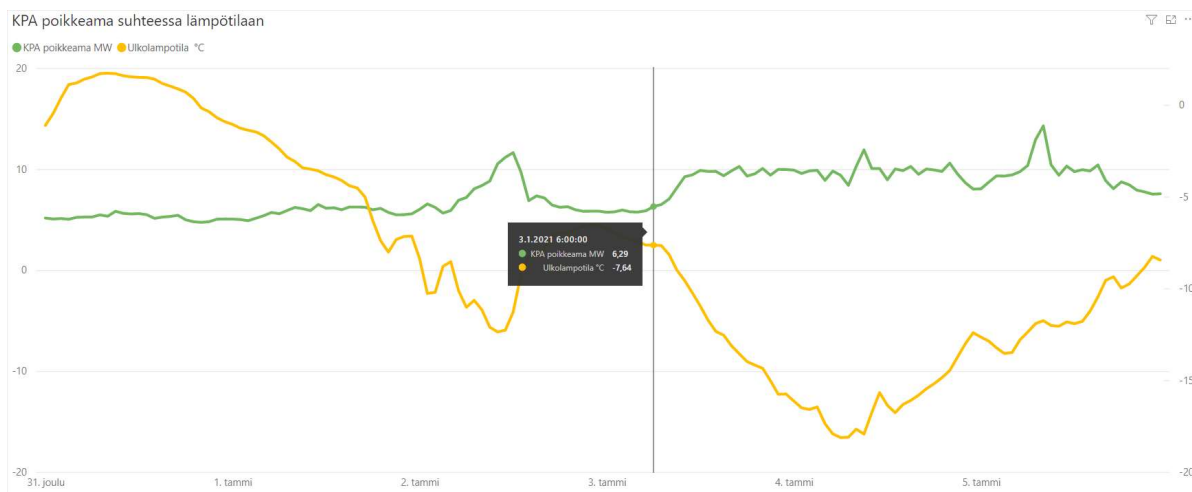
tuotannon tunnuslukuja työkaluvihjeeseen, voidaan tämänkin kuvaajan sisällä esittää lisätietoja, joiden avulla toteutunutta tilannetta voi tarkastella yksityiskohtaisemmin eri tuotantotapojen osalta.



Kuvio 37. Kannattavuusprosentin vertailu optimaaliseen kannattavuuteen

Päätöksenteon tueksi voidaan visualisoida myös lämmöntuotannon poikkeamaa optimaalisesta tuotantomäärästä suhteessa ulkolämpötilaan. Kuvio 38 esittää poikkeaman ja ulkolämpötilan suhdetta viivadiagrammin avulla. Kuvaajan avulla voidaan yrittää löytää tietynlaisia sääoloja, joissa poikkeama toteutuneen tuotannon ja optimaalisen tuotannon välillä on ollut erityisen suuri, jolloin voitaisiin pyrkiä havaintojen perusteella luomaan toimintamalleja vastaavanlaisiin tilanteisiin. Optimituotannon ja ulkolämpötilan visualisointiin voitaisiin tarvittaessa liittää myös muita säätilan mittauksia, kuten tuulen nopeutta ja suuntaa, sekä esimerkiksi auringon lämmitystehoa, joskin ulkolämpötila on selkeästi suurin vaikuttava tekijä lämmön tarpeen suhteen.

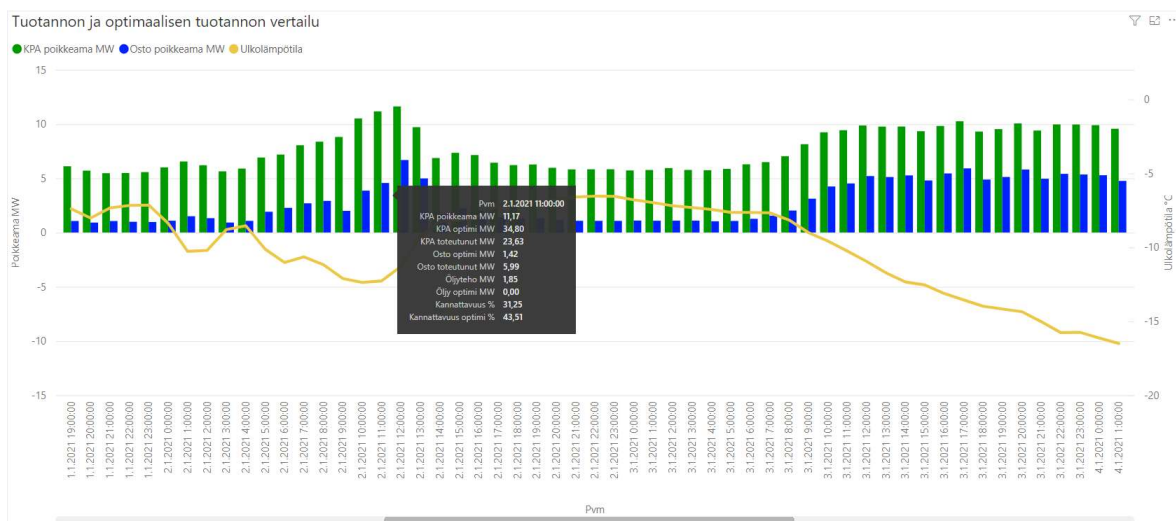




Kuvio 38. Kiinteän polttoaineen tuoton poikkeama optimaalisesta suhteessa lämpötilaan

#### 10.2.4 Tuotannon taloudellisuus suhteessa optimaaliseen tuotantotapaan

Päätöksenteon tukemisen kannalta on olennaista esittää myös tuotantotavan vaikutus taloudellisesti. Vertailemalla toteutunutta tuotantoa optimaaliseen tuotantoon, voidaan esittää kunkin tuotantomuodon poikkeama suhteessa optimaaliseen määrään ja samalla esittää poikkeaman vaikutus kannattavuuteen. Kuviossa 39 vertaillaan tuotannon poikkeaman määrä siten, että eri tuotantomuodot on jaettu omiksi pylväikseen diagrammissa. Tällä tavalla tarkastellessa voidaan pureutua tarkemmin tuotantomuotoihin poikkeamiin. Mikäli käytävissä on riittävästi talouslukuja, kuten tarkat tuotantohinnat sekä kaukolämpömyynnin tulot, voitaisiin kannattavuuden ohella esittää potentiaalinen säästö suoraan euro-määräisenä.



Kuvio 39. Toteutuneen kaukolämpötuotannon ja optimaalisen tuotannon vertailu

Visualisointien toteutuksessa Power BI:n tarjoamat visualisointityypit todettiin riittäviksi, eikä aiemmin käytettäväksi suunnitelluille R-ohjelmointikielellä toteutetuille visualisoinneille ilmennyt tarvetta. Vaikka Power BI asettaakin erilaisia reunaehdoja ja rajoitteita datan esittämistavoille, ne eivät aiheuttaneet suuria ongelmia ja sekä koontinäyttöjen että raportointinäkökymien kohdalla ongelmakohdat saatiin kierrettyä siten, ettei datan esitystavan kohdalla jouduttu tekemään kompromisseja.

### 10.3 Taulukkopohjainen raportointinäkökymä

Taulukkopohjaiset raportit mahdollistavat datan tarkastelun laitos- tai paikkakohtaisesti valittua aikajaksoa ja tarkasteluväliä käyttäen. Kuvassa 10 on nähtävissä Power BI Service – verkkopalveluun luotu monisivuinen, useita kymmeniä raporteja sisältävä raporttikokonaisuus. Näkökymän vaihtaminen eri raporttisivujen välillä tapahtuu sivun vasemmassa laidassa olevan listan kautta, joskin kokonaisuuden hahmottamista varten raporttinäkökymään luotiin myös erillinen etusivu, josta pääsee helposti mille tahansa raportin sivulle ja jossa raportit ovat kategorisoitu asiayhteyden mukaan.

Taulukkonäkökymät esittävät datan käyttäjän valitsemalla tarkkuudella siten, että taulukon yksi rivi vastaa joko tunnin, päivän tai kuukauden mittauksien keskiarvoa. Taulukoissa lähes kaikkien sarakkeiden alapuolelle on lisätty yksi tai useampi koontilaskenta. Tyypillisimpiä laskennallisia arvoja ovat eri mittauksien minimi-, maksimi- ja keskiarvomittaukset sekä tuotetun tehon summalaskennat. Lisäksi joidenkin sarakkeiden kohdalla lasketaan valitun aikajakson ensimmäisen ja viimeisen lukeman erotus. Tällaisia laskentoja ovat esimerkiksi kiinteän polttoaineen laitosten vedenkulutus sekä polttoaineiden kulutus- ja tankkausmäärät.

Etusivu

KPA 12MW

KAR YHT

KAR1

KAR2

KAR3

WIT

Takajärvi

Eteläntie

KPA1-KL

KPA1 päästöt

KPA1 Lauhde

KPA2-KL

KPA2 Lauhde

KPA2 Vesi

KPA2 polttoaine ja tuh...

KPA2 kattila ja palami...

KPA2 kattila ilmat

KPA2 kattila vesi

KPA2 pesuri

KPA2 tehot

Karjalahti-Vesi

Verkoston lämpötilat

Verkosto lämpötilaerot

**KPA 12MW**

Pvm	Ulkolämpötila °C	Käyntiaika h	Kattilaaenergia MW	Menoveden lt °C	Paluuv veden lt °C	Vesivirta m <sup>3</sup>	KL-energia MWh	Kulutettu öjy l
15.3.2021 15:00	1,40	0,00	0,00	86,95	87,72	0,00	0,00	0,00
15.3.2021 16:00	1,07	0,00	0,00	87,73	86,69	0,00	0,00	0,00
15.3.2021 17:00	0,53	0,00	0,00	86,94	87,71	0,00	0,00	0,00
15.3.2021 18:00	-0,06	0,00	0,00	89,77	88,61	0,00	0,00	0,00
15.3.2021 19:00	-0,92	0,00	0,00	90,22	89,16	0,00	0,00	0,00
15.3.2021 20:00	-1,15	0,00	0,00	90,76	89,71	0,00	0,00	0,00
15.3.2021 21:00	-1,61	0,00	0,00	90,96	89,87	0,00	0,00	0,00
15.3.2021 22:00	-1,58	0,00	0,00	91,60	90,48	0,00	0,00	0,00
15.3.2021 23:00	-2,29	0,00	0,00	91,79	90,72	0,00	0,00	0,00
15.3.2021 0:00	-1,65	0,00	0,00	89,14	84,00	0,00	0,00	0,00
<b>SUM</b>	<b>-14,46</b>	<b>39,83</b>	<b>0,00</b>	<b>67,32</b>	<b>42,66</b>	<b>0,00</b>	<b>0,00</b>	<b>21 922,07</b>
<b>MIN</b>	<b>1,40</b>	<b>0,00</b>	<b>0,00</b>	<b>86,95</b>	<b>87,72</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
<b>MAX</b>	<b>-2,29</b>	<b>0,00</b>	<b>0,00</b>	<b>91,79</b>	<b>90,48</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
<b>AVG</b>	<b>2,77</b>	<b>213,99</b>	<b>81,12</b>	<b>80,18</b>	<b>0,40</b>	<b>182,32</b>	<b>SUM</b>	<b>SUM</b>

**KPA 12MW päästöt**

Pvm	Öljymäärä l	Öljymäärä ML	Käyntiaika h	Hilidioksidit CO2 t	Rikkiidioksidit SO2 t	Hiukkaspäästöt t	Kokonaispöly g/s	Typen oksidit NO2 t	NOx g/s	Energia GWh	Energia Tj
15.3.2021 15:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 16:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 17:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 18:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 19:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 20:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 21:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 22:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
15.3.2021 23:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
16.3.2021 0:00	0,00	548 659,20	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,00	0,00
<b>SUM</b>	<b>21 922,07</b>	<b>39,83</b>	<b>58,05</b>	<b>0,0000</b>	<b>0,0110</b>	<b>0,08</b>	<b>0,0731</b>	<b>0,51</b>	<b>0,22</b>	<b>0,79</b>	<b>SUM</b>

Tarkasteluväli

15.3.2021

27.5.2021

Aikajaks

Kuukausi

Päivä

Tunti

KEMIN ENERGIA JA VESI

POLAR-AUTOMAATIO

INNOCENT

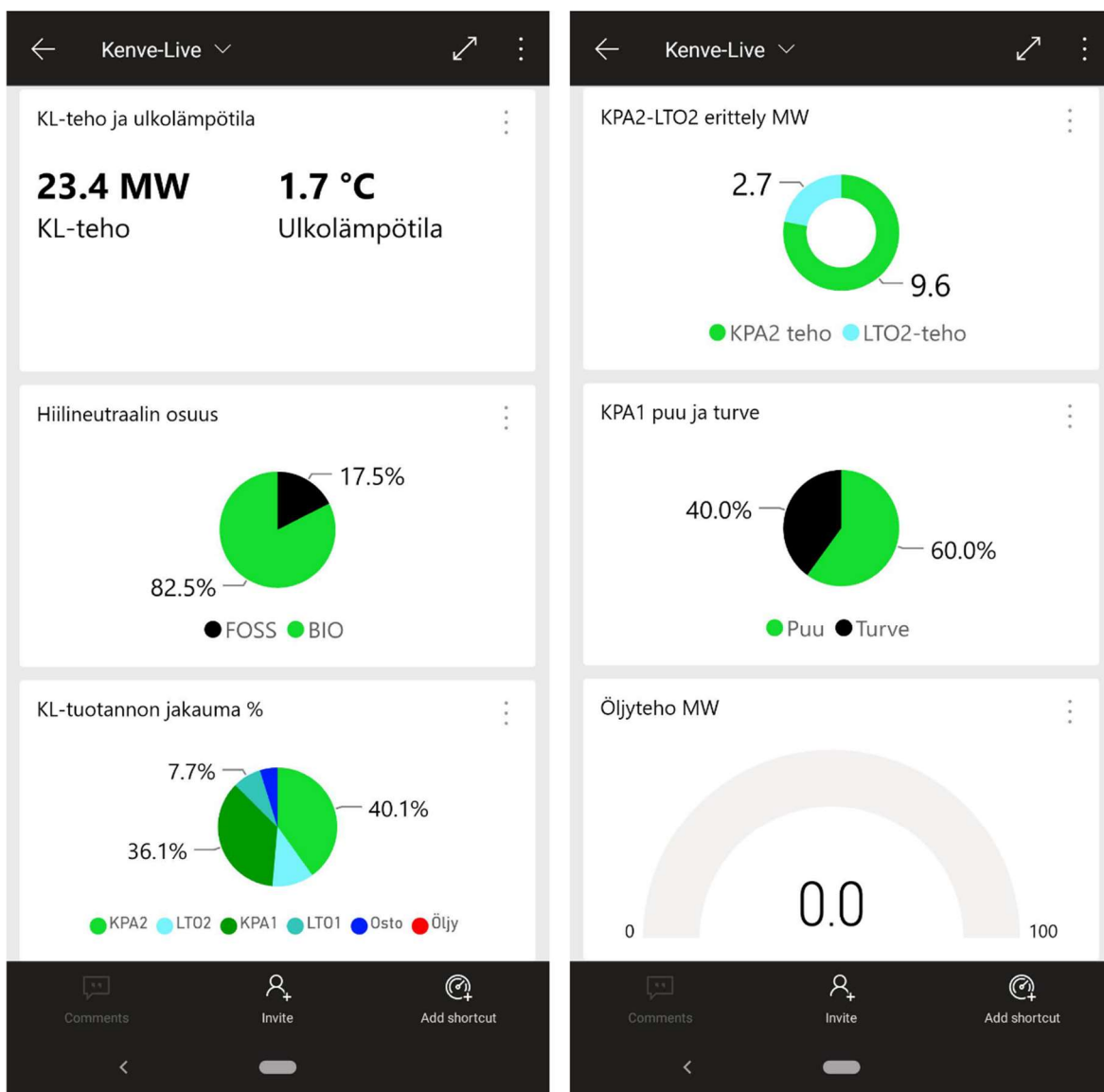
Kuva 10. Kuvakaappaus taulukkopohjaisesta raportointinäkymästä Power BI Service –palvelussa

Kaukolämpölaitosten raportit sisältävät tehomittausten ja päästölaskentojen ohella runsaasti tietoa kiinteää polttoainetta käyttävien laitosten vedenkulutuksesta, sekä olennaisia tietoja lauhdevedestä, kuten lauhteen PH-arvosta ja johtokyvystä. Uuden KPA2-laitoksen osalta raporteista on tarkasteltavissa myös tietoja laitoksen polttoaineen syötöstä sekä runsaasti tietoja palamisprosessin mittauksista. Lisäksi raporttinäkymä sisältää yhteenvetoraportin kaikkien lämmöntuotantomuotojen tehomittauksista.

Kaukolämpöverkosta koskevat raportit käsittävät tärkeimmät tiedot kaukolämpöverkoston lämpötiloista ja lämpötilaeroista, sekä vastaavat tiedot verkoston painemittauksista. Lisäksi raporteista on tarkasteltavissa öljysäiliöiden pinnanmittaustietoja, sekä saapuneiden tankkausten öljymäärät. Myös Kemin Energia ja Vesi Oy:n oman sääaseman tiedot on tarkasteltavissa raporteista.

## 10.4 Raporttinäkymien käyttö mobiililaitteilla

Kaikki järjestelmän reaaliaikaiset näytöt on pyritty toteuttamaan siten, että tuotannon avainluvut olisivat tarkasteltavissa myös mobiililaitteilla. Käytännössä lähes kaikki reaaliaikaisen tuotantonäkymän mittarit voidaan näyttää sellaisenaan Microsoftin Power BI –mobiilisovelluksessa (Microsoft 2020c). Kuvassa 11 on nähtävillä ruutukaappauksia reaaliaikaisten koontinäyttöjen mobiilinäkymistä. Reaaliaikanäyttöjen mobiilikäyttö oli osoittautunut jo vuoden 2019 uudistuksen yhteydessä erittäin toivotuksi ominaisuudeksi ja myös kaikki myöhemmin lisätyt KPA2-laitosta koskevat visualisoinnit, tuotannon ekologisuuden jakaumat sekä ennustenäkymät on tuotu mobiililaitteilla käytettäväksi.



Kuva 11. Ruutukaappauksia Microsoft Power BI –mobiilisovelluksesta

Huolimatta siitä, että raportointijärjestelmän taulukkopohjaisille näkymille ei asetettu ehtoja mobiilikäytön suhteen, kaikki raporttinäkymät ovat kuitenkin käytettävissä mobiilisovelluksen kautta, joskin taulukoiden sisältämien sarakkeiden suurehko lukumäärät aiheuttavat sen, ettei näkymät ole erityisen joustavia puhelin- tai tablettikäytössä. Etenkin puhelinten näyttöjen kapeahkojen kuvasuhteiden ja pienehköjen pikselimäärien vuoksi näkymien tarkastelu vaatii usein käyttäjältä näkymän suurentamista ja tarkastelua pienempi osa kerrallaan. Kehitysnäkökulmasta katsottuna taulukoiden mobiilikäyttöä voitaisiin helpottaa teemmällä raporteista erilliset mobiiliversiot, joskaan aivan kaikkia visualisointeja, kuten esimerkiksi taulukkojen sarakkeiden alapuolella olevia laskennallisia yhteenvetoja ei olisi mahdollista toteuttaa samalla tapaa kuten raporttien työpöytäversioissa. Erillisten mobiiliversioiden toteuttamisessa voitaisiin myös harkita taulukoiden sarakkeiden määrän rajaamista siten, että niissä joko näytettäisiin vain kaikkein tärkeimmät tiedot, tai että taulukot jaettaisiin useampaan, vähemmän sarakkeita sisältävään taulukkoon (Jarosz 2021).

## 11 Johtopäätökset ja pohdintaa

### 11.1 Käytetyn tutkimusmenetelmän soveltuvuus

Tutkimuksen tietoperusta luotiin kirjallisuuskatsauksen avulla ja käytännön tutkimus tehtiin suunnittelutieteen periaatteita noudattaen. Suunnittelutieteelle tyypillinen iteratiivinen kehittäminen sopi tällaiselle tutkimukselle, jossa haluttujen lopputulosten saavuttaminen vaati erilaisten menetelmien testaamista ja ratkaisun kehittämistä vaiheittain. Kirjallisuuskatsaus ja suunnittelutieteen menetelmät tukivat toisiaan: teoreettinen kirjallisuuskatsaus tuki sekä ensimmäisen, karkean ratkaisuvaihtoehdon luomista, että myöhempien ratkaisumallien iterointia. Ilman kirjallisuuskatsausta pelkän suunnittelutieteen menetelmien avulla luotu ratkaisu olisi voinut riippua liikaa tutkimuksen tekijän subjektiivisesta näkemyksestä, minkä painottuminen liikaa onkin yksi suunnittelutieteen keskeisimmistä heikkouksista.

### 11.2 Optimaalisen kaukolämmön tuotantokombinaation selvittäminen

Optimaalista kaukolämmön tuotantokombinaatiota yritettiin selvittää lineaarisen optimoinnin, geneettisen algoritmin sekä koneoppimisen menetelmiä käyttäen. Lineaarista optimointia suoritettiin Python-ohjelmointikielelle saatavalla PuLP-kirjastolla, jonka käyttö oli suoraviivaista. Tutkimuksen perusteella lineaarinen optimointi tuotti optimaalisen tuloksen varsin pienellä vaivalla. On kuitenkin huomioitava, että optimointitehtävän oikeanlainen mallintaminen on edellytys haluttujen lopputulosten saavuttamiselle. Optimoinnissa tehtiin erilaisia oletuksia tiettyjen muuttujien suhteen, jotka saattavat todellisuudessa olla entuudestaan tuntemattomia, kuten esimerkiksi polttoaineen laatu ja ostotehon saatavuus.

Geneettisellä algoritmilla voidaan lähtökohtaisesti ratkaista mikä tahansa optimointitehtävä. Tutkimuksen perusteella geneettinen algoritmi soveltui optimaalisen tuotantokombinaation selvittämiseen hyvin, mutta algoritmin toimintaa oli seurattava ja pohtia, tuottiko se tyydyttävää ratkaisua. Edullisempien tuotantomuotojen hinta on selkeästi alhaisempi kuin kalleimman tuotantomuodon, jonka johdosta algoritmi päätyi suosittamaan kysynnästä poikkeavaa tuotantokombinaatiota. Kysynnän täyttäminen on kuitenkin prioriteetiltaan tärkein seikka, joten algoritmin kohdefunktioon tuli lisätä virhetermi. Virhetermi huomioiden algoritmia kannustettiin täyttämään kysyntä, jonka myötä myös tuotantomuotojen välinen optimointi muodostui paremmaksi. Geneettisellä algoritmilla päästiin lopulta täsmälleen samaan tulokseen kuin lineaarisen optimoinnin menetelmällä.

Optimoinnilla saatu ratkaisu perustuu kuitenkin vain likimääräisiin arvioihin siitä, millaisia kustannuksia kunkin tuotantomuodon käytöstä aiheutuu. Näin ollen mallin mukainen kannattavuus ei välttämättä kuvasta todellisuutta parhaalla mahdollisessa tavalla. Optimoinnin lopputulokset luovat kuitenkin pohjan jatkokehitykselle.

### 11.3 Koneoppimisen menetelmien soveltuminen optimaalisen kaukolämmön tuotantokombinaation selvittämisessä

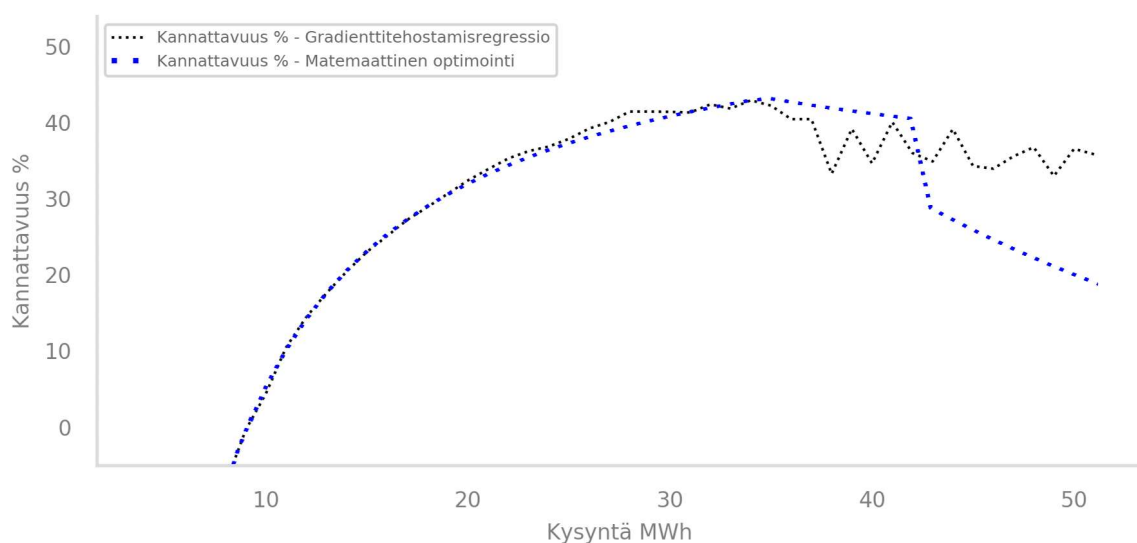
Koneoppimisen menetelmät perustuvat dataan, jota on kerätty kaukolämmön tuotantolaitoksen toiminnasta. Datan avulla pystytään mallintamaan muuttujien välisiä suhteita ja luomaan erilaisia ennustemalleja. Tässä tutkimuksessa tavoitteena oli tutkia miten koneoppimisen menetelmät soveltuvat optimointitehtävän ratkaisemiseen. Luonnollinen valinta jatkuvien muuttujien tapauksessa oli valita regressiotehtävään soveltuvia algoritmeja, joiden toimintaa vertailtiin. Vertailtavia malleja luotiin polynomiregressiolla, päätöspuulla sekä gradienttitehostamisregressiolla.

Data sisältää poikkeuksetta hajontaa, minkä takia koneoppimisen malleilla pyritään luomaan dataa yleistävä malli. Tällaisen mallin rakentaminen pitää sisällään matemaattisen optimoinnin menetelmiä, mutta sillä ei kuitenkaan toteuteta varsinaista optimointia. Tutkimuksen perusteella malleilla päästään lähelle optimaalista tasoa silloin, kun havaintoarvojen hajonta on pieni. Taulukossa 3 on listattu kaikkien vertailtujen koneoppimisen mallien tunnuslukuja. Vertailun perusteena käytetään ensisijaisesti selitysastetta ( $R^2$ ) ja keskineliövirheen neliötä (RMSE). Selitysaste osoittaa kuinka suuren osuuden selittävillä muuttujilla voidaan selittää selitettävän muuttujan vaihtelusta. Keskineliövirheen neliö puolestaan huomioi mallissa olevan virheen hajonnan. Kaikkien mallien opettamiseen käytettiin samoja muuttujia sekä samaa opetus- ja testidataa, joten myös selitysastetta voidaan pitää näiden mallien osalta vertailukelpoisena tunnuslukuna. Suurin selitysaste saavutettiin gradienttitehostamisregressiolla, jolla myös korjattu selitysaste on muita malleja parempi. Mallin hajonnan määrä on muista tunnusluvuista päätellen myös pienempi kuin muilla malleilla.

	Polynomi- regressio	Satunnaismetsä- regressio	Gradienttitehostamis- regressio
Selitysaste ( $R^2$ )	57 %	57 %	59 %
Korjattu selitysaste (adj. $R^2$ )	57 %	57 %	58 %
Keskineliövirhe (MSE)	25,250	25,290	24,440
Keskimääräinen absoluuttinen virhe (MAE)	2,622	2,553	2,432
Keskineliövirheen neliö (RMSE)	5,025	5,029	4,944

Taulukko 3. Koneoppimisen mallien tunnuslukuja

Kuviossa 40 on vertailtu matemaattisella optimoinnilla ja gradienttitehostamismenetelmällä tuotettua kannattavuutta kysynnän suhteen. Tämän perusteella voidaan todeta, että kauko-  
lämpötehon tuottaminen on suoraviivaista myös pelkistetyn kannattavuuslaskennan tapauksessa, kun tehon tarve on alle 35 MWh. Eroja alkaa muodostumaan, kun energiantuotannossa on käytettävä apuna osto- ja öljytehoa. Matemaattisen optimoinnin tapauksessa oli tehty oletus ostotehon saatavuudesta, joka poikkeaa datassa ilmenevän ostotehon käytöstä. Myös polttoaineen laadulla on vaikutusta siihen, kuinka paljon lämpöenergiaa voidaan ottaa talteen kiinteän polttoaineen kattilassa ja käyttää tätä lisätehona. Matemaattisessa optimoinnissa lämmöntalteenottoteho oli oletettu olevan maksimiosuus kiinteän polttoaineen tehosta.



Kuvio 40. Kannattavuuden vertailu gradienttitehostamisregression ja matemaattisen optimoinnin välillä



Koneoppimisen menetelmillä pystyttiin tuottamaan malli, jonka avulla voidaan estimoida tuotantomuotojen osuutta kysyntään nähden. Malli tuottaa vain keskimääräisen estimaatin, joten se ei välttämättä kuvasta optimaalista tilannetta. Datalla on suuri merkitys mallin muodostamisessa, joten mikäli dataa on mahdollista käsitellä korostaen kannattavimpia havaintoja, voitaisiin nämä tilanteet opettaa koneoppivalle mallille. Monimutkaisempi kannattavuuslaskenta saattaa edellyttää koneoppimisen hyödyntämistä rakentamalla mallin, joka huomioisi kaikki tarvittavat muuttujat. Tällainen malli toimisi kohdefunktiona, jonka avulla voitaisiin luokitella datan havainnot riittävän kannattaviksi ja samalla suodattaa kannattamattomimmat tilanteet pois.

Koneoppimisen malleilla saatuja tuloksia on vertailtu matemaattisilla optimointimenetelmillä saatuihin tuloksiin. Näiden kahden lähestymistavan tuottamia tuloksia ei kuitenkaan voida täysimääräisesti vertailla, sillä tämän tutkimuksen puitteissa matemaattista optimointia on toteutettu pelkistetyillä säännöillä. Koneoppimisen malli oppii säännöt kaukolämpötuotannon mittausdatasta, joka kuvastaa todellista toimintaympäristöä.

#### 11.4 Tiedon reaaliaikaisella visualisoinnilla saavutetut hyödyt

Jo aiemmin toteutetun järjestelmä uudistuksen ja reaaliaikaisen raportoinnin myötä saavutettiin tilanne, jossa tuotannon kannalta keskeisimmät tiedot saatiin tuotua niitä tarvitsevien henkilöiden käyttöön joko lähes välittömästi tai hitaimmillaankin kolmen tunnin sisällä. Muutos oli merkittävä verrattuna aiempaan tilanteeseen, jossa hajallaan olevasta datasta saatiin saada kokonaiskuva vasta useiden päivien tai pahimmillaan yli viikon viipeellä. Uudistusten myötä reagointiaikaa tuotannon tilanteen muutoksiin on saatu lyhennettyä merkittävästi, minkä myötä Kemin Energia ja Vesi Oy saavutti vuonna 2019 noin kymmenen prosentin vähennyksen CO<sub>2</sub>-päästöjen osalta.

Vaikka käytännössä tuotannon tarpeen arviointi perustuukin suurimmilta osin sääennusteiden seurantaan ja tuotannosta vastaavien henkilöiden ammattitaitoon, on tuotantoennusteen visualisointi koettu hyödylliseksi avuksi antamaan vahvistusta henkilöstön omalle arviolle tuotannon tarpeesta seuraavien vuorokausien aikana. Erityisen hyödyllisiksi ennusteet voidaan nähdä tilanteissa, joissa kaukolämpöyhtiön henkilöstöön tulee muutoksia, koska ennusteen avulla uusille työntekijöille muodostuu nopeammin käsitys kulloinkin tarvittavasta lämpötehomäärästä.

#### 11.5 Haasteet

Työn suurimmat haasteet aiheutuvat pääosin datan saatavuuteen liittyvistä ongelmista ja KPA2-laitoksen käyttöönoton viivästymisestä. Alun perin käytettäväksi suunniteltu talous-

ja kulutusdata jäi oletettua suppeammaksi, mistä johtuen osaa visualisoinneista ei voitu toteuttaa täysimääräisesti. Datan puutteista johtuen optimoinnin osalta kannattavuuden laskentaa tehtiin yksinkertaistetusti käyttäen vain muutamia, asiakkaan tekemään karkeaan arvioon perustuvia talouslukuja.

Joidenkin raporttien sisältämien laskennallisten arvojen kohdalla haasteita aiheutti myös itse automaatiojärjestelmän mittauslaitteet. Erityisesti öljysäiliöiden pinnanmittausanturit ovat herkkiä öljysäiliöiden lämpölaajenemisesta aiheutuville mittausvirheille. Säiliöiden vuotoisuuden ollessa kymmeniä tuhansia litroja, jo pienehkö lämpölaajenemisesta johtuva ero pinnanmittauksessa saattaa tehdä satojen litrojen suuruusluokkaa olevia poikkeamia mittauksiin ja sitä mukaa raporttien laskennallisiin öljynkulutuslukemiin. Raportoinnin tarkkuuden ja luotettavuuden kannalta olisi tarpeellista huomioida etukäteen todennäköisimmät mittausvirheiden aiheuttajat ja pohtia, millä tavalla eri mittauksien ongelmakohdat olisivat kierrettävissä, tai että voitaisiinko poikkeamat havainnollistaa sopivalla esitystavan valinnalla.

Alun perin alkuvuodesta 2021 käynnistettäväksi suunnitellun KPA2-laitoksen käyttöönotossa ilmeni viivästyksiä ja laitos käynnistyi vasta kevättalvella 2021. Viivästyksestä ja osittaisesta tietojen puutteesta johtuen kiinteää polttoainetta käsiteltiin optimoinnin kohdalla vain vanhemman KPA1-laitoksen osalta. Toisaalta laitosten suunnitellun käyttötavan myötä asialla ei olisi ollut optimointiratkaisujen hyödyntämisen kannalta suurtakaan merkitystä, koska KPA2-laitosta pyritään jatkossa käyttämään pääasiallisena jatkuvan kaukolämpötehon lähteenä ja KPA1-laitosta käytetään lisätehon tuottamiseen vain tarvittaessa, eli kiinteän polttoaineen laitosten toiminta voidaan nähdä tuotetun tehon kannalta yhtenä suurempana yksikkönä. Näin ollen myös kaukolämmön tuotannosta tähän saakka kerätty mittausdata ei kuvaa suunniteltua käyttötapaa ja uuden, molemmat laitokset huomioivan mallin kehittäminen edellyttää uuden datan keräämistä. Lisäksi matemaattisen optimoinnin osalta olisi syytä tarkentaa siinä käytettäviä rajoitteita ja muita kannattavuuteen vaikuttavia tekijöitä.

## 11.6 Jatkokehitys

Sekä optimointimallin kehittämisessä että raportointi- ja visualisointiratkaisuiden toteuttamisessa ilmeni runsaasti erilaisia jatkokehitysmahdollisuuksia. Jatkokehitystä tulisi viedä eteenpäin suunnittelutieteen periaatteiden mukaisesti ja arvioida tuotetun ratkaisun tuottamia tuloksia. Kehitystarpeita tulisi priorisoida arvioinnin perusteella syntyneiden johtopäätösten mukaisesti ja luoda yhdessä asiakkaan kanssa suunnitelma seuraavaa kehityskierrosta varten. Syntyvien artefaktien asianmukainen dokumentointi toimii kommunikoinnin välineenä eri sidosryhmien välillä.

### 11.6.1 Kannattavuuden laskenta ja optimointi

Kaukolämmön tuotannon kannattavuuteen vaikuttaa hyvin moni tekijä, mikä tekee sen arvioinnista monimutkaista. Tästä johtuen kannattavuuden laskentamalleja on pyrittävä yksinkertaistamaan. Saatavilla oleva data ja kannattavuuteen vaikuttavat tekijät olivat kuitenkin tässä tutkimuksessa suppeat. Pureutumalla syvemmälle kannattavuustekijöihin voitaisiin optimointia hyödyntää monimutkaisten suhteiden mallintamisessa ja tuottaa tarkempia tuloksia. Optimoinnissa voitaisiin huomioida myös tuotantolaitosten sammuttamiseen ja käynnistämiseen liittyvät reunaehdot, sillä lyhyt katko tuotannossa voi myös kasvattaa kustannuksia. Optimoinnilla voitaisiin tällöin osoittaa, kuinka kauan katkoa tulisi välttää, jotta kannattavuus olisi tasapainossa.

### 11.6.2 Optimointi tekoälyn menetelmillä

Koneoppiminen on tekoälyn osa-alue, jossa pyritään mallintamaan tiettyä toimintaperiaatetta datan avulla. Muilta osin tekoäly painottuu älykkäiden järjestelmien luontiin, joita hyödynnetään erilaisissa tehtävissä. Tekoälyllä voidaan toteuttaa malli, joka itsenäisesti oppii virheistään ja pystyy näin kehittymään tehtävässään. Optimointitehtävien tapauksessa olisi mahdollista tutkia esimerkiksi vahvistusoppimisen menetelmiä, joiden toiminta perustuu oppimisagentteihin. Agentit tekevät erilaisia toimintoja kohdeympäristössään ja niitä voidaan joko palkita tai rangaista niiden tekemien toimintojen perusteella. Neuroverkot ovat toinen yleinen menetelmä, jonka avulla on mahdollista mallintaa hyvin monimutkaisia tehtäviä. Neuroverkkojen tapauksessa tehtävän ratkaisuun liittyvien optimaalisten hyperparametrien löytäminen voi olla haastavaa, mutta niiden löytämiseksi on kehitetty erilaisia tehokkaita menetelmiä.

### 11.6.3 Matemaattisen optimoinnin ja koneoppimisen yhdistäminen

Matemaattisen optimoinnin menetelmät ja koneoppiminen voivat toimia myös yhdessä. Tällainen hybriditoteutus on tämän työn myötä periaatteessa jo olemassa, sillä optimointitehtävissä kaukolämpöenergian kysyntä saadaan koneoppimisen menetelmillä toteutetulta ennustemallilta. Koneoppimisen avulla selvitettyt muuttujien väliset suhteet voidaan huomioida matemaattisen optimoinnin menetelmissä monella tavalla. Tuotantomuotojen minimi- ja maksimiarvoja voitaisiin estimoida tietyissä tilanteissa, kuten esimerkiksi vuorokaudenajan perusteella, ja käyttää näitä arvoja matemaattisen optimoinnin rajoitteina. Tässä tutkimuksessa tehty oletus ostotehon saatavuudesta voitaisiin myös korvata koneoppimisen mallilla, jolla voitaisiin ennustaa sen saatavuutta tietyinä ajankohtana. Koneoppimisen mallit ovat käytännössä matemaattisia funktioita, joilla voidaan mallintaa monimutkaisiakin ongelmia. Näitä funktioita voisi sellaisenaan käyttää myös matemaattisen optimoinnin kohdefunktioina.

### 11.6.4 Optimointi- ja tuotantodatan visualisointi

Suurin osa raportointijärjestelmän eri näkymistä toteutettiin esittämällä data pääasiassa taulukkomuotoisena. Power BI tarjoaa laajan valikoiman erityyppisiä visualisointeja ja monipuolisemmalla visualisointien käytöllä raporttinäkymiä olisi mahdollista jatkokehittää siten, että numeromuotoisten tietojen väliset suhteet, tietojen muutokset ja muutosten suunta olisi helpommin hahmotettavissa. Käyttämällä erilaisia kuvaajia, mittareita, kaavioita tai karttoja voitaisiin visualisoinneilla saavuttaa merkittävää lisäarvoa, etenkin tietojen tulkitsemisen helppouden näkökulmasta. Vastaavasti myös kaukolämpöverkoston lämpötilat, paine-erot sekä verkoston kuntoon liittyvät mittaukset voitaisiin esittää karttanäkymässä, jonka avulla mittaukset saataisiin visuaalisesti yhdistettyä niiden todelliseen sijaintiin verkostossa. Verkoston mittausten karttanäkymää voitaisiin hyödyntää myös vikatilanteiden visualisoinnissa.

Datan optimointitietojen visualisoinnin osalta olennaisin jatkokehitystarve olisi liittää järjestelmä suoraan tietolähteisiin, joista voitaisiin hakea reaaliajassa tuotantoon liittyviä taloustietoja sekä tarkkoja mittauksia kaukolämpöasiakkaiden kuluttamasta lämpöenergiasta. Lisädatan avulla voitaisiin visualisoida ja ennustaa tuotannon kannattavuutta reaaliaikaisesti ja siten helpottaa päätöksentekoa. Olemassa olevan tuotanto- ja säädäntä liittämällä reaaliaikaiseen talous- ja kulutusdataan voitaisiin koko kaukolämmöntuotantoa tarkastella yhtenä kokonaistaloudellisena järjestelmänä, jossa eri tekijöiden vaikutukset toisiinsa olisi aiempaa helpommin tulkittavissa visuaalisesti.

### 11.6.5 Ennustedatan ja päästölaskentojen yhdistäminen

Tuotannon tarpeen ennustedataa ja raportointijärjestelmän päästölaskentoja yhdistämällä voitaisiin ennakoida myös tulevan tuotannon päästöjä. Mikäli päästöennusteeseen liitettäisiin myös talousdataa ja päästöoikeuksien hintatietoja, voitaisiin laskea ja visualisoida päästöjen aiheuttamat kustannukset ja sitä mukaa parantaa kokonaiskuvaa tuotannosta ja sen kustannuksista. Päästöennusteen pohjalta luodun visualisoinnin avulla olisi mahdollista koota näkymä, jossa on nähtävissä kunkin tuotantomuodon kokonaiskustannukset jaoteltuna varsinaisten tuotantokustannusten ja päästökaupan aiheuttamien kustannusten kesken.

### 11.6.6 Tuotannon tarpeen ennustamisen jatkokehitys

Tuotannon tarpeen ennustamisen keskeisin kehityskohde olisi ennusteen pituuden kasvattaminen. Sääennusteita tarjoavien tahojen ennusteiden pituuksissa voi olla suuria eroja. Lähipäivien ennustaminen on nykyajan säämalleilla helppoa estimoida, mutta pitemmän aikavälin ennusteet vaativat entistä kehittyneempien mallien käyttöä. Toteutuksessa voitaisiin hyödyntää ECMWF:n eli Euroopan keskipitkien sääennusteiden keskuksen tuottamia useiden vuorokausien päähän ulottuvia ennusteita, jotka ovat avoimesti käytettävissä ECMWF:n ylläpitämässä Copernicus-verkkopalvelussa (ECMWF, 2021). Vastaavasti ennusteita voitaisiin tehdä myös huomattavasti pidemmän, jopa viikkoja tai kuukausia käsittävistä ennusteista, joskin niiden sisältämä hajonta on usein melko suurta. Tällaisen ennustetiedon hyödyntämispotentiaali tuotannon ajamisen näkökulmasta on käytännössä melko vähäinen.

Tuotantoennusteen pituutta kasvattaessa olisi syytä visualisoida myös sen ennustettavuus, eli arvio siitä, kuinka luotettava ennuste on. Luotettavuus voidaan yksinkertaisimmillaan esittää esimerkiksi kolmeportaisella skaalalla sanallisesti, tai vaihtoehtoisesti esittämällä ennusteen osuvuuden todennäköisyys prosenttiasteikolla. Erityisesti tilanteissa, joissa ennustettavuus on heikkoa, voitaisiin hyödyntää ennusteen haarukointia, eli luoda useampi, erilaisilla lähtöarvoilla määritelty rinnakkainen ennuste ja esittää useita eri skenaarioita mahdollisista säätilan muutoksista. Esittämällä useita rinnakkaisia ennusteita voitaisiin lisätä ymmärrystä säätilan mahdollisista äkillisistäkin muutoksista ja siten parantaa erilaisiin tilanteisiin varautumista, mikäli tilanne olisi ollut jostakin saatavilla olevasta ennusteesta havaittavissa.

### 11.6.7 Polttoaine-erien riittävyyden ennustaminen

Optimointia voitaisiin laajentaa myös yhdistämällä polttoainekuormien punnitustietoja ja kuormista mitattuja energiasisältöjä pitkän aikavälin sääennusteisiin, jolloin voitaisiin laatia erilaisia ennustemalleja siitä, kuinka pitkälle tulevaisuuteen polttoainevarastojen sisältö riittää. Käytännössä polttoaineiden riittävyysennusteen laatiminen edellyttäisi tarkkaa tietoa siitä, mitä polttoainevarastoa ollaan milloinkin käyttämässä. Tämä aiheuttaa lisähaasteita, koska tilanteesta riippuen eri polttoainekuormia saatetaan sekoittaa keskenään, jolloin ei ole tarkkaa tietoa yhdistetyn polttoainemassan energiasisällöstä. Myös polttoainekuormien tilaamiseen ja kuljetusten ajankohtaan vaikuttaa useat eri säätilaan liittyvät tekijät. Esimerkiksi kevättalvella tyypillisesti tarkkaillaan pitkän ajan sääennusteiden kuukausien päähän ylettyviä ennusteita, koska säiden lämmitessä hake- ja puukuormien kuljetus tietyiltä alueilta vaikeutuu tai muuttuu mahdottomaksi maan routimisen ja sulamisen vuoksi.

### 11.6.8 Ennakoiva laitteiden kunnon seuranta

Koneoppimis- ja tekoälyratkaisut tarjoavat erittäin potentiaalisen työkalun kaukolämpöjärjestelmän ennakoivaan huoltotoimintaan (Vähäkainu ym. 2019, 37). Mittausdatasta tiettyjä poikkeavuuksia tai tapahtumasarjoja hakemalla voitaisiin pyrkiä tunnistamaan tilanteita, joissa jokin järjestelmän laite osoittaa hajoamisen merkkejä, mutta joita ihmisen olisi vaikeaa tai jopa mahdotonta havaita suuren datamassan seasta. Mahdollisia vikatilanteita voitaisiin yrittää tunnistaa esimerkiksi virtaus-, lämpötila- ja painemittausten tarkalla seurannalla, sekä esimerkiksi värinäanturien tuottaman datan tarkemmalla analysoinnilla.

Siinä missä tuotannon optimoinnin kannalta järjestelmän mittausten yhden minuutin intervalli on täysin riittävä, saattaa se olla ennakoivan huoltotoiminnan kannalta aivan liian pitkä. Mikäli mittauksia tapahtuu liian harvoin, saatetaan sulkea pois mahdollisuus havaita poikkeuksia ja virheitä, koska yksittäinen piikki mittausdatassa voi hävitä minuutin aikana tehtyjen mittausten keskiarvolaskentaan melko huomaamattomasti. Toisaalta ongelma olisi kiertävissä melko helposti sillä, että seurattavasta datasta tallennettaisiin myös minimi- ja maksiarvot, vaikkakaan näiden arvojen tarkkailu ei suoraan kerro mahdollisista toistuvista piikeistä mittauksissa.

Ennakoiva huoltotoiminta vaatii melko suurien datamäärien siirtoa ja käsittelyä, etenkin jos dataa kerätään lyhyissä intervaleissa. Seurannan kannalta pidemmälle viety ratkaisu olisi suorittaa mittausten tarkkailua suoraan kaukolämpöverkon mittauslaitteiden fyysisessä läheisyydessä, eli siirtää datan tulkinta- ja analysointitaakkaa paikallisesti verkossa sijaitseville Edge-laitteille. Tällä tavalla toimittaessa ei olisi tarvetta siirtää valtavaa datamassaa pilvipalveluun analysoitavaksi, vaan analyysi tapahtuisi lähellä datan lähdettä ja paikallinen

järjestelmä lähettäisi eteenpäin ainoastaan merkittäväksi tulkitsemansa datan. Ennakoivalla huoltotoiminnalla saavutettavia hyötyjä olisi esimerkiksi laitteiden eliniän ennustaminen, poikkeusten havainnointi sekä virhetilanteen eskaloitumisen estäminen ja sitä mukaa tuotannon katkosten minimointi. Edge-laitteita voitaisiin hyödyntää myös mittausdatan esikäsittelyssä ennen pilveen lähettämistä. (Microsoft 2019.)

## 12 Yhteenveto

Kaukolämpötuotanto on yleistynyt viimeisen kolmen vuosikymmenen aikana Suomessa merkittävästi. Samaan aikaan päästötavoitteet ovat kiristyneet jatkuvasti, minkä myötä kaukolämpötuotannolta edellytetään entistä ekologisempia sekä energia- ja kustannustehokkaampia ratkaisuja. Erityisesti päästökaupan vaikutukset kaukolämpötuotannon kannattavuuteen ovat merkittävän suuria.

Matemaattinen optimointi luo perustan moderneille menetelmille, kuten tekoälylle ja koneoppimiselle, joita nykyaikainen tiedolla johtaminen hyödyntää tehokkaasti. Perinteisillä matemaattisen optimoinnin menetelmillä on kuitenkin vielä käyttöä erilaisten ongelmien ratkaisemiseksi, sillä usein monimutkainenkin ongelma voidaan ratkaista yksinkertaisilla menetelmillä. Linearisella optimoinnilla on mahdollista ratkaista monia optimointitehtäviä, joissa muuttujien väliset suhteet ovat lineaariset. Evoluutioalgoritmeihin lukeutuvilla geneettisillä algoritmeilla voidaan myös ratkaista optimointitehtäviä ja niissä on potentiaalia monimutkaisempienkin tehtävien ratkaisemiseksi. Koneoppimisen menetelmillä voi myös saada hyviä tuloksia optimointitehtävien ratkaisuksi, mutta niissä tehtävä optimointi perustuu kuitenkin vain mallin sovittamiseen dataan. Matemaattisen optimoinnin ja koneoppimisen yhdistelmällä voidaan mahdollisesti luoda uudenlaista potentiaalia optimointitehtävien ratkaisemiseksi hyödyntäen olemassa olevaa dataa.

Jo pelkällä tuotantodatan visualisoinnilla ja nopealla raportoinnilla voidaan tukea ja helpottaa kaukolämpötuotannon ajamisen päätöksentekoa. Reaaliaikaisella ja selkeällä tilannekuvalla tuotannon tärkeimmät avainluvut saadaan koottua selkeiksi kokonaisuuksiksi, joiden avulla tuotannon muutostilanteisiin reagointi on nopeampaa. Liittämällä ennustedataa tuotannon mittauksiin voidaan tarjota ajantasainen ja päivittyvä kokonaiskuva siitä, mihin suuntaan tuotantoa tulisi missäkin tilanteessa ohjata. Tuotanto- ja ennustedataa visualisoidulla on mahdollista lisätä ymmärrystä koko tuotantoprosessin ja tuotannon tarpeen syy-yhteyksistä. Yhdistämällä ja visualisoimalla optimointi- ja talousdataa voidaan tuottaa ajan tasalla oleva kaukolämpötuotannon kokonaistaloudellinen tilannekuva, jonka avulla tuotantoa voidaan ohjata ympäristöystävällisempiin ja taloudellisesti tehokkaampiin ratkaisuihin.



## Lähteet

Anttila, P. & Rönkkönen M. 2017. Voimalaitoksen polttoaineen syötön optimointi. Saatavissa

<http://urn.fi/URN:NBN:fi:amk-2017090714881>

Duodecim Terveyskirjasto, Värisokeus ja poikkeava värinäkö 2020. Viitattu 8.5.2021. Saatavissa

<https://www.terveyskirjasto.fi/dlk00347>

Energiateollisuus ry 2020a, Kaukolämpötilasto 2019. Viitattu 15.4.2021. Saatavissa

[https://energia.fi/files/5384/Kaukolampotilasto\\_2019.pdf](https://energia.fi/files/5384/Kaukolampotilasto_2019.pdf)

Energiateollisuus ry 2020b, Kaukolämpötilasto vuositaulukot 2019. Viitattu 16.4.2021. Saatavissa

[https://energia.fi/files/5385/Vuositaulukot\\_19\\_FI.xlsx](https://energia.fi/files/5385/Vuositaulukot_19_FI.xlsx)

Energiateollisuus ry 2021. Energiantuotanto – Energiamaailma. Viitattu 23.5.2021. Saatavissa

<https://energiamaailma.fi/energiasta/energiantuotanto/>

Eskonen, H. 2020. Turpeen kiistely veroetu ei poistunutkaan, vaan tilalle ilmestyi yllättäen uusi mekanismi – riittääkö se puolittamaan turpeen polton? | Yle Uutiset | yle.fi. Viitattu 2.5.2021. Saatavissa

<https://yle.fi/uutiset/3-11547297>

Euroopan keskipitkien sääennusteiden keskus ECMWF. Copernicus Climate Data Store. Viitattu 14.4.2021. Saatavissa

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels?tab=overview>

Géron, A. 2017. Hands-On Machine Learning with Scikit-Learn & TensorFlow - Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol: O'Reilly Media Inc.

Hartikainen, J. 2021. Hiilipäästöjen hinnannousu Euroopassa on lyönyt kaikki ennusteet, ja sen seurauksena turpeesta tuli ”järkyttävän kallista” - Talous | HS.fi. Viitattu 2.5.2021. Saatavissa

<https://www.hs.fi/talous/art-2000007926977.html>

Hastie, T., Tibshirani, R. & Friedman, J. 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. painos. New York: Springer.

- Itkonen, J. 2020. 3.2 Päästökauppa | Välineet ilmastonmuutoksen hillitsemiseksi. Viitattu 2.5.2021. Saatavissa  
<https://www.ilmastoraportti.fi/paastokauppa.html>
- Kemin Energia ja Vesi Oy 2021a. Yhtiön historia. Viitattu 20.2.2021. Saatavissa  
<https://www.kenve.fi/yritys/historia/>
- Kemin Energia ja Vesi Oy 2021b. Kaukolämmön tuotanto. Viitattu 15.2.2021. Saatavissa  
<https://www.kenve.fi/palvelut/kaukolampo/kaukolammon-tuotanto/>
- Kemin Energia ja Vesi Oy 2021c. Vuosikertomus 2020. Viitattu 23.5.2021. Saatavissa  
<https://www.kenve.fi/wp-content/uploads/2021/05/KeEnVe-vuosikertomus-2020-net-1.pdf>
- Koponen, J. & Hildén, J. 2019. Data Visualization Handbook. Aalto Yliopisto. Espoo: Aalto ARTS Books
- Makkonen, I. 2014. Päästökauppa ja sen vaikutukset Etelä- Ja Keski-Pohjanmaalle. Viitattu 2.5.2021. Saatavissa  
<https://www.theseus.fi/bitstream/handle/10024/80850/B79.pdf>
- Elgabry, O. 2019. The Ultimate Guide to Data Cleaning. Viitattu 9.5.2021. Saatavissa  
<https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4>
- Jarosz M. 2021. 5 Practical Solutions to Make Responsive Data Tables. Viitattu 7.4.2021. Saatavissa  
<https://medium.com/appnroll-publication/5-practical-solutions-to-make-responsive-data-tables-ff031c48b122>
- Kosara R. 2007, Visualization Criticism – The Missing Link Between Information Visualization and Art. Saatavissa  
<https://kosara.net/papers/2007/Kosara-IV-2007.pdf>
- Mattila, M. 2003. Regressioanalyysin rajoitteet – KvantiMOTV. Viitattu 27.5.2021. Saatavissa  
<https://www.fsd.tuni.fi/menetelmaopetus/regressio/rajoitteet.html>
- Metsä Fibre 2021. Kemin biotuotetehdasprojekti. Viitattu 12.4.2021. Saatavissa  
<https://www.metsafibre.com/fi/yhtio/Kemin-biotuotetehdas/Pages/default.aspx>
- Microsoft 2021. Mikä Power BI on? Viitattu 26.2.2021. Saatavissa  
<https://docs.microsoft.com/fi-fi/power-bi/fundamentals/power-bi-overview>
- Microsoft 2020a. What is Power Query? Viitattu 8.5.2021. Saatavissa  
<https://docs.microsoft.com/en-us/power-query/power-query-what-is-power-query>

Microsoft 2020b. Data types in Power Query. Viitattu 8.5.2021. Saatavissa <https://docs.microsoft.com/en-us/power-query/data-types>

Microsoft 2020c. Optimize Power BI reports for the mobile app. Viitattu 4.5.2021. Saatavissa

<https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-create-phone-report>

Microsoft 2019, What is Azure IoT Edge? Viitattu 23.5.2021. Saatavissa <https://docs.microsoft.com/en-us/azure/iot-edge/about-iot-edge?view=iotedge-2020-11>

Montgomery, D., Peck, E. & Vining G. 2012. Introduction to Linear Regression Analysis. 5. painos. New Jersey: John Wiley & Sons Inc.

Motiva 2020. Uusiutuvan energian rahoitus. Viitattu 2.5.2021. Saatavissa [https://www.motiva.fi/ratkaisut/uusiutuva\\_energia/uusiutuva\\_energia\\_suomessa/uusiutuva\\_n\\_energian\\_rahointus](https://www.motiva.fi/ratkaisut/uusiutuva_energia/uusiutuva_energia_suomessa/uusiutuva_n_energian_rahointus)

Motiva 2019. Kaukolämmön hinta. Viitattu 24.5.2021. Saatavissa [https://www.motiva.fi/koti\\_ja\\_asuminen/rakentaminen/lammitysjarjestelman\\_valinta/lammitysmuodot/kaukolampo/kaukolammon\\_hinta](https://www.motiva.fi/koti_ja_asuminen/rakentaminen/lammitysjarjestelman_valinta/lammitysmuodot/kaukolampo/kaukolammon_hinta)

Mäkelä V-M & Tuunanen J. 2015. Suomalainen kaukolämmitys. Saatavissa <http://urn.fi/URN:ISBN:978-951-588-506-7>

Männistö, T. 2018. Methods for SE Research – Design Science. Viitattu 11.5.2021. Saatavissa

<https://courses.helsinki.fi/sites/default/files/course-material/4543078/MSER%20Design%20Science%2C%20CC%20BY-NC-SA.pdf>

Neittaanmäki, P. & Schroderus, O. 2010. Sähkön- ja lämmöntuotannon kustannussimulointi ja herkkyyshanalyysi. Viitattu 2.5.2021. Saatavissa

[http://www.mit.jyu.fi/scoma/energiasimulaattori/energiasimulointi\\_140610.pdf](http://www.mit.jyu.fi/scoma/energiasimulaattori/energiasimulointi_140610.pdf)

Nocedal, J. & Wright, S. 2006. Numerical Optimization. 2. painos. New York: Springer.

Simon, D. 2013. Evolutionary Optimization Algorithms - Biologically-Inspired and Population-Based Approaches to Computer Intelligence. New Jersey: John Wiley & Sons Inc.

Vähäkainu P., Kariluoto A. & Neittaanmäki P. 2019, Tekoäly ja rakennusten ennakoiva kunnossapito. Saatavissa

[https://www.jyu.fi/it/fi/tutkimus/julkaisut/it-julkaisut/tekoaly\\_ja\\_alykkaiden\\_rakennuksien\\_ennakoiva\\_kunnossapito.pdf](https://www.jyu.fi/it/fi/tutkimus/julkaisut/it-julkaisut/tekoaly_ja_alykkaiden_rakennuksien_ennakoiva_kunnossapito.pdf)

Tilastokeskus 2019. Sähkön ja lämmön tuotanto 2019. Viitattu 15.4.2021. Saatavissa [https://www.stat.fi/til/salatu/2019/salatu\\_2019\\_2020-11-03\\_tie\\_001\\_fi.html](https://www.stat.fi/til/salatu/2019/salatu_2019_2020-11-03_tie_001_fi.html)

Tilastokeskus 2021. Polttoaineluokitus 2021. Viitattu 20.3.2021. Saatavissa [https://www.stat.fi/tup/khkinv/khkaasut\\_polttoaineluokitus.html](https://www.stat.fi/tup/khkinv/khkaasut_polttoaineluokitus.html)

Työ- ja elinkeinoministeriö. Päästökauppa - Työ- ja elinkeinoministeriön verkkopalvelu. Viitattu 2.5.2021. Saatavissa <https://tem.fi/paastokauppa>

Visma 2019. Talouden tunnusluvut tutuksi - Opas lukujen tulkintaan. Viitattu 24.5.2021. Saatavissa <https://netvisor.fi/wp-content/uploads/2019/02/Tunnusluvut-tutuiksi-opas.pdf>

Visma a. Myyntikate (%). Viitattu 24.5.2021. Saatavissa [https://help.visma.net/fi\\_fi/financial-overview/content/online-help/kpi-gross-margin.htm](https://help.visma.net/fi_fi/financial-overview/content/online-help/kpi-gross-margin.htm)

Visma b. Käyttökate (%). Viitattu 24.5.2021. Saatavissa [https://help.visma.net/fi\\_fi/financial-overview/content/online-help/kpi-operating-margin-fi.htm](https://help.visma.net/fi_fi/financial-overview/content/online-help/kpi-operating-margin-fi.htm)

Ware, C 2021. Information Visualization: Perception for Design. Morgan Kaufmann. Cambridge, USA: Elsevier.

Witten, I., Frank, E. & Hall, M. 2011. Data Mining - Practical Machine Learning Tools and Techniques. 3. painos. Burlington: Elsevier Inc.

Yin M., Hollender D. Condelli L. Shewitz D. Duffy A, Movit M. 2014, The Power of Data Visualization: Advanced Presentations of NRS Data. Saatavissa [https://nrsweb.org/sites/default/files/Data\\_Visualization\\_Guide\\_508.pdf](https://nrsweb.org/sites/default/files/Data_Visualization_Guide_508.pdf)