

Heini Kaisto

**MASSADATAN KÄYTTÖ ASIAKASANALYTIKASSA
- CASE JYVÄSKYLÄN ENERGIA**



JYVÄSKYLÄN YLIOPISTO
TIETOJENKÄSITTELYTIETEIDEN LAITOS
2017

TIIVISTELMÄ

Kaisto, Heini

Massadatan käyttö asiakasanalytiikassa – case Jyväskylän Energia

Jyväskylä: Jyväskylän yliopisto, 2017, 84 s.

Tietojärjestelmätiede, Pro Gradu -tutkielma

Ohjaaja(t): Luoma, Eetu

Massadata on suuresti puhuttava lisäarvon luomisen lähde niin tieteellisille yhteisöille kuin yrityksillekin. Massadatan käsite elää kehittyvän teknologian mukana, mutta tavallisimmillaan sitä voi kuvata suurena määränä dataa, joka on monimuotoista ja kasvaa suurella nopeudella. Massadata mahdollistaa paljon liiketoiminnassa hyödynnettävässä analytiikassa, kuten liiketoimintatiedon hallinnan ja analytiikan prosesseissa sekä asiakasanalytiikassa. Asiakasanalytiikassa massadata mahdollistaa syvällisemmän asiakasymmärryksen saavuttamisen datasta ja sen hyödyntämisen asiakassuhdehallinnasta aina merkityksellisemmän asiakassegmentoinnin tekemiseen. Tutkimuksessa käsitellään ensin massadataa sen erilaisten määritelmien kautta sekä sen tunnettuja etuja ja hyötyjä, jonka jälkeen näitä käsitellään suhteessa massadatan hyödyntämiseen ja mahdollisuuksiin liiketoimintatiedon hallinnassa ja analytiikassa. Tutkielmassa käsitellään myös asiakasanalytiikkaa omana kokonaisuutenaan kattaen sen määritelmän, asiakassegmentoinnin suhde asiakasanalytiikkaan sekä parhaat keinot ja ongelmakohdat asiakassegmentoinnin toteuttamisessa. Seuraavaksi tutkimuksessa esitellään toteutettu empiirinen tutkimus, jonka tavoitteena oli määrittää, onko tapausyrityksellä kirjallisuudesta löydettyjen määritelmien mukaan massadataa ja jos on, löytyykö yrityksen tavasta hyödyntää massadataa yhtymäkohtia tai merkittäviä poikkeuksia kirjallisuudessa esitettyihin tapoihin ja kirjallisuuden pohjalta laadittuun viitekehukseen. Tutkimus toteutettiin teemahaastatteluna, johon haastateltaviksi valittiin tapausyrityksen työntekijöitä, jotka ovat tekemisissä asiakasdatan tai -analytiikan kanssa. Tutkimuksen tuloksina havaittiin, ettei tapausyrityksessä ole massadataa, mutta se voisi hyötyä joistain massadatalle ominaisista osaratkaisuksista.

Asiasanat: Massadata, Liiketoimintatiedon hallinta ja analytiikka, Datalähtöinen markkinointi, Asiakasanalytiikka, Asiakassegmentointi

ABSTRACT

Kaisto, Heini

Using Big Data in Customer Analytics – Case Jyväskylän Energia

Jyväskylä: University of Jyväskylä, 2017, 84 p.

Information Systems, Master's Thesis

Supervisor(s): Luoma, Eetu

Big Data is a widely-discussed topic on creating additional value for both scientific associations and companies. The definition of Big Data changes along with the technical environment, but usually it's described as a high volume of data that has variety and accumulates quickly. Big Data enables a lot in analytics processes in Business Intelligence and Analysis, as well as in customer analytics. In customer analytics, Big Data makes it possible to gain a deeper understanding of the customers, and use it to better the customer relationship functions as well as using it to make more meaningful customer groups in customer segmentation - process. The study first presents the definitions for Big Data and the problems and benefits with it. After this the study dedicates a whole chapter to customer analytics – its definition, how it's related to customer segmentation and the best-known practices and problems in customer segmentation. Next the framework for the empirical research is presented. The study was done as an interview survey where employees of the case company that dealt customer data or analyzed it were chosen for it. The research's aim was to find out whether or not the case study actually has Big Data and if so, does the way the case company uses Big Data for its customer analytics have similarities or remarkable differences to the ways presented in literature and the framework built in the study based on the literature. The results of the research were that the company did not have big data, but could benefit even in their current status from using some of the solutions used for Big Data.

Keywords: Big Data, Business Intelligence and Analysis, Data-Driven Marketing, Customer Analytics, Customer Segmentation

KUVIOT

KUVIO 1 Arvon luonnin prosessi segmentoinnissa (mukaiillen Kotler ym., 2008)	28
KUVIO 2 Datan hyödyntämismalli.....	38

TAULUKOT

TAULUKKO 1 Segmentoinnissa hyödynnettävät tekijät (mallintaen Kotler ym., 2008).....	31
TAULUKKO 2 Haastateltavien taustatiedot.....	49

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT

TAULUKOT

1	JOHDANTO.....	7
2	MASSADATA.....	10
	2.1 Massadatan määritelmä.....	10
	2.2 Massadatan etuja ja ongelmia.....	12
	2.3 Massadata liiketoimintatiedon hallinnassa ja analytiikassa.....	16
	2.4 Yhteenveto.....	22
3	ASIAKASANALYTIikka.....	23
	3.1 Asiakasanalytiikan määritelmä.....	23
	3.2 Asiakassegmentointi ja asiakasanalytiikka.....	25
	3.3 Asiakassegmentoinnin ongelmia ja parhaita ratkaisuja.....	28
	3.4 Yhteenveto ja viitekehysten rakentaminen.....	34
4	TEOREETTISEN OSUUDEN YHTEENVETO.....	41
5	EMPIIRINEN TUTKIMUS.....	45
	5.1 Tapaustutkimusyritys ja tutkimuksen tavoite.....	45
	5.2 Tutkimusmenetelmä ja tutkimuksen toteutus.....	46
	5.2.1 Kvalitatiivinen tutkimus.....	46
	5.2.2 Haastateltavien valitseminen.....	47
	5.2.3 Tutkimuksen toteutus.....	48
	5.3 Aineiston analyysi.....	49
6	TUTKIMUKSEN TULOKSET.....	51
	6.1 V-mallin tekijöiden esiintyminen yrityksessä.....	51
	6.1.1 Datan massa.....	51
	6.1.2 Datan monimuotoisuus ja rakenteettomuus.....	53
	6.1.3 Datan kertymisnopeus.....	54
	6.1.4 Datan Arvo.....	55
	6.2 Datan hyödyntäminen -viitekehysten toteutuminen yrityksessä....	56
	6.2.1 Datan kerääminen.....	56
	6.2.2 Datan varastointi.....	57
	6.2.3 Datan yhtenäistäminen ja koonti.....	58
	6.2.4 Datan mallinnus ja analyysi.....	58
	6.2.5 Datan tulosten tulkinta.....	59
	6.2.6 Datan käyttöönotto.....	59
	6.2.7 Datalinjaukset.....	60

6.3	Tapausyrityksen keräämä asiakasdata	61
6.3.1	Asiakkaan kotitalouden piirteet	61
6.3.2	Asiakkaan energiankulutusdata	62
6.3.3	Asiakkaan perusdata	62
6.3.4	Muu asiakkaaseen liittyvä informaatio	63
6.4	Yhteenveto	64
7	POHDINTA JA JOHTOPÄÄTÖKSET	66
7.1	Empiirisen tutkimuksen johtopäätökset	66
7.2	Empiirisen tutkimuksen suhde teoreettiseen tutkimukseen	69
7.3	Tutkimuksen rajoitukset ja käytettävyys	71
7.4	Eettinen tarkastelu	73
8	YHTEENVETO	74
8.1	Johtopäätökset	75
8.2	Jatkotutkimusaiheita	76
	LÄHTEET	78
	LIITE 1. TEEMAHAASTATTELURUNKO	82

1 JOHDANTO

Massadata on niin tieteellisissä yhteisöissä, liiketoiminnassa, kuin perinteisessä lehtimediassakin esiintynyt puheenaihe, jonka käsittelyssä tavallisesti nousee esille kaksi asiaa: se on jotain suurta ja uutta ja sen hyödyntämisellä on suuri potentiaali muuttaa nykyisiä data-analytiikan toimintoihin liittyviä prosesseja kaikilla toimialoilla. Teknologian kehittyessä myös yritysten keskinäinen kilpailu on kasvanut – yritykset eivät menesty enää pelkästään paikallisuutensa tai erityisen tuotetarjontansa perusteella. Sekä palveluja että tuotteita saa nykyisin tilattua internetistä ympäri maailman ja palveluntarjoajien on erottauduttava entistä tehokkaammin kohdennetun palvelutarjonnan sekä paremman asiakaspalvelukemuksen avulla. Sen vuoksi yrityksille on tärkeää ymmärtää, mitä heidän asiakkaansa oikein ovat, millaisia asioita he arvostavat ja mikä on saanut heidät valitsemaan juuri yrityksen palvelut. Massadatalla ja siihen pohjautuvalla asiakasanalytiikalla on potentiaalia tehostaa tätä asiakkaaseen liittyvän ymmärtämisen ja tiedon rikastamisen prosessia.

Massadatalla tarkoitetaan datamäärää, jota on nykyisillä metodeilla ja datan prosessointityökaluilla vielä mahdotonta käsitellä (Fan & Bifet, 2013). Massadataa voidaan jalostaa liiketoimintatiedon hallinnan ja analytiikan avulla, eli kehittämällä teknologioita, järjestelmiä, käytänteitä sekä sovelluksia data-analyysia varten, jolla pyritään saavuttamaan kriittistä tietämystä omasta liiketoiminnasta ja markkinoista (Lim, Chen & Chen, 2013). Liiketoimintatiedon hallintaa ja analytiikkaa hyödynnetään muun muassa datalähtöisessä markkinoinnissa, jolla tarkoitetaan markkinoinnin toimia, jotka pohjautuvat tiedonlouhintaan (Mulvenna, Norwood & Büchner, 1998). Asiakanalytiikka sen sijaan on kilpailuedun luomisen keino liikemaailmassa, jossa kehittyvä teknologia vähentää eroja eri palveluntarjoajien tuotteiden ja palveluiden välillä (Kolyshkina & Simoff, 2007). Asiakanalytiikkaa hyödynnetään asiakassegmentoinnissa, jonka avulla tunnistetaan ja arvotetaan eri asiakasryhmiä, joille tulee tuottaa erilaisia, heidän tarpeisiinsa ja haluihinsa soveltuvia ratkaisuja (Lynn, 2011).

Vaikka massadataa on tutkittu suhteellisen paljon, sen hyödyntämistä asiakasanalytiikassa on tutkittu suhteellisen vähän. Suurin osa tutkimuksia, jotka

osuvat tälle kohdealueelle, käsittelevät massadatan ja asiakasanalytiikan yhteyttä vain perinteisissä kuluttajayrityksissä, joissa asiakasanalytiikan toteutukseen on selkeitä, jo olemassa olevia malleja. Sen vuoksi on tärkeää tutkia massadatan laajempia hyödyntämismahdollisuuksia asiakasanalytiikassa myös yrityksissä, joiden palvelutarjonta ei kuulu perinteisiin kulutusmalleihin. Tämä tutkielma toteutettiin toimeksiantona Jyväskylän Energialle, joka on sähkö-, lämpö- ja vesipalveluita tarjoava Jyväskyläläinen yritys. Sen tarjonta ei kuulu perinteiseen kuluttajatarjontaan, jossa asiakkaiden käyttäytymistä voitaisiin analysoida pelkästään asiakkaan kulutuskäyttäytymisen ja asiointiaktiivisuuden mukaan. Yritys kuuluu siis juuri siihen kohderyhmään, jotka eivät saa nykyisistä tutkimuksista vastausta massadataa ja sen hyödyntämistä asiakasanalytiikassa koskeviin kysymyksiin. Lisäksi olemassa olevista tutkimuksista puuttuu yksinkertaiset mallit ja ratkaisut massadatan hallinnoimiseen ja hyödyntämiseen saatavilla olevien teknologioiden ja datan hyödyntämismallien avulla. Erityisesti teoksissa, joissa massadatan hyödyntämistä käsitellään asiakasanalytiikan näkökulmasta, teorit datan hyödyntämisestä ja potentiaalisista tehokkaista datalähteistä on ohitettu lähes täysin. Näitä aiemmasta tutkimuksesta puuttuvia osa-alueita on pyritty koostamaan kattavaksi, selkeämmin hahmotettaviksi kokonaisuuksiksi toteutetun teoreettisen tutkimuksen avulla.

Tutkimuksen tavoitteena oli koostaa olemassa olevasta massadataa, asiakasanalytiikkaa sekä massadatan hyödyntämistä asiakasanalytiikassa käsittelevästä kirjallisuudesta vastauksia ja ehdotuksia kysymyksiin siitä, mitä massadata on ja kuinka sitä voi hyödyntää. Tämän vuoksi tutkimus keskittyy avaamaan massadatan määritelmää ja luonnetta, avaten sen potentiaalia erityisesti liiketoiminnan saralla. Tämän lisäksi tutkimuksessa on käsitelty massadatan hyödyntämisen keskeisimmät ongelmat, käsitellen teknologian, organisaatioiden ja ihmisten sekä datan omistussuhteen epämääräisyyden sille asettamia rajoitteita. Tutkielma pyrkii avaamaan erilaisia osa-alueita ja prosesseja organisaatioissa, joissa massadataa voi hyödyntää ja jo hyödynnetään sekä esittelemään sen mahdollisuutta nykyisellään liiketoimintatiedon hallinnassa ja analytiikassa, johtaen tämän kautta sen mahdollisiin käyttöyhteyksiin erityisesti asiakasanalytiikassa ja siihen kuuluvassa asiakassegmentoinnissa. Tutkielmassa näihin aihealueisiin pyrittiin vastaamaan tutkielman tutkimuskysymyksen avulla:

- Kuinka massadata-pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa?

Tutkimuksen suorittamisessa hyödynnettiin seuraavia apukysymyksiä:

- Millä metodeilla yritys pystyy tehokkaasti tunnistamaan ja rikastamaan asiakasanalytiikan kannalta olennaista tietoa?
- Millainen asiakassegmentointimalli mahdollistaa yrityksen tunnistaa paremmin sekä potentiaalisia että jo palvelemaan asiakassegmenttejä?

Tutkielma toteutettiin perehtymällä massadataa ja asiakasanalytiikkaa käsittelevään kirjallisuuteen ja tutkimukseen, jonka jälkeen kirjallisuudesta tehtyjä havaintoja tutkittiin empiirisessä tutkimuksessa. Empiirinen tutkimus toteutettiin yrityksen työntekijöille toteutettuna teemahaastatteluna, jonka tavoitteena oli selvittää, onko yrityksellä massadataa ja jos on, kuinka he hyödyntävät sitä asiakasanalytiikassa ja onko hyödyntämisessä samankaltaisuuksia tai merkittäviä eroja kirjallisuudesta koostettuihin keinoihin.

Ensiksi tutkielmassa esitellään massadatan määritelmiä, sen hyötyjä ja suurimpia ongelmakohtia sekä liiketoimintatiedon hallinnan ja analytiikan roolia massadatan hyödyntämiseen. Seuraavaksi tutkielmassa käsitellään asiakasanalytiikkaa, avaten sen määritelmää, suhdetta asiakassegmentointiin sekä parhaita tapoja ja suurimpia ongelmakohtia asiakassegmentoinnin toteuttamisessa. Massadataa ja asiakasanalytiikkaa käsittelevien kappaleiden jälkeen tutkielmassa esitellään kirjallisuuden pohjalta laadittu viitekehys datan hyödyntämismallista, jota muun muassa toimeksiantajayrityksen kaltainen energia-alan yritys voi hyödyntää määritellessään nykyisiä datan keruu- ja käsittelyprosessejaan sekä mahdollisia keinoja toteuttaa asiakassegmentointia yrityksessä, joka ei tarjoa perinteiseen kulutuskäyttäytymiseen kuuluvia tuotteita tai palveluita. Teoreettisen osuuden yhteenvedon jälkeen tutkielmassa siirrytään käsittelemään toteutettua empiiristä tutkimusta, avaamalla tutkimukseen käytettyjä menetelmiä, kohde-ryhmiä ja tutkimuksesta tehtyjä havaintoja. Lopuksi tutkielmassa esitetään empiirisen tutkimuksen ja koko tutkielman yhteenvedo sekä tutkijan pohdinnat ja ehdotelmat jatkotutkimusaiheiksi aiheesta.

2 MASSADATA

Tutkielman ensimmäisessä pääluvussa keskitytään avaamaan massadatan käsitettä sekä tarkastelemaan massadatalle asetettuja odotuksia ja siihen liittyviä ongelmia. Tämän lisäksi kappaleessa käsitellään massadatalle olemassa sekä kehitteillä olevia työkaluja, joita massadataa kerryttävät tai sitä ostavat yritykset voivat hyödyntää liiketoimiensa edistämiseksi muun muassa liiketoimintatiedon hallinnan ja analytiikan kautta. Kappaleen yhteenvedossa esitellään tiivistettynä kappaleessa esitetyt asiat sekä tutkijan omat mielipiteet massadatanmääritelmästä, massadatalle esitettyjen potentiaalisten etujen ja sen nykyiseen hyödyntämiseen liittyvien ongelmien luonteesta sekä arvio massadatan käyttömahdollisuuksille liiketoimintatiedon hallinnassa ja analytiikassa.

2.1 Massadatan määritelmä

Massadata (eng. *Big Data*) on suhteellisen tuore käsite, joka vaikuttaa nousseen ajankohtaiseksi puheenaiheeksi vasta viime aikoina. Termi kuitenkin mainittiin ensimmäisen kerran jo vuonna 1998 Silicon Graphicsin (SGI) diaesityksessä otsikolla ”Massadata ja InfraStressin seuraava aalto” (Fan & Bifet, 2014). Massadata on seurausta maailmamme läpikäymästä *ICT-vallankumouksesta* (Nikulainen, 2013) ja *datafikaatiosta* (eng. *datafication*) (Ayankoya, Calitz & Greyling, 2014). ICT-vallankumouksen seurauksena on kehittynyt eräänlainen jatkuvasti enemmän ja nopeammalla vauhdilla dataa tuottava digitaalinen ekosysteemi (Nikulainen, 2013), jossa datafikaatiolla tarkoitetaan suurimman osan elämämme jokapäiväisistä toiminnoista muuttumista dataa tuottaviksi prosessiksi teknologian roolin kasvaessa. Datafikaation seurauksena dataa luodaan päivittäisellä tasolla kasvavissa määrin aiempaa enemmän (Ayankoya ym., 2014). Teknologiayritys IBM:n vuoden 2011 vuosiraportti konkretisoi hyvin datafikaation ja ICT-revoluution aiheuttamat muutokset maailmassa tuotettavan datan määrään – IBM:n raportin mukaan 90 prosenttia vuoden 2011 loppuun mennessä koko maailmassa tuotetusta datasta oli luotu sitä edeltäneen kahden vuoden aikana (IBM, 2011). Luodun datan määrän radikaalisen kasvun myötä on helppo ymmärtää, miksi massadata ja sen erilaiset hyödyntämismahdollisuudet kiinnostavat monenlaisia organisaatioita. Kun tämän lisäksi vielä datan käsittelyn ja varastoinnin kustannukset ovat laskeneet kehittyneen teknologian myötä, organisaatioilla on suurempi mahdollisuus kasvavaan määrään datan keräämistä ja hyödyntämistä. (Nikulainen, 2013). Tällaiset muutokset teknologisessa ympäristössämme ovat luoneet pohjan tutkimuksemme pääaiheelle – massadatalle.

Vaikka massadata on herättänyt paljon huomiota tieteellisten artikkeleiden ja tutkimusten lisäksi myös perinteisessä mediassa, se on silti yhä konsepti, jolle ei ole olemassa yhtä yhteistä sovittua ja yleisesti hyväksyttyä määritelmää. Osa

sitä käsittelevistä artikkeleista, kuten Fanin ja Bifetin vuonna 2013 julkaistu artikkeli, kuvaa sitä *datamääräksi, jota on nykyisillä metodeilla ja datan prosessointityökaluilla vielä mahdotonta käsitellä* (Fan & Bifet, 2013). Suurin osa massadatan kuvauksista kuitenkin keskittyvät kuvaamaan tapahtunutta muutosta datan keräämisessä, varastoisemisessa ja noutamisessa (Power, 2013). Tällaisille konkreettisimmille määritelmille on yhteistä, että massadataa kuvataan datana, joka mukailee 3V-mallia: datan määrä on suuri (eng. *volume*), data on monimuotoista ja strukturoimatonta (eng. *variety*) ja datan määrä kasvaa nopeasti (eng. *velocity*) (Szymczak, Zelik & Elm, 2014). Vaikka 3V-malli on suosittu massadataa käsittelevissä teoksissa, joidenkin tutkimuksien mukaan tämä ”perinteinen” 3V-malli on jo vanhentunut tapa kuvata massadataa. Näiden teosten mukaan kattavamman massadata -kuvauksen saamiseksi olemassa olevaan malliin on lisättävä vielä kaksi kuvaajaa, eli kaksi V-kirjainta lisää: datan oleellisuuden ja arvokkuuden tunnistaminen tietyn ongelman kannalta (eng. *veracity*) ja ajankohtaisuuden nopea vanheneminen alati kertyvän uuden datan myötä (eng. *volatility*) (Ayan koya ym., 2014). Näiden lisäkuvaajien ohelle on ehdotettu vielä kuudetta V-kirjainta, arvoa (eng. *value*), jolla tarkoitetaan piilevää potentiaalista arvoa suuressa määrää dataa, josta hyödyllisen datan noutamiseen tulisi löytää keino (Chen, Kazman, Haziyevev & Hrytsay, 2015). Massadata koostuukin perinteisen strukturoidun eli rakenteellisen, riveihin ja kolumneihin järjestettävissä olevan datan lisäksi, raakadatasta (kuten sensoridatasta), puolirakenteellisesta datasta (kuten web-logit) ja rakenteettomasta datasta (kuten twiittien tekstisisältö). Nämä perinteisestä poikkeavat datamuodot eivät sovellu perinteisten datasemantiikan ja tietovarastoinnin käsiteltäväksi (Martin, Avital & Jensen, 2013), vaatien erilaisten lähestymistapojen, työkalujen ja ratkaisujen löytämistä sen käsittelemiseen ja hyödyntämiseen, samalla kun se tarjoaa monimuotoisuudellaan mahdollisuuksia asioiden syvällisempään ja oivaltavampaan tutkimiseen.

Tärkeää massadatan käsitteessä on sen muodostavat englanninkieliset termit: *Big* ja *Data*. Siinä, missä data tarkoittaa kerättävissä olevan, varastoitavan ja päätöksen tekoon hyödynnettävää faktojen esiintymää (Ayan koya ym., 2014), massadatasta puhuttaessa on tärkeää ymmärtää, ettei etuliite *Big* (suom. *iso/suuri*) viittaa vain datan volyymiin, eli määrään, vaan *Big* käsittää myös datan monimuotoisuuden ja kasvunopeuden. (Frans, 2012, s.5). Tämä tarkoittaa, ettei suuri määrä dataa yksistään aina välttämättä ole massadataa, jos sen kasvunopeus ei ole huomattava ja datan kerääminen tapahtuu hyvin homogeenisestä lähteestä – tällöin kerättävä data ei ole monimuotoista ja strukturoimatonta, eli massadatan määritelmän monimuotoisuus ja kertymisnopeus -tekijät jäävät puuttumaan. Massadatan määrälle, eli volyymille ei ole kuitenkaan aiheutta käsittelevässä kirjallisuudessa käsitelty mitään numeraalista rajaa, jonka ylitettyään data täyttää massadatan määritelmän – koska massadata kuvastaa sellaisia datamääriä, joiden käsittelyyn vaaditaan uudenlaisia työkaluja ja varastointiratkaisuja sen hyödyntämiseksi, massadatan määritelmä muuttuu aina tietokoneiden varastointi- ja prosessointitehojen kasvamisen myötä. Hyvä tapa ymmärtää massadatan käsite on verrata sitä sen vastakohtaan – *Small Dataan*. Niin sanottu small data, eli

suomennettuna pieni data, kuvaa tietotekniikassa ja yrityksissä strukturoitua dataa, jota voidaan hallinnoida perinteisillä relaatiodatavarastoilla (Marton, Avital & Jensen, 2013). Massadata kuvaa siis dataa, joka ei sovellu small datan määrittelymään. Massadatan luonteen mahdollistamia hyötyjä ja sen aiheuttamia ongelmia mm. datan prosessoinnissa ja varastoinnissa käsitellään tarkemmin seuraavassa kappaleessa.

Massadata kuvaa siis datamääriä, joita on tämän hetkisillä ratkaisulla vielä haastavaa hyödyntää ja jota kuvataan datan määrän, kasvunopeuden ja monimuotoisuuden kautta. Datan määrä ei ole kuitenkaan tiettyyn datamäärään sitoutettu konsepti, vaan massadatan määrän käsite elää kehittyvän teknologian mukana, sitoutuen sen kykyyn varastoida kasvavia määriä dataa ja käsittelemään datamääriä nopeammin. Mitä enemmän teknologia kehittyy, sitä haastavammille asteikolle massadatan käsite myös siirtyy. Tämän perusteella massadataa voisikin kuvata vielä pääasiallisesti valjastamattomaksi potentiaaliksi hyödyntää datafikoituvan yhteiskunnan tuottamaa yhä monimuotoisempaa ja määrällisesti suurempaa dataa. Ratkaisuja ja ongelmia massadatan potentiaaliseen hyödyntämiseen käsitellään seuraavassa kappaleessa, avaten yrityksiä ja organisaatioita, jotka jo hyödyntävät massadataa palveluntarjonnassaan sekä datalähteiden kirjoa, jotka lukeutuvat massadataan ja ovat itsessään muuttaneet datasta rikastettavan tiedon hyödyntämismahdollisuuksia.

2.2 Massadatan etuja ja ongelmia

Massadataa on kuvattu informaatiovoimavaraksi, joka vaatii kustannustehokkaita ja innovatiivisia ratkaisuja informaatioprosessoinnissa tuottaakseen tehokkaita oivalluksia ja päätösten tekoa organisaatiolle (Gartner, 2012). Tämä kuvaus heijastaa hyvin sekä massadatan potentiaalisia etuja, että sen hyödyntämiseen liittyviä ongelmia. Vaikka esimerkiksi *massadatan hallinnointiin* (eng. *Big Data Management, BDM*) on olemassa jo useita työkaluja, massadataan pätee sama sääntö kuin mihin tahansa muuhun dataan – ilman oikeaa lähestymistä siihen, mitä datasta halutaan selville, olemassa olevat työkalut ovat hyödyttömiä (Frans, 2012, s. viii). Myös Reed ja Dangorra (2015) nostavat esille saman huolen *massadatan hyödyntämisen osaamisesta* käsitellessään massadataa suuresti puhuttavana potentiaalisena lisäarvon luomisen lähteenä eri toimialoille. He ovat artikkelissaan selkeästi sitä mieltä, että massadata on lähde arvotarjonnan edistämiseen, mutta sen hyödyntämisen ongelmaksi muodostuu, ettei massadatan tehokkaaseen hyödyntämiseen ole vielä löydetty toimivaa ja helposti implementoitavaa ratkaisua (Reed & Dongarra, 2015). Massadatan potentiaalisista eduista ja sen hyödyntämisen ongelmista puhuttaessa onkin tärkeää ymmärtää datan rooli hyötyjen toteuttamisessa - data toimii informaation ja faktojen keräämisen, järjestämisen ja kuvaamisen pohjana, joiden hallinta ja käsittäminen on sekä liiketoiminnan että tieteellisen tutkimisen elinehto. Tuo data muodostuu ja on sitoutunut yksittäis-

ten henkilöiden jokapäiväiseen elämään, jota keräämällä asiakaskuntaa ja tutkittavia kohteita ja ilmiöitä on mahdollisuus ymmärtää paremmin. (Ayankoya ym., 2014.)

Hyötyjen saaminen massadatan käytöstä onkin monivaiheinen prosessi, josta suurin osa tutkimuksista keskittyy vain muutamaaan vaiheeseen. Kokonaisuudessaan se kuitenkin kattaa datan keräämisen, tiedon erittelyn ja siistimisen, datan yhdentämisen, mallinnuksen ja analyysin, tulosten tulkinnan sekä käyttöönoton. (Jagadish, Gehrke, Labrinidis, Papakonstantinou, Patel, Ramakrishnan & Shahabi, 2014.) Benjamins (2014) kuitenkin esittelee tiiviissä artikkelissaan hyvin massadatan mahdollisuuksia yrityksen sisäisissä sekä ulkoisissa prosesseissa. Sisäisissä prosesseissa massadataa voi käyttää yrityksen liiketoimintaprosessien edistämiseen hyödyntäen yrityksen sisäisissä prosesseissa kertyvää dataa, joita yrityksen sisäinen analytiikka monissa tapauksissa jo tietyssä määrin valmiiksi analysoi. Tästä kertyvää tietoa voi käyttää päätöksenteon tukena mm. markkinoinnissa, kampanjoinnissa, poistuneiden asiakkaiden analysoinnissa ja prosessien tehostamisessa. Ulkoisissa prosesseissa yksi luonnollisista käyttöyhteyksistä massadatalle on mainonta, eli asiakasprofiilien luominen ja niiden hyödyntäminen kohdennetussa mainonnassa. Hyvä esimerkki yrityksistä, jotka ovat toteuttaneet tätä jo pidemmän aikaa, ovat hakukoneet Google ja Yahoo, jotka kohdentavat heiltä ostettuja mainospaikkoja juuri asianomaisille ihmisille suositeltujen hakutulosten ja sivumainoksien muodossa. Lisäksi, jos organisaatio itsessään kerää prosesseissaan ja toiminnassaan arvokasta tietoa esim. asiakaskannastaan, on sillä mahdollisuus kaupata tätä dataa ulkopuolisille toimijoille, jotka hyödyntävät sitä omissa prosesseissaan toimintansa ja asiakaskokemuksen edistämiseksi. (Benjamins, 2014). Myös massadatalle ominainen monimuotoisuus, useasta datalähteestä kerätty monimuotoinen rakenteeton ja puolirakenteinen data, antaa sen hyödyntämiselle uusia potentiaalisia ulottuvuuksia. Yuan ja Ho (2015) sekä Benjamins (2014) ovat listanneet massadatalle useita lähteitä, jotka poikkeavat perinteisistä datalähteistä ja voivat tuottaa lisäarvoa organisaatioille erilaisuudellaan. Yuan ja Ho (2015) ovat listanneet perinteisen internetistä saatavan datan lisäksi lähteiksi GPS-laitteet ja niiden antamat sijaintitiedot, mobiililaitteet, älykellot ja muut päälle puettavat älylaitteet, radiotaajuuksien tunnistuksen (eng. *radio frequency identification, RFID*), online-videot ja sosiaalisen median esimerkiksi massadatalle ominaisesta monimuotoisesta, osittain rakenteellisesta ja rakenteettomasta datasta (Yuan & Ho, 2015). Näiden lisäksi Benjamins (2014) esittää omassa listauksessaan myös koneiden välisestä vuorovaikutuksesta ja esineiden internetistä (eng. *Internet of Things, IoT*) muodostuvan sensoridatan, avoimen datan, mitä useimmiten hallitukset ja tieteelliset organisaatiot julkaisevat sekä liiketoimintadatan, kuten verkkokaupoista kertyvä datan. (Benjamins, 2014).

Massadatan hyötyihin liittyvässä tutkimuksessa esitellään useita erilaisia hyötyjä sen käytöstä. Yleisesti massadatasta keskustellessa esille nousee sen yleinen potentiaali - tutkimukset ovat osoittaneet organisaatioiden, jotka hyödyntävät massadataa olevan tuottavampia niin rahallisesti kuin tuotannollisesti verrattuna massadataa hyödyntämättömiin toimijoihin (Ayankoya ym., 2014).

Myös Nikulainen (2013) nostaa tämän tutkimuksessaan esille – yritysten mahdollisuus kerätä suurta määrää dataa ja tehdä näiden avulla uusia innovatiivisia ratkaisuja sekä ymmärtää heidän omaa liiketoiminta-alaansa paremmin antaa heille suuria mahdollisuuksia tuotannon lisäämiseen sekä tuotanto- ja palvelukustannusten pienentämiseen (Nikulainen, 2013).

Massadata mahdollistaa täysin uudenlaisten ja uudella tavalla hallinnoitavien yritysten muodostamisen, jossa tavallisesti ihmisten harkinnalle jääneet päätöksenteot korvataan dataan pohjautuvalla, algoritmeja hyödyntävällä analyttisellä päätöksenteolla (Ayankoya ym., 2014). Massadatan analytiikan soveltamisen potentiaali ylittää kaikki yrityksille, ihmisille ja yhteiskunnalle yhteisesti asetetut strategisen hallinnoinnin rajoitteet. Kehittämällä yksilöistä ennustuksia tekeviä algoritmeja hyödyntäen monimutkaisen rakenteen omaavaa ja useasta lähteestä kerättyä dataa, ihmisten tuottamia datajälkiä voidaan hyödyntää muun muassa yritysten työntekijöiden hallinnoinnissa, markkinatutkimuksissa, yritysten sisäisten prosessien tehostamisessa ja automatisoinnissa sekä palvelujen kohdistamisessa ja personalisoinnissa asiakkaille sopiviksi (Sodenkamp, Kozlovskiy & Staake, 2015).

Massadatan potentiaali ja sen luvatut hyödyt ovat siis suuret, mutta aiheesta tehty tutkimus ei esittele vain massadatan hyötyjä – massadatan hyödyntämisessä on vielä paljon teknisiä haasteita, joita on tutkittava, jotta massadataa voidaan hyödyntää sen täydessä potentiaalissa (Jagadish ym., 2014). Näitä ovat muun muassa datan varastoimisen, noutamisen ja muuntamisen haasteet, jotka on ratkaistava ennen kuin massadatan saadaan analytiikan työkaluilla hyödynnettävissä olevia tuloksia. (Kowsar ym., 2014). Jagadish ym. (2014) nostavat nämä haasteet esille omassa työssään ilmaistessaan Massadatan perinteiseen datan käsittelyyn tekemän muutoksen vaatimalla monimuotoisen datan menevän läpi kaikki viisi datankäsittelyn prosessia ilman, että niitä voi jakaa enää itsenäisiin eristettyihin silloihin datan käsittelemiseksi: massadatan käsittelyratkaisun tulee mahdollistaa 1. *Datan hankinta*, 2. *Tiedon rikastus ja siistiminen*, 3. *Datan integrointi, koostaminen ja esitys*, 4. *Datan mallinnus ja analysointi* sekä 5. *Tulosten tulkitseminen* raakadatan, puolirakenteellisesta ja rakenteellisesta datasta samalla tapaa, tehokkaasti ja nopeasti aiheuttamatta liikaa varastointi ongelmia datan suuresta määrästä johtuen (Jagadish ym., 2014). Lisäksi massadatan hyödyntämiseen liittyy yksi suuri ongelma ja aiheeseen liittyvissä tutkimuksissa usein esiintynyt aihe, jolle ei ole vielä saatu määritettyä, missä raja kulkee – käsiteltävän *datan suojaus ja yksityisyys*. Datan omistajuus, missä vaiheessa yksityisten henkilöiden tuottaman datan omistajuus siirtyy sen tuottajalta eri organisaatioille hyödynnettäväksi, on suuresti puhuttava puheenaihe data-analytiikan saralla tällä hetkellä ja ennen näiden ongelmien ratkaisemista massadatan ja sen hyödyntämisen mahdollistavan data-analytiikan kehittäminen ei voi lähteä nousuun (Benjamins, 2014).

Suurimmat ongelmat massadatan hyödyntämisessä liittyvät kuitenkin sen määritteleviin tekijöihin – sen määrään ja kasvunopeuteen. Nämä aiheuttavat ongelmia kaikissa perinteisissä datan käsittelyprosesseissa, joissa dataa yritetään siirtää paikallisiin datavarastoihin. Datansiirtoa rajoittaa verkon nopeus sekä

laitteiston kyky rinnakkaisen datan lähettämiseen, jonka vuoksi datan määrän kasvaessa datan lähettämisen ja siirtämisen hyötysuhde laskee. (Kowsar ym., 2014). Ayankoya ym. (2014) esitteleekin hyvin ne tekijät, jotka yrityksen tulee huomioida hyödyntääkseen massadataa yrityksen toimintojen edistämisessä, kuten implementoinnissa yrityksen liiketoimintatiedon hallintaan ja analytiikkaan:

- *Datakeskeisten toimintamallien omaksuminen:* Datan hyödyntämiseksi ongelmanratkaisussa ja päätöksenteossa näiden perinteisten prosessien on muututtava datakeskeisemmiksi. Tämä tarkoittaa, että esim. tavoitteita määrittäessä niiden tulee olla määritettynä siten että niihin päästään datapohjaisella lähestymistavalla.
- *Oleellisen datan tunnistaminen:* Massadata luo paljon tietoa globaalisti ympäri maailman, yleisestä tutkimustiedosta hallituksien ja hallintoelimien julkaisemiin globaaleihin tietopankkeihin. Ongelmana onkin hairautta kaiken saatavilla olevan datan käyttöön, mikä eliminoi datan paljoudesta, monimuotoisuudesta ja kertymänopeudesta saatavat edut. Tärkeintä datan hyödyntämisessä on toiminnoilleen ja tavoitteelleen oleellisen datan tunnistaminen, jonka tunnistamista voi helpottaa laatimalla datalinjauksen yrityksen toimintoja ja tavoitteita edistävästä datasta.
- *Osaamisen hyödyntäminen:* Massadatan käyttöönotto ja hyödyntäminen vaativat osaamista usealta eri osa-alueelta, kuten tietotekniikasta, statistiikasta, koneoppimisesta, datan visualisoinnista, viestinnästä ja verkkotunnusasiiantuntijuudesta. Yritysten tulee joko tunnistaa nämä lahjakkuudet oman organisaationsa sisältä ja luoda tiimejä heistä kattaakseen vaadittavien taitojen monimuotoisuuden massadatan asettamia haasteta käsitellessä tai palkata ulkopuolisia ammattilaisia, joilta löytyy osaaminen aiheesta jo valmiiksi. Molemissa tapauksissa parhaimman mahdollisen lopputuloksen saavuttamiseksi myös johdon tulee olla tietoinen datalähtöisen ajattelun keskeisyyden tärkeydestä ja kyetä itsekkin ajattelemaan ja hallinnoimaan asioita data-keskeisesti.

Tämä menestystekijälähestyminen erittelee hyvin myös ne ongelmat, joihin yrityksen tulee vastata hyödyntääkseen massadataa tehokkaasti – prosesseilleen ja tavoitteilleen keskeisten tietojen tunnistaminen suurista datamääristä, näiden tunnistamiseksi vaadittavien linjausten tekeminen ja tulosten siirtäminen osaksi tavallista päätöksentekoprosessia muuttamalla tavallisia prosesseja data-keskeisemmiksi massadatan vaatimien lahjakkuuksien omaavan ryhmän tai ulkopuolisten data-keskeisten lahjakkuuksien johdolla. (Ayankoya ym., 2014.)

Datan suuren määrän vuoksi sen käsittelyn tekniset haasteet ovat suurin ongelma sen potentiaalın hyödyntämiseksi. Datan määrä on massadatan hyödyntämisen ymmärrettävin ongelma, mutta myös massadatalle ominaiset datan määrän nopea kasvu ja monimuotoisuus ovat sen käsittelyyn vaikuttavia haasteita, joita ei saa unohtaa massadatan käsittelyä ja hyödyntämistä käsitellessä.

Varieteetti eli datan monimuotoisuus ja erilaiset datatyypit sekä sen semanttinen tulkinta asettavat datan hyödyntämiselle haasteita. Datan kasvunopeus taas aiheuttaa haasteita kertyvästä datasta jalostettavaan tietoon vaadittavasta reagointinopeudesta tietojen hyödyntämiseksi. Tämän vuoksi on erityisen tärkeää löytää ratkaisuja, kuten sen algoritmeja massadatan käsittelyyn, joka mahdollistaa nopeasti kertyvästä, monimuotoisesta, suuresta määrästä dataa hyödyllisten tietojen rikastamisen ilman, että juuri nämä aspektit koostuvat sille ongelmaksi varastoinnin ja käytettävissä olevan laitteiston suoritustehon puutteiden vuoksi. Massadatan potentiaali sitoutuukin suuressa osin sen ymmärtämiseen siitä, mitä dataa organisaatioiden tulisi pyrkiä hyödyntämään ja rikastamaan tiedoksi toimintojensa edistämiseksi ja ymmärtää myös datan hyödyntämisen organisaatiolle asettamat vaatimukset prosessien, toiminta- ja ajattelumallien muuttamiseksi datakeskeisemmiksi. Massadatan konkreettisempia mahdollisuuksia liiketoimintatiedon hallinnassa ja analytiikassa, sille tuomaa lisäarvoa ja erilaisia olemassa olevia ja kehitteillä olevia ratkaisuja käsitellään syvällisemmin seuraavassa osiossa.

2.3 Massadata liiketoimintatiedon hallinnassa ja analytiikassa

Liiketoimintatiedon hallinta ja analytiikka (eng. *Business Intelligence and Analysis*) on teknologioiden, järjestelmien, käytänteiden sekä sovellusten kehittämisestä data-analyysia varten, jolla pyritään saavuttamaan kriittistä tietämystä omasta liiketoiminnasta ja markkinoista (Lim, Chen & Chen, 2013). Liiketoimintatiedon hallinta ja analytiikka ovat datakeskeisiä lähestymistapoja, jotka hyödyntävät metodologioita, prosesseja, teknologioita ja eri työkaluja tiedon rikastamiseen datasta. (Ayankoya ym., 2014). BI&A toimii työkaluna kasvavien datamäärien hyödyntämiseksi resurssina liiketoiminnallisten toimien, kuten toiminnan tehostamisen, kustannustehokkuuden lisäämisen ja paremman tuotekehityksen ja kohdistamisen toteuttamiseksi (Krishnamoorthi & Mathew, 2015). Ayankoya, Calitz ja Greyling (2014) kuvaavat teoksessaan liiketoimintatiedon hallintaa ja analytiikka korvaamattomaksi tekijäksi liiketoiminnan päätöksenteossa ja datahallinnoissa. *Hallinta* -termillä kuvataan pääsyä käsiksi käsiteltävään dataan ja siitä raportointia, kun taas *analytiikka* tarjoaa mahdollisuuden tiedon rikastamiseen siitä (Ayankoya ym., 2014).

Data ja analyysi kulkevat aina käsi kädessä. Davenport, Harris ja Morison (2010) ovat todenneet, että parempien ratkaisujen ja oikeiden toimintojen suorittamiseksi, analytiikan käyttäminen on välttämätöntä ja analytiikan suorittamiseksi dataa ja analyysia on hyödynnettävä. Massadatan ja liiketoimintatiedon hallinnan ja analytiikan suhteessa on kuitenkin tärkeää muistaa, että massadatan käyttö itsessään ei automaattisesti tarkoita menestymistä – dataa analysoidessa, riippumatta sen koosta, täytyy erottaa ne resurssit ja lähteet jotka parhaiten edistävät yrityksen liiketoimintaa ja jotka datasta tiedoksi rikastettuna tuovat eniten lisäarvoa yritykselle (Frans, 2012, s.6).

Liiketoimintatiedon hallinnan ja analytiikan mahdollisuudet tällä saralla ovat kehittyneet sen evoluution myötä, sen kehittyessä pelkkiä tietovarastoja hyödyntävästä työkalusta mahdollisuudeksi hyödyntää myös strukturoimatonta ja web-pohjaista sisältöä. Tämä on mahdollistanut myös asiakkaan tuottaman datan hyödyntämisen mm. sosiaalisen median julkaisuista, mobiililaitteista, mobiiliapplikaatioista sekä erilaisista sensoreista. LaValle ym. (2011) esittävät massadatan laajenuksena liiketoimintatiedon analytiikalle. (LaValle, 2011; Constantiou & Kallinikos, 2015). Massadatan luonne määrittääkin nykyisellään sitä, kuinka liiketoimintatiedon hallintaa ja analytiikkaa käytetään ja missä kontekstissa sitä hyödynnetään – siinä, missä liiketoimintatiedon hallintaa ja analytiikkaa hyödynnettiin ennen selvästi rakenteellisten tiedonhallintajärjestelmien (eng. *database management system, DBSM*) yhteydessä, massadatan monimuotoisuuden vuoksi liiketoimintatiedon hallinnan ja analytiikan on pitänyt sopeutua käsittelemään myös strukturoimatonta tietoa, kuten web-pohjaista tietoa. Oikein käytettynä massadata yhdessä liiketoimintatiedon hallinnan ja analytiikan kanssa tarjoaa sille täysin uusia mahdollisuuksia - liiketoimintatiedon hallintaa ja analytiikkaa on ennen tavallisesti käytetty menneiden tapahtumien raportointiin ja ymmärtämiseen, massadata ja sen analytiikka tarjoavat mahdollisuuden tapahtumien ennustamiseen ja uusien liiketoimintavisioiden löytämiseen. (Ayankoya ym., 2014.)

Massadata -analytiikka voidaan määritellä sen vaatimuksesta joustavaan, korkean käsittelytason vaativaan tietokoneanalyysimetodeihin joita hyödynnetään suureen määrään heterogeenisiä datakokoelmia (Kowsar, Dashnov & Lonie, 2014). Massadatan käsittelyratkaisut vaihtelevat kokonaisratkaisuksista, jotka kattavat ratkaisut datan varastoinnista tiedon rikastamiseen, aina yksittäisiin ratkaisuihin saakka. Esimerkkejä prosesseista, joita ei voi yksistään hyödyntää massadatan käsittelyssä, mutta ovat tärkeässä osassa sen prosessia, ovat *ETL-prosessi* (Nakucçi, Theodorou, Jovanovic & Abelló, 2014) ja *tosiaikainen tiedon jalostus* (eng. *On-line analytical processing, OLAP*) (Lao & Tang, 2005).

ETL-prosessi, eli datatiedoston purku (eng. *Extract*), muuttaminen toiseen muotoon (eng. *Transform*), lataus (eng. *load*) -prosessi on datan käsittelyprosessi, jossa dataa haetaan useasta eri datalähteestä, muutetaan yhdeksi yhteiseksi dataformaatiksi siistimisen ja järjestämisen avulla ja ladataan tietokantaan. ETL on toimiva, mutta suhteellisesti hidaskäyttöprosessi, josta erityisesti datan muuntaminen vie aikaa, n. 80 % prosessin koko toteutuksesta. (Lao & Tang, 2005.) Prosessi on kuitenkin herkkä virheille, joka vaatii nykyisellään vielä suuren työpanoksen manuaalisesti sen hallinnoinnissa (Nakucçi ym., 2014).

Tosiaikainen tiedon jalostus sen sijaan on prosessi, jossa voi käsitellä ja rakentaa moniulotteista dataa välittömästi ja se soveltuu erityisesti datasta tehtäviin laskuihin, trendien ennustamiseen sekä hienovaraisen datan käsittelyyn (Lao & Tang, 2005). Termien sisäistäminen on tärkeää erityisesti käsitellessä massadatan teknisiä sovelluksia ja niiden soveltuvuutta hyödyntämiseen asiakasanalytiikassa. Esimerkiksi tosiaikaista tiedon jalostusta voidaan hyödyntää osana

massadatan rakenteellisen datan käsittelyä, mutta se ei vastaa niinkään massadatan strukturoimattoman datan käsittelyn tarpeisiin, toisin kuin ETL-prosessi, jonka avulla strukturoimatonta dataa voidaan yhtenäistää.

Muita massadatan käsittelemiseen soveltuvia, jo olemassa olevia ratkaisuja ovat *Hadoop* ja *MapReduce* (LeFevre, Sankaranarayanan, Hacigumus, Tatemura, Polyzotis & Carey, 2014), joita käsitellään suuressa osassa massadatan käsitteelyyn ja hyödyntämiseen liittyvissä artikkeleissa. Hadoop on avoimen lähdekoodin sovelluksen viitekehys, jota hyödynnetään hajautettujen tietokantojen ja hajautettujen tiedon prosessoinnin käsittelyssä suurilla datamäärillä klustereiden pohjalta (Wang, Qiu, Yang, Dong, Li & Li, 2009). Hadoop hyödyntää toiminnassaan Googlen MapReduce -ohjelmointimallia, joka mahdollistaa rinnakkaisen, samanaikaisen ohjelmoinnin ja tiedon käsittelyn, sovellusten ja tietokantojen itsenäisen virrehallinnoinnin sekä suuren skaalan datan helpon hallinnoinnin (Ullman, 2015). Hadoop ja MapReduce ovat kuitenkin teknisiä ratkaisuja, joita voi hyödyntää suunnitelmassa omia metodejaan tehokkaaseen ja toimivaan massadatan käsittelyyn, eivätkä niinkään valmiita algoritmeja tai sovellusratkaisuja, kuten esimerkiksi IBM:n *Watson Analytics*. *Watson Analytics* tarjoaa mahdollisuuden käsitellä massadatan varastointiongelmia ja yrityksen tavoitteiden kannalta oleellisten datakyselyjen tekemistä yksinkertaisesti, ilman suurempaa tietämystä skriptien kirjoittamisesta (IBM.com, 2016). Googlen *BigQuery* alusta on näistä vaihtoehdoista puoliväli-ratkaisu, joka toimii pilvessä SQL-tyyppisillä kyselyillä, mahdollistaen itse noudetun datan varastoinnin Googlen pilvivarastoihin ja datan ryhmittelyn ja käsittelyn käyttäjän tekemien tietokantahakujen pohjalta. (Fernandes & Bernardino, 2015.)

Massadatalta ja sen tehokkaasta hyödyntämisestä noussut tarve uusille käsittelyratkaisulle onkin herättänyt paljon huomiota informaatioteknologian ja tietotekniikan tutkijoiden keskuudessa, jotka ovat laatineet lukuisia erilaisia ehdotelmia ja jo käyttöön otettavia ratkaisuja jotka vastaavat joko yhteen tai useampaan massadatan käsittelyyn pääasiallisesti liittyvistä ongelmista – varastoinnin ongelma (volyymi), raakadatan, puolirakenteellisen ja rakenteettoman datan aiheuttamiin ongelmiin datan käsittelyssä ja vertailussa (varieteetti) sekä prosessoinnin keston ongelmaan nopeasti kertyvästä datasta ajankohtaisten tietojen saamiseksi (velositeetti). Esimerkiksi juuri LeFevre ym. (2014) ovat ehdottaneet MISO-algoritmia (MultiStore Online tuning algorithm), joka hyödyntää Hadoopin varastointijärjestelmää ulkoiselle massadatalle (*Hadoop Distributed File System*, HDFS) ja yrityksen sisäisesti kerryttämää tietoa, joka varastoidaan relaatiotietokantaan (*relational database management system*, RDBMS), hyödyntämällä näihin kohdistettujen kyselyiden sivutuotteita, jotka eivät pakota massadata-pohjaisen datan tuontia relaatiotietokantaan (LeFevre ym., 2014). Täten MISO pyrkii ratkaisemaan volyymin ja sen käsittelyyn vaadittavan prosessin ajankeston ongelmien ratkaisuja. Kowsar, Dashnov ja Lonie (2014) sen sijaan ehdottavat *datan lomitusta* (eng. *Data Interlocking*) ratkaisuksi massadatan käsittelemiseen. Datan lomituksessa data ja sen prosessointityökalut liitetään yhteen. Heidän ehdotelmansa koostuu neljästä osasta, jotka muodostavat massadatan käsittelykokonai-

suuden: Työpöytä (eng. *Workbench*), referenssidata (eng. *Reference Data*), objektiivarastointi (eng. *Object Storage*) ja datan lomitus. Työpöytä toimii datan prosessointityökaluna, joka hoitaa datan käsittelyn. Referenssidata on suuren toimijan tai yhteisön hyväksymä ja tuottama vertailukohde rikastettavalle datalle, joka on kriittinen datan analysoinnille asettamalla sille vertailupisteen. Objektivarastointi mahdollistaa suurien datamäärien varastoinnin ryhmiteltävinä objekteina, vastaten siten massadatalle ominaisiin *volyymin* ja *varieteetin* ongelmiin. Datan lomitus, joka on prosessin viimeinen osa ja koko prosessille nimen antava tekijä, tarkoittaa keinoja yhdistää data ja sen hallinnoinnin työkalut mahdollistamalla datan varastoinnin, muuntamisen ja tarkastelun saman palvelun käyttäjien kesken (Kowsar ym., 2014).

Pilvipalvelut tarjoavat ratkaisuja massadatan massan eli määrän kattamiseen. Pilvipalvelujen hyödyntäminen datan määrän aiheuttavan ongelman katkaisemisessa vaatii, että pilvipalvelun käyttäjällä on mahdollisuus välittömään sinne varastoidun tiedon hyödyntämiseen samalla, kun pilvipalvelun haltija vastaa sinne varastoitavan tiedon hallinnoinnista. Pilvipalvelun hyödyntäminen mahdollistaa sen, ettei massadataa hyödyntävä yritys turhaan liikuta suuria määriä dataa omilla palvelemillaan siitä hyödynnettävän datan rikastamiseksi, vaan käytettäväksi haluttava osuus voidaan noutaa pilvipalveluun varastoidusta datasta, sillä pilvipalvelujen rakenne mahdollistaa datan varastoimisen heti sen luomisen yhteydessä, riippumatta sen koosta, samalla kun se hyödyntää tarvittavia analyttisiä työkaluja datan tarkasteluun, vastaten siten massadatan tekniseen käsittelyyn liittyviin ongelmiin. Tällöin pilvi olisi toimijan lunastaman, ulkoinen palvelu, josta toimija maksaa vuokrapalveluna ulkoiselle palveluntarjoajalle. (Kowsar ym., 2014.) Pilvipalveluista on kolme erilaista palveluratkaisua: IaaS (*Infrastructure as a service*) kattaa serverit, verkostot ja varastointi, PaaS (*Platform as a service*) tarjoaa korkeamman suoritustehon kustomoituja sovelluksia ja SaaS (*Software as a Service*), joka on yksi tunnetuimmista pilvipalvelumalleista, jossa sovellukset toimivat suoraan pilvessä. Aiemmin esitelty BigQuery on SaaS-palvelu, jossa datan prosessointi ja jalostaminen tapahtuvat suoraan pilvipalvelimilla (Fernandes & Bernardino, 2015).

Vaikka massadata mahdollistaa datan käyttämisen laajemmalla alueella, kuin aikaisemmin on ollut mahdollista, on olemassa tilanteita, jossa analytiikka ei ole oikea ratkaisu päätöksentekemisen edessä. Davenport ym. (2010) ovat esittäneet viisi tilannetta, jossa analytiikka ja sen hyödyntäminen eivät toimi päätöksenteossa:

1. *Kun analyysin tekoon ei ole aikaa.*
2. *Kun asiasta ei ole aiempaa tietoa.*
3. *Kun historia on harhaanjohtavaa.*
4. *Kun päätöksentekijällä on huomattavaa kokemusta.*
5. *Kun muuttujat eivät ole mitattavissa.*

Nämä ovat poikkeustilanteita, joissa analytiikalla on heikompi asema päätöksenteon taustatekijänä. Ajanpuutteessa analytiikkaan liika luottaminen ja

analyysin suorittaminen väkisin kiireisen aikataulun alla voi tuottaa vääristyneitä tuloksia, jotka vain heikentävät tehtyjen päätöksiä kannattavuutta. Uuden suorittamisessa, josta ei ole aiempaa tutkittua tietoa ja kerättyä dataa, lähimpiin rinnastettaviin datatietoihin ja analyysiin luottaminen voi olla virhe, joka jälleen tuottaa vääristyneitä tietoja ja sen seurauksena vääristyneitä lopputuloksia. Harhaanjohtavalla historialla tarkoitetaan ennakkotapauksia, joiden muuttujat eivät ole täysin tiedossa ja mihin pohjautuvaan analyysiin ei voida luottaa sen puuttuvien tekijöiden vuoksi. Päätöksentekijän omatessa huomattavaa kokemusta tietystä päätöksentekoprosessista, eli kerättyään ja analysoituaan useasti samaa päätöksentekoa varten dataa, tämän prosessin uusiminen voi olla turhaa ja resurssien hukkaamista. Viimeinen poikkeustapaus, mittaamattomat muuttujat, tarkoittavat tilanteita jossa avainmuuttujat eivät ole luotettavasti mitattavista ja muutettavissa analysoitaviksi tekijöiksi, jolloin niiden muuttaminen väkisin analysoitaviksi formaateiksi voi johtaa virheelliseen tietoon. (Davenport ym., 2010.) Tämän lisäksi Deloitte (2012) on laatinut listauksen viidestä tavallisimmin analytiikan hyödyntämiseen liiketoimintatiedon analytiikassa liittyvistä virheistä, jotka Davenportista poiketen eivät johdu analytiikan soveltumattomuudesta poikkeustilanteissa hyödynnettäväksi, vaan analyysiin liittyvien odotuksien poikkeamiseen sen todellisesta luonteesta. Näitä ovat:

1. *Analytiikan väärinymmärtäminen*, analytiikan kokeminen pelkkänä ennustustyökaluna, jota voidaan toteuttaa jonkin valmiin ohjelmistoratkaisun kautta ymmärtämättä sen enempää datasta, johon tulevaisuuden ennustaminen pohjustetaan sekä epäily analyysin tehokkuudesta ja sen vuoksi sen kokonaisvaltaisen hyödyntämisen välttäminen ja kattamattomien tulosten saaminen käytöstä.
2. *Huoli datan laadusta*, yrityksen liiallinen keskittyminen datan eheyteen ja sen vuoksi liiallisten resurssien käyttäminen erillisten tietokantojen rakentamiseen ennen datan hyödyntämistä.
3. *Täydellisten statistiikkamallien metsästyminen*, eli datasta päätöksenteossa hyödynnettävän tiedon rikastamisen sijaan vastuussa oleva data-analytiikko pyrkii saavuttamaan mallinnoksissa statistiikalle ominaisia, matemaattisia absoluuttisia totuuksia jäljitteleviä malleja sen sijaan että käsitellystä datasta saataisiin tehokkaasti ja nopeasti hyödynnettäviä tietoja ulos.
4. *Yliluottavaiset data-analyysin suorittajat*, jotka luottavat liian sokeasti saamiinsa tuloksiin ilman tarvittavaa kriittisyyttä kaiken saatavan tiedon oleellisuudesta ja pätevyydestä eri päätöksentekoprosesseissa ja
5. *Kommunikoinnin puute data-asiantuntijoiden ja päätöksentekijöiden välillä.*

Liiketoimintatiedon hallintaan ja analytiikkaan massadatassa on tehnyt merkittävästi muutosta liiketoimintadatan keräämisen muuttunut luonne: perinteisesti organisaatiot ovat keränneet dataa vain liiketoimintaprosesseista, kuten liikevaihdosta, sisäisistä operaatioista (kuten markkinointi, myynti, rahoitus, tuotanto ja henkilöstöresurssit), mutta massadata ja kasvavan datan määrä on mahdollistanut organisaatioiden kerätä rakenteetonta dataa myös ulkoisesti.

Tämä muutos datan keräämisessä on mahdollistanut datasta oppimisen arvon kasvamisen liiketoiminnan päätöksenteossa (Ayankoya ym., 2014). Massadataa suositellaan otettavaksi osaksi perinteisiä päätöksentekoprosesseja, jotta se laajentaisi jo olemassa olevia tietolähteitä liiketoimintaympäristössä tapahtuvien muutosten seuraamiseksi ja yrityksen pysymiseksi ajan tasalla ja reagointivalmiudessa alan markkinoilla tapahtuviin muutoksiin (Constantiou & Kallinikos, 2015). Jo aiemmin mainittu kasvaneiden tiedonkeruukanavien määrä on mahdollistanut paljon liiketoimintatiedon hallinnassa ja analytiikassa – esimerkiksi alan ja asiakkaiden uutisia ja trendejä voi seurata analysoimalla sivustojen RSS-syötteitä, twiittejä ja yleisiä uutisvirtoja (Kowsar ym., 2014). Konkreettisimmin massadatan aiheuttamaa muutosta perinteiseen liiketoimintatiedon hallinnan ja analytiikkaan kuvaa Gartnerin (2012) ennustus alan vaatimista uusista osaajista – hänen mukaansa massadata vaatii globaalisti noin 4,4 miljoonaa uutta työntekijää, joista vain kolmasosan odotettiin täyttyvän. Lisäksi Wixom, Ariyachandra, Douglas, Goul, Gupta, Iyer, Kulkarni, Mooney, Phillips-Wren & Turetken (2014) painottavat teoksessaan akateemisten organisaatioiden tärkeyttä sisällyttää massadata osaksi liiketoimintatiedon hallinnan ja analytiikan koulutusohjelmia, joiden määrää myös yleisesti tulisi koulutuksessa lisätä muutoksen vaatimien uusien osaajien kouluttamiseksi työelämään (Wixom ym., 2014). Accenturen (2013) tekemän tutkimuksen mukaan kolmasosa yrityksistä on siirtynyt käyttämään analytiikka koko organisaationsa laajuudella uusien etuisuuksien saavuttamiseksi ja kaksi kolmasosaa yrityksistä on nimittänyt data-analytiikan tehtäviin erillisen päällikön johtamaan yrityksen toimintojen suuntautumista data-keskeisemmiksi yhden ja puolen vuoden sisällä aiheen noustua ajankohtaiseksi puheenaiheeksi vuonna 2011 (Accenture, 2013).

Massadatan mahdollistama ja toisaalta myös aiheuttama muutos liiketoimintatiedon hallinnassa ja analytiikassa on siis johtanut useisiin muutoksiin liiketoiminnassa. Se on vaikuttanut esimerkiksi useampien datalähteiden hyödyntämiseen yrityksen strategisessa päätöksenteossa ja muuttuvan toimintaympäristön tarkasteluun nopeasti kertyvän sekä useasta eri lähteestä kerätyn, liiketoimintaympäristön eri muuttujia kuvaavan monimuotoisen datan avulla. Nämä muutokset ovat tarjonneet mahdollisuuksia liiketoiminnalle sen eri osa-alueiden tapahtumien ennustamiseen, menneiden tapahtumien hyödyntämiseen ennustamisessa ja alalla tapahtuvien trendien havaitsemiseen mm. rakenteettomasta tekstipohjaisesta datasta, joita kertyy sosiaalisen median kanavista. Rakenteettonta dataa voidaan hyödyntää kuvaamaan ihmisten keskuudessa vallitsevia trendejä eli markkinoiden ihmisaspektia. Massadata on luonut tarvetta uusien osaajien kouluttamiseksi liiketoimintatiedon hallinnan ja analytiikan saralla (Wixom ym., 2014) sekä erillisten toimenkuvien luomiseksi massadatan aspektien hyödyntämiselle perinteisessä toiminnassa (Accenture, 2013). Massadatan hyödyntämisessä on tärkeää tunnistaa organisaation omien tavoitteiden kannalta oleellinen tieto ja luoda datalinjauksia siitä, millaista dataa rikastamalla ja hyödyntämällä yritys voi saavuttaa toivomiaan edistystä liiketoiminnan saralla. Massadata itsessään ei ole kuitenkaan aina vain onneen ja organisaatioiden keskittäessä toimintojaan massadatan luomien mahdollisuuksien ympärille heidän

on tärkeää muistaa poikkeustilanteet, jossa massadata analytiikka ei tarjoa luotettavia ratkaisuja ja joissa ihmislähtöisen ajattelun ja arvioinnin käyttäminen voi luoda oleellisempia lopputuloksia. Tutkimukset ovat kuitenkin lähes yksiselitteisesti sitä mieltä, että massadatan hyödyntäminen yrityksen liiketoiminnassa ja liiketoimintatiedon hallinnassa ja analytiikassa tuottaa niille merkittäviä lisäarvoja syvällisemmän ymmärryksen mahdollistajana. Tutkimuksen toimeksiantajayrityksen kaltaisissa tiettyä tuotepalvelua tarjoavissa yrityksissä onkin tärkeää hyödyntää massadataa ja liiketoimintatiedon hallintaa ja analytiikkaa kattavamman asiakasymmärryksen saavuttamiseksi, jotta he voivat kilpailla asiakkaille tuottamallaan kohdistetuilla palveluilla ja tuotteilla (Sodenkamp, Kozlovskiy & Staake, 2015). Aihetta käsitellään laajemmin seuraavassa kappaleessa ”Asiakasanalytiikka”.

2.4 Yhteenveto

Maailmamme pienentyessä kehittyvän teknologian tuodessa perinteisiä rajoituksia elämän eri toiminnoissa maailmanlaajuisesti alas tuotettavan datan määrä kasvaa. *Datafikaation* eli suurimman osaa jokapäiväisen elämämme toimintoja muuttuessa teknologia keskeiseksi tuotamme enemmän dataa ympäri vuorokauden omilla teknologisilla laitteillamme, eikä datan tuottaminen ole enää rajoittunut vain ihmisiin ja ihmisten välisiin vuorovaikutuksiin – koneiden yleistyessä ja kehittyessä myös koneet viestivät keskenään ja sensorit, jotka ovat laajentuneet suureksi osaksi maailmaa ja erilaisia teknologisia sovelluksia, tuottavat myös jatkuvasti dataa, kommunikoiden tuotettavasta datasta eri vastaanottajakoneiden kanssa. Datafikaation ansiosta jatkuvasti kasvavan datan määrällä on mahdollista saada merkittäviä tuloksia tutkimuksessa ja liiketoiminnassa, jos dataa osataan hyödyntää tehokkaasti. Massadata, eli määrältään, kasvunopeudeltaan ja monimuotoisuudeltaan rikas data on pohja mitä erilaisimmille käyttömahdollisuuksille ja massadata -työkalujen ja metodien kehittyessä sen käyttömahdollisuudet laajenevat. Tällä hetkellä suurin ongelma massadatan hyödyntämisessä on tietokoneiden prosessointitehon heikkous sekä ihmisten asenteet datan hyödyntämisen mahdollisuuksia kohtaan – kun ihminen, joka luo pohjan tiedonlouhinnan ja eri tiedonkäsittelykomentojen kautta eri työkaluille kerätä, yhdistellä ja rikastaa dataa sen yltiöpäisestä merestä ymmärtää, mitä dataa keräämällä saadaan oleellisempia tuloksia ulos, ovat organisaatiot jo askeleen lähempänä massadatalle esiteltyjen hyötyjen toteutumista. Tämän mahdollistamiseen massadata ja sen analytiikan käsittely sekä siitä tulevien ratkaisujen huomioiminen myös liiketoiminnassa käytännön tasolla täytyy tuoda osaksi koulutuksia – maailma tarvitsee uusia osaajia sekä massadataan liittyvien teknologisten ratkaisujen toteuttamisessa, että sen ymmärryksessä ja välittämisessä osaksi kaikkia organisaation prosesseja. Massadata toimii tällä hetkellä kilpailuvälineenä eri organisaatioiden välillä, mutta siitä seuraavat tekniset ratkaisut analytiikassa ja tiedonkäsittelyssä saattavat olla osa kaikkien organisaatioiden tulevaisuuden tiedonkäsittelyä ja data-analytiikkaa.

3 ASIAKASANALYTIikka

Kappaleessa käsitellään asiakasanalytiikka, sen määritelmää, sitä käsittelevissä tutkimuksissa esitettyjä hyödyntämismahdollisuuksia sekä asiakasanalytiikan ongelmakohtia. Kappale esittää yhtymäkohtia asiakasanalytiikan ja asiakassegmentoinnin välillä, joka selkeyttää aiemmin esitettyjä massadata analytiikan mahdollisia käyttömahdollisuuksia asiakassegmentoinnin osalla asiakassegmentoinnin toteuttamisessa usein hyödynnettävien analytiikka tekniikoiden kautta. Lopuksi kappale käsittelee yleisiä ongelmia asiakassegmentoinnissa sekä erilaisia mahdollisia keinoja ja viitekehyksiä asiakassegmentoinnin toteuttamiseen sekä kirjallisuuden asettamia vaatimuksia onnistuneelle asiakassegmentoinnille. Kappaleen lopuksi siinä käsitellyt asiat tiivistetään ja kirjallisuudesta esille tulleiden tehokkaiden ratkaisujen sekä olennaisen datan hyödyntämiseen liittyvien ongelmien pohjalta luodaan viitekehys asiakasanalytiikassa hyödynnettävän datan käsittelemiseksi.

3.1 Asiakasanalytiikan määritelmä

Asiakasanalytiikka (eng. *customer analytics*) on kilpailuedun luomisen keino liike-maailmassa, jossa kehittyvä teknologia vähentää eroja eri palveluntarjoajien tuotteiden ja palveluiden välillä (Kolyshkina & Simoff, 2007). Analytiikka on noussut suosituksi keskustelunaiheeksi liiketoimintamaailmassa sen mahdollistaman kilpailuedun vuoksi ja suuret konsultointiyritykset, kuten Accenture, Deloitte ja IBM ovat kaikki alkaneet harjoittaa toimintaa analytiikan ympärille (Frans, 2012, s. 14). Sodenkamp ym. (2015) esimerkiksi esittävät teoksessaan massadatan mahdollisuuksista energiateollisuudelle lisäarvon tuottamiseksi käsittelee massadata-pohjaisen asiakasanalytiikan hyödyntämistä visualisoituna tuloksina asiakaille heidän kulutuskäyttäytymisestään (Sodenkamp ym., 2015). Analytiikka voi vastata useisiin liiketoiminnan harjoittamiseen liittyviin kysymyksiin, eikä se rajoitu vain tämän hetkisen tiedon analyysiin – Davenport ym. mallin mukaan analytiikalla on potentiaalia vastata kysymyksiin menneistä tapahtumista aina tulevien ilmiöiden ennustamiseen. Kysymykset, joihin analytiikka vastaa, voidaan jakaa kahteen luokkaan: *informaatioon* ja *oivallukseen*, jotka molemmat jakaantuvat vielä menneeseen, nykyiseen ja tulevaan. *Mennyt informaatio*, mitä tapahtui, toteutuu raportoinnin kautta. *Nykyinen informaatio*, mitä tapahtuu tällä hetkellä, on vastattavissa ilmoitusten kautta. *Tulevaisuuden informaatio*, Mitä tulee tapahtumaan, on tutkittavissa ekstrapoloinnin, eli aiempien ilmiöiden luoman kaavan jatkamisella tapahtuvan ”ennustamisen” kautta. Sen sijaan oivalluksia, näkemyksiä, voidaan saada menneestä, eli miten ja miksi jotain tapahtui, mallintamisen ja kokeellisen suunnittelun avulla. *Oivalluksia nykyisyydestä*, eli mitä tulisi tehdä seuraavaksi, voi ratkaista suositusten avulla. *Oivalluksia tulevaisuudesta*,

eli mitä parhaimmassa/pahimmassa tapauksessa voi tapahtua, voi selventää ennustusten, optimisoinnin ja simulaatioiden avulla. (Davenport ym., 2010, s. 7.) Davenportin esittämiä analytiikan mahdollisuuksia informaation ja oivalluksen saralla voi myös hyödyntää asiakasanalytiikassa asiakkaan paremmaksi ymmärtämiseksi ja tiedon keräämiseksi, samoin kuin saatavilla olevan datan pohjalta tehtävien oivalluksien avulla voidaan ymmärtää esim. poistuneiden asiakkaiden motiiveja toiselle palveluntarjoajalle siirtymiseksi. Yksi mahdollisista työkaluista tällä saralla on tekstinlouhinta asiakaspuheluista, joiden avulla yritys voi tutkia, millaisia asiakkuudesta poistumisen merkkejä asiakas on osoittanut puhelujen aikana ja reagoida vastaavaisuudessa samankaltaisissa tilanteissa vielä asiakkuuteen kuuluvien kuluttajien sitoutumisessa takaisin palveluun (Kolyshkina & Simoff, 2007).

Sodenkamp ym. (2015) tutkivat artikkelissaan erityisesti asiakasanalytiikan tuomaa lisäarvoa energia-alan yritykselle. He mainitsevat teoksessaan yritykselle tärkeäksi muuttua pelkästä palveluntarjoajasta kokonaisvaltaisemmaksi palveluyritykseksi, joka luo lisäarvoa asiakkailleen hyödyntäen asiakasanalytiikka syvällisempään asiakasymmärryksen saavuttamiseksi. Erityisen tärkeässä roolissa tässä on yrityksen kerryttämä asiakkaiden virran kulutustiedot, joista yritykset jo valmiiksi kerryttävät dataa ja rikastaa tämä data data-analytiikan työkalujen ja koneoppimisen avulla. He ehdottavat teoksessaan data-analyysipohjaisten informaatiojärjestelmien hyödyntämistä yksityisille talouksille energiatehokkaiden käyttäytymismallien simuloinnissa. Yritys tarjoaa tällä tavoin asiakkaalle lisäarvoa mahdollistamalla näille syvällisemmän ymmärryksen oman energiankulutuskäyttäytymisensä vaikutuksesta kertyviin sähkölaskuihin samalla, kun yritys itsessään saavuttaa lisääntyntä asiakasuskollisuutta (loikkareiden välttäminen) ja saa mahdollisuuden toteuttaa lisämyyntiä (eng. *upselling*), eli ehdottaa asiakkaille, jotka eivät ole varmoja nykyisen ratkaisunsa sopivuudesta, kattavampia palvelupaketteja, jotka ovat kalliimpia, mutta heidän tarpeisiinsa sopivampia. (Sodenkamp ym., 2015.)

Asiakasanalytiikka on usein ensimmäisiä osa-alueita, josta analytiikan hyödyntäminen liiketoiminnassa aloitetaan, koska asiakkaat ovat se tekijä, jotka tuovat rahaa yrityksille. Asiakasanalytiikka voidaan hyödyntää usealla osa-alueella – asiakassegmentointiin ja tuottavimpien asiakkaiden tunnistamiseen, asiakaskäyttäytymisen ymmärtämiseen, asiakkaiden halujen ja tarpeiden ymmärtämiseen sekä näihin vastaavan tuotesuunnittelun ja mainostamisen suunnitteluun. Asiakassuhdehallinta (eng. *Customer relationship management, CRM*) on erityisen tärkeässä roolissa asiakkaiden ymmärtämiseksi ja heille lisäarvon luomiseksi mukautettuina tarjouksina, tuotteina ja yleisinä palveluratkaisuina, jonka toteuttamiseen vaaditaan asiakasprofiilien toteuttamista asiakasanalytiikan muodossa (Sicilia & García, 2003). Asiakasanalytiikka mahdollistaa kaikista tehokkaimman markkinoinnin suunnittelun, tuotteiden/palveluiden hintojen optimisoinnin parhaaseen tuottavuuteen sekä tunnistaa palveluntarjoajan vaihtamista harkitsevat asiakkaat ja aloittaa toimet näiden poistumisen välttämiseksi.

Asiakasanalytiikka toimii siis lisäarvon tuottamisen välineenä niin yritykselle itselleen asiakkaan syvällisemmäksi ymmärtämiseksi niin saatavilla olevan

datan rikastamisessa ilmiöitä selittäväksi ja ennustavaksi tiedoksi, kuin asiakkaiden asiakaspalvelukokemusta lisääväksi tiedoksi asiakastiedon visualisoinnin ja asiakassuhdehallinnoinnissa hyödynnettävän tiedon rikastamiseksi. Asiakasanalytiikan mahdollistava asiakkaan syvempää ymmärrystä voi hyödyntää myös asiakassegmentoinnissa, mitä käsitellään kattavammin seuraavassa kappaleessa.

3.2 Asiakassegmentointi ja asiakasanalytiikka

Asiakassegmentoinnin konseptin kehitti 1950-luvun puolella välissä Wendell R. Smith, amerikkalainen markkinoinnin ammattilainen (Wu & Lin, 2005). Asiakassegmentointi on oleellinen osa *markkinointia*, eli sosiaalisen ja johtamisen prosessia, jossa yksilöt ja ryhmät saavuttavat mitä he tarvitsevat ja haluavat luomalla ja vaihtamalla tuotteita ja arvoja toistensa kanssa (Kotler, Armstrong, Wong & Saunders, 2008). Asiakassegmentoinnin avulla tunnistetaan ja arvotetaan eri asiakasryhmiä, joille tulee tuottaa erilaisia, heidän tarpeisiinsa ja haluihinsa soveltuvia ratkaisuja. Segmentointi auttaa tunnistamaan potentiaalisia asiakasjoukkoja ja luokkia sekä erottamaan ryhmät, joiden palveleminen ei ole yrityksen kannalta tuotteliasta. Asiakassegmentointi on tärkeää eri segmenteille soveltuvien palvelujen luomiseksi sekä segmentoinnin mahdollistavan tuotekohdistuksen ja -sijoittelun vuoksi (Lynn, 2011). Yksi markkinoinnin merkkiteoksista, jota hyödynnetään paljon myös aiheen opiskelussa, Kotlerin *Principles of Marketing* -kirja nimeääkin segmentoinnin yhdeksi asiakaslähtöisen markkinoinnin avaintekijöistä (Kotler ym., 2008, s. 408). Lynnin määritelmä asiakassegmentoinnille on Kotlerin määritelmään verrattuna vain pintaraapaisu, jossa jätetään avaamatta segmentoinnin tärkeys käytännön tasolla – segmentoinnissa ei vain tunnisteta ja arvoteta potentiaalisia asiakasryhmiä, vaan sen avulla voidaan tunnistaa myös jo palveltujen asiakasryhmien lisäksi niitä ryhmiä, joita ei vielä palvella nykyisellä tarjonnalla sekä asiakasryhmiä, joilla voisi olla nykyiselle tarjonnalle kysyntää, mutta jolle ei vielä ole kohdennettu markkinointitoimia. Segmentointi auttaa myös arvioimaan, kuinka onnistuneita nykyiset ratkaisut eri ryhmien kohdalla ovat sekä onko eri ryhmille toteutettu erilaisia markkinointiratkaisuja, kuten erilaistettuja mainoskampanjoita ja tarjouksia jotka vastaavat juuri heidän motiiveihinsa palvelujen ostamiselle (Kotler ym., 2008, s. 410). Vaikka asiakassegmentoinnin toteuttamiseen on olemassa useita käytänteitä palveluntarjoajan liiketoiminnan luonteesta, käytettävissä olevasta teknologiasta ja tavoista riippuen, kaikkien luotujen asiakassegmenttien tulisi noudattaa seuraavia vaatimuksia olakseen toimiva ja hyödynnettävissä asiakasanalytiikassa:

1. Segmentin jäsenet ovat homogeenisiä, eli yksilöt saman segmentin sisällä jakavat samat tarpeet ja halut, joihin voi vastata yhdellä, heille räätälöidyllä ja kohdistetulla markkinointiratkaisulla.
2. Segmentti on helposti erotettavissa muista segmenteistä, jotta varmistetaan eri segmenttien tarpeiden erilaisuus.

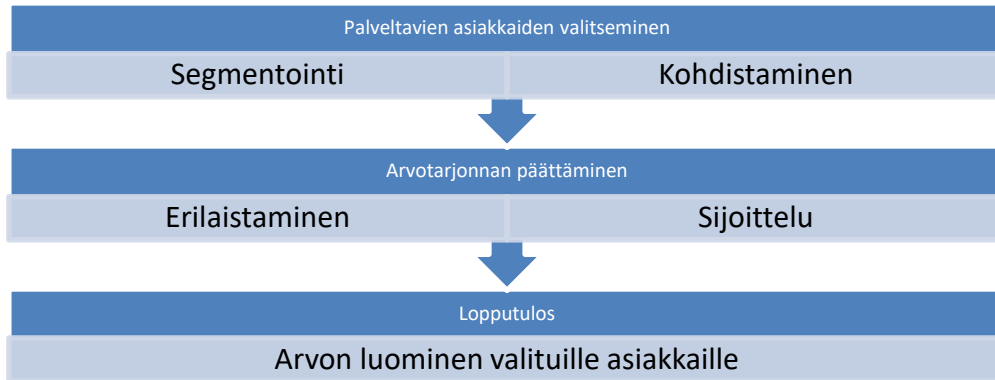
3. Segmenttiä voidaan käyttää perustana räätälöityyn kohdentamisstrategiaan.

Eli segmenttien vaatimukset ovat tiivistettävissä niiltä vaadittavaan homogeenisyyteen, tunnistettavuuteen ja tulkittavuuteen. (Haider, Chiarandini, & Brefeld, 2012.)

Asiakassegmentointi ja sen hyödyntäminen markkinoinnin suunnittelussa ja ratkaisussa on erityisen tärkeää, koska palvelua tarjoavat yritykset eivät voi vedota tarjonnallaan kaikkiin ostajiin tai ainakaan kaikkiin potentiaalisiin ostajiin samalla tavalla. (Kotler ym., 2008, s. 410). Jokainen asiakas eroaa toisistaan. Ostajien halut, resurssit, sijainti, ostotarpeet, ostokäyttäytyminen ja ostotarpeet voivat kaikki erota muista ostajista, tehden jokaisesta palveltavasta ostajasta oman yksilönsä massojen joukossa. Vaikka datan määrän kasvaminen ja erilaiset data-lähtöiset ratkaisumahdollisuudet markkinoinnin suorassa kohdentamisessa ovat kasvaneet, on yksittäisten asiakkaiden henkilökohtaisiin tarpeisiin vastaaminen suurelle asiakasryhmälle palveluja tarjottaessa edelleen haastavaa ja hyöty-vaste arvoltaan vielä sen verran tehotonta, että sitä yksistään voitaisiin hyödyntää koko asiakaskannan tarpeisiin vastattaessa. Tämän vuoksi segmentointi ja sen onnistunut käyttö markkinointiratkaisujen suunnittelussa on erityisen tärkeää. Kun asiakassegmentointi vastaa sille aiemmin esitettyihin vaatimuksiin eli sen avulla potentiaaliseen ostajaryhmään kuuluvista yksilöistä pystytään kokoamaan homogeenisiä, samankaltaisista asiakkaista koostuvia ryhmiä, joiden henkilökohtaisiin tarpeisiin ja motiiveihin pystytään vastaamaan yhdellä markkinointiratkaisulla, palveluntarjoaja pystyy toteuttamaan markkinoinnin ratkaisuja jotka kattavat tehokkaasti sen koko potentiaalista ostajakantaa.

Itse segmentointi asiakassegmentoinnin kokonaisprosessin toimenpiteissä onkin vain yksi osa kokonaisuutta. Segmentoinnin ympärille perinteisesti sijoittuvat toiminnot ovat jaoteltavissa kahteen yläluokkaan, *Palveltavien asiakkaiden valitsemiseen ja Arvotarjonnan päättämiseen*. Ensimmäisen luokan alaisuuteen kuuluu itse segmentointi, eli kokonaismarkkinan jakaminen pienempiin osioihin sekä *kohdentaminen* (eng. *targeting*), eli päätös siitä, mille segmentille/segmenteille palveluja ja tarjontaa lähdetään tarjoamaan. Arvotarjonnan päättämisen toimintoja sen sijaan ovat *erilaistaminen* (eng. *differentiating*), eli ylivoimaisen arvon tarjoamista asiakkailleen saatavilla olevia palveluja erilaistamalla sekä *sijoittelu* (eng. *positioning*), eli tuotteen järjestäminen selkeälle, erottuvalle ja kilpailukykyiselle paikalle kilpailevien tuotteiden rinnalle toivottujen asiakkaiden mielissä. Kaikilla näillä toiminnoilla pyritään edistämään markkinoinnin yhtä perusvaatimusta, eli *arvon luomista valituille asiakkaille*. (Kotler ym., 2008, s. 410).

KUVIO 1 - Arvon luonnin prosessi segmentoinnissa (mukaillen Kotler ym., 2008).



Jo palvelemissaan asiakassegmentit tunnistettuaan ja eriteltyään organisaatioilla on mahdollisuus tunnistaa asiakassegmenttejä, jotka eivät kuulu vielä sen kattamaan asiakaskantaan ja uudelleenarvioida, mille ryhmille markkinointia kannattaa kohdentaa, onko olemassa olevan tuotteen ominaisuuksissa jotain, joka vastaa asiakassegmentoinnin tarpeisiin, jota ei aiemmin ole pidetty kohde-ryhmänä tuotteelle ja kuinka mittavilla toiminnoilla tämän segmentin voisi ottaa mukaan kohdemarkkinoihin.

Analytiikan työkalut ovat tärkeitä sekä asiakassegmentoinnin toteuttamisessa, että segmenttien tietojen jälkijalostuksessa päätöksenteossa hyödynnettäväksi tiedoksi. Erityisen tärkeä osa tätä on *tiedonlouhinta* (eng. *data mining*), jota voidaan hyödyntää usealla eri tavalla tärkeiden asiakastietojen keräämiseen ja tulkintaan. Tiedonlouhinnasta saatuja tietoja hyödynnetään esimerkiksi asiakassuhdehallinnassa sen tuottojen maksimointiin ja kulujen vähentämiseen. (Vadivu & David, 2012). Yksi tärkeä ja paljon käsitelty osa-alue tiedonlouhinnan markkinoinnissa hyödynnettävistä tekniikoista on *klusterointi* (eng. *clustering*), eli aineistojen ryhmittelyä niiden samankaltaisuuden perusteella heterogeenisistä aineistoista homogeenisemmiksi. Klusteroinnin luonteen vuoksi sitä hyödynnetään usein asiakassegmenttien määrittämiseen (Singh, Rumantir, South, Bethwaite, 2014). Tiedonlouhinnan tuloksien jälkikäsitteilyllä, kuten visualisoinnilla ja tulosten kiinnostavuusperäisellä luokittelulla asiakastiedoista voidaan saada helposti tulkittavia tuloksia päätöksenteon tueksi. Yksi tärkeistä työkaluista asiakaskannan ymmärtämiseksi ja oman tarjontansa parantamiseksi asiakasanalytiikan avulla on tutkia, millaiset asiakkaat, eli mihin segmenttiin lukeutuvat entiset käyttäjät ovat vaihtaneet palveluntarjoajaa kilpailijoiden tarjontaan. Tätä kutsutaan *asiakkaiden hiertymiseksi/kulumaksi* (eng. *churning/attrition*). Selvittämällä, mitä mahdollisia motivaatioita tiettyyn segmenttiin kuuluvilla asiakkailla tai tiettyjä ominaisuuksia jakavilla asiakkailla on ollut vaihtaa palvelua, palveluntarjoaja pystyy tekemään päätöksiä tätä estävien toimintojen kehittämiseksi (Vadivu & David, 2012).

Datalähtöisellä markkinoinnilla (eng. *data-driven marketing*) tarkoitetaan markkinoinnin toimia, jotka pohjautuvat tiedonlouhintaan (Mulvenna, Norwood & Büchner, 1998) ja jonka avulla yritys pystyy kohdistamaan tarjouksiaan ja uutiskirjeitään asiakkailleen. Vaikka termi aiemmin käsitti suurimmassa osin vain online-tilassa tapahtuvan markkinoinnin ja verkkomyynnin toimet, sen merkitys

on nykyisin datan määrän kasvaessa ja liiketoiminnan digitalisoituessa laajentunut myös käyttöyhteyksiin, kuten asiakasstrategian luontiin datan avulla (Tanner, 2014). Vadivu & David (2012) käsittelevät asiakassuhdehallintaa ja kohdennettua markkinointia käsittelevässä teoksessaan myös datalähtöisen markkinoinnin käytön helpottumista ja sen tärkeyden kasvua, mutta käyttävät siitä nimeä *tietokantamarkkinointi* (eng. *database marketing*) (Vadivu & David, 2012). Asiakastietojen louhinta toimii tärkeässä roolissa asiakastietojen rikastamiseksi hyödynnettäviin muotoihin niin asiakasanalytiikan kohdalla, kuin asiakassegmentointia toteuttaessa. Kuitenkin myös siihen pätee sama analytiikkapohjainen ongelma, kuin aiemmassa kappaleessa esitetystä massadata -analytiikassa – vain ihmis pohjaisen päätöksenteon pohjalta tehtyjen oikeita, analyysille olennaista dataa noutavien algoritmien käyttö tuo hyötyjä asiakasanalytiikkaan ja sen avulla asiakassuhdehallintaan ja palveltujen sekä vielä palvelemattomien asiakassegmenttien ymmärtämiseen (Peng & Yuanyuan, 2008). Asiakkaiden ymmärtäminen onkin tärkeimmässä roolissa asiakassegmentoinnissa erityisesti palveluntarjoajayrityksillä ja toteutettavaksi eri keinoja ymmärtää ja jaotella asiakkaitaan paremmin.

Asiakasanalytiikka toimii siis työkaluna asiakassegmentoinnin mahdollistamiseksi mm. asiakkaihin liittyvän tiedonlouhinnan avulla. Asiakasryhmien syvällisempi ymmärrys mahdollistaa asiakkaiden ryhmittelyn yrityksen markkinointiratkaisujen tekemiseksi, joka on yksi asiakassegmentoinnin päätehtävistä. Tiedonlouhinnassa on kuitenkin ymmärrettävä, mitkä asiakasta kuvaavasta ja tämän kerryttämästä datasta on hyödynnettävissä asiakassegmentointiin, eli datalinjausten tekemisessä prosessien toteuttamisessa. Asiakassegmentoinnin toimivuutta, hyväksi todettuja käytäntöjä ja sen toteuttamisen viitekehyksiä samoin kuin asiakassegmentointiin liittyviä ongelmia käsitellään kattavammin seuraavassa kappaleessa.

3.3 Asiakassegmentoinnin ongelmia ja parhaita ratkaisuja

Kuten aiemmassa kappaleessa jo todettiin, asiakkaiden ymmärtäminen on tärkeässä roolissa asiakkaille sopivien ratkaisujen ja lähestymistapojen laatimiseksi eli markkinointiratkaisujen tekemiseksi. Erilasten asiakasryhmien ymmärtämiseksi ne on ensin tunnistettava segmentoinnin avulla. Sen vuoksi on tärkeää ymmärtää, mitkä ovat asiakassegmentoinnissa vallitsevat ongelmat, asiakassegmentoinnissa toteutettavat suositellut parhaat ratkaisut sekä yrityksen omaan liiketoiminnan luonteeseen sovellettavat ratkaisut, jotka huomioivat alakohtaiset tekijät, mitä kysymyksiä asiakassegmentoinnissa tulisi esittää ja mihin kysymyksiin tulisi hakea ratkaisuja yritykselle oleellisten segmenttien luomiseksi.

Asiakassegmentoinnin toteuttamisessa on omat ongelmansa, mm. oleellisten ja palveluntarjoajan omia tavoitteita ja asiakassegmenttejä todellisesti kuvaavien segmenttitekijöiden määrittely. Kuten ylläolevassa kappaleessa on esitelty, tiedonlouhinta voi hyödyntää suuresti asiakastietojen keräämisessä ja erittelemisessä. Tiedonlouhinnan kriteerien taustalla on kuitenkin lopulta aina ihminen,

joka päättää, mitkä tiedot asiakkaista on oleellisia kattavan asiakaskuvan ja -ymmärryksen luomiseksi. Tämä kuitenkin mahdollistaa, että päätetyt tekijät eivät olekaan asiakasryhmää kuvaavimpia ja tuottaa valheellisia oikeita tuloksia – ne vastaavat ihmisen tekemän algoritmien ehtoja ja palauttaa siten oikeita tuloksia, mutta ne eivät vastaa niitä arvoja ja tietoja, jotka parhaiten kuvaisivat asiakaskantaa, joita palveluntarjoaja tarvitsee todellisen asiakaskuvan ja -ymmärryksen luomiseksi. Näitä virheitä asiakassegmentoinnissa kutsutaan ”*valheellisen hyöksymiseksi ja totuuden hylkäämiseksi*” (Peng & Yuanyuan, 2008). Yksi ongelma asiakassegmentoinnissa on myös *päättämättömyys kohderyhmien valinnassa*. Palveluntarjoaja ei voi millään tarjota erilaistettuja markkinoinnin ratkaisuja kaikille potentiaalisille asiakkailleen, eikä erilaistaa tuotteitaan kaikille mahdollisille kohderyhmille, jonka vuoksi palveluntarjoajien onkin tärkeää päättää, minkä kohderyhmien palveleminen tarjonnan erilaistamisella ja kohdentamisella on palveluntarjoajalle itselleen kaikista hyödyllisintä. Kaikkien olemassa olevien potentiaalisten asiakkaiden kohdentaminen ei olisi mahdotonta pelkästään erilaistettujen markkinointiratkaisujen suunnittelun ja toteutuksen vaatiman ajan, työvoiman ja kustannuksien vuoksi, vaan myös siksi, että eri segmenttien tarpeet, halut ja arvot ovat aina jossain määrin toistensa kanssa ristiriidassa ja yhtä segmenttiä miellyttävät palveluratkaisut ja lähestyminen mainonnassa ei sovellu toiselle segmentille lainkaan ja saattaa vaikuttaa myös näiden asenteisiin palvelua kohtaan ja valmiuteen ostaa palvelua/tuotetta. Lisäksi asiakassegmentoinnin toteuttamisessa on ongelmana se, että asiakassegmentoinnin suunnittelu toteutetaan useimmiten aina palveluntarjoajan omien vaatimusten mukaan suoritettuna tutkimuksena, joiden tuloksia ja toteutustapoja ei ole julkistettu. (Lynn, 2011.)

Asiakassegmentoinnissa ei ole olemassa yhtä ainoaa absoluuttista ratkaisua. Kotler ym. (2008) kuitenkin esittävät viitekehyksen, jonka avulla segmentointia voidaan lähteä toteuttamaan. Heidän mallissaan on esitetty asiakasryhmän *maantieteellisen, demografisen, psykografisen* sekä *käytöksellisten* tekijöiden huomiointi segmentointia tehdessä. Maantieteellisiin tekijöihin kuuluvat *alueet*, kuten maanosat, maat ja maakohtaiset aluejaottelut, kuten *läänit, maakoot, kaupunkien koot, asuintiheys* ja *asuinilmastot*. Demografisiin tekijöihin kuuluvat *ikä, sukupuoli, perhekoot, perheen status*, kuten *nuori, nuori naimaton, naimisissa ilman lapsia, lapsiperheet* joiden nuorin lapsi on alle 6-vuotias, *lapsiperheet* joiden nuorin lapsi on yli 6-vuotias, *vanhat lapsettomat pariskunnat, vanhat lapselliset pariskunnat* joiden nuorin lapsi on yli 18-vuotias ja *vanha naimaton, tulotaso, ammatti, koulutus, uskonto, rotu* ja *kansallisuus*, psykografisiin *sosiaalinen luokka, elämäntyyli* kuten *saavuttaja, uskoja* ja *pyrkijä* sekä *persoonallisuus* ja *käytöksellisiin ostotilanne* (tavallinen ostos, erityistilanteen ostos), *etsitty hyöty* (laatu, palvelu, taloudellinen hyöty), *asiakkaan status* (ei asiakas, entinen asiakas, potentiaalinen asiakas, uusi asiakas, kanta-asiakas), *käyttöaste* (kevyt käyttö, keskimääräinen käyttö, runsas käyttö), *uskollisuusaste* (ei olemassa, keskivahva, vahva, absoluuttinen), *tietoisuusaste* (ei tunne tuotetta, on tietoinen tuotteesta, on saanut tietoa tuotteesta, kiinnostunut tuotteesta, haluaa ostaa tuotteen, aikoo ostaa tuotteen) ja *asenne tuotetta kohtaan* (innostunut, positiivinen, välinpitämätön, negatiivinen, vihamielinen). (Kotler ym., 2008, s. 412).

TAULUKKO 1 - Segmentoinnissa hyödynnettävät tekijät (mallintaen Kotler ym., 2008).

Geografiset tekijät	• alueet, maan koko, kaupunkien koko, asuintiheys, asuinilmasto
Demografiset tekijät	• Ikä, sukupuoli, perheen koko, siviilisääty, tulotaso, ammatti, koulutus, uskonto, rotu sekä kansallisuus
Psykografiset tekijät	• Sosiaalinen luokka, elämäntyyli ja persoonallisuus
Käytökselliset tekijät	• Ostotilanne, etsitty hyöty, asiakkaan status, käyttöaste, uskollisuusaste, tietoisuusaste, asenne tuotetta kohtaan.

Maantieteellisessä segmentoinnissa jaottelu eri segmenttien välillä voidaan viedä niin pieniin osa-alueisiin, kuin yrityksen tarjonnan kannalta on olennaista, aina maanosista pieniin kuntiin tai naapurustoihin. Erityisesti palveluntarjoajille, joiden tuotteet keskittyvät tiettyihin maantieteellisiin alueisiin, kuten lääneihin tai maakuntiin, on tärkeää, että tuotteen tarjontaa ja esilletuontia on mietitty pienempien maantieteellisten segmenttien kautta. Demografinen segmentointi on hyvä työkalu myös pienemmällä maantieteellisellä alueella palvelua tarjoavalle yritykselle, sillä lähes väistämättä jokaisella alueella on eri ikäluokkaa, sukupolvea, sukupuolta, tulo- ja koulutusluokkaa sekä perhestatusta edustavaa asiakaskuntaa. Demografiset tekijät ovatkin yksi suosituimmista työkaluista segmentointiin juuri tämän vuoksi. Vaikka ensisijainen segmentointi tehtäisiinkin muiden tekijöiden pohjalta, on tärkeää tietää eri segmenttien demografiset tekijät asiakasryhmien tunnistamiseksi, koon määrittämiseksi sekä tuotteiden tehokkaaksi kohdentamiseksi ja asiakkaiden tavoittamiseksi. Demografiset tekijät ovat erityisen tärkeitä tiedostaa asiakasryhmissä, koska asiakkaiden tarpeet ja ostomahdollisuudet vaihtelevat heidän ikänsä sekä perhestatuksen ja tuloluokan myötä. (Kotler ym., 2008, s. 411-413). Esimerkiksi opiskelijalla, joka elää pelkällä opintotuella ja lainalla, on erilaiset resurssit ja tarpeet palvelujen ostamiseksi kuin vakiintuneessa työssä olevalla perheellisellä asiakkaalla. Tämän vuoksi he tarvitsevat erilaisia lähestymistapoja ja palveluratkaisuja, jotka vastaavat juuri heidän tarpeisiinsa ja resursseihinsa. Psykografisessa segmentoinnissa keskitytään eri asiakassegmenttien sosiaaliseen statukseen, elämäntyyliin ja persoonallisuuteen ja erilaistetaan palveluja ja kohdentamista vastaamaan näihin tekijöihin. Kun näiden tekijöiden pohjalta tunnistetaan, millaisia asiakkaat ovat ja motiiveja minkä vuoksi he olisivat kiinnostuneita palveluista, *kuinka tuote sopii heidän elämäntyyliinsä ja toivomaansa elämäntyyliin*, sitä voidaan hyödyntää muun muassa tuotteen markkinoinnissa (Kotler ym., 2008, s. 415). Hyvä esimerkki tällaisesta segmentoinnista markkinoinnin toteutuksessa on tuotteiden mainostaminen joltain elämäntyyliä tai arvoa korostavana (Kotler ym., 2008, s. 417), kuten tuotteen ympäristöystävällisyyden tai reilun tuotannon osalta. Esimerkiksi vuonna 2015 suomalaisille tärkeisiin arvoihin kuului ympäristöstä huolehtiminen (T-Media, 2015). Käytöksellisiin tekijöihin pohjautuvassa segmentoinnissa otetaan huomioon ostajan tietoisuus tuotteista, asenteet niitä kohtaan, käyttötarpeet tuotteille sekä reaktiot niiden hyödyntämisestä. Erityisen tärkeitä kohtia Käytöksellisessä

segmentoinnissa ovat käyttäjän status, eli onko henkilö jo olemassa oleva asiakas sekä henkilön palveluiden käyttöaste ja uskollisuus palveluntarjoajaa kohtaan, mikä ilmenee palveluiden hankinnan keskittämisenä tietyille palveluntarjoajalle (Kotler ym., 2008, s. 417).

Segmentointia ei koskaan toteuteta yhden ainoan segmentointialueen avulla, vaan eri luokista otetaan huomioon eri tekijöitä kattavamman asiakaskuvan luomiseksi ja syvällisemmän asiakasymmärryksen saavuttamiseksi. Kohde-ryhmien syvällisemmän tunnistamisen avulla palveluiden erilaistamista ja kohdentamista voidaan lähteä konkreettisemmalla tasolla toteuttamaan verrattuna esim. pelkän segmentoinnin toteuttamiseen perhestatuksen pohjalta – kaksilapsinen perhe, jonka molemmat lapset ovat alle 18 ja asuvat kotona voi toimia ensimmäisenä askeleena asiakkaiden ryhmittelemiseksi, mutta perheetkin voivat vaihdella paljon varallisuuden ja arvojen osilta sekä ostomotivaatioiltaan palveluiden lunastamiseksi. Liian ympäröivää segmentointia tehdessä onkin vaara, ettei kokonaiskuvaa hahmota sen muodostavista yksittäisistä tekijöistä ja täysin erilaiset ostomotiivit, arvot ja resurssit omaavat asiakkaat luokitellaan yhden, heitä kuvaamattoman segmentin alle. Palveluntarjoajien onkin helpompaa menestyä tavoittelemalla eri segmenttejä kokonaisten markkinoiden sijasta (Kotler ym., 2008, s. 420). Lynn (2011) esittää omissa vaatimuksissaan ja pohjustuksissaan onnistuneelle segmentoinnille pääasiassa samat tekijät ja pääluokat, kuin Kotler, mutta Kotlerin kaltaisen tarkemman käsitteiden avaamisen sijaan Lynn keskittyy käsittelemään segmentoinnin monimuotoisuutta ja lukuisia mahdollisuuksia sekä tärkeyttä personoida organisaation oma segmentointi vastaamaan organisaation oman liiketoiminnan eri aspekteja, kuten toimialaa, markkinoita joilla toimitaan sekä asiakasryhmää. Näiden lisäksi Lynn (2011) esittää neljän kohdan vaatimukset onnistuneelle segmentoinnille:

1. Segmenttien tulee erota tavoilla, jotka mahdollistavat niiden koon ja saavutettavuuden helpon mittaamisen
2. Segmenttien tulee olla tarpeeksi suuria, jotta erilaistetut kohdennustoiminnot eri segmenttien kohdalla olisivat oikeutettuja ja kannattavia
3. Eri segmenttien tulee olla tavoitettavissa uniikkien medioiden ja markkinointikanavien kautta
4. Segmenttien tulee olla suhteellisen vakaita, eli niiden koko ei saa huomattavasti pienentyä ajan saatossa.

Lynnin (2011) onnistuneen segmentoinnin määritelmät pohjautuvat onnistuneen segmentoinnin perusvaatimuksille, eli siihen, kuinka segmentoinnin tulee olla hyödyksi organisaatiolle, segmentin sisäisten asiakkaiden tulee olla keskenään samanlaisia ja erota selkeästi muiden segmenttien asiakkaista suhteessa siihen, miten he vastaavat organisaation erilaiseen tarjontaan ja millaisia haluja ja vaateita heillä on. (Lynn, 2011). Lynnin määritelmä sisältää samat vaatimukset onnistuneelle segmentoinnille kuin Sicialian ja Garcían (2003) määritelmä, eli ryhmien *tunnistettavuuden*, tarpeeksi kattavan kokoluokan, saavutettavuuden hallin-

non markkinointiratkaisuilla, *pysyvyys* siten, ettei segmentti hajoa ennen segmentoinnista saatavien hyötyjen toteuttamista, *vastauslaajuuden*, eli kuinka laajasti segmentille voi toteuttaa markkinointiratkaisuja siten, että kaikki segmentin sisällä reagoivat niihin vielä toivotulla tavalla ja *toiminnallisuuden*, eli kuinka selkeästi segmentointi ohjaa oikeita markkinointiratkaisuja kohti, mukaan (Sicilia & García, 2003).

Kuten segmentoinnin ongelmakohtia käsittelevässä kappaleessa jo esiteltiin, useimmiten asiakassegmentoinnit toteutetaan yrityksen omina tutkimuksina, niissä tutkitut asiat eivät ole suuren yleisön saatavilla kirjallisuuden välityksellä. Muutamia esimerkkejä segmentoinnin suunnittelusta on kuitenkin julkaistu vuosien saatossa, kuten Swinyard ja Sturtmanin artikkeli vuodelta 1986, joka käsittelee paikallisen ravintolan asiakassegmentointia. (Swinyard & Sturtman, 1986). Swinyard & Sturtman lähtivät liikkeelle segmentoinnissaan tunnistamalla tekijöitä, jotka vaikuttavat mahdollisesti kiinnostukseen ja halukkuuteen ostaa palveluja ravintoloilta, joita olivat *sosiaalinen aspekti* (kuinka ravintolan sijainti ja luonne toimii potentiaalisena tapaamispaikkana), *laadullinen aspekti* (ruuan ja palvelun laatu pysyy samana käyntikerralta toiseen, samoin kuin ystävällinen asiakaspalvelu), *tuoteaspekti* (tarjoiltavat annokset ovat suuria) sekä *lisäarvoa tuovat aspektit* (Viehättävä henkilökunta, eloisa musiikki ja ravintolaympäristö, helppo sijainti). Tämän jälkeen asiakkailta kyseltiin näiden tekijöiden täytymistä kyseisessä ravintolassa sekä kuinka ne täytyisivät heidän ideaaliravintolassaan sekä millaisia viimeaikaisia ravintolassa käyntikokemuksia ja syitä ja motiiveja ravintolavierailuiden tiheyteen (kuten rahatilanteen vaikutus vierailutiheyteen tai oman kokkausinnostuksen vaikutus) ja lopuksi vastaajien tarkat demografiset tiedot. Näiden tietojen perusteella tehtiin asiakkaiden klusterointia vastaavalaisten ideaalien, motiivien ja demografisten piirteiden jakavien ryhmien kesken, josta muodostettiin ravintolan eri asiakassegmentit. (Swinyard & Sturtman, 1986.) Vaikka tässä esimerkissä olikin kyseessä ravintola-asiakkaiden segmentointi, heidän esittämiään аспекteja kysymyksissä voidaan hyödyntää myös muiden alojen asiakkaiden ymmärtämisessä ja sitä seuraavassa segmentoinnissa erilaisen asiakkaiden ryhmittelemiseksi.

Usein segmentoinnissa hyödynnettyjä tekniikoita ovat RFM-analyysi, eli analyysi jossa asiakkaan viimeisintä ostohetkeä (eng. *Recency*), ostojen toistuvuutta (eng. *Frequency*) ja ostoihin käytettyä rahamäärää (eng. *Monetary*) käytetään asiakkaan ostokäyttäytymisen analysointiin ja tämän kautta asiakassegmentointiin (Singh ym., 2014). Tämän lisäksi useat alat, joille palvelukokemus on keskeisessä osassa, ovat keskittäneet säännöllisten ostajien kohdentamiseen (eng. *Targeting frequent buyers*), jossa asiakas saa etuisuuksia toistuvista ostoistaan muodossa. Hyvä esimerkki tästä mallista ovat esimerkiksi lentoyhtiöt, joissa nykyään lähes jokaisessa on kanta-asiakasohjelmat usein lentäville, jonka ansioista he saavat tarjouksia lipuistaan plus-alennuksia uusista ostoksistaan. (Lynn, 2011.) Nämä mallit ovat kuitenkin toimeksiantaja yrityksen kaltaisissa palveluntarjoajayrityksissä tehottomia, sillä palveluiden ostamiseen sitoudutaan usein pitkiksi ajanjaksoiksi ja asiakas keskittyy ostamaan yhden pakettiratkaisun ongelmiinsa (tässä tapauksessa sähkösovimus), jota ei täydennetä satunnaisostoksilla, eikä

asiakkaiden ostoaktiivisuudet eroa toisistaan sähkösopimusten pakettimuotoisuuden vuoksi. Yksi käytetty segmentointiratkaisu on myös arvosegmentointi (eng. *value segmentation*), jossa asiakkaat jaotellaan eri segmentteihin heidän yritykselle tuoman arvonsa perusteella (Sicilia & García, 2003). Tämäkään ei kuitenkaan toimi kattavana segmentointimetodina energia-alan yritykselle, joiden palvelut on kohdistettu kaikilla sen toimialueella toimiville talouksille ja toimitiloille, jotka vaativat sähköä toimiakseen, eli tänä päivänä lähes kaikille rakennuksille ja asumuksille. Esimerkiksi Jyväskylän Energia tarjoaa kolmea erilaista sähkösopimusta, *määräaikaisen* (12/24 kuukautta, sama hinta koko sopimuskauden ajan, uusi tarjous sopimusajan loppuessa), *toistaiseksi voimassa olevan* (päättyy asiakkaan lopettaessa, hinnat päivittyvät 3 kuukauden välein, hinnat seurattavissa laskuista ja nettipalvelusta) sekä *Pörssisähkö* (hintaa perustuu pohjoismaisen NordPool-sähköpörssin joka tunnille määrittelemään spot-hintaan, asiakas säästää seuraamalla hinnan kehittymistä ja hyödyntämällä edulliset tunnit virran kulutuksessaan, lisäksi voimassa Jyväskylän Energian marginaali, palvelun käyttö edellyttää kuukausimaksun maksamisen) -sopimuksen (Jyväskylänenergia.fi, 2016). Sodenkamp ym. (2015) ehdottavatkin teoksessaan energia-alan yrityksille asiakasportfolioiden laatimista neljän tekijän mukaisesti: *asiakkaan kotitalouden piirteiden* (kuten talouden lämmitystyyppi, mahdollinen kiinnostus vihreän energian/biokaasun hyödyntämiseen sekä kotitalouden asuntotyyppi), *asiakkaan energiakulutustietojen, asiakkaan perustietojen* (kuten asukkaan osoite, ikä, perhestatus ja muu demografisiin segmentointitekijään lukeutuva data) ja muun *asiakkaaseen liittyvään informaation* (kuten naapurusto, sosiodemografinen tilastotietoa ja paikalliset sääolosuhteet) pohjalta. Näistä muodostuvan asiakasportfolion pohjalta yritys voi toteuttaa asiakasanalytiikan asiakaskokemukselle lisäarvoa tuovia ratkaisuja ja ymmärtää ja tunnistaa paremmin jo palvelemissaan segmenttejä, joissa energiakulutukseen vaikuttavat tekijät ja asiakkaan itsensä maksuvalmius ja arvot on huomioitu. (Sodenkamp ym., 2015.)

Asiakassegmentointiin ei ole olemassa yhtä, kaikille sopivaa ratkaisuja, jonka vuoksi useat asiakassegmentointitutkimukset toteutetaan yrityskohtaisesti näiden toimeksiantona. Segmentointiin on kuitenkin olemassa olevia malleja, kuten RFM-malli, usein-ostavien kohdentamisen -malli sekä arvosegmentointi, mutta nämä soveltuvat yrityksille, jotka tarjoavat tuotteita ja palveluita joissa ostajan ostotiheys ja ostomäärä sekä rahallisesti että määrällisesti tarjolla olevista tuotteista/palveluista voi vaihdella. Energiateollisuudessa sopimukset tehdään määräaikaisina sopimuksina, sopimus, johon päädytään, kattaa tavallisesti jo kokonaisuudessaan asiakkaan tarpeet tuotteelle eikä asiakkaalla ole esim. samaan kiinteistöön useampaa sähkösopimusta, jonka vuoksi nämä perinteikkäät kulutukseen pohjautuvat mallit eivät sovellu energiateollisuudessa käytettäväksi. Sen sijaan Kotler ym. (2008) esittämä viitekehys eri tekijöistä jotka auttavat segmentoimaan asiakkaita ja potentiaalista markkinaa, ovat hyödynnettävissä oikeastaan jokaisella liiketoiminta-alueella, joista Sodenkamp ym. (2015) on vielä tiivistänyt, mitkä Kotlerin esittämistä tekijöistä voisivat soveltua juuri energia-alan asiakkaiden jaottelemiseen. Lynn (2011) tarjoaa pohjan asiakassegmentoinnin

onnistuneisuuden arvioimiselle, esittämällä konkreettisia vaatimuksia eri segmenteille riippumatta niiden koostamiseen käytetyistä tekijöistä – kunhan segmenttien koot pystytään määrittämään, niiden koot ovat tarpeeksi suuria oikeutamaan erilaistetut tuotteet ja kohdentamistoimet, segmenttien tavoittelemiseen voidaan käyttää ryhmille uniikkeja kanavia ja segmentit eivät ole aikariippuvaisia (olemassa olevia tietyinä tapahtumahetkenä, mutta segmentti kuihtuu ajan myötä pois), segmenttien toimivuudelle on jo tarpeeksi perusteita sen hyödyntämiseksi markkinoinnin toimissa. Segmentoinnissa on tärkeää myös Swinyardin ja Sturtmanin (1986) esittelemä kysymysten kautta toteutettava segmentointitutkimus asiakassegmentoinnin toteuttamiseksi – vaikkei yrityksellä olisi mahdollista toteuttaa kyselyä asiakkailleen, esitettävien kysymysten laatiminen itsessään voi jo toimia avaimena ymmärrykselle siitä, mitkä ihmisten todelliset motiivit tarjottavan tuotteen tai palvelun käytössä on ja kuinka niihin voi vastata paremmin. Realisoitujen motivaatioiden ja arvojen pohjalta asiakkaita voi jaotella vielä merkityksellisempiin ryhmiin, kuin mitä esim. demografiset tekijät asiakkaistaan kertovat, ottaen huomioon ostohaluun vaikuttavat tarpeet ja arvot.

3.4 Yhteenveto ja viitekehysten rakentaminen

Asiakasanalytiikka toimii kilpailuetuna yrityksille teknologian vähentäessä perinteisiä kilpailuun vaikuttavia tekijöitä (Kolyshkina & Simoff, 2007). Asiakasanalytiikkaa voidaan hyödyntää esimerkiksi asiakasprofiilien luomiseksi, jonka avulla voidaan suorittaa datalähtöistä markkinointia suoraan eri asiakasprofiilin omaaville asiakkaille (Mulvenna ym., 1998) ja jakaa asiakkaita asiakassegmentteihin, jolle suoritetaan segmentille soveltuvia markkinointiratkaisuja palveluntarjonnan personalisoinnilla ja mainonnan kohdentamisella (Kotler ym., 2008). Asiakasanalytiikka voi tuottaa lisäarvoa yrityksille syvällisemmän asiakasymmärryksen muodossa, jota voi hyödyntää palveluyrityksille tärkeässä asiakassuhdehallinnossa (Sicilia & Garcia, 2003) sekä asiakkaalle tuotettavana lisäarvona esimerkiksi heidän kulutuskäyttäytymistään kuvaavan visualisoidun materiaalin avulla (Sodenkamp ym., 2015). Asiakassegmentointi on yksi tärkeimpiä toimia markkinoinnin edistämiseksi, mutta sen toteuttamiseen ei ole olemassa yhtä ainoaa kaikille soveltuvaa ratkaisua. Kotler ym. (2008) on kuitenkin esitellyt viitekehukset, jossa esitellään tärkeimmät asiakkaan luokitteluun merkityksellisiin ryhmiin hyödynnettävät tekijät: maantieteelliset, demografiset, psykografiset ja käytökselliset tekijät. Näiden lisäksi yritys voi hyödyntää olemassa olevia segmentointimalleja, jos heillä on mahdollisuus jaotella asiakkaansa pääasiallisesti heidän ostokäyttäytymisensä (Singh ym., 2014) ja yritykselleen tuottamansa arvon kautta (Sicilia & Garcia, 2003). Tapauksissa, jossa yrityksen liiketoiminnan luonne ei sovellu näiden mallien hyödyntämiseksi heidän liiketoiminnalleen olennaisten segmenttien tunnistamiseksi, Swinyardin ja Sturtmanin (1986) esittelevät keinon muodostaa asiakkaille kohdistettujen kysymysten kautta ymmärrystä eri asiakkaiden motiiveista käyttäen palvelua ja jaotella asiakkaat näiden motiivien ja halujen lisäksi asiakkaiden demografisten tietojen lisäksi

eri segmentteihin. Lynn (2011) asettaa vaatimukset onnistuneelle segmentoinnille pohjaten toteutettujen segmentoinnin käytön realisointiin markkinoinnin toiminnoissa. Tämän lisäksi Sodenkamp ym. (2015) esittävät vielä asiakasportfolioiden muodostamista pohjustuen energia-alalle oleellisiin asiakkaan maantieteellisiin, demografisiin, psykografisiin ja käytöksellisiin tekijöihin, joita voi hyödyntää energia-alalla asiakassegmentointia toteutettaessa.

Koska yrityksiä segmentoinnit toteutetaan usein heidän teetättämiensä tutkimuksien kautta, tämä Swinyardin ja Sturtmanin lähestymistapa asiakkaan perimmäisten motiivien ymmärtämiseen tuotteen käytössä ja käyttöhalukkuudessa mahdollistaa yrityksen luoda kysymysten muodossa viitekehys asiakkaidensa segmentoinnille, eli saavuttaa ymmärrys siitä, mitkä tekijät ovat asiakkaan kannalta merkittäviä ja sen avulla jakaa heidät ryhmiin. Hyödynnämme Sodenkampin (2015) ehdotelmaa energia-alalle soveltuvasta asiakasportfolioista sekä Kotlerin esittämiä segmentointitekijöitä asiakkaiden jaottelemisen helpottamiseksi. Pelkkien segmentointitekijöiden tunnistaminen ei kuitenkaan yksistään riitä toimivan segmentointimallin tekemiseksi, ellei ole selvillä ratkaisuja, joihin asiakassegmentointia halutaan hyödyntää - energiateollisuuden kaltaisessa palveluyrityksessä toivottavien ratkaisujen voisi kuvitella olevan kohdennettujen palvelujen ja kohdennetun mainonnan tavoittelu yhdessä asiakasymmärryksen parantamisen kanssa. Kun yritys tietää, mitä tarkoitusta varten sen tulee ryhmiä tunnistaa, sillä on jo suurempi mahdollisuus määrittää itselleen merkityksellisiä asiakassegmenttejä tarkoituksettomien ja käyttökeltottomien ryhmien sijaan. Koska sähköteollisuuden palvelut ovat tarpeen jokaiselle nykyaikaisissa olosuhteissa asuvalle ihmiselle, jokainen yrityksen toimialueella sijaitseva ihminen on potentiaalinen asiakas. Näistä ostopäätökseen vaikuttavat henkilöt määräytyvät perhestatuksen mukaan ja energian ostopäätös voi olla pariskunnissa yhteinen päätös tai kuulua toisen vastuualueeksi. Tarkempaa asiakasymmärrystä yritys voi hakea suorittamalla asiakastutkimuksen asiakkaiden kokemuksista palvelusta, toiveista ideaalille versiolle palvelua ja yleisillä kysymyksillä jotka avaavat asiakkaan motivaatiota juuri kyseisen palvelun käyttämiseen.

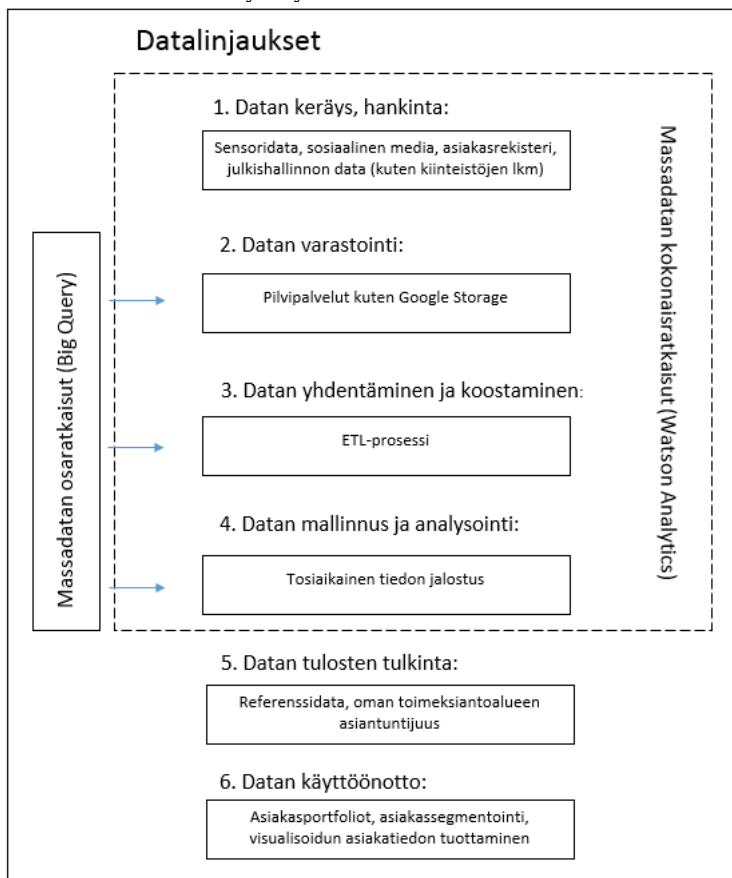
Edellä tehty yhteenveto asiakasanalytiikan kappaleessa osoittaa, millaista dataa keräämällä ja tietoa jalostamalla toimeksiantajayrityksen kaltainen palveluntarjoaja voi lähestyä asiakasanalytiikan ja asiakassegmentoinnin tekemistä datan pohjalta. Datan hyödyntämiseksi ja käsittelemiseksi tarvitaan kuitenkin ratkaisuja, jota tukemaan laaditaan massadatan käsittelyn viitekehys pohjustuen toisessa kappaleessa käsiteltyyn kirjallisuuteen. Kowsar ym. (2014) sekä Jagadish ym. (2014) ovat molemmat käsitelleet teoksissaan datan käsittelyn kokonaisprosessia datan hyödyntämiseksi. Heidän teoksiensa perusteella datan käsittelyn voi yksinkertaistaa kuuteen (6) päävaiheeseen:

1. Datan hankintaan, eli keräämiseen
2. Datan varastointiin
3. Datan yhdentämiseen ja koostamiseen
4. Datan mallinnukseen ja analysointiin
5. Datan tulosten tulkintaan
6. Datan käyttöönottoon

Kowsar ym. (2014) ja Jagadish ym. (2014) nostavat näistä käsittelyvaiheista erityisen haastaviksi datan varastoinnin ja yhtenäistämisen, jotka vastaavat suoraan massadatan V-mallin kahteen V-kirjaimeen: volyymiin ja varieteettiin. Tämän lisäksi Kowsar ym. (2014) nostaa esille myös huolen datan määrän nopeasta kasvusta (*Velocity*) joka on myös yksi V-mallin edustaja. Datan määrän nopea kasvu aiheuttaa ongelmia kaikissa perinteisissä datan käsittelyprosesseissa: datan hyödyntämisen esteeksi tulevat datansiirrossa verkon nopeus sekä laitteiston kyky rinnakkaisen datan lähettämiseen. Nämä rajoitteet aiheuttavat datan määrän kasvaessa datan lähettämisen ja siirtämisen hyötysuhteen laskua. (Kowsar ym., 2014.) Ayankoya ym. (2014) sen sijaan esittää teoksessaan datan todenmukaisuudesta (*veracity*) eli olennaisen datan löytämisestä suurista datamääristä sen hyödyntämiseen tarkoitettujen yhteyksien osalta. Ayankoya ym. (2014) esittää tähän ongelmaan ratkaisuksi datalinjauksia, joiden avulla varmistutaan, että toimintojen kannalta hyödyllistä dataa kerätään, varastoidaan, haetaan ja käsitellään varastoista, analysoidaan tavoitteiden kannalta olennaisella tavalla ja hyödynnetään liiketoiminnassa datakeskeisien toimintamallien avulla.

Nämä 4V-mallin edustajat luovat suurimmat ongelmat massadatan hyödyntämiseen samalla, kun ne määrittävät sitä. Alla esitetty viitekehys on kehitetty käsitellyn kirjallisuuden pohjalta vastaamaan näihin 4V-mallin aiheuttamiin ongelmiin, samalla kun se huomioi Kowsar ym. (2014) sekä Jagandish ym. (2014) esittämät vaiheet datan hyödyntämiseksi, samalla käyttäen esimerkkinä toimeksiantajayrityksen kaltaiselle yritykselle olennaisia kanavia asiakasanalytiikkaan ja -segmentointiin.

KUVIO 2 – Datan hyödyntämismalli



Viitekehyksessä esitetään datan hyödyntämisen eri askeleet sekä ratkaisut niiden käsittelyyn. Kaiken datan hyödyntämisen taustalla ovat datalinjaukset, jotka määrittävät, mitä dataa kerätään, mitä varastoidaan, mitä dataa käsitellään ja mallinnetaan rikastettavaksi tiedoksi tukeakseen liiketoiminnassa tehtäviä tulokintoja datan tulosten tulkinnan ja käyttöönoton osalta. Mm. Frans (2012) sekä Reed ja Dangorra (2015) ovat nostaneet töissään esille erityisen huolen datan hyödyntämisen ongelmista liittyen tietämättömyyteen siitä, millaista dataa tulisi hyödyntää, mihin datalinjaukset tarjoavat vastauksen. Massadatan hyödyntämisen kokonaisratkaisut, kuten Watson Analytics vastaavat tavallisesti neljästä ensimmäisestä datan hyödyntämisen osa-alueesta (IBM.com, 2016), kun taas BigQuery -palveluun tulee käyttäjän itse ladata hankkimansa data (Fernandes & Bernardino, 2015). Osa-alueiden alla olevat laatikot vastaavat kysymyksiin siitä, miten datan hyödyntämisen osa-alueita voi lähestyä yksittäisten ratkaisujen kautta – ETL-prosessi toimii rakenteettoman tiedon yhtenäistämiseen (Nakucçi ym., 2014), jonka jälkeen tosiakaisen tiedon jalostuksen ratkaisuja voidaan hyödyntää rakenteellisen tiedon analysointiin, mallintamiseen ja trendien ennustamiseen (Lao & Tang, 2005). Tämän jälkeen kaksi viimeisintä prosessia ovat ihmispohjaisen päätöstentien alaisuudessa, eli datan tuloksia tulkitaan, jotta tiedetään, millaisia toimia datan tulosten perusteella tulisi liiketoiminnassa tehdä.

Vaikka tutkimuksessa esitettiin muutamia esimerkkejä osa-ratkaisuista massadatan käsittelemiselle, kuten MISO-algoritmi (*MultiStore Online tuning algorithm*), joka ratkaisee datan volyymin ja nopean kertymisnopeuden ongelmia ulkoisen ja sisäisen tiedon erillisen varastoinnin ratkaisulla (LeFevre ym., 2014) sekä Hadoop viitekehys suurien datamäärien käsittelemiseksi klusteripohjaisesti (Wang ym., 2009) ja MapReduce -ohjelmointimalli rinnakkaisen ja samanaikaisen tiedon käsittelyn ja prosessoinnin (Ullman, 2015), näitä ratkaisuja ei otettu osaksi viitekehystä, sillä Hadoop ja Mapreduce ovat jo osittain käytössä muissa esitetyissä ratkaisuissa ja MISO-algoritmi on yksittäinen malli lukuisien muiden joukossa, joka myös hyödyntää Hadoopia suuresti toiminnassaan. Viitekehyksellä on tarkoitus luoda kuva erilaisista tavoista, kokonaisratkaisuista osa-ratkaisuihin, datan hyödyntämisen eri vaiheille. Viitekehysten tavoitteena on tarjota mahdollisuus tarkastella helposti sitä, mitä osa-vaiheiden ongelmia organisaatio on jo nykyisillä ratkaisuillaan tukenut ja kuinka paljon potentiaalisia hyötyjä yritys saisi mahdollisesta laajempien ratkaisujen käyttöönotosta.

Yrityksen määritettyä nykyisen datan hyödyntämisen tilansa viitekehyksessä esitettyyn malliin, sen tulisi myös määrittää, kuinka he voivat tehostaa nykyisiä toimiaan määrittääkseen parhaiden heitä palvelevat asiakassegmentit. Mahdollisessa asiakassegmentointimallissa yritys voisi hyödyntää ensin toimialueen asukas- ja yritysdataa selvittääkseen koko potentiaalisesta asiakaskannasta palvelemissa asiakkaat. Tällä tavoin yritys voisi määrittää, ketkä potentiaalisista asiakkaista ovat jonkin kilpailevan palveluntarjoajan palvelujen alaisuudessa. Seuraavaksi omista ja kilpailijan asiakkaista pitäisi pyrkiä löytämään merkityksellisiä ryhmiä tukeutuen asiakkaiden demografisiin tekijöihin. Lopuksi merkityksellisiä ryhmiä tulisi pyrkiä määrittämään pidemmälle ja arvottamaan asiakkaan energiankulutusdatan ja kiinteistötietojen kautta.

Koska energia-ala tarjoaa palvelua, joka on pakollinen kaikille nykyaikaisille toimijoille ja asukkaille jokapäiväisen elämän toimintojen mahdollistamiseksi, se saa tutkittua, löytyykö eri palveluntarjoajien alaisuudessa omalla toimialueellaan olevista henkilöistä/organisaatioista samojen muuttujien kautta ryhmiteltävissä olevia asiakasjoukkoja, segmenttejä, joiden tarpeisiin se ei vielä itse tarjonnallaan vastaa. Asiakkaiden yleistietoihin ja demografisiin tekijöihin tutustumalla palveluntarjoaja voi selvittää, omien ja kilpailijoiden asiakkaiden joukossa mahdollisesti ilmeneviä huomattavia eroavaisuuksia, jotka voivat paljastaa esim. palvelujen hinnoittelun vaikutuksesta tiettyjen asiakasryhmien ostopäätöksiin. Kun näitä tietoja verrataan oman ja kilpailijoiden asiakkaisiin kuuluvien energiankulutus ja kiinteistötietoihin, jotka ilmentävät yritykselle tuottoisimmat asiakkaat, voidaan tarkastella, mitkä samanlaisia demografisia tekijöitä kantavat segmentit ovat yritykselle potentiaalisimpia, palveleeko yritys jo kaikkia näistä onnistuneesti vai onko osa tuottoisimmista asiakkaista kilpailijoiden asiakkaina.

Potentiaalinen segmentointimalli voisi sisältää kolme Sodenkampin (2015) energia-alan asiakasportfolioissa ja segmentoinnissa hyödynnettäväksi ehdottamasta muuttujasta - asiakkaan perusdatan (demografiset tiedot), asiakkaan kiin-

teistön tiedot (kuten kiinteistön kokoluokka ja sen energiankulutukseen vaikuttavat muuttujat) sekä asiakkaan energiankulutusdatan. Ehdotelmassa tarvittavan datan keräämiseksi yrityksen tulisi oman datansa lisäksi hyödyntää eri organisaatioiden saataville asettamaa dataa sekä mahdollisesti datan ostamista relevanttia dataa kerryttäviltä yrityksiltä. Palveluntarjoaja voi hyödyntää ehdotelman ensimmäisessä kohdissa toimialueen hallintoelimien kerryttämää tietoa alueen asukkaista sekä saatavilla olevia yritysrekistereitä ja verrata näitä omaan asiakasrekisteriinsä kilpailijoilla olevien asiakkaiden määrittämiseksi. Omien asiakkaidensa merkittävimmät demografiset tiedot yritys saa todennäköisimmin omaan asiakasohjelmaansa keräämästään datasta, samoin kuin asiakkaan energiankulutusdatan eri sensoreiden ja niiden keräämän datan avulla. Asiakkaan kiinteistöjen tietojen keräämisessä voisi mahdollisesti hyödyntää kiinteistöalalla toimivien ja niistä tietoja keräävien organisaatioiden kerryttämää dataa joko näiden kerryttämän datan ostolla tai hyödyntämällä organisaatioiden julkisesti saataville asettamaa tietoa.

Kun yritys on perehtynyt näihin tietoihin ja tunnistanut asiakkaistaan samoja muuttujia jakavia ryhmiä, hän voi arvioida muodostuneiden segmenttien merkityksellisyyttä arvioimalla ryhmien jakamia arvoja, tarpeita ja resursseja palveluiden ostoon – täsmääkö alustavasti muodostuneiden segmenttien mahdollisuudet ja motiivit palveluiden ostoon toisiaan niin paljon, että ryhmää voi palvella yhdellä tuoteratkaisulla ja lähestyä samalla tapaa? Yritykselle merkityksellisten ja palveltavien segmenttien tunnistamiseksi alustavan segmentoinnin jälkeen voidaan hyödyntää Swinyardin ja Sturtmanin (1986) asiakasymmärrystä edistävien kysymysten laadintaa ryhmien tunnistamiseksi, jotka eivät yksistään selity demografisilla samanlaisuuksilla. Sen sijaan poikkeavalta vaikuttavia ilmentymiä muuten homogeenisissä ryhmissä voi tarkastella tutustumalla Sodenkampin (2015) asiakasportfolioihin hyödynnettäväksi esitettyyn neljänteen muuttujaan, jota ei segmentoinnissa hyödynnetty – asiakkaaseen liittyvät tekijät, kuten naapuruston vaikutus palvelun käyttöönottoon tai ostopäätökseen.

Näitä segmentointitekijöitä tarkasteltuaan yrityksellä tulisi olla mahdollisuus vertailla, mitä motivaatioita asiakkailla olisi valita kilpailijan palvelu omien palvelujensa sijasta – tarjoaako kilpailija jollekin erityiselle asiakassegmentille erilaistettua palveluja, jotka vastaavat heidän tarpeisiinsa tai onko kilpailijalla olemassa jokin kanta-asiakkuuden palkitsemisohjelma, jota yrityksellä itsellään ei ole käytössä? Poistuneiden asiakkaiden tietoja tutkimalla voi myös selvittää niitä arvoja ja odotuksia, mihin yritys itse ei tarjonnallaan voinut vastata ja muita mahdollisia syitä toisen palveluntarjoajan alaisuuteen siirtymiselle ja muuttaa asiakkaille tuotettuja arvoja ja palveluita sen mukaisesti.

Tutustumalla yleisiin trendi-ilmiöihin markkinoilla yrityksellä on myös mahdollisuus nähdä, mitä mahdollisia arvoja on ilmassa, joita he eivät tarjonnallaan tai markkinointiratkaisuillaan ole tuoneet tarpeeksi esille. Esimerkiksi sosiaalisen median analytiikkaa hyödyntämällä trendi-ilmiöitä on mahdollista seurata. Jos yritys myös huomaa segmentointia tehdessään asiakasryhmiä, joihin he eivät vielä ole mainonnallaan vedonneet tai joille tarjottavia palveluja he voisivat

erilaistaa vastaamaan enemmän heidän elämäntilanteisiinsa ja arvoihinsa, he voivat kehittää mainonnassa tehtäviä ratkaisuja näiden osalta.

Palveluntarjoaja yrityksenä energia-alan yrityksellä olisi kuitenkin mielestäni oleellisinta vastata tarjonnassaan eri elämäntilanteisiin ja vaatimuksiin. Esimerkiksi mökkiläiset kaipaavat sopimuksia, joilla he voivat hyödyntää asuntoaan edullisesti joko satunnaisesti ympäri vuoden tai vain mökkeilyn eri sesonkiaikoina. Kun yritys hyödyntää datan mahdollisuutta tuoda ilmi esimerkiksi alueella olevien kiinteistöjen aiheuttamia erilaisia tarpeita energiaratkaisuissa (mökkit) ja vallitsevat arvot ja trendit (kuten ympäristöystävällisyys ja suomalaisen työn tukeminen), yrityksen on mahdollista saada kattavampaa käsitystä asiakkaidensa sekä potentiaalisten asiakkaidensa mahdollisuuksista sekä motivoituneista palvelujen hyödyntämiseen ja hyödyntää myös näitä tekijöitä segmenttien kokoamisessa, arvottamisessa ja niille suoritettavien toimintojen suunnittelussa.

4 TOOREETTISEN OSUUDEN YHTEENVETO

Massadata kuvaa datamääriä, joita on nykyisillä ratkaisulla vielä haastavaa hyödyntää. Sille ei ole olemassa vielä yhtä yleisesti hyväksyttyä määritelmää, mutta yleisesti sitä kuvataan 3V-mallin avulla, joka sisältää datan määrän, kasvunopeuden ja monimuotoisuuden. Massadatan määrälle ei ole asetettua tiettyä rajaa, jonka ylitettyään data laskettaisiin massadataksi, vaan mitä enemmän teknologia kehittyy ja mitä nopeammin sillä pystyy käsittelemään dataa ja mitä tehokkaammin sillä voi varastoida suuria määriä dataa, sitä korkeammille datamäärille massadatan käsite myös siirtyy. Datan suuren määrän vuoksi sen käsittelyn tekniset haasteet ovat suurin ongelma massadatan potentiaalın hyödyntämiseksi, mutta myös massadatalle ominaiset datan monimuotoisuus ja datan määrän nopea kasvu ovat sen käsittelyyn vaikuttavia haasteita, jotka on huomioitava massadatan käsittelyä ja hyödyntämistä käsiteltäessä. Massadatan hyödyntämiseksi onkin tärkeää löytää helposti käytettäviä ratkaisuja, jotka mahdollistavat nopeasti kertyvästä, monimuotoisesta, suuresta määrästä dataa hyödyllisten tietojen rikastamisen ilman, että datan varastointi ja käytettävissä olevan laitteiston suoritusaste osoittautuvat ongelmiksi. Teknillisten haasteiden lisäksi massadatan potentiaali sitoutuukin suuressa osin datalinjauksien laatimisen puutteeseen tai niiden laatimiseen itselleen ja organisaationsa tavoitteita tukemattoman datan pohjalta.

Massadatan on mahdollistanut ja osaltaan myös aiheuttanut suuria muutoksia liiketoimintatiedon hallinnassa ja analytiikassa. Se on johtanut useampien datalähteiden hyödyntämiseen yrityksen strategisessa päätöksenteossa ja liiketoiminnan eri osa-alueiden tapahtumien ennustamiseen. Massadata on luonut tarvetta uusien osaajien kouluttamiseksi liiketoimintatiedon hallinnan ja analytiikan saralla sekä erillisten massadata keskeisten toimikuvien luomiseksi organisaatioihin. Tutkimukset ovat osoittaneet, että massadatan hyödyntäminen yrityksen liiketoiminnassa ja liiketoimintatiedon hallinnassa ja analytiikassa tuottaa niille merkittäviä tuloksia mm. lisäarvoja syvällisemmän ymmärryksen mahdollistajana, jota voi hyödyntää asiakasanalytiikassa. Asiakasanalytiikka toimii kilpailuetuna yrityksille teknologian vähentäessä perinteisiä kilpailuun vaikuttavia tekijöitä ja sen avulla voidaan esimerkiksi luoda asiakasprofiileja ja tuottaa asiakkaalle lisäarvoa heihin liittyvän tiedon visualisoinnilla. Asiakasanalytiikka mahdollistaa myös datalähtöisen markkinoinnin suorittamisen suoraan eri asiakasprofiilin omaaville asiakkaille ja asiakkaiden jakamisen asiakassegmentteihin. Asiakassegmentointia hyödynnetään suorittamalla segmentille soveltuvia markkinointiratkaisuja palveluntarjonnan personalisoinnilla sekä mainonnan kohdentamisella. Tärkeimmät asiakkaan luokitteluun merkityksellisiin ryhmiin hyödynnettävät tekijät ovat maantieteelliset, demografiset, psykografiset ja käytökselliset tekijät, joiden lisäksi yritys voi hyödyntää olemassa olevia segmentointimalleja, jos heillä on mahdollisuus jaotella asiakkaansa näiden ostokäyttäytymisensä ja yritykselleen tuottamansa arvon kautta. Palveluluontoisissa yrityksissä

nämä mallit eivät sovi asiakkaiden segmentointiin, jonka sijaan käsitystä eri asiakasryhmistä voi saada muodostamalla asiakkaan ostomotiiveja ja odotuksia avaavia kysymyksiä, joiden vastaukset ryhmittelemällä saadaan homogeenisiä asiakasryhmiä.

Tutkimuksen tutkimuskysymys oli: "Kuinka massadata -pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa?". Tutkimuksen kirjallisuuskatsaus avasi useita eri kanavia, jotka kerryttävät dataa, joita ei nykyisellään vielä hyödynnetä asiakasymmärryksen lisäämiseen. Massadatan todellinen potentiaali asiakasanalytiikalle onkin syvällisemmän asiakasymmärryksen saavuttaminen ja sen hyödyntäminen kohdennetussa markkinoinnissa ja asiakassegmentoinnissa. Aiheesta ei ole kuitenkaan tehty vielä tutkimusta, joka käsittelee massadatan asiakasanalytiikassa hyödyntämisen konkreettisia käyttötapoja, joten tämä havainto perustuu kirjallisuudessa esitettyihin kanaviin, jotka kerryttävät massadataa ja joita ei ole vielä hyödynnetty tehokkaasti asiakasanalytiikassa. Lisäksi asiakasanalytiikka voidaan tehostaa hyödyntämällä uusia kanavia, joilla on asiakasanalytiikalle annettavaa, mutta jotka eivät ole vielä yleisesti mukana asiakasdatan kerryttämisessä – massadatan perusluonteeseen määritellyt merkittävät massa ja monimuotoisuus sitoutuvat vahvasti yhteen muun muassa verkkokäyttäytymisen, trendien, mediaseurannan ja ulkopuolisen datan hyödyntämisen kanssa, jolla massadatan avulla voidaan tehostaa asiakkaista, sekä olemassa olevista että potentiaalisista, rikastettavaa ymmärrystä.

Tutkimuksen apukysymykset olivat: "Millä metodeilla yritys pystyy tehokkainten tunnistamaan ja rikastamaan asiakasanalytiikan kannalta olennaista tietoa?" sekä "Millainen asiakassegmentointi -malli mahdollistaa yrityksen tunnistaa paremmin sekä potentiaalisia että jo palvelemaan asiakassegmenttejä?". Asiakassegmentointi-aiheiseen kysymykseen vastattiin asiakasanalytiikan yhteenveto-osiossa esitellyssä viitekehyksessä, joka yhdessä asiakasanalytiikassa esitetyn poistuvien asiakkaiden analysoinnin työkalujen kautta avaa, millä tavoin asiakassegmentointi voi tuottaa hyötyä olemassa olevien asiakkaiden palvelemiseen sekä tunnistamaan niitä asiakkaita sekä niiden asiakkaiden piirteitä, jotka eivät vielä ole yrityksen asiakkaina. Asiakassegmentointimalli hyödyntää yrityksen omaa dataa, yleisesti saatavilla olevaa ja julkisesti jaettua tietoa ja dataa sekä mahdollisesti relevanttia dataa kerääviltä yrityksiltä datan ostoa. Malli tukeutuu energia-alalle tyypilliseen toimialuesitoutuneisuuteen, joka toimii ensimmäisenä askeleena omien asiakkaiden tunnistamiseen kaikkien potentiaalisten asiakkaiden joukosta. Tämän jälkeen asiakkaihin tutustutaan syvällisemmin näiden demografisten tietojen avulla, jonka jälkeen viimeisenä vaiheena on muodostettujen asiakasryhmien arvottaminen näiden energiankulutusikäyttäytymisen ja kiinteistöjen energiankulutukseen vaikuttavien tekijöiden antaman kulutustennusteen mukaan. Asiakasanalytiikan kannalta olennaisen tiedon tunnistaminen sen sijaan on aihe, jonka rajoitteet kulkevat käsi kädessä hyödynnettävien kanavien kanssa – kaikki potentiaaliset datalähteet kartoitettuaan yritys voi määrittää, minkä kanavien dataa jo hyödynnetään, millaista tietoa datasta saadaan rikastettua ja mitä lisäarvoa se tuo verrattuna muista kanavista saatuihin tietoi-

hin. Yrityksen sisäiset datalinjaukset, joissa määritetään yritysten tietyissä toiminnoissa hyödynnettävissä oleva data, ovat tässä tärkeässä osassa. Suurin osa kaikkiin analytiikkaan liittyvissä tutkimuksissa esitetyistä ongelmista liittyi vääränlaisen datan keräämiseen tai analytiikan käyttämiseen väärissä tilanteissa. Kirjallisuus itsessään ei kuitenkaan tarjoa suoria vastauksia siihen, mitä metodeja yrityksen tulisi hyödyntää asiakasanalytiikan kannalta relevanttimman datan keräämiseksi ja se vaikuttaakin enemmän aiheelta, johon kaikki potentiaaliset datakanavat kartoitettuaan yritys voi kokeilujen kautta löytää tehokkaimman ratkaisun, joka tuottaa juuri heidän tarpeisiinsa soveltuvia lopputuloksia. Olemassa olevia kanavia, joita suositellaan hyödynnettäväksi asiakasanalytiikan laatimisessa, on mm. sosiaalisen median kanavat, jotka kertovat paljon sitä käyttävien yksilöiden lisäksi myös kohderyhmän ostokäyttäytymiseen mahdollisesti vaikuttavista trendeistä.

Massadata ja sen hyödyntäminen on herättänyt paljon keskustelua niin tiedeyhteisöissä kuin perinteisessä mediassakin, mutta konkreettisen käsitteen sijasta massadata vaikuttaa toimivan ennemminkin kuvauksena muuttuvan ja digitalisoituvan yhteiskunnan aiheuttamalle datatulvalle, kuin konkreettisena konseptina, jolle voi luoda helposti implementoitavia ratkaisuja. Massadata kuvastaa kaikkea sitä potentiaalia, mikä digitalisaation seurauksena on syntynyt analytiikan saralle sekä sitä estettä, jonka datan siirron, prosessoinnin ja säilytyksen haasteet tuottavat teknologisesti. Data-analytiikka tarjoaa ratkaisuja ja kilpailuetuuksia niin yrityksissä kuin tieteellisissä piireissä, mutta siitä saatava hyöty vaikuttaa sitoutuvan suoraan siihen, kuinka datan rooli organisaatiossa hyväksytään – mitä datakeskeisempiä ja sitä ymmärrettäviä eri päättäjät ja toimijat organisaatiossa on, sitä todennäköisemmin ymmärretään mm. mikä data on toivottujen tulosten kannalta olennaista. Tämä on olennainen osa myös datalinjausten tekemistä, jotka toimivat tukena jokaisessa dataan liittyvässä prosessissa organisaatiossa. Asiakasanalytiikassa erityisesti asiakasymmärryksen kasvattaminen ja sitä kautta asiakkaisiin liittyvien palveluratkaisujen oivaltaminen massadatan avulla vaikuttaa olevan tärkeä ja potentiaalinen osa-alue, samoin kuin syvemmän asiakasymmärryksen hyödyntäminen asiakassegmentoinnissa entistäkin kuvaavampien ja oikeita ostomotivaatioita ja -resursseja kuvaavien segmenttien laatimiseen.

Massadataa käsittelevään tutkimukseen perehtyessä yksi asia nousi selvästi esille – nykyinen tutkimus on suurimmassa osin informaatioteknologian ja tietotekniikan osajilta toisille kirjoitettuja tutkimuksia, jotka keskittyvät massadatan potentiaaliin avaamiseen erilaisten teknillisten datankäsittely -mallien kautta. Tehdyistä tutkimuksista ja artikkeleista kuitenkin puuttuu teos, joka avaisi aiheesta kiinnostuneille, vähemmän teknologista osaamista omaaville ihmisille massadatan hyödyntämiseen liittyviä tärkeitä kysymyksiä: mitä aiheella tarkoitetaan, mihin kaikkiin prosesseihin ja yritysten liiketoimintafunktioihin sen avulla voidaan luoda lisäarvoa, mitkä ovat ensimmäiset askeleet yrityksen sisällä, jotta se voisi kehittyä massadatapohjaisia ratkaisuja hyödyntäväksi yritykseksi, tapahtui se sitten omatoimisesti vaadittavaa osaamista omaavan työvoiman

hankkimisella tai ulkoistamalla palveluiden implementointi massadataa tunteville IT-yrityksille. Kokisinkin ja ehdottaisin, että jo tunnistettujen ja vaadittujen tutkimusten lisäksi, jotka keskittyvät massadatan hyödyntämistä rajoittavien teknologisten ongelmien ratkomiseen, aiheesta tulisi tehdä lisää käytännönläheisempiä, kaupallisia ihmisiä hyödyntäviä tutkimuksia, jotka innostaisivat työntekijöitä ja päättäjiä massadatan hyödyntämistä mieltävän organisaatiossa toimimaan ja miettimään datakeskeisemmin. Massadatan hyödyntämisessä datan roolin väärinymmärtäminen on kuitenkin nostettu useasti suureksi tekijäksi sen käyttämisen esteenä, mistä syystä sen käsittely ja esittäminen vähemmän teknologiapainotteisesta näkökulmasta ehkä mahdollistaisi kokonaisten organisaatioiden asenteiden muutoksen ja käyttöhalukkuuden lisäämisen. Tämän lisäksi aiheesta olisi hyvä tehdä hyötysuhdetutkimuksia, jotka seuraavat massadataa hyödyntävän yrityksen massadatan aiheuttamien yritysmuutosten ajallisia ja rahoituskustannuksia verrattuna muutoksesta seuranneiden hyötyjen arvon kertymiseen tietyn ajan sisällä massadatan ensimmäisestä käyttöönotosta. Näin massadatan potentiaalista puhuttaessa sen toimivuus ei jäisi vain abstraktiksi käsitteeksi sen hyödyistä, vaan eri organisaatioilla olisi mahdollisuutta arvioida massadatan käyttöönoton hyötyjä verrattuna sen vaatimiin resursseihin.

5 EMPIIRINEN TUTKIMUS

Tässä kappaleessa esitellään empiirisen tutkimuksen tavoite, tapaustutkimusyri-
tys ja empiirisen tutkimuksen metodit sekä perustelut niiden valinnalle. Lisäksi
luvussa esitellään, kuinka tutkimuksessa saatu data validoidaan ja kuinka se ana-
lysoidaan.

5.1 Tapaustutkimusyritys ja tutkimuksen tavoite

Jyväskylän Energia on Jyväskylän kaupungin alueella toimiva energiantuotanto-
yhtiö. Yritys perustettiin vuonna 1902, jolloin se toimi Jyväskylän kaupungin
omistuksessa nimellä Sähkölaitos. Yritys aloitti kaukolämmön tuotannon
vuonna 1960 ja sen toiminta muuttui 1980 Energialaitos-nimen myötä liikelai-
tokseksi. Energialaitoksesta tehtiin vuonna 1997 osakeyhtiö, jolloin se sai nykyi-
sen nimensä, Jyväskylän Energia Oy ja vuonna 2006 yhtiö osti omistajaltaan ve-
siliiketoiminnan.

Nykyään Jyväskylän Energia tuottaa, myy ja jakelee sähköä, lämpöä ja vettä
omistamissaan verkoissa yhdessä tytäryhtiöidensä kanssa ja sen sähköä myy-
dään koko Suomen alueelle. Jyväskylän Energia vastaa Jyväskylän kaupungin
keskusta-alueen sähkön siirrosta (jyvaskylänenergia.fi, 2016). Tämän vuoksi yri-
tys tuottaa palveluja automaattisesti tietyn alueen väestölle huolimatta siitä,
ovatko he valinneet yritystä heidän pääasiallisesti sähköntuottajakseen. Lisäksi
yrityksen sähköntuotantoon liittyvät asiakkuudet, jotka ovat valinneet palvelu-
jen, eli sähkön, hankkimisen tapausyritykseltä, jakautuvat kahteen eri lohkoon –
suur- ja pieniasiakkuuksiin. Suurasiakkailla tarkoitetaan nimensä mukaisesti yk-
sityishenkilöistä poikkeavia asiakkaita, kuten yrityksiä.

Teoreettisen tutkimuksen mukaan massadataan on linkitettävissä sille omi-
naisia piirteitä, V-mallin tekijöitä, jotka ilmentävät massadatan olemassa oloa yri-
tyksessä. Sen lisäksi massadatan hyödyntämiseksi asetettiin vaatimuksia datan
prosessoinnin vaiheista ja osa-alueista, joita datan tehokas hyödyntäminen edel-
lyttää. Lisäksi kirjallisuudessa esiteltiin nimettyjä kanavia ja ohjeellisia ehdotuk-
sia kattavan asiakasdatan keräämisen laajuudelle sekä tehokkaita keinoja hyö-
dyntää näitä entistä rikkaamman asiakasymmärryksen saavuttamiseksi ja mer-
kityksellisten asiakassegmentoinnin toteuttamiseksi. Sen vuoksi suoritetussa tut-
kimuksessa selvitettiin tapausyrityksen keräämän, säilyttämän ja hyödyntämän
datan luonnetta ja laajuutta sekä sen hyödyntämiseen liittyviä ongelmia. Tutki-
muksen tavoitteena oli selvittää, miten yritys toteuttaa teorian pohjalta laaditun
datan hyödyntämismalli –viitekehityksessä esitettyjen osa-alueiden prosesseja
sekä sitä, löytyykö tapausyritykseltä näihin osa-alueisiin linkitettyjä massa-
data-ongelmia V-mallin tekijöiden mukaisesti tai eri osa-alueisiin linkitettyjä massa-
datan hyödyntämisen osa- tai kokonaisratkaisuja. Viitekehityksessä esitettyjä osa-

alueita olivat 1. datan keräys ja hankinta, 2. datan varastointi, 3. datan yhdentäminen ja koostaminen, 4. datan mallinnus ja analysointi, 5. datan tulosten tulkinta sekä 6. datan käyttöönotto ja tutkimukseen valitut V-mallin, eli Szymczak ym. (2014) ja Ayankoya ym. (2014) tutkimusten pohjalta yhdistetyn 4V-mallin tekijät olivat 1. massa, 2. monimuotoisuus ja rakenteettomuus, 3. kertymisnopeus ja 4. arvon löytäminen kerrytetystä datasta. Näitä osa-alueita lähdettiin tutkimaan tapaustutkimusyrityksessä datan hyödyntämismalli -viitekehyksen ja valittujen V-mallin tekijöiden pohjalta toteutetun haastattelurungon avulla (liite 1).

Haastattelututkimuksen pääasiallisena tavoitteena oli näiden välitavoitteiden kautta saada selville kaksi asiaa: 1. *onko yrityksellä massadataa* eli onko yrityksen havainnot datan käsittelyyn ja hyödyntämiseen asiakasanalytiikassa verrattavissa teoreettisessa osuudessa tehtyihin havaintoihin sekä 2. *käyttääkö yritys teoriakatsauksessa esiteltyjä massadatan osa- tai kokonaisratkaisuja asiakasdatan hyödyntämiseen*. Haastattelututkimuksen tulosten tavoitteena oli löytää vastaus tutkimuskysymykseen ja vertailla sitä teoriassa saatuun vastaukseen tutkielman tutkimuskysymykseen "Kuinka massadata -pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa" kartoittamalla yrityksen nykyinen datan luonne, datan käsittelyratkaisut ja datan hyödyntämissuhde ja esittelemällä tulokset tavalla, jotka ovat yleistettävissä muihin yrityksiin ja yhteisöihin, jotka kohtaavat toiminnassaan tapausyrityksen kaltaisia ongelmia ja miettivät asiakasanalytiikkansa tehostamista sekä sen mahdollistavia työkaluja.

5.2 Tutkimusmenetelmä ja tutkimuksen toteutus

Seuraavissa kappaleissa esitellään ja perustellaan valitun tutkimusmenetelmän ja tutkimuksen haastateltavien valitseminen sekä kerrotaan, millä tavalla tutkimus toteutettiin käytännössä.

5.2.1 Kvalitatiivinen tutkimus

Empiirinen tutkimus toteutettiin kvalitatiivisena tutkimuksena ja sen toteutustavaksi valittiin haastattelututkimus, tarkennettuna *teemahaastattelu*. *Kvalitatiivinen*, eli laadullinen tutkimus soveltuu parhaiten tutkimuksiin, jossa halutaan kerätä tietoa mittaamattomista muuttujista, kuten asian kanssa tekemisissä olevilta osapuolilta erilaisia mielipiteitä ja havainnoiteja kerätessä. Koska tutkimuksen tavoitteena on kerätä tietoa tapaustutkimusyrityksen asiakasdatan kanssa tekemisissä olevilta työntekijöiltä tietoa datan käsittelyn eri osavaiheista ja sen hyödyntämisestä tapausyrityksessä sekä kerätä tietoa massadataan teoriassa linkitettyjen ongelmien esiintymisestä yrityksessä, kvalitatiivinen tutkimusmetodi oli tutkimuksen kannalta olennaisten tulosten saamisen kannalta optimaalisin väline.

Haastattelututkimuksessa tutkija ja haastateltava keskustelevat tutkimusaiheisiin liittyvistä asioista. Perinteisestä keskustelusta poiketen haastattelulla on selkeä päämäärä, johon rakenteellinen haastattelutilaisuus tähtää, eli tutkimuskysymykseen vastaaminen. Haastattelun avulla kerrytetään tutkimusaineisto,

joka analysoinnin ja tulkinnan avulla käytetään tutkimustehtävän ratkaisemiseksi. (Hirsjärvi & Hurme 2000) Haastattelututkimuksen toteutustavaksi valittiin *teemahaastattelu*, joka on uutta kartoittava tutkimusmetodi, jossa kysymykset ovat avoimia. Tämän lisäksi teemahaastattelussa on mahdollista esittää lisäkysymyksiä, arvaamattomat vastaukset tallennetaan ja analysoidaan, eikä vastaajilla ole mahdollisuutta vastata kysymyksiin anonyymisti. Vaikka tässä haastattelun toteutustavassa työmäärä vastauksia kohti on suuri ja sen määrällinen analyysi on harvoin toteutettavissa, se antaa mahdollisuuden esittää kysymyksiä laajemmalla tasolla kuin tiukasti kyselypohjaan sidottujen haastattelujen toteuttaminen ja mahdollistaa siten kattavamman vastauspohjan kartuttamisen erilaiset käsityspohjan omaavilta haastateltavilta. (Hirsjärvi & Hurme, 2000.) Teemahaastattelu soveltuu tutkimusongelman kannalta olennaisen aineiston keräämiseen erityisesti siksi, koska tapausyrityksen haastateltavilla voi olla varsin erilaisia kokemuksia datan keräämisestä, käsittelemisestä ja hyödyntämisestä toisiinsa nähden sekä osa heistä saattaa tiedostamattaan tietää asioita ja ongelmia näiltä eri osa-alueilta, jotka saadaan esille vain vapaamman, teemoihin pohjautettujen kysymyksien ja lisäkysymyksien avulla. Lisäksi kun vastaajilla ei ole mahdollisuutta osallistua vastaustilanteeseen anonyymisti, voidaan varmistaa, että vain tutkimuksen kannalta olennaisien henkilöiden vastaukset vaikuttavat tehtäviin havaintoihin, analyysiin ja loppupäätelmiin. Kriteerejä tutkimushaastatteluun valituille tutkimuskohteille käsitellään tarkemmin seuraavassa luvussa.

5.2.2 Haastateltavien valitseminen

Koska tutkimuksen pääpaino on asiakasanalytiikassa ja massadata-analytiikan hyödyntämisessä siinä, haastateltaviksi valittiin tapausyrityksen työntekijöitä, jotka ovat mukana yrityksen asiakasdatan käsittelyn eri vaiheissa. Haastateltavien valitsemisessa hyödynnettiin toimeksiantajayrityksen ymmärrystä omien osastojensa roolissa asiakasdatan käsittelyn eri vaiheissa ja haastateltavia pyrittiin löytämään vähintään yksi (1) kappale jokaiselta eri osa-alueelta, joka kerää, varastoi, koostaa, mallintaa, analysoi, tulkitsee tai hyödyntää datan lopputuloksia työssään. Tällä tavoin haastateltavilta saatiin kattavammin selville yrityksen asiakasdatan tila sekä datasta aiheutuvien haasteiden ja niiden selvittämiseksi käytettyjen ratkaisujen esiintyminen yrityksessä.

Tapausyrityksen yhteyshenkilöiden kanssa käydyissä keskusteluista yrityksen rakenteesta ja datan linkaaresta yrityksen sisällä selvisi seuraavia asioita, joita hyödynnettiin haastateltavien henkilöiden määrittämisessä: yrityksessä on kolme (3) organisaatiota, jotka vastaavat asiakkaan osalta tehtävästä tiedonkeruusta. Näitä ovat uuden asiakkaan vastaanotto, mittaus, eli sensoridataorganisaatio sekä asiakkuuden hallinta, eli laskutus. Haastateltavien valintakriteereiksi muodostui siis ensisijaisesti näistä asiakkaan tiedonkeruuseen välittömästi linkittyvistä organisaatioiden edustajista haastateltavien hankkiminen, jonka lisäksi haastattelukohteita tuli saada vähintään yksi jokaisesta datan hankinnan ulkopuolisesta osastosta, jotta datan linkaaren kaikki eri osa-alueet yrityksen sisällä saataisiin katettua. Tämän vuoksi asiakasdatan hankinnan eri organisaatioiden

lisäksi haastatteluun sopiviksi osa-alueiksi yrityksen sisällä luokiteltiin mm. *markkinointi, talous ja tietohallinto*.

Haastattelututkimus suoritettiin 13 yrityksen edustajalle, joiden sijoittuminen tapaustutkimusyrityksen eri osastoihin jakaantui seuraavasti: tietohallinto, yksi (1) haastateltava, asiakkuudet, kolme (3) haastateltavaa, tuotehallinta, kolme haastateltavaa, verkot, kaksi haastateltavaa, markkinointi, yksi haastateltava ja talous, kolme haastateltavaa. Tutkittavien työskentelyhistoria tapaustutkimusyrityksessä vaihteli suuresti: haastateltavien työskentelyaikayrityksen alaisuudessa vaihteli <1-30 vuoden välillä, haastateltavien työskentelyajan yrityksessä keskiarvon ollessa 8 vuotta ja sen mediaanin ollessa 6 vuotta. Lisäksi kahdella haastateltavalla yrityksen sisäinen toimialue oli muuttunut merkittävästi alle vuoden sisällä haastattelun toteuttamisesta. Haastateltavien haastattelutulosille merkittävät taustatiedot on esitetty tarkemmin alla esiintyvässä taulukossa.

TAULUKKO 2 Haastateltavien taustatiedot

Toimialue	Aika yrityksessä	Viimeisin merkittävä toimenkuvan vaihdos
Asiakkuudet	1 vuosi	3 kuukautta
Asiakkuudet	9 vuotta	
Asiakkuudet	2 vuotta	
Markkinointi	8 vuotta	
Talous	7 vuotta	
Talous	6 vuotta	
Talous	3 vuotta	
Tietohallinto	30 vuotta	
Tuotehallinta	1 vuosi	
Tuotehallinta	4 vuotta	
Tuotehallinta	7 vuotta	< 1 vuosi
Verkot	6 vuotta	
Verkot	23 vuotta	3 vuotta

5.2.3 Tutkimuksen toteutus

Tutkimus toteutettiin yrityksen toimitiloissa yksilöhaastatteluina, eli yksittäiseen haastattelutilaisuuteen osallistui vuorollaan vain haastatteliija ja haastateltava. Koska tutkimus toteutettiin temahaastatteluna, tarkan kysymysrunгон sijaan haastattelua varten oli laadittu teoriaosuudessa laaditun datan hyödyntäminen -viitekehysten pohjalta kuudesta kysymyksestä muodostuva kysymyspohja, joista jokainen kysymys käsitteli yhtä erillistä osa-aluetta viitekehyksessä esitetyssä prosessissa. Tämän lisäksi haastattelussa kysyttiin jokaiselta haastateltavasta heidän kokemuksiaan massadatassa esiintyviin ongelmiin liitettyjen tekijöiden osalta, eli datan massasta, rakenteettomuudesta, kertymisnopeudesta sekä olennaisen arvon löytämisestä kerätystä datasta. Jokaiselta haastateltavalta kysyttiin

lisäksi heidän massadataan ja sen käsittelyyn liittyvän terminologian tunnistamista, jotta saatiin varmuus, että myös tiedostamattomat yhteydet massadataan liittyviin osa-ratkaisuihin saatiin selville. Haastattelun päätteeksi haastateltavia pyydettiin kuvaamaan massadataa heidän käsityksensä pohjalta, jotta tapausyrityksen käsitys tutkittavasta aiheesta sille olennaisien osastojen sisällä tulisi ilmi.

Jokaiselle haastattelulle oli varattu aikaa keskimäärin 45 minuuttia, jouston ollessa 15 minuuttia molempiin suuntiin. Tutkimuksen tietojen kerääminen toteutettiin tallentamalla tehdyt haastattelut kahteen eri nauhoituslaitteeseen, tietokoneeseen ja puhelimeen, jonka lisäksi haastatteluista tehtiin samalla kirjallisia muistiinpanoja, jonne lisättiin havaintoja haastateltavista. Erillisten muistiinpanojen avulla saatiin varmuus, että äänitteissä esiintymättömät eleet, ilmeet sekä epäilykset vastauksista ja aiheista saatiin myös osaksi kerättyä materiaalia.

5.3 Aineiston analyysi

Haastattelun tulokset analysoidaan teemoittelun avulla ja teemoina toimivat massadataa määrittelevien tekijöiden lisäksi näistä annetut havainnot tukemaan oletusta massadatan olemassaolosta yrityksellä sekä tukemaan oletusta siitä, ettei yrityksellä ole massadataa. Teemoittelu sopii haastatteluun analysointimetodiksi haastattelijoiden pienen määrän vuoksi ja haastattelussa selkeästi esiintyneiden teemojen vuoksi sekä siksi, että teemahaastattelun joustavuus, eli toteuttaminen ilman tarkkaa kysymyslomaketta, saataisiin parhaiten tuotua esille myös havainnoinnissa.

Hirsijärvin ja Hurmen (2000) mukaan haastatteluaineiston analyysin kuuluu neljä vaihetta: aineiston luenta, luokittelu, yhteyksien löytämien ja raportointi (Hirsjärvi & Hurme, 2000). Tätä määritelmää mukailleen aineiston analyysi aloitettiin muuttamalla nauhoitetut haastattelut tekstiksi, eli litteroimalla haastattelun äänitteet tekstimuotoon. Jokainen haastattelu muutettiin yksittäiseksi tekstitiedostoksi, jotka koostuivat haastattelun kysymyksistä ja vastauksista, eli haastattelun sisällöstä kokonaisuudessaan. Litteroinnin laadun varmistamiseksi jokainen haastattelu käytiin läpi vähintään kahdesti, jotta äänitteiden ja niiden tekstiversioiden sisällöt saatiin vastaamaan toisiaan täydellisesti. Haastattelujen luokittelu suoritettiin QDA Miner Lite-ohjelmalla. Luokittelussa tutkittua ilmiötä pyritään jäsentämään vertailemalla aineiston eri osia toisiinsa (Hirsjärvi & Hurme, 2000). Luokittelussa hyödynnettiin ohjelman koodauksien luontia, joilla haastateltavien vastauksissa esiintyviä teemoja merkittiin eri osa-alueille datan hyödyntämisen kokonaisprosessissa. Kaikki haastattelut käsiteltiin merkiten ohjelmalla vastauksissa esiintyviä tekijöitä liittyen kolmeen aihealueeseen: 1. yrityksessä esiintyviin V-mallin tekijöihin, 2. datan hyödyntämismallin osa-alueiden toteutumiseen yrityksessä sekä 3. tapausyrityksen keräämään ja hyödyntämään asiakasdataan. Koodit luotiin hyödyntäen kirjallisuuskatsauksessa havaittuja tekijöitä, luoden kaikille havaituille tekijöille oma koodi. Näillä koodeilla voitiin merkitä yksittäisiä lauseita, joissa eri teemoja ilmeni. Esimerkiksi automatisoitu datan mallinnus ja analyysi sekä manuaalisesti suoritettu datan mallinnus

ja analyysi toimivat yksittäisinä koodeina. Näin ohjelman avulla saatiin merkityä esimerkiksi seuraavan tyyppinen lause datan mallinnuksesta ja analyysistä yrityksessä, joka tehdään manuaalisesti:

”Ihminen lähtee, ei siinä sen enempää, siis joo pystytään [järjestelmän] puitteissa suodattamaan sitä Excelissä eri lailla ja näin, mutta ei siellä oo mitään automatiikkaa siellä takana, jos ei ihmistä lasketa automatiikaksi”.

Lisäksi tämä mahdollisti tutkimuksen haastattelurungon ulkopuolelle jääneiden tekijöiden havainnoinnin. Näistä saatiin luotua uusia teemoja, joita ei kirjallisuudessa noussut esille. Haastatteluissa tämä tuli ilmi esimerkiksi haastatteluvien huolena datan validiteetista, oikeellisuudesta ja eheydestä:

”.. se on se jatkuva riesa siinä, että ennekuin voi tehdä raporttia ja viedä sitä tietoa eteenpäin esimiehelle tai konsernin johtoryhmälle, niin täytyy aina varmistua siitä, että on kaikki tallessa ja verrata sitä reaalipintaa aiemmin tai viimekuussa, millaisia muutoksia siellä on käynyt, että sinänsä työläs ja aiheuttaa turhaa työtä...”

Kaikki eri teemoilla merkityt koodit, kuten esimerkiksi ”datan massa” koostettiin yksittäisiin tiedostoihin, jotka sisältävät vain kyseisellä koodilla merkittuja lauseita ja toimivat siten eri teemojen esiintymien yhdistelminä. Yhdistelyn tavoitteena on löytää teemojen ilmentymien väliltä samankaltaisuutta, vaihtelua ja poikkeuksia (Hirsjärvi & Hurme, 2000). Vastauksista yhdistetyistä teemojen koosteista etsittiin yhtäläisyyksiä, esimerkiksi haastattelujen kokemuksista datan kertymisnopeuden aiheuttamista ongelmista tapaus yrityksessä, kuten:

”Varmasti sitä on, että sitä kertyvää dataa ei ehditä käymään läpi, koska sitä tulee niin paljon, että se on vaan järkevä analysointi, että siihen ei oo riittävästi resursseja että varmaan aika iso osakin siitä kertyvästä datasta menee vähän niinku hukkaan tai silleen vähän niinku turhaan kerätään ja säilytetään kun sitä ei ehditä hyödyntää, niin sitä tapahtuu kyllä”.

Näitä vertailtiin kirjallisuudessa havaittuihin ilmentymiin datan kertymisnopeuden (*velocity*) aiheuttamista ongelmista, jotta voitiin havainnoida, miten V-mallin kertymisnopeutta esiintyi tapausyrityksessä. Esimerkiksi V-tekijöiden esiintymisiä yrityksessä jaoteltiin teemoittain ongelmia aiheuttaviin ja ongelmia aiheuttamattomiin havaintoihin. Datan hyödyntämismallin eri osa-alueisiin liittyvistä tekijöistä teemoitettiin vastaavuudet ja poikkeavuudet sekä puutokset yrityksessä. Asiakasdatan keräämisessä vastaukset teemoitettiin kirjallisuudessa ehdotettujen osa-alueiden ympärille, jonka jälkeen teeman sisältä luokiteltiin vastaavat ja poikkeavat havainnot. Näiden eri osa-alueiden teemoituksen ja teemojen jaottelun jälkeen saatiin kattavampi käsitys yrityksen V-mallin tekijöiden ilmentymistä, yrityksen datan hyödyntämisestä sekä asiakasdatan keräämisestä. Tämän analysoinnin pohjalta saatiin koottua kirjallinen raportti asiakasdata-analytiikan toteutumisesta ja sen hyödyntämiseen liittyvistä ongelmista yrityksessä. Nämä havainnot esitellään tutkimuksen tuloksina seuraavassa luvussa.

6 TUTKIMUKSEN TULOKSET

Tässä kappaleessa esitellään toteutetun, empiirisessä tutkimusosuudessa suoritettun haastattelututkimuksen havaintoja kolmeen eri teemaan jaoteltuna. Tapaustutkimuksen tarkoituksena oli saada tietoa tutkimuskohteena toimineen tapausyrityksen asiakasdatan käytöstä, eli sen elinkaaresta yrityksessä keräämisestä sen käyttöön sekä selvittää, onko yrityksellä massadatan määritelmiä vastaavia ongelmia ja käyttävätkö he joitain massadatan osa- tai kokonaisratkaisuja datan hyödyntämisessä. Tapaustutkimuksen suorittamiseen hyödynnettiin teoreettisessa osuudessa esitettyä V-mallia, joka koostui massadatan hyödyntämiseen ja käsittelemiseen liittyvistä ongelmista, kuten datan massasta (eng. *volume*) ja monimuotoisuudesta ja rakenteettomuudesta (eng. *variety*). Tämän lisäksi teemahaastattelun teemat koottiin kappaleen lopussa esitetyn Datan hyödyntämismalli -viitekehityksen ympärille, keskittyen sen datan prosessoinnin eri osaluokkiin ja niiden aiheuttamiin ongelmiin datan kokonaisvaltaisessa käsittelyssä ja hyödyntämisessä. Koska massadataa tutkitaan tutkielmassa asiakasdatan näkökulmasta, haastattelututkimuksessa huomattava paino oli myös tapausyrityksen keräämässä asiakasdatassa ja sen hyödyntämisessä. Seuraavissa alaluvuissa esitellään teemahaastattelututkimuksessa tehdyt havainnot V-mallin tekijöiden esiintymisestä yrityksessä, datan hyödyntämismallin toteutumisesta yrityksessä ja yrityksen keräämästä asiakasdatasta sekä kappaleen yhteenveto.

6.1 V-mallin tekijöiden esiintyminen yrityksessä

Tässä kappaleessa käsitellään haastattelusta tehtyjä havaintoja yrityksessä esiintyviin V-mallin tekijöihin liittyen, joista haastattelututkimukseen tarkasteltaviksi valittiin massa (*Volume*), rakenteettomuus ja monimuotoisuus (*Variety*), kertymisnopeus (*Velocity*) ja arvon löytäminen datasta (*Veracity*). Näitä tekijöitä ja niihin liittyvät ongelmat on usein liitetty massadatan esiintymiseen, joten selvittämällä tekijöiden esiintymistä yrityksessä voidaan saada käsitys massadatan esiintymisestä yrityksessä, jotka on jaoteltu kolmeen pääteemaan: 1. Massadatan esiintymistä yrityksessä indikoivat havaintoinnot, 2. Yrityksen dataprosessin eri vaiheet viitekehitykseen verrattuna 3. Kirjallisuudessa hyväksi havaittujen asiakasdatojen hyödyntäminen.

6.1.1 Datan massa

Datan volyymin käsiteltäessä kaikilta haastateltavilta kysyttiin seuraava kysymys:

”Onko teillä esiintynyt ongelmia datan volyymin esimerkiksi sen varastoinnissa ja käsittelyssä?”

Haastateltavista neljä (4) vastasi datan määrästä esiintyvän ongelmia ja yhdeksän (9) vastasi, ettei määrästä ole ainakaan heille itselleen ongelmia. Haastattelussa aiheesta nousi esille kaksi pääasiallista ongelmaa, jotka ovat suoraan seurausta yrityksen keräämän datan massasta: *prosessien hidastuminen* ja *varastointikapasiteetin riittämättömyys*. Haastattelussa kerrottiin havaintoja organisaation koneiden hidastumisesta erityisen suuria datamassoja käsitellessä. Havainnot keskittyivät pääasiallisesti osittaiseen järjestelmien hidastumiseen poikkeustilanteissa:

”Niin että määrä ois niin suuri, että se ois aiheuttanut ongelmia? (Mm-m). No on. [naurua]. On, yleensä se menee sillä tavalla, että joku järjestelmä alkaa hidastua ja se käyttökelpoisuus huononee ja kyllähän nyt ylipäänsä niinku jatkuvasti investoidaan lisää siihen datan säilyttämiseen että sehän kasvaa valtavaa vauhtia se data koko vauhtia mitä säilytetään ja kokoajan pitää laittaa lisää rahaa ja palvelimia että se saadaan säilytettyä, että on se tavallaan jatkuva ongelma”.

”Ei. Sen muistan, että muutaman kerran ollaan jouduttu tehoja lyömään koneisiin lisää sen datamassan suhteen. Että se on selvästi hidastanut joidenkin, esim. [yrityksen käyttämässä järjestelmässä] olevien laskentojen nopeutta, että on pitänyt tehoja nostaa palvelimissa tai muuta vastaavaa, ei oo muuten, tullut vastaan.”

”[...] on se tuolla tietohallinnossa ainakin näkynyt silleen, että vähän väliä joku itkee, että saadaanko lisää tehoja laitteisiin, että ollaan jouduttu sitten laitteisiin resursseja laittamaan lisää, että kyllä se aina pienen tuskan aiheuttaa sellainen ponnistus, että saadaan stabiiliksi se käyttötilanne. Aina välillä, muutaman kerran [yrityksen käyttämän järjestelmän] käytön aikana tullut vastaan, että resurssit on loppunut ja nykyin ollaan tehty elinkaarisuunnitelmia sellaiselle, että jatkossa se homma kehittyy.”

Suuremmista, yrityksen laajuisesta koneiden prosessointitehon riittämättömyydestä esimerkiksi toistuvien toimintakatkoksien muodossa ei yrityksen edustajilla ollut kuitenkaan havaintoja. Sen sijaan osa prosessointitehon riittämättömyyteen liittyvistä havainnoista sisälsi ongelmia, kuten ihmisresurssit osana järjestelmätehojen tasapainottamista:

”On. Käytännössä järjestelmät ja yhteydet ja datan käsittelijät eli resurssit, eli ne, jotka käsittelee niitä, niin jatkuvasti painitaan niiden kanssa että pitää miettiä miten saadaan kaikki ne toimimaan. Palvelimet, tiedon siirto, kaikki. Tää on ihan jokapäiväistä.”

Tämän lisäksi datan massa aiheutti ongelmia puolilla, jossa käsiteltiin tekstipohjaista dataa johtuen siitä, että sen käsittely tehtiin manuaalisesti:

”No siis, ei nyt silleen [ole ongelmia datan volyymin kanssa dataa varastoidessa tai käsitellessä], mutta sen takanahan on se, ettei ihminen pysty niinku määrättömästi sitä käsittelemään, että tavallaan sen verran, kun pystytään sitä vastaanottamaan niin sitä [dataa] on, että jos se ois jotenkin automatisoitu tai jalostettu niin sittenhän se ois ihan eri juttu.”

Kuitenkin suuri osa haastateltavista koki, ettei heillä ole ongelmia kerryttämänsä datan volyymin, eli määrän kanssa. Haastateltavista huomattava osa vastasi, että jollain muulla osastolla saattaisi olla ongelmia, mutta että heidän omalla osastollaan ei esiintynyt ongelmia datan määrän suhteen. Ei ongelmia -vastasten joukossa oli myös suoraa vastauksia ongelmattomuuteen viitaten.

”Ja no sitten taas siinä automaatiojärjestelmän puolella, siellä ei olla ehkä niin paljoa sitä isoa dataa pyöritetty, mutta vastaavat ongelmathan siellä tulee, jos se laitetaan monia muuttujia pyörittään, että tuossa [yrityksen käyttämässä järjestelmässä] puhutaan, että jos otetaan tuntisarja, niin se on 8760 saraketta ja sitten jos meillä on tuota, vaikka se meidän kaikki käyttöpaikat joku 5000 niin se on jo aika iso määrä. Et siellä automaatiopuolella se on tietysti katotaan niitä muutamia muuttujia, niin en osaa sanoa tarkemmin onko siellä ollut sitten ongelmia”.

”No ei oikeestaan. Tai kyllä mä koen, että ne ongelmat on enemmänkin siellä se saadun datan hyödyntämispuolella ja jatkojalostuksessa.”

Datan massaan liittyvät vastaukset keskittyivätkin pääasiallisesti hetkellisten poikkeustapausten, muiden osastojen kerryttämään dataan liittyvien mahdollisten ongelmien ja prosessoinnin manuaalisuuteen liittyvään mielikuvaan datan määrän paljoudesta.

6.1.2 Datan monimuotoisuus ja rakenteettomuus

Datan rakenteettomuuteen liittyvät havainnot jakautuivat datan monimuotoisuuteen ja datan yhteensopimattomuuteen eri järjestelmien välillä. Datan monimuotoisuutta ja rakenteettomuutta selvitettiin kysymällä haastateltavilta:

”Onko teillä esiintynyt ongelmia datan monimuotoisuuden ja rakenteettomuuden vuoksi?”

Datan rakenteettomuudesta selkeitä havaintoja oli huomattavasti vähemmän: kolme (3) vastasi sen aiheuttavan ongelmia, kun taas loput kymmenen (10) haastateltavista vastasi, ettei monimuotoisuudesta ja rakenteettomuudesta ole heille ongelmia. Datan monimuotoisuudesta ongelmia esiintyi pääasiassa tekstimuotoista asiakasdataa käsittelevien markkinoinnin osa-alueen edustajien havainnoissa.

”On. Mä oon ainakin marmattanut siitä pitkän aikaa, että en mä esimerkiksi suoraan sieltä meidän [yrityksen käyttämästä tietojärjestelmästä] pysty sitä dataa hyödyntämään millään tavalla.”

Sen sijaan osa rakenteettomuuteen liittyneistä havainnoista linkittyi suoraan organisaation käyttämien järjestelmien välisiin rajapintoihin sekä niissä tapahtuvaan datan ”rikkoutumiseen”, jossa osa datasta ei siirry kunnolla järjestelmästä toiseen. Haastateltavien keskuudessa esiintyi suurta huolta datan validiteetin laskemisesta järjestelmien välisen datan siirtymisen ja pirstoutumisen seurauksena ja viidessä (5) suoritetuista haastatteluista nousi esiin selkeää huolta datan luotettavuudesta:

”[Tietystä yrityksen käyttämästä järjestelmästä] tulee vähän kaikennäköistä dataa joutuun tietysti siitä mitä siellä on siellä järjestelmässä, miten sinne on sitä tietoa viety, miten sitä on syötetty eri kenttiin. On varmasti tehty paljon virheitä ja sitten on tietysti se rajapinta, mistä sitä tietoa tulee eri järjestelmiin [...] Se taas sitten johtuu kolmanesta osapuolesta, osittain siitä rajapinnasta ja sen madonrei’istä ja ongelmista. Paljon

on ollut ongelmia siinä. Vaikka se periaatteessa se data on ihan selkeätä, mitä sieltä pitäis tulla mutta sieltä aina puuttuu jotain. Kyllä, sitä ihmetellään, missä on vika, mistä se löytyy se data vai löytyykö sitä, minne se on hävinnyt.”

”Se on rakennettu ehkä vähän huonosti se ja se tieto ei oo aina luotettavaa. Että pitää tietää tavallaan mitä tekee”.

Suurin osa vastaajista koki, ettei heillä ole ongelmia rakenteettomuuden ja monimuotoisuuden kanssa, sillä heidän keräämänsä data on lähes täysin rakenteellista ja rakenteettoman datan kerääminen on pääasiallisesti poikkeus heidän yleisessä datapolitiikassaan.

”No ei, ei meillä käytännössä ihan semmoista [ongelmaa ole], koska ne on aika tarkkaan mietitty ne protokollat, miten ne tiedot siirtyy siellä, niitä on niin valtavat määrät [ettei] siinä ole oikein varaa siihen että siinä tulisi tämmöisiä [ongelmia]”.

”Joo ei meillä varmaan tuo oo ollut kyllä ongelma, että tietysti on kiinnitetty jo alkuvaiheessa huomiota, että se on semmoisessa muodossa, että sitä pystyy sitten käsittelemään. Muotojahan on monia, et se tehdään sitten siellä järjestelmän sisällä se”.

”No ei, että se on nimenomaan sturkturoitua dataa mitä nyt ollaan kerätty ja jalostettu, että ei silleen ole ollut ongelmia”.

Haastatteluiden perusteella yrityksessä ei siis ainakaan pääasiallisesti kerätä tai hyödynnetä rakenteetonta ja monimuotoista dataa.

6.1.3 Datan kertymisnopeus

Kertymisnopeuteen liittyviä, massadatan olemassa olevia tukevia havaintoja oli organisaatiossa selvästi vähiten. Syyksi tähän haastattelussa annettiin mm. kaikista hektisimpien dataa keräävien järjestelmien syklit, joiden perusrakenteessa on jo varauduttu datan nopeaan kertymiseen yrityksen järjestelmissä. Haastateltavista yksi (1) kertoi kertymisnopeudesta seuraavista ongelmista, kun taas 9 vastasi ”ei” ja kolme vastasi, etteivät osaa sanoa. Havainnoissa nousi kuitenkin esille muutamia huolia datan nopeasta kertymisestä sekä kertymisestä tiedostettujen, erityisen nopeasti ja paljon dataa kertyvien järjestelmien ulkopuolella havainnoitavaksi ja tulkittavaksi. Nämä havainnot liittyivät aiemman, massaa käsittelevän kappaleen havainnon tavoin ihmispohjaisen työn aiheuttamaan mielikuvaan datan luonteesta:

”No ehkä se varmaan, jos mennään tämmöisen mediaseurannan tai noiden julkaisujen seurannalla ja uutistenseuranta, niin sitähan tulee aivan valtavasti, että siinä ehkä on just se, että mitkä on ne oikeat jutut mitä seurata, että sitähan voi netissä pyöriä vaikka koko ajan ja etsiä uusia artikkeleita, mutta sitten että niitä löytyy ihan hirveästi, niin sitten mikä on se todellinen hyöty, että missä vaiheessa siihen on perehdytty tarpeeksi... ehkä se siellä puolen se ähky on mahdollista saada”.

Kertymisnopeudesta haastateltavilla oli pääasiallisesti havaintoja siitä, ettei kertymisnopeus aiheuta heidän organisaatiossaan ongelmia. Nopeinten yrityksessä dataa ympärivuorokautisesti kerryttävät järjestelmät ovat poikkeuksellisia

yrittäjien yleiseen datan keräämiseen verrattuna ja kyseiset järjestelmät on syklitetty käsittelemään dataa säännöllisinä purskauksina:

”No ei silleen voi sanoa että tuohon kysymykseen ei voi vastata muuta kuin että ei”.

”No ei että kyllä meillä ihan fiksit prosessit siellä on, että käytännössä se tapahtuu päivittäin, että tuntipohjainen mittausdata jossa ne suurimmat datat on päivittyy päivittäin tietovarastoon, ettei siinä päivän aikana ehdi niin hirveää määrää tulla. Suurimmat ongelmat liittyy aina siihen, että kun ei ole aina oikein validia se data, että joudutaan kuitenkin joskus jopa useita päiviä historiaa korjailemaan, että se on se suurin ongelma tuossa, että sehän ei sinällään liity varastointiin tuossa, että täytyy myöskin sinne varastoon korjata sitten se data validiksi, kun se sinne operatiivisiin järjestelmissä on korjattu”.

Myös datan kertymisnopeudesta keskusteltaessa datan luotettavuus siirtyessä järjestelmästä toiseen nousi esille haastateltavia huolestuttavana osa-alueena.

6.1.4 Datan Arvo

Haastattelussa datan arvoa ja sen löytämiseen mahdollisesti liittyviä ongelmia lähdettiin selvittämään mm. seuraavan kysymyksen avulla:

”Onko teillä esiintynyt ongelmia olennaisen arvon löytämisessä keräämästänne ja käsittelemästä datasta?”

Kaikista massadataa indikoitavista teema-alueista eniten havaintoja saatiin tältä osa-alueelta. Arvokkaan datan löytämiseen liittyvistä ongelmista havaintoja oli yli puolella haastateltavista, seitsemällä (7) vastaajalla, neljä (4) ei osannut sanoa ja vain kaksi (2) vastasi, ettei arvon löytäminen kerätystä datasta aiheuta heille ongelmia. Haastateltavien kokema vakavuus tämän osa-alueen osalta vaihteli suuresti haastateltavien kesken:

”On. Sitä on jouduttu paljon miettiä, mitä on ne asiat jotka sieltä pitää niinku löytää, että päästään siihen toivottuun lopputulokseen. Se on jatkuvaa mietintää, että löydetään... Jos tiedetään, että jotakin ongelmia on jossain kohtaa, niin meidän pitää löytää se, että millä tavalla sieltä tiedosta saadaan ulos se, mitä me haetaan..”

”No silleen varmaan on ongelmaa, että meillä ei oo tehty semmoista tarkkaa suunnitelmaa vaan siitä että mitä asiakastietoa me lähdettäis keräämään, että kuten sanoin niitä puheluita tulee kuitenkin paljon ja me ollaan paljon asiakaskantaan yhteyksissä ja tekemisissä ja muuta niin, jos meillä ois joku selkeä suunnitelma että jotakin tietoa lähdetään rikastamaan tai säännönmukaisesti asiakkaalta keräämään niin meillä ois se mahdollisuus, mutta tällaista suunnitelmaa ei olla vielä tehty. Sitä suunnitelmaa ei varmaan sen takia olla tehty, kun ei oo sitten, niinku ensimmäisenä pitäis varmaan selvittää, että mitä me sillä tiedolla sitten tehdään. Mihin me tarvitaan jotain tietoa, että miksi sitä pitäisi kerätä. Semmoista ei olla koskaan tehty.”

”Siis kaivaa se olennainen tieto? No toki, sehän on aina se ongelma, kun on paljon dataa, että mikä on olennainen tieto ja miten sitä käsitellään.”

”No on siis mulla ainakin sillä lailla, että jos mä etsin jotain markkinointiin liittyviä trendejä ja niistä tulee semmoinen tietty kooste, niin sen pureskeleminen ja jakaminen

koskettamaan meitä, että sieltä löytys jotain, niin siihen ei ole mitään oikotietä, se [etsiminen ja käsitteleminen] on tämmöstä 'kotikutosta'."

"Sitä [kehittämistä] pitäis varmaan tehdä, että miten saada se olennainen tieto oikeasti, koska sitä tietoa on jo nyt, sitä on paljon, nyt se on helposti saatavilla, nyt siitä kehitetään niitä raportteja, niin tässä vaiheessa pitäis löytää se olennainen [data] just millä on oikeasti jotain merkitystä."

Haastattelun vastaajat, jotka kokivat, että arvon löytäminen kerätystä datasta aiheutti heille ongelmia, kokivat, ettei heidän osastollaan ollut selkeitä datalinjauksia. Nämä haastateltavat kokivat, että arvon löytämiseen ja tiedon rikastamisen kannalta tehokkaimpien algoritmien eteen oli käytettävä ns. "yritys ja erehtyminen" -taktiikkaa tarkempien linjauksien puuttuessa.

Vain harva kyselyyn vastanneista koki, ettei heille aiheudu arvokkaan datan löytämisestä ongelmia. Haastateltavat, jotka kokivat vahvasti, ettei heillä ole ongelmia löytää arvokasta dataa heidän kerryttämänsä datan joukosta, asettivat tästä kiitoksen mm. heidän tiiminsä sisäisille datalinjauksille, joita katsastettiin säännöllisesti maksimaalisen suoritusnopeuden saavuttamiseksi. Vastauksien vahvuudet arvon löytämisen ongelmallisuuteen liittyen kuitenkin vaihtelivat myös vastaajien kesken ja osa vastaajista koki, ettei heillä ole tullut vatsaan erityisiä ongelmia henkilökohtaisesti asian osalta, mutta korostivat, etteivät osanneet puhua koko osastonsa puolesta vastatessaan.

6.2 Datan hyödyntäminen -viitekehityksen toteutuminen yrityksessä

Tässä kappaleessa esitetään havaintoja tapausyrityksestä siitä, kuinka teoriaosuudessa laaditun datan hyödyntämisen viitekehityksen osa-alueet toteutuvat yrityksen asiakasdataa keräävillä sekä asiakasanalytiikkaa käsittelevillä osastoilla. Neljännen kappaleen lopussa luodun datan keräämisen ja hyödyntämisen viitekehityksen eri vaiheet olivat 1. datan kerääminen, 2. datan varastointi, 3. datan yhtenäistäminen ja koonti, 4. datan mallinnus ja analyysi, 5. datan tulosten tulkinta ja 6. datan käyttöönotto, jotka kaikki linkittyvät kohdan 7. datalinjausten sisään, käsitellen linjaukset liittyen kaikkiin näihin datan käsittelyn eri osa-alueisiin.

6.2.1 Datan kerääminen

Pääasiallisesti yritys hyödynsi asiakasanalytiikassaan itse keräämäänsä dataa asiakkaisiinsa liittyen. Dataa kerättiin sekä relaatiotietokannoissa käsiteltävässä muodossa sekä monimuotoisessa ja rakenteettomassa muodossa. Sisäisesti kerrytetystä datasta suurin osa datan keräämisestä liittyi asiakkaan kulutukseen liittyvään dataan:

"[...] mittaus tieto, lähinnä sähkö ja kaukolämpö tulee tuntidatana, sähköstä kerätään lämpötilatietoja, sähkökatko-logia ja energiamäärää".

Rakenteetonta dataa kerättiin mm. asiakaspuhelujen, asiakkaan lähettämien sähköpostien sekä muiden asiakaspalautteiden kautta. Tapausyritys hyödyntää ulkopuolista dataa sekä varastoitavassa että uudelleen hyödynnettävässä muodossa ja raportteina, joita ei erikseen tallenneta heidän omiin tietokantoihinsa jatkojalostamista ja -hyödyntämistä varten. Tällaisista tapauksista saatiin seuraavanlaisia havaintoja mm. uutisseurantaan ja sosiaaliseen mediaan liittyen:

"[...] meillä on tuommonen, uutisista kerätään, MeltWater -niminen palvelu muistaakseni, joka tavallaan suodattaa eri uutiskanavista mitä, no varmaan Suomessa on, tuskin se maailmalta sitä haravoi vaan suodattaa Suomessa olevista uutisista niin semmoset, missä esiintyy joitakin avainsanoja jotka on meille kiinnostavia, siis se tuuppaa sen tuonne meidän Intranettiin sen niinku tällasen uutisvirran".

Sen sijaan ulkopuolista dataa, jota kerättiin ja varastoitiin myös jatko- ja hyödyntämistä varten, oli muun muassa asiakkaiden kiinteistötiedot, sähköntuotantoon liittyvät hinta- ja kilpailijatiedot, toimialaan liittyvät laki- ja asetusmuutokset sekä säätiedot:

"[...] tällä hetkellä esimerkiksi on käytetty sitä tietoa, mikä on kunkin kiinteistön lämmitysjärjestelmä ja sen perusteella on sitten markkinoinnin ja myynnin toimenpiteitä kohdistettu".

6.2.2 Datan varastointi

Tapausyrityksessä datan varastointiin liittyvissä vastauksissa esiintyi eniten erilaisia vastauksia haastateltavien havainnoissa. Suuri osa vastaajista kertoi lähes kaiken heidän keräämänsä datan olevan pilvipalvelimilla:

"Varmaan se pilvi on, se on niin massiivinen se datavarasto siellä".

Osa vastaajista taas kertoi datan varastoinnin jakautumisesta lähes puolitain talon sisäisiin varastointiratkaisuihin ja pilviratkaisuihin:

"[...] mutta meillä on myös pilvessä olevia palvelimia, että meillä on sellanne siirtymävaihe oikeastaan siinä".

Kuitenkin haastateltavat, jotka olivat eniten tekemisissä yrityksen tietovarastointiratkaisujen kanssa, vastasivat yrityksen datan varastoinnin tapahtuvan pääasiallisesti talon sisällä muutamien ulkopuolisen palveluntarjoajan pilveen sijoitettuja ratkaisuja lukuun ottamatta:

"Että kyllä meillä on meidän järjestelmien infrastruktuuri niinku oma konesali ja suurin osa järjestelmistä on siellä, näitä asioita palveluita lukuun ottamatta varastoidaan meidän omissa järjestelmissä".

Kaikkiin kolmeen vastausluokkaan kuului useampia haastateltavia, jonka perusteella kaikille kolmelle havainnolle löytyy yrityksen sisältä tukijoita.

6.2.3 Datan yhtenäistäminen ja koonti

Suurella osalla haastatteluun osallistuneilla ei ollut minkäänlaisia havaintoja datan yhtenäistämistä ja koostamisesta johtuen yrityksen monimuotoisen datan keräämisen puutteesta. Suurin osa kerättyyn dataan liittyvästä yhtenäistämistä vaikuttaisi tapahtuvan järjestelmien sisällä:

”Meillä on tosiaan meidän kumppanit, jonka kanssa ollaan rakennettu tää tekniikka ja toteutus ja ne prosessit sinne ja sitten tietysti niitä kuutioidaan ja julkaistaan eri välineillä sitten sitä dataa. Mut kyllä ihan tällöinen perinteinen tietovarastointiratkaisu se on, ettei siinä mitään eksoottista ole”.

Pääasiallisesti haastattelijoilla ei kuitenkaan ollut kommentoitavaa tähän osa-alueeseen, jonka sijaan suurin osa vastauksista liittyi siihen, kuinka yrityksessä ei ole esiintynyt ongelmia datan yhtenäistämisen ja koonnin osa-alueella, sillä lähes kaikki yrityksen tällä hetkellä keräämästä datasta on käsiteltävissä reaaliaikaisissa:

” En oo kuullut, en tiedä, että meillä ois semmoista käytössä [...] Meillä se on hyvin tarkasti tuommoista mitattua, tuotettua, myytyä, aika selkeää se data kuitenkin”.

6.2.4 Datan mallinnus ja analyysi

Haastateltavien yleisin vastaus datan mallinnuksesta ja analyysistä liittyi Excelin ja Excel-pohjaisten ratkaisujen hyödyntämiseen datan mallinnuksessa ja analyysissä sekä automatisoidussa ja manuaalisessa työstössä:

”No pääasiallinen työkalu on Pivot, eli sillä Excelin Pivot-ominaisuudella pureudutaan sitten tänne tietovarastoon, pyöritellään siinä sitten niitä kymmeniä dimensioita ja useita faktoja eri näkökulmista,”

Tämän lisäksi yrityksellä oli käytössä ohjelmia, joita oli mahdollista hyödyntää datan automatisoituun mallintamiseen rakentamalla järjestelmään kaavoja, jotka mahdollistivat automatisoidun analyysin:

”Pääasiallisesti se järjestelmä käsittelee, mutta sitten ne joita järjestelmä ei osaa analysoida tai tulkita tai sanomat ei jostain syystä mene, niin ne tulee sitten manuaaliseen käsittelyyn, joka on osittain automatisoitu sillä tavalla, että sitten jollain järjestelmän palikalla pystytään vielä sitten analysoimaan ja tekemään korjauksia. Sitten jää ihan tietty osa semmoiset, jotka pitää vaan manuaalisesti sitä tietoa sitten tulkita ja korjata ja muuta”.

Kuitenkin valmiiden pohjien koettiin kattavan vain vähäinen osa kaikkea mallinnusta ja analyysia vaativaa tietoa ja osa-alueita. Suurin osa analyysistä tapahtui joko laatimalla hyödynnettävät mallit alusta lähtien itse, manuaalisesti olemassa olevia pohjia hyödyntäen tai määrittämällä erikseen yrityksen käyttämään ratkaisuun puoleksi automatisoituja ratkaisuja hyväksi havaittujen mallien pohjalle:

”No vähemmän meillä mitään valmiita pohjia on, ehkä me ollaan meillä siinä asemassa tuolla, että me ensin itse haetaan niitä malleja ja sitten kenties löydetään jotain

semmoisia, mitä jatkossa tarvitaan ja voidaan siitä automatisoida joku semmoinen raportti, että se on enemmän sellaista testausta ja kokeilemistä, että mikä ois semmoinen hyvä juttu.”

Harva yrityksen ratkaisusta kuitenkin mahdollisesti datan automatisoidun mallintamisen, jonka sijaan kaikki visuaalinen tuotanto rikastetusta datasta tehtiin pääasiallisesti käsin. Tämä koski erityisesti datan hyödyntämisestä sisäisesti, eli asiakkaan ymmärtämistä koskevan datan visualisointia. Haastatteluissa kuitenkin nousi esille, että yrityksellä oli käytössä ratkaisuja, jotka automatisoidusti visualisoivat dataa asiakaskulutuksesta asiakkaille itselleen verkkosivustoilla.

6.2.5 Datat tulosten tulkinta

Datan tuloksia tulkittiin tapausyrityksessä pääasiallisesti eri osa-alueiden asiantuntijoiden ja omien kokemusten pohjalta. Suurin osa haastateltavista vastasi datan tulosten tulkinnan tapahtuvan manuaalisesti, jolloin sisäistä ja ulkopuolista referenssidataa hyödynnettiin pääasiassa silloin, jos alalla oli tapahtunut poikkeavuuksia, jokin kilpailija oli pärjännyt erityisen hyvin tai tulkitsijan mielestä uusissa tuloksissa oli erityisiä poikkeamia aiempiin tuloksiin verrattuna. Pääasiallinen arvo asetettiin usein asiantuntijuutta ja referenssidataa vertailtaessa asiantuntijoille:

”No väittäisin että aika monesti, että jos me havaitaan joku mittarivirhe niin asiantuntijahan sen ottaa käsittelyyn, mä puhun taas niistä ja käyttämällä omaa asiantuntijuuttaan selvittää aiheuttaako tää lisätoimenpiteitä vai voidaanko me vaan tällöinen poikkeama täällä datassa hyväksyä ja kyllä se aina jos se poikkeama hyväksytään niin se perustuu siihen asiantuntemukseen ja kokemukseen”.

Datan tulkinnan väliaikoja ja jatkuvuutta selvitettäessä haastateltavat vastasivat pääasiallisesti tarkastelevansa dataa ja tekevänsä siitä tulkintoja kuukauden välein:

”Kyllähän se sitten kuitenkin useimmiten on näissä isommissa asioissa sitten se kuukausitaso, että esimerkiksi meidän täytyy toimittaa joka kuukausi tiettyjä asioita meidän ulkoiselle asiakaspalvelulle että sinne lähtee joka kuukausi tiettyjä tällöisiä tiedostoja ja sitten muualle myös usein, kuukausitaso meillä se, millä tavallisesti kerätään sitä tietoa”.

Vain harva vastattavista vastasi käyttävänsä dataa tukena päivittäisen päätöksenteon ja toiminnanohjauksen välineenä. Osa haastateltavissa myös kommentoi dataa saatujen tulosten tulkinnan tapahtuvan niin sanotusti ”vain tarpeen tullen”, jolloin dataa saatuja tuloksia tulkittiin vain, kun niistä etsittiin vastauksia ilmenneihin poikkeavuuksiin tai erikoisiin tapahtumiin haastateltavien vastuualueilla.

6.2.6 Datat käyttöönotto

Suurimmalla osalla haastateltavista datan käyttöönotto kuului jollain tasolla osaksi heidän rooliaan yrityksessä. Vain kaksi (2) haastateltavaa vastasi vastuunsa ulottuvan vain datan tulosten tulkintaan, joista ei tehty edes raportteja

muulloin, kuin esiintyneissä poikkeustapauksissa. Pääasiallisesti datan käyttöönotto tapahtui koko organisaation laajuudella raporttien laatimisen muodossa, mutta harvalla osa-alueella datan käyttöönotto tapahtui omaan toimintaan ja prosesseihin liittyvien muutosten toteuttamiseksi:

”Ei se oikeastaan ole kuin tuollainen seuranta, että ei me tuota tuommoista tietoa mihinkään hankinnan suunnitteluun tai isompien linjauksien tekemiseen juurikaan tehdä”

”Meillä on varmaan osa meidän toiminnoista sillä tavalla kypsymättömiä, että siellä ei välttämättä kaikkea olennaista dataa osata seurata ja joskushan törmää tällaiseenkin ilmiöön että jos sä haluat kerätä tarkempaa dataa jostain toiminnasta niin se onki, vastustetaan, niinku "ei saa" ja keksitään vaikka mitä uhkakuvia ja riskejä ja mitä siihen datan keruuseen liittyy ja tuota yleensä siihen on jonkinlainen pelko taustalla että kytätään entistä tarkemmin ja näin”.

Huomattava osa vastaajista kertoi vastuualueellaan datan käyttöönoton liittyvän ennusteiden laatimiseen, joita käytettyyn tuotannon ja sijoittamisen ohjenuorena:

”Ja onhan tuossa sitten sähkökauppahan elää siinä päivärytmissä, että sehän on koko ajan kun joudutaan aina ennustamaan seuraava päivä, niinku jokaisen tunnin tuotanto joudutaan myöskin sähkö asiakkaiden kulutus joudutaan ennustamaan jonkun tunnin tarkkuudella, niin siihen palataan sitten seuraavana päivänä aina että miten hyvin se ennuste meni.”

Huomionarvoinen ilmiö datan käyttöönottoon liittyvistä vastauksista ei olekaan se, millä kaikilla tavoilla datan käyttöönotto yrityksessä haastatteluista tehtävien havaintojen perusteella tapahtuu, vaan se, mille osa-alueelle yksikään ratkaisusta ei suoraan liittynyt – asiakkaan lähestymiseen ja tuotteiden kohdistamiseen liittyvien muutoksien tekemiseksi näiden suoriutumistehon parantamiseksi ja asiakkaan oletettujen toiveiden sekä asiakaskokemuksen ehostamiseksi. Asiakasanalytiikka hyödynnettiin pääasiallisesti tietoisesti tuotantoon liittyvien tekijöiden muuttamiseksi, joka vaikuttaa asiakkaaseen sähköhinnan muodostumisen kautta, mutta lukuun ottamatta visualisoituja tilastoja asiakkaan sähkökulutuksesta pääasiallisesti datan käyttöönotto tapahtui asiakkaan arvokokemukseen liittymättömillä tavoilla.

6.2.7 Datalinjaukset

Vain yhdellä haastateltavalla 13:sta oli vastatessa selkeä havainto heidän osastoillaan olemassa olevista datalinjauksista, jotka ohjasivat datan hyödyntämisprosessin kaikkia osa-alueita:

”[...] muistan, että muutaman kerran ollaan, me ollaan kerran vuodessa tai puolen vuoden välein semmoinen workshop, niin siellä ollaan keskusteltu näistäkin aiheista, että mikä ois meillä, mitä työkaluja me voitais hyödyntää sieltä mittaustiedosta tai asiakastiedosta, löytääksemme niitä poikkeamia.”

Suurimmalla osalla haastateltavia tämän osa-alueen puuttuminen osasto-kohtaisesti sekä yrityskohtaisesti heijastui heidän vastauksistaan ongelmissa

olennaisen tiedon rikastamisesta kerätystä datasta sekä hyödyntämättömän datan keräämisestä osana datan keräysprosessia:

”Enemmänkin ollut haasteena se, ettei olla vielä kaikkea dataa kerätty, mikä ois tarkoituksenmukaista ja tarpeellista”.

”[...] aika paljon just niitä yllätyksiä, että miksi tämmöistä oikeasti kerätään, että eiks tämmöstä oo [...] toi yleensä, mitkä mielletään tuotteeksi ja palveluksi, niin se on selänen, että niistäkään ei ole selkeää linjausta, porukka toimii niissä omissa lokeroissaan. Et siinäkin tulee just se, okei tiedetään asioista, dataa kerätään ehkä, mut se ei siirry laajemmalle ja ristiin, että siitä oltais koko konsernissa tietosia”.

Haastattelussa suurin osa ei tunnistanut termiä ”datalinjaus”, jonka lisäksi termin avaamisen jälkeen haastateltavilla ei ollut havaintoja datan hyödyntämisen prosessiin liittyviin osa-alueisiin linkitettyjen ohjeistuksen tai asiakasanalytiikkaan liittyvän terminologian yhtäläisyydestä koko yrityksen sisällä.

6.3 Tapausyrityksen keräämä asiakasdata

Tässä kappaleessa käsitellään havaintoja tapausyrityksen keräämästä asiakasdatasta. Tutkimuksessa hyödynnettyyn datan hyödyntämisen viitekehukseen on sisällytetty myös Sodenkamp (2015) ehdottamia osa-alueita datasta, jota energia-alan yritysten tulisi kerätä asiakkaistaan paremman asiakasymmärryksen saavuttamiseksi ja heitä parhaiten vastaavien palveluiden ja tuotteiden kehittämiseksi. Näitä olivat muun muassa *asiakkaan kotitalouden piirteet* (kuten talouden lämmitystyyppi, mahdollinen kiinnostus vihreään energian/biokaasun hyödyntämiseen sekä kotitalouden asuntotyyppi), *asiakkaan energiakulutusdata*, *asiakkaan perusdata* (kuten asukkaan osoite, ikä, perhestatus ja muu demografisiin segmentointitekijään lukeutuva data) ja *muu asiakkaaseen liittyvän informaatio* (kuten naapurusto, sosiodemografinen statistiikka ja paikalliset sääolosuhteet), joiden esiintymisiä ja esiintymisen puutteita yrityksessä käsitellään kirjallisuudessa suositeltujen aihealueiden nimiä kantavien otsikoiden alla.

6.3.1 Asiakkaan kotitalouden piirteet

Yrityksen kerryttämästä asiakasdatasta esille nousi vahvasti asiakkaan perustietojen kerääminen sopimuksissa, johon liittyi myös asiakkaan kotitalouden piirteiden kartoittaminen. Muun muassa asiakkaan kiinteistön tietojen keräämisen kuuluminen osaksi asiakkaasta kerättäviä tietoja nousi esille haastattelussa:

”[...] me saadaan aina mm. rakennuskuutiot, huoneistomäärät, periaatteessa saatais asukamäärät mutta niitä ei päivitetä, liittymisvuodet, rakennustyyppit, ainakin näitä”

Lisäksi asiakkaan kiinteistöön liittyviä tietoja kerättiin toimialueen kaupunkien kerryttämää dataa hyödyntäen:

”[Yrityksen tämän hetkiseen paikkajärjestelmään] saadaan mm. Jyväskylän alueen kartta, johon voidaan laittaa esimerkiksi ”tämä kiinteistö on kiinnostunut liittymään kaukolämpöön, tää on se lämmitysjärjestelmä”, ”Tällön on saneerattu”. Se toimii sitten taas erillään näistä kaikista muista ja siihen [järjestelmään] on yhdistetty sitten kaupungin tietojärjestelmän tietoja. Ja tota sielä on taas sitten huikeesti tietoa”

Paikkatietoja käsittelevä järjestelmä oli kuitenkin poistumassa käytöstä ja sen tilalle oli tulossa ratkaisu, jossa kohteita koskevaan analytiikkaan ei ollut mahdollista lisätä aiemman ratkaisun kaltaisia lisätietoja.

6.3.2 Asiakkaan energiankulutusdata

Tapausyrityksen vahvimpia osa-alueita oli asiakkaan energiankulutusdatan kerääminen, jonka keräämisen syklityksessä se oli lähiaikoina tehnyt muutoksia. Uudistuksen jälkeen kulutukseen liittyvää dataa alettiin kerätä tuntikohtaisesti:

”[Yrityksen käyttämä järjestelmä] kerää meidän asiakasdatan sähköstä ja lämmöstä, tunnin tarkkuudella tulee se mittausdata”.

Lisäksi asiakkaan energiakulutukseen liittyvä data oli saatettu asiakkaiden saataville visuaalisessa muodossa:

”Haluttiin meidän asiakkaille esitellä heidän oma kulutuksensa tuntitasolla. Asiakshan pääse netin kautta itse kätteleen, kuten varmaan lähes joka ikisellä energiayhtiöllä on jo tänä päivänä käytössä, että pystyvät sellaisen palvelun asiakkailleen tarjoamaan”.

6.3.3 Asiakkaan perusdata

Tapausyritys keräsi asiakkaan perusdataan liittyvää tietoa laadittujen sopimusten avulla, joissa pääasiallisina kerättävinä tietoina on asiakkaan nimi, henkilötunnus ja sopimuskohteen osoite. Tämän lisäksi sopimuksessa kerätään tietoja myös asiakkaiden puhelinnumerosta ja sähköpostista, mutta sen hyödynnettävyys esim. markkinoinnissa jää vähäiseksi, sillä asiakkaalla itsellään on mahdollisuus estää tietojen näkyminen:

”Me tarvitaan se tieto, että kuka sen sopimuksen tekee ja nimi ja henkilötunnus on meille ne tärkeimmät tiedot, että totta kai puhelinnumero ja sähköpostikin me kysytään, mutta että ne pystyy tavallaan asiakas blokkamaan, että ei tarvi niitä silleen täyttää”.

Sopimuksessa kerättävien osa-alueiden lisäksi asiakkaihin liittyviä perustietoja kerätään myös online-palvelujen kautta, jossa asiakkailla on mahdollisuus täydentää pakollisten tietojen lisäksi lisätietoja itsestään ja kiinteistöstään täydentämällä esimerkiksi perheensä kokoa ja lisätietoja kiinteistöön liittyen:

”Ja siellä onlinepalvelu-puolella niin siellä sitten asiakas pystyy myöskin profiloimaan sitä ikään kuin sitä asuntoaan, tää koskee kuluttaja-asiakkaita siis vain, eli pystyy kertomaan talouden henkilömäärän ja no talouden lämmitysmuoto meillä onkin jo siellä tiedossa ja tota, sitten pystyy jonkin verran määrittelemään onko jotain lattialämmitystä tai saunaa tai muuta tämmöistä jotka on sitten taas kulutuksen vertailutietoa, mutta tavallaan tuo sitten meille myös sitä profilointitietoa.”

6.3.4 Muu asiakkaaseen liittyvä informaatio

Tapausyritys keräsi muuta asiakkaaseen liittyvää dataa sosiaalisen median ja uutisseurannan avulla, jonka lisäksi asiakkaan kanssa tapahtuneista yhteydenotoista kerättiin dataa liittyen muun muassa asiakastyytyväisyyteen. Uutisseuranta ja sosiaalisen median seuranta toteutettiin pääasiallisesti ulkopuolisen osapuolen toteuttamina raportteina, joista dataa ei jäänyt yrityksille muussa muodossa talteen:

”Meillä on tuommoinen niinku uutisista, uutisista kerätään, MeltWater -niminen palvelu muistaakseni, joka niinku tavallaan suodattaa eri uutiskanavista mitä, no varmaan Suomessa on, tuskin se maailmalta sitä haravoi, vaan suodattaa Suomessa olevista uutisista semmoiset, missä esiintyy joitakin avainsanoja, jotka on meille kiinnostavia, siis se niinku tuuppaa sen tuonne meidän Intranettiin sen niinku tällaisen uutisvirran”

Alaan ja yritykseen liittyvien mediasta tehtävien raporttien lisäksi yritys toteutti asiakaspalautekyselyjä, joiden avulla yritys yritti paremmin kartoittaa asiakkaiden kokemuksia palveluista, asiakkaiden luonteesta ja heille tärkeistä arvoista. Asiakaspalautteita kerättiin useita kanavia pitkin, mutta vielä nykyisessä muodossaan sieltä ei voida takautuvasti tai yleisesti etsiä ilmentymiä tiettyyn teemaan liittyen, joka vaikeuttaa sen käytettävyyttä tarkemman asiakasymmärrettävyyden saavuttamisessa:

”Meillä on asiakaspalautejärjestelmä, jonne tavallaan jokainen asiakkaan yhteydenotto tulee sitten puhelimitse tai sähköpostitse tai mitä tahansa kanavaa. Jokainen kirjataan sinne asiakaspalautejärjestelmään ja jos tulee yksi puhelimella niin sitten se kirjataan sinne, että mitä se asiakas on sanonut ja siitä seurataan niitä määriä, että pystytään raportoimaan ja varmaan jonku aiheen perusteella pystytään varmaan vähän jottelemaan, että mitä se on sitten tavallaan siihen sisältöön enemmän sitten. Varmaan oo mitään, että sieltä osattais poimia sitten sen tarkemmin”.

Sen lisäksi yritys keräsi säätietoja tehdäkseen ennustuksia asiakkaidensa kulutuksista tulevaisuudessa:

”[...] tää säätieto esimerkiksi vaikuttaa siihen tuotannon suunnitteluun, että kuinka paljon tehdään lähitulevaisuudessa, että sehän on suunnilleen viikon päähän se säätieto kuranttia, että totta kai se on tuotannon suunnittelua, tehdään vuositasolla, että milloin ajetaan mitäkin voimalaitosta ja kuinka paljon tarvitaan lämpöä kaupunkiin, niin siitähän voi sellaisen karkean vuosisuunnitelman tehdä. Mutta kun mennään tarkemmalle tasolle, että viikkotasolle ja päivätasolle, niin sinne tarvitaan sitä sääennustetta, että saadaan ne tarkemman tason tiedot”.

Vaikka säätiedot linkittyvät asiakkaisiin heille toteutettavan palvelun tuotannon suunnittelussa, se ei kuitenkaan ole suoraan hyödynnettävissä asiakasanalytiikkaan.

6.4 Yhteenvedo

Vaikka vastauksissa oli havaittavissa lievää eroa eri osastojen välillä liittyen datan määrään, monimuotoisuuteen, kertymisnopeuteen ja arvoon, pääasiallisesti vastauksissa toistui vahvasti samat teemat. Kerätyn massan volyyymi eli määrä aiheutti selvästi eniten huolta vastaajien keskuudessa, vaikka suoria havaintoja sen aiheuttamista ongelmista yrityksessä ei erityisesti esiintynyt. Selkeästi vähiten ongelmalliseksi osa-alueeksi koettiin datan monimuotoisuus ja strukturoimattomuus, sillä yrityksessä ei erityisesti kerätty muuta kuin relaatiotietokannoissa käsiteltävissä olevaa dataa. Datan kertymisnopeus oli myös osa-alue, joka herätti huolta vastaajien keskuudessa, mutta josta ei ollut selviä havaintoja – suurin osa vastaajista, jotka vastasivat sen tuottavan ongelmia, vastasivat spekuloiden muiden osastojen tilannetta datan kanssa. Vastaajat olettivat usein verkkojen puolella olevan ongelmia datan kertymisnopeuden kanssa, mutta osaston omat kommentit pääasiallisesti kumosivat muiden osastojen edustajien oletukset. Selkeästi ongelmallisimmaksi osa-alueeksi paljastui kerätystä datasta arvokkaan tiedon rikastaminen, josta suurimmalla osalla vastaajista oli omia, osastokohtaisia kokemuksia.

Massadataan liitettyjen, ilmentävien ja ongelmia aiheuttavien osa-alueiden havainnot, eli tutkimuksessa käytetyt 4V-tekijää tukivat hyvin myös datan hyödyntämismallista tehtyjä havaintoja – suurin osa datasta varastoidaan edelleen yrityksen omissa konesaleissa, eikä yritys tarvitse erillistä ratkaisua datan yhtenäistämiseen, sillä yrityksessä ei vielä kerätä monimuotoista ja rakenteetonta dataa. Havainnot datalinjausten puutteesta tukevat myös havaintoa arvon löytämisestä datasta yhtenä suurimpana ongelma-alueena datan hyödyntämisessä.

Yritys kerää dataa asiakkaistaan pääasiallisesti yhteistyösopimuksiin liittyvien tietokenttien avulla, jonka lisäksi asiakkaiden intressejä seurattiin ja lisäarvoa pyrittiin tuottamaan nettisivukäyttäytymistä seuraamalla sekä tarjoamalla visualisoituja ratkaisuja omien tuotteidensa, eli sähkön, veden tai lämmön kulutuksen seuraamiseen. Pääasiallisesti kuitenkin etenkin demografisten tietojen seuraaminen oli havaintojen mukaan suhteellisen pintapuolista, eikä ulkoisia datalähteitä hyödynnetty erityisesti kohderyhmien määrittämiseen esimerkiksi tarkastelemalla kiinteistötietojen avulla jo asiakkaina olevia asuinalueita ja katveja asiakaskartassa. Yritys hyödyntää kuitenkin analytiikassaan koosteita alakohtaisista keskusteluista ja heitä koskevista uutisista mediassa MeltWaterin mediamonitorointipalvelujen avulla. Muu mm. sosiaalisen mediaan liittyvä seuranta heidän omalla tilillään kuitenkin toteutettiin pääasiallisesti sitoutumisen seurannalla julkaisuihin ja yrityksen sivuihin, eikä seuraajien käyttäytymistä kanavalla hyödynnetty muilla tavoin kohderyhmää kiinnostavien trendien tai aihealueiden seuraamiseen. Tähän mennessä kuitenkin asiakkaisiin liittyvää dataa ei ole käytetty analytiikassa esimerkiksi kohderyhmien tai asiakaskunnasta selkeästi erottuvien segmenttien määrittämiseen. Tällaista segmentointia voitaisiin hyö-

dyntää esimerkiksi markkinoinnin kohdentamiseen sekä tunnistamaan mahdollisia kohderyhmiä, joiden tarpeisiin nykyinen palvelutarjonta tai markkinointi eivät vielä vastaa.

7 POHDINTA JA JOHTOPÄÄTÖKSET

Tässä kappaleessa tutkimuksen teoriaosuus ja empiirinen tutkimus kootaan yhtenäiseksi kokonaisuudeksi käsittelemällä tutkimushaastattelusta saatujen havaintojen johtopäätökset sekä niiden suhdetta teoriaosuudessa tehtyihin havaintoihin.

Tutkimuksen tavoitteena oli selvittää keinoja hyödyntää massadataa asiakasanalytiikassa avaamalla massadatan käsitteistöä ja siihen liittyviä ratkaisuja. Tutkimustavoitteesta muodostettiin tutkimuksen tutkimuskysymys ja apukysymykset, jotka olivat:

- Kuinka massadata -pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa?
- Millä metodeilla yritys pystyy tehokkainten tunnistamaan ja rikastamaan asiakasanalytiikan kannalta olennaista tietoa?
- Millainen asiakassegmentointimalli mahdollistaa yrityksen tunnistaa paremmin sekä potentiaalisia että jo palvelemaan asiakassegmenttejä

7.1 Empiirisen tutkimuksen johtopäätökset

Haastattelututkimuksen tavoitteena oli määrittää, hyödyntääkö yritys teorialähtöisessä tutkimuksessa löydettyjä tapoja käyttää massadataa asiakasanalytiikassa ja erityisesti tältä pohjalta toteutetussa asiakassegmentoinnissa. Näitä ilmiöitä tutkittiin massadataa, liiketoimintatiedon hallintaa ja analytiikkaa sekä asiakasanalytiikkaa käsittelevän kirjallisuuden pohjalta laadittua datan hyödyntämismalli -viitekehystä hyödyntäen. Haastattelututkimuksessa pyrittiin myös selvittämään, kuinka viitekehys toteutuu yrityksessä. Lisäksi tutkimuksen tavoitteena oli selvittää, onko yrityksellä teorian pohjalta tehtyjen määritelmien mukaan massadataa ja käyttääkö yritys teoriakatsauksessa esiteltyjä massadatan osa- tai kokonaisratkaisuja asiakasdatan hyödyntämiseen.

Haastattelututkimuksen havaintojen pohjalta yrityksellä ei vielä nykyisellään ole massadataa, vaikka yrityksessä esiintyy V-mallin tekijöihin liitettäviä lieveilmiöitä. Szymczakin, Zelikin ja Elmin vuonna 2014 määrittämästä 3V-mallista, eli datan massasta, monimuotoisuudesta ja kertymisnopeudesta yrityksessä esiintyi huolia pääasiallisesti vain massan, eli datan volyymin suhteen. Ayan-koya ym. vuonna 2014 määrittelemistä lisätekijöistä 4V-malliin mukaan otettu arvon löytäminen datasta esiintyi kuitenkin vahvinten yrityksessä ja vastaajilla oli eniten henkilökohtaisia kokemuksia juuri tästä tekijästä. Vaikka massa koettiin haastatteluissa suurimmaksi ongelmakohtaksi, vain harvalla vastaajista oli siitä henkilökohtaisia kokemuksia ja pääasiallisesti huoli liittyi hetkittäisiin järjestelmien hidastumisiin. Arvon löytämisen osa-alueella ongelmaksi ei niinkään

muodostunut datan suuri määrä, josta oleellista arvoa oli vaikeaa löytää, vaan saatavilla olevien järjestelmien hyödyntäminen kaiken hyödyllisen tiedon rikastamiseen sekä yrityksen toiminnoille oleellisen datan tunnistaminen. Tähän selityksenä vaikuttaisi olevan muun muassa selkeiden datalinjausten puuttuminen osastoilla, joka on yksi tärkeimmistä osa-alueista datan kattavan hyödyntämisen mahdollistamiseksi (Ayankoya ym., 2014). Koska yritys ei haastatteluhetkellä kerryttänyt monimuotoista ja strukturoimatonta dataa eikä datan kertymisnopeudesta ollut merkittäviä havaintoja, yrityksen kerryttämä ja käsittelemä data on lueteltavissa massadataksi vasta, kun yrityksen 4V-tekijät alkavat vaatia selkeästi erillisiä järjestelmäratkaisuja datamassan käsittelemiseksi.

Empiirisessä tutkimuksessa selvitettiin myös, miten datan hyödyntämismallin osa-alueet toteutuivat tapausyrityksessä ja mikä sen hyödyntämisaste oli. Datan hyödyntämismallin pohjana käytettiin erityisesti Kowsarin ym. (2014) sekä Jagadishin ym. (2014) teoriaa datan hyödyntämisen yksinkertaistamisesta kuuteen osa-vaiheeseen, jotka olivat 1. datan kerääminen, 2. datan varastointi, 3. datan yhtenäistäminen, 4. datan mallennus ja analyysi, 5. datan tulkinta ja 6. datan käyttöönotto. Näitä osa-alueita tapausyrityksessä tutkimalla selvisi, ettei yritys käsittele dataa hyödyntämismallin jokaisella osa-alueella. Erityisesti datan yhtenäistäminen, mallennus ja analyysi sekä käyttöönotto olivat osa-alueita, jotka eivät toteutuneet kaikilla haastateltavien osastoilla esimerkkiprosessin mukaisesti. Merkittävin poikkeus mallin toteutumisessa yrityksessä oli datan käyttöönotossa. Suurin osa vastaajista vastasi osastollaan käyttöönoton tapahtuvan raportoinnin tai laskutuksen muodossa, eli käsiteltyä ja tulkittua dataa ei käytetty osaston prosessien tai tarjoamien palveluiden aktiiviseen kehittämiseen. Jotta yritys voisi toteuttaa asiakasanalytiikkaa ja asiakkaille tuotettuja palveluja tehokkaammin, heidän tulisi tehostaa datan käyttöönottoa tukemaan omien prosessiensa tehostamista ja asiakkaille tarjottavien palveluiden kehittämistä sekä tehokkaammin hyödyntää kaikkea dataa asiakasymmärryksen kasvattamiseen. Tämän avulla he saisivat selkeämmät kehukset asiakassegmentoinnin toteuttamiselle. Asiakasanalytiikan tehostamista ajatellen yrityksen tulisi myös tehokkaammin luoda mallinnuksia ja automatisoitua vertailua asiakkaisiin liittyvien tilastojen välillä, jotta esimerkiksi asiakaspoistumia, käyttäytymistä sosiaalisessa mediassa ja kulutustottumusten muuttumista voitaisiin tarkastella yrityksessä laajemmalla aikajänteellä ja tehdä siten tehokkaampia tulkintoja ja ennustuksia kohderyhmiin liittyen. Suureen rooliin toteutetussa tutkimuksessa nousi jo aiemmin mainittu arvon löytäminen kerätystä datasta. Tämä havainto tukee datan hyödyntämismalli -viitekehukseen kuuluvan datalinjausten laatimisen tärkeyttä merkityksellisten datalähteiden määrittämiseksi sekä yhteisöjen ja organisaatioiden laajuudella hyödynnettävän analyysin toteuttamiseksi – yksi tärkeistä osa-alueista datalinjauksissa on esimerkiksi käytetyn terminologian yhdenmukaisuus, joka tapausyrityksessä tehdyssä tutkimuksessa nousi esille toteutumattomana kohtana. Erityisesti yrityksissä, jotka keräävät jo massadatan määritelmää vastaavaa määrää ja monimuotoista dataa, tämä osa-alue nousee erittäin tärkeäksi myös asiakasanalytiikan toteutuksessa.

Tutkimuksessa pohjana toimineessa viitekehyksessä kerättävää asiakasdataa ja sen kattavuutta tutkittiin käyttämällä tekijöitä asiakkaan luokittelusta merkityksellisiin ryhmiin (Kotler ym., 2008) sekä energia-alan yrityksen portfolion tekijöitä (Sodenkamp ym., 2012) vertauskohtina kerättävän datan kattavuuteen. Asiakkaan luokittelussa merkityksellisiä ryhmiä olivat maantieteelliset, demografiset, psykografiset ja käytökselliset tekijät, kun taas energia-alan yrityksen portfolion tekijöitä, jotka toimivat tarkennuksena asiakkaan luokittelun merkityksellisiin tekijöihin, olivat asiakkaan kotitalouden piirteet, asiakkaan energiakulutusdata, asiakkaan perusdata ja muu asiakkaaseen liittyvä informaatio. Yrityksessä asiakkaisiin liittyvää tietoa kerättiin eniten asiakkaan kulutusdatan ja henkilötietojen muodossa. Henkilötiedot kerättiin pääasiallisesti yrityksen asiakkaiden laatimista sopimuksista. Huomionarvoinen asiakasanalytiikassa mahdollisesti hyödynnettävä tekijä, jota yritys keräsi säännöllisesti ulkopuoliselta palveluntarjoajalta tilattuna palveluna, oli yritykseen liittyvä mediaseuranta. Mediaraportit kattoivat yritykseen ja sille olennaisiin avainsanoihin liittyvän sosiaalisen median lisäksi myös yritykseen ja alaan liittyvän perinteisen mediaseurannan, josta yritys sai säännöllisesti raportin. Raporttien hyödyntäminen tapausyrityksessä oli kuitenkin vielä rajallista, sillä esiintyneistä trendeistä ei vielä toteutettu pitkän aikajänteen seurantaa, sillä raportteja ei siirretty dokumenttina vastaanottamisen jälkeen analytiikkaa mahdollistaviin järjestelmiin. Koska yritys ei vielä kerrytä monimuotoista ja strukturoimatonta dataa, se ei nykyisellään hyötyisi massadatan analytiikkaan tarkoitetuista ratkaisuista, eikä massadataa hyödyntävästä asiakasanalytiikasta asiakassegmentointinsa toteuttamisessa.

Jotta yrityksessä voitaisiin tulkittavan olevan massadataa ja perusteita sen hyödyntämiseen tarkoitettujen ratkaisujen hankkimiselle, yrityksen tulisi laajentaa sen asiakasdatan laajuutta ja lähteitä kattamaan ainakin muutamaa osa-aluetta kaikista viitekehyksen pohjana käytetyistä asiakkaan luokitteluun merkillisistä tekijöistä. Teoriassa kuitenkin yrityksen selkeimpiä katvealueita kerrytettävässä datassa ja ongelmallisuutta asiakassegmentoinnin toteuttamisessa voidaan tulkita siten, että ulkopuolista ja rakenteetonta dataa keräämällä yritys voisi saada kattavampaa ymmärrystä sekä jo palvelemistaan että vielä tavoittamattomista kohderyhmistä. Tapausyrityksen ja sen kaltaisten yritysten kannalta massadatapohjaisen asiakasanalytiikan hyödyntäminen asiakassegmentoinnissa tulisi toteuttaa erityisesti siis kerrytettävän datan lähteitä laajentamalla. Ulkopuolisen tiedon hyödyntämisellä yrityksellä voisi olla myös helpompi määrittää, keitä toiminta-alueensa kaikista potentiaalisista asiakkaista he eivät vielä palvele ja tutkia, löytyykö tästä ryhmästä joitain merkittäviä yhtäläisyyksiä, jotka selittävät muilta palveluntarjoajilta ostamisen motiiveja. Koska analytiikkaa on vaikea toteuttaa hyödyntämällä perinteisesti asiakassegmentoinnissa käytettyjä analyyssejä, kuten asiakkaan viimeisintä ostohetkeä (*Recency*), ostojen toistuvuutta (*Frequency*) ja ostoihin käytettyä rahamäärää (*Monetary*) hyödyntävää RFM-analyysia käyttäen (Singh ym., 2014), kohderyhmien yhtäläisyyksiä ja eroavaisuuksia tulisi tarkastella muita taustatietoja ja käyttäytymisestä löytyvää dataa hyödyntäen.

Vaikka empiirisen tutkimuksen perusteella yrityksellä ei ole vielä massadataa, jota voisi hyödyntää asiakasanalytiikassa, tutkimuksessa paljastui osa-alueita, joilla tapausyritys voisi hyötyä datan hyödyntämismalli -viitekehityksen osa-alueista. 4V-tekijöistä eniten huolta aiheuttaneen datan massaa eli määrää olisi luonnollista lähteä purkamaan ottamalla yrityksessä käyttöön pilvivarastoja korvaamaan nykyisiä konesaleja, jonne dataa pääasiallisesti varastoidaan. Lisäksi tapausyrityksessä olisi kannattavaa laajentaa datan lähteiden määrää vastaamaan asiakkaan luokitteluun merkillisiä tekijöitä kattavaksi. Tapausyrityksen datankäsittelyprosessi saataisiin tehokkaammin tukemaan datan hyödyntämistä käyttöönottamalla liiketoimintatiedon hallintaan ja analytiikkaan tarkoitettuja ratkaisuja, joilla osa nykyisin manuaalisesti suoritettavista prosesseista saataisiin automatisoitua. Lisäksi tapausyritys saisi merkityksellisempiä tuloksia keräämälleen ja käsittelemälleen datalle tehostamalla sen käyttöönottoa tukemaan yrityksen prosessien ja toimintojen kehittymistä sekä asiakkaiden syvällisemmän ymmärtämisen laajentamista. Selkeillä datalinjauksilla yrityksen osastot pystyisivät tehokkaammin määrittelemään heidän toiminnalleen merkillisten datalähteiden määrittämisen sekä sen, mitä dataa yrityksen ja sen osastojen tulisi kerätä asiakasanalyysin ja tehokkaamman toiminnan mahdollistamiseksi. Ilman varhaisessa vaiheessa määritettyjä datalinjauksia yrityksen on sitä vaikeampi rikastaa kerätystä datasta oleellista tietoa, mitä enemmän kerätyn datan määrä ja lähteet kasvavat. Myös yrityksen sisäinen raportointi ja osastojen välinen viestintä tehostuisi, jos kaikki osastot puhuisivat samoista asioista samoilla termeillä - tällöin kerätty data ja sen pohjalta laaditut raportit olisivat myös tehokkaammin hyödynnettävissä koko yrityksen laajuisesti, myös asiakasanalytiikka toteutettaessa käsitellessä asiakastyyppejä ja näiden ominaisuuksia.

7.2 Empiirisen tutkimuksen suhde teoreettiseen tutkimukseen

Tässä kappaleessa käsitellään, mitä teoreettisesta tutkimuksesta saatuja havaintoja empiirinen tutkimus tuki, mitä merkittäviä eroavaisuuksia tutkimuksien välillä oli sekä millä tapaa empiirisen tutkimuksen vastaus tutkimuskysymykseen vastaa teoreettisessa tutkimuksessa saatua vastausta.

Koska tapausyrityksellä ei ollut vielä käytössään massadatan määritelmiä vastaavaa dataa, siitä saatuja havaintoja, niiden tulkintoja ja tutkimuksen perusteella tehtyjä johtopäätöksiä ei voida käyttää tukemaan teoreettisessa osuudessa tehtyjä havaintoja keinoista massadatan hyödyntämiselle asiakasanalytiikassa. Massadatan sijasta yrityksen data vastaa *small datan* määritelmää, eli dataa voidaan käsitellä yrityksessä perinteisissä relaatiotietokannoissa (Marton ym., 2013). Yrityksessä esiintyneet kehityskohdat asiakasdatan hyödyntämisessä ja asiakassegmentointimallin laatimisessa kuitenkin tukevat kirjallisuudesta löydettyjä määritelmiä kerättävän asiakasdatan tekijöille ja tarvetta laajemmille lähteille valituissa datankeruukanavissa. Erityisesti laajennettavaa löytyi Kotler ym. (2008) määrittelemistä asiakkaan merkillisen luokittelun tekijöistä, joista tapaus-

yritys keräsi pääasiallisesti demografisten tekijöiden alle lukeutuvaa tietoa. Sodenkamp. ym (2012) määrittelemistä, energia-alan yrityksille tarkoitetuissa tarkemmin määritellyistä asiakasportfolion osa-alueista yritys keräsi pääasiallisesti asiakkaan tietoja, kulutusdataa sekä muusta asiakkaaseen liittyvästä tiedosta pääasiallisesti säätietoja oman resursointinsa suunnitteluun. Myös Constantiu ja Kallinikos (2015) suosittelevat käytettävien lähteiden laajentamista ja olemassa olevien tietolähteiden kattavaa hyödyntämistä sekä niiden ottamista osaksi perinteisiä päätöksentekoprosesseja liiketoimintaympäristössä tapahtuvien muutosten seuraamiseksi, yrityksen pysymiseksi ajan tasalla ja reagointivalmiudessa alan markkinoilla tapahtuviin muutoksiin.

Koska tutkimuksessa havaittiin myös poikkeuksia Kowsarin ym. (2014) sekä Jagadishin ym. (2014) teorioissa datan hyödyntämisen yksinkertaistamisesta kuuteen osa-vaiheeseen, joita hyödynnettiin empiirisessä tutkimuksessa viitekehyksenä käytetyn datan hyödyntämismallin laadinnassa, tämä voisi viitata siihen, että yrityksen muuttaessa prosessejaan ja toimintojaan vastaamaan enemmän datan hyödyntämismallia, se voisi tehokkaammin prosessoida ja hyödyntää asiakasdataa asiakasanalytiikan ja asiakassegmentoinnin toteuttamisessa. Asiakasanalytiikan toteuttamisen ja hyödyntämisen ongelmien sekä teoriassa laaditun datan hyödyntämismallin toteutumisen poikkeavuuksien välillä ei kuitenkaan ole suoraa kausaalisuhteita, joten on mahdollista, että asiakasanalytiikan ongelmallisuus sekä mallin esittelemien prosessien ja ratkaisujen toteutumisen poikkeukset kirjallisuudessa esitettyyn malliin yrityksessä eivät liity suoraan toisiinsa.

Koska yrityksellä ei ollut vielä massadataa tai käytössä olemassa olevia asiakassegmentointimalleja ja asiakasanalytiikkaa toteutettiin kirjallisuudessa esitettyihin malleihin verrattuna vielä suppeasti, empiirisen tutkimuksen suhdetta teoreettiseen tutkimukseen ei voi niinkään tarkastella prosessien yhtäläisyyksien kuin eroavaisuuksien kautta. Suurimmaksi eroavaisuudeksi empiirisen tutkimuksen havaintojen ja teoreettisen tutkimuksen välillä nousi *datan luotettavuus ja yhtenäisyys* sen siirtyessä käytettävien järjestelmien välillä. Vaikka teoriaosuudessa yhdeksi massadataa määritteleviksi V-tekijöiksi esiteltiin olennaisen datan löytäminen kerrytettävästä datasta (*Veracity*) (Ayankoya ym., 2014), se ei vastaa tutkimushaastattelussa ilmennyttä ongelmaa tiedon käytettävyyden heikentymisestä sen siirtyessä järjestelmästä toiseen. Jos yrityksellä olisi jo käytössään massadatan nykyisiä määritelmiä vastaavaa dataa, tämä tekijä ja sen aiheuttamien ongelmien minimoiminen olisi entistä tärkeämmässä roolissa datan käytettävyyden takaamisessa.

Tutkimuksen tutkimuskysymys oli ”Kuinka massadata -pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa?”. Teoreettisen yhteenvedon tuloksena kysymyksen vastaus oli yrityksen kerryttämän asiakasdatan lähteiden laajentaminen kattamaan asiakkaiden luokittelun kannalta merkityksellisiä tekijöitä. Vaikkei tapaustutkimusyrityksellä ollut käytössään massadatan määritelmää vastaavaa dataa, myös empiirisessä tutkimuksessa nousi ilmi tarve yrityksen hyödyntämien datalähteiden laajentamiselle syvällisemmän asiakasymmär-

ryksen saavuttamiseksi. Vaikka massadatan keräämiseen, varastointiin ja käsittelyyn on olemassa jo useita ratkaisuja, joiden käyttöönotto ja hyödyntäminen kuulostavat ensimmäiseltä askeleelta suurten ja monimuotoisten datamäärien hyödyntämiseen asiakasanalytiikassa, massadatan hyödyntämistä tulisi lähestyä tavoite- eikä järjestelmälähtöisesti. On tärkeämpää lähestyä massadatan hyödyntämistä määrittelemällä, mitä dataa yrityksen tulisi kerätä ja analysoida syventääkseen asiakasymmärrystään, eli laatimalla datalinjaukset. Massadatan hyödyntämisen mahdollistamana asiakasymmärrystä voidaan syventää tutkimalla laajoja ilmiöitä kohderyhmissä muun muassa arvojen kehittymisenä sekä ennustamalla tulevia trendejä ja mukauttamalla tarjontaansa niihin vastaaviksi. Määrittämällä asiakasymmärtämisen syventämiselle olennaiset ja potentiaaliset datalähteet yritys voi tehokkaammin hyödyntää asiakkaiden segmentointia myös yrityksissä, joiden liiketoiminta ei ole perinteistä palveluliiketoimintaa, kuten energia-alan yrityksissä. Asiakkaan syvällisemmän ymmärryksen kautta heidän luokitteluaan voidaan perinteisten demografisten ja kulutustietojen lisäksi lähteä jaottelemaan heidän ostoon vaikuttavien tekijöiden, eli ideaalien ja motiivien kautta, kuten Swinyard ja Sturtman (1986) ehdottivat. Massadatan todellinen potentiaali asiakassegmentoinnin toteuttamisessa onkin suurien massamäärien hyödyntäminen syvällisemmän asiakasymmärryksen saavuttamisessa tutkimalla asiakkaiden ostopäätöksiä taustalla olevia ideoita ja motiiveja, jotka eivät tulisi yrityksen perinteisesti keräämästä asiakasdatasta esille.

Vaikka empiirisen tutkimuksen yhteydessä tehdyt havainnot tarpeesta kerrottavan datan lähteiden laajentamisen tarpeelle tukevat kirjallisuuden havainnot aiheesta, empiiristä tutkimusta ja siitä tehtyjä havainnoja ei kuitenkaan voida käyttää suoraan tukena kirjallisuudesta tehtyihin havaintoihin, koska tapaustutkimusyritys ei täyttänyt tutkimusasetelman vaatimuksia olemassa olevalle massadatalle. Sen vuoksi empiirisessä tutkimuksessa tehtyjen havaintojen suhdetta teoreettisessa tutkimuksessa tehtyihin havaintoihin tulee tarkastella korkeintaan suuntaa antavina.

7.3 Tutkimuksen rajoitukset ja käytettävyys

Tutkimuksen käytettävyyttä eli sen luotettavuutta voidaan tarkastella hyödyntämällä Lincolnin ja Cuban vuonna 1985 laatimaa määritelmää laadullisen tutkimuksen luotettavuuskriteereistä. Kriteereitä ovat tutkimuksen siirrettävyys/sovellettavuus, totuudellisuus, vahvistettavuus, uskottavuus sekä tutkijan subjektiivisuus ja objektiivisuus. Tutkimuksen siirrettävyydellä ja sovellettavuudella tarkoitetaan, tutkimusaineiston rikasta kuvailua, jonka avulla lukija voi päätellä, kuinka paljon tutkimuksen löydöksiä voi soveltaa muihin tutkimuskohteisiin. Totuudellisuudella tarkoitetaan sitä, kuinka paljon löydökset eli tutkijan johtopäätökset vastaavat tutkimuskohteen todellista tilaa. Vahvistettavuudella tarkoitetaan tehtyjen tulkintojen saamaa tukea toisista tutkimuksista. Uskottavuudella tarkoitetaan tutkijan omia ennako-odotuksia ja johtopäätösten oikeellisuutta, johon linkittyä vahvasti tutkijan subjektiivisuus ja

objektiivisuus, eli kuinka tutkijan oma persoona on mukana tutkimuksessa, säilyttäen samalla neutraali ote tutkimuksen löydöksiä kohtaan. (Lincoln & Cuba, 1985.)

Näistä luotettavuutta ja käytettävyyttä määrittelevistä tekijöistä tutkimuksen rajoituksiksi nousee tutkimuksen tulosten soveltuminen hyödynnettäväksi muissa tutkimuksissa, sillä tutkimus kohdennettiin selkeästi perinteisistä palveluntarjoajista erottuville yrityksille. Koska haastattelu toteutettiin Jyväskylässä paikallisesti toimivalle yritykselle ja heidän 13 työntekijälleen, tutkimus ei välttämättä vastaa muiden alan yritysten tilaa. Lisäksi tutkimuksen rajoitteena toimii tutkimuksessa käytetty tutkimusmenetelmä - vaikka tutkimuksen toteuttaminen teemahaastatteluna mahdollistaa syvällisten havaintojen tekemisen tutkimusalueella, se pohjautuu haastateltavien henkilökohtaisiin kokemuksiin tutkittavasta aiheesta ja voi aiheuttaa rajoitteita tutkimuksen tulosten yleistämiselle ja käytettävyydelle.

Tutkimuksen toteutus ja analyysi suoritettiin pitkällä aikavälillä, joka on yksi tunnetuimmista totuudellisuuteen vaikuttavista tekijöistä, jonka vuoksi myös se voi vaikuttaa tutkimuksen käytettävyyteen. Tutkimuksen käytettävyydessä on myös otettava huomioon se, että tapaustutkimusyrityksen tila poikkesi tutkimusasettelusta paljastuessa, ettei yrityksellä ollut vielä käytössään massadataa. Tämän vuoksi teoreettisessa tutkimuksessa tehtyjä havaintoja ei voi tukea empiirisen tutkimuksen havainnoilla.

Koska aiheesta on tehty vasta vähän tutkimusta, erityisesti tapausyrityksen kaltaisille yrityksille, tutkimuksella ei ole vielä vertailukohtia tukemaan tutkimuksessa tehtyjä havaintoja. Uskottavuuteen voi vaikuttaa tutkijan henkilökohtaiset ennakkoluulot empiirisessä tutkimuksessa tehtyihin vastauksiin liittyen, joka voi vaikuttaa tulosten tulkintaan ja havaintojen yhdentämiseen.

Myös tutkimukseen valittujen osallistujien määrä voidaan nähdä rajoituksena tutkimukselle. Tutkimuksessa haasteltiin 13 edustajaa ja vaikka edustajien valinnassa keskitettiin tapausyrityksen sisällä määrän sijasta laatuun ja oleellisuuteen käsiteltävien aiheiden kannalta, haastateltavien pieni määrä rajoittaa haastattelututkimuksen tulosten yleistettävyyttä. Haastateltavien määrässä on kuitenkin otettava huomioon tutkimuksen toteutustapa tapaus tutkimuksena tietylle yritykselle, joka vaikuttaa haastateltavien otannan mahdolliseen enimmäismäärään.

Lisäksi tutkimuksen hyödynnettävyydessä tulee ottaa huomioon toteutetun tutkimuksen tapausyritys ja sen harjoittaman liiketoiminnan luonne, joka voi rajoittaa tulosten yleistettävyyttä - Jyväskylän Energia on Keski-Suomen alueella toimiva sähkö-, lämpö- ja vesipalveluita tarjoava yritys, jonka asiakkaiden ostokäyttäytyminen ja palveluihin sitoutuminen ovat heikosti verrattavissa esimerkiksi perinteistä kuluttajaliiketoimintaa harjoittaviin yrityksiin.

7.4 Eettinen tarkastelu

Tutkimuksen toteutusta tulee tarkastella myös sen tutkimuseettisestä näkökulmasta, eli kuinka haastattelututkimukseen osallistuvia haastateltavia on ja heidän yksityisyyttään on käsitelty haastattelututkimuksen toteuttamisessa. Hirsjärvi, Remes ja Sajavaara (2009) määrittelevät eettisen tutkimuksen lähtökohdiksi ihmisarvon kunnioittamisen sekä huolellisuuden ja tarkkuuden noudattamisen tutkimustyössä. Lisäksi he panostavat eettisesti kestävien tiedonhallinta-, tutkimus- ja arviointimenetelmien käyttämistä sekä avoimuutta tutkimuksen toteuttamisessa (Hirsjärvi, Remes ja Sajavaara, 2009).

Haastattelututkimuksen tapauksessa on tärkeää, että jokaiselta haastateltuun osallistuvalla on saatu suostumus tutkimukseen osallistumisesta ja heidän anonymiteettinsa ja tutkimuksen luottamuksellisuus on taattu (Hirsjärvi ym., 2009). Kaikki tutkimukseen osallistuvat antoivat suostumuksensa tutkimukseen osallistumiseen sekä haastattelujen tallentamiseen. Osallistujien anonymiteetista ja haastattelutallenteista sekä niistä tehdyistä litteroinneista huolehdittiin varastoimalla ja käsittelemällä tutkimusaineistoa luottamuksellisesti ja tuhoamalla haastattelumateriaalit, kun niiden säilyttäminen ei ollut enää tarpeellista tutkimuksen toteuttamisen kannalta. Tutkimuksen raportoinnissa tutkimuksen kulua ja tutkimukseen liittyviä puutteita on pyritty kuvaamaan tarkasti ja rehellisesti. Tämän perusteella tutkimus täyttää Hirsjärven, Remeksen ja Sajavaaran vaatimukset eettiselle tutkimukselle.

8 YHTEENVETO

Massadata, sen luonne sekä potentiaali liiketoiminnassa hyödyntämiselle kiinnostavat monia. Sen vuoksi on tärkeää määritellä tapoja, työkaluja ja prosesseja sen hyödyntämiseen liiketoiminnan analytiikassa. Teknologian kehityksen vähentäessä perinteisiä kilpailuedun tekijöitä yritysten on erityisen tärkeää luoda merkityksellisiä suhteita asiakkaidensa kanssa tunnistamalla heille merkityksellisiä asiakaskohderyhmiä sekä keinoja tuottaa entistä enemmän heille kohdennettuja palveluja ja markkinointia asiakasanalytiikan kautta.

Tämä Pro-Gradu tutkimus toteutettiin toimeksiantona Jyväskylän Energialle, joka on Jyväskylän alueella toimiva sähkö-, lämpö- ja vesipalveluita tarjoava yritys. Tutkimuksen tavoitteena oli avata massadatan määritelmää ja luonnetta sekä sen potentiaalia erityisesti liiketoiminnan alueella tutkimalla keinoja hyödyntää sitä asiakasanalytiikassa. Näitä asioita tutkittiin käsittelemällä tutkimuksen teoreettisessa osuudessa massadatan määritelmiä sekä massadatan potentiaalia ja hyödyntämisen keskeisimpiä ongelmia. Seuraavaksi tutkimuksessa avattiin liiketoimintatiedon hallintaa ja analytiikkaa, sekä massadatan suhdetta osa-alueen ratkaisuihin. Lopuksi kirjallisessa osuudessa käsiteltiin asiakasanalytiikkaa ja asiakassegmentointia sekä esiteltiin teoreettisen tutkimuksen pohjalta rakennettu datan hyödyntämismalli -viitekehys, jota hyödynnettiin empiirisen tutkimuksen toteutuksessa.

Empiirisessä tutkimuksessa vastausta tutkielman tutkimuskysymykseen ”Kuinka massadata -pohjaista analytiikkaa voidaan hyödyntää asiakassegmentoinnissa” tutkittiin tutkimushaastattelulla haastatellen tapaustutkimusyrityksen työntekijöitä. Tutkimuksen tavoitteena oli selvittää, miten datan hyödyntämismalli toteutuu viitekehyksessä ja esiintyykö yrityksessä teoriaosuudesta löydettyjä massadataan liitettäviä 4V-mallin ongelmia, jotka ilmentävät myös massadatan olemamassa oloa yrityksessä. Tutkimus toteutettiin teemahaastatteluna 13:lle yrityksen työntekijälle, jonka jälkeen tutkimusaineisto analysoitiin teemoittamalla vastaukset 4V-mallin tekijöiden, datan hyödyntämismallin viitekehysten toteutumisen sekä kerätyn asiakasdatan luokkiin.

Empiirisen haastattelun tuloksia ei voitu käyttää tukemaan teoreettisessa tutkimuksessa tehtyjä havaintoja, koska tutkimuksen tuloksena paljastui, ettei yrityksellä ole käytössään vielä massadatan määritelmää vastaavaa dataa. Tutkittaessa 4V-mallin tekijöiden ilmentymiä yrityksissä suurin osa haastateltavista vastasi, ettei heillä ole omakohtaisia havaintoja suurimmasta osasta tekijöihin liitettävistä ongelmista. Tehdyt havainnot kuitenkin tukivat kirjallisuuden havainnot siitä, että massadatan hyödyntämisessä suurimmaksi ongelmaksi nousee datan hyödyntämisen prosessien puutteellisuus sekä datan roolin väärinymmärtäminen. Havainnot tukivat myös teoreettisessa osuudessa tehtyjä havaintoja siitä, että käytettävien datalähteiden laajentaminen yrityksen tavoitteelle merkityksellisten datalinjausten puitteissa mahdollistaa kattavamman asiakasymmärryksen ja sen pohjalta tehtävien asiakkaiden motivaatioiden ja ideaalien hahmottamisen, jota voidaan käyttää tukena asiakassegmentoinnin laatimisessa.

Tutkimuksen apukysymykset olivat ”Millä metodeilla yritys pystyy tehokainten tunnistamaan ja rikastamaan asiakasanalytiikan kannalta olennaista tietoa?” sekä ”Millainen asiakassegmentointimalli mahdollistaa yrityksen tunnistaa paremmin sekä potentiaalisia että jo palvelemaan asiakassegmenttejä?”. Empiirisen tutkimuksen havainnot datalinjausten puuttumisesta tukivat teoreettisessa osuudessa tehtyjä havaintoja datalinjausten tärkeydelle oleellisen datan tunnistamisessa ja rikastamisessa asiakasanalytiikan kannalta. Koska yrityksellä ei ollut vielä käytössään olemassa olevia asiakassegmentointimalleja, yrityksen kerryttämän datalähteiden suppeuden voidaan tulkita tukevan teoreettisessa osuudessa tehtyjä havaintoja lähteiden laajentamisen tarpeelle mm. ulkopuolista dataa hyödyntämällä.

8.1 Johtopäätökset

Tässä tutkimuksessa saatujen tulosten perusteella massadataan pohjautuva analytiikka itsessään ei tuota yrityksille lisäarvoa ja syvällisempää ymmärrystä asiakkaista, ellei yritys ymmärrä, millaista dataa keräämällä ja jalostamalla asiakasymmärrystä voidaan laajentaa. Asiakasanalytiikassa hyödynnettävien lähteiden tulisi kattaa mahdollisimman monipuolisia osa-alueita, jotta asiakkaan demografisten tietojen lisäksi saataisiin käsityksiä asiakkaiden arvoista ja muihin ostopäätöksiin vaikuttavista tekijöistä, asiakkaiden ideaaleista ja motivaatioista. Myös trendien seuraaminen ja ennustaminen lisää yrityksen mahdollisuuksia tuottaa entistä enemmän asiakkaille kohdennettuja ratkaisuja, auttaen yritystä erottumaan kilpailijoistaan ja luomaan merkittävemmän yhteyden asiakkaisiinsa. Massadatapohjaisen analytiikan mahdollistama laajaan historiadataan perustava ennakoiva analytiikka onkin yksi sen merkittävimpiä etuisuuksia suuren määrän dataa vaativien laajempien asiakkaisiin liittyvien ilmiöiden, yhtäläisyyksien ja eroavaisuuksien analysoinnin ja sen pohjalta tehtävän asiakassegmentoinnin ohella.

Tutkimuksen kirjallisuuskatsaus avasi useita eri kanavia, jotka kerryttävät dataa, joita ei nykyisellään vielä hyödynnetä asiakasymmärryksen lisäämiseen. Massadatan todellinen potentiaali asiakasanalytiikalle onkin syvällisemmän asiakasymmärryksen saavuttaminen ja sen hyödyntäminen kohdennetussa markkinoinnissa ja asiakassegmentoinnissa. Haastattelututkimus tukee teoriassa tehtyjä havaintoja siitä, että asiakasanalytiikka voidaan tehostaa hyödyntämällä uusia kanavia, joilla on asiakasanalytiikalle annettavaa, mutta jotka eivät ole vielä yleisesti mukana asiakasdatan kerryttämisessä – massadatan perusluonteen määritellyt merkittävä volyyymi ja monimuotoisuus sitoutuvat vahvasti yhteen muun muassa verkkokäyttäytymisen, trendien, mediaseurannan sekä ulkopuolisen datan hyödyntämisen kanssa, jolla massadatan avulla voidaan tehostaa asiakkaista, sekä olemassa olevista että potentiaalisista, saatavaa ymmärrystä.

Tutkimuksen perusteella tärkein osa-alue massadatan hyödyntämisessä asiakasanalytiikassa olisikin yrityksen tavoitteisiin sidottujen datalinjausten luominen ja datan tehokas hyödyntäminen jatkuvina, merkityksellisinä käyttöönottoina yrityksen toiminnan tehostamiseksi sekä asiakkaille suunnattujen palvelujen kehittämiseksi. Kun asiakkaiden segmentointi suoritetaan tarkkailemalla heidän motivaatioitaan ja ideaalejaan demografisiin tekijöihin liitettynä, yrityksen on helpompi ennustaa muutoksia kohderyhmissä asiakasanalytiikan ja asiakkaisiin liittyvien ilmiöiden, kuten trendien ja arvojen muutosten avulla. Tätä kautta yrityksellä olisi myös mahdollisuus jatkuvasti kohdentaa ja kehittää palvelujaan ja markkinointiaan eri kohderyhmille sopiviksi. Massadatan hyödyntämisellä asiakasanalytiikassa on suuri potentiaali luoda lisäarvoa yrityksille, kunhan sen hyödyntämisessä keskitytään pelkkien teknologioiden valjastamisen sijaan asiakkaan syvällisemmälle ymmärtämiselle ja kattavan asiakassegmentoinnin toteuttamiselle merkillisen asiakasdatan määrittämiseen.

8.2 Jatkotutkimusaiheita

Yleisesti tarkasteltaessa massadatan hyödyntämistä asiakasanalytiikassa jatkotutkimuksien kannalta olisi tärkeää saada yhtenäinen, yleisesti hyväksytty määritelmä sille, mitä massadata on. Nykyisessä määritelmässään, jossa sen määritelmä ei sisällä konkreettisia, määrällisiä rajoitteita sen määrittämiseksi, on vaikea sitoa olemassa olevia käytänteitä, prosesseja tai liiketoimintatiedon hallinnan ja analytiikan -ratkaisuja massadatan hyödyntämiseen asiakasanalytiikassa. Jos jatkotutkimuksissa massadatalle voitaisiin todeta selkeämpi määritelmä, kuten datan määrään, strukturoimattomuuteen tai kertymisnopeuteen liittyviä rajoitteita, joiden täyttymisen jälkeen data luetellaan massadatakksi, siihen vaadittavia ratkaisuja olisi mahdollista määritellä konkreettisimmin. Massadatan tarkempi määritelmä mahdollistaisi myös yritysten ja yhteisöjen tarkastella paremmin kerjättyänsä datan luonnetta sekä sen mahdollisesti vaatimia uusia ratkaisuja. Yksittäisesti esimerkiksi datan määrään tai monimuotoisuuteen löytyy ratkaisuja pilvivarastoinnin ja erilaisten liiketoiminnan analytiikan työkalujen avulla, mutta yrityksille, joiden data vastaa massadatan 4V-tekijöitä niiden ongelma-alueiden kautta voisi olla ajankohtaisempaa hyödyntää suoraan massadatalle tarkoitettuja kokonaisratkaisuja yksittäisten, hajautettujen ratkaisujen sijasta.

Myös massadatan hyödyntämiseen ja tehokkaaseen rikastamiseen tiedoksi saatavilla olevista ratkaisuista tulisi tehdä enemmän sekä esittelevää että vertailevaa tutkimusta tulevaisuudessa. Tällaisen tutkimuksen vähäisyys ja teknologiapuolelle painottuneisuus on ymmärrettävissä aiheeseen vasta viime vuosina kohdistuneen mielenkiinnon ja kehittyneiden teknologisten ratkaisujen vuoksi, mutta tälle aihealueelle keskittyvä tutkimus olisi erityisen hyödyllinen yrityksille, jotka ovat kiinnostuneita valjastamaan massadatan suoraan johonkin konkreettiseen käyttötarkoitukseen yrityksessä helposti ja tehokkaasti. Esimerkiksi juuri asiakasanalytiikkaa ja sen tehostamista harkitseville organisaatioille tulisi olla saatavilla puolueetonta tutkimusta, joka helpottaisi mahdollisesti muun muassa

massadatalle suunniteltujen ratkaisujen eroavaisuuksia perinteisiin liiketoiminnan hallinnan ja analytiikan ratkaisuihin.

Lisäksi haastattelututkimuksessa poikkeukselliseksi osa-alueeksi nousut datan yhtenäisyys käytetyistä ratkaisuista toiseen siirtyessä tulisi olla tarkemmin käsitelty jatkotutkimuksissa, kuten myös keinot siihen liittyvien ongelmien välttämiseksi erityisesti asiakasanalytiikkaa tutkittaessa.

Kun taas tutkimusta tarkastellaan tapausyrityksen kannalta, eli energiapalveluita tuottavan yrityksen asiakasanalytiikkaa ajatellen, alaa koskien tulisi tehdä enemmän tutkimusta juuri asiakasanalytiikkaan liittyen. Koska alan kohderyhmien ostokäyttäytyminen eroaa suuresti yleisesti asiakasanalytiikkaa käsittelevissä tutkimuksissa käytetyistä ostomalleista ja teorioista, yleistä kirjallisuutta asiakasanalytiikassa voi hyödyntää alalla heikosti.

LÄHTEET

- Accenture (2013). "Analytics in Action: Breakthroughs and Barriers on the Journey to ROI." Haettu 03.04.2016 osoitteessa: https://www.accenture.com/fi-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Technology_6/Accenture-Analytics-In-Action-Survey.pdf.
- Ayankoya, K., Calitz, A. & Greyling, J. (2014). Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication. In *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology (SAICSIT '14)*, ACM, New York, NY, USA, , Pages 192 , 7 pages.
- Benjamins, V. R. (2014). Big Data: from Hype to Reality? In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14) (WIMS '14)*. ACM, New York, NY, USA.
- Chen H-M., Kazman, R., Haziyevev, S. & Hrytsay, O. (2015). Big data system development: an embedded case study with a global outsourcing firm. In *Proceedings of the First International Workshop on BIG Data Software Engineering (BIGDSE '15)*. IEEE Press, s. 44-50. Piscataway, NJ, USA.
- Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, 30(1), s. 44-57.
- Cuzzocrea, A., Bellatreche, L. & Song, I-Y. (2013). Data warehousing and OLAP over big data: current challenges and future research directions. In *Proceedings of the sixteenth international workshop on Data warehousing and OLAP (DOLAP '13)*. ACM, New York, USA, s. 67-70.
- Davenport, T. H. (2006). Competing on Analytics. *Harvard Business Review*.
- Davenport, T. H., Harris, J. G. & Morison, R. (2010). Analytics at Work : Smart Decisions, Better Results. *Harvard Business Press*, Boston, Massachusetts.
- Deloitte. (2012). 5 Reasons Analytics Projects Can Fail. *CIO Journal*. Haettu 3.4.2016 osoitteesta: <http://deloitte.wsj.com/cio/2012/06/06/5-reasons-analytics-projects-can-fail/>
- Franks, B. (2012). Taming the Big Data Tidal Wave - Finding Opportunities in Huge Data Streams with Advanced Analytics. *John Wiley & Sons*.
- Fernandes, S. & Bernardino, J. (2015). What is BigQuery? In *Proceedings of the 19th International Database Engineering & Applications Symposium (IDEAS '15)*. ACM, New York, USA, s. 202-203.
- Gartner (2012). Big Data. Haettu 23.01.2016 osoitteessa <http://www.gartner.com/it-glossary/big-data>. 22.12.2015.
- Haider, P., Chiarandini, P. and Brefeld, U. (2012). Discriminative clustering for market segmentation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12)*. ACM, New York, NY, USA, s. 417-425.
- Hirsjärvi, S., & Hurme, H. (2000). Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö. Helsinki: Yliopistopaino.

- Hirsjärvi, S., Remes, P. & Sajavaara, P. (2009). Tutki ja kirjoita (15. uud. painos). Helsinki: Tammi.
- IBM (2011). IBM Annual Report. Haettu 20.12.2015 osoitteesta: https://www.ibm.com/annualreport/2011/bin/assets/2011_ibm_annual.pdf
- IBM.com (2015). Asiakasanalytiikka. Haettu 21.12.2015 osoitteesta: <https://www-01.ibm.com/software/fi/analytics/rte/an/customer-analytics/>
- IBM.com (2016). What is Watso Analytics? Haettu 3.4.2016 osoitteesta: http://www.ibm.com/analytics/watson-analytics/?cm_mmc=search-gsn-ROW-_branded-watson-_ibm%20data%20analytics-Broad-_NDC-WW-WA-mkt-oww
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R. & Shahabi, C. (2014). Big data and its technical challenges. *Commun. ACM* 57, 7 (July 2014), s. 86-94.
- Jyvaskylanenergia.fi, (2016a). Sähkön hinta. Haettu 02.04.2016 osoitteesta: http://www.jyvaskylanenergia.fi/sahko/sahkon-hinta?gclid=Cj0KEQjwoYi4BRDF_PHHu6rI7NMBEiQAKZ-JuJ_yasGbwTiozwWfalqZuKL-vpuDfANQITNMqzsUOVYaAral8P8HAQ
- Jyvaskylanenergia.fi, (2016b). Sähkön jakeluyhtiöt Jyväskylässä. Haettu 24.11.2016 osoitteesta: http://www.jyvaskylanenergia.fi/filebank/466-Jyvaskylan_sahkonjakeluverkkojen_haltijoiden_vastuualueet_02_2013.pdf
- Kolyshkina, I. & Simoff, S. (2007). Customer analytics projects: addressing existing problems with a process that leads to success. In *Proceedings of the sixth Australasian conference on Data mining and analytics volume 70*, s. 13-19. Darlinghurst, Australia.
- Kotler, P & Armstrong, G. (2010). Principles of Marketing. *Pearson Education, 9th Edition*. s. 5. Haettu 22.12.2015 osoitteesta: http://202.74.245.22:8080/xmlui/bitstream/handle/123456789/61/Kotler_Principles%20of%20marketing%209th%20edition.pdf
- Kowsar, Y., Dashnow, H. & Lonie, A. (2014). Data Interlocking: Coupling Analytics to the Data. In *Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing (UCC '14)*. IEEE Computer Society, Washington, DC, USA, s. 696-701.
- Krishnamoorthi, S. & Mathew, S. (2015). Business Analytics and Business Value: A Case Study. *Thirty Sixth International Conference on Information Systems, Fort Worth 2015*.
- Labrinidis, A. & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proc. VLDB Endow.* 5, 12 (August 2012), s. 2032-2033.
- Lao, G. & Tang, Y. (2005). The Application of Data Warehousing in E-business Environment and Case Study. *ICEC'05*, Elokuu 15-17, s. 815-817.
- LeFevre, J., Sankaranarayanan, J., Hacigumus, H., Tatemura, J., Polyzotis, N. & Carey, M. J. (2014). MISO: souping up big data query processing with a multistore system. *SIGMOD '14: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*. ACM, New York, NY, USA, s. 1591-1602.

- Lim, E-P., Chen, H. & Chen, G. (2013). Business Intelligence and Analytics: Research Directions. *ACM Trans. Manage. Inf. Syst.* 3, 4, Article 17 (January 2013), s. 2032-2041.
- Lincoln, YS. & Guba, EG. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.
- Loshin, D. & Reifer, A. (2013). Using Information to Develop a Culture of Customer Centricity: Customer Centricity, Analytics, and Information Utilization. *Morgan Kaufmann Publishers Inc.*, San Francisco, CA, USA.
- Lynn, M. (2011). Segmenting and targeting your market: Strategies and limitations. *Cornell University, School of Hospitality Administration* Haettu 22.12.2015, osoitteesta: <http://scholarship.sha.cornell.edu/articles/243>
- Manyika, J., Chui, M., Brown, B., Bughin J., Dobbs, R. Roxburgh, C. & Byers A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, Toukokuu 2011.
- Marton, A., Avital, M. & Jensen, T. B. (2013). Reframing Open Big Data. *ECIS 2013 Completed Research*. Paperi 146.
- Miao, X. (2014). Big Data and Smart Grid. In *Proceedings of the 2014 International Conference on Big Data Science and Computing (BigDataScience '14)*. ACM, New York, NY, USA.
- Mulvenna, M., Norwood M. & Büchner, A. (1998). Data-Driven Marketing. *NetAcedamy.com*.
- Nakucçi, E., Theodorou, V., Jovanovic, P. & Abelló, A. (2014). Bijoux: Data Generator for Evaluating ETL Process Quality. In *Proceedings of the 17th International Workshop on Data Warehousing and OLAP (DOLAP '14)*. ACM, New York, NY, USA, s. 23-32.
- Nikulainen, T. (2013). Big Data Revolution – What Is It? *ETLA Brief No 10*. Haettu 28.2.2016 osoitteesta: <http://pub.etla.fi/ETLA-Muistio-Brief-10.pdf>
- Peng, Z. & Yuanyuan, H. (2008). The method for solving two types of errors in customer segmentation on unbalanced data. In *Proceedings of the 10th international conference on Electronic commerce (ICEC '08)*. ACM, New York, NY, USA, Artikkel 16.
- Power, D. (2013). Using Big Data for Analytics and Decision Support. *MWAIS 2013 Proceedings*.
- Reed, D.A & Dongarra, J. (2015). Exascale computing and Big Data. *Communications of the ACM*, 58(7). Kesäkuu 2015.
- Saaranen-Kauppinen, A. & Puusniekka, A. (2006). KvaliMOTV - Menetelmäopetuksen tietovaranto. *Tampere: Yhteiskuntatieteellinen tietoaarkisto*. Haettu 7.5.2017 osoitteesta: <http://www.fsd.uta.fi/menetelmaopetus/>
- Sicilia, M-Á & García, E. (2003). On fuzziness in relationship value segmentation: applications to personalized e-commerce. *SIGecom Exch.* 4(2), s. 1-10.
- Singh, A., Rumantir, G., South, A. & Bethwaite, B. (2014). Clustering Experiments on Big Transaction Data for Market Segmentation. In *Proceedings of the 2014 International Conference on Big Data Science and Computing (BigDataScience '14)*. ACM, New York, NY, USA.

- Sodenkamp, M., Kozlovskiy, I. & Staake, T. (2015). Gaining IS Business Value through Big Data Analytics: A Case Study of the Energy Sector. *AIS Electronic Library (AISeL) - ICIS 2015 Proceedings*.
- Swinyard, W. R., & Struman, K. D. (1986). Market Segmentation: Finding the Heart of Your Restaurant's Market. *Cornell Hotel and Restaurant Administration Quarterly*, 37(1), s. 88-96.
- Szymczak, S., Zelik D.J. & Elm, W. (2014). Support for Big Data's Limiting Resource - Human Attention. *HCBDR '14*, 01. - 03. Huhtikuuta, 2014, Raleigh, NC, USA.
- T-Media. (2015). Kansan arvot 2015 -tutkimus. Haettu 4.4.2016 osoitteesta: <http://www.t-media.fi/kansan-arvot-2015/>
- Tanner, J. F. (2014). Analytics and Dynamic Customer Strategy: Big Profits from Big Data. *John Wiley & Sons, Inc, Hoboken, New Jersey*.
- Thomson, R., Lebiere, C. & Bennati, S. (2014). Human, Model and Machine: A Complementary Approach to Big Data. In *Proceedings of the 2014 Workshop on Human Centered Big Data Research (HCBDR '14)*. ACM, New York, NY, USA, s. 27-31.
- Ullman, J. (2015). MapReduce Algorithms. In *Proceedings of the 2nd IKDD Conference on Data Sciences (CODS-IKDD '15)*. ACM, New York, NY, USA.
- Vadivu, P. S. & David, V. K. (2012). Optimized feature extraction and actionable knowledge discovery for Customer Relationship Management (CRM). In *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology (CCSEIT '12)*. ACM, New York, NY, USA, s. 275-282.
- Wang, F., Qiu, J., Yang, J., Dong, B., Li, X. & Li, Y. (2009). Hadoop high availability through metadata replication. In *Proceedings of the first international workshop on Cloud data management (CloudDB '09)*. ACM, New York, USA. s. 37-44.
- Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Kulkarni, U., Mooney, B. J. G., Phillips-Wren, G. & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, 34(1), s. 1-13.
- Wu, J. & Lin, Z. (2005). Research on customer segmentation model by clustering. In *Proceedings of the 7th international conference on Electronic commerce (ICEC '05)*. ACM, New York, NY, USA, s. 316-318.
- Yuan, Y-L. & Ho, C-I. (2015). Rethinking the Destination Marketing Organization Management in the Big Data Era. In *Proceedings of the ASE BigData & SocialInformatics 2015 (ASE BD&SI '15)*. ACM, New York, NY, USA.

LIITE 1. TEEMAHAASTATTELURUNKO

1. Haastateltavan historia yrityksessä

- Kuinka kauan aikaa ollut töissä yrityksessä?
- Millä osastoilla ja toimenkuvissa?
- Kuinka kauan aikaa työskennellyt nykyisessä toimenkuvassaan?

2. Datan hyödyntämismallin toteutuminen ja kerätty asiakasanalytiikka

- Datan kerääminen: "Mitä kaikkea dataa asiakkaisiinne liittyen keräätte tällä hetkellä?"
- Datan varastointi: "Millaisia datan varastointiratkaisuja teillä on käytössä? Löytyykö teiltä kaikki tieto oman talon sisältä?"
- Datan yhtenäistäminen ja koostaminen: "Millä tavalla te käsittelette dataa ennen sen mallinnusta tai muuta tulkittavaan muotoon muuntamista?"
- Datan mallinnus ja analysointi: "Minkälaisia työkaluja ja metodeja te tavallisesti käytätte datan analysointiin?"
- Datan tulosten tulkinta: "Millä tavalla teillä tarkastellaan uutta dataa suhteessa vanhasta datasta tehtyihin päätöksiin?"
- Datan käyttöönotto: "Kuinka tehokkaasti teillä tehdään muutoksia toimintaan ja perinteiseen prosesseihin uuden datan pohjalta?"

3. Kokemukset ja havainnot massadataan liitettävien 4V-mallin tekijöistä

- Onko teillä esiintynyt ongelmia datan volyymin kasvamisesta esimerkiksi sen varastoinnissa ja käsittelyssä?
- Onko teillä esiintynyt ongelmia datan monimuotoisuuden ja rakenteettomuuden vuoksi?
- Onko teillä esiintynyt ongelmia datan kertymisnopeuden vuoksi?
- Onko teillä esiintynyt ongelmia olennaisen arvon löytämisessä keräämästänne ja käsittelemästänne datasta?