

Emil Vartiainen

**MASSADATAN VAIKUTUKSET YRITYSTEN
PÄÄTÖKSENTEKOON**



JYVÄSKYLÄN YLIOPISTO
TIETOJENKÄSITTELYTIETEIDEN LAITOS
2017

TIIVISTELMÄ

Vartiainen, Emil

Massadatan vaikutukset yritysten päätöksentekoon

Jyväskylä: Jyväskylän yliopisto, 2017, 29 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Luoma, Eetu

Teknologinen kehitys on mahdollistanut ympärillämme suurella nopeudella syntyvän, monimuotoisien ja määrällisesti valtavan massadatan keräämisen ja hyödyntämisen. Yritykset eivät kuitenkaan pysty hyödyntämään kaikkea käytettävissä olevaa dataa. Tässä kirjallisuuskatsauksena toteutetussa tutkielmassa tarkasteltiin massadatan käyttöä yritysten päätöksenteon näkökulmasta. Tarkoituksena oli tarkastella massadatan vaikutuksia yritysten päätöksentekoon. Lisäksi käsiteltiin massadata-analytiikkaa, sekä massadatan tuomia mahdollisuuksia ja haasteita yrityksille yleisellä tasolla. Tutkielmassa käytettiin vertaisarvioituja tieteellisiä artikkeleita, jotka ovat pääosin kirjoitettu 2010-luvulla. Tutkielman keskeisimpänä tuloksena havaittiin, että massadatalla on useita vaikutuksia yritysten päätöksentekoon. Massadatan avulla tehdyt päätökset nähtiin paremmiksi päätöksiksi, kuin perinteisin menetelmin tehdyt. Eniten hyötyjä syntyy, kun massadatan avulla automatisoidaan usein toistuvia päätöksiä. Massadatan käyttö päätöksentekovaiheessa edellyttää yrityksen johdon tukea ja datan käyttöön perustuvaa yrityskulttuuria. Tuloksista voidaan päätellä, että yritysten kannattaa käyttää dataa mahdollisimman paljon päätöksentekonsa tukena ja lisäksi muovata yrityskulttuuria dataohjautuvampaan suuntaan.

Asiasanat: massadata, big data, massadata-analytiikka, dataohjautuva päätöksenteko

ABSTRACT

Vartiainen, Emil

Big data's impacts on organizational decision making

Jyväskylä: University of Jyväskylä, 2017, 29 p.

Information Systems, Bachelor's Thesis

Supervisor: Luoma, Eetu

Technological development in recent years has enabled collecting and analyzing high velocity and large volume big data which comes from variety of sources. However, organizations lack to benefit from all the data available. This study was carried out as literature review. Its objective was to find how big data impacts on organizational decision making. In addition, this paper contains review of concept of big data, big data analytics and big data's opportunities and challenges to organizations in general. The literature used was peer evaluated research papers mostly written in 2010's. The key finding of this paper was that big data has several impacts on organizational decision making. Decisions made by using big data were considered better decisions than decisions based on intuition. The most beneficial is to automate repetitive decisions. Using big data in decision making requires support from top management and data-driven organizational culture. Conclusion from the findings is that organizations should use big data to support their decision making and also shape their organizational culture more data-driven way.

Keywords: big data, big data analytics, data-driven decision making

KUVIOT

KUVIO 1 Massadatan kolme V:tä (Russom, 2011)	9
KUVIO 2 Prosessi tietämyksen erottamiseksi massadatasta (Gandomi & Haider, 2015).....	13
KUVIO 3 Liiketoimintatiedon hallinnan suhde muihin järjestelmiin (Negash, 2004).....	22

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT

1	JOHDANTO.....	6
2	MASSADATA.....	9
	2.1 Massadatan käsite.....	9
	2.2 Massadata-analytiikka	11
	2.3 Massadatan mahdollisuuksia	14
	2.4 Massadatan haasteita	16
3	MASSADATA YRITYSTEN PÄÄTÖKSENTEOSSA	19
	3.1 Päätöksentekoprosessi	19
	3.2 Päätöksentekoa tukevat järjestelmät	20
	3.3 Massadatan vaikutukset yritysten päätöksentekoon	22
4	YHTEENVETO	25
	LÄHTEET	28

1 JOHDANTO

Ympärillämme syntyy jatkuvasti valtavat määrät uutta dataa. Jokainen meistä osallistuu datan tuottamiseen joka päivä lähettämällä sähköposteja, lataamalla valokuvia sosiaaliseen mediaan, selailemalla uutisia verkkosivustoilla, asioimalla ruokakaupassa, maksamalla laskuja verkkopankissa tai tekemällä lähes mitä tahansa (Power, 2013). Tallennettua muotoa tästä datasta kutsutaan massadatakseksi (eng. big data). Tämän tutkimuksen tarkoituksena on selvittää massadatan vaikutuksia yritysten päätöksentekoon.

Massadatalle ominaista on sen suuri määrä, monimuotoisuus ja synty nopeus. Nämä ominaisuudet muodostavat yleisimmän käytössä olevan määritelmän massadatalle. Edellä mainittujen ominaisuuksien lisäksi joihinkin määritelmiin on lisätty todenmukaisuus, vaihtelevuus ja arvo. Yhtenäistä massadatan määritelmää ei kuitenkaan ole olemassa. (Gandomi & Haider, 2015.) Kuitenkin yleisin seikka, joka erottaa massadatan tavallisesta datasta, on sen rakenne. Tavallinen data on strukturoitua, ennalta määritetyn muotoista ja tallentuu tietokantoihin sellaisenaan. Tavallista dataa pystytään yleensä prosessoimaan käsittelemättä sitä. Suurin osa massadatasta on strukturoimatonta, kuten videoita, kuvia tai sosiaalisen median syötteitä. Tällaisen datan käsittely on haastavaa, eikä siitä ei voida suoraan tehdä hakuja sen sisällön perusteella. (McAfee & Brynjolfsson, 2012.) Gandomi ja Haider (2015) esittävät strukturoimattoman datan osuuden kaikesta datasta olevan jopa 95 %.

Yritykset ovat jo kymmeniä vuosia käyttäneet erilaisia järjestelmiä analysoimaan dataa ja käyttämään saamaansa informaatiota päätöksenteon tukena. Viimeisimpänä käytössä ovat olleet liiketoimintatiedon hallintajärjestelmät, eli tutummin kutsuttuna business intelligence -järjestelmät. Nämä järjestelmät eivät kuitenkaan ole kykeneväisiä käsittelemään massadataa sen ominaisuuksien vuoksi (Provost & Fawcett 2013; Gandomi & Haider, 2015). Niinpä onkin täytynyt kehittää uusia teknologioita hallitsemaan ja käsittelemään massadataa. Walker (2015) esittää, että massadata on muunnettavissa taloudelliseksi hyödyksi. Hyödyn saaminen massadatasta on kuitenkin monivaiheinen prosessi. Prosessin

vaiheet ovat datan keruu, datan puhdistaminen ja informaation erottaminen, datan yhtenäistäminen, mallintaminen ja analysointi, sekä tulkinta. (Jagadish ym., 2014.)

Yrityksillä on valtavat määrät omaa, sekä julkista dataa käytettävissään. Lisäksi jatkuvasti syntyy uusia datalähteitä hyödynnettäväksi (Chen, Chiang & Storey, 2012). Haasteena yrityksillä on muuntaa tämä data informaatioksi ja sitä kautta hyödyksi päätöksentekovaiheessa (Gandomi & Haider; 2015, Power 2013). Massadata tarjoaa yritysten eri osille mahdollisuuden kehittää toimintaansa tai jopa tarjota asiakkailleen täysin uudenlaisia palveluita. McAfee ja Brynjolfsson (2012) esittävät massadataa hyödyntävien yritysten olevan paljon ketterämpiä, kuin heidän kilpailijoidensa. Gandomi ja Haider (2015) taas sanovat massadatatknologioita käyttävien yritysten pystyvän reaaliaikaiseen datan hyödyntämiseen suuristakin datamassoista. Myös useat case-tutkimukset osoittavat massadataa hyödyntämällä suuren lisäarvon olevan saatavilla yrityksille (Jagadish ym., 2014). Edellä mainituista seikoista johtuen yritykset ovatkin erittäin kiinnostuneita tavoista hyödyntää entistä suurempia ja monimuotoisempia datamassoja.

Yrityksillä on joka päivä useita suuria ja vielä useampia pienempiä päätöksiä tehtävänä. Monet näistä päätöksistä ovat rahalliselta arvoltaan todella merkittäviä ja vaikuttavat useisiin henkilöihin niin yrityksen sisällä, kuin sidosryhmissäkin. Niinpä yrityksillä on halu tehdä paras mahdollinen päätös. McAfee ja Brynjolfsson (2012) esittävät massadatan analysoinnin olevan tehokkaampaa kuin päätöksentekomenetelmien, joita käytettiin aiemmin. Päätökset, jotka tehtiin aiemmin arvaamalla paras vaihtoehto kokemusten perusteella, voidaan nykyisin tehdä käyttämällä matemaattisia datamalleja (Jagadish ym., 2014). Tarvitaan kuitenkin uudenlaista osaamista, jotta massadataa hyödyntävää päätöksentekoa voidaan käyttää yrityksissä.

Tässä tutkielmassa selvitettiin yritysten massadatan käyttöä päätöksenteon näkökulmasta. Aihetta oli mielekästä tutkia, koska massadatan vaikutuksia päätöksentekoon on suurimmaksi osaksi ainoastaan sivuttu kirjallisuudessa. Aiheesta ei myöskään ollut suomenkielistä tutkimustietoa saatavilla. Lisäksi, tutkielmassa tutustuttiin massadatan tuomiin mahdollisuuksiin ja haasteisiin yrityksille yleisellä tasolla. Tutkielman tavoitteena oli vastata tutkimuskysymyksiin:

- Mitä vaikutuksia massadatalta on yritysten päätöksentekoon?

Apukysymyksinä olivat:

- Miten massadataa analysoidaan?
- Mitä mahdollisuuksia ja haasteita massadataan liittyy?

Tutkielma toteutettiin kirjallisuuskatsauksena. Tutkielman lähteet haettiin pääosin Google Scholar-hakukoneesta, sekä AISEL-tietokannasta hakusanoilla "big data", "big data analytics", sekä "data-driven decision making". Tutkielmaan valittiin noin 30 artikkelia hakusanojen osuvuuden, siteerausmäärien, sekä Julkaisufoorumin julkaisukanavahaun arvosanan perusteella. Tämän lisäksi

käytettiin muutamia ulkopuolisia lähteitä. Tutkielma kirjoitettiin käyttäen Okolin ja Schbramin (2010) esittelemää menetelmää.

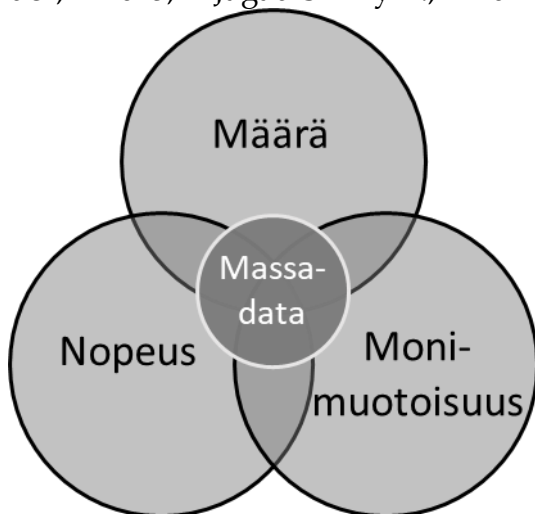
Tutkielma rakentuu johdantoluvusta, kahdesta sisältöluvusta ja yhteenvetoluvusta. Ensimmäisessä sisältöluvussa käsitellään massadataa ja sen käsitettä. Lisäksi avataan massadata-analytiikan prosessia ja käsitellään massadatan tuomia mahdollisuuksia ja haasteita. Ensimmäisessä sisältöluvussa vastataan tutkielman apukysymyksiin: "Miten massadataa analysoidaan?" ja "Mitä mahdollisuuksia ja haasteita massadataan liittyy?" Toisessa sisältöluvussa käsitellään massadataa yritysten päätöksentekomenetelmien ja -prosessien näkökulmista. Toisessa sisältöluvussa vastataan tutkimuskysymykseen "Mitä vaikutuksia massadatalta on yritysten päätöksentekoon?" Viimeisenä tutkielmassa on yhteenvetoluku, jossa teen yhteenvedon käsitellyistä aiheista, esitän tutkimuksen tulokset ja omat johtopäätökseni, sekä nostan esille kiinnostavia jatkotutkimusaiheita.

2 MASSADATA

Meidän jokaisen arkipäiväiset toimet synnyttävät valtavan määrän dataa, joka tallentuu jossain muodossa digimaailman syövereihin. Puhelimet, verkon selaaminen, sosiaalinen media, pikaviestisovellukset, GPS-anturit ja esineiden internet (eng. IoT, Internet of Things) tuottavat dataa osana niiden normaalia toimintaa (McAfee & Brynjolfsson, 2012). Eikä datan määrän kasvu ole hidastumassa, vaan ennemminkin dataa oletetaan syntyvän koko ajan enemmän ja nopeammin (Gao, Koronios & Selle, 2015; Fan & Bifet, 2013). Syntyvää dataa kutsutaan massadatakksi (eng. big data). Massadatan uudenlaiset ominaisuudet erottavat sen perinteisestä datasta. Massadatalla ei siis tarkoiteta ainoastaan dataa, joka on kooltaan suurta.

2.1 Massadatan käsite

Massadatalle ei ole kirjallisuudessa osoitettu yhtä ja ainoaa määritelmää (Gandomi & Haider, 2015; Ward & Barker, 2013). Ylivoimaisesti käytetyin määritelmä on niin sanottu kolmen V:n malli (kuvio 1). Kolme V:tä muodostuvat englanninkielisistä sanoista määrä (eng. volume), monimuotoisuus (eng. variety) ja nopeus (eng. velocity) (McAfee & Brynjolfsson, 2012; Fan & Bifet, 2013; Gandomi & Haider, 2015; Jagadish ym., 2014; Wu, Zhu, Wu & Ding, 2014).



KUVIO 1 Massadatan kolme V:tä (Russom, 2011)

Datan määrällä tarkoitetaan datan fyysistä määrää. Massadatalle ominaista on sen valtava koko. Massadatan määrästä puhuttaessa numero- tai tavumäärien luetteleminen ei ole mielekäästä, sillä mikä tänä päivänä suurta, ei välttämättä ole sitä enää tulevaisuudessa. Osumammin määrää kuvaa sen olevan niin suurta, että perinteiset tietojenkäsittelytyökalut eivät pysty sitä käsittelemään. (Gandomi & Haider, 2015.)

Datan monimuotoisuudella tarkoitetaan datan olevan peräisin useista erilaisista lähteistä, sekä tallentuneen useisiin eri tiedostomuotoihin. Data voi olla rakenteeltaan strukturoitua, osittain strukturoitua tai strukturoimatonta (Gandomi & Haider, 2015; Power, 2013; Kambatla, Kollias, Kumar & Grama, 2014). Perinteinen data on strukturoitua, ennalta määrätyn muotoista, kuten relaatio-tietokannat tai datatietueet (Manyika ym., 2011). Strukturoituun dataan on myös helppo tehdä hakuja, koska tiedot ovat määrämuotoisia (Sagiroglu & Sinanc, 2013). Kaikesta datasta kuitenkin arviolta 95% on strukturoimatonta dataa, kuten valokuvia, äänitiedostoja, videoita tai tekstiä (Gandomi & Haider, 2015). Strukturoimatonta dataa ei voida tallentaa perinteisiin tietovarastoihin, eikä sitä voida analysoida perinteisin menetelmin. Osittain strukturoitu data on strukturoimatonta dataa, joka sisältää jonkinlaisia tunnistetietoja, joiden avulla dataa on mahdollista prosessoida (Manyika ym., 2011). Esimerkki tällaisesta datasta on XML-tekstitiedostot, jossa teksti on luokiteltu sisällön mukaan.

Datan nopeudella tarkoitetaan sekä sen syntynopeutta, että analysointinopeutta (Gandomi & Haider, 2015; Jagadish ym., 2014; Sagiroglu & Sinanc, 2013; Fan & Bifet, 2013). Esimerkiksi sensorit tuottavat dataa jatkuvalla syötöllä, jopa useita kertoja sekunnissa. Tällaisen datan välitön analysointi on tärkeää, koska jos datan analysoinnissa aikailaan, on uutta dataa kerennyt syntyä jo valtava määrä ja reagointia vaatinut data on jo kerennyt vanhentumaan. Reaaliaikainen tai lähes reaaliaikainen informaation analysointi tekee yrityksestä ketterämmän kuin heidän kilpailijansa (McAfee & Brynjolfsson, 2012; Corte-Real, Oliveira & Ruivo, 2016). Lisäksi Sagiroglu ja Sinanc (2013) esittävät massadatan välittömän analysoinnin maksimoivan yritysten datasta saatavan arvon.

Vaikka nämä kolme ominaisuutta esiintyvät useimmissa massadatan määritelmässä, yleistä mittaria, joka määrittäisi massadatan määrän, monimuotoisuuden tai nopeuden mukaan, ei ole olemassa (Gandomi & Haider, 2015). He tyytyvät toteamaan, että kun yksi ominaisuus kasvaa, niin kasvaa myös todennäköisyys muiden kasvamiselle. Russom (2011) toteaa ainakin monimuotoisuuden ja määrän lisääntymisen korreloivan toistensa kanssa.

Kolmen V:n mallin ollessa yleisin massadatan käytetty määritelmä, myös muita määritelmiä on esitelty. Yleisimmät näistä laajentavat kolmen V:n mallia lisäämällä siihen uusia ominaisuuksia. Yleisimmät näistä lisätyistä ominaisuuksista ovat todenmukaisuus (eng. veracity), vaihtelevuus (eng. variability) ja arvo (eng. value).

Datan todenmukaisuudella tarkoitetaan joistain lähteistä kerätyn massadatan olevan laadullisesti epäluotettavampaa, kuin perinteisen datan (Gandomi & Haider, 2015). Esimerkiksi kuvasta päätelty tunnetila on epäluotettavampi, kuin kyselyllä selvitetty tekstimuotoinen vastaus. Näin ollen joissain tapauksissa massadata ei ole yhtä luotettavaa, kuin perinteinen data (Boyd & Crawford, 2012).

Datan vaihtelevuudella tarkoitetaan datan syntynopeuden vaihtelevuutta (Gandomi & Haider, 2015). Mikäli datan syntyminen on kiinnitetty johonkin tapahtumaan, esimerkiksi tavaratalossa liukuovien aukeamiseen, niin datan syntynopeus voi vaihdella todella paljon. Arki-iltapäivänä dataa syntyy jatkuvasti, mutta yöllä kun tavaratalo on suljettuna, ei dataa synny ollenkaan.

Datan arvolla tarkoitetaan datan sisältämää rahallista hyötyä (Gandomi & Haider, 2015; Fan & Bifet, 2013). Fan ja Bifet (2013) perustelevat näkemystään sillä, että massadataa hyödyntävät yritykset pystyvät tekemään päätöksiä vastaamalla kysymyksiin, joita ei ennen osattu esittää. Gandomi ja Haider (2015) taas esittävät, että arvo on saatavilla analysoimalla isoja määriä vähäarvoista dataa.

Davenport, Barth ja Bean (2012) esittävät eroavan näkemyksen yleisestä kolmen tai useamman V:n mallista määrittelemällä massadatan olevan liian suurta tai rakenteetonta hallittavaksi ja analysoitavaksi perinteisille järjestelmille. Provost ja Fawcett (2013) ovat samaa mieltä sanomalla, että massadatalla tarkoitetaan dataa, joka on liian suurta tai monimuotoista perinteisille järjestelmille ja sen vuoksi tarvitsevan uusia teknologioita sen käsittelemistä varten. Sagiroglu ja Sinanc (2013) yhtyvät näkemykseen toteamalla, että massadata on termi suurille datamassoille, jotka ovat monimuotoisia rakenteeltaan, ja joita on vaikea säilöä, analysoida tai visualisoida. Myös Fan ja Bifet (2012) allekirjoittavat tämän näkemyksen toteamalla massadatan olevan termi datamassoille, joita me emme voi hallita nykyisillä metodeilla tai työkaluilla niiden koon ja monimuotoisuuden vuoksi. Lisäksi Power (2013) toteaa massadatan olevan värikäs termi merkitykselliselle muutokselle datan keräyksessä, säilömisessä ja hyödyntämisessä.

2.2 Massadata-analytiikka

Jotta massadatatista saadaan hyötyä, täytyy dataa keräämisen lisäksi myös analysoida. Power (2013) toteaa massadatatista olevan hyötyä ainoastaan, jos sitä analysoidaan. McAfee ja Brynjolfsson (2012) esittävät massadata-analytiikan (eng. big data analytics) olevan informaation etsimistä datasta ja sen liiketoiminnallista hyödyntämistä. Russom (2011) määrittelee massadata-analytiikan yksinkertaisesti edistyksellisiksi tekniikoiksi massadatan käsittelemiseksi. Chen, Chiang ja Storey (2012) jatkavat samoilla linjoilla kuvaamalla massadata-analytiikan olevan tekniikoita ja sovelluksia datalle, joka vaatii laajuudellaan ja monimutkaisuudellaan uniikkeja datan varastointi-, hallinta-, analyysi- ja visualisointiteknologioita. Massadata-analytiikka voidaan myös nähdä iteratiivisena prosessina, jossa on viisi eri vaihetta. Vaiheet ovat: datan keruu, datan puhdistaminen ja informaation erottaminen, datan yhdentäminen, mallintaminen ja analysointi, sekä tulkinta. (Jagadish ym., 2014.) Gandomi ja Haider (2015) jakavat edellä mainitun prosessin vaiheet kahteen osaan: datan hallintaan ja datan analysointiin (kuvio 2). He toteavat kolmen ensimmäisen vaiheen olevan osa datan hallintaa ja kahden viimeisen analysointia. Fan ja Bifet (2013) eivät halua erottaa massadata-analytiikkaa tavallisesta data-analytiikasta, sillä datan määrä jatkaa kasvuaan myös tulevaisuudessa ja paluuta perinteiseen pienen datamäärän analysointiin ei enää ole.

Massadatan analysointiprosessi katsotaan alkavan datan keruusta. Kuten jo todettua, dataa syntyy nykyisin enemmän ja nopeammin kuin koskaan aikaisemmin (Fan & Bifet, 2013). Walker (2015) kuvailee jokaisen meistä olevan kävelevä

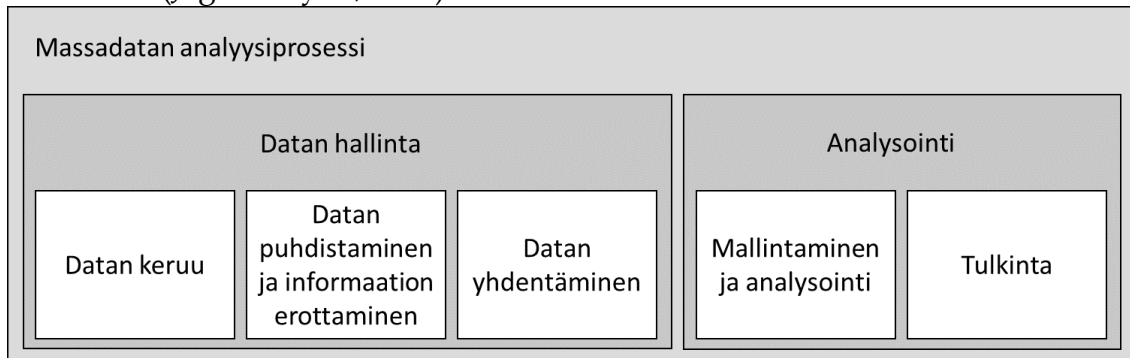
datalähde, jota joku seuraa ja kerää meistä syntyvän datan talteen hyödyntääkseen sitä tulevaisuudessa. Walker myös esittää digitalisaation mahdollistavan henkilökohtaisten asioiden, kuten ihmissuhteiden, kokemusten tai tunnetilojen dataksi muuntamisen. Datan voidaan katsoa olevan yrityksen sisäistä, kuten transaktiodata, tai ulkoista, kuten verkkoon julkaistavat sosiaalisen median päivitykset. Jos ulkoinen data on vapaata kaikkien saatavilla olevaa dataa, puhutaan avoimesta datasta. Li, Zhang ja Hu (2017) esittävät sisäisen datan olevan sisälöltään heikompaa verrattuna ulkoiseen dataan.

Yleensä yritykset aloittavat massadataprojektinsa keräämällä ja käsittelemällä kaikkea saatavissa olevaa dataa. Ongelmana tässä lähestymistavassa on resurssien hukkaaminen datalähteisiin, joista ei loppujen lopuksi ole edes hyötyä analyysivaiheessa. (LaValle, Lesser, Shockley, Hopkins & Kruschwitz, 2011; Jagadish ym., 2014.) Yritysten kannattaisi aloittaa prosessi määrittelemällä oikeat kysymykset, jotka vastaavat haluttuun liiketoiminnalliseen tarkoitukseen ja vasta sen jälkeen määritellä kerättävä data (LaValle ym., 2011). Toinen datan keräämiseen liittyvä ongelma on rajata analysoinnissa käytettävä aineisto. Koska datamassat ovat valtavia, niin jo teknologisen suorituskyvyn takia saatetaan joutua tekemään rajauksia dataan. Dataa kerätessä ei haluta poistaa tai rajata arvoja, jotka vaikuttaisivat datasta saatavaan informaatioon (Jagadish ym., 2014). Fan ja Bifet (2013) nostavat myös esiin kysymyksen uusista menetelmistä ja teknologioista datan keräämiseksi, sillä massadatan ominaisuudet määrä, monimuotoisuus ja nopeus tekevät siitä vaikeaa, jollei mahdotonta nykyisille työkaluille.

Datan keräämistä seuraa datan puhdistaminen ja informaation erottaminen. Koska massadata on kerätty useista eri lähteistä, joista osa on laadukkaampia ja luotettavampia kuin toiset, täytyy kerätty data puhdistaa virheistä. Mitä enemmän dataa on kerätty, sitä enemmän se sisältää virheitä (Walker, 2015). Labrinidis ja Jagadish (2012) toteavat kerätyn datan harvoin olevan valmista analysoitavaksi sellaisenaan. Koska kerätty data ei yleensä ole strukturoitua, on data ensin muunnettava käsiteltävään muotoon. Tämä vaihe on hyvin sovelluskohtainen, sillä strukturoimattomista lähteistä kerättävät merkitykselliset havainnot voivat vaihdella todella paljon datasta, sekä analyysin lähtökohdista ja tavoitteista riippuen. (Jagadish ym., 2014.) Gandomi ja Haider (2015) käyttävät esimerkkinä informaation erottamisesta tunnistavaa tekstinlukijaa, joka erottaa lääkärin määräämästä reseptistä automaattisesti lääkkeen nimen, annostuksen ja käyttöohjeen. Sovelluskohtaisuus käy esimerkistä hyvin ilmi, koska samaa menetelmää ei voida käyttää esimerkiksi säätilojen poimimiseen sääennustuksesta, vaikka analysoitava data onkin samankaltaista.

Kun data on kerätty ja puhdistettu, sekä informaatio erotettu, on vuorossa datan yhdentäminen. Tällä vaiheella tarkoitetaan useista eri lähteistä kerätyn datan yhdentämistä tai yhteen tuomista analysointia varten. Usein useista lähteistä kerätyllä ja analysoidulla datalla saadaan tarkempia tuloksia analysoimalla kutakin datalähdettä erikseen. (Manyika ym., 2011.) Useat lähteet auttavat saavuttamaan paremman kokonais kuvan analysoitavasta kohteesta, esimerkiksi potilastietojen yhdistäminen ympäristömuuttujiin auttaa lääkäreitä tulkitsemaan potilaan oireiden perisyyn (Jagadish ym., 2014). Jokaisen datalähteen täytyy auttaa

saavuttamaan vastaus massadatan etsittäväseen kysymykseen, sillä muuten datan käsitteleminen on pelkkää resurssien haaskausta (LaValle ym., 2011). Lähihistoriassa tietovarastot ovat olleet merkittävässä roolissa datan analysoinnin kannalta. Näiden toimintamahdollisuudet ovat kuitenkin rajalliset uudessa, useita datalähteitä ja -formaatteja sisältävässä massadatan maailmassa. (Gao ym., 2015.) Tietovarastojen suunnittelu onkin nykymaailmassa erittäin tärkeää, jotta data olisi tehokkaasti käytettävissä oikeaan aikaan, oikeassa paikassa ja oikeassa muodossa (Jagadish ym., 2014).



KUVIO 2 Prosessi tietämyksen erottamiseksi massadatatista (Gandomi & Haider, 2015)

Datan yhdentämisen jälkeen vuorossa on mallintaminen ja analysointi. Datan analysoinnin voi jakaa karkeasti kolmeen luokkaan: kuvaileva analytiikka (eng. descriptive analytics), ennakoiva analytiikka (eng. predictive analytics) ja ohjaileva analytiikka (eng. prescriptive analytics) (LaValle ym., 2011; Power, 2013). Kaikille näille on oma käyttötarkoituksensa. Power (2013) esittää kuvailevan analytiikan olevan historiadatan tarkastelua tilastollisin menetelmin, jonka avulla pyritään tekemään päätelmiä tulevaisuuteen. Power kutsuu tätä osa-aluetta liiketoimintatiedon hallitsemiseksi (eng. business intelligence). Massadatan vaikutukset tähän analytiikan muotoon, on tarjota yhä suurempia datamassoja analysoimista varten.

Ennakoivalla analytiikalla tarkoitetaan erilaisia tekniikoita, jotka ennustavat tulevaa historiallisen ja nykyisen datan perusteella (Gandomi & Haider, 2015). Power (2013) esittää ennakoivan analytiikan tarkoittavan skenaariomallien tekemistä historiadatan pohjalta. Massadata mahdollistaa näiden mallien olevan ajankohtaisempia ja tarkempia kuin koskaan aikaisemmin. Gandomi ja Haider (2015) nostavat esiin tarpeen uusille tilastollisille menetelmille, koska perinteiset menetelmät ovat suunniteltu osajoukoista tehtäviin johtopäätöksiin. Massadata mahdollistaa kuitenkin myös merkittävän, jopa koko joukkoa käsittelevän datan keräämisen. Mitä enemmän dataa on määrällisesti, sen tarkempi lopputulos saadaan tilastollisia menetelmiä käyttämällä (Russom, 2011).

Kaikkein kehittynein analytiikan menetelmä on ohjaileva analytiikka. Ohjaileva analytiikka nimensä mukaisesti ohjaa toimintaa. Käytännössä se tarkoittaa reaaliaikaista tietovirran seuraamista ja analysoimista (Davenport ym., 2012; Boyd & Crawford, 2012; Fan & Bifet, 2013). Russom (2011) toteaa haasteen olevan kaiken tarpeellisen datan reaaliaikaisessa analysoinnissa ja tarvittavien toimen-

piteiden tunnistamisessa. Davenport ym. (2012) esittävät, että dataa ei enää säilötä tietovarastoissa entiseen malliin, ja että datan virtauksen välitön ja tarkka seuraaminen on tärkeämpää, kuin historiallisen datan analysointi. He toteavat massadatan parissa työskentelemisessä tärkeää olevan kyky nopeaan päätöksentekoon ja useisiin korjaaviin toimenpiteisiin. Tehokkaasti toimiessaan kaiken täytyisi tapahtua reaaliaikaisuuden lisäksi täysin automaattisesti (Labrinidis & Jagadish, 2012). Boyd ja Crawford (2012) käyttävät esimerkkinä ihmisten reaktioiden analysoimista juuri vaalitulosten julkistamisen jälkeen. Koska kyseinen reaktio on hetkellinen, se täytyy tunnistaa ja analysoida välittömästi, sekä ryhtyä toimenpiteisiin, mikäli tilanne niin vaatii.

Datan analysoinnissa käytettävät mallit elävät ja kuolevat liiketoiminnallisten tarpeiden mukaan. Malleja tulee kehittää ja tarkentaa jatkuvalla prosessilla. (LaValle ym., 2011.) Yksinkertaiset mallit ja massadata päihittävät monimutkaisemmat ja kehittyneemmät mallit, joita käytetään perinteiseen dataan (McAfee & Brynjolfsson, 2012).

Analysointiprosessin viimeinen vaihe on datan tulkinta. Massadatan analysoinnilla saatava hyöty on rajallinen, mikäli käyttäjät eivät ymmärrä analyysin tuloksia (Labrinidis & Jagadish, 2012). Päätöksentekijän täytyy osata tulkita analyysin tuloksia tehdäkseen harkitun päätöksen. Yleensä tämä tarkoittaa, että päätöksentekijä yrittää toistaa ja tällä tavoin varmistaa tietokoneen toimittaman tuloksen. (Jagadish ym., 2014.) Manyika ym. (2011) näkevät tulkitsemisen osaamisen puutteen haasteena, joka hidastaa massadatan hyödyntämistä. McAfee ja Brynjolfsson (2012) toteavatkin, että parhaat datatieteilijät (eng. data scientist) ymmärtävät myös yritysten liiketoiminnallista puolta ja osaavat auttaa johtajia tulkitsemaan analyysin tuloksia. Jotta analyysin tuloksista saataisiin käyttäjäystävällisiä, täytyy käyttää erityistä huomiota tulosten visualisointiin (Fan & Bifet, 2013). Yritysjohdajat haluavat ymmärtää saamaansa informaatiota nopeasti ja he näkevät datan visualisoinnin olevan paras keino tarkastella tuloksia. Esimerkkejä datan visualisoimisesta ovat raportointinäkyvä (eng. dashboard) tai tulokortti (eng. scorecard). (LaValle ym., 2011.)

2.3 Massadatan mahdollisuuksia

Massadata mahdollistaa täysin uudenlaisen datan keräämisen, varastoimisen ja analysoimisen (Power, 2013). Chen ym. (2012) toteavat dataa olevan kaikkialla ympärillämme, sekä sen olevan edullisempaa kerätä kuin koskaan aikaisemmin. Massadatan käytön nähdään tuovan yritykselle moninkertaisen hyödyn kustannuksiin verrattuna (Jagadish ym., 2014). Provost ja Fawcett (2013) huomasivat, että massadata-teknologioiden käyttö korreloi positiivisesti tuottavuuden kasvun kanssa. Corte-Real ym. (2016) esittävät massadata-analytiikan lisäävän yritysten ketteryyttä kolmella tapaa: yritykset ovat valmiimpia havaitsemaan uhkia ja mahdollisuuksia, yritykset tarttuvat herkemmin uusiin mahdollisuuksiin ja yritykset mukautuvat paremmin teknologiseen ympäristöön saavuttaakseen kilpailuetua.

Sagioglu ja Sinanc (2013) luettelevat pitkän listan massadatan tuottamia hyötyjä. Näitä hyötyjä ovat: osuvampi markkinointi, tarkempi näkemys liiketoiminnasta, asiakaslähtöinen erikoistuminen, myynti- ja markkinamahdollisuuksien tunnistaminen, automatisoitu päätöksenteko, asiakkaan käyttäytymisen ymmärtäminen, investointien lisääntynyt tuottosuhde, riskien ja markkinoiden trendien tunnistaminen, liiketoiminnan muutoksen ymmärtäminen, parempi suunnittelu- ja ennustuskyky, sekä tuotannon tuottojen lisääminen. Tiivistetysti he toteavat massadatalla olevan potentiaalia muuttaa jokaista liiketoiminnan osa-aluetta. Russom (2011) esittää massadata-analytiikan hyödyttävän erityisesti asiakassuhteiden, sekä liiketoimintatiedon hallintaa. Massadata voidaan nähdä myös potentiaalisena työkaluna vakavien yhteiskunnallisten ongelmien, kuten sairauksien, terrorismin ja ilmastonmuutoksen ratkaisemiseksi (Boyd & Crawford, 2012).

LaValle ym. (2011) huomasivat tutkimuksissaan, että parhaiten pärjäävät yritykset käyttivät analytiikkaa viisi kertaa enemmän kuin huonosti pärjäävät. He myös huomasivat, että yritykset, jotka tiedostivat oman tilansa analytiikan käytön suhteen, olivat valmiimpia kääntämään kohtaamansa haasteet mahdollisuuksiksi. Heidän mukaansa yritysten prosesseissa on useita kohtia, joissa analytiikka voi lisätä liiketoiminnallista arvoa. Heidän mielestään analytiikan tuomat mahdollisuudet ovat keskeisiä menestyksen saavuttamiseksi ja lisäarvon tuottamiseksi. Ajan myötä datapohjainen päätöksenteko leviää koko organisaatioon, joka johtaa analytiikan tuoman lisäarvon nopeaan lisääntymiseen. Nopein keino aloittaa massadatan hyödyntäminen on keskittyä isoihin liiketoiminnallisiin ongelmiin, jotka voidaan ratkoa analytiikan avulla, kehittää osaamista, jota organisaatiossa jo on, sekä viedä saavutettua uutta näkemystä osaksi päivittäistä toimintaa. (LaValle ym., 2011.)

McAfee ja Brynjolfsson (2012) toteavat yhä useampien liiketoiminnallisten prosessien digitalisoitumisen, uusien datalähteiden syntymisen ja alati halpenevan teknologian tuovan meidät uudelle aikakaudelle. Uudet, perinteisistä poikkeavat, datalähteet sisältävät valtavasti potentiaalista analysoitavaa (Power, 2013). Digi-aikana syntyneet yritykset osaavat jo hyödyntää massadataa, mutta massadatalla on valtavasti potentiaalia tarjota kilpailuetua myös olemassa olleille yrityksille. Massadataa hyödyntämällä myös perinteiset yritykset voivat saada valtavasti lisäarvoa toimintaansa. (McAfee & Brynjolfsson, 2012.) Gandomi ja Haider (2015) huomasivat jopa pienten ja keskisuurten yritysten hyödyntävän massadata-analytiikasta esimerkiksi verkkosivujen ulkoasun suunnittelussa, myynnin lisäämisessä nykyisille asiakkaille tai yksilöityjen tuote-ehtotusten tekemisessä. Tarjolla on jatkuvasti useampia liiketoiminnallisia mahdollisuuksia hyödyntää massadataa, joihin yritysten kannattaisi tarttua (Russom, 2011).

Lähivuosien teknologinen kehitys on ollut valtavaa. Työkalut massadatan hallitsemiseksi ovat kehittyneet suurin harppauksin. Lisäksi suurin osa ohjelmistoista perustuu avoimeen lähdekoodiin (eng. open source), joka mahdollistaa työkalujen käytön suurten yritysten ulkopuolellakin. (McAfee & Brynjolfsson, 2012.) Uudet datan hallintaan ja analysointiin kehitetyt teknologiat mahdollista-

vat organisaatioiden hyödyntävän dataa osana liiketoiminnan prosesseja. Massadata-teknologiat mahdollistavat esimerkiksi tunneanalyysin tekemisen miljoonista hyödyllisistä sosiaalisen median julkaisuista. (Gandomi & Haider, 2015.) Massadata tarjoaa myös hyvän mahdollisuuden teknologisten innovaatioiden kehittämiseen yrityksissä (Li ym., 2017).

Massadatan avulla voidaan löytää rakeisia yksityiskohtia liiketoiminnoista ja asiakaskohtaamisista, jotka löytävät harvoin tiensä tavalliselle raportille. Jos yritys haluaa nähdä jotain täysin uutta liiketoiminnastaan, kannattaa sen käyttää dataa, jota ei ole koskaan aiemmin käytetty liiketoimintatiedon hallinnassa tai analytiikassa. (Russom, 2011.) Verkottumisen, sekä datan keruun ja varastoinnin nopean kehityksen myötä massadata leviää vauhdilla kaikille tieteen ja tekniikan aloille (Wu ym., 2014). Strategisena resurssina käytettynä massadatasta on saata- vissa suuri taloudellinen hyöty (Li ym., 2017).

2.4 Massadatan haasteita

Massadatan käyttöön liittyy myös monia haasteita. Massadatan kuvitellaan olevan aina oikeassa, mutta se ei pidä paikkaansa (Labrinidis & Jagadish, 2012). Vaikka datasta nähdään olevan valtavasti hyötyä, yritykset eivät kykene hyödyntämään sitä (Gao ym., 2015; Wu ym., 2014). LaValle ym. (2011) toteuttamassa kyselyssä kuusi kymmenestä vastaajasta totesi yrityksillään olevan enemmän dataa kuin he pystyvät tehokkaasti hyödyntämään. Yleisimmät haasteet liittyvät massadata-teknologioiden tai -menetelmien osaamiseen, datan monimuotoisuuteen tai muihin ominaisuuksiin, järjestelmiin, johtamiseen, yksityisyyteen, omistajuuteen tai käytettävyyteen. Massadata ei aina ole parempaa dataa. Se riippuu datan todenmukaisuudesta ja kyvystä tarjota oikeaa informaatiota. (Fan & Bifet, 2013.) Silti on tärkeää muistaa, että yritysten kohtaamat suurimmat ongelmat ovat samalla myös niiden suurimpia mahdollisuuksia ja massadatan tarjoaa uusia keinoja niiden ratkaisemiseksi (LaValle ym., 2011).

Massadata-teknologioiden ja -menetelmien osaaminen nousi esiin haasteena useissa tutkimuksissa (McAfee & Brynjolfsson, 2012; Labrinidis & Jagadish, 2012; Sagioglu & Sinanc, 2013; Gao ym., 2015; LaValle ym., 2011; Davenport ym., 2012; Russom, 2011). Sagioglu ja Sinanc (2013) toteavat osaamisen puutteen olevan haaste massadata-analytiikan käyttöönottovaiheessa. Massadata-teknologioiden osaajat ovat vaikeita löytää ja kalliita palkata (Gao ym., 2015; Sagioglu & Sinanc, 2013). LaValle ym. (2011) nostavat suurimmaksi ongelmaksi ymmärryksen puutteen data-analytiikan vaikutuksista liiketoimintaan. On vaikea hyödyntää teknologioita ja menetelmiä, joiden käytön tarkoitusta ei ymmärrä. Koska datan keruu ja käsittely ovat ensiarvoisen tärkeitä vaiheita datan analysoimista varten, henkilöillä, jotka työskentelevät massadatan parissa, tulee olla oleellista ja luovaa IT-osaamista (Davenport ym., 2012).

Massadatan ominaisuudet luovat useita haasteita sen hyödyntämiseen. Jo datan määrä tuottaa haasteita (Jagadish ym., 2014). Tavoitteena on kerätä talteen

ainoastaan data, joka sisältää arvoa ja olla keräämättä loppuosaa. Haasteena onkin tunnistaa arvokas osa datasta, jotta mitään olennaista osaa ei menetettäisi. (Labrinidis & Jagadish, 2012.) Tarpeettoman datan poistamiseen sisältyy aina riskejä (Chen, Mao & Liu, 2014). Datan määrän lisäksi myös muut ominaisuudet, kuten monimuotoisuus ja todenmukaisuus tuovat ongelmia. On todennäköistä, että osa strukturoimattoman datan sisältämästä hyödyllisestä informaatiosta katoaa datan käsittelyvaiheessa (Gao ym., 2015). Strukturoimaton data on muutenkin vähemmän arvokasta ja vaatii nopeaa, jopa reaaliaikaista, analysointia tuottaakseen hyötyä (Li ym., 2017). Boyd ja Crawford (2012) toteavat suurten internetistä louhittujen datamassojen sisältävän virheellisyyttä ja virheellisyyden keräytävän, kun useita datamassoja käytetään yhdessä. Virheellisyyttä aiheuttaa esimerkiksi puuttuvat tai epämääräiset arvot (Wu ym., 2014; Jagadish ym., 2014). Onkin tärkeää, että käytetty data on laadukasta, koska massadatan käytöllä voi olla merkittäviä seurauksia yrityksissä (Gao ym., 2015).

Myös järjestelmät luovat useita massadataan liittyviä haasteita. Gao ym. (2015) toteavat datan massiivisen määrän säilytyksen tulevan ongelmaksi ennemmin tai myöhemmin. Perinteiset tietovarastot eivät pysty varastoimaan strukturoimatonta tai osittain strukturoitua dataa (Russom, 2011; Chen ym., 2014). Onkin tarve kehittää uusia teknologioita tai menetelmiä, kuten datan yksinkertaistaminen datan varastointia varten (Li ym., 2017). Järjestelmät ovat myös kohteita hyökkäyksille ja toimintahäiriöille (Wu ym., 2014). Lisäksi Sagiroglu ja Sinanc (2013) nostavat esiin ongelman prosessoritehon riittävyyden suhteen. Datan määrä lisääntyy reilusti nopeammin kuin prosessoriteho kasvaa Mooren lain mukaan (Jagadish ym., 2014). Kambatla ym. (2014) esittelevät haasteita olevan suurten datamäärien varastoinnissa, maantieteellisesti hajautetun datan yhteen saamisessa, sekä dataan tehtävässä laskennassa.

Massadatan hyödyntäminen vaatii uudenlaista johtajuutta. LaValle ym. (2011) nimeävätkin suurimmiksi haasteiksi massadatan käytölle johtajuuden ja yrityskulttuurin. Haaste on suuri varsinkin johtajille, jotka ovat tottuneet vastaamaan kaikesta päätöksenteosta itse. Massadatan hyödyntäminen vaatii joskus päätäntävällän antamista datalle (McAfee & Brynjolfsson, 2012). Eikä datasta informaation tuottaminen ole yksinkertaista (Sagiroglu & Sinanc, 2013). Kun data on kerätty ja analysoitu, se toimitetaan päätöksentekijöille. Yleensä nämä henkilöt ovat yrityksissä johtavissa asemissa. Jos he jättävät käyttämättä tuotetun informaation, eivät heidän alaisensaakaan todennäköisesti käytä dataa hyödykseen. (Labrinidis & Jagadish, 2012.)

Yksityisyys nousi esiin yhtenä suurimmista massadatan tuomista haasteista (Labrinidis & Jagadish, 2012; Sagiroglu & Sinanc, 2013; Wu ym., 2014; Gao ym., 2015). Tietosuoja on merkittävä huolenaihe, joka korostuu massadatan kontekstissa (Jagadish ym., 2014). Chen ym. (2014) jakavat massadataa koskevan tietosuojan kahteen osaan: tietosuojaan datan keruuvaiheessa ja tietosuojaan datan säilöntävaiheessa. Datan keruuvaiheessa saatetaan kerätä tietoja, joita käyttäjä ei halunnut jakaa tai ole edes tietoinen niiden keräämisestä. Datan säilöntävaiheessa on tärkeää suojata data niin, ettei yksittäistä lähdettä pystytä tunnistamaan, sekä niin, että tiedot säilyvät suojattuina tietovarastoissa. Jagadish ym.

(2014) näkevät ongelmaksi useista lähteistä saatavan datan yhdistelemisen, jolloin pystytään mahdollisesti yksilöimään datan lähde. Boyd ja Crawford (2012) toivat esiin pelottavan skenaarion, jossa massadataa käytetään valtiovallan valvovana silmänä, joka rikkoo kansalaisten yksityisyyttä, rajaa kansalaisoikeuksia ja lisää valtaapitävien otetta. Giber (2016) tuokin ilmi huolensa massadatan epäeettisestä käytöstä ja toivoo ihmisten kiinnittävän huomiota myös yksityisyyteen massadataprojekteissa.

Haasteena nähdään myös massadatan omistajuus. Yritykset haluavat kuvitella omistavansa kaiken keräämänsä datan. Yksityishenkilö saattaisi kuitenkin haluta vaihtaa palveluntarjoajaa ja siirtää kerätyn datan kilpailevan yrityksen hallintaan tai jopa poistattaa omat tietonsa kokonaan. Yrityksille haasteena on säilyttää sopiva suhde kilpailuedun ja asiakastyytyväisyyden välillä, kun taas yksityishenkilöille olla tietoinen heistä kerättävästä datasta ja saada riittävä korvaus siitä. (Cole, Nelson & McDaniel, 2015.)

3 MASSADATA YRITYSTEN PÄÄTÖKSENTEOSSA

Yritykset kohtaavat päivittäin valtavan määrän tilanteita, joissa tarvitaan päätöksentekoa. Yritysten täytyy tietää mitä tapahtuu nyt, mitä todennäköisesti tapahtuu seuraavaksi ja mitä toimia täytyy tehdä saavuttaakseen paras mahdollinen lopputulos (LaValle ym., 2011). Päätöksenteko voi kuitenkin olla erittäin haastavaa. Erilaisia vaihtoehtoja, joiden väliltä päätös täytyy tehdä, voi olla satoja tai jopa tuhansia (Aronson, Liang & Turban, 2005, s. 39). Massadata voi auttaa yrityksiä tekemään parempia päätöksiä. Datapohjainen tai dataohjautuva päätöksenteko (eng. data-driven decision making) tarkoittaa päätöksentekoa data-analyysin perusteella intuition sijasta (Provost & Fawcett, 2013). McAfee ja Brynjolfsson (2012) väittävät datapohjaiset päätösten olevan parempia päätöksiä, kuin perinteisin menetelmin tehtyjen. Provost ja Fawcett (2013) huomasivat, että mitä enemmän yritykset olivat dataohjautuvia, sen tuottavampia ne olivat.

3.1 Päätöksentekoprosessi

Päätöksenteko on prosessi, jossa valitaan useiden vaihtoehtojen väliltä tarkoituksena saavuttaa tietty tavoite tai tavoitteet (Aronson ym., 2005, s. 39). He esittävät Simonin päätöksentekomallin, jossa päätöksentekoprosessi koostuu neljästä vaiheesta. Vaiheet ovat: tiedustelu (eng. intelligence), suunnittelu (eng. design), valinta (eng. choice) ja toteutus (eng. implementation). Heidän mukaansa tämä malli on kaikkein suppein, mutta silti valmis kuvaus päätöksentekoprosessista. Mallin mukaan vaiheet seuraavat toisiaan, mutta jokaisesta vaiheesta voidaan myös palata edelliseen.

Simonin mallin mukainen päätöksentekoprosessi alkaa tiedusteluvaiheella. Tiedusteluvaiheessa ympäristöstä etsitään ongelmia tai mahdollisuuksia joko ajoittain tai jatkuvasti. Vaiheeseen kuuluu organisaation tavoitteiden tunnistaminen, datan keruu, ongelmien tunnistaminen, ongelmien omistajuuden kohdistaminen, ongelmien jaottelu, sekä ongelmien kuvaaminen. Tässä vaiheessa tehdään myös yleisoletukset ongelmien yksinkertaistamiseksi ja niiden käsittelyn helpottamiseksi seuraavissa vaiheissa. (Aronson ym., 2005, s. 51.)

Seuraava vaihe Simonin mallin mukaan on suunnittelu. Suunnitteluvaiheessa etsitään ja kuvataan vaihtoehtoja ongelman ratkaisemiseksi. Tässä vaiheessa päätetään myös kriteerit, joiden mukaan vaihtoehtoja tarkastellaan. Näiden perusteella vaihtoehtoja vertaillaan ja mahdollista lopputulemaa yritetään ennustaa. (Aronson ym., 2005, s. 56.)

Kolmas vaihe mallin mukaan on valinta. Tässä vaiheessa esitetyistä ja arvioiduista vaihtoehdoista valitaan paras tai sopivin ja sen toteuttamiskelpoisuus testataan. Laaditaan myös suunnitelma vaihtoehdon toteuttamiselle. (Aronson ym., 2005, s. 69.)

Viimeisenä mallin mukaisena vaiheena on toteutus. Toteutusvaiheessa suunnitelma viedään käytäntöön ja sen vaikutuksia seurataan. Toteutus on onnistunut, kun se ratkaisee ensimmäisessä vaiheessa määritellyn ongelman. Epäonnistunut toteutus johtaa palaamiseen mallin edeltäviin vaiheisiin. (Aronson ym., 2005, s. 70.)

Mintzberg ja Westley (2001) esittävät näkemyksen, jonka mukaan päätöksiä ei aina tehdä edellä kuvatun rationaalisen mallin mukaisesti, jossa päätöksenteko perustui ajatteluun ja useiden vaihtoehtojen läpikäymiseen. Heidän mukaansa päätöksiä voidaan tehdä myös näkemys edellä tai yrityksen ja erehdyksen kautta. Näkemys edellä tehtävissä päätöksissä päätöksentekijä on nähnyt mielessään, kuinka jokin ongelma ratkeaa. Tämänkaltaiset päätökset ovat yleisiä, kun etsitään luovia ratkaisuja. Yrityksen ja erehdyksen kautta tehtävät päätökset tehdään yksinkertaisesti kokeilemalla eri vaihtoehtoja ja valitsemalla parhaiten toimivat vaihtoehdot. Tällaista päätöksentekoa hyödynnetään usein, kun tilanne on uusi tai monimutkainen ja informaatiota on saatavilla huonosti. (Mintzberg & Westley, 2001.)

Myös lukuisia muita malleja päätöksentekoprosessista on olemassa. Eräs nykymaailmaan sopivimmista on HTPT-silmukka (eng. OODA-loop), jossa prosessin vaiheet toistavat itseään jatkuvasti muodostaen ikään kuin silmukan. Prosessin vaiheet ovat havainnointi (eng. observe), tilanearvio (eng. orient), päätös (eng. decide) ja toiminta (eng. act). Jokaisesta vaiheesta kerätään palautetta, joka vaikuttaa seuraavaan silmukan läpikulkuun. (Boyd, 1996.) HTPT-silmukka sopii nykymaailmaan erityisen hyvin, koska päätöksiä vaaditaan jatkuvasti ja kilpailuedun saavuttaminen markkinoilla on yhä haastavampaa. Yritysten täytyy olla kykeneväisiä nopeisiin päätöksiin ja valmiina reagoimaan muutoksiin tyydyttääkseen asiakkaidensa tarpeet. (Mayhew, Saleh & Williams, 2016.)

3.2 Päätöksentekoa tukevat järjestelmät

Päätöksenteon tueksi rakennetut järjestelmät eivät ole uusi keksintö. Ensimmäiset päätöksentekoa tukevat järjestelmät (eng. decision-support systems) ovat olleet käytössä jo 1970-luvulla (Watson & Wixom, 2007). Alkuvaiheessa johtajat eivät kuitenkaan luottaneet uusiin järjestelmiin, vaan tekivät päätöksiä entiseen tapaan oman intuitionsa pohjalta (Aronson ym., 2005). Yksinkertaistettuna päätöksentekoa tukevilla järjestelmillä tarkoitetaan tietoteknisiä ratkaisuita, joilla voidaan tukea monimutkaista päätöksentekoa ja ongelmanratkaisua. Järjestelmillä pyritään parantamaan päätöksenteon tehokkuutta tai lisäämään päätöksen vaikuttavuutta. (Shim ym., 2002.)

Alun perin päätöksentekoa tukevat järjestelmät olivat ainoastaan harvojen päättävissä asemassa olevien henkilöiden henkilökohtaisessa käytössä. Kuitenkin 1980-luvulla kehittyivät ryhmäpäätöksentekoa tukevat järjestelmät (eng. group support systems) tarjotakseen myös ryhmille ongelmanratkaisualustan. (Shim ym., 2002.) Ryhmien tai tiimien saaminen samaan tilaan ja aikaan on haasteellista ja kallista, joten järjestelmät tarjoavat siihen helpotetun ratkaisun. Yleisimpiä

ryhmäpäättöksentekoa tukevan järjestelmän ominaisuuksia ovat video- ja äänikonferenssit, elektroninen aivoriivialusta, äänestystyökalut ja dokumenttien jako. (Aronson ym., 2005, s. 374). Internetin yleistyminen on lisännyt järjestelmien käyttöä valtavasti tehden prosesseista tehokkaampia ja laajemmin saatavilla olevia, koska suurin osa sovelluksista käyttää verkkoselainta käyttöliittymänään (Shim ym., 2002).

Lähemmäs 2000-lukua tultaessa päätöksentekoa tukevat järjestelmät olivat kehittyneet merkittävästi. Uusia merkittäviä työkaluja olivat tietovarastot (eng. data warehouse), OLAP (eng. on-line analytical processing) ja tiedonlouhinta (eng. data mining). (Shim ym., 2002.)

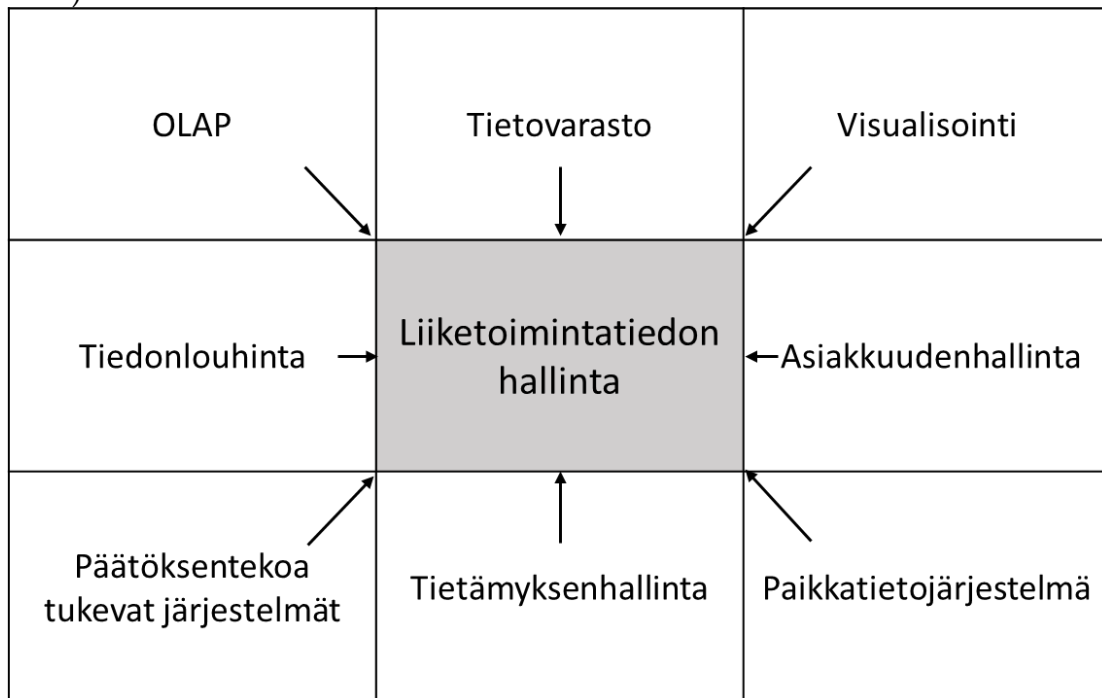
Tietovarastoinnilla tarkoitetaan datan tallentamista lähteestä tai lähteistä yhdennettyn tietovarastoon. Dataa voidaan kerätä sisäisistä järjestelmistä tai ulkoisista lähteistä. (Watson & Wixom, 2007.) Shim ym. (2002) määrittelevät tietovaraston aiheuuntuneeksi, yhdenntyksi, aika-muuntuvaksi, datan pysyvässäilöksi. Heidän mukaansa laajan tietovaraston rakentaminen johtaa yleensä kasvaneeseen mielenkiintoon analysoida ja hyödyntää kertynyttä historiadataa. Watson ja Wixom (2007) toteavat metadatan olevan tärkeä osa tietovarastointia. Heidän mukaansa oikea metadata lisää läpinäkyvyyttä prosessissa, jossa data siirtyy datan lähteestä tietovaraston kautta loppukäyttäjälle.

Yksi vaihtoehto tietovarastossa olevan datan analysoinnille on OLAP. Shim ym. (2007) kuvaavat OLAP:in työkaluina, joilla käyttäjä voi tarkastella interaktiivisesti dataa useista eri reaali maailman ulottuvuuksista. OLAP voidaankin nähdä työkaluina, joilla tarkastella dataa moniulotteisesti, sekä suorittaa toimintoja, kuten suodatus, ryhmittely, syventyminen, sekä pivointi (Chaudhuri, Dayal & Narasayya, 2011). OLAP:in avulla käyttäjä voi esimerkiksi aluksi tarkastella jonkin maanosan valitun ominaisuuden arvoa, josta käyttäjä siirtyy katselmaan tietyn maan kyseisen ominaisuuden arvoa ja yhä edelleen tietyn kaupungin arvoa. OLAP tarjoaakin mallinnus- ja visualisointimahdollisuuksia suurille datamassoille (Aronson ym., 2005, s. 257). OLAP on yhä yleisin datan analysointityökalu ja sellaisena sen odotetaan pysyvänkin (Russom, 2011).

Tiedonlouhintaa voi kuvailla myös termeillä tietokantatutkimus tai informaation ja tietämyksen etsintä. Tiedonlouhinnassa datasta etsitään toistuvia sarjoja, joista johdetaan sääntöjä analysointia varten. (Shim ym., 2002.) Aronson ym. (2005, s. 263) kuvaavat tiedonlouhinnan prosessiksi, jossa käytetään tilastotieteellisiä ja matemaattisia tekniikoita hyödyllisen informaation tunnistamiseen ja erottamiseen laajoista tietokannoista. Heidän mukaansa tiedonlouhintaprosessi on automatisoitavissa, jotta datasta saadaan hyödyllistä tietoa nopeammin. Tiedonlouhinta on myös tarkempi keino analysoida dataa kuin OLAP, koska sen avulla on mahdollista rakentaa ennustavia malleja (Chaudhuri ym., 2011).

Päättöksentekoa tukevista järjestelmistä nykypäivänä kaikkein tunnetuimpia ja käytetyimpiä ovat liiketoimintatiedon hallintajärjestelmät. Ne yhdistävät datan keruun, varastoinnin ja tietämyksenhallinnan analyttisten työkalujen kanssa esittääkseen kilpailuetua tarjoavaa informaatiota suunnittelijoille ja päättöksentekijöille (Negash, 2004). Chaudhuri ym. (2011) kuvaavat liiketoimintatiedon hallintajärjestelmiä kokoelmaksi päätöksentekoa tukevia teknologioita, joita

yritysten tietotyöntekijät voivat käyttää tehdäkseen parempia ja nopeampia päätöksiä. Watson ja Wixom (2007) kuvaavat liiketoimintatiedon hallintaa prosessina, jossa on kaksi pääaktiviteettia: datan järjestelmään tuominen ja informaation järjestelmästä vieminen. Heidän mukaansa datan tuomisesta saatava arvo on marginaalinen verrattuna järjestelmästä viedyn informaation käyttämiseen päätöksenteossa ja yritykset keskittyvätkin pääasiassa tämän vaiheen kehittämiseen. Liiketoimintatiedon hallintajärjestelmiä käytetään, jotta ymmärrettäisiin yrityksen nykytila ja havaittaisiin saatavilla olevia mahdollisuuksia. Liiketoimintatiedon hallintajärjestelmät voidaankin nähdä kokonaisuutena, joka yhdistää aikaisemmin käytössä olleet työkalut yhdeksi kokonaisuudeksi (kuvio 3). (Negash, 2004).



KUVIO 3 Liiketoimintatiedon hallinnan suhde muihin järjestelmiin (Negash, 2004)

3.3 Massadatan vaikutukset yritysten päätöksentekoon

Massadata ei ole uusi ilmiö, mutta sen tehokas hyödyntäminen on (Russom, 2011). Monilla toimialoilla yritykset tarjoavat samanlaisia tuotteita ja käyttävät samoja teknologioita, jolloin liiketoimintaprosessit jäävät ainoiksi tavoiksi erottautua kilpailijoista ja hankkia kilpailuetua. Analytiikkaa käyttävät yritykset pystyvät virittämään prosessinsa tehokkaammiksi kuin heidän kilpailijansa. (Davenport, 2006.) Yritysten johtajat pohtivat saavatko he kaiken irti jättimäisistä datamassoista, joita heidän yrityksillään jo on käytettävissään ja etsivät jatkuvasti uusia tapoja hyödyntää dataa, sekä hankkia kilpailuetua (LaValle ym., 2011). Massadatan ansiosta johtajat voivat mitata enemmän asioita ja siten tuntea pa-

remmin heidän yrityksensä, sekä hyödyntää saatua informaatiota päätöksenteossa ja yrityksensä suorituskyvyn kasvattamisessa. Yksinkertaisesti kuvattuna massadatan käyttö johtaa parempiin ennusteisiin ja paremmat ennusteet parempiin päätöksiin. (McAfee & Brynjolfsson, 2012.)

Yritysten johtajat haluavat heidän yritystensä toiminnan perustuvan data-pohjaiseen päätöksentekoon (LaValle ym., 2011). Päätökset, jotka tehtiin ennen arvaamalla tai mallintamalla käsin, voidaan tehdä nykyisin automaattisten matemaattisten datamallien avulla (Jagadish ym., 2012; McAfee & Brynjolfsson, 2012). McAfee ja Brynjolfsson (2012) huomasivat tutkimuksessaan, että mitä enemmän yritykset luonnehtivat itseään dataohjautuviksi, sitä paremmin he pärjäsivät taloudellisilla ja toiminnallisilla mittareilla mitattuina. Heidän tulostensa mukaan kunkin toimialan paras kolmannes eniten itseään dataohjautuvaksi kutsuvista yrityksistä oli keskimäärin 5% tuottavampia ja 6% kannattavampia kuin heidän kilpailijansa. LaValle ym. (2011) huomasivat, että yritykset jotka kokivat liiketoimintatiedon ja analytiikan tuovan heille kilpailuetua, olivat kaksi kertaa todennäköisemmin parhaiten menestyvien yritysten joukossa. Heidän mukaansa johtajat haluavat ymmärtää parhaan mahdollisen ratkaisun monimuotoisten liiketoiminnallisten parametrien tai uuden informaation perusteella ja he haluavat olla valmiina reagoimaan nopeasti. Dataohjautuvuudella onkin mahdollista parantaa yritysten tuottavuutta (Jagadish ym., 2012; Provost & Fawcett, 2013).

Provost ja Fawcett (2013) jakavat päätökset kahteen kategoriaan: päätöksiin, joissa datasta pitää löytää jotakin uutta informaatiota ja päätöksiin, jotka ovat toistuvia. Heidän mukaansa erityisesti toistuvien päätösten tarkkuuden parantaminen massadata-analyysin avulla voi johtaa suuriin tuottoihin. Esimerkiksi jos yritys pystyy dataa analysoimalla löytämään keinoja, joilla kasvattaa edes hieman todennäköisyyttä, jolla asiakas uusii tilauksensa, voi sillä olla suuri vaikutus yrityksen liikevaihtoon. Yritysten täytyykin palkata datatieteilijöitä, jotka osaa- vat etsiä datasta toistuvia kaavoja ja muuttavat nämä hyödylliseksi liiketoimintatiedoksi (McAfee & Brynjolfsson, 2012). Power (2013) nostaakin esiin huomion, että yritysjohtajien täytyy ymmärtää mitä datatieteilijä tekee ja miksi hänet täytyy palkata.

Analytiikan on ennen nähty olevan prosessien valvomista ja poikkeamien etsimistä. Massadatan ansiosta tämä näkemys kääntyy pääläelleen, koska massadata mahdollistaa ennakoivan analytiikan. Yritykset, jotka huomaavat muutokset ja pystyvät reagoimaan nopeasti saavat kilpailuetua. (Davenport ym., 2012.) Gandomi ja Haider (2015) toteavat organisaatioiden tarvitsevan tehokkaita prosesseja suurten ja nopeasti syntyvien datamassojen hallitsemiseksi ja hyödylliseksi informaatioksi saattamiseksi. Kilpailuedun saavuttamiseksi, yritysten nähdään muuntavan omia ydinliiketoimintojaan dataohjautuviksi, jotka pohjautuvat jatkuvasti syntyvään monimuotoiseen massadataan (Jagadish ym., 2014; Davenport ym., 2012). Fan ja Bifet (2013) muistuttavat, että yritysten on tärkeä arvioida saatua muutosta myös tilastollisten tulosten perusteella. Tulosten ei tarvitse aina olla taloudellisia, sillä joskus tavoitteena voi olla jokin muu mittari, kuten asiakastyytyväisyyden parantaminen.

Jotta dataohjautuvaa päätöksentekoa käytettäisiin yrityksissä, täytyy uudet mallit olla tarpeeksi yksinkertaisia ymmärtää loppukäyttäjälle, tiiviisti liitettyjä yrityksen strategiaan tavoitteisiin ja sulavasti upotettuja yrityksen sisäisiin prosesseihin (LaValle ym., 2011). Parhaiten massadataa hyödyntävillä yrityksillä ei ole eniten dataa, vaan johto, joka on asettanut selkeät liiketoiminnan tavoitteet, määritellyt menestyksen tekijät ja ymmärtänyt mitä massadatalta voidaan saavuttaa. Massadatan voima ei poista tarvetta ihmisen näkemykselle ja kokemukseksi. (McAfee & Brynjolfsson, 2012; Gao ym., 2015.) Walker (2012) toteaaakin, ettei massadatan käyttö vaimenna ihmisten luovuutta, arvauksia tai näkemyksiä, vaan tarjoaa niille uudenlaista käyttöä.

Menestyminen nykypäivän datasuuntautuneessa liiketoimintaympäristössä vaatii kykyä soveltaa massadateknologioita käytännön liiketoiminnallisiin ongelmiin (Provost & Fawcett, 2013). Sagiroglu ja Sinanc (2013) toteavat menestyksen vaativan merkittäviä panostuksia avainhenkilöiden kouluttamiseen. Vaikkakin useat yritysjohtajat ovat aidosti dataohjautuvia ja valmiina uskomaan ristiriidassa heidän oman näkemyksensä kanssa olevaa dataa, luottavat ihmiset silti liikaa omaan kokemukseensa ja intuitioonsa, ennemmin kuin saatavilla olevaan dataan. Kulttuurillinen muutos vaatii johdolta toimia ja merkittävimpänä niistä on nähdä johdon käyttävän datapohjaista päätöksentekoa ja syrjäyttävän oman näkemyksensä datan tarjoamalla ratkaisulla. (McAfee & Brynjolfsson, 2012.)

McAfee ja Brynjolfsson (2012) toteavat tehokkaan organisaation tuovan käytettävissä olevan informaation ja oikeat henkilöt yhteen. Heidän mukaansa informaation ja päätöksentekijöiden lisäksi prosessiin täytyy osallistua ongelmanratkoojia, jotka osaavat käsitellä massadataa. Massadatan käsittely ei tuota lisäarvoa, jos päätöksentekijät eivät ymmärrä analyysin tuloksia (Labrinidis & Jagdish, 2012). Provost ja Fawcett (2013) nostavatkin datatieteilijät yhdistäväksi tekijäksi massadatan käsittelyn ja datapohjaisen päätöksenteon väliin. Henkilöt, prosessit ja teknologiat täytyy kaikki ottaa huomioon massadateknologioiden käyttöä suunniteltaessa (Gao ym., 2015). Russom (2011) toteaa analytiikan käytön ja käyttöönoton vaihtelevan yritysten eri osastojen välillä. Hänen mukaansa jotkin osastot kehittävät omia massadataratkaisujaan, jos heillä on niille vahva liiketoiminnallinen tarve.

Datapohjainen päätöksenteko mahdollistaa myös täysin automatisoidun päätöksenteon tekemisen. Provost ja Fawcett (2013) toteavat, että automatisoidun päätöksenteon avulla yritys voi ryhtyä toimiin välittömästi, kun järjestelmät löytävät datasta poikkeaman tai etukäteen mallinnetun kaavamaisuuden. Yleisimpänä esimerkkinä automatisoidusta päätöksenteosta voidaan pitää pankkien luottokorttipetosten valvomista. Jos korttia käytetään poikkeavassa paikassa, kuten toisella puolella maapalloa, niin pankki jäädyttää kortin automaattisesti, kunnes se saa asiakkaaltaan varmistuksen kortin käyttäjästä. Automatisoidussa päätöksenteossa seurataankin datan sisällön sijasta datan virtausta (Davenport ym., 2012).

4 YHTEENVETO

Datan määrä on kasvanut valtavasti lähivuosina, eikä kehityksen odoteta hidastuvan tulevaisuudessakaan. Suurin osa tästä syntyvästä datasta täyttää massadatan kriteerit ollen määrältään todella suurta, syntyvän valtavalla tai vaihtelevalla nopeudella, olevan rakenteeltaan strukturoimatonta, sisältävän virheellisiä arvoja tai olevan muunnettavissa rahalliseksi arvoksi. Lisäksi syntyvää dataa ei pystytä käsittelemään perinteisin menetelmin eikä järjestelmin, vaan sitä varten tarvitaan uusia teknologioita. Käsitellyn kirjallisuuden perusteella voinkin todeta, että massadatan yksiselitteinen määrittäminen on hankalaa ja määritelmä kontekstisidonnaista.

Tässä tutkielmassa tutustuttiin massadataan yritysten päätöksenteon näkökulmasta. Aiheen käsittelemiseksi tutustuttiin myös massadatan analysoimiseen, massadatan tuomiin mahdollisuuksiin ja haasteisiin, sekä päätöksentekoprosessiin ja päätöksentekojärjestelmiin. Tutkimuskysymyksenä oli ”Mitä vaikutuksia massadatalla on yritysten päätöksentekoon?” ja apukysymyksinä ”Miten massadataa analysoidaan?” ja ”Mitä mahdollisuuksia ja haasteita massadataan liittyy?” Tutkielman haasteena oli suoraan tutkimuskysymykseen vastaavien lähteiden vähäinen määrä, joten tutkielman tulokset on johdettu yleisesti massadatan vaikutuksia käsittelevistä lähteistä.

Massadatalla nähdään olevan useita vaikutuksia yritysten päätöksentekoon. Yritysjohtajat haluavat yritystensä päätöksenteon perustuvan dataan intuition sijasta. Johtajat voivat massadatan avulla mitata tarkemmin yritystensä nykytilaa ja käyttää datasta saatua informaatiota päätöksenteon tukena. Massadata mahdollistaa ennakoivan analytiikan käytön, jonka avulla yritykset voivat luoda tarkempia ennusteita tulevaisuudesta ja näiden ennusteiden avulla tehdä parempia päätöksiä. Massadatan käyttö päätöksenteon tukena on kaikkein hyödyllisintä usein toistuvissa päätöksissä, joissa kerran kehitettyä mallia voidaan käyttää yhä uudestaan ja uudestaan ja näin saada jatkuvaa hyötyä. Massadatan avulla pystytään myös täysin automatisoimaan osa päätöksenteosta. Massadatarjestelmä voidaan kytkeä seuraamaan datan virtaa ja ryhtymään toimiin automaattisesti huomattavissa poikkeavuuksissa.

Huomattavaa oli, että tuloksia yritysten tuottavuudessa ja kannattavuudessa havaittiin jo pelkästään perustuen yritysten omaan näkemykseen heidän dataohjautuvuudesta, eikä varsinaisen teknologian käytön perusteella. Yritykset, jotka mielsivät itsensä dataohjautuviksi, olivat tuottavampia ja kannattavampia kuin perinteiset yritykset. Tästä vedän johtopäätöksen, että massadatasta saatavat hyödyt liittyvät myös yrityskulttuurin muutokseen ja että itsensä dataohjautuviksi mieltävät yritykset myös aidosti käyttävät datapohjaista päätöksentekoa.

Datapohjainen päätöksenteko leviää nopeasti yritysten sisällä ja näin ollen käytön lisääntyessä myös hyödyt moninkertaistuvat. Haasteena yrityksillä on saada dataohjautuva päätöksenteko leviämään koko organisaatioonsa. Massadatan käytön tuleekin olla sidottuna yrityksen niin strategiaan kuin liiketoiminnallisiin tavoitteisiin ja prosesseihin, jotta yritysten työntekijät omaksuisivat uusien

menetelmien käytön. Johdon käyttäytyminen ja heidän dataohjautuvuutensa näyttäytyy merkittävänä tekijänä yrityskulttuurin muovaamisessa.

Massadata-analytiikka on yksinkertaistettuna tiedon etsimistä ja erottamista massadatesta. Massadata-analytiikkaan kuuluu kaksi osaa, datan hallinta ja datan analysointi. Se voidaan edelleen jakaa prosessiksi, jossa on viisi vaihetta: datan keruu, datan puhdistaminen ja informaation erottaminen, datan yhdentäminen, datan mallintaminen ja analysointi, sekä datan tulkinta. Kolmen ensimmäisen vaiheen katsotaan sisältyvän datan hallintaan ja kahden viimeisen datan analysoimiseen. Yksistään massadatan keruun ei nähdä tuottavan lisäarvoa, vaan kerätty data täytyy käsitellä aina edellä kuvatun prosessin mukaisesti. Tarvemmin massadata-analytiikkaa on käsitelty kappaleessa 2.2.

Massadata tarjoaa useita mahdollisuuksia yrityksille. Koska massadata tulee yleensä täysin uusista lähteistä, se tarjoaa informaatiota, jota ei ole aiemmin ollut yritysten saatavilla. Uuden informaation avulla yritys pystyy lisäämään tietoisuuttaan omasta yrityksestään, asiakkaistaan, sekä markkinoista, joilla se toimii. Lisääntyneen tietoisuuden ansiosta yritykset pystyvät havaitsemaan entistä paremmin uusia liiketoiminnallisia mahdollisuuksia tai liiketoimintaan vaikuttavia uhkakuvia ja mukauttamaan toimintaansa niiden perusteella.

Massadatalle on myös tuottavuuden kasvua tukeva vaikutus. Massadatan avulla pystytään löytämään tuotantoprosessista pullonkauloja, jotka hidastavat tuotantoa. Lisäksi pystytään jopa automatisoimaan tuotannon virheiden havainnointiprosessi ja näin ollen vähentämään tuotannon virheellisyyttä. Massadatan tarjoamat hyödyt koskevatkin kaikkia yrityksiä, eivätkä pelkästään digiaikana syntyneitä. Suurin potentiaali hyödyntää massadataa onkin perinteisillä yrityksillä, koska he ovat saaneet liiketoimintansa kannattavaksi jo ilman massadataa.

Massadatan hyödyntämisessä on myös useita haasteita. Massadata-tekniologioiden osajia on hankala löytää ja kallista palkata. Lisäksi heidän liiketoiminnallinen ymmärrys on usein puutteellista. Haasteita on myös järjestelmissä. Koska massadataa syntyy ja kerätään valtavia määriä, sen säilyttäminen on ongelmallista. Järjestelmät ovat ja tulevat aina olemaan alttiita häiriöille tai hyökkäyksille. Myös prosessoritehon riittävyys voi tulla vastaan valtavia datamassoja käsiteltäessä. Lisäksi ihmisten yksityisyyden säilyttäminen on haaste, niin dataa kerätessä, varastoidessa, kuin käsitellessäkin.

Myöskin massadatan ominaisuudet aiheuttavat haasteita yrityksille. Massadatan valtavan määrän keräämisen ja säilömistä helpottamiseksi, täytyy tunnistaa hyödyllinen osa datasta ja tallentaa ainoastaan se. Riskinä on, että osa tärkeästä datasta menetetään rajausta tehtäessä. Kerättävä data voi myös olla jo alun perin virheellistä tai sisältää tyhjiä arvoja. Datan virheellisyys kertaantuu useita datalähteitä käytettäessä. Lisäksi useat datalähteet aiheuttavat käsittelyongelmia, koska data on usein monimuotoista ja strukturoimatonta. Kaiken lisäksi massadataa syntyy koko ajan lisää suurella nopeudella, joten dataa täytyy pystyä analysoimaan nopeasti, jopa reaaliaikaisesti.

Haasteeksi luetaan myös johtaminen. Massadatan käyttö edellyttää johdon tukea ja useimmissa tapauksissa se tarkoittaa koko yrityskulttuurin muutosta.

Johtajien on päästävä irti vanhoista tavoistaan tehdä päätöksiä perustuen kokeemukseen tai intuition ja korvattava ne datapohjaisilla päätöksentekomalleilla. Datapohjaisen päätöksenteon hyödyntäminen edellyttää kuitenkin osaamista tulkita datasta johdettua informaatiota. Johtajien onkin opeteltava tulkitsemaan uudenlaista informaatiota tai käytettävä apunaan henkilöitä, jotka osaavat kääntää tuotetun informaation liiketoiminnallisesti hyödylliseen muotoon.

Tutkielman kannalta kiinnostavaa oli huomata, että osa yrityksistä on jo herännyt hyödyntämään massadataa. Kiinnostavaa myös oli, että nämä yritykset pärjäsivät erityisen hyvin useilla eri mittareilla mitattuna. Koska tutkimuksen lähteet olivat pääosin 2010-luvun alkupuolelta, olisi mielenkiintoista lukea ajankohtainen katsaus massadataa hyödyntävistä yrityksistä ja heidän massadatan avulla saavuttamista hyödyistään. Lisäksi, koska datapohjaisen päätöksenteon nähdään rajoittavan ihmisten luovuutta päätöksentekovaiheessa, mielenkiintoinen jatkotutkimusaihe voisi olla tutkia missä tilanteissa ja kuinka dataohjautuvat yritykset käyttävät luovuutta hyödykseen.

LÄHTEET

- Aronson, J. E., Liang, T., & Turban, E. (2005). *Decision support systems and intelligent systems*. Pearson Prentice-Hall.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Boyd, J. R. (1996). The essence of winning and losing. *Unpublished Lecture Notes*, 12(23), 123-125.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Cole, D., Nelson, J., & McDaniel, B. (2015). Benefits and risks of big data. *Proceedings of SAIS 2015*. 1-5.
- Corte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in european firms. *Journal of Business Research*, 70, 379-390.
- Davenport, T. H. (2006). Competing on analytics. *Harvard business review*, 84(1), 98-108.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43-46.
- Fan, W., & Bifet, A. (2013). Mining big data: Current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter*, 14(2), 1-5.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in big data analytics projects. *Proceeding of Twenty-first Americas Conference on Information Systems* (1-14). Puerto Rico
- Giber, L. (2016). Ethical framework of big data application. *Proceedings of International Conference Information Systems 2016* (1-6).
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21-31.
- Li, M., Zhang, Z., & Hu, Z. (2017). Big data-driven technology innovation: Concept and key problems. *Proceedings of WHICEB 2017* (490-496). Wuhan

- 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*
- Mayhew, H., Saleh, T. & Williams, S. (2016) Making data analytics work for you—instead of the other way around. Haettu 20.11.2017 osoitteesta <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/making-data-analytics-work-for-you-instead-of-the-other-way-around>
- McAfee, A. & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-68.
- Mintzberg, H., & Westley, F. (2001). Decision making: It's not what you think. *MIT Sloan Management Review*, 42(3), 89-93.
- Negash, S. (2004). Business intelligence. *The Communications of the Association for Information Systems*, 13(1), 177-195.
- Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems*, 10(26), 1-49.
- Power, D. J. (2014). Using 'Big data' for analytics and decision support. *Journal of Decision Systems*, 23(2), 222-228.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.
- Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, 19, 40.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *International Conference on Collaboration Technologies and Systems (CTS)*, 42-47.
- Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, 33(2), 111-126.
- Walker, J. (2014). Big Data: A Revolution that Will Transform how we Live, Work, and Think. *International Journal of Advertising*, 33(1), 181-183.
- Ward, J. S., & Barker, A. (2013). Undefined by data: A survey of big data definitions. *arXiv Preprint arXiv:1309.5821*, 1-2.
- Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9), 96-99.
- Wu, X., Zhu, X., Wu, G., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.