



Open your mind. LUT.

Lappeenranta University of Technology

**Lappeenrannan teknillinen yliopisto**

LUT School of Business and Management

Tietojohtaminen ja Johtajuus

# **Tietoperusteisen päätöksenteon merkitys yrityksen menestykseen**

Pro Gradu tutkielma

29.4.2018

Lassi Ropponen

Työn ohjaaja: Aino Kianto

2. tarkastaja: Henri Hussinki

## TIIVISTELMÄ

<b>Tekijä:</b>	Lassi Ropponen
<b>Tutkielman nimi:</b>	Tietoperusteisen päätöksenteon merkitys yrityksen menestykseen
<b>Tiedekunta:</b>	LUT School of Business and Management
<b>Maisteriohjelma:</b>	Tietojohtaminen ja johtajuus
<b>Vuosi:</b>	2018
<b>Pro gradu -tutkielma:</b>	Lappeenrannan teknillinen yliopisto 76 sivua, 5 liitesivua
<b>Tarkastajat:</b>	Professori Aino Kianto Tutkijatohtori Henri Hussinki
<b>Hakusanat:</b>	Big Data, tietoperusteinen päätöksenteko, data-analytiikka

Big Data ja Big Data analytiikka ovat nousseet 2010 -luvulta lähtien erittäin keskeiseksi tutkimusalueeksi. Big Datan merkityksestä menestykseen on kuitenkin vain vähän aiempaa tutkimusta. Aikaisemmat tutkimukset ovat osoittaneet datan arvon syntyvän tehokkaamman päätöksenteon kautta. Tämän vuoksi koin erittäin tärkeäksi tutkia tätä aihealuetta. Tämän tutkielman tarkoituksena on tarkastella data-analytiikan vaikutuksia yrityksen menestykseen tietoperusteisen päätöksenteon kautta.

Tutkielman empiirinen osio toteutettiin hyödyntäen Lappeenrannan teknillisen yliopiston ”Aineeton pääoma ja arvonluonti” -projektin vuonna 2017 kyselylomakkeilla kerättyä aineistoa. Tutkielman teoreettinen osa käsittelee dataa, data-analytiikkaa ja päätöksentekoa. Näiden pohjalta tutkielmaan luotiin tutkimusmalli. Tutkimus ei onnistunut todistamaan tietoperusteisen päätöksenteon vaikuttavan suoraan yrityksen menestykseen, mutta esille nousi erittäin mielenkiintoisia tuloksia teknologisten ja strategisten kyvykkyyksien vaikutuksesta tietoperusteiseen päätöksentekoon. Tutkimus osoitti esimerkiksi teknologisten kyvykkyyksien roolin kasvavan, kun yritys ei vielä hyödynnä Big Dataa.

## ABSTRACT

<b>Author:</b>	Lassi Ropponen
<b>Title:</b>	Effect of knowledge-based decision-making to success
<b>Faculty:</b>	LUT School of Business and Management
<b>Major:</b>	Knowledge management and Leadership
<b>Year:</b>	2018
<b>Master's Thesis:</b>	Lappeenranta University of Technology 76 pages, 5 appendices
<b>Examiners:</b>	Professor Aino Kianto Post-doctoral Researcher Henri Hussinki
<b>Keywords:</b>	Big Data, knowledge-based decision-making, data analytics

Big Data and Big Data analytics have risen to a meaningful position in the research field of the 21<sup>st</sup> century. Effect of Big Data to success has still only a little previous research. Previous researches have shown that the value of data arises from efficient decision-making. For this reason, I find the research subject very interesting. Purpose of this study is to examine effects of data analytics to success from the knowledge-based decision-making point of view.

The empirical part of the study was done by using Lappeenranta University of Technology's "Intangible capital and value creation" project's in year 2017 gathered data. Theoretical part of this study covers subjects of data, data analytics and decision-making. Research model is based on these subjects. This study could not confirm straight effect of knowledge-based decision-making to success in any direction. From the results highly interesting findings were found regarding technological and strategic capabilities of the firm. For example, this study showed that the meaning of technological capabilities grows when firm does not yet use Big Data as its source of data.

## ALKUSANAT

Viisi vuotta kestäneet yliopisto-opintoni ovat tämän Pro Gradun myötä nyt valmiit. Olen erittäin kiitollinen siitä, että sain suorittaa opintoni Lappeenrannan teknillisen yliopiston laadukkaassa opetuksessa ja ohjauksessa. Tämän lopputyön tekeminen on ollut erittäin opettavainen ja vaiherikas kokemus. Kaikki vastoinkäymiset ja onnistumisen tunteet ovat osaltaan muokannet tästä työstä itseni näköisen.

Haluan ensiksi osoittaa suuret kiitokset Pro Graduni ohjaajalle Aino Kiannolle ja Aineeton pääoma ja arvonluonti -projektin muille jäsenille. Ilman heitä en olisi koskaan päässyt käsiksi näin tärkeään ja ajankohtaiseen aiheeseen. Osoitan myös suuret kiitokset tyttöystävälleni Annelle, vanhemmilleni ja ystävilleni kaikesta tuesta, jota olen opintojeni ohella saanut. Heidän tukensa on tuonut tarvittavaa tasapainoa myös tämän Pro Gradun tekemiseen.

Vantaalla 29.4.2018

Lassi Ropponen

## SISÄLLYSLUETTELO

1	JOHDANTO .....	1
1.1	Tutkimuskysymykset .....	5
1.2	Työn rakenne .....	7
2	TEOREETTINEN TAUSTA .....	8
2.1	Big Data ja Big Data -analytiikka .....	8
2.2	Päätöksentekoprosessi .....	14
2.3	Big Datan ja päätöksenteon yhteys .....	19
2.4	Yhteenveto .....	25
3	TUTKIMUSMALLI JA HYPOTEEESIT .....	28
4	TUTKIMUSMENETELMÄT .....	33
4.1	Menetelmät.....	33
4.2	Aineiston kuvaus .....	35
4.3	Datan seulonta .....	36
4.4	Tutkiva faktorianalyysi, EFA .....	38
4.5	Konfirmatorinen faktorianalyysi, CFA .....	41
4.6	Rakenneyhtälömalli, SEM .....	48
5	TUTKIMUSTULOKSET .....	52
5.1	Tietoperusteisen päätöksenteon vaikutus menestykseen .....	52
5.2	Kyvykkyyksien vaikutus tietoperusteiseen päätöksentekoon.....	54
5.3	Tulosten yhteenveto .....	60
6	JOHTOPÄÄTÖKSET .....	64
	LÄHTEET .....	72

## TAULUKOT

Taulukko 1: Päätöksentekoa määrittävät tekijät.	17
Taulukko 2: Yrityksen menestys	31
Taulukko 3: Tietoperusteinen päätöksenteko	31
Taulukko 4: Johdon datakyvykkyudet	32
Taulukko 5: Teknologiset kyvykkyudet	32
Taulukko 6: KMO ja Bartlettin testi	39
Taulukko 7: Reliabiliteetit	40
Taulukko 8: Mallien sopivuusluvut	43
Taulukko 9: Validiteetti ja reliabiliteetti yhtenäiselle mallille	45
Taulukko 10: Validiteetti ja reliabiliteetti jaetulle mallille	45
Taulukko 11: Sopivuustestit kausaalisille malleille	50
Taulukko 12: Tulokset, yhtenäinen malli	50
Taulukko 13: Tulokset, suuri runsaan ja monimuotoisen datan hyödyntämisen taso	51
Taulukko 14: Tulokset, pieni runsaan ja monimuotoisen datan hyödyntämisen taso	51
Taulukko 15: Analyysin tulokset	52

## KUVIOT

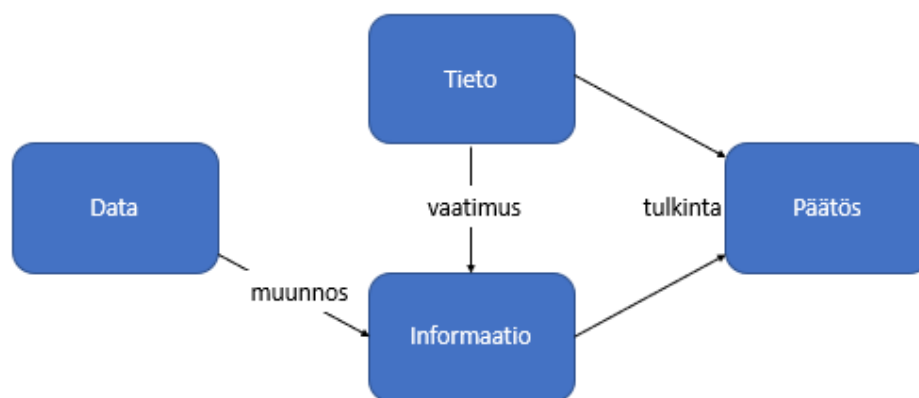
Kuvio 1: Datan, Informaation, Tiedon ja Päätöksen suhde	2
Kuvio 2: Big Data	9
Kuvio 3: Big Data -analytiikka	11
Kuvio 4: Päätöksentekosysteemi	15
Kuvio 5: Big Datan ja päätöksenteon yhteys	19
Kuvio 6: Reaktioajan viive	21
Kuvio 7: Yhteenveto	25
Kuvio 8: Tutkimusmalli	28
Kuvio 9: Cookin etäisyys latenteille muuttujille	48
Kuvio 10: H1, H2 ja H3	53
Kuvio 11: H4, H5 ja H6	55
Kuvio 12: Yhdistetyn mallin tulokset	62
Kuvio 13: Suuren monimuotoisen ja runsaan datan hyödyntämisen tulokset	62
Kuvio 14: Vähäisen monimuotoisen ja runsaan datan hyödyntämisen tulokset	63

## 1 JOHDANTO

Janssenin, Van Der Voortin ja Wahyudin (2017) mukaan aikaisemmat tutkimukset ovat osoittaneet, että Big Datan ja Big Data -analytiikan arvo muodostuu päätöksenteon laadun parantumisena. Aiheen tärkeydestä huolimatta, Big Datan hyödyntämisestä päätöksenteossa on heidän mukaansa erittäin vähän aiempaa tutkimusta. Usein saatetaan olettaa Big Datan ja sen analysoinnin luovan itsestään arvoa, mutta tämä saattaa olla liian pelkistetty oletus. Koska monet eri tekijät vaikuttavat päätöksenteon laatuun, Big Data ei itsessään ole ratkaisu optimaaliseen päätöksentekoon. Datan hyödyntäminen yrityksen arvonluonnissa tapahtuu kuitenkin päätöksenteon tehostumisen kautta.

Päätöksenteon tehokkuutta on tutkittu jo hyvin pitkään. Esimerkiksi Keller (1987) tutki informaation laadun ja määrän vaikutuksia päätöksentekoon. Hänen tutkimuksensa osoitti, että tiedon laadulla ja määrällä on selvästi merkitystä päätöksenteon tehokkuudelle. Lisäksi tutkimus osoitti näillä tekijöillä olevan vaikutusta myös keskenään. Vuosien varrella organisaatioiden päätöksenteko on muuttunut ja tarjolla olevan tiedon määrä on kasvanut. Big Data -analytiikka (BDA) on näiden syiden vuoksi noussut viime vuosina esiin hyödylliseksi työkaluksi.

Data voidaan määritellä käsitteellisesti monella tavalla. Tässä työssä datalla viitataan kerättyyn datamassaan, joka koostuu usein vain ykkösistä ja nolista. Kun datalle annetaan jokin tarkoitus, esimerkiksi se luokitellaan tarkoittamaan yksittäisen asiakkaan ostokertoja vuodessa, data muuttuu informaatioksi. Informaatiolla on siis kirjaimellisesti informaatioarvoa. Informaatio ei kuitenkaan itsessään luo arvoa yritykselle. Yhdistelemällä ja analysoimalla informaatiota eri lähteistä, muotoutuu informaatio tiedoksi. Data toimii siis lopulta tiedon lähteenä. Esimerkiksi analysoimalla kaikkien asiakkaiden ostokertoja ja näiden ostotapahtumien fyysisiä sijainteja, syntyy tietoa siitä, mitkä toimipaikat ovat kaikista tehokkaimpia tai tehottomimpia. Tietoa ja informaatiota voidaan tämän jälkeen hyödyntää päätöksenteossa, jotta toimintaa voidaan ohjata haluttuun suuntaan. Tämä suhde on kuvattu kuviossa 1.



**Kuvio 1.** Datan, Informaation, Tiedon ja Päätöksen suhde

Päätöksenteon tehokkuuden lisäksi Big Datan avulla saatu tieto luo yritykselle arvoa kilpailuedun kautta. Tietoperusteisen yritysnäkemyksen mukaan tieto toimiikin yritysten tärkeimpänä strategisena resurssina (Grant 1996). Barney (1991) kuvaa tiedon toimivan kestävän kilpailuedun lähteenä. Hän määrittää kestävää kilpailuetua omaavan yrityksen hyödyntävän sisäisiä vahvuuksiaan, vastaamalla kilpailuympäristön mahdollisuuksiin, neutralisoiden uhat ja ohittaen sisäiset heikkoudet. Yrityksellä on kilpailuetua, kun se toteuttaa arvoa tuottavaa strategiaa, jota kukaan muu sen nykyinen tai potentiaalinen kilpailija ei hyödynnä. Kestäväksi kilpailuetu muuttuu, kun kilpailijat eivät pysty toistamaan tämän strategian hyötyjä (Barney 1991). Tämä ei kuitenkaan tarkoita sitä, että etu kestäisi ikuisesti. Big Data on nykypäivän edistyksellisimmän tiedon lähde, joten se on resurssi, johon on syytä panostaa.

Tieto toimii itsessään resurssina, jota on erittäin vaikea jäljitellä, koska se on luonteeltaan usein hyvin yksilöllistä. Yrityksen kerätessä itse dataa kahdelta vuodelta nykyaikaisilla menetelmillä, menee myös kilpailijalla luonnollisesti muutama vuosi saman datamäärän keräämiseen. Tieto on siis arvokas resurssi, joka voi olla aina askeleen edellä kilpailijoita.



Big Datan ja sen arvoa luovan vaikutuksen tutkiminen on erittäin ajankohtaista sen trendikkyiden vuoksi. Se on monien yritysten ja tutkijoiden kuuma puheenaihe ja sen tuoma kilpailullinen potentiaali on jo laajasti sisäistetty. Özkösen, Arin ja Gencerin (2015) tutkimus Big Datan tutkimuksellisen kiinnostuksen kehittymisestä osoittaa 2011 vuoden olleen käännekohta, jonka jälkeen kiinnostus Big Dataa kohtaan on kasvanut räjähdysmäisesti, eikä kasvu ole vielä hiipumassa. Myös Chenin, Chiangin ja Storeyn (2012) tutkimuksessa tämä käännekohta vuoden 2011 kohdalla oli havaittavissa. Big Datan lisäksi he kuvaavat kiinnostuksen liiketoimintanalytiikan tutkimista kohtaan lähteneen samanaikaisesti kasvuun.

Myös Powerin (2015) tutkimukset ovat osoittaneet, että data-analytiikan tärkeä merkitys on erittäin laajalti tiedostettua. Big Datan merkitys paremmassa arvonluonnissa, on kuitenkin saanut osakseen myös kyseenalaistusta. Big Datan todellista merkitystä tulevaisuuden arvonluonnissa ei siis ole osoitettu vielä tarpeeksi hyvin. Power (2015) kuvaakin Big Datan ja päätöksenteon tutkimuksen olevan vasta alkuvaiheessa. Tutkimusta tarvitaan lisää, jotta voidaan paremmin ymmärtää Big Datan hyödyntämistä eri konteksteissa. Tämän vuoksi koen tärkeäksi, että aiheetä tutkitaan tarkemmin.

Özkösen et al. (2015) ennustavat Big Data -tutkimusten jatkavan kasvuaan. Tämä puolestaan johtaa teknologiseen kehitykseen, asiakastytyväisyyteen, parempaan terveydenhuoltoon ja yleisen hintatason laskuun. Kyse on siis erittäin tärkeästä aiheesta. Parempi ymmärrys Big Datan hyödyntämisestä yritysten arvonluojana kasvattaa aiheen kiinnostavuutta ja lähestyttävyyttä entisestään.

Edellä kuvattujen syiden vuoksi, valitsin tutkimukseni aiheeksi tietolähtöisen päätöksenteon ja datan merkityksen yrityksen menestykseen. Pro Gradu -tutkielmani tarkoituksena on syventyä teoreettisesti tarkastelemaan arvon syntymistä Big Datasta tehokkaamman päätöksenteon kautta. Big Datan ja BDA:n osalta keskeistä tässä työssä on niiden avulla saavutettu tieto, jota päätöksenteossa voidaan puolestaan hyödyntää. Työn tavoitteena on siis tuottaa tutkimusta myös tietojohdami-

sen saralle. Big Dataa käsitellään tässä työssä edistyksellisimpänä datan muotona. Koen itse Big Datan avaimeksi laadukkaampaan tietoon, jonka avulla yritykset voivat olla entistä menestyksekkäämpiä.

Pro Gradun empiirisen osion tavoitteena on tutkia datasta saadun tiedon merkitystä ja vaikutuksia päätöksenteon kautta tapahtuvaan arvonluontiin. Arvonluontia mitataan tässä työssä yrityksen menestymisen kautta. Big Datan hyödyntäminen yrityksissä on Suomessa vasta aluillaan. Empiirisen tutkimuksen osalta onkin tärkeämpää tutkia Big Datan sijaan yleisemmin datan merkitystä menestykseen. Data-analytiikan onnistuminen vaatii yritykseltä sekä teknologisia että strategisia valmiuksia. Näiden merkitystä käsitellään tarkemmin tutkielman teoriaosuudessa. Tutkielmani empiirinen osio tutkii yleisesti datan merkitystä menestykseen tietoperusteisen päätöksenteon ja yrityksen datakyvykkyyksien kautta.

Tutkielman empiirisen osion analysointi toteutettiin rakenneyhtälömallinnuksen avulla. Menetelmän avulla voidaan tutkia useiden eri tekijöiden keskeisiä yhteyksiä. Rakenneyhtälömallinnus mahdollistaa myös tekijöiden yhteyksien rakenteen tarkastelun. Käytännössä menetelmä alkaa tutkimusmallin muodostamisella. Tutkimusmalli perustuu teoreettiseen ymmärrykseen käsiteltävästä aiheesta. Koska mallista on tarkoituksena tutkia muuttujien välisiä suhteita, muodostetaan hypoteesit mallin suhteiden mukaisiksi.

Varsinainen analyysi aloitetaan tutkivalla faktorianalyysillä, jonka avulla aineistosta pyritään löytämään olemassa olevia faktorirakenteita. Faktorirakennetta pyritään muokkaamaan niin, että se vastaisi parhaiten teoreettista mallia. Analyysiä jatketaan tämän jälkeen konfirmatorisella faktorianalyysillä. Tämän analyysin tarkoituksena on testata teoreettisen mallin sopivuutta tutkimusaineistoon. Tarkoituksena on siis parantaa mallia empiiriseen tietoon pohjautuen, pitämällä kuitenkin teoreettiset lähtökohdat mukana. Analyysin pohjalta mallista muodostetaan rakenneyhtälömalli, jonka avulla asetettuja hypoteeseja testataan. Tutkimusmenetelmää käsitellään tarkemmin vielä tämän tutkielman kappaleessa 4.

Tietojohtamisen opiskelijana olen havainnut tiedon keskeisen merkityksen kestäväen kilpailuedun ja arvonluomisessa. Jatkuvasti kilpailullisemmat markkinat vaativat yrityksiltä tavallisia analysointimenetelmiä vaativampia työkaluja. Toisaalta painostusta monipuolisemmalle data-analytiikalle luovat jatkuvasti kasvavat datavirrat, joiden avulla on mahdollista saada täysin uudenlaisia vastauksia päätöksentekoa edistäviin kysymyksiin. Näiden tekijöiden vuoksi, aihe tiedon roolista ja tärkeydestä yritysten menestykselle on oman mielenkiintoni polttopisteessä.

## 1.1 Tutkimuskysymykset

Tämän kappaleen tarkoituksena on kuvata tutkielman tutkimuskysymykset. Koko tutkimus kiteytyy alla esitettyyn päätutkimuskysymykseen, jolla tavoitellaan syvällisempää ymmärrystä datan vaikutuksesta yrityksen menestykseen. Alatutkimuskysymysten tavoitteena on puolestaan tarkentaa ja täydentää päätutkimuskysymystä teoriaosuudessa perustelluista näkökulmista. Lisäksi tässä kappaleessa käsitellään tutkimuksen rajauksia.

Päätutkimuskysymys:

*Miten datan hyödyntäminen päätöksenteossa vaikuttaa yrityksen menestymiseen?*

Alatutkimuskysymykset

*Miten data- ja teknologiakyvykkyudet vaikuttavat yrityksen menestymiseen?*

*Miten data- ja teknologiakyvykkyudet vaikuttavat tietoperusteiseen päätöksentekoon?*

*Miten runsaan ja monipuolisen datan hyödyntäminen vaikuttaa data- ja teknologiakyvykkyysiin?*

Janssen et al. (2017) mukaan Big Datan arvo muodostuu päätöksenteon tehostamisen kautta. Tämän vuoksi tutkielman päätutkimuskysymyksen tarkoituksena on saada selville, miten dataa päätöksenteossa hyödyntävät yritykset menestyvät. Kysymyksen tarkoituksena on tutkia tarkemmin tietoperusteisen päätöksenteon merkitystä menestymiselle. Tavoitteena on saada parempi käsitys tietoperusteisen päätöksenteon ja data-analytiikan roolista yritysten tehokkaalle toiminnalle. Päätutkimuskysymystä lähestytään sekä teoreettisesta että empiirisestä näkökulmasta.

Choon (1996) mukaan kyvykkyudet rajoittavat rationaalista päätöksentekoa. Akter et al. (2017) kokevat puolestaan yritysten tietoon perustuvien kyvykkyyksien muodostuvan teknologisista ja johdon strategisista kyvykkyyksistä. Asian syvällisemmän ymmärryksen saavuttamiseksi tutkielman alatutkimuskysymykset pureutuvat yritysten teknologisten ja strategisten datakyvykkyyksien merkitykseen tietoperusteiselle päätöksenteolle sekä menestykselle. Myös alatutkimuskysymyksiä tarkastellaan sekä teoreettisesti että empiirisesti.

Tutkielmassani käsitellään teoreettisena viitekehyksenä Big Datan, Big Data -analytiikan ja päätöksenteon teorioita sekä niiden välisiä yhteyksiä. Datan rooli nähdään tutkimuksessa päätöksentekoon tarvittavan tiedon lähteenä. Big Datan osalta tutkimus rajautuu käsittelemään BDA-menetelmin aikaansaatuja eksplisiittistä tietoa. Työn tarkoituksena on tutkia eksplisiittisen tiedon liiketoiminnallisen arvon muodostumista ja vaikutusta yrityksen menestymiseen. Päätöksenteon osalta tutkimuksen tarkoituksena on keskittyä tietoperusteisen päätöksenteon vaikutuksiin menestykselle. Päätöksentekoa käsitellään tässä työssä tietoperusteisesta näkökulmasta. Päätöksentekoa määrittää siis käytettävissä oleva tieto.

Tutkimuksesta rajautuu myöhemmin tutkimusmallissa esitelty tärkeä elementti eli laatu. Tutkielman empiirinen tutkimus ei siis syvenny tarkastelemaan datan ja analytiikan laadun merkitystä tietoperusteiselle päätöksenteolle. Tutkimus rajautuu tarkastelemaan tutkimuskohdetta ottamatta kantaa siihen, että datan ja analytiikan laatu saattaa vaihdella todella voimakkaasti eri yrityksissä.

## 1.2 Työn rakenne

Tutkielman toinen kappale käsittelee työn teoreettisen taustan. Kappaleen tarkoituksena on kuvata Big Dataa, Big Data -analytiikkaa, päätöksentekoa ja näiden välisiä yhteyksiä aikaisempaan tutkimustietoon perustuen. Teoriaosuuden tarkoituksena on kuvata teoreettiset lähtökohdat arvon syntymiselle datasta. Tämän perustella aiheesta muodostetaan empiirinen tutkimusmalli. Edellä mainitut käsitteet muodostavat myös tutkimuksen viitekehyksen, johon tutkimusmalli perustuu. Teoreettisten käsitteiden yhteenveto on kuvattuna kappaleen kaksi lopussa.

Kolmannen kappaleen tarkoituksena on kuvata tutkimuksessa käytettävä tutkimusmalli ja tutkittavat hypoteesit. Tutkielman neljäs kappale käsittelee tutkimuksessa käytetyt tutkimusmenetelmät. Kappaleen tarkoituksena on kuvata tutkimuksen kulkua ja tutkimuksen kohde. Kappaleessa esitetään lisäksi tutkimuksessa käytetyn aineiston keruu ja tutkimuksen aineiston analysointi. Tutkielman viides kappale kokoaa analyysin tulokset yhteen ja vastaa asetettuihin hypoteeseihin. Kappaleen tarkoituksena on käsitellä aineiston esittämää vastausta tutkittavaan ongelmaan. Työn kuudes ja viimeinen kappale käsittelee tutkielman johtopäätöksiä. Kappaleessa vastataan tutkimuskysymyksiin sekä pohditaan tulosten merkitystä.

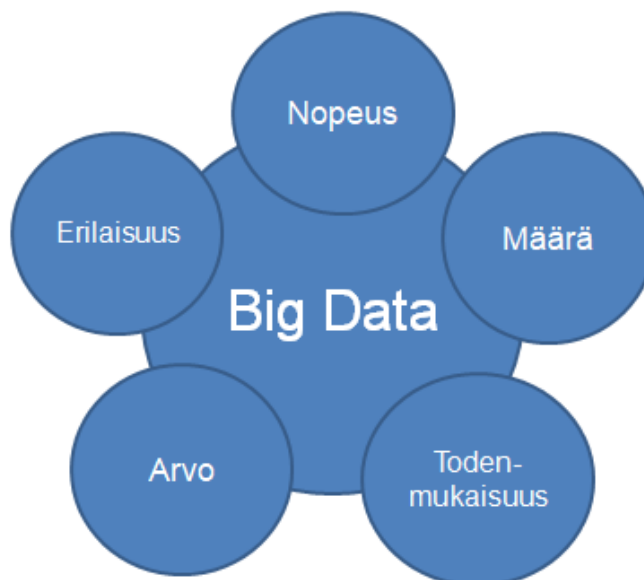
## 2 TEOREETTINEN TAUSTA

Tämän kappaleen tarkoituksena on määritellä ja kuvata tutkielmassani käyttämäni teoreettiset lähtökohdat. Aihealueet toimivat työn viitekehyksenä, johon empiirissä osiossa tarkasteltu tutkimusmalli perustuu. Tämä tutkielma käsittelee Big Dataa ja Big Data -analytiikkaa, päätöksentekoa ja näiden välistä yhteyttä. Päätöksenteon ja Big Datan välistä yhteyttä tarkastellaan aiemman tutkimustiedon valossa. Tutkittavat kohteet muodostavat tutkimuksen viitekehyksen. Niiden yhteenveto kuvataan tämän kappaleen lopussa.

### 2.1 Big Data ja Big Data -analytiikka

Bayraki (2015) mukaan liiketoiminta-analytiikka on noussut potentiaalisesti liiketoiminnan mahdollistajaksi julkisella ja yksityisellä sektorilla. Lisäksi se on yksi nopeimmin kasvavista tutkimusalueista. Analytiikkaa hyödyntämällä organisaatioiden päätöksentekijät voivat integroida irrallisia datalähteitä trendien tunnistamiseksi, tehokkuuden parantamiseksi, uuden liiketoiminnan tunnistamiseksi ja parempien päätösten tekemiseksi (Bayrak 2015). Data-analytiikasta on näin mahdollista hyödyntää paremman päätöksenteon myötä, kun päätökset eivät perustu vain oletusten varaan.

Digitaaliset teknologiat ovat kuitenkin muuttaneet sitä, miten organisaatiot rakentuvat ja toimivat. Datan tallentamismahdollisuudet ovat kasvaneet eksponentiaalisesti muuttaen myös datan keräystapoja. Dataa muodostuu jokainen sekunti suuria määriä ja useista eri lähteistä. Tämä data vaatii uusia menetelmiä sen varastointiin ja analysointiin. (Elgandy & Elragal 2015) Tätä suurta ja monimuotoista datamassaa kutsutaan Big Dataksi.



**Kuvio 2.** Big Data (Özkösen et al. 2015)

Özkösen et al. (2015) mukaan Big Datalle on kirjallisuudessa useita eri määritelmiä. Big Data -määritelmässä yhdistävänä tekijänä on kuitenkin se, että sillä kuvataan korkean nopeuden, määrän ja erilaisuuden omaavaa datamassaa, jota on haastavaa analysoida perinteisillä menetelmillä. Näiden tekijöiden lisäksi osa määritelmistä lisää tekijöiksi vielä datan suuren arvon ja todenmukaisuuden. Nämä tekijät on kuvattu kuviossa 2. Tutkijoiden mukaan Big Datalle on ominaista myös se, että data on monimutkaista. Power (2014) kuvaa Big Datan olevan värikäs fraasi merkitykselliselle muutokselle datan keräämisessä, varastoinnissa ja haussa.

Tärkeimmät Big Dataa määrittävät tekijät ovat erilaisuus (Variety), nopeus (Velocity) ja määrä (Volume). Datan erilaisuudella viitataan rakenteelliseen heterogeenisyyteen kohdedatassa (Gandomi & Haider 2015). Gandomin ja Haiderin (2015) mukaan teknologinen kehitys on mahdollistanut yritysten käyttää strukturoitua, puolistrukturoitua ja strukturoimatonta dataa, jotka ovat tyyliltään hyvin erilaisia. He kuvaavat tämän hyvin olennaiseksi, sillä olemassa olevasta datasta vain 5% on strukturoidussa muodossa. BDA-menetelmät mahdollistavat siis kaiken mahdollisen datan hyödyntämisen. Teksti, kuvat, äänitteet ja videot ovat esimerkkejä strukturoimattomasta datasta.

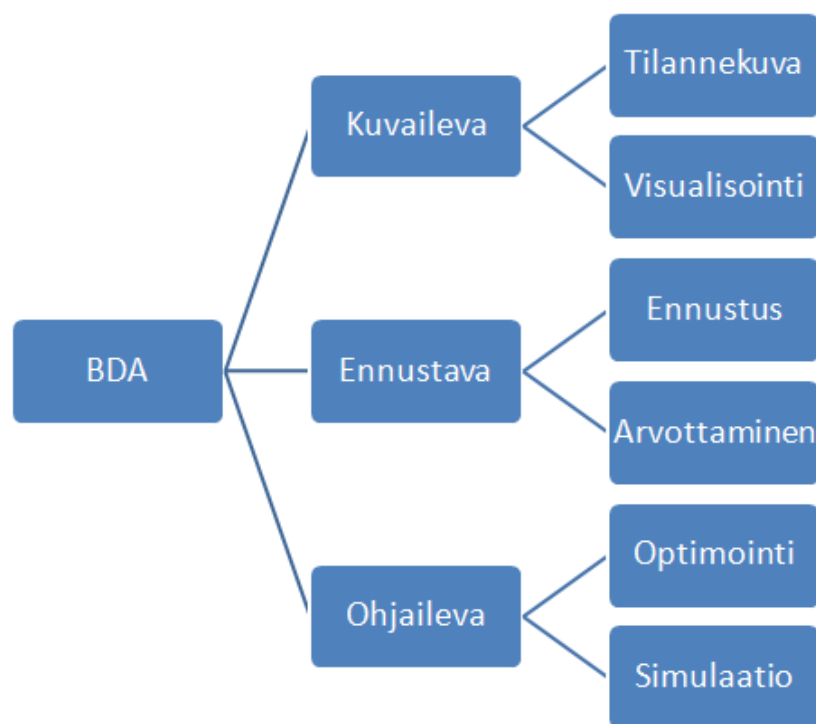
Datan nopeudella viitataan tahtiin, jolla uutta dataa syntyy ja uutta dataa tulisi käsitellä (Gandomi & Haider 2015). Nykyään dataa muodostuu useista eri lähteistä ja sitä syntyy joka sekunti erittäin suuria määriä. Gandomin ja Haiderin (2015) mukaan tämä on nostanut esiin tarpeen reaaliaikaiselle analytiikalle. Esimerkiksi mobiiliapplikaatioiden läpi muodostuva data, mahdollistaa reaaliaikaisen ja personalisoitujen tarjousten luomisen asiakkaille. BDA-menetelmät mahdollistavat siis arvonluonnin sellaisista datalähteistä, joita perinteisillä menetelmillä on mahdotonta analysoida tarpeeksi nopeasti.

Datan syntymisen suuri nopeus on syy sille, miksi dataa on myös erittäin suuria määriä. Dataa on usein jopa tera- tai petatavuja. Teratavallinen dataa mahtuu esimerkiksi 1500 CD-levyyn tai 220 DVD:hen. Big Datan määrittäminen määrän perusteella on kuitenkin erittäin suhteellista ja se vaihtelee tekijöistä, kuten ajasta ja datan tyypistä riippuen. Datan tyyppi määrittää erittäin paljon sitä, mikä koetaan suureksi määräksi kyseistä dataa. Tämän vuoksi Big Datan määrittäminen määrän perusteella koetaan osittain epäkäytännölliseksi. (Gandomi & Haider 2015) Tärkeintä määrän osalta on kuitenkin se, että suuri datamäärä sisältää usein tavanomaista tarkempaa informaatiota tutkittavasta kohteesta. Tämän vuoksi suurten datamäärien käyttäminen päätöksenteossa on erittäin mielekästä.

Jotkut määritelmät lisäävät arvon (Value) yhdeksi tekijäksi. Vaikka Big Data ei ole tyypiltään arvotiheää, analysoimalla sitä suurissa määrissä, voidaan siitä saada aikaan erittäin arvokasta tietoa (Gandomi & Haider 2015). Özkösen et al. (2015) määrittelevät Big Datan tarkemmin sen luokkien kautta. Heidän määritelmänsä mukaan Big Datan lähteenä voi olla esimerkiksi Web ja sosiaalinen media, laite tai IoT (Internet of Things) -sensori. Muodoltaan data voi olla strukturoitua, puolistrukturoitua tai epämuodollista. Data varastoidaan dokumentteihin, kolumneihin, graafeihin tai avainlukuihin. Datan prosessointi puolestaan voidaan toteuttaa, joko erissä tai reaaliajassa. Data voi siis olla erittäin monimuotoista ja sitä voidaan kerätä useasta eri lähteestä.



Wamba, Gunasekaran, Akter, Ren, Dubey, Childe (2017) toteavat Big Data -analytiikan eli BDA:n olevan avaintekijänä liiketoiminnan tehokkuuden ja vaikuttavuuden näkökulmista, sen operatiivisen ja strategisen potentiaalin perusteella. BDA toimii vaiheena, joka luo arvoa Big Datalle. Aikaisempien tutkimusten perusteella, BDA:n hyödyntämisellä on osoitettu olevan positiivinen vaikutus yrityksen suorituskykyyn (Germann & Lilien et al. 2014). BDA mahdollistaa yrityksen kehittymisen entistä ketterämmäksi ja tulevaisuusorientoituneemmaksi. BDA-menetelmät ovat myös ainoa tapa valjastaa Big Datan arvo yrityksen käytettäväksi.



**Kuvio 3.** Big Data -analytiikka (Pyne, Rao & Rao, 2016)

Pynen, Raon ja Raon (2016) mukaan BDA jakautuu kuvailevaan (Descriptive), ennustavaan (Predictive) ja ohjailevaan (Prescriptive) analyysiin. Nämä kolme pääluokkaa ovat käytännössä myös toistensa jatkumoa. Nämä luokat on kuvattu kuviossa 3. Kuvaileva analyysi louhii suurista datamääristä potentiaalisia yhteneväisyyksiä (Pyne, Rao et al. 2016). Tätä kutsutaan termillä Datan louhinta (Data mi-

ning). Kuvailevan analyysin avulla on mahdollista kuvata tai jopa visualisoida mennyttä tai nykyistä tilaa. BDA mahdollistaa siis myös reaaliaikaisen tilanteen seurannan esimerkiksi tuotantolaitteen tilaa mittaavien sensorien avulla. Kuvailevalla analyysillä pyritään vastaamaan kysymykseen ”*Mitä on tapahtunut?*” (Pyne, Rao et al. 2016).

Seuraava keskeinen BDA:n luokka on ennustava analyysi. Tämän tyyppiset analysointimenetelmät siirtyvät menneen kuvaamisesta tulevaisuuden ennustamiseen. Ennustava analytiikka pyrkii tyypillisesti vastaamaan kysymykseen ”*Mitä todennäköisesti tulevaisuudessa tapahtuu?*” (Pyne, Rao et al. 2016). Waller ja Fawcett (2013) kuvaavat ennustavan analytiikan pyrkivän nopeisiin ja kustannustehokkaisiin arvioihin eri muuttujien suhteiden välillä johtopäätösten muodostamiseksi. Ennustava analytiikka ei siis voi koskaan tuottaa sataprosenttisesti varmaa tulevaisuutta (Pyne, Rao et al. 2016). Se kuitenkin osoittaa menneiden tapahtumien perusteella vaihtoehtoisia tulevaisuuksia ja näiden toteutumisprosentteja. Näitä voidaan puolestaan hyödyntää tietolähtöisemmässä päätöksenteossa. Ennustavan analytiikan avulla on esimerkiksi mahdollista saada ennusteita tulevasta tai arvottaa asiakkaan riskiprofiilia.

Ohjaileva BDA menee ennustavasta analytiikasta vielä askeleen eteenpäin. Se pyrkii vastaamaan kysymykseen ”*Miten saamme sen toteutumaan?*” (Pyne, Rao et al. 2016). Tarkoituksena on siis tutkia sitä, mitä toimenpiteitä yrityksen tulee tehdä, jotta se saavuttaa tietyn tavoitteen analyysin ennustamassa tulevaisuudessa. Yleisimpiä työvälineitä tähän ovat optimointi ja simulaatiot, joiden avulla etsitään optimaalisinta reittiä saavuttaa haluttu tulevaisuus. Tämä BDA-menetelmä on kuitenkin verrattain uusi. Päätöksenteossa ohjailevaa analytiikkaa käyttää vain 3 % yrityksistä (Pyne, Rao et al. 2016). Optimaalisimman ratkaisun löytämiseksi ohjailevassa analytiikassa hyödynnetään sekä kuvailevaa että ennustavaa analytiikkaa (Hamilton 2015).

Tiivistetysti Big Datan ja BDA:n osalta keskeistä on datan monimuotoisuus ja se, että tarkoituksena on etsiä trendejä ja yhteneväisyyksiä, joita on erittäin vaikea tai mahdotonta havaita muita menetelmiä käyttäen. Big Datan ja BDA:n tulevaisuus on Özkösen et al. (2015) mukaan nousujohteista. Yritykset, joilla on edellytykset suurten datamäärien käsittelyyn, vievät oman työnsä askeleen edelle muita toimijoita. Esimerkkiyrityksiksi he kuvaavat Googlen, Amazonin, Facebookin ja eBayn. Näillä kaikilla on kilpailu- ja kehittymisetua suurten datamäärien vuoksi.

Bizerin, Bonczin, Bordien ja Erlingin (2011) mukaan Big Datan maailmassa on kaksi suurta haasteluokkaa. Ensimmäisen haasteen asettaa tekniikka, sillä käsiteltävää dataa on uskomattoman paljon. Toisena haasteena on puolestaan semantiikka. Datasta on löydettävä merkittävää tietoa ja tätä tietoa on yhdisteltävä ongelman vaatimalla tavalla. Ilman tarkoituksenmukaista datan hyödyntämistä tekniikka ja työkalut ovat vain hienoja temppuja.

Big Data -analytiikan osalta on myös tärkeää ymmärtää sen arvon syntyvän varsinaisen työkalun lisäksi monien muiden tekijöiden summana. Akter, Wamba, Gunasekaran, Dubey ja Childe (2016) kuvaavat Big Data -analytiikkakyvykkyyksien muodostuvan BDA johdon, teknologian ja taidon kyvykkyyksistä. Johdon kyvykkydet pitävät sisällään BDA:n suunnittelun, investoinnit, koordinaation ja kontrollin. Teknologiset kyvykkydet puolestaan muodostuvat järjestelmän yhdistettävyydestä, yhteensopivuudesta ja modulaarisuudesta. BDA taidot viittaavat puolestaan yksilön tekniseen osaamiseen ja liiketoiminta ymmärrykseen. Näiden tekijöiden summana muodostuvat BDA-kyvykkydet tehokkuuden luomiseksi. Pelkän BDA-työkalun käyttöön ottaminen ilman kykyä hyödyntää sitä oikein ei siis takaa toimivaa arvonluontia datasta.

Esimerkkejä BDA:n tehokkaalla hyödyntämiselle on paljon. Big Datan avulla voidaan esimerkiksi laskea riskiportfolioita ja ennustaa tulevaa riskien minimoimiseksi. Asiakasdatasta on mahdollista löytää oivalluksia strategisen suunnan löytämiseksi, kampanjoiden optimoimiseksi tai tärkeimpien asiakkaiden tunnistamiseksi. BDA:n

avulla voidaan myös lähettää räätälöityjä suosituksia, kun asiakkaat ovat oikeassa paikassa tarjouksen hyödyntämiseksi. Big Data saattaa myös paljastaa epäonnistumisten tai ongelmien syyt. (Power 2014)

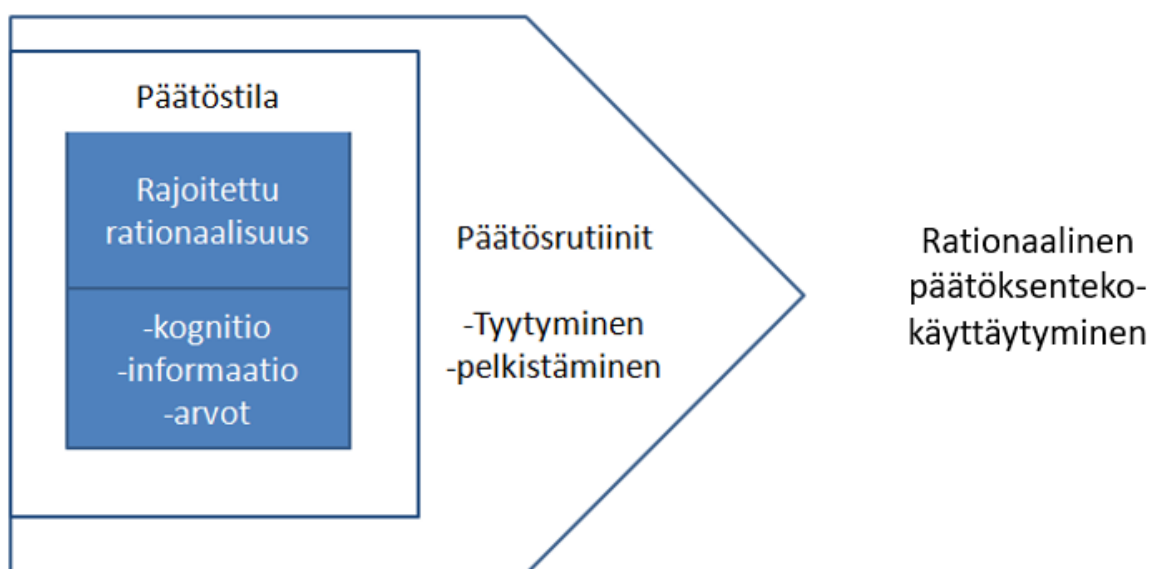
## 2.2 Päätöksentekoprosessi

Kaikkien päätösten tarkoituksena on löytää ja valita toiminnalle suunta, joilla saavutetaan tietyt tavoitteet. Päätösten tekemisen vaikeus riippuu usein siitä, kuinka selkeitä nämä tavoitteet ovat ja kuinka hyvin tiedämme eri vaihtoehdoista, joilla tavoitteet voidaan saavuttaa. (Choo & Johnston 2004) Tämän vuoksi Big Dataan perustuva päätöksenteko ei itsessään eroa muusta päätöksenteosta. Päätöksiä tehdään useilla eri organisaatiotasoilla ja ne saattavat olla tyypiltään pieniä tai suuria, riippuen tavoitteista, joihin päätöksellä pyritään. Päätöksen suuruus voi vaikuttaa myös päätöksen vaatimaan informaation määrään. Big Data mahdollistaa erittäin monipuolisen lähteen päätöstentoon perustaksi.

Ideaalissa maailmassa rationaalinen päätös vaatisi täyden tutkimuksen mahdollisista vaihtoehdoista, luotettavaa tietoa eri vaihtoehtojen seurauksista ja mahdollisuutta arvioida niiden lopputulemia. Todellisessa maailmassa tällaiset vaatimukset informaation keräämiselle ja prosessoinnille ovat epärealistisia. Objekttiivisen rationaalisuuden sijaan päätöksenteko organisaatioissa tulisi nähdä rajoitetun rationaalisuuden käsitteen kautta. (Simons 1957) Päätöksenteossa on siis tehtävä pelkistyksiä ja välillä tyydyttävä siihen tietoon mitä on saatavilla.

Choon (1996) tutki myös Simonsin rajoitetun rationaalisuuden käsitettä, jonka mukaan ihmismielen kapasiteettiä rationaaliselle päätöksenteolle rajoittaa kyvykkyydet, tavat ja refleksit, tiedon ja informaation määrä sekä arvot, jotka saattavat syntyä organisaation tavoitteista. Rajoitetusta rationaalisuudesta johtuen Choo (1996) esittää kahtiajaon päätöksenteon tavoille. Ensiksi päätöksentekijä tyytyy. Tämä tarkoittaa, että päätöksentekijä etsii tyydyttäviä tai tarpeeksi hyviä vaihtoehtoja op-

timaalisen ratkaisun sijaan. Tyypillisesti suurin osa päätöksenteosta on tyydyttävien valintojen etsimistä ja valintaa. Toiseksi päätöksentekijä *pelkistää* päätöksentekoprosessia. Päätöksentekijä siis seuraa rutiineja ja hyödyntää nyrkkisääntöjä välttääkseen epävarmuutta ja vähentääkseen monimutkaisuutta.



**Kuvio 4.** Päätöksentekosysteemi (Choo 1996).

Kuvio 4 kuvaa avaintekijöitä organisaatiosta päätöksentekosysteeminä. Organisaatiot etsivät rationaalista käyttäytymistä toimenpiteille, jotka vaikuttavat sen tavoitteisiin. Yksittäiset ihmiset tekevät kuitenkin lopulta päätökset ja heidän käytöstään rajoittaa henkilökohtaiset kognitiot, informaatio ja arvot. Tapa ylittää kuilu organisaation rationaalisuuden ja yksilöiden rajoitetun rationaalisuuden välillä on luoda päätöstila ja -rutiinit, jotka ohjaavat ja kontrolloivat yksilön käytöstä. (Choo 1996)

Nutt (1984) tutki lähes kahdeksaakymmentä tapaustutkimusta päätöksenteosta tunnistaakseen prosessin luonteen. Tutkimus paljasti useita eri prosessityyppejä arvioivasta ja tutkivasta historialliseen malliin. Nämä mallit eroavat esimerkiksi ideoiden luonnin tavan ja prosessin perustelujen osalta. Tutkimus osoitti myös sen, että johtajat eivät käytä tutkijoiden esittämiä ohjeellisia metodeja parempaan päätöksentekoon. Suurimman osan päätöksenteosta huomattiin olevan ratkaisukes-

keistä. Tämä nähtiin rajoittavan innovaatioita ja rajoittavan huomioon otettavien vaihtoehtojen määrää sekä edistävän kyseenalaisten taktiikoiden käyttämistä.

Schwarber (2005) kuvaa päätöksenteon tapahtuvan keskusteluiden kautta, joita ihmiset käyvät joko yksittäisten ihmisten, tiimien tai toimintojen välillä. Ongelmaksi näissä keskusteluissa, hän näkee ihmisten tavan puhua eri tarkoituksista, jakaen tietoa umpimähkään ja käyden samaa asiaa läpi uudelleen ja uudelleen, pääsemättä lopputulokseen. Päätöksenteko on paljon tehokkaampaa, kun se on fokusoitua ja keskustelu ohjattua.

Kolmeksi kriittiseksi päätöksenteon tekijäksi Schwarber (2005) listaa tavoitteen, vaihtoehdot ja riskit. Nämä elementit on löydyttävä päätöksenteosta, jotta se voi olla onnistunutta. Hänen mukaansa hyvä päätöksenteko ei ala vaihtoehdoilla vaan tavoitteilla. Tätä käsitystä kuitenkin haastaa BDA:an liittyvä Data mining eli datan louhinta. Datan louhinnassa suurista datamassoista etsitään tarkoituksella uusia trendejä ja yhteneväisyyksiä. Datan louhinnan avulla päätöksenteko on siis mahdollista aloittaa vaihtoehdoista, joita datasta löydetään.

Kun tavoitteet on asetettu, havaittuja vaihtoehtoja tulee riskiarvioida ja parhaita työstää eteenpäin. Lopullinen valinta eli päätös syntyy, kun nämä kolme elementtiä on otettu huomioon. (Schwarber 2005) Näiden tekijöiden osalta Big Data -analytiikka tarjoaa paljon uusia mahdollisuuksia ja helpotuksia. Big Data -analytiikan avulla on mahdollista saavuttaa käsitys kohteena olevan asian tilasta tasolla, jota yksittäinen ihminen ei pysty käsittämään. Analytiikka saattaa myös tuoda esiin sellaisia vaihtoehtoja, jotka jäisivät muuten täysin käsittelemättä.

Lyhyesti Nutt (1984) linjaa päätöksenteon olevan valikoima toimintaa, joka alkaa päätöksenteon kohteena olevan asian tunnistamisella ja päättyy toteutukseen. Tämä määritelmä toimii myös tämän tutkielman päätöksenteon lyhyenä määritelmänä. Työn kannalta ei siis ole olennaista, mitä päätöksenteon mallia yksittäinen yritys käyttää, sillä päätöksentekoon liittyy aina yhteisiä tekijöitä.

Tarve reagoida ja sitä seuraava kiire tehdä päätöksiä
Päätöksentekijän tietovaranto. Tarkennettuna informaation prosessointi ja hyödyntämiskyky
Päätöksentekijän persoonallisuus, kuten älykkyys, vastuunottokyky, kyky työskennellä yhteistyössä ja päästä sopimukseen, tavoitteellisuus ja itsenäisyys
Päätöksentekijän näkemys tavoitteiden käyttökelpoisuudesta
Tieto käsiteltävästä ongelmasta. Päätöksenteon oikeellisuus vaatii erilaista informaatiota tapahtumista, faktoista ja ilmiöistä
Päätöksentekoon tarvittavan informaation keräämisen ja prosessoinnin kustannus, joka ei voi olla korkeampi kuin tavoiteltu hyöty
Normit
Päätöksenteon tila, johon liittyy varmuus potentiaalisista vaihtoehdoista ja niihin liittyvistä riskeistä
Mahdollisuus tehdä päätöksiä asiaankuuluvalla vapauden marginaalilla

**Taulukko 1.** Päätöksentekoa määrittävät tekijät. (Kościelniak, Puto 2015)

Tämän työn kannalta merkittävimmät päätöksentekoa määrittävät tekijät on listattu taulukkoon 1. Kościelniakin ja Puton (2015) mukaan nämä tekijät määrittävät päätöksentekoa ja päätöksenteon tehokkuutta. Kuviossa 3 kuvatun päätöksentekosysteemin kannalta nämä tekijät määrittävät tarkemmin päätöstilaa ja rutiineja. Tekijät liittyvät siis vallitsevaan käsitykseen päätöksenteon kohteesta, tavoista toteuttaa päätöksentekoa ja kyvystä ymmärtää mahdollisia ratkaisuja. Päätöksiä on pystyttävä tekemään nopeasti ja laadukkaasti hyödyntäen mahdollisimman tarkkaa tietoa. Tekijöistä on havaittavissa se, että Big Datalla pystytään vaikuttamaan niistä useisiin. Big Datasta saatu entistä tarkempi tieto esimerkiksi vähentää inhimillisten tekijöiden merkitystä, kun kerättyyn dataan on mahdollista luottaa enemmän.

Kościelniak ja Puto (2015) tarkentavat vielä päätöksenteon tekijöiden vaikutuksia päätöksentekoprosessin laatuun neljän säännön kautta. Säännöistä ensimmäinen on fokus (Focus), joka viittaa siihen, että päätöksentekijän tulisi keskittyä aiheisiin,

jotka ovat sille kaikista tutuimpia. Tämä synnyttää yksilökohtaisia kompetensseja päätöksentekoon. Toinen sääntö on nopeus (Fast). Nopein mahdollinen reagointi ulkoisiin ja sisäisiin signaaleihin vahvistaa päätösten laatua. Kolmanneksi säännöksi tutkijat esittävät parhaan sijoituksen (First), jolla he viittaavat vaatimukseen täyttää asiakkaiden tarpeet parhaiten markkinoilla. Neljäntenä sääntönä on jousto (Flexibility), jonka mukaan innovatiivisuus ja muutoskyky resursseissa ja rakenteissa parantavat päätöksentekokyvykkyyksiä. Big Data on nykypäivän kuuma puheenaihe. Sen avulla on mahdollista havaita esimerkiksi trendejä reaaliajassa. Big Datan hyödyntäminen päätöksenteossa vastaa näihin sääntöihin erittäin hyvin.

Tietoperusteinen päätöksenteko perustuu tiedon hyödyntämiseen päätöksenteon lähteenä. Kaner ja Karni (2004) esittävät tiedon johtamiselle neljä aktiviteettitasoa, jotka ohjaavat tiedon käsittelyä päätöksenteossa. Ensimmäinen taso on tiedon hankinnan taso, jolla viitataan tiedon luomiseen ja hankkimiseen esimerkiksi innovaatioiden, päätöksenteon ja tukijärjestelmien kautta. Toinen aktiviteettitaso on tiedon järjestelyn taso, jossa tietoa säilytetään, suojataan ja organisoidaan. Kolmantena tietoa arvioidaan vakuutustasolla. Tasolla pyritään vakuuttumaan siitä, että tieto on relevanttia, tarkkaa, yhtenevää ja arvokasta. Viimeisellä tiedon käyttämisen tasolla tietoa valitaan, siirretään, fokusoidaan ja hyödynnetään ongelmien ratkaisemiseksi ja päätösten tekemiseksi.

Tietoperusteisesta päätöksenteosta käytetään myös nimitystä dataperusteinen päätöksenteko (data-driven decision making). Provost ja Fawcett (2013) käyttävät tätä termiä korostaakseen tiedon datalähtöisyyttä ja Big Datan merkitystä nykymaailmassa. Esimerkiksi he esittävät markkinoijan voivan valita käytettävät mainokset perustuen hänen pitkään kokemukseensa alalla ja hänen omaan arvioonsa siitä, mikä toimii. Hän voi myös perustaa valinnan data-analyysiin, joka osoittaa kuinka asiakkaat reagoivat tiettyihin mainoksiin. Tutkijoiden mukaan valinta todellisuudessa usein perustuu kuitenkin kolmanteen tapaan, joka on näiden lähestymistapojen yhdistäminen. Tutkijat eivät pidä dataperusteisuutta kaikki tai ei mitään -menetelmänä. Henkilön kokemukset luovat hiljaista tietoa, jota on erittäin vaikea

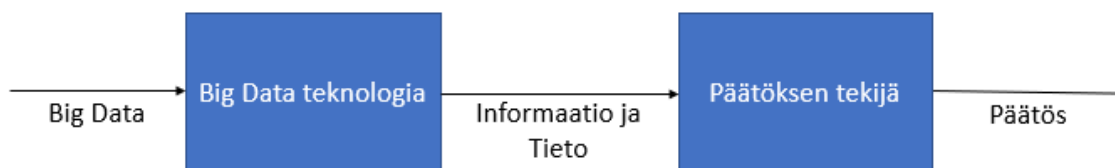


sisällyttää data-analyyseihin. Kokemuksen luoma tieto saadaan tarvittaessa päätökseen mukaan päätöksentekijän kautta.

### 2.3 Big Datan ja päätöksenteon yhteys

Tutkimusta Big Datan ja päätöksenteon välisistä yhteyksistä on vielä verrattain vähän. Tämän kappaleen tarkoituksena on kuitenkin syventyä siihen, mitä aikaisemmat tutkimukset ovat aiheesta havainneet. Kappaleen tarkoituksena on esitellä tämän tutkielman empiirisen osion kannalta keskeisimpiä havaintoja Big Datan hyödyntämisestä päätöksenteossa.

Informaatio on avainasemassa päätöksenteon tuloksellisuudelle, erityisesti päätösten laadulle. Data koetaan 2000-luvun raakamateriaaliksi, jota analysoimalla on mahdollista saada arvokasta tietoa. Tarkemmin sanottuna, päätöksentekijöiden on pystyttävä saavuttamaan toiminnan kannalta arvokkaita oivalluksia hyödyntämällä datamassoja, jotka ovat suuria koon, nopeuden ja datan erilaisuuden vuoksi. (Elgandy & Elragal 2015) Tämä analyysin tuottama arvokas tieto toimii päätöksentekoa tukevana elementtinä. Päätökset voivat perustua siis enemmän tietoon, kuin ihmisten oletuksiin ja kokemuksiin. Big Datan luonteen vuoksi, sitä hyödyntämällä on mahdollista ottaa huomioon paljon useampia tekijöitä, kuin mitä yksittäinen ihminen pystyy näkemään tai kokemaan.



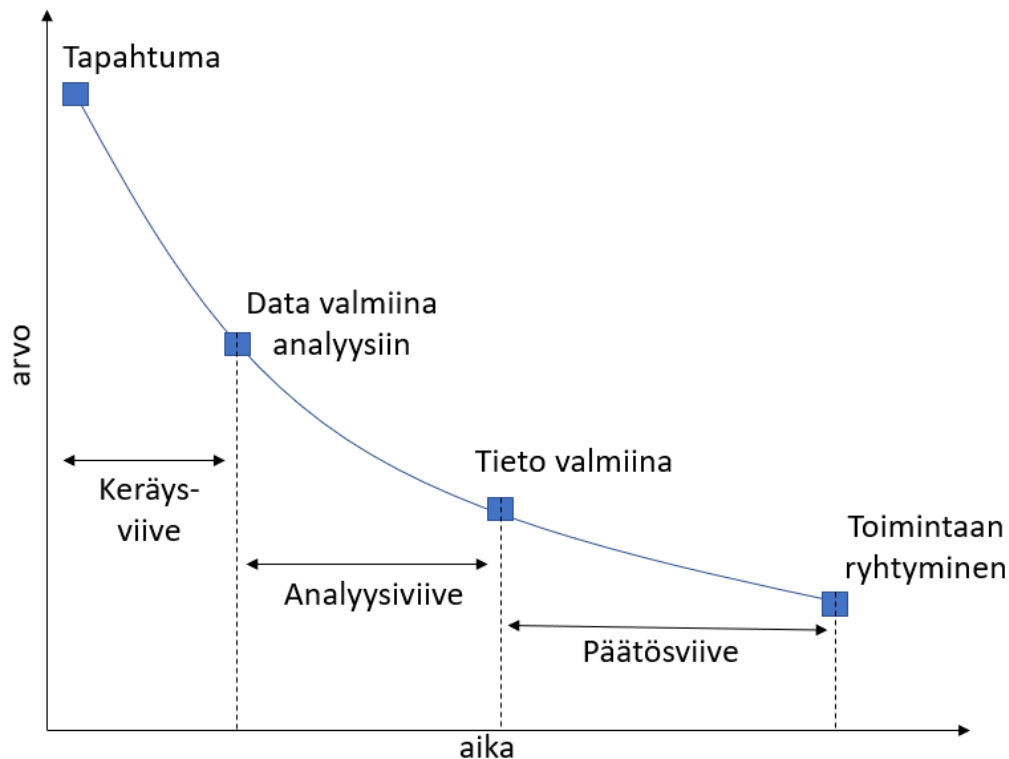
**Kuvio 5.** Big Datan ja päätöksenteon yhteys (Raghunathan 1999)

Raghunathan (1999) käytti tutkimuksessaan kuviota 5 vastaavaa syöte-prosessi-tuotos -mallia, kuvaamaan datan yhteyttä päätöksentekoon. Mallin osalta hän esittää tärkeäksi sen, että tuotoksen eli päätöksen osalta laatu riippuu syötteen ja pro-

sessin laadusta. Koska datasta saatua tietoa hyödynnetään päätöksenteossa, johdavat huonot datalähteet ja huono data huonoihin päätöksiin. Mogesin, Vlasselaerin, Lemahietun ja Baesensin (2016) mukaan datan laadun vaikutus päätöksenteon laatuun on tiedetty fakta ja se on erittäin kriittinen elementti, kun dataa on määrällisesti todella paljon. Tämän vuoksi datan laatu on noussut myös huolenaiheeksi organisaatioissa, jotka hyödyntävät Big Dataa.

Raghunathanin (1999) tutkimuksen mukaan informaation ja päätösten laadun välillä on selkeitä yhteyksiä. Esimerkiksi parempi informaation laatu parantaa myös päätösten laatua, mikäli valittu päätöksentekoprosessi on tilanteeseen soveltuva. Toisaalta hänen tutkimuksensa osoitti, että mikäli päätöksentekoprosessi on valittu väärin operatiivisissa ongelmissa, informaation laatu heikentää päätöksen laatua. Hän kuvaa informaation laadun olevan todella moniulotteinen käsite. Informaation laatuun vaikuttaa esimerkiksi datan olennaisuus, konteksti, merkittävyys ja saavutettavuus.

Informaation ja päätösten laadun lisäksi Kubina, Varmus ja Kubinova (2015) esittävät Big Datan ja Big Data -teknologioiden olevan tärkeässä asemassa yritysten päätöksenteolle, koska ne auttavat yrityksiä ymmärtämään ja määrittämään tarkoituksia suuresta määrästä maailmaa koskevaa informaatiota. He korostavat datan käytön olevan hyödyn tuoja. Tutkijat kuten Power (2014) kuvaavat datan olevan suuri- tai pienikokoisena arvotonta itsessään. Data ei siis suoraan takaa arvoa tai menestystä yritykselle. Power (2014) esittää datan arvon avautuvan vasta kontekstin ja esitystavan kautta. Tämän vuoksi hän korostaakin sitä, että ensin on ymmärrettävä mitä datalähteillä voidaan ja halutaan tehdä. Laadun ja teknologian lisäksi Big Datalle on siis asetettava tehtävä ja selkeät tavoitteet. Haasteena Big Datan hyödyntämiselle on kuitenkin usein yritysjohton haluttomuus panostaa laitteistoihin ja ohjelmistoihin datan varastoinniksi, joka saattaa osoittautua hyödyttömäksi.



**Kuvio 6.** Reaktioajan viive (Pigni et al. 2016)

Big Datan ja päätöksenteon välille asettuu vielä yksi keskeinen tekijä, joka on aika. Pigni, Piccoli ja Watson (2016) tuovat esiin aiempaan tutkimukseen perustuvan reaktioajan viiveen käsitteen, joka kuvaa nopeuden merkitystä arvonluonnille Big Datasta. Tämä on kuvattu kuviossa 6. Malli kuvaa liiketoiminta-arvon pienentymisen, joka johtuu eriasteisista viiveistä. Tarvittavan datan kerääminen tutkittavasta tapahtumasta aiheuttaa keräysviiveen. Datan analysoimiseen kulunut aika puolestaan luo analyysiviiveen. Analyysin avulla saadun tiedon ja toimintaan johtavan päätöksen aikaansaamisen välille jää puolestaan päätösviive. Mitä enemmän reagointiin kuluu aikaa, sitä arvottomammaksi tieto tapahtumasta muuttuu. Tämän vuoksi Pigni et al. (2016) nostavat esiin myös ennustavan data-analytiikan edun reaktioajoissa.

Reaktioajan viive -mallista on tärkeää havaita se, että pienentämällä viiveitä jokaisella tasolla, saavutetaan parhaiten arvoa. Teknologian kehittäminen esimerkiksi Big Data -tasolle pienentää keräys- ja analyysiviivettä muun muassa reaaliaikaisen analyysin avulla, mahdollistaen samalla päätöksenteon tehokkuuden parantumisen esimerkiksi päätöksenteon laadun ja automaation kautta.

Big Data kuitenkin luo mahdollisuuden arvonluonnille päätöksenteossa. Aikaisempi tutkimus osoittaaakin Big Datalla olevan selviä hyötyjä päätöksentekoon. Yksi tärkeimmistä tavoista, joilla yritys voi luoda arvoa Big Datasta, on Big Data -analyysin tuottama tuki päätöksenteolle (Kubina, Varmus et al. 2015). Analyysi voi esimerkiksi paljastaa piileviä korrelaatioita tai riskejä. Big Datan avulla matalan tason päätöksenteko on mahdollista automatisoida jopa kokonaan. Kubina et al. (2015) toteavat myös, että Big Datan käyttäminen mahdollistaa korkeamman informaation läpinäkyvyyden organisaatioiden sisällä. Laajempi, syvempi ja tarkempi käsitys tarkasteltavasta kohteesta parantaa siis päätöksentekoa.

Brown, Chui ja Manyika (2011) kuvaavat Big Datan opastavan yrityksiä päätöksentekoprosessissa. Big Data -analytiikan avulla on mahdollista esimerkiksi testata eri päätösten vaikutusta, joka puolestaan voi johtaa parempaan kilpailuasemaan. Tutkijat toteavat johtavien verkkoyritysten olevan aktiivisia testaajia. Esimerkiksi fyysisiä tuotteita myyvä yritys voi hyödyntää testitilanteita päätöksenteon tukena, mutta Big Data -tasaisen tiedon hyödyntäminen nostaa tämän aivan uudelle tasolle.

Brown et al. (2011) korostavat myös sitä, että aina kontrolloitujen kokeiden tekeminen ei ole mahdollista. Yritykset voivat kuitenkin tehdä luonnollisempia testejä esimerkiksi tehokkuuden vaihteluiden tutkimiseksi. He nostavat esimerkiksi tilanteen, jossa julkisen sektorin yritys testasi samanlaista työtä, mutta eri kohteissa tekevien työntekijöiden tehokkuutta. Pelkkä datan nostaminen avoimeksi nosti tehottomien työntekijöiden tehokkuutta. Chaen, Yangin, Olsonin ja Sheun (2014) tutkimus datan ja edistyksellisen analytiikan hyödyistä osoitti niillä olevan selvä yhteys liiketoiminta-arvoon operatiivisen tehokkuuden kautta. Edistyksellisellä data-analytiikalla

viitataan heidän tutkimuksessaan esimerkiksi optimointiin, joka toimii osana ohjaavaa Big Data analytiikkaa.

Big Datan hyödyntäminen voi jopa muuttaa päätöksentekoprosessin pääläelleen, kun suuresta datamäärästä on mahdollista etsiä trendejä, jotka voivat osoittautua organisaatiolle mahdollisuuksiksi tai heikkouksiksi (Economist Intelligence Unit. 2012). Big Data on muuttunut valtavirran aktiviteetiksi organisaatioissa mahdollistaen täysin uudenlaiset perusteet päätöksenteolle (Janssen, Van der Voort et al. 2017). Päätöksentekoa ei enää ole pakko aloittaa kysymyksellä, johon etsitään vastauksia. Trendit ja yhteneväisyydet kerätyssä datassa voivat osoittaa yritykselle täysin uusia toimintavaihtoehtoja.

Big Data on uusi kilpailuedun lähde kaikilla toimialoilla ja kaikissa yrityksen toiminnoissa. Big Data -ratkaisulla on potentiaalia antaa yrityksille ennennäkemätöntä tietoa asiakkaiden käytöksestä ja epävakaasta kilpailuympäristöstä. Tämä mahdollistaa yrityksille kilpailijoita nopeammat ja tehokkaammat liiketoimintapäätökset. Monilähteyinen data erilaisista sensoreista, Webistä tai sosiaalisesta mediasta, parantaa analyttisiä kyvykkyyksiä ja liiketoimintaketteryttä. (Ziora 2015)

Economist Intelligence Unitin (2012) tutkimuksen mukaan Big Datan hyödyntäminen oli jo nostanut organisaatioiden tehokkuutta keskimäärin 26 % vuosina 2009–2012. Tutkimuksen mukaan Big Dataa hyödynnetään päätöksenteon tukena 58 % kerroista ja 29 % käytöstä tapahtuu päätöksenteon automaationa. Big Data on siis jo osa monen yrityksen päätöksentekoa. Siitä on ollut selvää hyötyä päätöksentekoprosessin tehokkuudelle. Tämä on lupaavaa tietoa niille yrityksille, jotka kärsivät hitaan päätöksenteon aiheuttamista ongelmista.

Kościelniakin ja Puton (2015) tutkimuksen mukaan suurten ja pienten datajoukkojen tehokas käyttö vaikuttaa päätösten laatuun yritysten nykyaikaisissa toiminnoissa. He eivät kuitenkaan kuvaa tätä helpoksi tehtäväksi, koska nykyaikaiset infor-

maatio- ja kommunikaatioteknologiat ovat hyvin monimutkaisia laitteiston, ohjelmiston ja datan prosessointiratkaisujen osalta.

Data- ja tietoperusteisesta päätöksentekonäkökulmasta Power (2015) kuvaa parhaiden päätösten olevan dataan perustuvia ja tämän vuoksi dataa on analysoitava enemmän. Hän kertoo, että yhdeksän kymmenestä johtajasta uskoo datan olevan neljäs tuotannontekijä pääoman, työvoiman ja maan kanssa. Big data siis edustaa perusteellista muutosta yritysten päätöksenteolle.

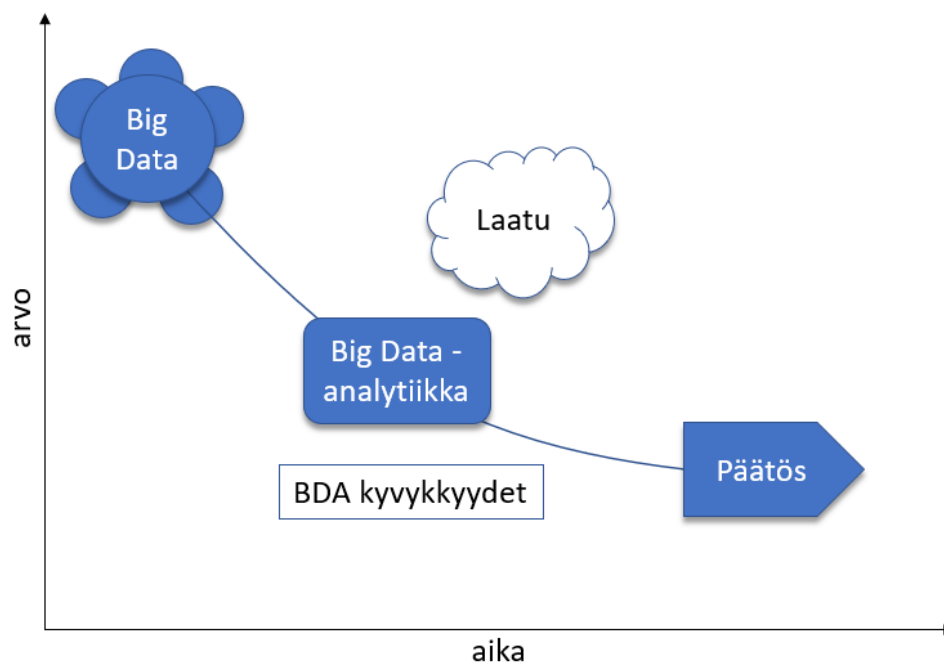
Powerin (2015) mukaan johdon ja informaatioteknologian ammattilaisten on erittäin tärkeää ymmärtää datavetoisia päätöksentekojärjestelmiä ja kuinka sellaiset järjestelmät voivat tuoda liiketoimintatietoa ja parantaa tehokkuuden mittaamista. Power (2015) toteaa suurimman osan ymmärtävän data-analytiikan johtavan tehokkaaseen päätöksentekoon. Hän korostaakin monien epäilevän Big Datan merkitystä suuren arvon saavuttamiselle. Provost ja Fawcett (2013) määrittävät dataperusteisen päätöksenteon olevan päätösten perustamista datan analyysille pelkän intuition sijaan. He kokevat dataperusteisen päätöksenteon hyötyjen olevan osoitettu jo pitävästi.

Tietoperusteisesta näkökulmasta päätöksentekoon käytettävän tiedon osalta on olemassa yksi keskeinen haaste. Tieto voi olla eksplisiittisen tiedon lisäksi hiljaista tietoa, joka on esimerkiksi ihmisissä olevaa kokemusten tuomaa tietoa. Hiljaista tietoa on erittäin vaikea muotoilla kirjoitettuun muotoon. Tämän vuoksi sitä on myös erittäin hankala analysoida sellaisenaan. Esimerkiksi tietoperusteisesta näkemyksestä Grant (1996) kuvaa ylemmän tason päätösten olevan usein riippuvaisia alemman tason tiedosta, joten hierarkia köyhdyttää ylemmän tason päätösten laatua. Hän toteaa myös, että mikäli tieto on yrityksen ensisijainen resurssi ja tieto sijoittuu yksilöihin, silloin työntekijät omistavat suuren osan yrityksen resursseista. Tämä puolestaan haastaa teoreettista perustaa osakkeenomistajien arvon kasvattamiselle.

Big Data -analytiikan hyödyntäminen voi kuitenkin olla myös ratkaisu tähän hiljaisen tiedon asettamaan ongelmaan. Myös Grantin (1996) ajatus perustui siihen, että päätöksen laatu perustuu päätösten tekemiseen käytettävän tiedon relevanttuteen ja että kaikki oppiminen tapahtuu yksittäisissä ihmisissä. Big Data mahdollistaa kuitenkin entistä tarkemman eksplisiittisen tiedon luomisen, joten päätöksissä käytettävän tiedon painoarvoa olisi mahdollista siirtää enemmän eksplisiittiseen tietoon. Lisäksi nykyteknologia mahdollistaa koneoppimisen hyödyntämisen. Tämä tarkoittaa, että ohjelmiston on algoritmien avulla mahdollista oppia tutkimaan dataa niin hyvin, että se pystyy tekemään esimerkiksi päätöksiä kokonaan uuteen dataan perustuen (Tynan 2017).

## 2.4 Yhteenveto

Tämän tutkielman toisen kappaleen teoreettiset taustat on koottu kuvioon 7 yhdistäen aiheet keskenään. Yhteenveto kokoaa siis edellä esitetyt teoriat Big Datan, BDA:n ja päätöksenteon osalta ja niiden yhteydestä toisiinsa. Yhteenveto toimii tutkielman empiirisen osion tutkimusmallin perustana.



**Kuvio 7.** Yhteenveto

Kuvion 7 kuvaama yhteenvetomalli alkaa Big Datasta, joka toimii uudenaikaisena raaka-aineena arvonluonnille. Big Data on luonteeltaan monimutkaista ja sitä on erittäin suuria määriä. Lisäksi Big Dataa on mahdollista kerätä esimerkiksi reaaliaikaisesti eli dataa syntyy jatkuvasti lisää. Nämä ominaisuudet mahdollistavat kuitenkin sen, että Big Data sisältää tarkkaa ja piilevää tietoa keräyksen kohteesta ja datan sisäisistä yhteyksistä.

Kerättyä Big Dataa analysoidaan BDA-menetelmin, jotta datassa piilevä arvo saadaan esiin. Analyysin avulla datasta saadaan informaatiota ja tietoa päätöksenteon tueksi. Kuvaileva analytiikka mahdollistaa menneen ajan tutkimisen tulevaisuuden suunnan löytämiseksi. Ennustava analytiikka puolestaan kertoo menneeseen aikaan perustuen siitä, millainen tulevaisuus todennäköisesti on. Ohjaileva analytiikka menee asiassa vielä askeleen edelle. Se esittää ohjeita toiminnalle ennustetussa tulevaisuudessa, jotta haluttu päämäärä saavutetaan. Nämä eri analyysitavat tuovat Big Datasta esiin päätöksenteolle tärkeää tietoa.

Seuraavaa vaihetta eli päätöksentekoa määrittää päätöksentekosysteemi ja päätöksentekoon liittyvät tekijät. Nämä määrittävät sitä, miten ja miksi saatua tietoa hyödynnetään. Päätöksentekoprosessin pohjalta syntyy rajoitetun rationaalisuuden määrittämä päätös. Analyysin tuottama tieto voi nopeuttaa päätöksentekoa päätösten tietoperusteisuuden kautta. Päätökseen vaikuttavien tekijöiden painoarvoa voidaan siirtää tiedolle, sen ollessa entistä tarkempaa. Periaatteessa ohjaileva analytiikka tarjoaa mahdollisuuden tiettyjen päätösten automatisoinnille.

Kuvion 7 kuvaaman prosessin arvo määrittyy kahden tärkeän tekijän yhdistelmänä. Laatu ja aika määrittävät erittäin vahvasti sitä, kuinka arvokasta prosessin mukainen päätöksenteko on liiketoiminnalle. Laatu määrittää kokonaisvaltaisesti jokaista mallin osa-aluetta. Syötetyn datan ja käytetyn menetelmän laatu ja oikeellisuus vaikuttavat erittäin paljon myös päätöksen laatuun. Organisaation, järjestelmän ja yksilön tasolla laatu vaikuttaa kompetenssiin ja organisatorisiin kyvykkyyksiin hyödyntää järjestelmiä ja niiden luomaa tietoa.

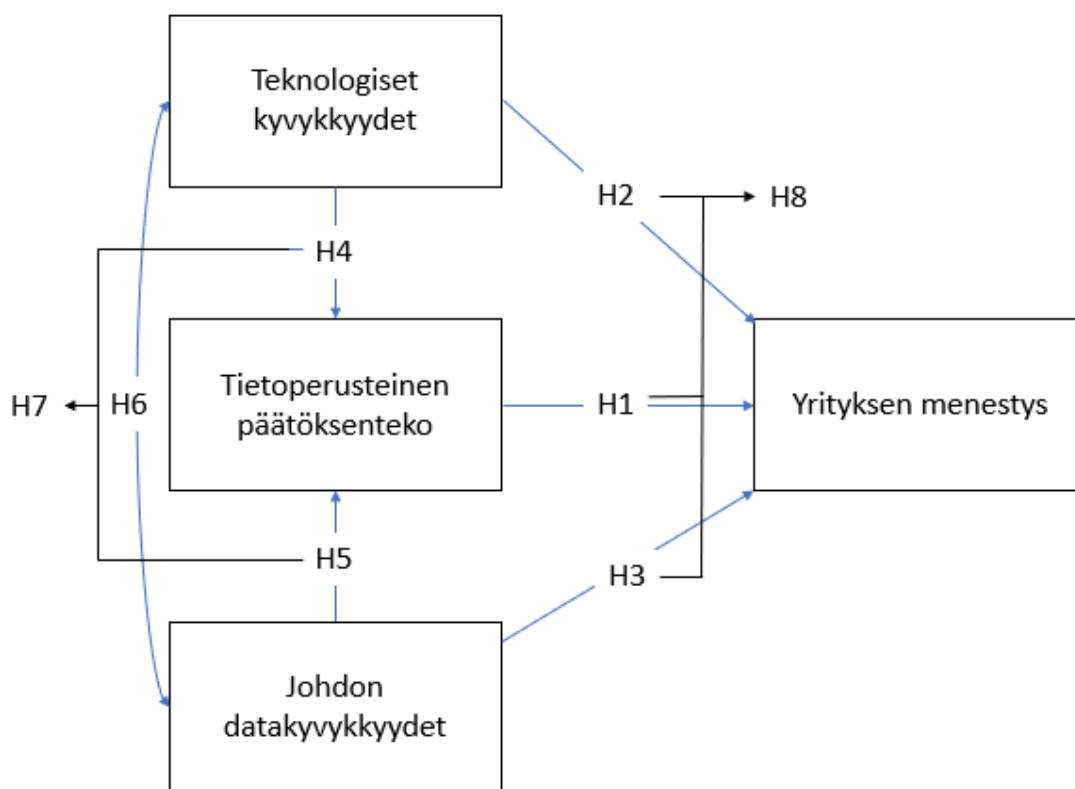


Aikaa puolestaan määrittää nopeus hyödyntää Big Datan sisältämää informaatiota. Big Data on arvokkainta, kun sitä pystytään hyödyntämään nopeasti. Ajasta on mahdollista säästää prosessin jokaisessa kohdassa. Mitä nopeammin Data saadaan kerättyä ja analysoitua, sitä nopeammin voidaan tehdä tietoon perustuvia päätöksiä. Myös nopeuden osalta on huomioon otettava myös laadun säilyttäminen.

BDA-kyvykkyydet toimivat puolestaan koko prosessin mahdollistajana, jotta prosessi pystyy tuottamaan kilpailuetua luovaa tehokasta päätöksentekoa. Esimerkiksi johdon kyvykkyyksien avulla varmistetaan Big Datan oikeellinen ja tarkoituksenmukainen kerääminen, analysointi ja hyödyntäminen päätöksenteossa. Teknologiset kyvykkyydet puolestaan varmistavat datan tehokkaan analysoinnin, joka tuottaa halutun laista informaatiota ja tietoa. Yhdessä nämä kaikki kuviossa 7 kuvatut kohdat, luovat liiketoiminnallista arvoa ja kilpailuetua menestyksen takaamiseksi. Tämän mallin mukaisesti data saa siis päätöksenteon kautta arvon. Tämän data-arvon yhdistäminen muuhun liiketoiminnalliseen arvoon johtaa lopulta yrityksen menestymiseen.

### 3 TUTKIMUSMALLI JA HYPOTEESIT

Tämän kappaleen tarkoituksena on esitellä viitekehyksestä johdettu tutkimusmalli ja sen pohjalta asetetut hypoteesit. Tutkimusmalli on kuvattu kuviossa 8. Malli on johdettu teoreettisista lähtökohdista niin, että tietoperusteinen päätöksenteko ja siihen liittyvät teknologiset ja johdon datakyvykkyudet toimivat arvoa luovina ja arvonluontia mahdollistavina tekijöinä. Tutkimuksen tarkoituksena on tarkastella miten nämä tekijät vaikuttavat yrityksen menestykseen. Samalla tarkoituksena on myös tarkastella tekijöiden välisiä suhteita. Niiden avulla pyritään saamaan tietoa kyvykkyysien merkityksestä tietoperusteiseen päätöksentekoon.



**Kuvio 8:** Tutkimusmalli

Tämän tutkielman teoreettisiin lähtökohtiin perustuen arvonluonti dataperusteisesti tapahtuu laadukkaan ja nopean päätöksenteon seurauksena niin, että data ja tieto toimivat päätösten perustana. Yrityksen kyky luoda arvoa johtaa puolestaan menestymiseen. Jotta tietoperusteinen päätöksenteko toteutuu, vaaditaan yritykseltä sekä teknologisia että strategisia kyvykkyyksiä hyödyntää dataa. Kun kyvykkyyksien luomat lähtökohdat ovat kunnossa ja päätöksiä tehdään tietoperusteisesti, yritysten on mahdollista menestyä paremmin. Tämän vuoksi tälle tutkimukselle on asetettu seuraavat hypoteesit.

### **Hypoteesit:**

**H1:** *Tietoperusteisen päätöksenteon hyödyntäminen johtaa yrityksen parempaan menestymiseen.*

**H2:** *Teknologisten kyvykkyyksien parempi taso johtaa parempaan menestymiseen.*

**H3:** *Johdon datakyvykkyyksien parempi taso johtaa parempaan menestymiseen.*

Tutkimuksen kolme ensimmäistä hypoteesia (H1, H2 ja H3) pyrkivät todentamaan kyvykkyyksien ja tietoperusteisen päätöksenteon merkitystä yrityksen menestykseen. Tutkimus tarkastelee tutkimusmallin mukaisesti näiden muuttujien vaikutuksia erikseen, mutta kuitenkin niin, että H1:n osalta myös kyvykkyyksien vaikutus on otettu huomioon. Hypoteesit testaavat tietoperusteisen päätöksenteon ja kyvykkyyksien vaikutuksia menestykseen tutkittavassa aineistossa.

**H4:** *Teknologiset kyvykkyydet vaikuttavat positiivisesti tietoperusteiseen päätöksentekoon.*

**H5:** *Johdon datakyvykkyydet vaikuttavat positiivisesti tietoperusteisen päätöksentekoon.*

**H6:** *Teknologiset kyvykkyydet ja Johdon datakyvykkyydet tukevat toinen toisiaan.*

Kolme seuraavaa hypoteesia (H4, H5 ja H6) muodostavat tutkimuksen toisen kokonaisuuden. Tietoperusteinen päätöksenteko ja kyvykkyydet ovat teoreettisesti erittäin tiukassa yhteydessä toisiinsa. Näiden hypoteesien tarkoituksena on testata kyseisten muuttujien keskinäisiä vaikutuksia. H4 ja H5 testaavat kyvykkyyksien merkitystä tietoperusteisen päätöksenteon olemassaoloon. H6 tarkoituksena on puolestaan tarkastella kyvykkyyksien keskinäistä yhteyttä.

**H7:** *Monimuotoisen ja runsaan datan hyödyntäminen vahvistaa a) teknologisten kyvykkyyksien ja b) johdon datakyvykkyyksien merkitystä tietoperusteiseen päätöksentekoon sekä c) vahvistaa kyvykkyyksien välistä yhteyttä.*

**H8:** *Monimuotoisen ja runsaan datan hyödyntäminen vahvistaa a) tietoperusteisen päätöksenteon ja b) kyvykkyyksien vaikutusta menestykseen.*

Hypoteesit H7 ja H8 pyrkivät luomaan tutkimukseen yhteyden Big Dataan. H7:n osalta tarkasteltavana ovat uudelleen H4, H5 ja H6. Hypoteeseja tarkastellaan kuitenkin erikseen ryhmissä, joissa runsaan ja monimuotoisen datan hyödyntäminen on vahvaa ja puolestaan heikkoa. Vastaavasti H8 lähestyy hypoteeseja H1, H2 ja H3 samoissa ryhmissä. H7:n ja H8:n tarkoituksena on siis tutkia runsaan ja monimuotoisen datan hyödyntämisen vaikutusta tutkimusmalliin.

Teorian pohjalta luodun tutkimusmallin mukaisesti tutkimuksessa tutkittiin neljää muuttujaa. Tutkimuksessa muuttujista kolme viimeistä pyrkivät selittämään ensimmäistä. Muuttujista ensimmäinen on Yrityksen menestys (FIRMPER), joka on kuvattu taulukossa 2. Tätä muuttujaa kuvattiin neljän tekijän avulla. Nämä olivat yrityksen taloudellinen tilanne, kasvu, työntekijöiden tuottavuus ja työntekijöiden pysyvyys. Muuttujan tarkoituksena on saada käsitys yrityksen menestyksen tilasta.

Muuttuja	ID	Tekijä
Yrityksen menestys (FIRMPER)	FIRMPER1	Taloudellinen tilanne
	FIRMPER2	Yrityksen kasvu
	FIRMPER3	Työntekijöiden tuottavuus
	FIRMPER4	Työntekijöiden pysyvyys

**Taulukko 2.** Yrityksen menestys

Toisena muuttujista on Tietoperusteinen päätöksenteko (ANACULT). Tämä muuttuja mittaa yrityksen kykyä hyödyntää dataa päätöksenteon perustana. Muuttujaa kuvattiin viiden tekijän avulla, jotka ovat kuvattu taulukossa 3.

Muuttuja	ID	Tekijä
Tietoperusteinen päätöksenteko (ANACULT)	ANACULT1	Yrityksessämme dataa pidetään arvokkaana voimavarana.
	ANACULT2	Teemme päätökset datan, emme oletusten, perusteella.
	ANACULT3	Jos data on ristiriidassa intuitiomme kanssa, uskomme dataan.
	ANACULT4	Tarkastelemme ja kehitämme toimintatapojamme perustuen datasta jalostettuun tietoon.
	ANACULT5	Valmennamme työntekijöitämme tekemään päätöksiä datan perusteella.

**Taulukko 3.** Tietoperusteinen päätöksenteko

Kolmas muuttuja on Johdon datakyvykkyudet (STRATKM). Tämä muuttuja kuvaa yrityksen johdon strategisia valmiuksia tietoperusteiseen päätöksentekoon. Muuttujaa kuvattiin viiden tekijän avulla, jotka ovat kuvattu taulukossa 4.

Muuttuja	ID	Tekijä
<b>Johdon datakyvykkyudet (STRATKM)</b>	STRATKM1	Strategiaa suunnitellaan ja päivitetään tiedon ja osaamisen pohjalta.
	STRATKM2	Strategiassa otetaan huomioon yrityksemme tiedon ja osaamisen kehittäminen.
	STRATKM3	Strategista osaamista verrataan systemaattisesti kilpailijoihin.
	STRATKM4	Tiedon ja osaamisen johtamisstrategiaamme viestitään henkilöstölle selkeästi ja kattavasti.
	STRATKM5	Yrityksessämme on määritelty selkeästi kuka vastaa tiedon strategiseen johtamiseen liittyvistä asioista.

**Taulukko 4.** Johdon datakyvykkyudet

Viimeisenä muuttujana ovat yrityksen Teknologiset kyvykkyudet (ANATECH). Muuttuja kuvaa nimensä mukaisesti teknologisia kyvykkyksiä datan käsittelyyn tietoperusteisen päätöksenteon tueksi. Muuttuja kuvattiin viiden tekijän kautta. Nämä ovat kuvattuna taulukossa 5.

Muuttuja	ID	Tekijä
<b>Teknologiset kyvykkyudet (ANATECH)</b>	ANATECH1	Yrityksemme hyödyntää modernia teknologiaa datan käsittelyyn.
	ANATECH2	Yrityksemme hyödyntää erilaisia datan visualisointityökaluja.
	ANATECH3	Yrityksemme hyödyntää pilvipalveluita datan käsittelyyn ja analysointiin.
	ANATECH4	Yrityksemme hyödyntää avoimen lähdekoodin ohjelmistoja datan analysointiin.
	ANATECH5	Yrityksemme hyödyntää modernia teknologiaa datan tallentamiseen ja säilömiseen.

**Taulukko 5.** Teknologiset kyvykkyudet

Näiden muuttujien lisäksi tutkimuksessa tarkasteltiin Big Datan hyödyntämistä ANADAT1 tekijällä, joka kuvasi runsaan ja monimuotoisen datan olemassaoloa (Tekijä: Saatavillamme on runsaasti monimuotoista dataa). Tätä ANADAT1 tekijää hyödynnetään datan analysoinnissa luokittelevana tekijänä etsittäessä vastauksia tutkimuskysymyksiin.

## 4 TUTKIMUSMENETELMÄT

Tämän kappaleen tarkoituksena on kuvata ensin tutkielman empiirisessä osiossa käytetyt tutkimusmenetelmät. Kappaleessa kuvataan lisäksi tutkimuksen aineisto ja sen keruu sekä aineiston varsinainen analysointi. Analyysi on toteutettu käyttäen rakenneyhtälömallinnusta. Rakenneyhtälöiden avulla on mahdollista tutkia monimutkaisia, useiden eri tekijöiden välisiä yhteyksiä, kuten tässä työssä on tarkoitus. Analyysi toteutettiin käyttäen IBM:n SPSS Statistics ja SPSS Amos -ohjelmistoja.

### 4.1 Menetelmät

Tämän tutkielman empiirinen tutkimus analysoitiin rakenneyhtälömallinnusta (SEM, Structural Equation Models) hyödyntäen. Rakenneyhtälömallinnuksen avulla on mahdollista tutkia, miten hyvin teoriaan pohjautuvat hypoteesit tai mallit sopivat tutkittavaan aineistoon (Schreiber, Nora et al. 2006). Rakenteellisesti SEM koostuu kahdesta osa-alueesta: 1) mittamallista, joka määrittää mitattujen muuttujien suhdetta latentteihin muuttujiin sekä 2) rakennemallista, jonka avulla esitetään latenttien muuttujien välisiä suhteita, ottaen mittaukseen liittyvät virheet huomioon (Jöreskog, Sörbom 2001).

Tutkimus toteutettiin Hairin (2010) kirjassaan kuvaamalla SEM prosessilla. Tutkimuksen ensimmäinen vaihe on datan seulonta (Data screening). Tämän vaiheen tarkoituksena on tarkastella kerättyä data-aineistoa ja poistaa siitä analyysiä vääristäviä datarivejä, kuten rivit, joissa ei ole tarpeeksi vastauksia tai vastaaja on vastannut vain yhdellä numerolla kaikkiin kohtiin. Lisäksi datan seulontavaiheessa voidaan tehdä havaintoja muuttujien kyvystä selittää mitattavaa asiaa. Tässä hyödynnetään esimerkiksi muuttujasta puuttuvien vastausten määrää sekä muuttujan hui-pukkuutta ja vinoutta. Oletuksena on, että selitettävien muuttujien tulisi olla liki normaalijakautuneita.

Datan seulontavaiheen jälkeen muuttujien välisiä suhteita mitataan tutkivan faktori-analyysin (EFA, Exploratory Factor Analysis) avulla. EFA:n on tilastollinen lähes-tymistapa muuttujien välisten korrelaatioiden havaitsemiseksi. Tämän vaiheen tar-koituksena on antaa analyysin vapaasti muodostaa aineiston muuttujista faktorit, joilla on vahva korrelaation keskenään. Muodostuneita faktoreita voidaan muuttaa vastaamaan paremmin haluttua teoreettista mallia poistamalla analyysistä huonosti korreloivia muuttujia. EFA siis osoittaa faktoreiden olemassaolon lisäksi heikot muuttujat, jotta ne eivät vääristä lopullista analyysiä. Tämä vaihe toteutettiin tutki-muksessa SPSS Statistics ohjelmistolla.

Prosessin seuraava vaihe on konfirmatorinen faktorianalyysi (CFA, Confirmatory Factor Analysis), joka on mittamallin viimeinen vaihe. Tämän vaiheen tarkoitukse-na on testata kerätyn aineiston sopivuutta teoriaan. Vaiheessa tarkastellaan mallin sopivuutta, reliabiliteettia ja validiteettia teoreettisesti muodostettuun malliin. CFA:ssa muuttujat pakotetaan koskemaan tiettyä muuttujaa, joten muuttujien va-linnassa käytetään apuna EFA:n tuloksia. Käytännössä tarkoituksena on varmistaa EFA:n tuottaman faktorirakenteen olemassaolo tutkittavassa aineistossa.

Prosessin viimeinen vaihe on varsinainen rakenneyhtälömallinnus (SEM, Structural Equation Modeling), jossa teoreettisesti asetettuja hypoteeseja testataan mallin latenttien muuttujien välisten suhteiden avulla. Tässä vaiheessa malli muutetaan teoreettista mallia vastaavaksi rakennemalliksi. Rakennemallin avulla on mahdollis-ta tutkia esimerkiksi välillisiä ja suoria vaikutuksia latenttien muuttujien välillä. CFA ja rakenneyhtälömallinnus toteutettiin tässä tutkimuksessa SPSS Amos ohjelmis-tolla.



## 4.2 Aineiston kuvaus

Tutkimuksen hypoteeseja testattiin Lappeenrannan teknillisen yliopiston ”Aineeton pääoma ja arvonluonti” -projektin vuonna 2017 kyselylomakkeilla kerätyllä aineistolla. Aineisto kerättiin strukturoidulla kyselylomakkeella hyödyntäen avaininformanttitekniikkaa. Tutkittu populaatio koostui toimialarajat ylittävästi palveluita ja tuotteita valmistavista yrityksistä, joissa oli vähintään 100 työntekijää. Yritysten valinnassa hyödynnettiin Intellian tietokantaa.

Kokonaisuudessaan kysely toteutettiin kahdessa osassa. Projektissa potentiaalisia yrityksiä alkuperäisen aineiston keräämiseksi löydettiin yhteensä 1711. Näihin yrityksiin oltiin ensin yhteydessä puhelimitse sopivien vastaajien löytämiseksi. 1711 potentiaalisesta yrityksestä vastauksia lopulliseen aineistoon saatiin 221. Vastausprosentti oli näin ollen 12,9 %.

Projektin aineistossa tarkasteltiin aineetonta pääomaa ja arvonluontia monesta eri näkökulmasta, mutta tämän tutkielman tutkimuksen toteuttamiseksi aineistosta käytettiin vain sille tarpeellisia osa-alueita. Käytetyt osa-alueet käsittelivät yrityksen tuloksellisuutta, teknologisia kyvykkyyksiä, strategisen tason tietojohdamista ja tietoperusteisen päätöksenteon kulttuuria. Kaikkia käytettyjä mittareita tutkittiin asteikolla 1 - 7. Yrityksen tuloksellisuutta mittaavien kysymysten osalta 1 viittasi ”selvästi alle keskitasoon” ja 7 ”selvästi yli keskitason”. Muiden osa-alueiden osalta 1 viittasi vastaukseen ”täysin eri mieltä” ja 7 vastaukseen ”täysin samaa mieltä”. Tämä helpotti myös tulosten keskinäistä vertailemista. Tässä tutkielmassa käytetyt osa-alueet ovat kuvattuna tarkemmin kappaleessa 3.

### 4.3 Datan seulonta

Tutkimuksessa käytetyn tutkimusmenetelmän mukaisesti neljävaiheinen SEM-prosessi alkoi datanseulonnalla. Datan seulonnan tarkoituksena on muokata data-aineistoa niin, että siitä poistetaan analyysiä mahdollisesti vääristävät vastaukset. Samalla aineistosta pyritään havaitsemaan muita analyysiin mahdollisesti vaikuttavia tekijöitä, kuten tutkittavien muuttujien jakaumien huipukkuutta ja kaltevuutta. Datan seulonnassa käsitellään myös samalla puuttuvat datatiedot.

Datan seulonta alkoi puuttuvan datan käsittelyllä. Rakenneyhtälömallinnus toimii paremmin, kun data on kokonaista, mahdollistaen esimerkiksi muutosindeksien hyödyntämisen. Tämän vuoksi 221 vastaajan aineistosta poistettiin ensin kaikki vastaajat, jotka olivat jättäneet yli 10 % vastauksista tyhjiksi. Analyysistä pudotetuista vastaajista suuri osa oli jättänyt vastaamatta kokonaisuun aihealueisiin, joita muuttujat edustivat. Analyysin luotettavuuden parantamiseksi aineistosta poistettiin 42 vastaajaa, jolloin tutkittava otoskoko oli 179 vastaajaa. Aineistoon ei jäänyt jäljelle yhtään vastaajaa, joka olisi vastannut kaikkiin kohtiin samalla numerolla vääristäen mahdollisesti tulosta, joten enempää vastaajia ei vähennetty. Myöskään poikkeavien havaintojen perusteella ei aineistosta tehty poistoja. Koska kyselyn kaikkien kysymysten vastausten asteikko oli 1-7, ei tässä vaiheessa ole perusteltua pitää havaintoja muista poikkeavina.

Lopulliseen aineistoon jäi vielä puuttuvia datapisteitä, jotka ovat kuvattuna liitteessä 1 kohdassa "missing". Erittäin huomattavana aineistosta nousi esiin teknologisten kyvykkyyksien muuttuja ANATECH4, jossa puuttuvia vastauksia oli edelleen 59. Muuttujan puutteellisen datan vuoksi, se poistettiin analyysistä tässä vaiheessa. Muiden muuttujien osalta puuttuvat datapisteet päädyttiin täyttämään muuttujan omalla keskiarvolla, jotta data-aineistoon ei jäänyt yhtään tyhjiä vastauksia. Tyhjät vastaukset päädyttiin täyttämään muuttujien keskiarvoilla, jotta muuttujat pysyisivät muutoksesta huolimatta mahdollisimman samanlaisina.

Datan seulonnan osalta muuttujista on vielä tutkittava niiden jakaumia. Rakenneyhtälömallinnuksen toimimisen kannalta muuttujien tulisi olla lähellä normaalijakaumaa. Muuttujien jakaumia tutkittiin huipukkuuden ja vinouden osalta. Muuttujien vinouden (skewnes) ja huipukkuuden (kurtosis) arvot on listattu Liitteeseen 1. Vinouden sekä huipukkuuden osalta ainoaksi ongelmamuuttujaksi nousi johdon datavykykyksiä kuvaava STRATKM2. Se sai huipukkuusarvoksi 2,186, joka viittaa melko terävähuippuiseen jakaumaan. Lisäksi sen vinouden arvo oli -1,137, joka viittaa lievästi vasemmalle kaltevaan jakaumaan. Toisena huomioitavana muuttujana oli yrityksen menestystä kuvaava FIRMPER4, jonka huipukkuusarvo oli 1,436. Jakauma on siis normaalijakaumaa hieman terävähuippuisempi.

Jakaumien perusteella analyysistä ei kuitenkaan vielä pudotettu muuttujia. STRATKM2 ja FIRMPER4 osalta niiden mahdollinen vaikutus analyysin onnistumiseen otetaan huomioon SEM prosessin seuraavassa vaiheessa, joka on tutkiva faktorianalyysi. Mikäli mallin sopivuudessa on ongelmia, nämä muuttujat ovat mahdollisesti analyysistä poistettavissa. Datan seulonnan pohjalta tutkivaa faktorianalyysiä varten aineistona toimii kokonainen 179 vastaajan aineisto. Aineistosta on lisäksi pudotettu ANATECH4 muuttuja, perustuen suureen puutteellisen datan määrään.

#### 4.4 Tutkiva faktorianalyysi, EFA

Aineiston varsinainen analyysi alkaa tutkivalla faktorianalyysillä, jonka tarkoituksena on antaa faktorien syntyä vapaasti aineiston muuttujien korrelaatioiden perusteella. Tarkoituksena on siis selvittää latautuvatko muuttujat tutkimusmallin mukaisille faktoreille, jotka ovat mallin latentit muuttujat. Tutkivan faktorianalyysin osalta keskeistä on tarkastella latausten lisäksi faktoreiden sopivuutta, validiteettia ja reliabiliteettia.

EFA:n ensimmäinen vaihe oli muuttujien ja faktoreiden iterointi niin, että saavutettiin mahdollisimman sopiva faktorimalli. Tähän hyödynnettiin SPSS Statistics ohjelmiston faktorianalyysiä, joka toteutettiin käyttäen suurimman uskottavuuden estimointimenetelmää. Suurimman uskottavuuden estimointimenetelmä oli soveltuvin EFA-vaiheeseen, koska myös konfirmatorinen faktorianalyysi SPSS Amos ohjelmistossa hyödyntää tätä menetelmää. Lisäksi otoksen koko on riittävän suuri estimoinnin toteuttamiseksi ja estimointi osoittaa mallin sopivuutta Khiin neliötestillä.

Kaikki analyysin muuttujat mukaan otettuna, faktorianalyysi osoitti selvästi odotettujen faktoreiden olemassaolon. Ainoa muuttuja, joka ei latautunut mihinkään faktoriin kunnolla, oli FIRMPER4. Kyseisessä muuttujassa oli havaittu ongelmia jo hui-pukkuuden osalta. Kyseisen muuttujan lisäksi lopullisesta faktorianalyysistä poistettiin tietoperusteisen päätöksenteon osalta ANACULT1 ja ANACULT5, jotka latautuivat huonosti faktoriin.

FIRMPER3 latautui faktorianalyysissä myös hieman huonosti yrityksen menestysfaktoriin. Muuttuja säilytettiin analyysissä, koska rakenneyhtälömallinnuksen konfirmatorinen faktorianalyysi SPSS Amos -ohjelmistolla vaatii vähintään kolmen muuttujan selittävän latenttia muuttujaa, jotta analyysi on mahdollista toteuttaa. FIRMPER3:n mukaan ottaminen analyysiin on siis otettava huomioon arvioitaessa tutkimuksen validiteettia. Lopulliset faktoreiden lataukset ovat nähtävissä liitteestä 2.

EFA:n avulla saadun faktorimallin sopivuutta estimoitiin Khiin neliöarvon avulla, joka on hieman otoskoriippuvainen. Estimointi ilmoittaa Khiin neliöarvon lisäksi vapausasteen sekä testin p-arvon, jonka tulee olla alle 0,05, jotta tulos on merkittävä. Estimoinnissa Khiin neliöarvoksi saatiin 110,525, vapausaste oli 51 ja p-arvo 0,000. Khiin neliöarvon perusteella saatu faktoriryhmä oli sopiva. Sopivuutta testattiin kuitenkin vielä soveltuvuustestien, validiteetin ja reliabiliteetin kautta.

### Soveltuvuustestit

KMO ja Bartlettin Testi		
Kaiser-Meyer-Olkin.		0,814
Bartlettin Testi	Chi-Square	1170,937
	df	105
	p-arvo	0,000

**Taulukko 6.** KMO ja Bartlettin testi

Aineiston soveltuvuutta tutkivaan faktorianalyysiin testattiin Kaiser-Meyer-Olkin (KMO) otoksen soveltuvuustestillä ja Bartlettin testillä. Hairin (2010) mukaan yli 0,8 KMO arvot osoittavat tarkasteltujen muuttujien ennustavan hyvin faktoreita. Bartlettin testi puolestaan osoittaa, että olemassa on pätevät korrelaatiot muuttujien välillä, mikäli p-arvo on alle 0,05. Taulukossa 6 on esiteltynä testien tulokset. Näiden testien perusteella EFA:n KMO arvo 0,814 ja Bartlettin testin p-arvo 0,000 osoittavat aineiston soveltuvan hyvin ennustamaan löydettyjä faktoreita.

Edellä kuvattujen soveltuvuustestien lisäksi saadun faktorimallin toimivuutta arvioitiin kommunaliteettien avulla. Kommunaliteettiarvot on listattu liitteeseen 3. Hairin (2010) mukaan kommunaliteetti osoittaa sen, kuinka suuri osuus yksittäisen muuttujan vaihtelusta selittyy havaittujen faktoreiden avulla. Kommunaliteetin osalta ainoa heikko muuttuja oli FIRMPER3, jonka heikkous havaittiin jo faktoreiden lataus-

vaiheessa. Muuttujaa ei kuitenkaan pudotettu analyysistä, jotta konfirmatorinen faktorianalyysi oli mahdollista toteuttaa. Kaikki muut muuttujat saavuttivat hyväksyttävän kommunaliteetin tason.

### Reliabiliteetti

Latentti muuttuja	Cronbachin alfa ( $\alpha$ )	Muuttujien määrä
FIRMPER	0,686	3
ANACULT	0,834	3
STRATKM	0,827	5
ANATECH	0,835	4

Taulukko 7. Reliabiliteetit

EFA:n tuottamien faktoreiden reliabiliteettia tarkasteltiin Cronbachin alfa ( $\alpha$ ) arvojen avulla, jotka ovat näkyvissä taulukossa 7. Cronbachin alfa-testin alimpana hyväksyttävänä arvona pidetään 0,6 ja hyvänä reliabiliteettina pidetään yli 0,7 arvoja. Latenteista muuttujista FIRMPER ylitti alimman sallitun reliabiliteetin rajan 0,6. Se ei kuitenkaan ylittänyt hyvän reliabiliteetin rajaa 0,7, kuten muut latentit muuttujat.

### Validiteetti

Hair (2010) esittää, että yhtäpitävä validiteetti osoittaa yksittäisen faktorin muuttujien olevan korkeasti korreloivia. Faktorin latauksen tulisi olla yli 0,35, jotta tavoitettu merkitsevyys saavutettaisiin. Faktoreiden lataukset ovat nähtävissä liitteestä 2. Faktorien muuttujien latauksista vain haasteita tuonut FIRMPER3 ei aivan yltänyt tähän tasoon. Reliabiliteettitestauksen myötä FIRMPER faktori on kuitenkin hyväksyttävällä tasolla.

Diskriminantin validiteetin Hair (2010) kuvaa puolestaan osoittavan faktoreiden erillaisuutta ja keskinäistä korreloimattomuutta. Analyysistä ei pudotettu yhtään muuttujaa, joka olisi latautunut ristiin useampaan faktoriin. Muuttujat eivät siis latautuneet voimakkaasti useampiin faktoreihin. Faktoreiden keskinäiset korrelaatiot ovat nähtävissä liitteestä 2. Faktoreista tietoperusteinen päätöksenteko ja teknologiset

kyvykkyydet korreloivat voimakkaimmin keskenään. Ne eivät kuitenkaan ylittäneet liian suurta korreloinnin arvoa 0,7. Ristilatausten ja korrelaatioiden osalta validiteetti on siis kunnossa. Faktorit latautuivat erittäin odotetulla tavalla.

#### **4.5 Konfirmatorinen faktorianalyysi, CFA**

Seuraava vaihe rakenneyhtälömallinnukselle oli konfirmatorinen faktorianalyysi, jonka tavoitteena on testata aineiston sopivuutta faktoreihin pakottamalla muuttujat teoriaan perustuviin latentteihin muuttujiin. Tämän tutkimuksen osalta CFA-latentit muuttujat vastaavat EFA:n osoittamia faktoreita, jotka vastasivat teoreettisesti tavoiteltuja faktoreita. Tutkivasta faktorianalyysistä poiketen tämän tutkimuksen konfirmatorisessa faktorianalyysissä tarkastellaan koko aineiston lisäksi aineistoa niin, että aineisto on jaettu runsasta ja monimuotoista dataa paljon ja vähän hyödyntäviin. Mallia testataan siis kaksi kertaa, ensin yhtenäisellä aineistolla ja sen jälkeen jaetulla. Jaetun aineiston mallissa aineisto jakautui ryhmään, jossa runsaan ja monimuotoisen datan hyödyntäminen oli tasolla 6-7 (93 vastaajaa), sekä ryhmään, jossa runsaan ja monimuotoisen datan hyödyntäminen oli tasolla 1-5 (86 vastaajaa).

CFA vaihe alkoi faktorimallin sopivuuden testaamisella, joka toteutettiin tavallisella mallin hyvyystestillä (goodness-of-fit). Hair (2010) suosittelee mallin sopivuustestauksen tapahtuvan sekä absoluuttisilla että inkrementaalisilla sopivuusmittareilla. Absoluuttiset mittarit mittaavat mallin kokonaissopivuutta. Inkrementaaliset mittarit puolestaan osoittavat kuinka hyvin kyseinen malli sopii verrattuna tilastolliseen nolamalliin. Rakenneyhtälömallinnukselle tyypillisesti tässä tutkimuksessa absoluuttisina mittareina käytettiin Khiin neliötä ja sen p-arvoa, RMSEA, PCLOSE, SRMR ja GFI -mittareita. Inkrementaalisista mittareista tutkimuksessa käytettiin CFI ja TLI arvoja. Mallien sopivuustestauksen tarkoituksena on varmistaa, että mallit sopivat rakenneyhtälömallinnuksen jatkoanalyysiin.

Konfirmatorinen faktorianalyysi SPSS Amos -ohjelmistolla antaa mallin sopivuustietojen lisäksi muutosehdotuksia mallin sopivuuden parantamiseksi. Testaus osoitti, että STRATKM faktorissa STRATKM3 ja STRATKM5 muuttujat aiheuttivat keskenään epäjohdonmukaisuuden, joka heikensi mallin sopivuutta. Muutosten tekeminen malliin empiiristen havaintojen perusteella tulee olla harkittua, jotta teoreettiset lähtökohdat säilyvät mallissa (Hair 2010). Koska kyseinen faktori on latenteista muuttujista suurin, oli perustellusti mahdollista poistaa tässä vaiheessa STRATKM5 muuttuja analyysistä. STRATKM5 latautui näistä muuttujista heikommaksi faktoriin. Tämän muutoksen jälkeen CFA ei esittänyt enempää muutosehdotuksia mallin sopivuuden parantamiseksi.

CFA-mallin lataukset menivät EFA:n pohjalta odotetusti. Muuttujat latautuivat faktoreille hyvin lukuun ottamatta FIRMPER3 muuttujaa, jonka huono latautuminen oli havaittu jo EFA-vaiheessa. CFA-vaihetta ei kuitenkaan ollut mahdollista toteuttaa alle kolmen muuttujan faktoreilla SPSS Amos -ohjelmistolla. Tämä ongelmallinen muuttuja ei onneksi vaikuttanut mallin kokonaissopivuuteen.

### **Sopivuustestit**

Lopullisen mallin sopivuustestien luvut ovat näkyvissä taulukossa 8. Yhtenäinen malli sekä jaettumalli täyttävät vaaditut arvot testatuilla asteikoilla suurimmilta osiltaan. Seuraavaksi tarkastellaan hieman tarkemmin saatuja mallin sopivuustietoja. Khiin neliön osalta p-arvo yli 0,05 osoittaa teoreettisten ja estimoitujen kovarianssien olevan tilastollisesti samanlaisia eli hyvän sopivuuden omaavia (Hair 2010). P-arvo on kuitenkin erittäin herkkä otoskoolle, joka saattaa olla osasyynä sille, että mallien p-arvot eivät ylittäneet kyseistä rajaa. Schreiberin (2006) mukaan otoskoon on oltava riittävä, mutta isoilla otoskoilla testi hylkää helposti mallin. Lisäksi Hooper et al. (2008) toteavat muuttujien huipukkuuden vaikuttavan tähän testiin. Mallissa STRATKM2 oli datan seulonnan vaiheessa osoitettu huomattavan huipukkaaksi. Näiden syiden vuoksi mallin yhteensopivuutta tutkittiin myös muiden mittarien avulla.



Mittari	Yhtenäinen malli	Jaettumalli	Raja-arvo
Khiin neliö	115,924	187,832	-
DF	71	142	-
p-arvo	0,001	0,006	> 0,05
Khiin neliö / DF	1,633	1,323	< 3
CFI	0,956	0,952	> 0,90
TLI	0,944	0,939	> 0,90
RMSEA	0,060	0,043	< 0,07
PCLOSE	0,203	0,764	> 0,05
SRMR	0,048	0,064	< 0,08
GFI	0,917	0,893	< 0,90

**Taulukko 8.** Mallien sopivuusluvut

Khiin neliötestin rajoittuneisuuden vuoksi Hooper (2008) kuvaa useiden rakenneyhtälömalleja hyödyntävien tutkimusten käyttävän p-arvon sijaan Khiin neliön arvoa jaettuna vapausasteella (DF). Alle 3:n arvo kyseisessä mittarissa viittaa hyvään mallin yhteensopivuuteen. Taulukon 8 mukaisesti tämä testi osoittaa mallin olevan yhteensopiva.

Schreiber (2006) kuvaa CFI:n (comparative fit index) testaavan puolestaan mallin yleistä riittävyyttä vertaamalla estimoitua mallia nollamalliin. CFI:n raja-arvoksi on esitetty > 0,90 arvoa. Vastaavasti Schreiber (2006) esittää TLI (Tucker-Lewis index) arvioivan suhteotokseen vaikutusta, vaikka Khiin neliötesti hylkäisi mallin. Se ottaa samalla huomioon otokseen vaikutuksen ja vapausasteiden määrän. Tämän yhteensopivuusindeksin tulisi myös olla > 0,90. Molemmat tämän tutkimuksen mallit ylsivät näissä inkrementaalisissa mittareissa tavoiteltujen raja-arvojen yli, osoittaen hyvää sopivuutta.

Absoluuttisista malleista RMSEA (root mean square error of approximation) testaa mallin riittävyyttä vertaamalla teoreettista mallia täydelliseen malliin. Se on yksi suosituimmista sopivuustesteistä. Sen raja-arvosta on käyty paljon keskustelua, mutta hyväksyttävänä raja-arvona on pidetty Hooperin (2008) ja Steigerin (2007)

mukaan  $< 0,07$  arvoa. PCLOSE ( $p$  of Close Fit) mittaa puolestaan yksipuolisesti nollahypoteesia, jossa oletetaan RMSEA:n olevan  $0,05$  eli hyvin sopiva malli. Arvo osoittaa siis myös mallin sopivuutta ja sen raja-arvona on  $> 0,05$ . Molemmat tarkastellut mallit ovat myös näillä mittareilla mitattuna sopivuutensa puolesta hyvällä tasolla.

SRMR (standardized root mean square residual) kuvaa myös tutkittavan mallin riittävyttä tarkastelemalla keskimääräistä jäännöskorrelaatiota. SRMR arvo ilmoittaa standardoidun suhdeluvun. Siinä tunnusluku on suhteutettu otoskorrelaatioon. SRMR hyväksyttävyyden raja-arvona on  $< 0,08$ . (Schreiber, Nora et al. 2006) SRMR arvojen osalta mallien sopivuudet olivat hyvällä tasolla.

Mallien sopivuustestien viimeinen testi oli GFI (Goodness of Fit Index) testi, joka nimensä mukaisesti testaa mallin sopivuuden hyvyttä. Raja-arvoksi GFI:lle on esitetty  $> 0,9$  arvoa. Jaettumalli ei aivan yltänyt asteikolla hyvin sopivan mallin tasolle. GFI mittarin osalta on kuitenkin ollut puhetta sen hyödyttömyydestä, koska otoskoko vaikuttaa siihen rajusti. Tämän vuoksi esimerkiksi Sharma et al. (2005) eivät suosittele tämän mittarin käyttöä. Jaetussa mallissa aineisto on jaettu kahteen lähes yhtä suureen osaan, joka selvästi näkyi GFI-tuloksessa. Kokonaisuutena mallien sopivuustestit osoittivat hyvää sopivuutta. Malleihin ei siis ollut tarvetta tehdä enempää muutoksia ja analyysiä voitiin jatkaa.

### **Validiteetti ja reliabiliteetti**

Sopivuustestien puolesta hyväksytyjä malleja tarkasteltiin seuraavaksi validiteetin ja reliabiliteetin osalta. Muutoksena EFA-vaiheen malliin, malleja oli nyt kaksi ja niissä STRATKM5 muuttuja oli pudotettu analyysistä. Mallien faktoreiden reliabiliteettia arvioitiin yhdistetyllä reliabiliteetti arvolla (CR, Composite Reliability), jonka raja-arvo on Hun ja Bentlerin (1999) mukaan  $> 0,7$ . Mallien faktoreiden validiteettia tarkasteltiin puolestaan AVE (Average Variance Extracted) arvolla, jonka raja-arvo Hun ja Bentlerin (1999) mukaan  $> 0,5$ .

	CR	AVE	MSV	MaxR(H)	STRATKM	ANATECH	ANACULT	FIRMPER
STRATKM	0,832	0,557	0,24	0,86	<b>0,747</b>			
ANATECH	0,846	0,583	0,488	0,871	0,414***	<b>0,764</b>		
ANACULT	0,842	0,644	0,488	0,875	0,490***	0,699***	<b>0,802</b>	
FIRMPER	<b>0,658</b>	<b>0,41</b>	0,023	0,723	0,126	0,124	0,151	<b>0,641</b>

**Taulukko 9.** Validiteetti ja reliabiliteetti yhtenäiselle mallille.

	CR	AVE	MSV	MaxR(H)	STRATKM	ANATECH	ANACULT	FIRMPER
STRATKM	0,84	0,572	0,392	0,87	<b>0,756</b>			
ANATECH	0,826	0,549	0,498	0,854	0,590***	<b>0,741</b>		
ANACULT	0,849	0,653	0,498	0,874	0,626***	0,705***	<b>0,808</b>	
FIRMPER	<b>0,68</b>	<b>0,44</b>	0,039	0,785	0,196	0,063	0,18	<b>0,663</b>

**Taulukko 10.** Validiteetti ja reliabiliteetti jaetulle mallille.

Taulukko 9 ja taulukko 10 kuvaavat testin osoittamat validiteetti ja reliabiliteettiarvot testatuissa malleissa. Validiteettia arvioitaessa tarkasteltiin AVE arvoja ja faktoreiden välisiä korrelaatioita. Molemmissa malleissa FIRMPER faktorin validiteetti ei yllä hyväksyttävälle tasolle. Tämä johtuu jo aiemmin faktorissa esiintyneistä ongelmista. Tämä tulee ottaa huomioon arvioitaessa tuloksia suhteessa FIRMPER faktoriin. Muiden faktoreiden osalta validiteetti on riittävällä tasolla molemmissa malleissa. Validiteetin osalta MSV (Maximum Shared Variances) arvot ovat myös kummassakin mallissa ja kaikissa faktoreissa AVE arvoja pienempiä, joka Hairin (2010) mukaan edelleen osoittaa hyvää validiteettia.

Reliabiliteettiarvot ovat myös listattuna taulukkoihin 9 ja 10. Myös reliabiliteettiarvot (CR) ovat FIRMPER-faktorin osalta kummassakin mallissa hieman alle raja-arvon 0,7. Huomattavaa on, että STRATKM, ANACULT ja FIRMPER -faktoreiden validiteetti ja reliabiliteetti ovat parempia jaetussa mallissa.

## **Invarianssi**

Koska tutkimuksessa tarkastellaan kausaalisessa rakenneyhtälömallissa jaettua mallia, jossa runsaan ja monipuolisen datan hyödyntäminen toimii ryhmittelevänä muuttujana, CFA mallista on vielä tarkasteltava invariansseja. Tämän vaiheen tarkoituksena on varmistaa, että faktorirakenne ja lataukset ovat riittävän samanlaisia molemmissa ryhmissä. Muuten jaetun mallin ryhmät eivät mittaa samaa latenttien muuttujien rakennetta.

Invariansseja tarkasteltiin SPSS Amos ohjelmassa vertaamalla mallin sopivuutta rajoitettuna ja rajoittamattomana. Jaetun mallin sopivuustestien tulokset esiteltiin jo taulukossa 8. Mallin hyvät sopivuusluvut osoittavat konfiguroidun invarianssin olevan hyvällä tasolla. Tämän jälkeen mallia testattiin rajoittamalla reitit latentteihin muuttujiin yhtä suuriksi. Rajoitetun ja rajoittamattoman mallin regressiopainoarvoja vertailtaessa Khiin neliön eroavaisuustesti osoitti metrisen invarianssin olemassaolon (p-arvo 0,996). Rakenteellisen invarianssin osalta ANATECH-faktorista jouduttiin vapauttamaan ANATECH5-muuttuja, jotta invarianssi saavutettiin. Tämä saattaa vaikuttaa tulosten arviointiin, kun vertaillaan ryhmienvälisiä eroja ANATECH-faktorin osalta.

## **Yhtenäinen menetelmäpoikkeama**

Podsakoff et al. (2003) kuvaavat yhtenäisellä menetelmäpoikkeamalla (CMB, common method bias) viitattavan tilanteeseen, jossa dataan vaikuttaa virheelliset korrelaatiot. Heidän mukaan menetelmäpoikkeamat ovat yksi suurimmista syistä mittausvirheisiin, jotka puolestaan uhkaavat johtopäätösten validiteettia. Tässä tutkimuksessa poikkeamaa mitattiin Harmanin yhden faktorin testillä (Harman's single Factor test). Testi osoittaa sen, selittykö suurin osa varianssista yhdellä faktorilla.

Menetelmäpoikkeamaa testattiin kaikkien muuttujien faktorianalyysillä. Analyysissä käytettiin Varimax-rotatiota, jossa faktoreiden määrä pakotettiin yhteen. Datassa olisi menetelmäpoikkeama, mikäli yksi faktori selittäisi suurinta osaa mallin varianssista. Testin perusteella tutkimusaineistosta yksi faktori selitti 30 % varianssista. Näin ollen menetelmäpoikkeama ei todennäköisesti ole ongelma käsitellyssä aineistossa.

### **EFA ja CFA vaiheiden yhteenveto**

Tutkivan ja konfirmatorisen faktorianalyysin aikana analysoitavasta mallista pudotettiin viisi muuttujaa. Muuttujat olivat FIRMPER4, ANACULT1, ANACULT5, ANATECH4 ja STRATKM5. Kyseiset muuttujat aiheuttivat validiteetti-, mallin sopivuus- ja johdonmukaisuusongelmia malliin. Muuttujia poistettaessa huomioitiin teoreettisten lähtökohtien säilyminen mallissa.

Latenteista muuttujista, jotka muodostettiin faktoreista, yrityksen menestystä kuvaava FIRMPER sisältää validiteetti- ja reliabiliteetti-ongelmia. Kyseistä latenttia muuttujaa ei kuitenkaan ole aineiston perusteella mahdollista muokata enempää, jotta rakenneyhtälömalli on mahdollista toteuttaa. Tämä on siis otettava huomioon tehtäessä johtopäätöksiä FIRMPER osalta.

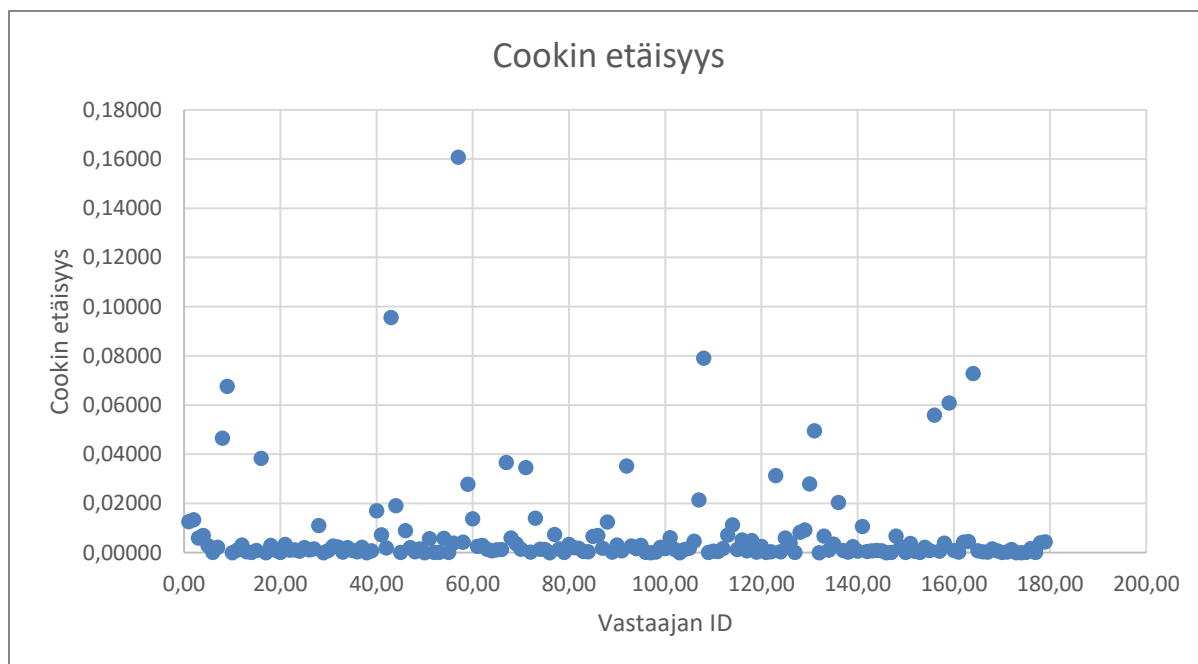
Tutkivaan faktorianalyysiin pohjautuva konfirmatorinen faktorianalyysin muodostama lopullinen mittamalli on nähtävissä liitteessä 4. Konfirmatorisen faktorianalyysin tarkoituksena oli varmistaa, että teoreettisesti asetettu tutkimusmalli sopii aineistoon. EFA ja CFA-vaiheet todistivat faktoreiden eli latenttien muuttujien olemassaolon aineistossa sekä mittamallin kokonaissopivuuden aineistoon. Seuraavana vaiheena rakenneyhtälömallinnukselle on rakennemallin luominen mittamallin pohjalta.

## 4.6 Rakenneyhtälömalli, SEM

Rakenneyhtälömallinnuksessa (SEM) viimeisenä vaiheena on mittamallin muuttaminen rakennemalliksi. Käytännössä tämä tarkoittaa mittamallin latenttien muuttujien välisten suhteiden muuttamista tutkimusmallia vastaavaksi kausaaliseksi malliksi. Kappaleessa 3 asetetut hypoteesit osoittavat siis millaisia yhteyksiä mallissa tutkitaan. Rakenneyhtälömalli rakennetaan SEM-vaiheessa samanlaisiksi yhtenäiselle ja jaetulle mallille. Tarkoituksena on siis testata hypoteeseja tutkimusmallin mukaisesti yhtenäisessä sekä jaetussa aineistossa.

### Monimuuttujien oletukset

Ennen rakenneyhtälön rakentamista tulee vielä tarkastella monimuuttujien oletusten osalta mahdollisia analyysiin vaikuttavia poikkeavia havaintoja sekä multikollinearisuutta. Analyysiin mahdollisesti vaikuttavien poikkeavien havaintojen olemassaoloa selittävissä muuttujissa tutkittiin SPSS Statistics ohjelmistossa Cookin etäisyysanalyysillä (Cook's distance), joka on kuvattuna kuviossa 9.



**Kuvio 9.** Cookin etäisyys latenteille muuttujille

Kuviosta 9 nähdään, että yksikään vastaaja ei ylitä Cookin etäisyyden osalta arvoa 1,0, joka viittaisi voimakkaaseen monimuuttuja vaikuttajaan. Analyysin osalta yhden vastaajan etäisyysarvo on 0,16077. Vaikka se poikkeaaakin eniten muista vastaajista, ei se kuitenkaan tule vaikuttamaan tuloksiin huomattavasti. Cookin etäisyysanalyysin perusteella vaikuttavia poikkeavia vastauksia ei siis ole.

Multikollineaarisuutta tarkasteltiin SPSS Statistics -ohjelmiston lineaarisen regressiion kollineaarisuustilastoilla. Multikollineaarisuustestin tulokset ovat nähtävissä liitteestä 5. Saatujen tulosten osalta VIF (Variable Inflation Factor) -arvot, eli muuttujan inflaatiofaktoriarvot ovat tarkastelun kohteena. Yhdenkään selittävän muuttujan kohdalla VIF-arvo ei ylittänyt erittäin ongelmallisen muuttujan rajaa 10. Kaikki VIF-arvot olivat alle arvon 3, jota esimerkiksi Hair (2010) pitää hyvänä raja-arvona. Multikollineaarisuudesta ei siis aiheudu analyysiin ongelmia.

### **Rakenneyhtälömalli**

Varsinainen rakenneyhtälömallin luominen toteutettiin SPSS Amos -ohjelmistossa, jossa mittamallin pohjalta rakennettiin uusi kausaalinen malli. Rakenneyhtälömalliin asetettiin teoreettisesti luodut yhteydet ja virhetermit. Rakenneyhtälömallin sopivuutta testattiin samojen mittarien valossa kuin analyysin CFA-vaiheessa. Sopivuustestien tulokset ovat nähtävissä taulukossa 11. Tavoitteena oli varmistaa, että mallien sopivuus ei ole muuttunut rakenneyhtälömallin rakennuksen myötä.

Koska tässä tutkimuksessa ei tarkastella minkään välillisen muuttujan vaikutusta malleihin, niihin ei lisätty uusia muuttujia. Tämän vuoksi mallien sopivuuteen ei myöskään tullut muutoksia. Sopivuustestien tulokset kuitenkin osoittavat, että rakenneyhtälömalli on mittamallin pohjalta toimiva. Testit eivät myöskään osoittaneet, että malleihin tulisi tehdä muutosindeksien pohjalta muutoksia sopivuuden parantamiseksi.

Mittari	Yhtenäinen malli	Jaettumalli	Raja-arvo
Khiin neliö	115,924	187,832	-
DF	71	142	-
p-arvo	0,001	0,006	> 0,05
Khiin neliö / DF	1,633	1,323	< 3
CFI	0,956	0,952	> 0,90
TLI	0,944	0,939	> 0,90
RMSEA	0,060	0,043	< 0,07
PCLOSE	0,203	0,764	> 0,05
SRMR	0,048	0,064	< 0,08
GFI	0,917	0,893	< 0,90

**Taulukko 11.** Sopivuustestit kausaalisille malleille

Seuraavana vaiheena oli hypoteesien testaus rakenneyhtälömallin avulla. Mallien avulla testattiin hypoteeseissa asetettuja muuttujien suoria yhteyksiä. Lisäksi jaetun mallin avulla testattiin runsaan ja monimuotoisen datan hyödyntämisen vaikutuksia suoriin yhteyksiin. Jaetun mallin osalta tulokset esitetään erikseen suuren ja pienen datan hyödyntämisen osalta.

Käytännössä rakenneyhtälömalleja tarkasteltiin lineaarisen regressioanalyysin avulla, jonka SPSS Amos -ohjelma toteuttaa rakenneyhtälömallille. Tarkastelu siis kohdistui hypoteesien perusteella vastemuuttujien riippuvuuteen selittävistä muuttujista. Analyysin tuottamat standardoidut  $\beta$ -estimaatit ovat esiteltynä taulukoissa 12 ja 13 ja 14.

Hypoteesi	Estimaatti
H1 ANACULT → FIRMPER	0,100
H2 ANATECH → FIRMPER	0,027
H3 STRATKM → FIRMPER	0,066
H4 ANATECH → ANACULT	<b>0,598***</b>
H5 STRATKM → ANACULT	<b>0,243**</b>
H6 ANATECH ↔ STRATKM	<b>0,414***</b>

Huomio: Lihavoidut luvut ovat merkitseviä tasoilla \*p < 0,1; \*\*p < 0,05; \*\*\*p < 0,01.

**Taulukko 12.** Tulokset, yhtenäinen malli



Hypoteesi		Estimaatti
H1	ANACULT → FIRMPER	0,087
H2	ANATECH → FIRMPER	-0,205
H3	STRATKM → FIRMPER	0,209
H4	ANATECH → ANACULT	<b>0,552***</b>
H5	STRATKM → ANACULT	<b>0,275**</b>
H6	ANATECH ↔ STRATKM	<b>0,563***</b>

Huomio: Lihavoidut luvut ovat merkitseviä tasoilla \*p < 0,1; \*\*p < 0,05; \*\*\*p < 0,01.

**Taulukko 13.** Tulokset, suuri runsaan ja monimuotoisen datan hyödyntämisen taso

Hypoteesi		Estimaatti
H1	ANACULT → FIRMPER	0,130
H2	ANATECH → FIRMPER	0,079
H3	STRATKM → FIRMPER	0,045
H4	ANATECH → ANACULT	<b>0,602***</b>
H5	STRATKM → ANACULT	0,252
H6	ANATECH ↔ STRATKM	0,152

Huomio: Lihavoidut luvut ovat merkitseviä tasoilla \*p < 0,1; \*\*p < 0,05; \*\*\*p < 0,01.

**Taulukko 14.** Tulokset, pieni runsaan ja monimuotoisen datan hyödyntämisen taso

## 5 TUTKIMUSTULOKSET

Tämän kappaleen tarkoituksena on kuvata tutkielman empiirisen analyysin tulokset. Kappaleessa tarkastellaan hypoteesien kautta sitä, millaisen vastauksen analyysi antoi tutkittavaan ongelmaan. Kappaleessa keskitytään tarkastelemaan myös tulosten merkitystä ja uskottavuutta. Analyysin antamat tulokset ovat koottuna taulukkoon 15.

Hypoteesi		Estimaatti		
		Yhtenäinen malli	Suuri Big Data	Pieni Big Data
H1	ANACULT → FIRMPER	0,100	0,087	0,130
H2	ANATECH → FIRMPER	0,027	-0,205	0,079
H3	STRATKM → FIRMPER	0,066	0,209	0,045
H4	ANATECH → ANACULT	<b>0,598***</b>	<b>0,552***</b>	<b>0,602***</b>
H5	STRATKM → ANACULT	<b>0,243**</b>	<b>0,275**</b>	0,252
H6	ANATECH ↔ STRATKM	<b>0,414***</b>	<b>0,563***</b>	0,152

Huomio: Lihavoidut luvut ovat merkitseviä tasoilla \*p < 0,1; \*\*p < 0,05; \*\*\*p < 0,01.

**Taulukko 15.** Analyysin tulokset

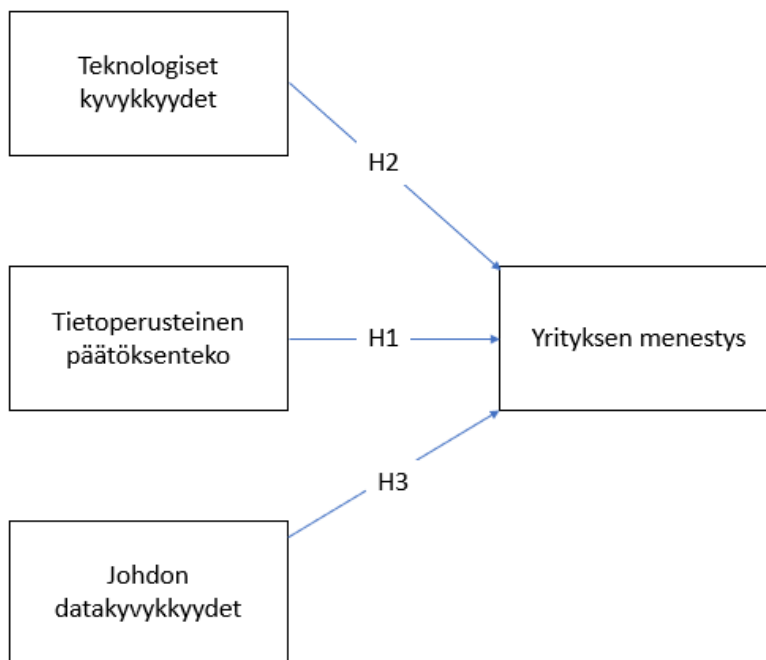
### 5.1 Tietoperusteisen päätöksenteon vaikutus menestykseen

Tietoperusteisen päätöksenteon, sekä siihen liittyvien teknologisten ja johdon strategisten datakyvykkyyksien merkitystä tarkasteltiin koko aineiston osalta yhtenäisen mallin avulla. Tarkoituksena oli saada vastaus hypoteeseihin 1, 2 ja 3. Hypoteesit pohjautuivat teoreettiseen tutkimusmalliin kuvion 10 mukaisesti.

**H1:** *Tietoperusteisen päätöksenteon hyödyntäminen johtaa yrityksen parempaan menestymiseen.*

**H2:** *Teknologisten kyvykkyyksien parempi taso johtaa parempaan menestymiseen.*

**H3:** *Johdon datakyvykkyyksien parempi taso johtaa parempaan menestymiseen.*



**Kuvio 10.** H1, H2 ja H3

Analyysin perusteella yksikään yrityksen menestystä koskeva hypoteesi ei saanut merkitsevää p-arvoa. Tämän vuoksi hypoteesit 1, 2 ja 3 joudutaan hylkäämään. Tulos ei kuitenkaan ole täysin yksiselitteinen. Näin on siksi, että tutkimuksessa yrityksen menestystä mitattiin ja analysoitiin erittäin subjektiivisesti yrityksen taloudellisen tilanteen, kasvun ja työntekijöiden tuottavuuden perusteella. Näiden mittarien keskinäinen vertailtavuus ei yllä objektiivisen vastineen tasolle, jota voitaisiin mitata vertailemalla esimerkiksi yritysten kokoon suhteutettuja tilinpäätöstietoja.

Subjektiivisen menestyksen vertailu saattoi olla syy toiseen ongelmaan. Yrityksen menestyksen mittari nimittäin epäonnistui validiteetin ja reliabiliteetin osalta selittämään muita latentteja muuttujia. Tämän vuoksi hypoteesien hylkääminen ei osoita hypoteesien mukaisten yhteyksien puutosta. Validiteetti- ja reliabiliteetti-ongelmaan menestysmittarin osalta olisi mahdollisesti voitu vaikuttaa lisäämällä tutkimukseen muita menestystä mittaavia tekijöitä. Rakenneyhtälömallinnus ei mahdollistanut yrityksen menestystä mittaavan latentin muuttujan supistamista, joka heikensi selvästi sekä validiteettia että reliabiliteettia.

Koko aineiston tarkastelun lisäksi tietoperusteisen päätöksenteon vaikutusta menestykseen tarkasteltiin myös Big Datan hyödyntämisen valossa. Big Dataa tarkasteltiin runsaan ja monimuotoisen datan hyödyntämisen kautta. Hypoteesin 8 tarkoituksena oli vertailla hypoteesien 1, 2 ja 3 eroja jaetussa aineistossa.

**H8:** *Monimuotoisen ja runsaan datan hyödyntäminen vahvistaa a) tietoperusteisen päätöksenteon ja b) kyvykkyyksien vaikutusta menestykseen.*

Myös jaetun aineiston osalta hypoteesit 1, 2 ja 3 joudutaan analyysin perusteella hylkäämään. Tähän vaikuttaa samat yrityksen menestysfaktoria koskevat validiteetti- ja reliabiliteettiongelmat. Menestyksen mittaaminen on erittäin hankalaa, koska siihen vaikuttaa erittäin monet tekijät. Vaikka datasta onnistuttaisiin luomaan arvoa suunnitellusti, ei sekään välttämättä takaa yrityksen menestystä.

Tietoperusteisen päätöksenteon vaikutusta menestykseen on syytä tarkastella vielä tutkielman teoriaosuuden kautta, jotta aiemman tutkimuksen kontribuutio näkyy tuloksissa. Tätä tarkastellaan tarkemmin johtopäätökset -kappaleessa. Teoreettisesta näkökulmasta datan arvo syntyy päätöksenteon tehokkuuden kautta. Reaktioajan viiveteorian mukaisesti olennaista arvon syntymiselle on aika, joka jää datan keräämisen ja päätöksenteon väliin. Reaktioajan viiveteoriasta johdettu tutkielman teoreettisen osion yhteenvetomalli osoittaa kyvykkyyksien merkityksen datan arvonluonnissa. Kyvykkyyksiin liittyviä tutkimuksen tuloksia tarkastellaan seuraavaksi.

## **5.2 Kyvykkyyksien vaikutus tietoperusteiseen päätöksentekoon**

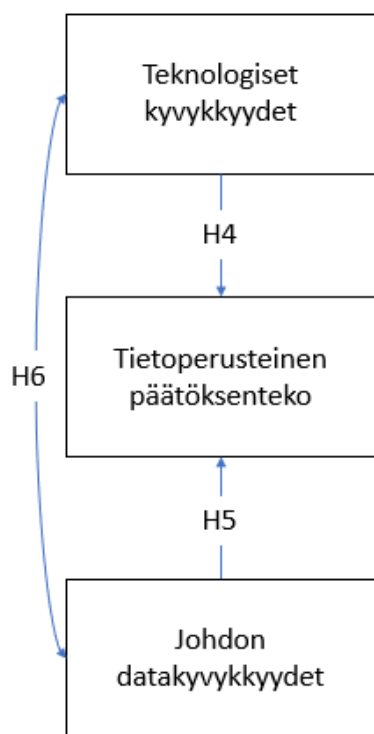
Reaktioajan viiveeseen pystytään vaikuttamaan teknologisten kyvykkyyksien avulla. Vastaavasti johdon datakyvykkyydet vaikuttavat datan hyödyntämispotentiaaliin. Teknologiat mahdollistavat datan oikeanlaisen ja nopean käsittelyn päätöksenteon perustaksi. Johdon datakyvykkyydet puolestaan mahdollistavat tietoperusteisen

päätöksenteon olemassaolon. Kyvykkyyksien vaikutusta tietoperusteiseen päätöksentekoon tutkittiin tutkimuksessa hypoteesien 4, 5 ja 6 kautta.

**H4:** *Teknologiset kyvykkyydet vaikuttavat positiivisesti tietoperusteiseen päätöksentekoon.*

**H5:** *Johdon datakyvykkyydet vaikuttavat positiivisesti tietoperusteisen päätöksentekoon.*

**H6:** *Teknologiset kyvykkyydet ja Johdon datakyvykkyydet tukevat toinen toisiaan.*



**Kuvio 11.** H4, H5 ja H6

Tutkimustulosten osalta tutkielman varsinainen kontribuutio kohdistuu kuvion 11 mukaisesti tutkimusmallin tietoperusteisen päätöksenteon osa-alueeseen. Näihin tutkittaviin ongelmiin saatiin tutkimuksesta luotettavampia tuloksia. Työn teoreettinen osa-alue kiteytyy myös tähän kokonaisuuteen.

Teknologisten kyvykkyyksien vaikutusta tietoperusteiseen päätöksentekoon tarkasteltiin hypoteesin 4 avulla. Standardoitu  $\beta$ -estimaatti hypoteesille oli **0,598\*\*\*** (\*\*p < 0,01). Tulos siis tukee hypoteesia saaden erittäin hyvän merkitsevyysarvon. Tutkimustuloksen perusteella teknologisilla kyvykkyyksillä on positiivinen merkitys tietoperusteiseen päätöksentekoon.

Teknologisten kyvykkyyksien osalta lopullisessa analyysissä tutkittiin modernin teknologian, datan visualisointityökalujen ja pilvipalveluiden hyödyntämistä datan tallentamiseen, säilömiseen, käsittelyyn ja analysointiin. Tietoperusteista päätöksentekoa tarkasteltiin puolestaan kyvyllä tehdä päätöksiä dataan ja siitä jalostettuun tietoon perustuen sekä sisäisesti että ulkoisesti. Lisäksi faktori mittasi luottamusta dataan tehtäessä päätöksiä. Analyysistä saatu tulos osoittaa, että teknologiset kyvykkyydet asettavat selväsi valmiudet tietoperusteiselle päätöksenteolle. Tulosten osalta teknologiset kyvykkyydet asettuivat merkittävimpään asemaan selitettäessä tietoperusteista päätöksentekoa.

Saatu tulos on tulkittavissa myös niin, että teknologiset kyvykkyydet ohjaavat päätöksiä tietoperusteiseen suuntaan. Yritysten teknologiset valmiudet helpottavat datasta jalostetun tiedon saantia. Näin ollen yritysten on helpommin mahdollista perustella päätöksiään dataan kuin intuitioonsa perustuen. Samalla luottamus datasta jalostetun tiedon hyödyntämiselle päätöksenteossa kasvaa entisestään.

Hypoteesin 5 tavoitteena oli tutkia johdon strategisten datakyvykkyyksien vaikutusta tietoperusteiseen päätöksentekoon. Standardoitu  $\beta$ -estimaatti hypoteesille oli **0,243\*\*** (\*\*p < 0,05). Analyysi siis tukee asetettua hypoteesia. Myös johdon datakyvykkyydet vaikuttavat teknologisten kyvykkyyksien tavoin tietoperusteiseen päätöksentekoon. Yritysten kaikki toiminta perustuu johdon asettamiin strategisiin tavoitteisiin. Johdon datakyvykkyyksiä mitattiin analyysissä strategian tietoperustaisen suunnittelun, kehittämisen ja viestinnän näkökulmista.

Se, että tiedon merkitys on myös otettu strategisella tasolla huomioon, edesauttaa tietoperusteisen päätöksenteon olemassaoloa. Strategian kehittäminen tiedon ja osaamisen pohjalta viittaa puolestaan suoraan tietoperusteiseen strategiseen päätöksentekoon. Johdon datakyvykkyydet siis vaikuttavat positiivisesti tietoperusteiseen päätöksentekoon.

Analyysin tuloksen perusteella johdon datakyvykkyyksien merkitys tietoperusteiselle päätöksenteolle ei ollut yhtä vahvaa kuin teknologisten kyvykkyyksien. Tämä tulos ei kuitenkaan ole täysin yksiselitteinen. Ilman teknologisia kyvykkyyksiä yritys ei pystyisi jalostamaan dataa ollenkaan tietoperusteiseen päätöksentekoon. Vastaavasti ilman johdon strategista tukea teknologioihin ei panosteta tai strategiaa ohjata tietoperustaiseen suuntaan. Aikaisemmat tutkimukset olivat lisäksi osoittaneet, että yritysjohdon haluttomuus panostaa teknologioihin, jotka keräävät mahdollisesti myös turhaa dataa, oli yksi isoimmista syistä tietoperusteisen päätöksenteon epäonnistumiselle. Ilman toista ei siis voi olla toista. Tämän vuoksi kyvykkyyksien välistä yhteyttä tutkittiin mallissa hypoteesilla 6.

Hypoteesi 6 tarkasteli teknologisten kyvykkyyksien ja johdon datakyvykkyyksien välistä suhdetta. Standardoitu  $\beta$ -estimaatti hypoteesille oli **0,414\*\*\*** (\*\*p < 0,01). Kyvykkyyksien välillä on näin positiivinen yhteys. Ne selittävät toistensa vaihteluita melko hyvin. Tämä tulos osoittaa, että teknologiset kyvykkyydet ja strategiset kyvykkyydet tukevat toinen toisiaan, tukien ajatusta siitä, että molempia tarvitaan tietoperusteisen päätöksenteon olemassaoloon ja onnistumiseen.

Tulokset tukevat asetettuja hypoteeseja osoittaen, että kyvykkyydet asettavat edellytykset tietoperusteiselle päätöksenteolle. Samalla ne ohjaavat päätöksentekoa tietoperusteiseen suuntaan. Tutkimuksessa asiaan perehdyttiin kuitenkin vielä tarkemmin tarkastelemalla Big Datan hyödyntämisen vaikutusta jakamalla tutkittava aineisto perustuen runsaan ja monimuotoisen datan hyödyntämiseen. Tätä tutkittiin hypoteesilla 7, jonka tarkoituksena oli tutkia hypoteesien 4, 5 ja 6 muutoksia jae-tussa mallissa.

**H7:** *Monimuotoisen ja runsaan datan hyödyntäminen vahvistaa a) teknologisten kyvykkyyksien ja b) johdon datakyvykkyyksien merkitystä tietoperusteiseen päätöksentekoon sekä c) vahvistaa kyvykkyyksien välistä yhteyttä.*

Tarkasteltaessa ensin teknologisten kyvykkyyksien vaikutusta tietoperusteiseen päätöksentekoon (H7a) nähdään jo tuloksissa pientä eroa. Hypoteesin standardoitu  $\beta$ -estimaatti suurelle monimuotoisen datan hyödyntämiselle oli **0,552\*\*\*** (\*\*p < 0,01) ja vastaavasti pienelle **0,602\*\*\*** (\*\*p < 0,01). Kummassakin tapauksessa teknologiset kyvykkyydet selittävät erittäin hyvin tietoperusteisen päätöksenteon vaihteluita. Tulokset saivat lisäksi erittäin hyvät merkitsevyysarvot.

Mielenkiintoinen havainto on kuitenkin se, että vähäisemmän monimuotoisen datan hyödyntämisen osalta teknologiset kyvykkyydet vaikuttavat hieman enemmän tietoperusteiseen päätöksentekoon. Ero ei ole suuri. Se kuvaa hyvin sitä, että vähäisemmällä monimuotoisen datan hyödyntämisen tasolla teknologioiden ja teknologisten kyvykkyyksien merkitys korostuu tietoperusteisessa päätöksenteossa. Vaikka yrityksessä ei strategisella tasolla oltaisi vielä kovinkaan pitkällä datan tai Big Datan hyödyntämisessä, voivat teknologiat tarjota tietoa, jota halutaan hyödyntää päätöksissä. Vastaavasti tulos osoittaa, että teknologiset kyvykkyydet ohjaavat kummassakin tapauksessa päätöksentekoa tietoperusteiseen suuntaan.

Hypoteesi 7a joudutaan siis tulosten valossa hylkäämään. Monimuotoisen ja runsaan datan hyödyntäminen ei vahvista teknologisten kyvykkyyksien merkitystä tietoperusteiseen päätöksentekoon. Teknologisten kyvykkyyksien merkitys on jopa suurempaa, kun runsaan ja monimuotoisen datan hyödyntäminen on vähäistä.

Runsaan ja monimuotoisen datan hyödyntämisen vaikutusta johdon datakyvykkyyksien merkitykseen tietoperusteiselle päätöksenteolle tutkittiin hypoteesilla H7b. Hypoteesin standardoitu  $\beta$ -estimaatti suurelle monimuotoisen datan hyödyntämiselle oli **0,275\*\*** (\*\*p < 0,05) ja vastaavasti pienelle 0,252 (p > 0,1). Hypoteesi sai merkitsevän p-arvon suuren monimuotoisen ja runsaan datan hyödyntämisen mal-



lissa. Vastaavasti pienen monimuotoisen ja runsaan datan hyödyntämisen mallin p-arvo hypoteesille ei ollut merkitsevä, p-arvon ollessa yli 0,1.

Tulokset osoittavat, että johdon datakyvykkyudet selittävät tietoperusteista päätöksentekoa melko hyvin yrityksen hyödyntäessä monimuotoista ja runsasta dataa eli Big Dataa. Vastaavasti yrityksissä, joissa monimuotoista dataa hyödynnetään vain vähän tai ei ollenkaan, tilastollista vahvistusta johdon datakyvykkyysien vaikutuksesta ei ole.

Tutkimus ei ota kantaa siihen, millä tasolla tietoperusteista päätöksentekoa toteutetaan tutkittavissa yrityksissä. Tämän vuoksi tulokset heijastavat selvästi sitä, että runsasta ja monipuolista dataa paljon käyttävät yritykset tarvitsevat johdon strategista tukea, jotta tietoperusteinen päätöksenteko on mahdollista. Tämä johtuu siitä, että tietoperusteinen päätöksenteko on näissä yrityksissä paljon korkeammalla tasolla. Kuten tässä tutkielmassa on teoreettisesti kuvattu, Big Data on nykymittareilla todella korkeatasoista dataa. Sen hyödyntäminen ei onnistu perinteisillä menetelmillä, joille se on liian raskasta.

Tuloksen perusteella H7b hyväksytään. Runsaan ja monimuotoisen datan hyödyntäminen vahvistaa johdon datakyvykkyysien merkitystä tietoperusteiseen päätöksentekoon. Toisin sanottuna Big Dataa hyödyntävät yritykset tarvitsevat johdon strategista tukea, jotta päätöksiä voidaan tehdä tietoperusteisesti. Vastaavasti johdon datakyvykkyysien parempi taso johtaa korkeampaan päätösten tietoperusteisuuteen.

Runsaan ja monimuotoisen datan hyödyntämisen vaikutusta tutkittiin vielä hypoteesin H7c avulla, jonka tarkoituksena on verrata kyvykkyysien välistä yhteyttä jaetussa mallissa. Hypoteesin standardoitu  $\beta$ -estimaatti suurelle monimuotoisen datan hyödyntämiselle oli **0,563\*\*\*** (\*\*\*)  $p < 0,01$ ) ja vastaavasti pienelle 0,152 ( $p > 0,1$ ). Tämän hypoteesin tulosten osalta ero on erittäin selvä.

Suurella monimuotoisen ja runsaan datan hyödyntämisen tasolla teknologisten kyvykkyyksien ja johdon datakyvykkyyksien välinen yhteys on erittäin huomattavaa. Tulos sai myös todella hyvän merkitsevyysarvon. Kyvykkyydet siis tukevat toisiaan vahvasti. Vähäisen monimuotoisen datan hyödyntämisen mallissa tilastollisesti merkitsevää yhteyttä kyvykkyyksien välille ei löytynyt. Teknologiset kyvykkyydet ovat siis täysin erillisiä johdon kyvykkyyksiin nähden.

Hypoteesi H7c hyväksytään tulosten perusteella. Runsaan ja monimuotoisen datan hyödyntäminen vahvistaa kyvykkyyksien välistä yhteyttä. Big Data hyödyntävissä yrityksissä teknologiset ja strategiset kyvykkyydet tukevat toisiaan voimakkaasti. Myös tämä tulos vahvistaa ajatusta siitä, että tietoperusteisen päätöksenteon korkeammalla tasolla johdon strategisten tavoitteiden merkitys korostuu.

### **5.3 Tulosten yhteenveto**

Rakenneyhtälömallinnusta hyödyntäneen regressioanalyysin tulosten perusteella tutkimuksen hypoteesit 1, 2, 3 ja 8 hylättiin. Tutkimus ei siis pystynyt osoittamaan, että tietoperusteisella päätöksenteolla tai siihen keskeisesti liittyvillä teknologisilla ja johdon datakyvykkyyksillä olisi suoraa yhteyttä yrityksen menestymiseen. Tulos ei kuitenkaan todellisuudessa kerro yhteyden puuttumisesta, sillä yrityksen menestystä mitannut faktori epäonnistui validiteetin ja reliabiliteetin osalta. Aikaisemmat tutkimukset ovat osoittaneet, että tiedon yhteys arvonluontiin tapahtuu kuitenkin tehokkaamman päätöksenteon myötä. Yrityksen menestykseen vaikuttaa päätöksentekoon pohjautuvan arvonluonnin lisäksi paljon muitakin tekijöitä, joita tässä tutkimuksessa ei oteta huomioon. Vaikka yritys onnistuisi luomaan arvoa datasta, se ei itsessään ole avain menestykseen.

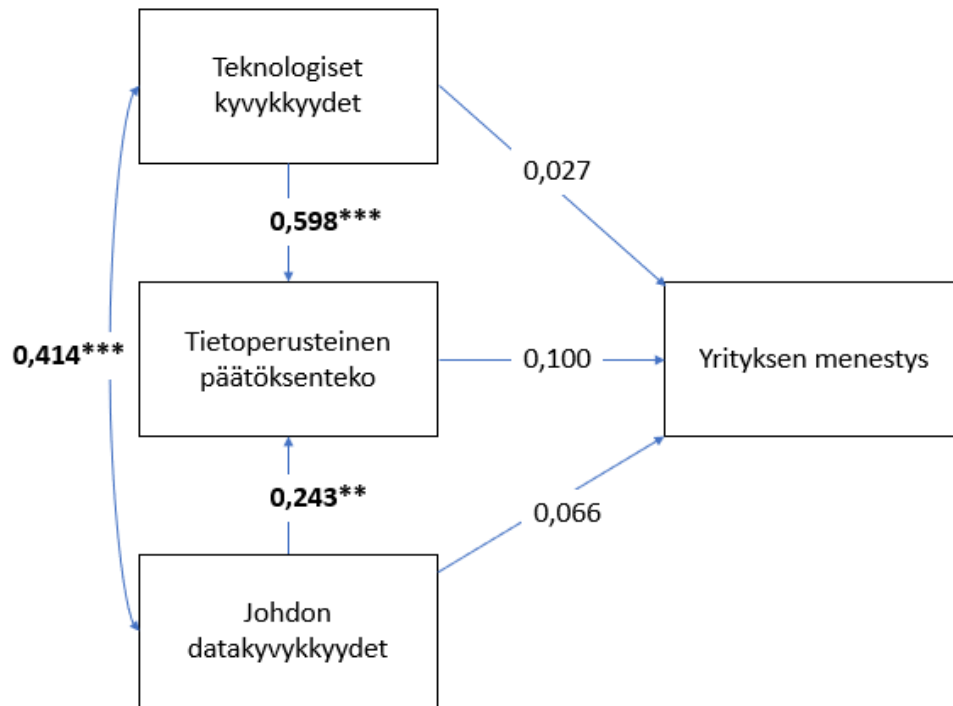
Big Datan ja datan yleisesti on osoitettu olevan itsessään arvotonta. Datan arvo syntyy jalostamalla sitä tiedoksi, jota puolestaan voidaan hyödyntää päätöksenteossa. Myös tämän tutkimuksen tulokset osoittavat samaa, sillä teknologisten ky-

vykykyksien merkitys tietoperusteiseen päätöksentekoon oli tulosten perusteella erittäin voimakasta. Teknologiat mahdollistavat datan jalostamisen tiedoksi. Ne ovat suorastaan perusedellytys tietoperusteiseen päätöksentekoon. Tutkimuksen tulokset ovat vielä kuvattuna kuvioissa 12, 13 ja 14. Kaikissa tutkituissa tapauksissa teknologiset kyvykkyudet selittivät tietoperusteista päätöksentekoa voimakkaasti, mikä on nähtävissä kuvioista. Tulosten perusteella hypoteesi 4 hyväksyttiin.

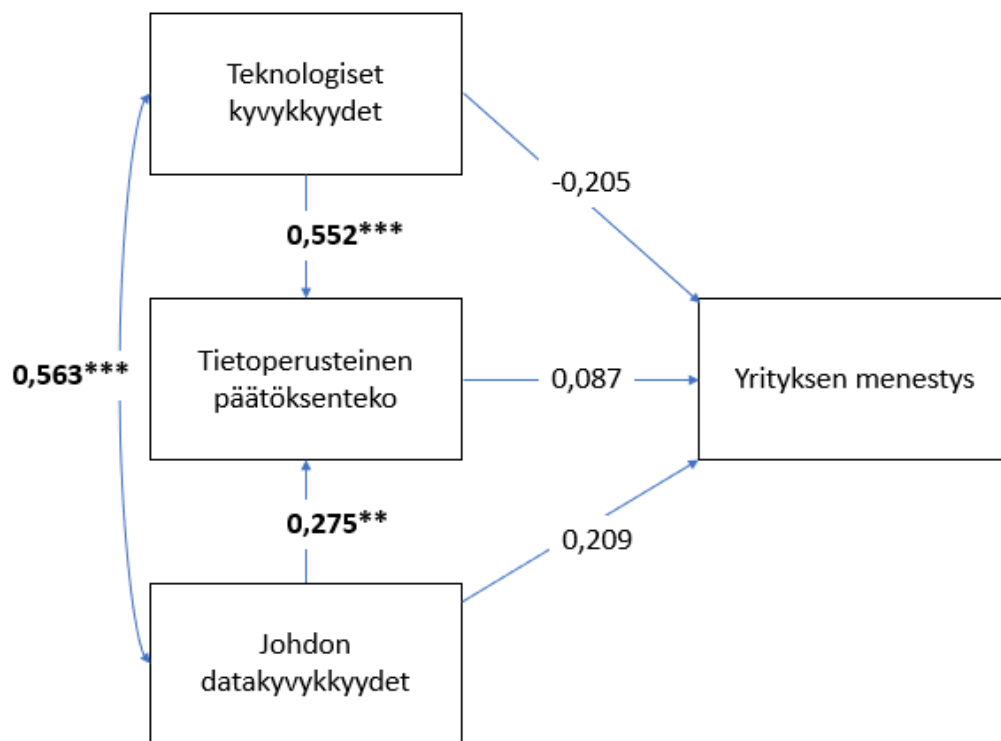
Teknologisten kyvykkyyksien osalta mielenkiintoinen tulos kävi ilmi vertailtaessa Big Datan hyödyntämisen vaikutusta malliin. Vähäisemmän Big Datan hyödyntämisen tapauksessa teknologisten kyvykkyyksien merkitys korostui entisestään. Tämän vuoksi hypoteesi 7a jouduttiin hylkäämään.

Strategisten johdon datakyvykkyyksien positiivinen vaikutus tietoperusteiseen päätöksentekoon oli tutkimuksen tulosten perusteella myös ilmeistä. Tietoperusteinen päätöksenteko tarvitsee johdon strategista tukea onnistuakseen. Tämä yhteys näkyy kuvioista 12. Hypoteesi 5 hyväksyttiin saatujen tulosten perusteella. Tuloksista oli ilmeistä, että johdon strategisen tuen tarve kasvaa, kun yritys hyödyntää korkeatasoisempaa data-analytiikkaa tietoperusteiseen päätöksentekoon. Hypoteesi 7b hyväksyttiin tulosten perusteella.

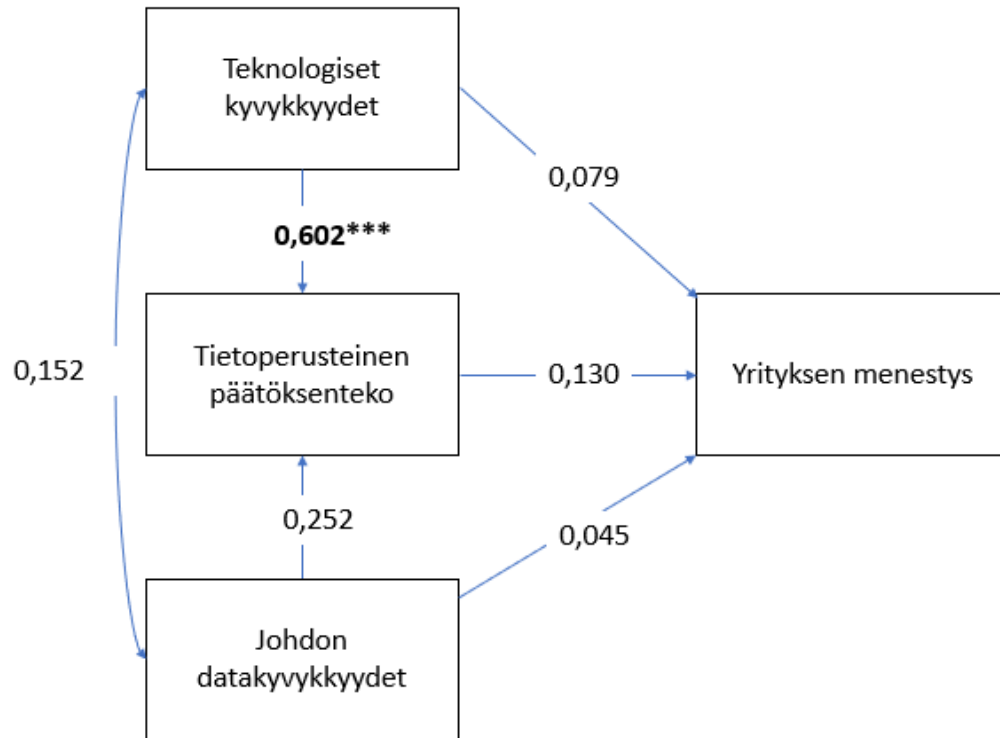
Teknologisten kyvykkyyksien ja johdon datakyvykkyyksien keskinäistä yhteyttä tutkittiin hypoteesilla 6. Tulokset osoittivat niiden tukevan selvästi toisiaan (kuvio 12). Tulosten kannalta mielenkiintoinen havainto syntyi jälleen monimuotoisen ja runsaan datan hyödyntämisestä. Kuvioista 13 ja 14 nähdään, että kyvykkyudet tukevat toisiaan erittäin voimakkaasti, kun monimuotoisen datan hyödyntäminen on suurta. Vastaavasti tätä tilastollista yhteyttä ei löytynyt vähäisen hyödyntämisen osalta. Tämä tukee edelleen ajatusta korkeamman tason analytiikan strategisen tuen tarpeesta. Tulosten perusteella hypoteesit 6 ja 7c hyväksytään.



**Kuvio 12.** Yhdistetyn mallin tulokset



**Kuvio 13.** Suuren monimuotoisen ja runsaan datan hyödyntämisen tulokset



**Kuvio 14.** Vähäisen monimuotoisen ja runsaan datan hyödyntämisen tulokset

## 6 JOHTOPÄÄTÖKSET

Tämän kappaleen tarkoituksena on kerätä teoreettiset ja empiiriset havainnot sekä tulokset yhteen. Tuloksia tarkastellaan tutkielman viitekehysten valossa. Tässä kappaleessa vastataan myös tutkielman alussa asetettuihin tutkimuskysymyksiin ja pohditaan tutkimuksen kontribuutiota. Kappaleen lopussa esitetään lisäksi aiheeseen liittyviä jatkotutkimusehdotuksia.

Teoreettisesta tarkastelusta tämän tutkielman aihe rakentui kappaleessa 2 kuvattujen teoreettisten lähtökohtien ympärille. Teoria kuvasi datasta syntyvän arvon muodostuvan prosessissa, joka alkaa datan keräämisestä. Dataa on puolestaan mahdollista kerätä erittäin monimuotoisesti ja runsaasti. Tämän vuoksi tutkimuksissa on vahvasti noussut esiin Big Datan käsite, jolla viitataan luontinopeudeltaan, monimuotoisuudeltaan ja kooltaan suureen datamäärään. Big Datan osalta olennaista on sen asettamat uudet vaatimukset, kun perinteiset menetelmät eivät enää riitä sen analysointiin. Arvonluontiprosessin kannalta Big Data ei kuitenkaan ole välttämätöntä. Arvoa on mahdollista luoda samoilla periaatteilla myös pienemmistä datamääristä. Big Data kuitenkin mahdollistaa tarkemman ja kattavamman analysoinnin, joten siitä on mahdollista työstää enemmän arvoa.

Arvonluontiprosessin toinen vaihe on datan analysointi yrityksessä käytettävissä olevien teknologioiden avulla. Teoreettisesti teknologiat mahdollistavat jopa Big Datan reaaliaikaisen analysoinnin. Tämä analysoinnin vaihe on teoreettisesti arvonluonnille tärkein, koska datan arvo syntyy vasta jalostamalla sitä informaatioksi ja tiedoksi. Teknologiset kyvykkyydet ovat myös reaktioajanviiveteorian perusteella erittäin keskeisiä. Koko arvonluontiprosessiin vaikuttaa siihen käytetty aika. Datan potentiaalinen arvo laskee ajan kuluessa, koska se kuvaa aina tietyn hetken tilaa. Tarpeeksi vanha data saattaa muuttua jopa täysin arvottomaksi, kun datan keräämisen ja analysoinnin aikana on syntynyt jo paljon uutta dataa.

Viimeisen vaiheen arvonluontiprosessille asettaa analyysin tuottaman informaation ja tiedon hyödyntäminen päätöksenteossa. Tähän vaikuttaa teoreettisesti yrityksen tietoperustaisten päätösten kulttuuri sekä yrityksen johdon strategiset kyvykkyydet ohjata päätöksentekoa tietoperustaiseen suuntaan. Mikäli päätöksissä luotetaan analyysin tuottamaan tietoon, voidaan päätöksiä tehdä välittömästi analyysin valmistuttua. Päätöksentekoprosessi sisältää yleisesti paljon sellaisia tekijöitä, joihin datasta jalostetulla tiedolla voidaan vaikuttaa. Näin päätöksentekoa saadaan tehostettua erittäin paljon.

Viitekehyksen mukaisesti asetettuja tutkimuskysymyksiä testattiin empiirisessä aineistossa hypoteeseilla. Tutkimuksen tarkoituksena oli pyrkiä vastaamaan asetettuihin tutkimuskysymyksiin. Tutkielman päätutkimuskysymyksenä oli:

*Miten datan hyödyntäminen päätöksenteossa vaikuttaa yrityksen menestymiseen?*

Aiemman tutkimuksen perusteella datan hyödyntäminen päätöksenteossa vaikuttaa menestykseen tämän tutkielman mukaisesti arvonluonnin kautta. Esimerkiksi Kubina et al. (2015) näkivät vahvan yhteyden Big Datan ja päätöksentekoon perustuvan arvonluonnin välillä. Yritysten on siis data-analytiikan avulla helpompi mitata sisäistä ja ympäröivää liiketoiminnan kenttää ja tehdä liiketoimintaa edistäviä päätöksiä tulosten perusteella. Lisäksi data toimii resurssina, joka luo kestävästä kilpailuetua (Grant 1996).

Yritysten keräämä data on hyvin usein yksilöllistä ja luonteeltaan vaikeasti kopioitavaa. Mikäli yrityksellä on esimerkiksi kolmen vuoden ajalta dataa asiakkaistaan, menee uudella kilpailijalla vastaavaan datan keräämiseen kolme vuotta. Tässä vaiheessa yritys on kuitenkin jälleen kolme vuotta edellä datan keräyksessä. Datan avulla on siis mahdollista olla aina askeleen edellä kilpailijaa saavuttaen kestävästä kilpailuetua. Teknologiset valmiudet puolestaan määrittävät sitä, kuinka paljon ja nopeasti datasta saadaan jalostettua tietoa. Tämä tarkoittaa myös sitä, että datan avulla saavutettu kilpailuetu vaatii yritykseltä tarpeeksi monimutkaista dataa, jotta

datan kopiointi tai vastaavan datamäärän kerääminen olisi vaikeaa kilpailijalle. Esimerkiksi Big Data toimii tähän vastauksena.

Tutkielman empiirisessä osiossa pyrittiin tarkastelemaan tietoperusteisen päätöksenteon, teknologisten kyvykkyyksien ja johdon datakyvykkyyksien vaikutusta subjektiivisesti mitattuun menestykseen. Kyvykkyyksien osalta tutkimuksessa oli asetettuna alatutkimuskysymys, johon vastausta etsittiin yhdessä päätutkimuskysymyksen kanssa. Alatutkimuskysymyksenä oli:

*Miten data- ja teknologiakyvykkyudet vaikuttavat yrityksen menestymiseen?*

Menestyksen osalta pyrittiin tutkimaan myös runsaan ja monimuotoisen datan hyödyntämisen vaikutusta tietoperusteiseen päätöksentekoon ja kyvykkyyksiin. Menestysmittarin validiteetti- ja reliabiliteettiongelmiensa vuoksi tietoperusteisen päätöksenteon tai siihen liittyvien kyvykkyyksien suoria vaikutuksia menestykseen ei pystytty tutkimuksen perusteella osoittamaan kumpaankaan suuntaan. Aiemmat tutkimukset kuten Akter et al. (2016) ja Wamba et al. (2017) ovat kuitenkin osoittaneet, että Big Data -kyvykkyyksillä on positiivinen vaikutus yritysten menestymiseen. Esimerkiksi Wamba et al. (2017) tulokset osoittivat Big Data kyvykkyyksien selittävän 65 % menestysfaktorin varianssista.

Tietoperusteisen päätöksenteon on aiemmissa tutkimuksissa osoitettu mahdollistavan arvonluontia (Raghunathan 1999). Tämä ei kuitenkaan tarkoita sitä, että tietoperusteinen päätöksenteko toimisi itsessään menestyksen luojana. Tietoperusteinen päätöksenteko on yrityksen varsinaista liiketoimintaa tukevaa toimintaa. Tietoperusteisella päätöksenteolla on siis osoitettu olevan menestystä vauhdittava vaikutus. Vaikka suoria yhteyksiä ei pystytty empiirisessä tutkimuksessa osoittamaan, tietoperusteista päätöksentekoa tutkittiin vielä kahdella alatutkimuskysymyksellä, joilla pyrittiin saamaan parempi käsitys myös päätutkimuskysymykseen. Alatutkimuskysymykset olivat:



*Miten data- ja teknologiakyvykkyudet vaikuttavat tietoperusteiseen päätöksentekoon?*

*Miten runsaan ja monipuolisen datan hyödyntäminen vaikuttaa data- ja teknologiakyvykkyysiin?*

Tämän tutkielman varsinainen kontribuutio syntyiikin tutkittaessa näitä osa-alueita tietoperusteisesta päätöksenteosta. Tutkimus osoitti, että teknologiset kyvykkyudet ja johdon datakyvykkyudet ovat erittäin keskeisessä asemassa tietoperusteista päätöksentekoa selitettäessä, kuten myös teorian pohjalta pystyttiin oletamaan. Tätä ajatusta tukee esimerkiksi Akter et al (2016) tutkimuksen tulokset. Ne osoittivat Big Data -kyvykkyyksillä olevan positiivinen yhteys Big Datan hyödyntämistaitoihin.

Teoreettisesta tarkastelusta katsottuna kyvykkyudet toimivat koko prosessin mahdollistavana tekijänä. Tutkimuksen tulokset osoittivat, että teknologiset kyvykkyudet selittävät tietoperusteista päätöksentekoa riippumatta siitä, millä tasolla runsaan ja monimuotoisen datan hyödyntäminen on. Myös Akter et al. (2016) tulokset osoittivat teknologisten kyvykkyyksien olevan avainkomponentti Big Data -kyvykkyksiä tarkasteltaessa. Heidän tuloksensa mukaan Big Dataa hyödyntävien yritysten tulisi panostaa kyvykkyysiin. Tämän tutkimuksen tulosten perusteella vähäisellä Big Datan tasolla teknologiat selittivät jopa enemmän tietoperusteista päätöksentekoa. Tämä oli erittäin mielenkiintoinen havainto, sillä hypoteesilla kuvattuna alkuoletuksena oli, että runsaan ja monimuotoisen datan suuremmalla tasolla teknologiset kyvykkyudet selittäisivät enemmän tietoperusteista päätöksentekoa.

Kun yritys käyttää tiedon lähteenä Big Dataa kevyempää dataa, asettuvat teknologiset kyvykkyudet tutkimuksen perusteella kaikista keskeisimpään rooliin. Yritykset siis tekevät päätöksiä tietoperusteisesti, kunhan teknologiset valmiudet ovat olemassa. Tämän tason tiedon hyödyntäminen päätöksissä ei tutkimuksen perusteella

tarvitse johdon strategista tukea. Tulos on myös tulkittavissa niin, että teknologiset kyvykkyudet ohjaavat yritystä tekemään päätöksiä tietoperusteisesti. Tämä vaikuttaa loogiselta, koska teknologiat mahdollistavat tietomäärän kasvattamisen päätösten perustaksi. Tutkimuksen perusteella varteenotettava ensiaskel tietoperusteiseen päätöksentekoon on asettaa ensin teknologiset kyvykkyudet hyvälle tasolle.

Vastaavasti tämä tutkimus osoitti, että runsasta ja monipuolista dataa hyödyntävien yritysten osalta johdon strategiset datakyvykkyudet nousivat teknologisten kyvykkyysien tavoin tärkeäksi. Kun data-analytiikan tasoa nostetaan, tarvitsee tietoperusteinen päätöksenteko myös strategiatason tukea onnistuakseen. Tähän voi osaltaan vaikuttaa se, että monipuolisemmasta datasta on mahdollista saada tarkempaa tietoa kaikkien päätöstasojen tueksi. Saadun tiedon tarkentuessa, sitä on näin mahdollista käyttää laajemmin ja tietoon on mahdollista luottaa enemmän. Tämä puolestaan saattaa vaatia yrityksen johdon tukea tietojohdamisen osalta.

Tuloksissa on otettava huomioon myös se tosiasia, että Big Data ja sen hyödyntäminen on varsinkin Suomessa vielä alkutekijöissään. Big Dataa on tutkittu runsaasti vasta 2010-luvulla, joten sen vaikutukset ovat vasta alussa. Big Dataan perustuva tietoperusteinen päätöksenteko on toistaiseksi tulevaisuuden kasvava trendi, jota kilpailu tulee mahdollisesti vaatimaan.

Päätutkimuskysymykseen vastaten tietoperusteinen päätöksenteko mahdollistaa datan jalostamisen arvokkaaksi tiedoksi, jota yritys voi hyödyntää menestyksen saavuttamiseksi. Teknologiset kyvykkyudet ja johdon datakyvykkyudet vaikuttavat välillisesti menestykseen mahdollistamalla ja parantamalla tietoperusteista päätöksentekoa. Kyvykkyudet siis määrittävät erittäin paljon sitä, kuinka tietoperusteinen päätöksenteko onnistuu. Runsaan ja monimuotoisen datan hyödyntäminen asettaa puolestaan strategisia vaatimuksia tietoperusteiselle päätöksenteolle.

Kaiken kaikkiaan tämä tutkimus lisää ymmärrystä tietoperusteisen päätöksenteon lähtökohtiin. Tutkimus tuo teoreettista kontribuutiota tietojohdamiseen, havainnoimalla tietoperusteiseen päätöksentekoon liittyvien kyvykkyyksien merkitystä tutkittu aineistossa. Tutkimus nosti esiin teknologisten kyvykkyyksien tärkeän merkityksen tietoperusteisen päätöksenteon onnistumiselle.

Lisäksi johdon datakyvykkyyksien merkityksen osoitettiin kasvavan päätöksenteon perustaksi analysoitavan datamäärän kasvaessa ja muuttuessa monimuotoisemmaksi. Tutkimus toi siis esiin tietojohdamisen keskeistä roolia korkeatasoisen data-analytiikan hyödyntämisessä, johon tutkielmassa viitataan johdon datakyvykkyyksillä. Tämän vuoksi tutkimus myös vahvistaa ymmärrystä tietojohdamisen tarpeesta.

Tämän tutkimuksen tulokset kasvattavat ymmärrystä tietoperusteiseen päätöksentekoon liittyvistä tekijöistä, joihin kyvykkyyksillä voidaan vaikuttaa. Vaikka tämän tutkimuksen tulokset eivät pystyneet osoittamaan tietoperusteisen päätöksenteon suoria vaikutuksia menestykseen, voidaan aikaisempiin tutkimuksiin vedoten todeta yhteyden tietoperusteisen päätöksenteon ja arvonluonnin välillä olevan ilmeistä. Tulokset tietoperusteisen päätöksenteon, teknologisten kyvykkyyksien ja johdon datakyvykkyyksien välisistä yhteyksistä ovat kuitenkin erittäin arvokkaita yrityksille, jotka haluavat parantaa omaa tietoperusteisen päätöksenteon tasoaan.

Teknologiset kyvykkyydet toimivat tulosten perusteella kaikista tärkeimpänä linkkinä kaiken tasoisen datan ja päätöksenteon välillä. Yritysjohdon tulisi uskaltaa panostaa teknologioihin, joilla dataa kerätään, varastoidaan ja analysoidaan. Tässä tutkimuksessa ei otettu kantaa yhteen viitekehyksessä esiin nousseeseen tekijään, nimittäin laatuun. Datan ja siitä jalostetun tiedon laatu asettuvat kuitenkin keskeiseen rooliin varsinkin, kun data saavuttaa monimutkaisuudeltaan Big Datan tason. Laatu on myös otettava huomioon teknologisia panostuksia tehtäessä. Yritysten on tarkkaan arvioitava, miten dataa analysoidaan sekä mistä ja millaista dataa kerätään. Tähän puolestaan voidaan vaikuttaa parantamalla tietojohdamista johdon da-

takyvykkyksiä kasvattamalla. Käytännössä asioiden oikein tekemisen lisäksi on osattava tehdä oikeita asioita.

Tämän tutkimuksen tuloksille asettuu myös valitusta tutkimustavasta ja kontekstista johtuvia rajoitteita, jotka on otettava huomioon. Rajoitteet toimivat kuitenkin pohjana aiheeseen liittyville uusille tutkimuksille. Tässä tutkimuksessa empiirinen aineisto koostui yli 100 työntekijää työllistävistä suomalaisista yrityksistä. Tuloksia ei tämän vuoksi voida välttämättä yleistää maailmanlaajuiseen kontekstiin. Tutkimuksen tuloksia rajoittaa myös se, että kyselylomakkeessa yksittäinen vastaaja vastasi kaikkia osa-alueita koskeviin kysymyksiin. Tämä näkyi esimerkiksi tutkimuksen analysointivaiheessa aineiston tyhjinä vastauksina. Tämä tutkimus ei myöskään huomionnut sitä, että eri yrityksissä ja varsinkin eri toimialoilla tietoa tarvitaan ja hyödynnetään täysin eri tasoilla. Saadut tulokset saattaisivat siis tarkentua esimerkiksi tutkimalla toimialakohtaisia eroja.

Suomessa on lisäksi erittäin paljon tutkittua populaatiota pienempiä yrityksiä, jotka hyötyvät myös tietoperusteisesta päätöksenteosta. Tuloksia on vaikea yleistää pienempiin yrityksiin, joissa kyvykkyysien rooli saattaa olla hyvin erilainen. Tulevaisuudessa tätä tutkimusta vastaava tutkimus olisi mielenkiintoista toteuttaa esimerkiksi start-up -yrityksiin tai yleisesti pienempiin yrityksiin.

Kuten tässä tutkielmassa esitetään, tietoperusteinen päätöksenteko Suomessa sekä yleisesti ei ole nähnyt lähellekään huippuaan. Tämä tutkimus kuvaa vain aineiston keruun aikaista tilannetta käsitelystä aiheesta. Reliabiliteetin parantamiseksi asiaa tulisi tutkia uudelleen, jotta myös ajan myötä tapahtuva kehitys tulee huomioon otetuksi. Lopuksi huomioon tulee ottaa tutkimuksessa käytettyjen mittareiden subjektiivisuus. Tulosten validiteettia voitaisiin parantaa lisäämällä tutkimukseen enemmän objektiivisia mittareita.

Tämä tutkimus oli erittäin mielenkiintoista toteuttaa hyödyntäen Lappeenrannan teknillisen yliopiston ”Aineeton pääoma ja arvonluonti” -projektin aineistoa. Tässä tutkielmassa tutkittu aihe on erittäin ajankohtainen ja mielenkiintoinen. Tutkimusta tehdessä nousi esiin muutamia erittäin mielenkiintoisia jatkotutkimusaiheita. Kuten edellä on kuvattu, vastaavanlainen tutkimus olisi mielenkiintoista toteuttaa uudelleen ja käyttäen objektiivisempia mittareita, jotta esimerkiksi menestysmittari saavuttaisi paremman validiteetin ja reliabiliteetin. Tarkastelua olisi myös mielekästä kohdentaa pienempiin yrityksiin.

Edellä kuvattujen aiheiden lisäksi olisi erittäin mielenkiintoista tarkentaa tarkastelua esimerkiksi laadullisesti päätöksentekoon liittyviin elementteihin, joihin tiedolla voidaan vaikuttaa. Jatkotutkimusta olisi myös mielekästä toteuttaa tässä tutkielmassa tarkasteltujen kyvykkyyksien osalta. Teknologiset kyvykkyydet nousivat tuloksissa avainasemaan, joten niiden merkitystä olisi syytä tutkia vielä tarkemmin.

## LÄHTEET

AKTER, S., WAMBA, S.F., GUNASEKARAN, A., DUBEY RAMESHWAR and CHILDE, S.J., 2016. *How to improve firm performance using big data analytics capability and business strategy alignment?* .

BARNEY, J., 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management*, **17**(1), pp. 99-120.

BAYRAK, T., 2015. A Review of Business Analytics: A Business Enabler or Another Passing Fad. *Procedia - Social and Behavioral Sciences*, **195**, pp. 230-239.

BIZER, C., BONCZ, P., BRODIE, M. and ERLING, O., 2011. The Meaningful Use of Big Data: Four Perspectives – Four Challenges. *ACM SIGMOD Record*, **40**(4), pp. 56-60.

BROWN, B., CHUI, M. and MANYIKA, J., 2011. Are you ready for the era of 'big data'. *McKinsey Quarterly*, **4**, pp. 24-35.

CHAE, B.(., YANG, C., OLSON, D. and SHEU, C., 2014. The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. *Decision Support Systems*, **59**, pp. 119-126.

CHEN, H., CHIANG, R.H.L. and STOREY, V.C., 2012. Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, **36**(4), pp. 1165.

CHOO, C.W., 1996. The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, **16**(5), pp. 329-340.

CHOO, C.W. and JOHNSTON, R., 2004. Innovation in the knowing organization: a case study of an e-commerce initiative. *Journal of Knowledge Management*, **8**(6), pp. 77-92.

ECONOMIST INTELLIGENCE UNIT., 2012-last update, The deciding factor: Big data & decision making [Homepage of Economist Intelligence Unit.], [Online].

ELGENDY, N. and ELRAGAL, A., 2015. Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, **100**, pp. 1071-1084.

GANDOMI, A. and HAIDER, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, **35**(2), pp. 137-144.

GERMANN, F., LILIEN, G.L., FIEDLER, L. and KRAUS, M., 2014. Do retailers benefit from deploying customer analytics? *Journal of Retailing*, **90**, pp. 587-593.

- GRANT, R.M., 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal*, **17**(Winter special), pp. 109-122.
- HAIR, J.F., 2010. *Multivariate data analysis*. 7. ed., global ed. edn. Upper Saddle River ; Munich [u.a.]: Pearson.
- HAMILTON, B., 2015. From predictive to prescriptive analytics. *Strategic finance*, , pp. 62-63.
- HOOPER, D., COUGHLAN, J. and MULLEN, M.R., 2008. Structural Equation Modelling: Guidelines for Determining Model Fit.
- HU, L. and BENTLER, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**(1), pp. 1-55.
- JANSSEN, M., VAN DER VOORT, H. and WAHYUDI, A., 2017. Factors influencing big data decision-making quality. *Journal of business research*, **70**, pp. 338-345.
- JÖRESKOG, K.G. and SÖRBOM, D., 2001. *LISREL 8 user's reference guide*. 2. ed. edn. Lincolnwood: Scientific Software International.
- KANER, M. and KARNI, R., 2004. A Capability Maturity Model for Knowledge-Based Decisionmaking. *Information Knowledge Systems Management*, **3**(4), pp. 225-252.
- KELLER, K.L., 1987. Effects of quality and quantity of information on decision effectiveness. *Journal of consumer research*, **14**(2), pp. 200-213.
- KOŚCIELNIAK, H. and PUTO, A., 2015. BIG DATA in Decision Making Processes of Enterprises. *Procedia Computer Science*, **65**, pp. 1052-1058.
- KUBINA, M., VARMUS, M. and KUBINOVA, I., 2015. Use of Big Data for Competitive Advantage of Company. *Procedia Economics and Finance*, **26**, pp. 561-565.
- MOGES, H., VLASSELAER, V.V., LEMAHIEU, W. and BAESENS, B., 2016. Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes — An exploratory study. *Decision Support Systems*, **83**, pp. 32.
- NUTT, P.C., 1984. *Types of Organizational Decision Processes*.
- ÖZKÖSE, H., ARI, E.S. and GENCER, C., 2015. Yesterday, Today and Tomorrow of Big Data. *Procedia - Social and Behavioral Sciences*, **195**, pp. 1042-1050.

PIGNI, F., PICCOLI, G. and WATSON, R., 2016. Digital Data Streams: Creating Value from the Real-Time Flow of Big Data. *California Management Review*, **58**(3), pp. 5-25.

PODSAKOFF, P.M., MACKENZIE, S.B., LEE, J. and PODSAKOFF, N.P., 2003. Common Method Biases in Behavioral Research. *Journal of Applied Psychology*, **88**(5), pp. 879-903.

POWER, D.J., 2015 'Big Data' Decision Making Use Cases, 2015, Springer, pp. 1-9.

POWER, D.J., 2014. Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, **23**(2), pp. 222-228.

PROVOST, F. and FAWCETT, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, **1**(1), pp. 51-59.

PYNE, S., RAO, B. and RAO, S.B., 2016. *Big data analytics: methods and applications*. New Delhi: Springer.

RAGHUNATHAN, S., 1999. Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. *Decision Support Systems*, **26**(4), pp. 275-286.

SCHREIBER, J.B., NORA, A., STAGE, F.K., BARLOW, E.A. and KING, J., 2006. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, **99**(6), pp. 323-337.

SCHWARBER, P.D., 2005. Leaders and the decision-making process. *Management decision*, **43**, pp. 1086-1092.

SHARMA, S., MUKHERJEE, S., KUMAR, A. and DILLON, W.R., 2005. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, **58**(7), pp. 935-943.

SIMONS, H., 1957. *Models of Man: Social and Rational*. New York: John Wiley.

STEIGER, J.H., 2007. Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, **42**(5), pp. 893-898.

TYNAN, D., 2017. What is machine learning? Software derived from data. *InfoWorld.com*, .



WALLER, M.A. and FAWCETT, S.E., 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, **34**(2), pp. 77-84.

WAMBA, S.F., GUNASEKARAN, A., AKTER, S., REN, S.J., DUBEY, R. and CHILDE, S.J., 2017. Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, **70**, pp. 356-365.

ZIORA, A.C.L., 2015. The Role of Big Data Solutions in the Management of Organizations. Review of Selected Practical Examples. *Procedia Computer Science*, **65**, pp. 1006-1012.

## Liitteet

### Liite 1. Aineiston kuvaus

		FIRMPER F1	FIRMPER F2	FIRMPER F3	FIRMPER F4	STRATKM 1	STRATKM 2	STRATKM 3	STRATKM 4	STRATKM 5	ANADAT1
N	Valid	178	178	177	179	178	179	174	179	179	174
	Missing	1	1	2	0	1	0	5	0	0	5
Mean		5,03	4,69	4,81	5,46	5,35	5,51	4,70	4,41	4,41	5,31
Median		5,00	5,00	5,00	6,00	5,00	6,00	5,00	5,00	4,00	6,00
Skewness		-0,917	-0,243	-0,327	-0,953	-0,736	-1,137	-0,561	-0,423	-0,148	-0,747
Std. Error of		0,182	0,182	0,183	0,182	0,182	0,182	0,184	0,182	0,182	0,184
Kurtosis		0,826	-0,365	0,336	1,436	0,844	2,186	-0,037	-0,056	-0,836	0,742
Std. Error of Kurtosis		0,362	0,362	0,363	0,361	0,362	0,361	0,366	0,361	0,361	0,366

		ANATECH 1	ANATECH 2	ANATECH 3	ANATECH 4	ANATECH 5	ANACULT 1	ANACULT 2	ANACULT 3	ANACULT 4	ANACULT 5
N	Valid	177	174	170	120	167	179	178	169	178	174
	Missing	2	5	9	59	12	0	1	10	1	5
Mean		4,57	4,27	4,39	3,10	4,66	5,47	5,25	4,75	5,07	4,36
Median		5,00	4,00	5,00	3,00	5,00	6,00	5,00	5,00	5,00	4,00
Skewness		-0,388	-0,200	-0,339	0,501	-0,368	-0,776	-0,764	-0,413	-0,633	-0,461
Std. Error of		0,183	0,184	0,186	0,221	0,188	0,182	0,182	0,187	0,182	0,184
Kurtosis		-0,489	-0,636	-0,612	-0,373	-0,410	0,940	0,463	-0,632	0,321	-0,246
Std. Error of Kurtosis		0,363	0,366	0,370	0,438	0,374	0,361	0,362	0,371	0,362	0,366

## Liite 2. EFA faktoreiden lataukset ja korrelaatiot

<b>Pattern Matrix<sup>a</sup></b>				
	Factor			
	1	2	3	4
STRATKM2	0,847			
STRATKM1	0,778			
STRATKM4	0,704			
STRATKM3	0,660			
STRATKM5	0,510			
ANATECH1		0,873		
ANATECH2		0,846		
ANATECH5		0,703		
ANATECH3		0,640		
ANACULT3			0,826	
ANACULT4			0,822	
ANACULT2			0,659	
FIRMPERF2				0,795
FIRMPERF1				0,702
FIRMPERF3				0,342

Extraction Method: Maximum Likelihood.  
a. Rotation converged in 5 iterations.

<b>Factor Correlation Matrix</b>				
Factor	1	2	3	4
1	1,000	0,423	0,480	0,124
2	0,423	1,000	0,636	0,127
3	0,480	0,636	1,000	0,141
4	0,124	0,127	0,141	1,000

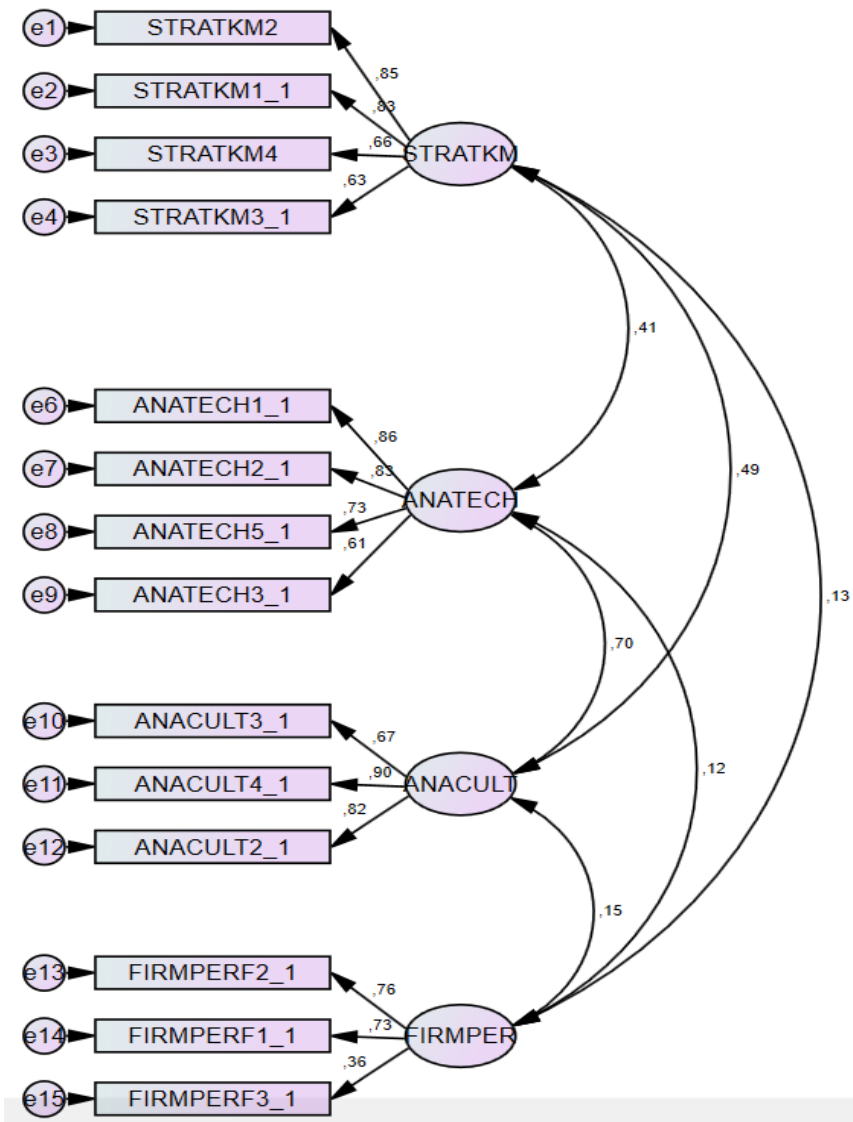
Extraction Method: Maximum Likelihood.  
Rotation Method: Promax with Kaiser Normalization.

## Liite 3. Kommunaliteetit

<b>Kommunaliteetti</b>		
	Initial	Extraction
ANATECH1	0,650	0,751
ANATECH2	0,637	0,691
ANATECH3	0,391	0,386
ANATECH5	0,533	0,551
STRATKM4	0,499	0,486
STRATKM5	0,400	0,326
STRATKM1	0,598	0,638
STRATKM3	0,428	0,431
FIRMPERF1	0,366	0,505
FIRMPERF2	0,387	0,638
FIRMPERF3	0,194	0,131
ANACULT2	0,635	0,655
ANACULT3	0,463	0,540
ANACULT4	0,673	0,805
STRATKM2	0,614	0,695

Extraction Method: Maximum

Liite 4. CFA malli



## Liite 5. Multikollinearisuus

Coefficients <sup>a</sup>								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3,204	0,419		7,643	0,000		
	ANACULT	0,144	0,141	0,126	1,024	0,307	0,361	2,770
	ANATECH	0,017	0,099	0,020	0,173	0,863	0,404	2,476
	STRATKM	0,076	0,088	0,077	0,867	0,387	0,692	1,444

a. Dependent Variable: FIRMPER