



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

DANIEL JOKELA
UUSIEN MENETELMIEN HYÖDYNTÄMINEN TEKNOLOGIADIF-
FUUSION ENNUSTAMISESSA
Diplomityö

Tarkastajat:
professori Matti Sommarberg ja yli-
opistonlehtori Juha T. Tantt
Tarkastaja ja aihe hyväksytty
26. maaliskuuta 2018

TIIVISTELMÄ

DANIEL JOKELA: Uusien menetelmien hyödyntäminen teknologiadiffuusion ennustamisessa

Tampereen teknillinen yliopisto

Diplomityö, 123 sivua, 26 liitesivua

Marraskuu 2018

Tuotantotalouden diplomi-insinöörin tutkinto-ohjelma

Pääaine: Tuotantotalous

Tarkastajat: professori Matti Sommarberg ja yliopistonlehtori Juha T. Tantt

Avainsanat: Teknologiajohtaminen, teknologiadiffuusio

Uusien teknologioiden ja teknologiadiffuusion kehityksen ennustaminen varhaisessa vaiheessa sisältää epävarmuustekijöitä. Investointipäätös sisältää riskin ja päätöksen ajoitus ilman tietoa perustuu usein kokemuksiin tai mielipiteisiin.

Työssä tutkittiin teknologiaennustamisen epävarmuutta ja pyrittiin löytämään ratkaisuja ennustamistarkkuuden parantamiseen sekä parempaan investointipäätökseen ja -ajoitukseen. Tämän diplomityön tarkoituksena oli selvittää huomioitavia asioita teknologioiden ennustamisessa ja vaatimuksia teknologiadiffuusion mitattaville asioille sekä löytää teknologiadiffuusion kehittymisen ennustamiseen sopivia menetelmiä.

Tutkimusmenetelmänä oli kirjallisuustutkimus sekä kvantitatiivinen tutkimus eri aineistoihin perustuvasta teknologiadiffuusion kehittymisen ennustamisesta. Kirjallisuustutkimuksessa koottiin orastavien teknologioiden kehittymisen ennustamiseksi sopivia lähteitä, ennustusmalleja sekä teoriaa. Kirjallisuustutkimuksen perusteella valittiin ennustettavalle kohteelle sopivat lähteet, laskentamenetelmä ja -malli. Kirjallisuustutkimuksen tuloksia sovellettiin kvantitatiivisessa tutkimuksessa. Kvantitatiivisessa osuudessa toteutettiin teknologiadiffuusion kehittymisennuste tutkimuksen kohteena olevaan ilmiöön.

Tutkimustuloksena kehitettiin orastavien teknologioiden teknologiadiffuusion kehittymiseen sopiva ennustemenetelmä. Menetelmässä integroidaan useita eri ennustuslähteitä sekä hyödynnetään Bassin diffuusiolaskentamallia. Laskentamallissa integroidaan eri ennustuslähteiden datan käyttäytymistä kuvaavat parametrit Bassin laskentamallin keskeisiin laskentaparametreihin. Menetelmä sopii orastavien teknologioiden kehittymisen arviointiin. Integroidulla Bassin laskentamallilla voidaan arvioida teknologian kehittymistä tai taantumista sekä arvioida oikea-aikaisia liiketoimintapäätöksiä ennen merkittävää teknologiadiffuusiota.

ABSTRACT

Daniel Jokela: Using the new methods in forecasting of technology diffusion
Tampere University of Technology
Master of Science Thesis, 123 pages, 26 Appendix pages
November 2018
Master's Degree Programme in industrial Management
Major: Industrial management
Examiners: Professor Matti Sommarberg, university lecturer Juha T. Tantt

Keywords: Technology forecasting, technology diffusion

Forecasting the development of new technologies and technology diffusion contains uncertainties. Making an investment decision includes risks and timing the decision without information is usually based on former experiences or mere opinions.

This study examines the uncertainty in technical forecasting, and seeks to find solutions to improve forecasting accuracy in order to time investment decisions better.

The purpose of this thesis is to investigate the issues to consider when predicting technologies and the requirements for measuring technology diffusion, and to find suitable methods for predicting the development of technology diffusion.

As research methods, a literature research and a quantitative study based on various sources were conducted on predicting the development of technology diffusion. In the literature research, appropriate sources, prediction models and theory were compiled to predict the development of emerging technologies. A suitable calculation method and calculation model were selected for prediction.

The results from the literature research were applied in a quantitative study. In quantitative terms, the developmental trend of technology diffusion was implemented in predicting a phenomenon that was subject of the study.

As a result of this study, a suitable forecasting method for the development of the technology diffusion of emerging technologies was developed. The method integrates several different prediction sources with the Bass diffusion model. The calculation model integrates the parameters of the behavior of data of different prediction sources into the key calculation parameters of the Bass calculation model.

This method is suitable for evaluating the development of emerging technologies. The integrated Bass computation model can be used to evaluate technology development or decline, and to evaluate timely business decisions before major technology diffusion.

ALKUSANAT

Opinnäytetyö on tehty Tampereen teknilliselle yliopistolle. Diplomityössä selvitetään mahdollisuuksia orastavan teknologian tarkempaan diffuusioennustukseen. Haluan kiittää ohjaajina toimineita professori Matti Sommarbergia ja yliopistonlehtori Juha Tanttua työtä edistävästä keskusteluista ja rakentavista kommentteista. Mielipiteenne ovat olleet erittäin arvokkaita ja johtaneet työn aikana esille tulleiden haasteiden kohtaamiseen.

Lämmin kiitos kuuluu myös useille Tampereen teknillisen yliopiston professoreille, yliopistonopettajille ja muulle henkilökunnalle työssä tarvittavien taitojen saavuttamisessa. Ystäviä ja perhettä kiitän tuesta ja aiheeseen liittyvistä keskusteluista. Mielipiteenne ovat olleet erityisen arvokkaita opiskelujen eri vaiheissa.

Helsingissä, 19.11.2018

Daniel Jokela

SISÄLLYSLUETTELO

1.	JOHDANTO	1
	Diplomityön tavoitteet.....	2
	Tutkimusasetelma	2
	Tutkimuksen rakenne	4
2.	TAUSTA TEKNOGIAN ENNUSTAMISESSA.....	6
	2.1 Tekninen tausta teknologiadiffuusion ennustamisen menetelmissä	6
	2.2 Teoreettisia taustoja teknologiadiffuusion ennustamisen menetelmissä.....	11
	2.3 Tekoälyteknologioiden riippuvuus big datasta ja datatieteistä	15
	2.4 Innovaatio- ja teknologiadiffuusiomalli	20
	2.5 Diffuusion kehittymisen ennakkoinnin tietolähteiden ajantasaisuus	30
	2.6 Big data -ennustusmenetelmien tarkempi tarkastelu.....	37
	2.7 Teknologian kehittymisen aaltoliikkeet	39
	2.8 Uusien teknologioiden kehittymisen nopeutuminen	45
3.	TEKNOLOGIADIFFUUSION KASVUN ENNUSTAMINEN	50
	3.1 Teknologiadiffuusion ennustuksen menetelmävalinta	50
	3.2 Teknologianennustamisen aineistot kvantitatiivisessa tutkimuksessa	51
	3.3 Teknologiaennustamisen tietolähteiden valinta	58
	3.4 Teknologian ennustaminen muilla menetelmillä	59
	3.5 Big data -teknologian tunnistaminen	60
4.	TUTKIMUSMENETELMÄT JA AINEISTO.....	63
	4.1 Kirjallinen tutkimusmenetelmä ja aineisto.....	63
	4.2 Kvantitatiivinen tutkimusmenetelmä ja aineisto	67
	4.3 Patenttitietokantahaut ja aineistojen käsittely	68
5.	TULOKSET JA POHDINTA	70
	5.1 Kirjallisuustutkimuksen tuloksia.....	70
	5.2 Kvantitatiivinen tutkimus big data -ilmiön kehitymisestä	75
	5.3 Arvio tulosten luotettavuudelle otoksien perusteella patenttitietokannasta .	92
	5.4 Laskenta hypergeometrisella jakaumalla	95
	5.5 Patenttitietokannan aineistolliset ominaisuudet ja niiden haasteet	98
	5.6 Yhteenveto patenttitietokantatutkimuksen aineistosta	107
	5.7 Saavutettujen tulosten ja niiden merkityksen esittely	108
6.	YHTEENVETO JA JATKOTUTKIMUKSIEN AIHEET	112
	LÄHTEET.....	115

LIITE A: TEKNOLOGIAN KEHITTÄMISEN ENNUSTUSMALLEJA

LIITE B: OTOS GOOGLE TRENDS -PALVELUSTA

LIITE C: SOVELLUS KRIITTISEN MASSAPISTEEN ESITTÄMISTÄ VARTEN (P JA Q-PARAMETREILLA)

LIITE D: SOVELLUS KRIITTISEN MASSAPISTEEN LASKENTAA VARTEN, JOSSA RATKAISTAAN P JA Q-PARAMETRIIT

LIITE E: CHO JA DAIM (2016) TUTKIMUSTULOKSIEN MALLINTAMINEN MATLAB-OHJELMISTOLLA HYÖDYNTÄEN BASS (1969) LASKENTAMALLIA

LIITE F: MADDISON PROJEKTISTA LASKETUT INFLAATIOKORJAUKSET JA SUHTEELLISEN VAIHTELUN LASKENTATULOKSET

LIITE G: LASKENTA, JOLLA LASKETAAN PATENTTITIEKANNAN OTOKSEN POIMINTAVIRHEIDEN ESIINTYVYYDEN TODENNÄKÖISYYSJAKAUMA

LIITE H: PATENTTITIEKANNAN OTOKSEN POIMINTAVIRHEIDEN ESIINTYVYYDEN TODENNÄKÖISYYSJAKAUMAN KERTYMÄFUNKTIO

LIITE I: HYPERGEOMETRISEN JAKAUMAN LASKENTA 4% VIRHEARVIOLE

LIITE J: EXCEL -OHJELMISTON HYÖDYNTÄMINEN JA TYÖVAIHEET PATENTTITIEKANNAN TIETOJEN KÄSITTELYSSÄ

LIITE K: PATENTTIEN KÄSITTELYAIKOJEN KESKIMÄÄRÄINEN VIIVE AJAN FUNKTIONA

LIITE L: PATENTTIEN KÄSITTELYAIKAA KUVAAVA HISTOGRAMMI

LIITE L: MATLAB-KOODI JOUKON A JA B LASKENTAAN JA KUVAAMISEEN

LIITE N: MATLAB-KOODI, JOLLA LASKETAAN JA ESITETÄÄN LAATIKKOJANA KUVAAJA JOUKOSTA, JOILLA PATENTTIEN KÄSITTELYAIKA ON VÄHINTÄÄN 400 PÄIVÄÄ

LIITE O: ESPACENET -TIEKANNASSA ESIINTYNEET IPC-KOODIT BIG DATAAN LIITTYEN KK \ LKM. (08/2009-2/2018)

LIITE P: PYLVÄSKUVAAJAESITYS IPC-KOODILLISISTA PATENTTIEN MÄÄRÄLLISISTÄ OSUUKSISTA (KPL) PATENTEISTA KODIEN LUKUMÄÄRÄN MUKAAN

LIITE Q: MATLAB-OHJELMA EPÄLINEAARISEN SYSTEEMIN PARAMETRIEN SELVITTÄMISEEN

LIITE R: BASSIN INNOVAATIODIFFUUSIO LASKENTA

LYHENTEET JA MERKINNÄT

BDA	Big data analytiikka (Big data analytics)
Big Data 3v's	Big dataa määrittelevät ominaisuudet (Volume, Variety ja Versality)
BKT	Bruttokansantuote
CMP	Kriittinen massapiste eli kriittinen käyttäjien määrä teknologia-diffuusiassa (Critical Mass Point)
CPC	Patenttiluokitus (Cooperative Patent Classification, EPO & USPTO)
CPTO	Kiinan patenttitoimisto (China Patent and Trademark Office)
Diffuusio	Teknologian kehittymisen vaihe, jossa teknologiaa hyödynnetään. Diffuusiassa käytön tai hyödyntämisen määrä on muutoksessa
Disruptiivinen	Uusi teknologia voi olla disruptiivinen, jolloin se tuotantotavalla tai menetelmällä muuttaa tuottavuutta tai lopputuotetta tehden aikaisemman teknologian tarpeettomaksi
EPO	Euroopan patenttitoimisto (European patent office)
GAN	Tekoälyn neuroverkkoratkaisu (Generative adversarial network)
HMSAM	Adaptiomalli (Hedonic-motivation system adoption model)
IDT	Innovaatiodiffuusioteoria (Innovation diffusion theory)
Innovaatio	Tehdä asioita uudella tavalla ja soveltaa keksintöä
IPC	Kansainvälinen patenttiluokitus (International Patent Classification)
JPO	Japanin patenttitoimisto (Japan Patent Office)
KIPO	Korean patenttitoimisto (Korean Intellectual Property Office)
K-means	Tekoälyssä hyödynnettävä hierarkkinen klusterointi
LED	Valo ja mm. näyttöteknologia (Light-emitting diode)
MACBETH	Käyttäytymismallinnuksessa hyödynnettävä mittausmenetelmä (Measuring Attractiveness by a Categorical Based Evaluation Technique -methodology)
MCDM	Monimuuttujiin perustuva päätöskriteeristö (Multiple-criteria decision-making)
NSF	Yhdysvaltojen kansallinen tiedesäätiö (National Science Foundation)
OECD	Taloudellisen yhteistyön ja kehityksen järjestö (The Organisation for Economic Co-operation and Development)
OLED	Näyttöteknologia (An organic light-emitting diode)
PCT	Kansainvälinen patenttijärjestelmä (Patent Cooperation Treaty)
PESTEL	Ulkoista toimintaympäristöä analysoiva ja mittaava menetelmä, joka huomio poliittiset, ekonomiset, sosiaaliset, teknologiset, ekologiset ja lailliset tekijät.
RBM	Tekoälyn neuroverkkoratkaisu (Restricted Boltzmann machine)
RGDP	Reaalibruttokansantalouden tuotto (Real gross domestic product)
SRMO	Tutkimusorganisaatio (SAGE Research Methods Online)
SVM	Koneoppimiseen liittyvä tukivektorikone (Support vector machine)
TAM	Käyttäytymisteoria (Technology acceptance model)
TPB	Käyttäytymisteoria (Theory of planned behavior: Ajzen)
TRA	Käyttäytymisteoria (Theory of reasoned action)
USPTO	Patenttitoimisto (United States Patent and Trademark Office)

1. JOHDANTO

Tulevia teknologioita voidaan ennustaa eri ilmiöiden aikaisemman yleistymisen ja teknologian leviämisen perusteella. Teknologian kehittyminen voidaan jakaa *keksintö – innovaatio – diffuusio* -vaiheisiin, joista diffuusiovaihe kuvaa teknologian yleistymistä ja leviämistä (Schumpeter 1934; ref. Jamasb & Kohler 2007). Godin (2010) kirjoitti myös William F. Ogburnin (1886-1959) tutkineen teknologiainnovaatiota ja -diffuusiota. Godin totesi Ogburnin esittäneen innovaatioteorian jakautuvan *keksintö ja sen kehittäminen – diffuusio ja käyttö – innovaation vaikutukset* -vaiheisiin (Ogburn 1937; ref. Godin 2010). Tulevista teknologioista on usein saatavilla rajallisesti tietoa. Teknologian käyttöönottopäätös sisältää epävarmuuden ja eri tekijöistä johtuvia mahdollisuuksia ja riskejä. Kilpailijoita tarkempi tulevien ja hiipuvien teknologioiden ennustaminen voi tuoda strategista kilpailuetua. Innovaatiodiffuusioteorian (IDT) varhaisina vaikuttajina voidaan pitää myös Gabriel Tardea. Tarde esitti teknologian leviämisen määrää kuvaavan S-muotoisen diffuusiokäyrän jo vuonna 1903 (Kaminski 2011).

Diplomityö asemoituu uusien teknologioiden ennustamiseen, teknologidiffuusion varhaiseen arviointiin ja investointipäätösten ajoittamiseen. Työssä käsitellään teknologiaennustamisen epävarmuutta ja pyritään löytämään ratkaisuja ennustamistarkkuuden parantamiseen. Tuloksissa esitetään eri ennustamismenetelmien ja lähdeaineistojen eroja. Oikeilla valinnoilla saavutetaan tarkempi ennustus tulevista teknologioista. Työssä huomioidaan uusien teknologiaennustamismenetelmien ja tietolähteiden hyödyntäminen sekä niiden yhdistäminen teknologiadiffuusion ennustamiseen. Työssä verrataan eri tietolähteistä johdettuja teknologiadiffuusioennusteita ja teknologian kehittymisen eri vaiheita.

Tutkimusongelmana on teknologiadiffuusion ennustamisen epävarmuus. Uusien teknologioiden diffuusio voi olla nopeaa, kehittyä alkuvaiheessa erittäin hitaasti tai teknologia voi korvautua toisella ilman näkyvää syytä. Päätetty teknologiavalinta voi myöhemmin osoittautua virhearvioinniksi. Investointipäätöksestä voi kulua pitkä aika tuotantokäyttöön ja ennustuksen tulisi olla riittävän luotettava, jotta investointi on strategisesti kannattava.

Diplomityön tavoitteet

Tässä diplomityössä on keskeistä löytää ratkaisuja uuden ja nopeasti kehittyvän ilmiön tai teknologian ennustamiseen. Työn tavoitteena on vastata asetettuun tutkimusongelmaan sekä tutkimuskysymyksiin. Tutkimuskysymykset, joihin tässä työssä etsitään vastausta ovat

- 1) Mitä asioita tulisi huomioida teknologioiden ennustamisessa?
- 2) Mitä vaatimuksia teknologiadiffuusion mitattaville asioille tulee asettaa?
- 3) Miten teknologiadiffuusion menetelmävalinnat vaikuttavat ennustustarkkuuteen?

Tutkimusasetelma

Ensimmäiseen tutkimuskysymykseen: ”Mitä asioita tulisi huomioida teknologioiden ennustamisessa?” vastataan käsittelemällä teorioita, aineistovalintoja, ennustamisen menetelmiä, teknologiadiffuusion ja teknologian hyväksymisen taustoja. Tutkimusmenetelmänä on kirjallisuustutkimus sekä valittuihin teknologioihin kohdistuva kvantitatiivinen tutkimus. Kvantitatiivisessa tutkimuksessa poimitaan otoksia muun muassa talous-, web-haku-, artikkeli- ja patenttitietokannoista. Kvantitatiivisia tuloksia reflektoidaan kirjallisuustutkimuksen tuloksiin.

Toiseen tutkimuskysymykseen: ”Mitä vaatimuksia teknologiadiffuusion mitattaville asioille tulee asettaa?” vastataan kirjallisessa ja kvantitatiivisessa tutkimuksessa. Kirjallisuustutkimuksen avulla etsitään uusien teknologioiden ennustamiseen soveltuvia tietolähteitä ja menetelmiä. Kirjallisuudessa esitetty tietolähteiden kronologinen järjestys koostuu eri tutkimuksista. Kvantitatiivisessa tutkimuksessa valitaan tietolähteet, mitattavat asiat, mittaustapa ja -menetelmä. Kvantitatiivisen tutkimuksen tuloksissa kuvataan teknologian kehittymisen vaihe ja arvioidaan teknologian kehittyminen.

Kolmannen tutkimuskysymyksen: ”Miten teknologiadiffuusion menetelmävalinnat vaikuttavat ennustustarkkuuteen?” sisältöä ja vastausta käsitellään huomioiden aineistovalintoja. Teknologian ennustamisessa voidaan hyödyntää mm. markkinadataa, rahoitus-, web-haku-, patenti-, artikkeli-, julkaisu-, ja konferenssijulkaisutietokantoja. Tutkimuskohteeksi rajataan big data -ilmiö, jota moni teknologia voi hyödyntää. Big data -ilmiön kehittyminen riippuu liitännäisteknologioiden kehittymisestä. Liitännäisteknologioita ovat datankäsittely ja -varastointiteknologia sekä mm. tietoliikenneteknologiat. Teknologioiden kehittymisen vuorovaikutusta verrataan liitännäisteknologioiden kehittymiseen. Kirjallisuustutkimustulosten sekä eri tietokannoista saatavien tietojen pohjalta tehdään arvio ilmiön kehittymisestä sekä sen mahdollistamien teknologioiden kehittymisestä.

Kvantitatiivisessa tutkimuksessa on tehty rajaus, jossa keskitytään diffuusiomallien mukaiseen tulosten esittämiseen. Tässä työssä koostetaan ja hyödynnetään eri tutkimuksissa

ennustukseen käytettyjä tietolähteitä. Kirjallisuustutkimuksen avulla löydettyjä tietolähteitä järjestetään ja arvioidaan niiden sopivuus ennustamiseen. Patenttitietokantaa koskevassa osuudessa osoitetaan aineistojen ongelmia. Patenttitietokantaa koskevan osuuden työvaiheet on esitetty taulukossa 1. Aineisto-ongelmien avulla tuodaan esille aineistojen valintojen sekä tulkinnan vaikutus teknologiadiffuusion ennustukseen. Kvantitatiivisten aineistojen laatuongelmien osalta tämä työ keskittyy patenttitietokannasta saatujen otoksien ominaisuuksiin.

Taulukko 1. Patenttitietokantaa koskevan osuuden työvaiheet tutkimuksessa.

Vaihe	Tehtävä
1.	Asetetaan hakukriteerit tietolähteelle.
2.	Haetaan tiedot asetettujen kriteerien perusteella patenttitietokannasta.
3.	Luokitellaan kohteet osa-alueittain.
4.	Tarkastetaan hakutulokset.
5.	Selvitetään patenttien määrä kuukausittain.
6.	Kuvataan teknologian kasvukäyrät elinkaarelle (kumulatiivinen määrä).
7.	Lasketaan/mallinnetaan teknologian diffuusionopeus.
8.	Selvitetään patenttikoodien kokonaismäärä kuukausittain.
9.	Arvioidaan patenttikoodien käyttökelpoisuutta ennustamisessa.
10.	Lasketaan kokonaismäärä eri patenttikooduille.
11.	Lasketaan kasvumahdollisuudet.
12.	Toistetaan otanta ja verrataan tuloksia aikaisempaan otokseen.
13.	Toistetaan otanta ja verrataan tuloksia kahteen aikaisempaan otokseen.
14.	Arvioidaan syitä otoksista saatuihin eroihin ja esitetään tulokset eri otoksista.
15.	Suodatetaan data ja oikaistaan tulokset.
16.	Lasketaan datan käyttäytymiseen perustuvat innovaattori sekä seuraajatekijät osuu-
17.	Toteutetaan teknologiadiffuusionennuste sopivalla laskentamallilla.
18.	Lasketaan investointipäätöksen ennusteeseen perustuva optimiajoitus.

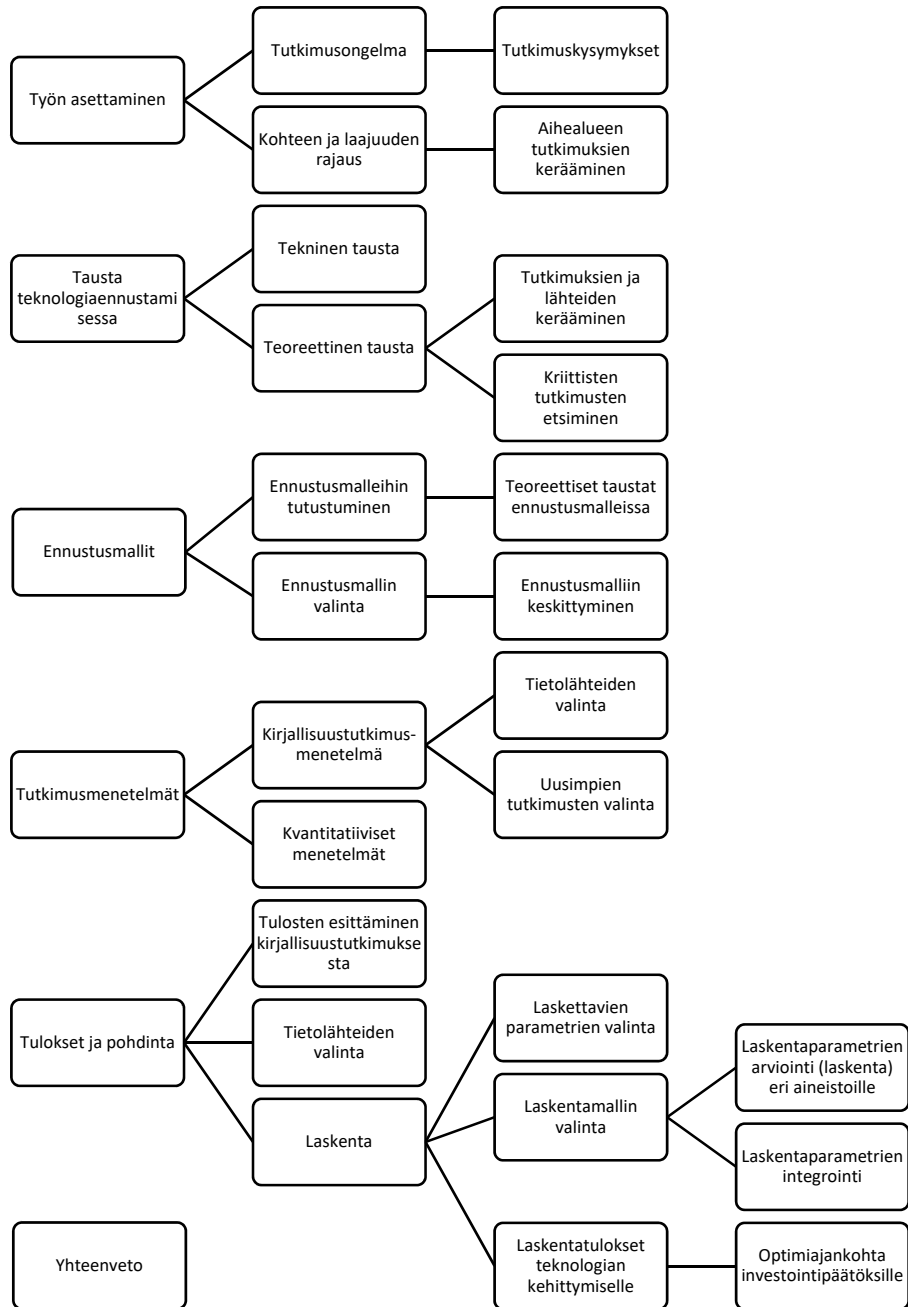
Kirjallisuudesta esitettyjä tietolähteitä kootaan ja valittuihin tietolähteisiin kohdistetaan hakuja. Tällaisia tietolähteitä ovat esimerkiksi konferenssi-, rahoitus-, markkinointi-, web-haku-, artikkeli-, talous-, ja patenttitietokannat. Tietolähteiden kronologinen ajantasaisuus pyritään selvittämään hakutuloksien perusteella. Tietolähteiden käyttökelpoisuutta arvioidaan tässä työssä tarkoitettussa kontekstissa. Tulokset eri tietolähteistä pyritään yhteismitallistamaan ja järjestämään kronologisesti. Kronologisella järjestämisellä pyritään osoittamaan tietolähteen valinnan riippuvuus käyttötarkoituksesta ja ajoituksesta.

Tutkimuksen rakenne

Työn ensimmäisessä vaiheessa tehdään kirjallisuustutkimus. Kirjallisuustutkimuksessa tuodaan esille *teknologiadiffuusion vaikuttavia asioita, joita tulee huomioida teknologioiden ennustamisessa*. Kirjallisuustutkimuksessa käydään läpi mm. ulkoisia ympäristötekijöitä, tietolähteitä ja niiden ajantasaisuutta sekä teknologian tuntemisen tarpeita. Kirjallisuuden avulla pyritään kuvaamaan teknologian kehittymiseen vaikuttavia asioita, jotka vaikuttavat ennustettavuuteen. Kirjallisuustutkimuksessa käydään läpi *vaatimuksia teknologiadiffuusiosta mitattaville asioille*. Asioiden huomioiminen ja niiden mittaaminen edellyttää etukäteen tehtäviä valintoja ja rajouksia mittauskohteista. Nämä valinnat ja rajoukset vaikuttavat teknologiadiffuusion mallintamiseen. Oikeilla mitattavien asioiden valinnoilla ja vaatimuksilla mahdollistetaan kehittyvän teknologiadiffuusion mallintaminen. Tutkimusongelman kolmannessa tutkimuskysymyksessä, ”*miten teknologiadiffuusion menetelmävalinnat vaikuttavat ennustustarkkuuteen?*” vastataan kirjallisuustutkimuksessa. Kirjallisuustutkimuksessa esitetyt tietolähteet pyritään järjestämään kronologisesti. Kirjallisuuden perusteella valitaan kvantitatiivisessa tutkimuksessa käytettävät tietolähteet. Kvantitatiivisessa tutkimuksessa saatujen tuloksia verrataan kirjallisuudessa esitettyihin tuloksiin.

Diplomityössä esitellään joitakin teknologian ennustamismalleja ja valitaan teknologian ennustamislähteet ja -malli. Haku kohdistetaan big data -ilmiöön sanahakuina, johon liittyviä internethakumääriä, artikkeleita, konferenssijulkaisuja ja patenteja kerätään satunnaisotoksiin. Saatua hakutuloksien ajallista ja laadullista vertailukelpoisuutta selvitetään kvantitatiivisin menetelmin. Tuloksien perusteella mallinnetaan big data -ilmiön esiintymien diffuusion kehittymistä eri tietolähteiden ja tekijöiden perusteella. Ennusteessa esitetään liitännäis- ja osateknologioiden kehittymisen vuorovaikutus varsinaisen ratkaisun kehittymisen kanssa. Yhteenvedossa esitetään arvio eri tekijöiden vaikutuksesta ennustukseen. Työn tuloksena esitetään valintoja ja eri menetelmiä teknologiaennustamiseen. Ennusteen tekeminen teknologiadiffuusion varhaisessa vaiheessa edellyttää onnistuneita lähde-, menetelmä- ja ennustusmallivalintoja.

Työn rakenne pyritään säilyttämään alkuperäisen diplomityöohjeen mukaelmana. Kuvassa 1 esitetään rakenteellisesti tämän diplomityön työvaiheita.



Kuva 1. Diplomityön työvaiheet kronologisessa järjestyksessä (ylhäältä alas).

Diplomityön työvaiheet on esitetty kuvassa 1 yleisellä tarkkuudella. Työvaiheet koostuvat kuudesta kokonaisuudesta.

2. TAUSTA TEKNOLOGIAN ENNUSTAMISESSA

2.1 Tekninen tausta teknologiadiffuusion ennustamisen menetelmissä

Ennustaminen voi perustua kvalitatiiviseen tai kvantitatiiviseen tutkimukseen. Caplice ja Sheffi (2006) esittivät, että rajaamalla ennustaminen kvantitatiiviseksi, rajataan samalla muuttujien määrä rajoitettuun joukkoon. He toteavat, että rajattu tutkimusjoukko vastaa harvoin todellisuutta. Ennustamisessa on etua, että hyödynnetään useita menetelmiä ja mahdollisimman monia tekijöitä yhtäaikaaisesti (Caplice & Sheffi 2006). Tässä työssä tehdään rajauksia menetelmiin, mutta säilytetään laaja-alainen näkökulma. Kvantitatiivisia menetelmiä hyödynnetään aineistojen keräämisessä, käsittelyssä ja analysoinnissa. Aineistojen rivimäärän vuoksi koko aineiston manuaalinen käsittely ilman kvantitatiivisia menetelmiä ei ole mielekäs ($n > 10\,000$). Eri menetelmien avulla voidaan arvioida aineiston laatua, vaikka arvioitava populaatio olisi suuri. Tilastollinen näkökulma tuodaan myös esille, koska kvantitatiivinen tutkimusosuus perustuu hakutuloksista koostettujen satunnaisotoksiin. Mittaustapa ja mittauksien tulkinta sekä monet muut tekijät aiheuttavat sen, että tulokset eivät ole tarkkoja. Tuloksista lasketaan sekä esitetään erilaisia tilastollisia tunnuslukuja.

Jyväskylän yliopisto tuo esille, että SAGE Research Methods Online (SRMO) -tietokantaa voi hyödyntää tutkimusmenetelmälahteenä (Jyväskylän yliopisto 2015). SRMO (2018) palvelussa esitetään, että tutkimusmetodeina voidaan käyttää datan keräämistä ja kvantitatiivista datan analysointia. Palvelussa esitetään kvantitatiivisen tutkimuksen mahdollisuudesta hyödyntää useita menetelmiä. Näytteenotossa huomioidaan esimerkiksi satunnaisotanta, populaatio ja parametrisissa menetelmissä muun muassa normaalijakauma, populaation keskiarvo sekä varianssi. Kokeellisissa aineistoanalyysissä ja data-analyysissä voidaan tutkia tuloksia esimerkiksi esittämällä:

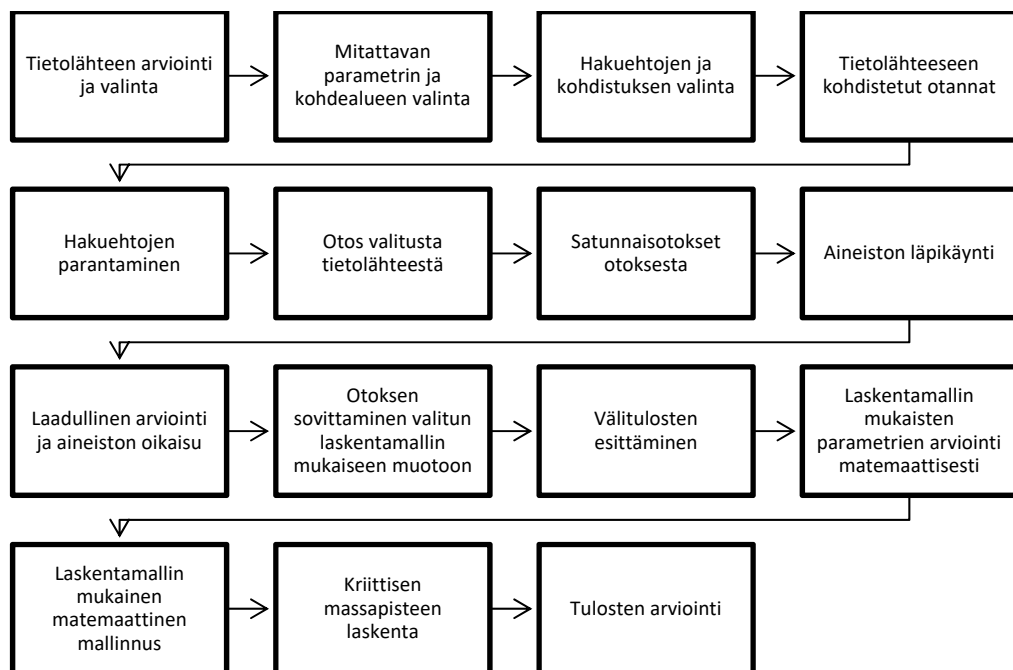
- laatikkojanakuvaajia
- datan visualisointia (mm. pylväskuvaaja ja histogrammi)
- mittauksia sekä mittausvirheitä
- poikkeavia havaintoja (outliers).

(SRMO 2018)

Tässä hyödynnetään SRMO (2018) esittämiä tutkimusmenetelmiä. Ennustaminen perustuu valittujen muuttujien mittaamiseen satunnaisotoksilla. Ennustamisessa keskeinen laskennallinen komponentti on teknologian esiintyvyys kussakin valitussa lähteessä. Ennustamisessa on olennaista pyrkiä esittämään teknologian esiintyvyyden kehittyminen ajan funktiona, joka jatkuu ajallisesti tulevaisuuteen. Tuloksia kuvataan esimerkiksi

- histogrammilla (jakaumien ja osajoukkojen tunnistaminen)
- laatikkojanakuvaajalla (mediaanin ja hajonnan kuvaaminen)
- vertailemalla tuloksia normaalijakaumaan
- arvioimalla tuloksien virherajoja ja -tasoja
- vertaamalla eri otoksia
- esittämällä virheitä
- teknologiadiffuusion S-muotoisilla käyrillä (kumuloituva summa ajan funktiona)
- ennustamalla teknologian kehittymistä S-muotoisilla käyrillä ajan funktiona.

Kuvassa 2 on esitetty käytetty kvantitatiivisen tutkimuksen yleinen periaate riippumatta valitusta lähdeaineistosta.



Kuva 2. Kvantitatiivisessa tutkimuksessa käytetty tekninen menetelmä.

Teknologiaa voidaan ennustaa useiden tietolähteiden perusteella. Tietolähteiden laatu, kuten esimerkiksi lähteen sopivuus, eheys, luotettavuus, saatavuus ja kronologinen järjestys voi vaikuttaa ennustamisen tarkkuuteen sekä ennustettavuuteen. Kirjallisuustutkimuksessa pyritään järjestämään eri aineistot kronologiseen järjestykseen. Teknisessä osuudessa tutkitaan aineistojen käyttäytymistä, ominaisuuksia ja vastaavuutta kirjallisuustutkimuksen tuloksiin. Teknologiadiffuusion kasvukäyrissä esitetään usein Bassin kehittämää kasvukäyrää, mutta innovaatio- ja teknologiadiffuusion kasvukäyrämalleja on jo kehitetty aikaisemmin (Bass 1969). Cho ja Daim (2016) koostivat tutkimuksessaan 10 eri yleistä laskentamallia teknologiadiffuusion kehittymisen ennustamiseen (Liite A).

Bass (1969) esitteli uusien tuotteiden leviämistä kuvaavan innovatiodiffuusiota kuvaavan laskentamallin (jatkossa Bassin laskentamalli). Hän esitti teorian, jossa yhdistyivät

aikaisemmat teorit. Bassin tutkimuksissa ennustamismallia testataan empiirisesti kerättyyn 11 eri kuluttajatuotteen yleistymiseen. Kirjoituksessaan Bass käsittelee adoption ja diffuusion olleen alun perin Rogersin (1962) esittelemä. Bass painottaakin, että kyse ei ole vain innovaatioidiffuusioteoriasta (IDT), vaan *adaption ajoituksesta* (Bass 1969).

Bass esitti perusteluina mallissaan, että todennäköisyys diffuusiolle on esitettävissä (Bass 1969). Kaavassa 1 on esitetty mukaelma Bassin esittämistä kaavoista. Kaavassa 1 osoitetaan innovaattoreiden tärkeys diffuusion varhaisessa kehittämisessä. Diffuusion alussa kaavan seuraajia kuvaavien tekijöiden tulo on nolla ($(q/m) * A(t) = 0$). Tällöin on vahvasti osoitettu, että tekijällä (q/m) ei ole diffuusion alussa merkitystä. Todennäköisyys adaptionille hetkellä t voidaan laskea kaavalla (1).

$$P(t) = \frac{f(t)}{1 - F(t)} = p + \frac{q}{m} * A(t) = p + q * F(t), \quad (1)$$

jossa

$P(t)$ on todennäköisyys adaptionille (t :n eli ajan funktiona)

$f(t)$ on todennäköisyys adaptionille ajan hetkellä t eli adaptionin nopeus

q on q -parametri eli seuraajia kuvaava vakio

m on adaption kokonaismäärä, johon adaption voi kasvaa

p on p -parametri eli innovaattoreita kuvaava vakio

$A(t)$ on adaption määrä ajan hetkellä t eli esimerkiksi saavutettu myynti

$F(t)$ on $\int_0^t f(t)$ eli diffuusion taso ajan hetkellä t . Tällöin $F(0) = 0$, kun adaption on alussa.

(Bass 1969)

Bassin kaava (1) voidaan esittää differentiaaliyhtälönä. Adaptionin määrän muutosnopeus on laskettavissa kaavalla (1.5). Differentiaaliyhtälö voidaan johtaa välivaiheina vaiheiden (1.1 - 1.4) kautta, jossa ratkaistaan todennäköisyyden tiheysfunktio $f(t)$. Integroimalla tiheysfunktion saadaan todennäköisyyden kertymäfunktio. Kertymäfunktio kuvaa ajan hetkeen t mennessä adaptionien määrää kaavassa (1.6).

$$\frac{f(t)}{1 - F(t)} = p + q * F(t) \quad (1.1)$$

$$f(t) = (1 - F(t)) * (p + q * F(t)) \quad (1.2)$$

$$f(t) = p + q * F(t) - p * F(t) - q * F(t)^2 \quad (1.3)$$

$$f(t) = p + (q - p) F(t) - q * F(t)^2 \quad (1.4)$$

$$f(t) = \frac{dF}{dt} \quad (1.5)$$

$$F(t) = \int_0^t f(t) dt \quad (1.6)$$

Bass (2004) tarkensi keskeisiä asioita vuonna 1969 esittämästään innovaatioidiffuusio-mallista (IDT). Bassin (2004) mukaan adaptiomäärä eli diffuusio eri ajanhetkille on laskettavissa. Kaavassa (2) on mukaelma alun perin Bassin (1969) esittämistä periaatteista laskennasta.

$$n(t) = \frac{dN(t)}{d(t)} = p [m - N(t)] + \frac{q}{m} N(t) [m - N(t)], \quad (2)$$

jossa $n(t)$ on adaptoinnin määrä ajanhetkellä t sekä $N(t)$ on adoptoinnin kumuloitunut määrä. Adaptiomäärän huippu aikajaksossa saavutetaan, kun diffuusiokäyrässä esitetty adaptiomäärä on suurin tarkastelujaksolla. Adaption määrän huippuajankohta T_{max} voidaan laskea p ja q -parametrien avulla kaavalla (3).

$$T_{max} = \frac{1}{p+q} \ln \left(\frac{q}{p} \right) \quad (3)$$

Adaptoijien määrän huippuajankohta voidaan vastaavasti ratkaista kaavalla (4)

$$f(T_{max}) = \frac{(p+q)^2}{4q} = \frac{1}{4q} (p + q)^2 \quad (4)$$

Bassin mallissa myynnin kumulatiivisella määrällä kuvataan adaptoijien määrää. Teknologiadiffuusio voi kasvaa myös toisella markkina- ja/tai toimialalla. Kaavassa (1) on oletettu, että markkinoiden koko (m) säilyy samana eli markkina-alue ei kasva ja teknologia *adaptoidaan vain kerran*. Kaavassa (5) on esitetty Bassin laskentamalli, jolla voidaan ennustaa teknologiadiffuusion kehittymistä näillä reunaehdoilla. Bass (2004) esittää adaption kumulatiivisen määrän $F(t)$ olevan laskettavissa kaavalla (5). Kaavasta (5) voidaan johtaa kaava (3 ja 4) kaavoilla (5.1-5.3) sekä kaava (4) kaavoilla (5.4 – 5.5).

$$F(t) = m * \frac{[1 - e^{-(p+q)(t-t_0)}]}{[1 + (\frac{q}{p})e^{-(p+q)(t-t_0)}]} \quad (5)$$

$$F(t) = \frac{[1 - e^{-(p+q)(t-t_0)}]}{[1 + (\frac{q}{p})e^{-(p+q)(t-t_0)}]} \quad (5.1)$$

$$F(t) = \frac{[\frac{(p+q)^2}{p} - e^{-(p+q)(t-t_0)}]}{[1 + (\frac{q}{p})e^{-(p+q)(t-t_0)}]} \quad (5.2)$$

$$F(t) = \frac{\ln \left[\frac{(p+q)^2}{p} - e^{-(p+q)(t-t_0)} \right]}{\ln \left[1 + (\frac{q}{p})e^{-(p+q)(t-t_0)} \right]} \quad (5.3)$$

$$T_{max} = \frac{1}{p+q} \ln \left(\frac{q}{p} \right) \quad (3)$$

$$f(T_{max}) = \frac{\left[\frac{(p+q)^2}{p} - e^{-\ln\left(\frac{q}{p}\right)} \right]}{\left[\left(1 + \frac{q}{p}\right) e^{-\ln\left(\frac{q}{p}\right)} \right]^2} \quad (5.4)$$

$$f(T_{max}) = \frac{\left[\frac{(p+q)^2}{p} \left(\frac{p}{q}\right) \right]}{\left[\left(1 + \frac{q}{p}\right) \left(\frac{p}{q}\right) \right]^2} \quad (5.5)$$

Kaavoissa (1 ja 2) tarkoitettu diffuusion kehittyminen voidaan diskretoida kaavan (6) summalauseella. Aikajakso voi olla esimerkiksi viikko, kuukausi tai vuosi. Tulosten esittäminen yksinkertaistuu, jos aikajakso on yhtä suuri kuin tarkastelujakso. Innovaatiodiffuusion kumulatiivisen määrän kehittyminen voidaan ilmaista kaavalla (6).

$$F(t) = \sum_0^n S(t), \quad (6)$$

jossa n on aikajakso eli laskentaoperaattori (esim. viikko, kuukausi tai vuosi) ja $S(t)$ on teknologian adaptoineen käyttäjien määrä aikajaksolla. Bassin laskentamallissa potentiaalisten adaptoivien käyttäjien määrä vähenee adaptoijien määrän lisääntyessä. Kaavassa (7) esitetään potentiaalisten käyttäjämäärien $P(t)$ väheneminen eri ajan hetkillä summalauseen avulla. P tarkoittaa potentiaalista käyttäjämäärää alussa ($t=0$).

$$P(t) = P - F(t) = P - \sum S(t) \quad (7)$$

Summalauseen periaate voidaan esittää myös Matlab-ohjelman avulla. Ohjelmassa 1 hyödynnetään p ja q -parametreja sekä kaavoja (1, 2, 6 ja 7). Ohjelma 1 on mukaelma Bassin kaavoista ja Ostojicin (2010) esittämästä ohjelmasta.

```
clear variables, close all;
iteraatiot=1000;
laskentavali=0.1;
q=0.40;% q-parametri      (kuva 23 Bassin laskentamallin q-parametri)
P=100000;
F(1)=0;
p=0.0055; % p-parametri (kuva 23 Bassin laskentamallin p-parametri)
t(1)=0;
for i=2:iteraatiot
    s(i)=(p +q*F(i-1)/P)*(P-F(i-1))*laskentavali;
    F(i)=s(i)+F(i-1);
    t(i)=t(i-1)+laskentavali;
end
plot(t,F,'b')
hold on
```

Ohjelma 1. Teknologiadiffuusion ennustamiseen soveltuva Bassin laskentamallin muokattu ohjelmakoodi esitettynä Matlab 2017b -ohjelmistoa varten. Ohjelman toiminta hyödyntää mm. kaavoja (1,2,6 ja 7). (Bass 1969; Bass 2004; Ostojic 2010)

Ennustusmalli tulisi validoida esimerkiksi puolittamalla populaatiosta poimittu otos ja tutkimalla toisen puolen otosta ensimmäisen puolen otoksen perusteella tehdyllä ennusteella. Tällöin voidaan todentaa ennustettavuus ilman suoraa riippuvuutta tulokseen samassa populaatiossa. Mikäli teknologian ennustamiseen on useampia tietolähteitä käytettävissä, on lähteiden yhtäaikaisesta käyttämisestä hyötyä. Useissa tutkimuksissa eri ennakointiin käytettyjen tietolähteiden esiintymien diffuusiokasvukäyrät toistuvat kronologisessa järjestyksessä. Ennustustarkkuutta on tutkimuksissa parannettu merkittävästi hyödyntämällä jo kahta eri tietolähdettä yhtäaikaisesti. Laskentamallit vaativat usein laskentaparametrien arvioimisen. Parker (1994) toi esille, että ennustusmallin laskentaparametrien arvioimiseksi voidaan hyödyntää esimerkiksi regressioanalyysia.

2.2 Teoreettisia taustoja teknologiadiffuusion ennustamisen menetelmissä

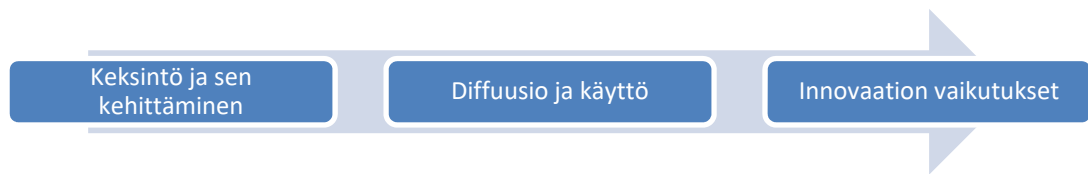
Menetelmävalinta on osa teknologiaennustamista. Caplice ja Sheffi (2006) kirjoittivat laajemman näkökulman ennustamisessa parantavan ennustamisen varmuutta. He esittävät uusien tuotteiden ennustamisen olevan enemmän prosessi kuin tekniikkoja. Wang, Liu, Zhou ja Wen (2018) esittivät, että *orastavien teknologioiden* ennakointia voidaan parantaa menetelmällä, jossa kirjallisuus, patenttitietokanta ja teknologiatiekartta integroidaan ennakoinnissa. Heidän mukaansa kirjallisuustutkimuksella saadaan tieto kehittyvän teknologian tutkimuksen nykytilanteesta. Patenttitietokannan tutkimus antaa tilannekuvan teknologian sovelluksien määrän kehityksestä sekä teknologiatiekartta kuvaa kehittyvien teknologioiden riippuvuuksia (Wang *et al.* 2018).



Kuva 3. Tulkinta Schumpeterin esittämän ”*invention-innovation-diffusion paradigm*” -teorian vaiheiden kronologisesta järjestyksestä (Schumpeter 1934; ref. Jamasb & Kohler 2007).

Teknologia tarvitsee innovaatioita teknologian leviämiseen eli teknologiadiffuusioon. Latinan kielen sana ”*innovare*” tarkoittaa ”tehdä jotain uutta”, joka ei tarkoita keksintöä. Kuvassa 3 on esitetty Schumpeterin esittämä luokittelu, joka kuvaa hyvin tätä eroa (Jamasb & Kohler 2007). Malli tunnetaan myös ”*linear model of innovations*” -mallina, jota Maclaurin kehitti (Maclaurin 1940-luku; ref. Godin 2008). Godin (2011) myöhemmin kirjoitti, että malli perustuisi kuitenkin Maurice Hollandin jo vuonna 1928 esittelemään ”*research cycle*” -malliin. Tidd ja Bessant (2013) esittävät, että innovaatio tarkoittaa taloudellisesti hyödynnettävää uuden toimintatavan tai esimerkiksi keksinnön soveltamista. Innovaatio voi olla uusi tuote, palvelu, prosessi tai tekniikka.

Keksinnön sekä innovaation erottaminen ei ole aina ollut tarkkarajaista. Godin (2015) on julkaissut runsaasti innovaation historiaa käsitteleviä julkaisuja. Godin (2008) esitti toisen näkökulman innovaatiotutkimuksen varhaisiin vaikuttajiin. Godin (2008) esitti teknologiainnovaatiotutkimuksen pioneeriksi W. Rupert Maclaurinia (1907-1959), joka jatkokehitti Schumpeterin ajatuksia. Maclaurin kehitti myös kuvassa 3 esitettyä teknologiainnovaation teoriaa sekä ensimmäiset taksonomiat teknologiainnovaation mittaamiseen (Godin 2008). Godin (2010) julkaisi myös Ogburnin työtä kuvaavan julkaisun. Julkaisussa todetaan Schumpeterin lisäksi William F. Ogburnin ja hänen kollegansa S. Colum Gilfillanin olleen teknologiainnovaatiotutkimuksen varhaisia vaikuttajia (Godin 2010).



Kuva 4. Tulkinta teknologiainnovaatiotutkimukseen soveltuvan Ogburnin esittämien teorian vaiheiden kronologisesta järjestyksestä (Ogburn 1937; ref. Godin 2010).

Taulukko 2. Mukaelma Godinin tutkimuksessaan esittämästä koosteesta Ogburnin teknologiainnovaation kronologisille vaiheille (Godin 2010).

Ajankohta	Vaiheet
Ogburn ja Gilfillan (1933)	1) Idea → 2) kokeiltava laite → 3) malli tai suunnitelma → 4) ensimmäinen demonstraatio → 5) käytännön laite → 6) vakiintunut käyttö → 7) laaja-alainen hyväksyntä
Ogburn (1937b)	1) Idea → 2) malli → 3) testi → 4) kehittäminen → 5) markkinointi → 6) myynti → 7) käyttö → 8) vaikutukset
Ogburn (1937a)	1) Idea → 2) suunnitelma tai malli → 3) design → 4) parannukset → 5) myynti → 6) markkinointi → 7) laajan mittakaavan tuotanto
Ogburn ja Nimkoff (1940)	1) Idea → 2) kehittäminen → 3) malli → 4) keksintö → 5) parannukset → 6) markkinointi
Ogburn (1941)	1) Idea → 2) suunnitelma → 3) konkreettinen muoto → 4) parannukset → 5) tuotanto → 6) markkinointi → 7) myynti

(Godin 2010)

Kuvassa 4 on esitetty tulkinta eräästä Ogburnin esittämästä teoriasta, jossa tarkastelunäkökulmina olivat keksintö ja sen kehittäminen, diffuusio ja käyttö sekä innovaation vaikutukset (Ogburn 1937; ref. Godin 2010). Kuvassa 3 ja 4 on esitetty kaksi mallia esitetyistä teknologiatutkimuksen vaiheista, miten keksintö, innovaatio ja diffuusio sijoittuvat kronologisesti toisiinsa nähden. Godin (2010) on koostanut Ogburnin esittämiä teknologiainnovaatiotutkimuksen tunnistettuja vaiheita taulukossa 2. Taulukossa olevat tut-

kimukset osoittavat eri näkökulmien erottelua ja runsautta jo historiassa. Godinin tutkimuksien mukaan kymmenet muutkin ovat osallistuneen varhaiseen innovaatio- ja teknologiatutkimukseen Schumpeterin lisäksi (Godin 2015; Godin 2010; Godin 2008).

Rose, Jones ja Furneaux (2015) esittävät korkean teknologian kasvattaneen keksintöjen ja innovaatioiden määrää. He kiteyttävätkin korkean teknologian aloilla olevan riippuvuus innovaatioista, joka mahdollistaa yritysten nopean kasvun. Rose *et al.* (2015) kirjoittivat Yhdysvaltojen kansallisen tiedesäätiön (NSF) tutkineen innovaatioiden määrää. Kyseisessä NSF:n raportissa mainittiin, että uudet teknologiat ja teollisuuden alat ovat merkittäviä Yhdysvaltojen talouden kehitykselle (NSF 2014). Raportin mukaan innovaatioiden määrä ja talouden kasvu muodostuvat pienien yritysten liiketoiminnasta, johon suuryritykset investoivat (NSF 2014). Raportissa kirjoitettiin, että investoinnin kohteena olevissa yrityksissä keskeinen toimiala liittyy *korkeaan teknologiaan* (NSF 2014). Rose *et al.* (2015) kirjoittivat, että NSF:n raportin mukaan informaation ja teknologian aloilla oli suurempi innovaatioiden määrä kuin muilla teollisuuden aloilla (Rose *et al.* 2015). Sganzerla, Seixas Filho ja Conti (2016) esittävät, että perinteisten teollisuuden aloilla on siirrytty strategioista käytäntöihin. Heidän mukaansa digitaalisuuden innovaatioita sovelletaan käytännössä (Digital Transformation). He esittävät myös uusien liiketoiminta- ja toimintamallien luonnin nopeutuneen aikaisemmasta. Sganzerla *et al.* (2016) mukaan muutoksen aiheuttajina ovat olleet mm. globaali kysyntä, erilaiset ansaintamallit, markkinoiden vaihtelu ja pitkäaikaisen omaisuuden elinkaariajattelu. He kirjoittivat myös digitaalisten innovaatioiden parantaneen toiminnan tehokkuutta. (Rose *et al.* 2015; NSF 2014; Sganzerla *et al.* 2016)

Tidd ja Bessant (2013) tutkivat innovaatioiden luokittelua. Tutkimuksessaan he luokittelivat innovaatioita monella eri tavoilla. Bessant (2015) kirjoitti innovaatioiden olevan jaettavissa neljään eri ryhmään seuraavasti:

- 1) *Tuoteinnovaatiot* ovat muutokset tuotteissa ja palveluissa, joita organisaatiot tarjoavat.
- 2) *Prosessi-innovaatiot* ovat muutokset tavoissa, miten yrityksen tarjoamaa tuotetaan ja toimitetaan asiakkaalle.
- 3) *Asemointi-innovaatiot* ovat muutokset kontekstissa, joissa tuotteita ja palveluita otetaan käyttöön. Tällainen voi olla esimerkiksi tuotteen käyttökohde.
- 4) *Paradigmainnovaatiot* ovat aikaisempaa mielikuvia rikkovia innovaatioita, mitä organisaatioiden ei ole oletettu voivan tehdä ja toimittaa asiakkaalle.

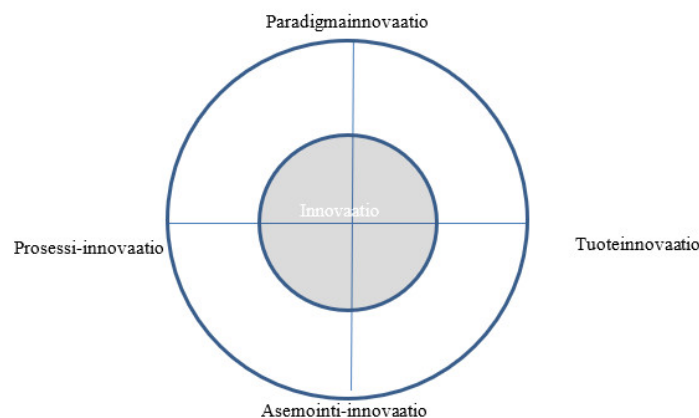
(Bessant 2015)

Tidd ja Bessant (2013) esittivät edellä mainittujen innovaatioiden mahdollisuudesta luokitella eri tavoilla riippuen muutoksen määrästä tai innovaation osuudesta kokonaisuudessa. Innovaatiot voidaan myös jakaa inkrementaaliseen tai radikaaliin innovaatioon

sekä luokitella modulaariseksi tai arkkitehtuuriseksi innovaatioksi. Innovaation radikaalisuutta kuvaa luokittelu, jossa innovaatio luokitellaan paradigman säilyttäviin, muuttaviin tai murtaviin innovaatioihin. (Tidd & Bessant 2013)

Al-Sabri ja Al-Mashari (2017) tutkivat prosessi-innovaatiota. He kirjoittivat prosessien kehittämisen kasvattaneen viime vuosina merkittävästi suosiotaan. Heidän mukaansa liiketoimintaprosessien kehittämiseen voidaan hyödyntää valmiita menetelmiä. Tutkimuksen yhteenvedossa todetaan, että prosessi-innovaatioiden kehittämiseen käytetty menetelmä vaikuttaa kehittämisen kohteeseen. Eri menetelmät vaikuttavat prosesseihin eri tavoilla. Osa menetelmistä parantaa prosessin tarkkuutta ja osa muuttaa prosessia täydellisemmäksi tai johdonmukaisemmaksi. (Al-Sabri & Al-Mashari 2017)

Kanagal (2015) tutki tuoteinnovaatioita. Hänen mukaansa tuote sisältää yhden tai useamman seuraavista ominaisuuksista: sisältö, tuotteen attribuutit, edut, hyödyt, tuoteominaisuudet, toiminnallisuus, suorituskyky, liiketoimintamalli, käyttökokemus ja kuluttajakokemus (Kanagal 2015). Kanagal (2015) esitti tuoteinnovaation tarkoittavan innovaatiota, jolla muutetaan edellä lueteltuja tuotteen ominaisuuksia, kuten tuotteen elinkaarta tai muuta ominaisuutta. Hän kirjoitti myöhemmin myös tuoteinnovaation markkinavalinnoista, jossa esitettiin disruptiivisten uutuustuotteiden tuominen ensin kehittyville markkinoille (Kanagal 2017). Myös Winter ja Govindarajan (2017) esittävät useiden tuotteiden osalta markkinastrategiaa, jossa voimakkaasti muuttuvat tuoteinnovaatiot tuotiin ensin kehittyville markkinoille.



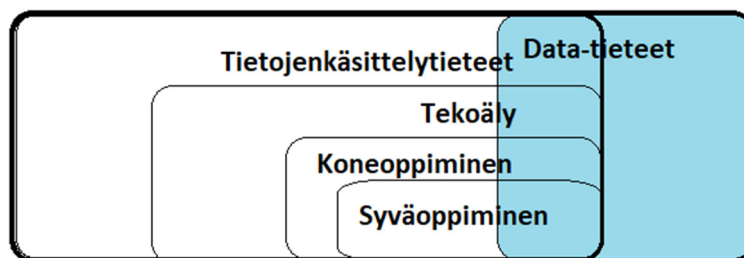
Kuva 5. Mukaelma innovaatiotyypin esityksestä, jossa innovaatio voi edustaa useampaa innovaatioryhmää yhtäaikaisesti (Tidd & Bessant 2013; Jokela 2016).

Tidd ja Bessant (2013) esittivät, että innovaatio voi kuulua useampaan innovaatioryhmään yhtäaikaisesti. Heidän mukaansa uusi tuotekokonaisuus voi edustaa asemointi-, prosessi- ja tuoteinnovaatiota yhtäaikaisesti (kuva 5). Radikaaleista asemointi-innovaatioista voidaan tuoda esille mm. halpalentoyhtiöt sekä paradigmmainnovaatioista fyysisen musiikkimedian tarpeen syrjäyttänyt iTunes alusta (Tidd & Bessant 2013).

2.3 Tekoälyteknologioiden riippuvuus big datasta ja datatieteistä

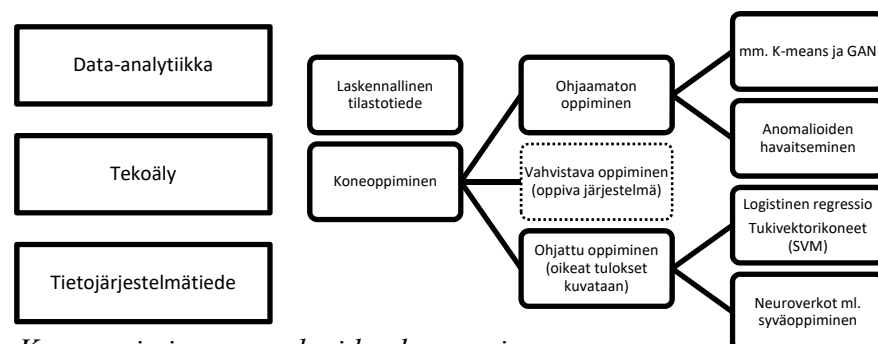
Tässä työssä tutkittavana kohteena on big data -ilmiö, joka vaatii useampia teknologioita toteutuakseen. Olennaista palveluissa on se, että palveluilla voidaan tuottaa asiakkaille toimitettavaa arvoa (Davis 2008). Big data ei itsessään tuota palveluita, jotka tuottaisivat käyttäjille arvoa. Tavallista onkin, että big dataa hyödynnetään eri teknologioilla, kuten data-analytiikalla ja tekoälyllä. Koneoppiminen ja syväoppiminen hyödyntävät eri tyyppisten dataa (Roos & Hagström 2018). Big dataa voidaan tältä osin pitää ilmiönä, joka vahvistaa tekoälyn hyödyntämisen mahdollisuuksia erityyppisissä teknologiaratkaisuissa.

Roos ja Hagström (2018) esittävät syväoppimisen kuuluvan koneoppimiseen ja molempien olevan osa tekoälyyn kuuluvia teknologioita (kuva 6), jotka hyödyntävät datatieteitä. Big data eroaa perinteisistä dataratkaisuista sen osalta, että siinä käytetään erityyppisiä dataa suurina määrinä, jonka käsittely on nopeaa (Davenport 2014).



Kuva 6. Datatieteiden tietojenkäsittelytieteiden riippuvuus. Kuva on mukaelma ”Elements of AI”-kurssin sisällöstä, jossa osoitetaan eri tekoälyosa-alueiden riippuvuus datatieteistä. (Roos & Hagström 2018)

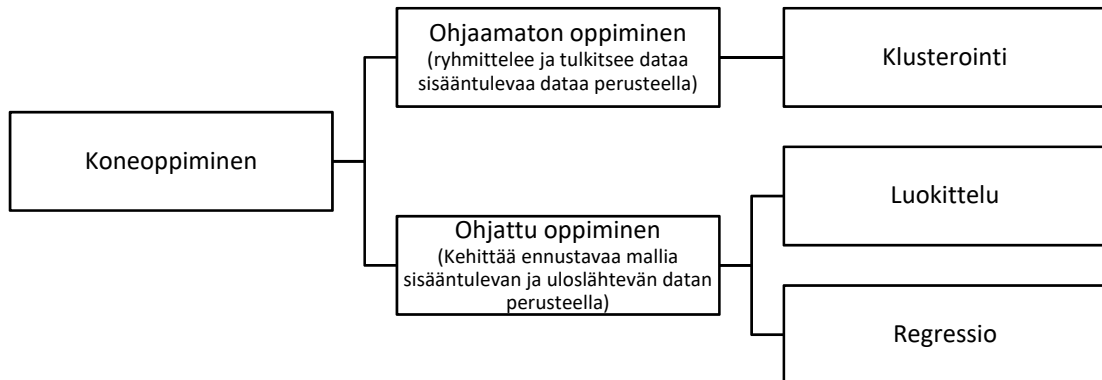
Ailisto, Heikkilä, Helaakoski, Neuvonen ja Seppälä (2018) kuvasivat tekoälyratkaisujen kokonaisuudesta (kuva 7) tarkentaakseen käsitteitä ja niiden välisiä suhteita.



Kuva 7. Koneoppimisen osa-alueiden kuvaaminen mukaeltuna Ailisto et al. (2018) esittämästä kuvasta.

Työssään he luokittelivat koneoppimisen sisältävän ohjatun-, ohjaamattoman ja vahvistavan oppimisen (Ailisto et al. 2018). Koneoppiminen (Machine learning) voidaan luokitella myös menetelmien osalta MathWorks (2016) tuottaman opetusmateriaalin mukaisesti (kuva 8), jossa koneoppiminen on jaettu vain kahteen osa-alueeseen.

Kuvassa 8 on käytetty luokittelua, jossa vahvistavaa oppimista ei ole huomioitu, jonka mm. Ailisto *et al.* (2018) ja Sims (2015) toivat esille esittämissään jaotteluissa (Kuva 7).



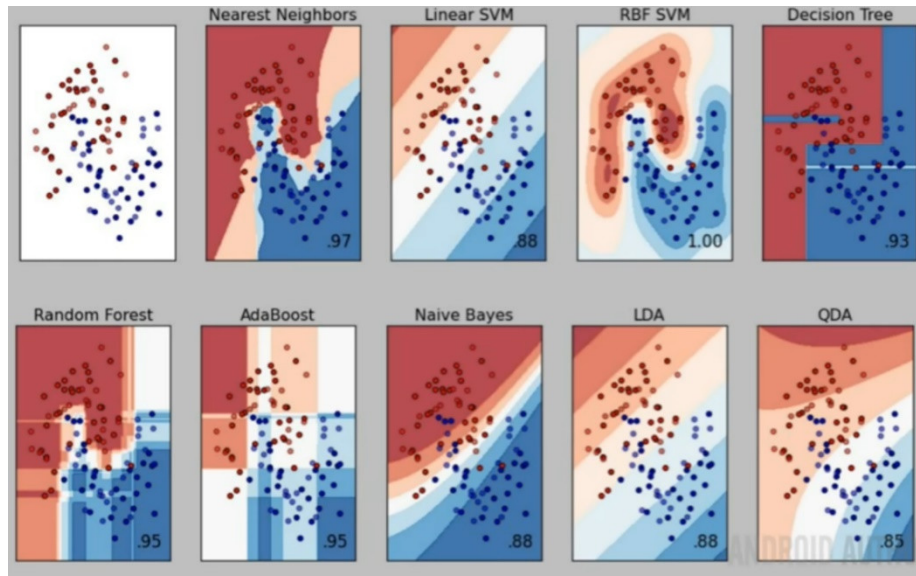
Kuva 8. *Matworks (2016) esittämän kaavion mukaelma, jossa on jaoteltu koneoppiminen vain kahteen osa-alueeseen ja osoitettu sen riippuvuus datasta.*

Ailisto *et al.* (2018) mukaan ohjaamaton ja ohjattu oppiminen kuuluvat tekoälyn konnektionismien koulukuntaan, joka hyödyntää datapohjaisia menetelmiä. Konnektionismissa ajattelu, tunteet, tietoisuus ovat esitetty yksinkertaisten yksiköiden avulla. Yksiköt kuten neuronit kytkeytyvät synapsien avulla toisiinsa muodostavat neuroverkkoja. Tekoälyn osalta neuroverkot ovat keinotekoisia. Neuroverkkoja voidaan hyödyntää esimerkiksi kasvojen tunnistuksessa. (Ailisto *et al.* 2018)

Ailisto *et al.* (2018) esittivät koneoppimisen olevan tietotekniikan osa-alue, jossa yleensä hyödynnetään tilastotiedettä. Koneoppimisessa tarkoitetaan tietokoneen kykyä parantaa suoritustaan *datan avulla* ilman ohjelmointia, jolla määritellään uusia käsittelysääntöjä (Ailisto *et al.* 2018). Samoin Roos ja Hagström (2018) esittivät koneoppimisen riippuvuuden datatieteistä sekä Sims (2015) toi esille koneoppimisen riippuvuuden datasta ja sen saatavuudesta. (Ailisto *et al.* 2018; Roos & Hagström 2018; Sims 2015)

Kuvassa 7 ja 8 esitettyjen datan luokitteluun, regressioon ja klusterointiin tarvittavan oikean algoritmin valinnassa ei ole valmista vastausta (Mathworks 2016). Algoritmi voidaan valita kokeilemalla sen sopivuutta dataan, mutta valintaa voi ohjata myös datan tyyppi tai datajoukon koko (Mathworks 2016). Sims (2015) esittää, että koneoppimisen osa-alueena *ohjaamaton oppiminen* perustuu dataan, jota ei ole etukäteen luokiteltu. Roarkin (2018) mukaan ohjaamattoman oppimisen avulla voidaan *löytää rakenne* rakenteettomalle datalle. Sims (2015) esityksessä on visualisoitu eri algoritmien sopivuutta. Visualisointi on havainnollinen tapa kuvata algoritmien toimintaa ja sopivuutta kyseiselle datalle (kuva 9). Sims (2015) esittämässä kuvan 9 vasemmassa yläkulmassa on mallinnettava data.

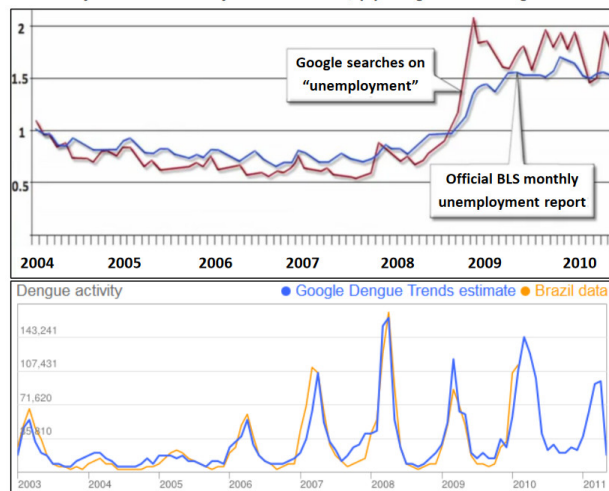
Kuvassa on vertailtu erilaisia luokittelualgoritmeja. Tässä tapauksessa on vertailtu eri malleja lineaarisen tukivektorikoneen avulla. ”RBF SVM” -luokittelu on saanut kuvassa 9 arvon 1.00, jolloin mallin laskennallisesti kuvattu alue peittää sopivimmin data-alkiot.



Kuva 9. Sims (2015) esittelee visuaalisesti eri luokittelualgoritmien sopivuutta yhdelle esimerkille. Kuvaamalla dataa ja luokittelualgoritmeja voidaan värityksellä osoittaa mallin sopivuutta dataan. (Sims 2015)

Martin (2013) tutki artikkelissaan big datan hyödyntämismahdollisuuksia. Yhtenä aihealueena hän vertasi internetistä saatavia sanahakuja tutkimustuloksiin.

Figure 2: Real-time Prediction: (a) Google searches on unemployment vs. official government statistics from the Bureau of Labor Statistics; (b) Google Brazil Dengue Activities



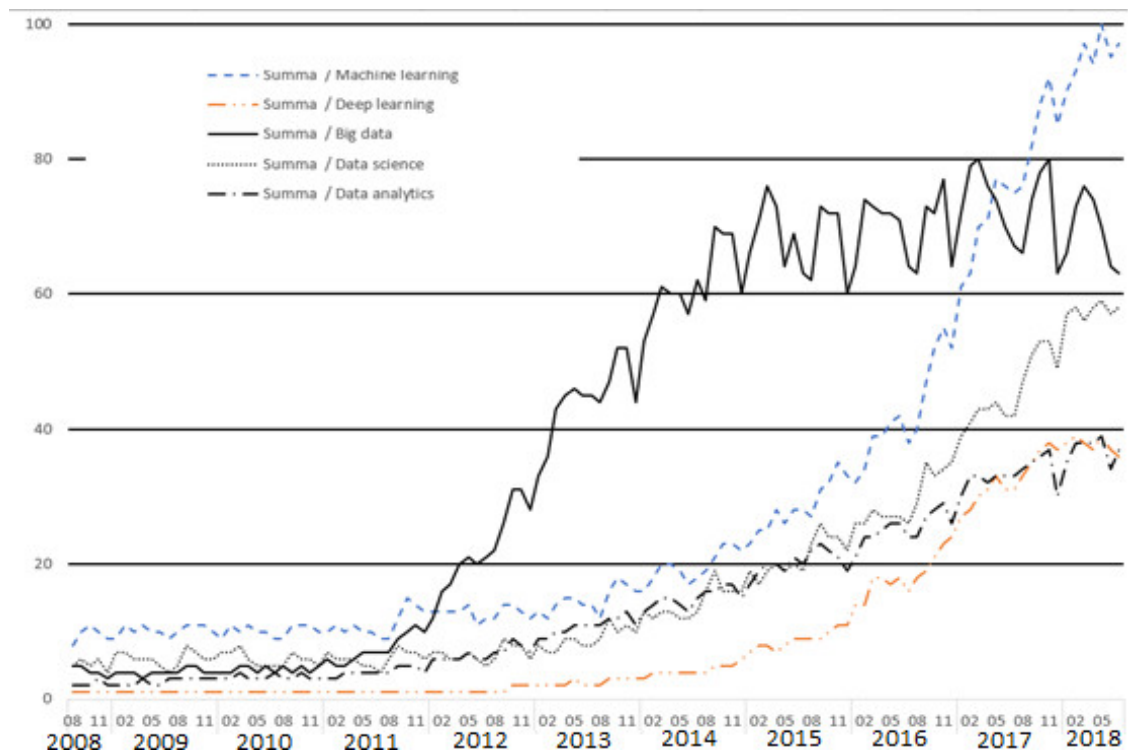
Source: Hubbard, 2011; <http://www.hubbardresearch.com>; Google correlate, <http://www.google.org/denquetrends/about/how.html>

Kuva 10. Tutkimuksessa esitetty ”Real-time Prediction”-kuvaus (Martin 2013).

Martin (2013) kirjoitti kommenttien, hakujen tuottavan lähes samoja tilastollisia tuloksia, kuin kotitalouskyselyt ja äänestykset. Kuvassa 10 Martin (2013) osoittaa kahdessa eri

tapauksessa, miten Googlesta tehdyt internethaut ja tilastolliset tutkimukset korreloivat toisiaan. Kuvassa 10 on verrattu Googlesta tehtyä sanahakua tutkimustuloksiin kahden eri tilaston osalta. Hänen mukaansa Google hakutulokset ovat olleet käytettävissä mak- sutta, reaaliaikaisesti sekä helposti (Martin 2013). Artikkelissaan Martin (2013) osoitti Google hakujen määrien sekä tilastotutkimustuloksien korreloivan kahdessa eri tapauk- sessa. Kahden tapauksen perusteella ei voida tehdä johtopäätöstä, että internetsanahaku on sopiva useisiin tilastollisesti tutkittaviin asioihin. Vastaavasti artikkelissa osoitetaan, että tilastotutkimuksen tulos on saatu viiveellä johtuen tilastollisen työn hitaudesta (Mar- tin 2013).

Dataratkaisujen kehittyminen vaikuttaa datasta riippuvien teknologioiden kehittymiseen. Riippuvien ilmiöiden vaikutusta voidaan arvioida olettamalla, että tekoälyn kehittämi- selle oli etua big data -ilmiön kehittämisestä. Tekoälyn ja big datan välistä riippuvuutta voidaan tutkia internetsanahakujen kehittämisellä. Tässä taustoitetaan asiaa big datan - ilmiön osalta. Taustoituksessa hyödynnetään Martin (2013) vahvistamaa hypoteesia, että Google -hakutulos korreloi tilastollista asian ilmentymistä.



Kuva 11. Suhteelliset hakusanamäärät kohteissa, joissa haku on kohdistettu koneoppimisen, syväoppimisen, big datan, datatieteiden ja data-analytiikkaan. Kuva on esitetty ajan funktiona (kk, vuosi).

Kuvassa 11 pyritään kuvaamaan Roosin ja Hagströmin (2018) sekä Simsin (2015) esittä- mää tekoälyn ja datan riippuvuutta Google Trends -palvelun avulla (Liite B). Tietojen keräämisessä on hyödynnetty Google Trends -palvelua. Kuvan 11 pystyakseli (Y-akseli) kuvaa hakutermien esiintymien suhteellista osuutta (0-100 prosenttia) ja vaaka-akseli (x- akseli) kuvaa aikaa.

Simsin (2015), Davenportin (2014) sekä Gandomin ja Haiderin (2015) mukaan big data sisältää pääosin järjestämätöntä ja rakenteetonta dataa. Big datan lisäksi kuvassa 11 esitettyyn poimintaan on valittu Roosin ja Hagströmin (2018) kuvassa 6 esittämiä riippuvuuksia. He esittivät riippuvuuden tekoälyn, koneoppimisen, syväoppimisen, datatieteiden ja data-analytiikan välillä. Kuvassa 11 on valittu parametreiksi edellä mainitut termit ja toteutettu internethakujen määristä poiminta (7/2018). Tietojen keräämisessä on käytetty Google Trends -palvelua. Kuvan 11 lähtödata on peräisin esitetystä Google Trends -poimintatuloksesta (liite B). Poimintatulosta on jatkokäsitelty Excel -ohjelmistolla aineiston järjestämistä ja kuvaamista varten. Kuvassa 11 esitettyjen osuuksien kehittymistä vertaamalla voidaan todeta, että big data on ilmiönä yleistynyt voimakkaammin ennen koneoppimista, syväoppimista, datatieteitä ja data-analytiikkaa. Ilman tarkempaa tutkimista, voidaan todeta, että otoksen perusteella big data -ilmiö on kehittynyt mm. koneoppimista aikaisemmin. Koneoppimisen teknologiarippuvuutta big datasta ei voi päätellä yhden satunnaisen otoksen perusteella. Voidaan kuitenkin todeta, että tämän otoksen perusteella big data -ilmiö ja kyseiset teknologiat ovat kehittyneet voimakkaasti 2010-luvulla, mutta niillä ei ole välttämättä riippuvuutta keskenään. Teknologian tuntemisesta voi olla etua seurattavien riippuvuuksien liittämis-, osateknologioiden ja ilmiöiden valinnassa. Kirjallisuustutkimus on yksi yleinen keino uuteen teknologiaan tutustumisessa. Kirjallisuudesta voi löytyä vastauksia *teknologian riippuvuuksien tunnistamiseksi*. Aktuaali kirjallisuus kuvaa myös asioiden muuttumista. Kuusi (2018) antaa esimerkin oppivista neuroverkoista, joka on yksi lupaavista teknologioista. Oppivia neuroverkkoja ei muutama vuosi aikaisemmin nähty mahdollisena teknologioiden haastajana.

Ailisto *at al.* (2018) toivat aikaisemmin esille, että tekoälyyn perustuvat ohjelmistoratkaisut parantaisivat suoritustaan ilman ohjelmointia. Tämän perusteella innovaation luokittelun mukaan big dataan liittyvät tekoälyratkaisut voidaan arvioida edustavan prosessi-, tuote- asemointi- ja paradigmainnovaatioita yhtäaikaisesti. Big datassa on luovuttu datan tarkasta määrittelystä ja luotetaan rakenteettoman datan tuovan tuloksia. Innovaation arkkitehtuuri vs. modulaarisuusluokittelussa big data sekä sitä hyödyntävät tekoälyratkaisut edustavat arkkitehtuurimuutosta. Teknologian sekä sen käyttökohteiden tunteminen mahdollistaa asioiden huomioimisen teknologiaennusteessa. Big dataan perustuvilla ratkaisuilla on mahdollista hyödyntää tekoälyä, kuten koneoppimista, syväoppimista tai vahvennettua oppimista sekä organisaatioiden ulkopuolisia tietoja. Big data sekä sen liitännäisteknologiat ovat muuttaneet käsitystä tietojen käsittelystä. *Teknologiadiffuusion laskentamallin valintaan vaikuttaa teknologian innovatiivisuus (Bassin malli) tai esimerkiksi teknologian korvaavuus (Fisher-Pry -laskentamalli)*.

Tutkittavan ilmiön tai ratkaisun syvempi tunteminen voi edellyttää liitännäisteknologioiden riippuvuuden ymmärtämisen ja erottamisen aikaisemmista ratkaisuista sekä osateknologioista. Davenport (2014) toi kirjassaan useasti esille big dataksi esitettyjä ratkaisuja, jotka eivät edustaneet big dataa. Hän esittikin useilla tavoilla, miten big data eroaa perinteisistä ratkaisuista ja miten analytiikkaa on sovellettu eri aikakausina. Davenport (2014)

luokitteli eri aikakausina käytettyä terminologiaa (taulukko 3), jolla voidaan erottaa ilmiöitä toisistaan.

Taulukko 3. *Datakäytössä olevien termien käyttö käyttötarkoituksittain tiedollajoh-
tamisessa (Davenport 2014).*

Käytetty termi	Ajanjakso	Selite
Päätöksen tuki (Decision support)	1970-1985	Data-analyysin hyödyntäminen päättämisessä
Johdon tuki (Executive support)	1980-1990	Data-analytiikan hyödyntäminen johdon rapor- toinnissa
Käytönaikainen raportointi (OLAP)	1990-2000	Ohjelmistot moniulotteisen datan analysointiin
BI-järjestelmät (Business Intelli- gence)	1989-2005	Työkaluohjelmistot tietopohjaiseen päätöksen- tekoon eli liiketoimintatiedon hallinta
Analytiikka (Analytics)	2005-2010	Keskittyminen tilastolliseen ja matemaattiseen analysointiin päätöksenteossa
Big data	2010-	Keskittyminen isoihin rakeenteettomiin ja no- peasti liikkuviin datoihin (big data 3v's)

2.4 Innovaatio- ja teknologiadiffuusiomalli

Vuoden 1945 jälkeen tutkittiin uusien tuotteiden markkinoille tuloa (Fourt & Woodlock 1960). Heidän tutkimuksessaan esitettiin, että neljä viidestä uudesta tuotteesta epäonnistu-
tuu markkinoille tulossa. He kehittivät myös matemaattisen mallin, jolla voidaan enna-
koida uuden tuotteen markkinoille pääsyä (Fourt & Woodlock 1960). Mansfield (1961)
tutki, miten nopeasti uusi teknologian siirtyy yritykseltä toiselle. Hänen tutkimuksessa
keskityttiin esimerkin vaikutukseen uusien tuotteiden markkinoille tulossa. Tutkimuk-
sessa kehitettiin matemaattinen malli, jolla selitetään innovaation siirtymisen määrien
vaihteluja (Mansfield 1961). Teknologiadiffuusion kehityskäyrä muistuttaa usein biolo-
gisen kasvun kehityskäyrää eli S-muotoista käyrää, joka lähenee ajan kuluessa saturaat-
iotasoa (Marinakis 2012). Bass kehitti vuonna 1969 innovaatioidiffuusiomallin, jolla voi
kuvata uusien tuotteiden tai palveluiden kehittymistä (Mahajan, Muller & Bass 1995).
Bassin malli on keskeinen matemaattinen menetelmä innovaatioidiffuusion ennustami-
sessa (Wang Y., Pei L. & Wang Z. 2017). Wang *et al.* (2017) mukaan Bassin mallin etu
verrattuna aikaisempiin malleihin oli se, että mallissa on yhdistetty ulkoisien tekijöiden
(Fourt & Woodlock 1960) ja sisäisten tekijöiden (Mansfield 1961) vaikutukset samaan
malliin.

Bass (1969) jakoi tutkimuksessaan innovaation adaptoijat omiin ryhmiin, joissa huomioidaan ulkoisten tekijöiden ja sisäisten tekijöiden vaikutukset. Mahajan *et al.* (1995) mukaan Bassin diffuusiomallin etuna on sen helppo esittäminen. He toteavat diffuusiomallin olevan kuvattavissa esimerkiksi graafisesti, matemaattisesti ja symbolisesti. Bassin diffuusiomalli, joka muistuttaa S-käyrän muotoista kuviota (kuva 12) täytti toivotut vaatimukset (Mahajan *et al.* 1995). Mahajan *et al.* (1995) mukaan Bassin diffuusiomallilla voidaan selittää mm. luonnollista kasvua, tietokoneiden, talouden kehittymistä ja useita muita asioita. He kiteyttivät kuitenkin, että diffuusiomallit ovat kasvumalleja, mutta kaikki kasvumallit eivät välttämättä ole diffuusiomalleja.

Bassin (1969) esittämästä innovaatioidiffuusiomallin (IDT) julkaisuajankohdasta on esitetty erilaisia näkemyksiä. Bass's Basement Research Institute (2010) mukaan vuonna 1969 esitelty malli perustuisi Bass jo vuonna 1963 esittämään malliin. Bass (1963) tutki aikaisemmin markkinakäyttäytymistä. Tässä tutkimuksessa kehitettiin markkinadiffuusiomallia, jolla selvitettiin markkinoinnin vaikutusta uusien asiakkaiden ostokäyttäytymiseen.

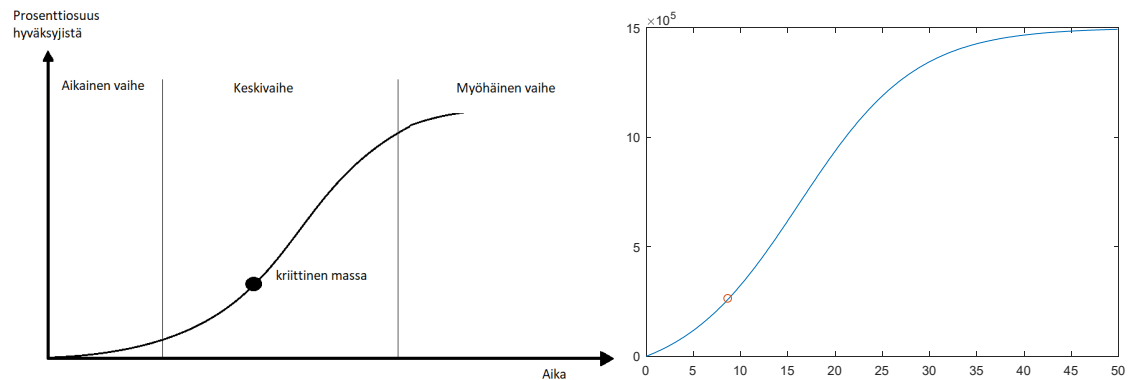
Yleisesti empiiriset kasvumallit soveltuvat myös teknologiadiffuusion kuvaamiseen. Teknologian ennustamisessa on merkitystä ennustusmallin valinnalla. Teknologian korvautuminen tai syrjäytyminen voidaan huomioida teknologiadiffuusiota kuvaavissa malleissa. Marinakis (2012) esitti, että Fisher-Pry -malli sopii erityisesti teknologiadiffuusion kehittymiseen (Liite A). Marinakis toi tutkimuksessa esille Fisher-Pry mallin paremman sopivuuden korvaavan teknologian teknologiaennustamiseen, jossa huomioidaan kilpailu toisen teknologian kanssa (Fisher & Pry 1971; Marinakis 2012). Myös Fisher-Pry mallin muoto muistuttaa S-käyrää eli biologisen kasvun mallia (Fisher & Pry 1971). Vastaavasti teknologian syrjäytymiseen toisen teknologian tieltä voidaan käyttää myös Gompertzin kuolleisuusmallia. (Marinakis 2012)

Cho ja Daim (2016) esittivät tutkimuksessaan *kriittisen massan* merkittävydestä markkinoille siirtymisessä. Heidän mukaansa kiihtyvän kasvun vaiheen ja kriittisen massapisteen välisenä aikana (innovaatioidiffuusiossa) on *kannattavinta tulla markkinoille*. He perustelivat väitettään sillä, että tässä vaiheessa yritykset tukevat riittävästi asiakkaiden tuotteita, riskit ovat pienemmät ja mahdollisuudet paremmat kuin muissa teknologian kehittymisen vaiheissa (Cho & Daim 2016). Sundqvist, Frank, Puumalainen ja Kämäräinen (2002) esittivät teknologiadiffuusiota kuvaavan S-käyrän kiihtyvässä vaiheessa olevan kriittisen massapisteen laskettavissa kaavalla (8).

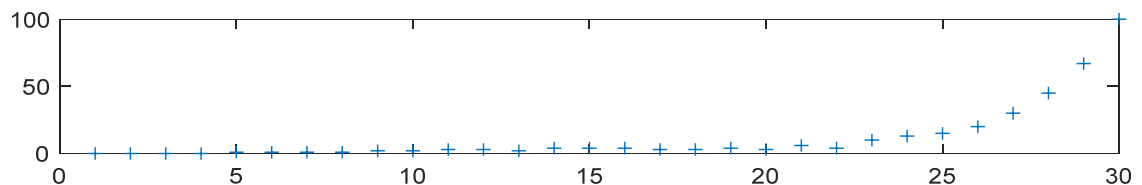
$$t = \frac{\ln((2+\sqrt{3}) * p/q)}{-(p+q)}, \quad (8)$$

jossa t on ajankohta kriittiselle massapisteelle, p on p -parametri eli innovaattoreita kuvaava vakio ja q on q -parametri eli seuraajia kuvaava vakio. Sundqvist *et al.* (2002) ovat esittäneet, että kriittinen massapiste (CMP) saavutetaan keskimäärin, kun 10-20 prosenttia diffuusion saturaatiotasosta on saavutettu. Matemaattisesti CMP sijoittuu riippuen p

ja q -parametreista eli innovaattoreiden ja seuraajien suhteellisista määristä. CMP:n on esitetty olevan toisen derivaatan kohdalla. CMP riippuu näistä parametreista, vaikka se saavutetaan S-muotoisessa käyrän Y-akselilla yleensä 10 ja 20 prosentin välillä. Sundqvist *et al.* (2002) tutkimustuloksissa CMP sijaitsi 1 ja 34 prosentin välillä keskiarvon ollessa noin 13 prosenttia. He esittivät, että maantieteelliset erot olivat merkittäviä syitä osuuksien eroissa. Sundqvist *et al.* (2002) tuovat esille, että verkostovaikutus vahvistaa diffuusiota voimakkaasti.



Kuva 12. Vasemmalla mukaelma innovaatioidiffuusion S-käyrästä, joka perustuu Bass (1969) esittämään malliin (Xiong, Payne & Kinsella 2016). Oikealla laskettu Sundqvist *et al.* (2002) mukaisella laskentakaavan kohdalla oleva CMP. CMP saavutetaan x-akselin arvolla = 8.6 (Liite C).

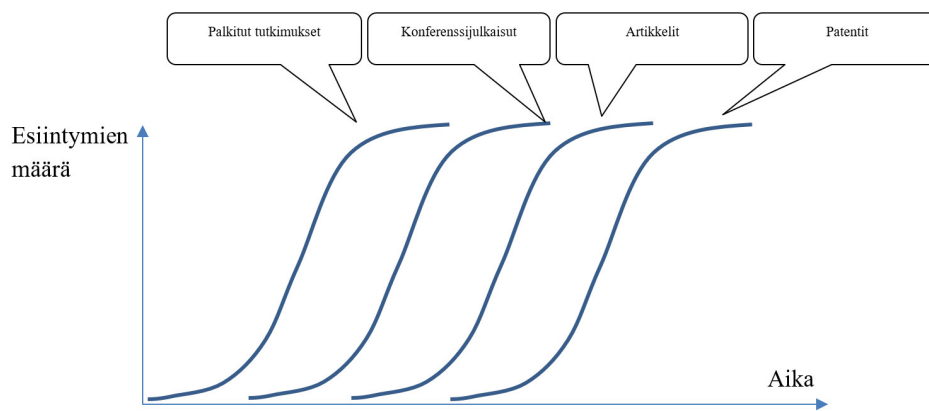


Kuva 13. Tulevaisuudessa olevan CMP:n laskenta kuvitteellisen datan avulla, jossa data loppuu x-akselin arvoon 30 (Liite D). Sundqvist *et al.* kaavan mukaan CMP saavutettaisiin x-akselin arvolla 32, jos kehitys säilyisi laskentamallin mukaisena. Toinen derivaatta saavutetaan vasta arvolla 35.

Kuvassa 13 on esitetty CMP:n laskenta tulevaisuuden ajankohtaan. CMP on merkittävä investoinnin kannalta (Sundqvist *et al.* 2002). Sundqvist *et al.* (2002) esittivät adaptio-määrän vähenevän, jos kriittistä massapistettä ei saavuteta. Vastaavasti he tuovat esille adaption kiihtyvän kriittisen massapisteen saavuttamisen jälkeen. Parametrien p ja q arvojen perusteella diffuusio voidaan jakaa kehittymisen kahteen ryhmään. Ensimmäisessä ryhmässä *menestyvällä tuotteella* seuraajien määrää kuvaava q -parametri on p -parametria suurempi ($p < q$) Jackson (2013). Toisessa ryhmässä seuraajia on vähemmän kuin innovaattoreita, jolloin tuote ei menesty Jackson (2013). Tässä työssä keskitytään kriittisen massapisteen saavuttaviin kasvumalleihin. Näissä kasvumalleissa seuraajien osuutta kuvaava kerroin (q -parametri) on innovaattoreiden osuutta kuvaava (p -parametri) määrää suurempi. Kaminski (2011) mukaan *massapiste sijaitsee varhaisten omaksujien ja varhaisen enemmistön välissä*. Hän esittää varhaisten omaksujien johtavan mielipiteillä, asettavan trendejä sekä roolimalleja teknologian myöhemmin adaptoiville.

Huomioiden Kaminskin esityksen massapisteen sijainnista: Sundqvist *et al.* (2002), Mahajan *et al.* (1995) ja Rogers (2003) esittävät lähes yhtä suuren keskimääräisen suhteellisen arvon kriittiselle massapisteelle (kuva 12 ja 16).

Linden ja Fenn (2003) esittävät, että yritysten osalta adaptoinnin optimaika riippuu yritystyyppistä. He jakoivat yritykset A, B ja C -tyyppiin ja esittivät edellä mainituille yritystyypeille eri optimaajoitukset adaptoida uutta teknologiaa. Tuotteiden yleistymisen ja optimaajoituksen ajankohta eivät tule yllätyksenä. Cho ja Daim (2016) kuvasivat patenttien määrän kasvamisen tapahtuvan (kuva 14) ajallisesti myöhäisessä vaiheessa sekä mm. patenttien määrien kasvun olevan kuvattavissa Bassin esittämällä diffuusion kasvukäyrällä. Kronologinen tulkinta tutkimuksessa esitetystä järjestyksessä on esitetty kuvassa 15 (Cho & Daim 2016).



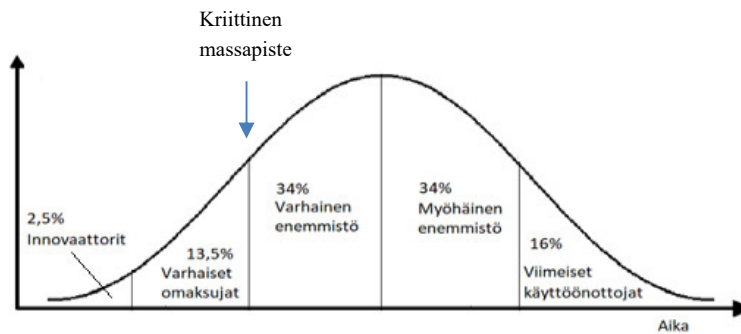
Kuva 14. Tulkinta patenttien esiintymien määrän sijoittumisesta muiden aineistojen esiintymien määrän kehittymiseen (Cho & Daim 2016).



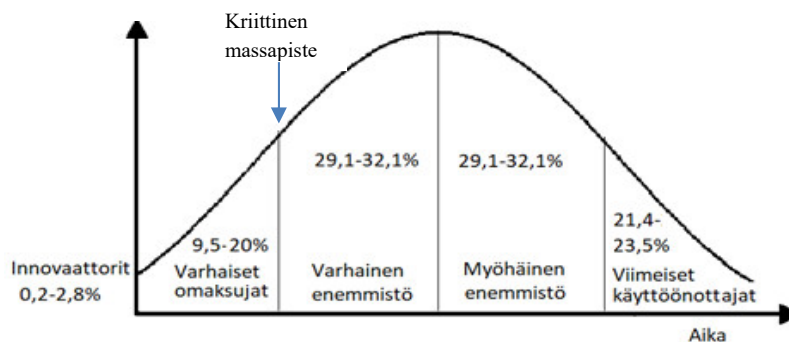
Kuva 15. Kronologinen tulkinta Chon ja Daimin (2016) esittämästä järjestyksestä teknologiaa indikoivista tietolähteistä.

Eri käyttäjäryhmät adaptoivat innovaatiot eri vaiheissa. Rogers (1962) esitti edelleen käyttökelpoisen innovaatioidiffuusiomallin (Rogers 2003; Rogers 1962). Rogers tutki innovaatioidiffuusiota ja kehitti mallin, jossa on jaettu innovaation kehittymisen vaiheet käyttäjäryhmien mukaisesti. Rogersin malli on tunnettu luokittelusta, jossa hän jakaa innovaation omaksujat kuvassa 16 esitettyyn viiteen ryhmään. Kuvassa 17 on Bassin esittämät prosenttiosuudet innovaation omaksujista. Kaminskin (2011) mukaan Ryan ja Gross ovat esittäneet vastaavan luokittelutavan jo vuonna 1943. Vertaamalla kuvaa 16 ja kuvaa 17 huomataan yhteneväisyydet eri tutkijoiden esittämistä käyttäjäryhmistä. Esitetyt ryhmät ovat yhtenevät (kuva 18), mutta Bassin (1969) ja Rogersin (2003) esittämien käyttäjäryhmien prosenttiosuuksissa on eroja eli prosenttiosuudet eri käyttäjäryhmille eivät

ole tarkkoja lukuja. Myös Jackson (2013) totesi, että innovaattoreiden ja seuraajien osuudet vaihtelevat, joka vaikuttaa uuden tuotteen menestymiseen markkinoilla. Huomioiden Sundqvist *et al.* (2002) esittämät prosenttiosuudet ja Kaminski (2011) esittämä kriittisen massapisteen sijainti, voidaan todeta prosenttiosuuksien vaihtelevan runsaasti. Osuuksiin vaikuttavat muun muassa markkina-alue, tuote, palvelu tai verkostovaikutus.



Kuva 16. Mukaelma innovaatioadaption prosenttiosuuksista eri käyttäjäryhmille Rogersin (2003) sekä kriittisen massapisteen sijainti Kaminskin (2011) mukaan.



Kuva 17. Mukaelma innovaatioadaption prosenttiosuuksista eri käyttäjäryhmillä, jossa Kaminskin esittämä kriittisen massapisteen sijainti (Mahajan *et al.* 1995; Kaminski 2011).



Kuva 18. Bass (1969) sekä Rogers (2003) tutkimusten perusteella tehty kronologinen järjestys luokitelluille käyttäjäryhmille innovaatioidiffuusiassa.

Bass (1969) esitti tutkimuksessaan, että kirjallisuudessa luokitellut käyttäjäryhmät voidaan jakaa ensin kahteen ryhmään tuotteen käyttöönoton ajoituksen perusteella. Ryhmät (kuva 19) jakautuvat tällöin hänen esityksessään *innovaattoreihin ja seuraajiin* (Bass 1969). Bassin mukaan ensimmäiseen ryhmään kuuluvat innovaattorit ovat henkilöitä, jotka ovat ottaneet innovaation käyttöön ilman muiden henkilöiden päätöksiä. Bass esittää innovaattoreiden käyttäytymisen eroavan seuraajista. Hän kirjoitti innovaattoreihin eivät vaikuta muiden yksilöiden päätökset tuotteen käyttöönottopäätöksissä. Tämä siitä

huolimatta, vaikka he olisivat vuorovaikutuksessa keskenään. Tutkimuksessa esitettiin, että toiseen ryhmään (seuraajat) kuuluvat luokat 2-5 eli *varhaiset omaksujat, varhainen ja myöhäinen enemmistö sekä viimeiset käyttöönottajat*, jotka kaikki seuraavat muiden päätöksiä omista päätöksissään. (Bass 1969)

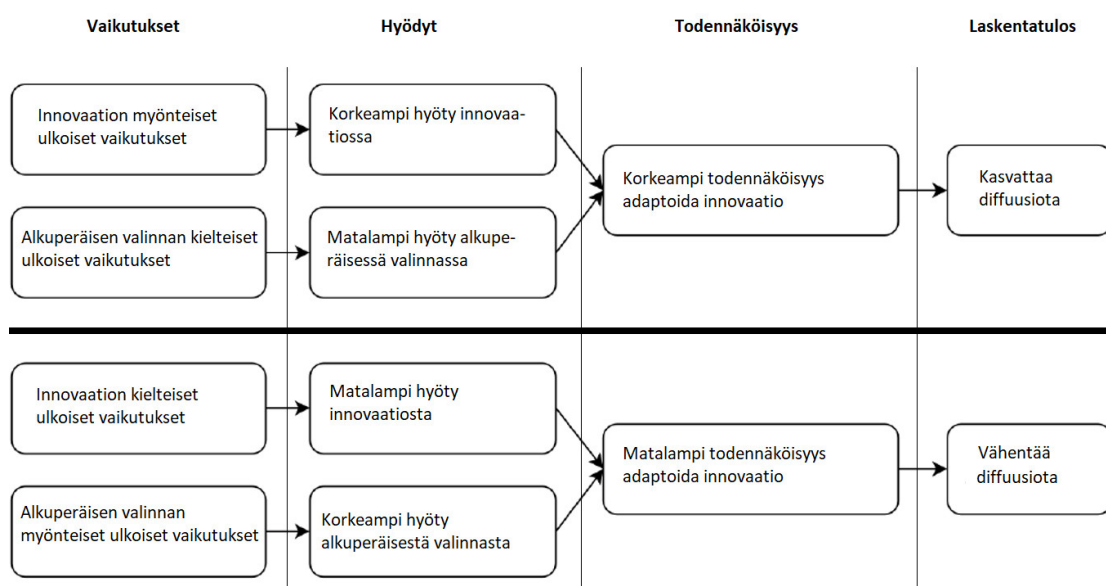


Kuva 19. Bassin esittämä kronologinen järjestys innovaatioadaptiolle, jossa on käyttäjät jaettu vain kahteen ryhmään (Bass 1969).

Harva teknologia on ollut ikuista ja usein tutkimukset keskittyvät teknologiadiffuusion kasvuun eli teknologian käytön hyväksyntään. Myös laskentamallit ovat usein kasvumalleja, joissa teknologia lähenee asympotoottisesti diffuusion saturaatiota. Teknologian kielteiset vaikutukset, mutta myös kilpaileva, syrjäyttävä tai edullisempi teknologia vähentää teknologian käyttöä ja kehittymistä. Poiketen useammista tutkimuksista Xiong *et al.* (2016) ottivat tutkimuksessaan huomioon teknologiadiffuusion negatiivisia vaikutuksia. He toivat esille, että suuri osa tutkimuksista on keskittynyt teknologiadiffuusion voimistaviin tekijöihin ja negatiivisesti teknologiadiffuusion vaikuttavia ulkoisia tekijöitä on tutkittu vähemmän. (Xiong *et al.* 2016)

Vaikka Bassin laskentamallia (1969) pidetään hyvänä perusmallina, on muissa laskentamalleissa tarkasteltu innovaatioidiffuusiota eri tarkastelunäkökulmista. Eri tutkimustulokset voivat vaihdella osittain johtuen tarkastelunäkökulmasta tai eri asioiden tutkimisesta. Tutkimustulokset eivät ole vertailukelpoisia keskenään, jolleivät ne käsittele asioita yhteismitallisesti. Xiong *et al.* (2016) tutkivat innovaatioiden leviämisessä *vertaisvaikutuksen teoriaa* ja simuloivat sitä tutkimuksessaan (Xiong *et al.* 2016). He tarkastelivat teknologian leviämistä tutkimuksessaan *vertaisvaikutuksen* kontekstissa (kuva 20) ja havaitsivat yksilöiden välisen vuorovaikutuksen ja vertaisvaikutuksen vaikuttavan merkittävästi teknologiadiffuusion (Xiong *et al.* 2016). Myös Bollinger ja Gillingham (2012) tutkivat vertaisvuorovaikutusta eli sosiaalista vuorovaikutuksen osuutta käyttöönottopäätöksissä. Heidän tutkimuksessaan keskityttiin aurinkopaneelien käyttöönottoon Kaliforniassa. Heidän mukaansa *uusissa tuotteissa sosiaalisen vuorovaikutukseen osuus* käyttöönottopäätöksessä on merkittävä tekijä (Bollinger & Gillingham 2012).

Xiong *et al.* (2016) vertaisvaikutustutkimuksen otokseen valittiin kaksi eri käyttäjäryhmää, joita tutkittiin omina kokonaisuuksinaan. Xiong *et al.* (2016) tutkimustuloksen mukaan muun muassa vertaisvaikutus on liian heikosti huomioitu innovaatioidiffuusiassa.



Kuva 20. Innovaatioidiffuusion ulkoisen vaikutuksen kaavio vertaisvaikutuksessa. Kaaviossa keskiviivan yläpuolinen osuus edistää innovaatioidiffuusiota ja alapuolinen osuus heikentää sitä. Lähde: Mukaelma Xiong *et al.* (2016) kuvasta ”Externality Effect in the Diffusion of Innovation”.

Tutkimustuloksien mukaan innovaatioidiffuusioon *vertaisvaikutuksen osalta* vaikuttavat kolme seuraavaa tekijää:

- tiedon vaikutus (information effect)
- kokemusten vaikutus (experience effect)
- ulkoisten tekijöiden vaikutus (externality effect).

(Xiong *et al.* 2016)

Koskien tietoisia valintoja innovaation tai teknologian käyttöönotosta, on tiedon vaikutus ensimmäinen tekijä innovaation adaptiossa vertaisvaikutuksen kontekstissa (Xiong *et al.* 2016). He perustelevat tutkimuksessaan väitettään sillä että, sellaista jota ei tiedä olevan olemassa ei voi valita käyttöön. Asiayhteyden irroittaminen kontekstista ja innovaation sekä teknologian suora vertaaminen johtaa helposti virheellisiin päätelmiin. Yleisesti on hyväksytty, että innovaatio ja teknologiaa voidaan tarjota myös osana palvelua. Tällöin käyttäjä ei tiedä hyödyntävänsä kyseistä teknologiaa. Teknologia ei tässä tapauksessa vaadi käyttäjältä erillistä päätöstä sen hyödyntämisestä eli käyttäjän tietoista teknologia-adaptiota. Uusissa teknologioissa, kuten sosiaalisessa mediassa hyödynnetään uusimpia teknologioita. Loppukäyttäjät eivät ole tehneet tietoista päätöstä monimutkaisten

teknologioiden adaptiosta. Käytön helppous ja koettu hyöty ovat vaikuttaneet uuden teknologian adaptioon.

Ulkoisten tekijöiden osalta vertaisvaikutusta voidaan arvioida kuvan 20 mallilla. Mallissa arvioidaan globaalien verkkojen, vaatimusten mukaisuuden sekä sosiaalisten normien ja asetusten vaikutusta innovaatioidiffuusioon. Tutkimuksessa huomattiin informaatiovaikutuksen osalta (Taulukko 4), että informaatiovaikutus riittää käyttöönottopäätökseen, jos innovaatio on helppo ottaa käyttöön. Havaittiin myös, että lisätieto kustannustehokkuudesta oli innovaation käyttöönottopäätökselle merkityksellinen. (Xiong *et al.* 2016)

Taulukko 4. *Vertaisvaikutuksissa olevia vaikutuskanavia. Mukaelma Xiong et al. (2016) Channels of Specific Peer Effects -taulukosta.*

	Informaatiovaikutus	Kokemuksen vaikutus	Ulkoinen vaikutus
Kanavat	<ul style="list-style-type: none"> tietoisuus asiasta tieto yleisesti kustannustehokkuudesta tieto kustannustehokkuudesta kyseisessä tapauksessa. 	<ul style="list-style-type: none"> ensimmäisten kokeilijoiden kokemukset, erityinen tekniikka, tietotaito tai osaaminen suorat mielipiteet käytettävissä olevat materiaaliressurit. 	<ul style="list-style-type: none"> henkilöiden välinen vaikutus verkoston paikallinen vaikutus globaalien verkkojen vaikutus vaatimusten mukaisuus eli oletus, että muutkin käyttävät (Band wagon effect) sosiaaliset normit eli kirjoittamattomat säännöt käyttämisestä.

Xiong *et al.* (2016) toteuttaman tutkimuksen tulos ei ole vertailukelpoinen Bassin diffuusiomalliin, koska Xiong *et al.* tutkivat innovaatioidiffuusion vertaisvuorovaikutuksia. Vertaisvuorovaikutus on innovaatioidiffuusiosta yksi tekijä. Xiong *et al.* (2016) mukaan innovaatioidiffuusion aikaisessa vaiheessa vaikuttaa tiedon siirtyminen sosiaalisessa vuorovaikutuksessa. Vertaisvaikutuksen keskivaiheessa saavutetaan kriittinen massa, jolloin innovaation leviämiseen vaikuttavat kokemukset. Ulkoinen vaikutus on vertaisvaikutuksista viimeisessä vaiheessa (taulukko 4). (Xiong *et al.* 2016)

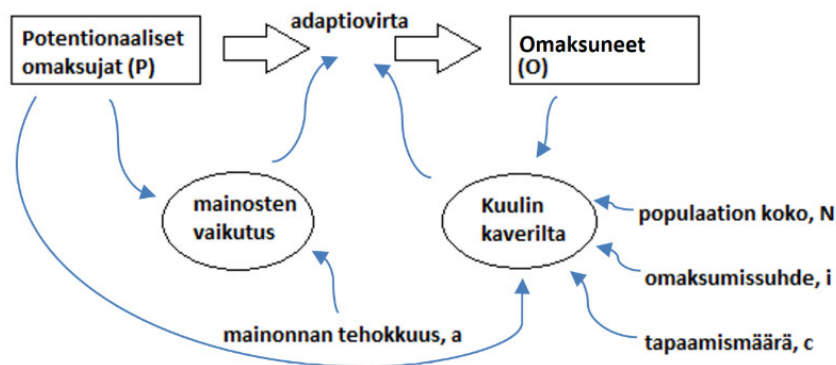


Kuva 21. *Arvio Xiong et al. (2016) tutkimuksen teknologian adaptiovaikutusten kronologinen järjestys vertaisvaikutuksessa.*

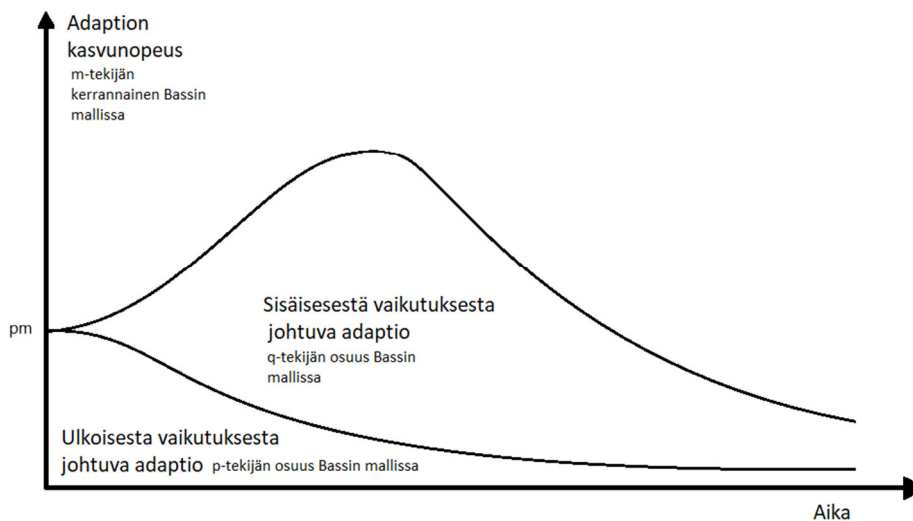
Xiong *et al.* (2016) tutkimuksessa esitettiin järjestys vertaisvuorovaikutuksen osalta. Kuvassa 21 on kuvattu kronologisessa järjestyksessä vertaisvuorovaikutuksen adaptiovaikutukset. Xiong tutki vertaisvaikutusta myös toisessa tutkimusryhmässä (Xiong, Puqing & Bobashev 2018). Tässä tutkimuksessa tuotiin esille, että *monimutkaisten sosiaalisten verkkojen* vaikutusta innovaatioidiffuusioon ei ole tutkittu riittävästi (Xiong *et al.* 2018).

Innovaatiodiffuusio on tarvinnut toistaiseksi ihmisiä levitäkseen. 'Theory of reasoned action' (TRA) ja sen laajennos 'Theory of planned behavior' (TPB) kuvaavat ihmisen käyttäytymistä tarkoituksenmukaisessa tai suunnitellussa toiminnassa. Ajzen ja Fishbein tutkivat ihmisen käyttäytymistä, jolla osaltaan selitetään teknologian hyväksymistä (Trafimow 2009). Teknologian hyväksyminen vaikuttaa merkittävästi teknologiadiffuusioon. Tässä työssä ei keskitytä teknologian hyväksymismalleihin.

Forsström (2011) käsittelee tutkimuksessaan ulkoisen vaikutuksen ja vertaisvaikutuksen eroja. Hän esittää kuvan 22 avulla ulkoisen mainonnan "Mainosten vaikutus" sekä vertaisvaikutuksen eron eli "kuulin kaverilta". Kuva 22 on havainnollinen tapa osoittaa Bassin laskentamallin vaikutus adaptiovirtaan.



Kuva 22. Bassin laskentamalli, jossa on huomioitu populaation määrä. Mukaelma Forsström (2011) esittämästä matemaattista mallia esittävästä kuvasta.

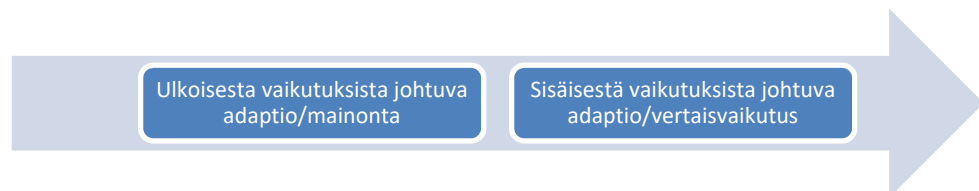


Kuva 23. Mukaelma ulkoisen ja sisäisten tekijöiden vaikutuksen huomioimisesta Bassin laskentamallissa (Mahajan et al. 1995). Alussa adaptio johtuu ainoastaan innovaattoreiden osuudesta ja mahdollisesta adaptioijien määrästä.

Kuvat 22 ja 23 täydentävät toisiansa. Kuvan 23 informaatiota on tässä työssä lisätty Bassin laskentamallin tekijä m sekä p ja q -parametreilla, jotta eri tekijöiden merkitys ja vaikutukset ovat selvemmin kuvattavissa. Kuvan 23 vasemmassa reunassa on pm -merkintä,

joka kertoo lähtötilanteesta sen, että seuraajilla ei ole alussa vaikutusta. Kuvista 22 ja 23 voidaan havaita Bassin laskentamallin periaate. Kuvassa 22 esitetyt P ja O -muuttujat eivät ole p ja q -parametreja.

Kuvassa 24 on järjestetty ulkoisen vaikutuksen ja vertaisvaikutukset kronologiseen järjestykseen. Kuvan 22, 23 ja 24 perusteella mainonnan vaikutus on merkittävä diffuusion alkuvaiheissa.



Kuva 24. Arvio innovaatioadaption kronologisesta järjestyksestä Bass (1969) mallissa Mahajan *et al.* (1995) sekä Forsström (2011) lähteitä hyödyntäen.

Myös Wang *et al.* (2017) mainitsevat Bassin mallin etuna olevan yleiskäyttöisyys ja yksinkertaisuus. He esittävät, että Bassin malli sopii kuvaamaan liike-elämän, uusien tuotteiden ja teknologian ennustamista. Bassin alkuperäinen malli on perusmalli, joka voi huomioida riittävästi kaikkia tekijöitä. Bassin mallissa on kuitenkin käsitelty asiaa monipuolisesti, jota on myös empiirisesti testattu kuluttajatuotteilla (Bass 1969). Useammassa tutkimuksessa nykyäänkin hyödynnetään modifioitua Bassin mallia, joka huomioi tietyn tarkastelunäkökulman. Wang, Dang ja Pei (2011) kehittivät Bassin mallia ja lisäsivät malliin osan, joka huomioi markkinoiden koon. Tätä mallia kutsutaan Grey Bass Model -nimellä. Grey etuliite tulee Grey järjestelmäteoriasta, joka on suunniteltu pienien otoksien mallintamiseen. (Wang *et al.* 2017)

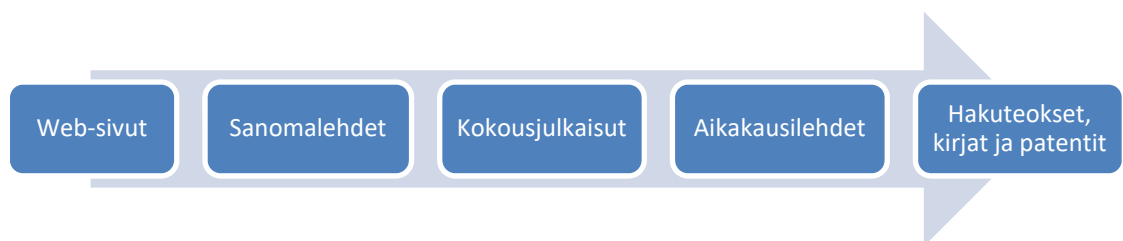
2.5 Diffuusion kehittymisen ennakkoinnin tietolähteiden ajantasaisuus

Diffuusion ennakkointiin käytetty lähde sijoittuu kronologisesti lähteelle ominaiseen ajanjaksoon. Sassali (2014) käsitteli eri julkaisulähteiden ajantasaisuutta ja esitti taulukon 5 mukaista järjestystä tiedon julkaisuviiveille.

Taulukko 5. *Informaatiokanavien julkaisuviiveet Oulun yliopiston tiedonhallintakurssimateriaalin mukaan (Sassali 2014).*

Julkaisutyyppi	Reaaliaikaisuus
web-sivut	saman päivän aikana
sanomalehdet	yksi päivä
kokousjulkaisut	viikko
aikakausilehdet ja tieteelliset aikakausilehdet	kuukausista vuosiin
hakuteokset, kirjat ja patentit.	useita vuosia.

On hyvä huomata, että Sassalin (2014) sekä Chon ja Daimin (2016) näkemykset julkaisujärjestyksestä (kuva 25 ja kuva 26) ovat yhdensuuntaisia, vaikka vaiheiden käyttötarkoitus eroavat toisistaan. Molempien mukaan patentit ja tarkemmat viimeistellyt julkaisut sisältävät kronologisesti suurimman viiveen julkaisuajankohdiltaan. Web-sivutkaan eivät ennusta tulevaisuutta saman tiedon osalta, koska ovat tietolähteenä lähes reaaliaikaisia. Web-haut saattavat ennustaa toisen ilmiö esiintymistä, kuten hypeilmiötä. Chon ja Daimin (2016) mukaan perustutkimuksen määrä voi ennakoida tulevaa soveltavien tutkimusten määrää sekä patenttien määrää. Sassalin (2014) tarkastelunnäkökulma (taulukko 5) on informaation jakelu ja hän esittääkin erilaisien informaation jakelukanavien huomioivan kohderyhmän tarpeita ja informaation reaaliaikaisuutta.

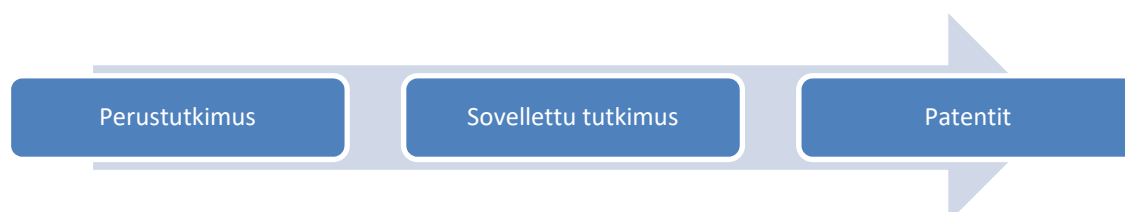


Kuva 25. *Tulkinta Sassalin (2014) esittämän näkemyksen kronologisesta järjestyksestä informaation jakelun kontekstissa.*

Useamman lähteen yhtäaikainen hyödyntäminen voi parantaa diffuusion ennakkoinnin tarkkuutta. Teknologian tarkalle ennakkoinnille on suuri merkitys menestyvälle liiketoiminnalle. Lee, Choi ja Sohn (2018) tutkivat teknologiadiffuusiota patenttitietokannan ja web-hakuliikenteen sisällön perusteella. Tutkimuksessaan pelkäävät web-hakutulosten

osalta Bassin diffuusioennustusmallilla ei saatu tarkkoja tuloksia (Lee *et al.* 2018). Patenttietokannan huomioiminen web-hakuliikenteen kanssa yhtäaikaaisesti Bassin diffuusiomallissa toi ennakointiin lisää tarkkuutta (Lee *et al.* 2018). Sassali (2014) sekä Cho ja Daim (2016) ovat todenneet patenttien hakuprosessin sisältävän viiveitä. Lee *et al.* (2018) pyrkivät tutkimuksessaan yhdistämään eri käytettävien tietolähteiden parhaita puolia omalla mallillaan eli tuomaan patenttietokannan tarkkuuden web-hakujen ajantasaisuuden yhteyteen. Heidän tutkimuksessaan saavutettiin tuloksena yhtäaikaisesti nopea sekä tarkka teknologiaennustus kahdesta eri teknologiasta, joista toinen teknologia edusti teollisuutta ja toinen kuluttajamarkkinoita (Lee *et al.* 2018). Lee *et al.* (2018) tutkimuksen tuloksista tulee huomioida, että tutkimus toteutettiin tietyssä markkina-alueessa ja kontekstissa, jolloin tulokset eivät välttämättä ole sovellettavissa suoraan muualle. (Cho & Daim 2016; Lee *et al.* 2018; Sassali 2014)

Cho ja Daim (2016) kirjoittivat, että tutkimuksessa on hyvä huomioida tutkimuksen konteksti. Cho ja Daim (2016) tutkivat 'Light-emitted diod' (LED) -näyttötekniikan kehittymistä. Heidän tutkimuksessaan hyödynnettiin patenttietokantoja, koska patenttietokanta on teknologiateollisuudessa yleisesti hyödynnetty tietolähde. Cho ja Daim (2016) käyttivät taulukon 6 tietolähteitä tutkimuksessaan. Taulukossa 6 kuvatut tietolähteet on järjestetty kronologiseen järjestykseen kuvaan 26. Kuvassa 26 perustutkimus on asetettu ajallisesti ensimmäiseksi kronologiseksi kehittyväksi vaiheeksi näyttötekniikan 'An organic light-emitting diode' -tekniikan (OLED) yhteydessä.



Kuva 26. Tulkinta Chon ja Daimin (2016) tutkimuksen tietolähteiden kronologisesta järjestyksestä tutkimuksen ja kehittämisen näkökulmasta, jossa huomioidaan tutkimus- ja kehitystoiminnan vaiheet.

Cho ja Daim (2016) esittivät tekniikan ilmentymisen määrien kehittymistä hyödyntämällä Bassin laskentamallia (S-muotoisilla käyrillä). Tutkimustuloksissa verrattiin eri tietolähteiden välistä ajallista erotusta (aika/vuotta). Kuvassa 27 on esitetty ajallista viivettä perustutkimuksen ja Yhdysvaltojen patenttitoimiston patenttietojen (USPTO) välillä. Diffusion kehittymistä kuvaavien käyrien ajallinen ero oli noin 5-6 vuotta. Cho ja Daim (2016) toistivat tutkimuksen, jossa huomioitiin useita patenttietokantoja lähteenä. Patenttietokantoja valittiin Euroopan, Kansainvälisen, Kiinan, Japanin ja Korean -patenttitoimistoista (EPO, CPTO, JPO, KIPO ja PCT). Heidän tutkimuksessa tulokset eivät juuri muuttuneet eli useamman patenttietokannan sekä perustutkimuksen välillä ollut ajallinen ero säilyi. (Cho & Daim 2016)

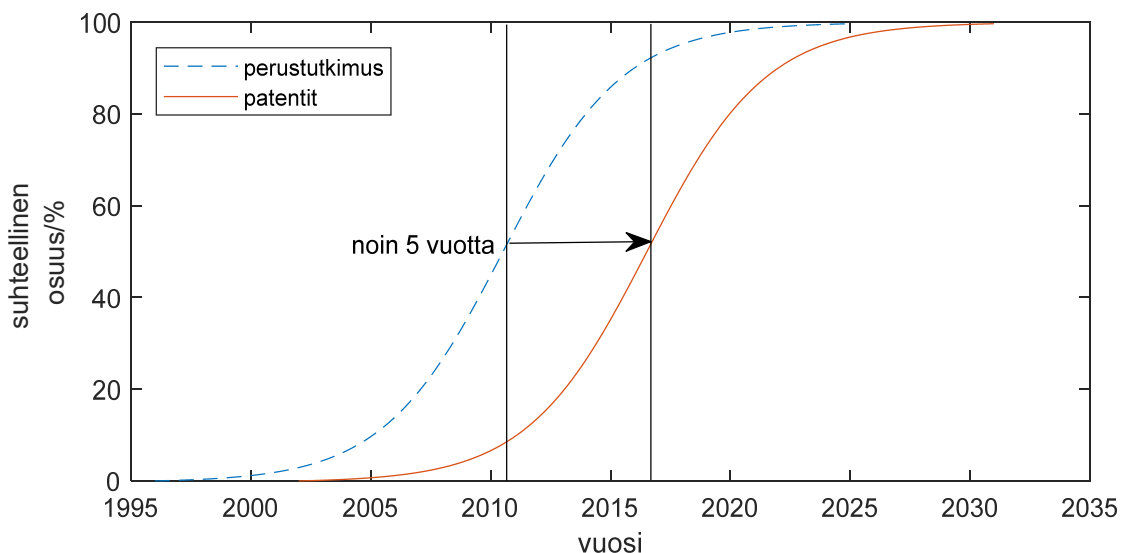
Taulukossa 6 on esitetty Chon ja Daimin (2016) tutkimuksessa käytetyt tietolähteet. Perustutkimuksen he ovat arvioituneen indikoituvan Web of science -artikkelitietokannan

tieteellisen viittausindeksin perusteella. Vastaavasti kehittämistä he arvioivat patenttitietokannan indeksien perusteella.

Taulukko 6. Cho ja Daim (2016) teknologiaennustamisessa käytetyt tietolähteet.

T & K:n vaihe	Tyypillinen lähde	Tietokanta
Perustutkimus	Tieteellisten viittausten indeksi	Web of science
Sovellettu tutkimus	Tekniikan alan indeksi	Compendex
Kehittäminen	Patentit	USPTO, PCT, EPO, JPO, KIPO, CPTO

Kuvassa 27 on mallinnettu Chon ja Daimin (2016) esittämä laskentatulokset. Laskenta perustuu Bassin laskentamalliin (Liite E). Laskentamallin lähtötiedot, kuten lähtö-, loppuvuodet sekä p ja q -parametrit on iteroitu laskennalla Chon ja Daimin (2016) tutkimustuloksien kuvan avulla. Laskennalla varmistettiin aikaisemmin esitetyn mallin soveltaminen ja sen soveltuvuus Chon ja Daimin (2016) mallintamiin tutkimustuloksiin.



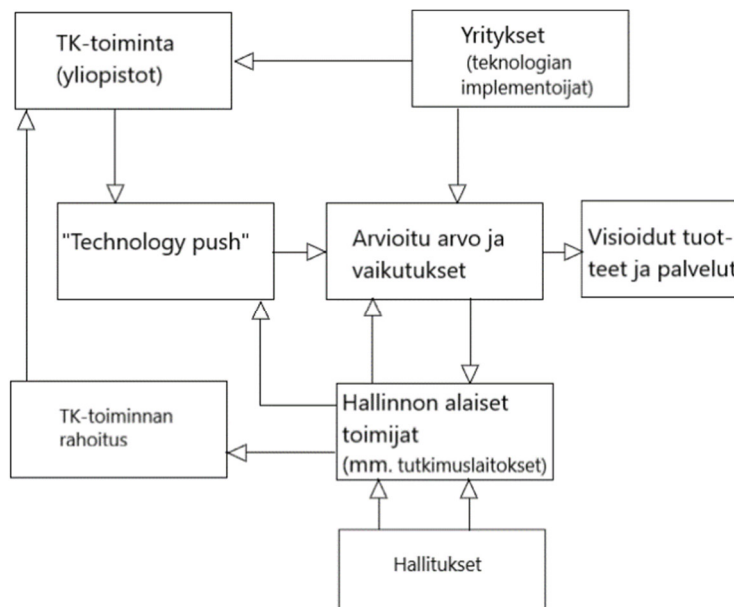
Kuva 27. Matemaattinen mallinnus OLED-tekniikan kehittämisestä. Mallissa havainnollistetaan perustutkimuksen- ja patenttitietokannan perusteella saatujen poimintojen ajallisia eroja. Lähtötiedot ovat saatu Chon ja Daimin (2016) tutkimuksen kuvatulokinnan sovituksella. Bassin mallin mukainen laskenta on esitetty liitteessä (Liite E).

Kuvassa 27 on simuloitu LED-tekniikkaa korvaavan OLED-tekniikan diffuusion kehittymistä Bassin laskentamallilla. Korvaavan tekniikadiffuusion ennakoitiin sopii paremmin Fisher-Pry ennustusmalli. Cho ja Daim (2016) esittävätkin, että OLED-tekniikadiffuusion arviointiin on hyödyllistä vertailla Fisher-Pryn sekä Bassin kasvumallin soveltuvuutta. Tarkempi Chon ja Daimin (2016) esittämien tulosten tarkastelu osoittaa, että kuvassa 27 kasvukäyrien välinen ero säilyy noin viitenä vuotena. Kasvuvaiheessa

olevien teknologiadiffuusiota kuvaavien S-käyrien muodot korreloivat toisiansa ja aika-ero S-käyrien välillä säilyi lähes yhtä suurena eri vaiheissa. Korreloinnista huolimatta perustutkimuksen, sovelletun tutkimuksen tai patenttitietokantaan perustuvan tutkimuksen S-muotoisien kasvukäyrien muodot vastaavat vain visuaalisesti toisiansa. Viiveet vaihtelevat teknologiadiffuusion eri ajankohtina, kuten myös teknologian varhaisessa vaiheessa. Kasvukäyrien erilainen kehittyminen johtuu innovaattoreiden ja seuraajien osuuksista sekä potentiaalisten adaptiivien määristä. Kuvasta voidaan myös huomata, että patenttitietokannoissa ei juuri ole patenteja, kun perustutkimuksen määrä on ylittänyt jo 20 prosentin osuuden saturaatioon tarvittavasta määrästä. Chon ja Daimin (2016) tutkimuksessa aineistojen yhtäaikainen käyttö tarkensi patenttien määrän kehittymisen ennustusta. Vuoden 2014 kohdalla perustutkimuksen määrä on vakiintumassa, mutta patenttien määrä (kaupallisesti suojattavat ratkaisut) ovat investoinnin kannalta erittäin otollisessa vaiheessa, jolloin diffuusio on kiihtyvästi kasvava ja lähestytään kriittistä massapistettä.

Cho ja Daim (2016), Lee *et al.* (2018) ja Huang *et al.* (2018) tutkivat ennakointimenetelmillä, joissa hyödynnetään uuden teknologian ilmenemistä varhaisessa vaiheessa. Näissä tutkimuksissa hyödynnettiin useita tietolähteitä yhtäaikaaisesti. Näissä tutkimuksissa hyödynnettiin teknologian ennustamisen lähteinä muun muassa

- perustutkimuksien määrä: Cho ja Daim (2016)
- web-hakuliikenteen määrä: Lee *et al.* (2018)
- markkinadata ja tutkimusrahoituspäätöksien määrä: Huang *et al.* (2018).



Kuva 28. Mukaelma teknologian kehittymisen mallista (Huang *et al.* 2018).
Tutkimuksessa esitettiin rahoituksen vaikutus teknologian kehittymiseen.

Huang *et al.* (2018) esittivät tutkimuksessaan mallin, jossa kuvataan kokonaisuutta teknologian kehittämisestä (kuva 28). Kuvan 28 mallissa osateknologiat aiheuttavat riippu-

vuuden teknologian kehittymiseen ja erityyppisillä rahoituksilla on merkittävä rooli teknologian kehittymisille. Riippumattomuuden säilyttäminen tutkimuksessa voi vaikeutua ulkopuolisen kaupallisen toimijan rahoittaessa tutkimusta. Toisaalta ilman rahoitusta ei tutkimusta välttämättä tehdä ja perusteita kehitystoimintaan ei saavuteta. Rahoitusta kohdentamalla tietyn teknologian suosiminen mahdollistuu ja tutkimustulokset voivat olla myös kaupallisesti paremmin hyödynnettävissä, joka kasvattaa teknologiadiffuusiota.

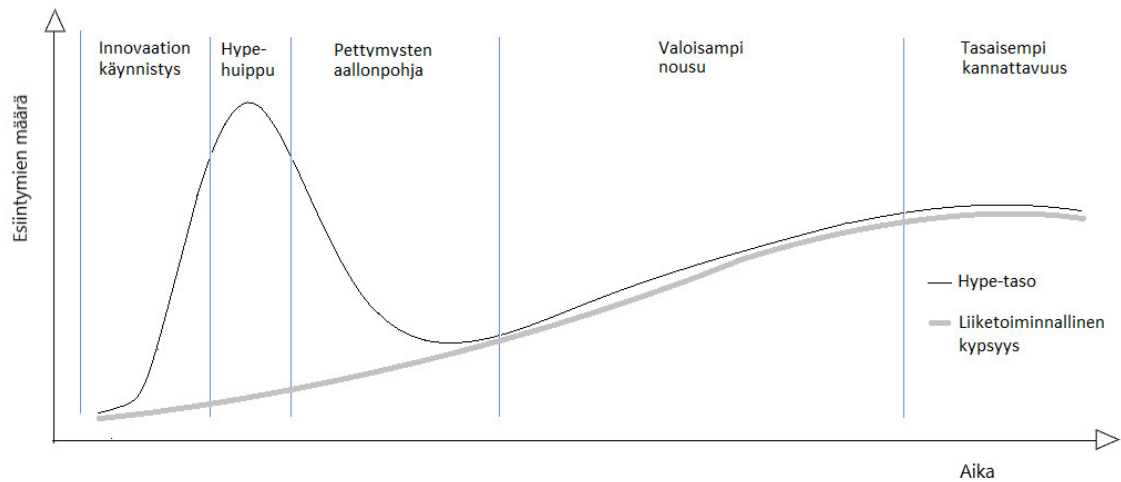
Davenport (2014) esitti eri tavoilla uusien teknologioiden hypeä, jossa uusi teknologia halutaan liittää voimakkaasti kehittyvään teknologiaan. Hypeä usein esitetään Gartnerin hypekäyrällä (kuvat 29 ja 30), jonka hypeilmioiteorian (Amaran laki) kantaisänä on pidetty Roy Amaran (1925–2007) esittämää teoriaa. Amaran laissa lyhyellä aikajaksolla odotuksemme ovat alussa toteumaa suuremmat, mutta pitkällä aikajaksolla odotukset ovat kuitenkin toteumaa pienemmät (Ratcliffe 2016).

Gartner tutkii kaupallisena yrityksenä tulevia ilmiöitä. Gartnerin hypekäyrä (kuva 29) on uusien teknologioiden osalta tunnettu esitystapa ja kuvaa teknologioiden ilmentymistä ajan funktiona. Brescanin ja Epplerin (2010) mukaan hypekäyrän (hype cycle) ensimmäinen esittäjä oli Jackie Fenn (1995) työskennellessään Gartnerilla vuonna 1995. Linden ja Fenn (2003) esittivät, että hypekäyrän ensimmäisillä vaiheilla kuvataan enemmän ihmisten asenteita, kuin teknologian kehittymistä. Heidän mukaansa Gartner jakaa yritykset teknologian adaptoinnin perusteella kolmeen seuraavaan tyyppiin:

1. Tyypin A yritykset ovat aggressiivisia riskinottajia, jotka ottavat teknologiat ensimmäisinä käyttöön.
2. Tyypin B yritykset seuraavat tyypin A yrityksiä, säilyttäen teknologisen edun kilpailijoihin nähden.
3. Tyypin C yritykset toimivat mukavuusalueellaan ja menettävät mahdollisuuksia kilpailla uusista teknologioista.

Linden ja Fenn (2003) esittävät, että riskiä ottavat aggressiiviset tyypin A yritykset voivat aloittaa teknologian *pilotoinnin* jo hypehuipussa. Optimiajankohta tyypin B yrityksille välittömästi hypehuipun loputtua ja tyypin C yrityksille vasta uuden nousun alkaessa (Linden ja Fenn 2003).

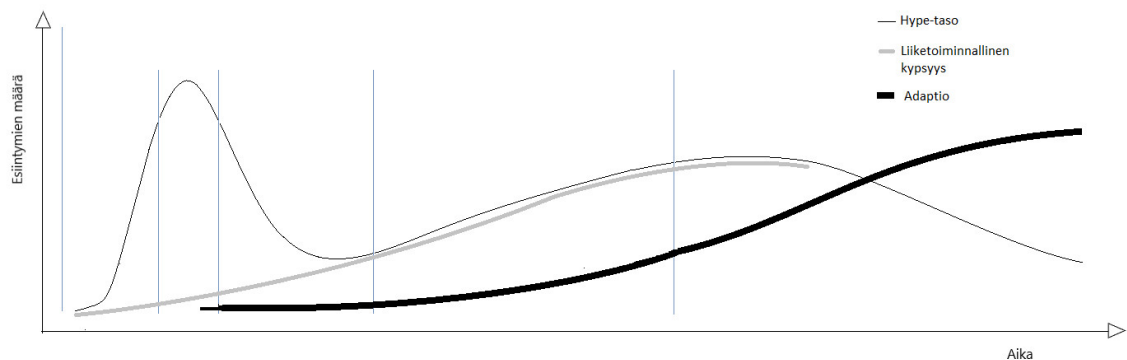
Riskiä ottavat tyypin A yritykset aloittavat teknologioiden käyttöönotot hypehuipun (kuva 29) jälkeen (Linden ja Fenn 2003).



Kuva 29. Mukaella Gartnerin esittämästä hypekäyrästä, jolla kuvataan tulevia teknologioita (Bresciani & Eppler 2010; Gartner 2017).

Lindenin ja Fennin (2003) mukaan teknologia tai innovaatio käy läpi hypeen liittyviä vaiheita ennen tuotantoon tuloa (kuva 29), joita ovat

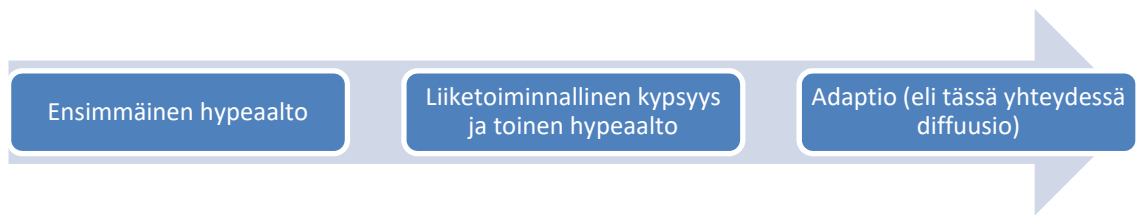
- innovaation käynnistys eli teknologian löytyminen
- hypehuippu
- pettymysten aallonpohja
- valoisampi nousu
- tasaisempi kannattavuus.



Kuva 30. Gartnerin usein esittämän laajennetun hypekäyrän ja teknologian adaption välinen ajallinen riippuvuus. Kuva on mukaella Brescianin ja Epplerin (2010) sekä Lindenin ja Fennin (2003) tutkimuksista.

Linden ja Fenn (2003) muistuttavat tutkimuksessaan, että tyypin A yritykset valitsevat varhaisessa vaiheessa käyttöönotettavat teknologiat huolellisesti. He esittävät, että vain strategisesti tärkeitä teknologioita tulee ottaa käyttöön varhaisessa vaiheessa. Linden ja Fenn (2003) tarkentavat yritystyyppien B ja C osalta, että nämä yritykset harkitsevat tek-

nologian varhaista omaksumista, jos varhainen teknologian adaptointi on liiketoiminnallisesti olennaista. Ensimmäisestä hypeaalosta voi kulua pitkä aika teknologia-adaption alkamiseen. Kuvassa 30 on esitetty hypekäyrän vaiheiden mukaelma, joiden kronologinen tulkinta esitetään kuvassa 31. Kuvaan 30 on lisätty arvio adaption kronologisesta sijainnista suhteessa laajennettuun hypekäyrään. Kuvassa 29 ja 30 on yleistetty aaltojen vaiheita. Nopeasti tai hitaasti kehittyvillä teknologioilla aallot ovat eri muotoisia. Linden ja Fenn (2003) sekä Brescian ja Eppler (2010) esittivät kolmelle eri yritystyyppille (A, B ja C) eri adaptioajankohdat, jolloin oikea adaptioaika riippuu yrityksen tyyppistä ja otettava riskistä.



Kuva 31. *Kronoginen tulkinta vaiheista sekä hypeaaltojen kronogisesta sijainnista (Bresciani & Eppler 2010; Linden & Fenn 2003).*

Teknologiat ovat kehittymisen nopeuden mukaan jaettavissa taulukon 7 mukaisesti nopeasti ja hitaasti kehittyviin teknologioihin (Linden & Fenn 2003).

Taulukko 7. *Teknologioiden jakaminen hitaasti ja nopeasti kehittyviin.*

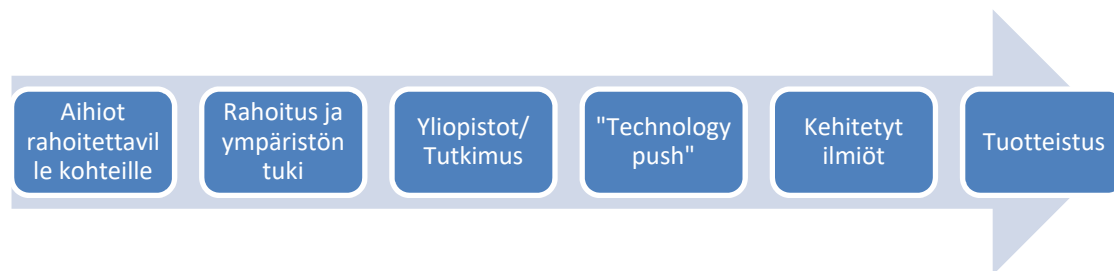
Nopeasti kehittyvien teknologioiden indikaattoreita (fast track)	Hitaasti kehittyvien teknologioiden indikaattoreita (long fuse)
Teknologian korkea arvo	Lähtökohtaisesti monimutkainen teknologia, joka edellyttää teknologian osaamista (esim. kvanttilaskentaa)
Yksinkertainen hyödyntäminen käyttäjillä ja organisaatioissa	Käyttäjien hyväksyntä tai säätelykysymykset (biometriikka, koneellinen päättely ja digitaaliset allekirjoitukset)
Useampia vahvoja toimittajia, jotka tukevat teknologiaa	Uusien infrastruktuurien tai ekosysteemien käyttöönotto, jotka tarvitsevat aikaa kehittymiseen (esim. digitaaliset allekirjoitukset ja julkisen avaimien infra)
Nykyisen infrastruktuurin hyödyntäminen	Ammattitaidon katvealueet (esimerkiksi tiedonhankinta, simulointi tai monimutkainen suunnittelu)
Nopea siirtymä kuluttajien käytöstä yritysten käyttöön.	Suuret muutokset liiketoimintaprosesseihin tai uuden liiketoimintamallin luominen (esimerkiksi asiakkuuksien hallinta).

(Linden ja Fenn 2003).

Yleistä *nopeasti kehittyvistä* teknologioiden osalta hypekäyrän muoto on erilainen ja hypehuipun jälkeen ”pettymysten aallon pohja on lyhyt”. Sundqvist *et al.* (2002) tutkivat teknologiadiffuusion kehittymistä langattomien verkkojen yhteydessä. Tutkimuksessaan

he esittivät valmiin infrastruktuurin yhteydessä verkkovaikutuksen nopeuttavan teknologian kehittymistä voimakkaasti (Sundqvist *et al.* 2002).

Tutkittavan teknologian kontekstilla on merkitystä. Tutkimuksessaan Cho ja Daim (2016) huomioivat teknologialähtöisyyden sopivien tietolähteiden valinnassa (Taulukko 6). Vastaavasti Huang *et al.* (2018) hyödynsivät big dataan liittyvässä tutkimuksessa taulukon 7 mukaisia lähteitä teknologian ennustamiseen (Huang *et al.* 2018).



Kuva 32. Arvio Huang *et al.* (2018) tutkimuksen kohteiden kronologisesta järjestyksestä.

Huomioiden Huang *et al.* (2018) esittämän (kuva 28) kaavion järjestyksen, voidaan teknologian kehittymistä arvioida kuvan 32 mukaisesti, jossa on tulkittu kaaviossa esitetyt asioiden kronologista järjestystä Huang *et al.* (2018) mukaan.

Taulukko 8. Teknologiaennustamisessa käytetyt tietolähteet (Huang *et al.* 2018).

Tieto	Tietolähde	Määrä (knl)
Ohjeet/käytännöt	Valkoinen talo (https://www.whitehouse.gov)	~10
Tieteelliset julkaisut	Web of Science (http://apps.webofknowledge.com)	1599
Patenttihakemukset	Thomson Innovation (https://www.thomsoninnovation.com)	765
Rahoituspäätökset	National Science Foundation (https://www.nsf.gov)	830
Kaupalliset tietueet	ProQuest (http://search.proquest.com)	9997

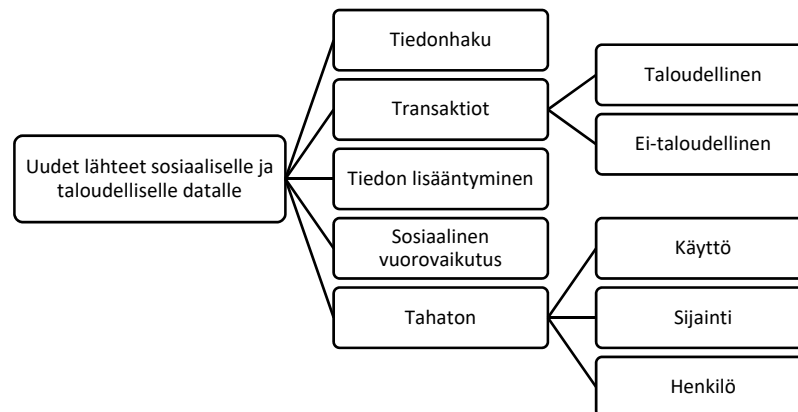
Taulukossa 8 on esitetty Huang *et al.* (2018) tutkimuksessa käytettyjä tietolähteitä. Huang *et al.* (2018) sekä Chon ja Daimin (2016) tutkimuksien hyvinä puolina voidaan todeta selkeä esitys eri tietolähteiden kronologisesta järjestyksestä ja tietolähteiden saatavuudesta.

2.6 Big data -ennustusmenetelmien tarkempi tarkastelu

Big datan hyödyntäminen on saavuttamassa merkittävää asemaa analytiikassa ja ennustamisessa. Blazquez ja Domenech (2018) tutkivat teknologian ennakointiin sopivia tietolähteitä. He kirjoittivat tiedon määrän nopean kasvamisen ja rakenteettomuuden tuovan

ennustettavuutta. Tutkimuksessaan he esittivät sosiaalisten ja kaupallisten lähteiden sopivuuden ennakointiin. Big datassa voidaan yhdistellä suuri määrä rakenteettomia ja erilaisia datalähteitä, joka mahdollistaa uusien asioiden ennustamisen. Käsityönä suurien rakenteettomien datamäärien yhdisteleminen on mahdotonta. Big dataa on hyödynnänyt myös Liu, Tseng, F. ja Tseng, Y. (2018) sekä Schuelke-Leech (2018). Näissä tutkimuksissa tuodaan esille monimutkaisuus ennakoinnissa. Tutkimuksissa hyödynnetään big dataan perustuvia ennustamisteknologioita. Big dataa on hyödynnetty myös tutkimuksessa, jossa Roßmann, Canzaniello, von der Gracht ja Hartmann (2018) tutkivat big data-ennustusmenetelmiä. Heidän tutkimuksessa big data-analytiikka (BDA) oli keskeinen ennustamiseen käytetty menetelmä.

Schuelke-Leech (2018) esittää tutkimuksessaan tietotekniikan, materiaali- ja bioteknologian välisiä riippuvuuksia. Huomionarvoista on, että big datalla on tässä esityksessä riippuvuus useisiin uusiin keskeisiin ilmiöihin ja teknologioihin (Schuelke-Leech 2018). Tutkimuksessaan he esittivät big datalla olevan riippuvuus autonomisiin järjestelmiin, tekoälyyn, kuten koneoppimiseen sekä sosiaaliseen mediaan, liikenteeseen ja ennakointiin (Schuelke-Leech 2018). Big datalle ominainen ”pienen” datan tallentamisen ja siirtämisen helppous osaltaan on mahdollistanut uusien tietolähteiden hyödyntämisen osana profiloituja käyttäjäryhmiä, teknologioita tai muita ilmiöitä. Esimerkkinä ”pienestä” rakenteettomasta datasta on esitetty Blazquezin ja Domenechin (2018) tekemässä tutkimuksessa, jossa rakenteeton data pyritään koneellisesti luokittelemaan kuvan 33 rakenteeseen.



Kuva 33. Mukaelma Blazquez ja Domenech (2018) esittämästä rakenteettoman datan luokittelusta.

Kuvassa 33 on esitetty mukaelma heidän esittämästään luokittelusta, joka on rakenteinen ja voidaan hyödyntää big dataan pohjautuvassa ratkaisussa (Blazquez & Domenech 2018).

Viime vuosina Big dataan perustuvia ennustusmenetelmiä on tutkittu useissa tutkimuksissa. Alla olevassa taulukossa 9 on koostettu esimerkin vuoksi viisi eri big data -tutkimusta tuloksineen.

Taulukko 9. *Satunnaisia big data -ennusmenetelmiin liittyviä tutkimuksia.*

Tutkimus	Tekijät	Tutkimustulos
Forecasting with Big Data: A Review	Hassani ja Silva (2015)	Big data tuo tulevaisuudessa mahdollisuuden tehdä tarkkoja ja tuottoisia ennusteita, mutta se vaatii osaamista uusista menetelmistä, joka on yksi suurimmista haasteista
Concept and benchmark results for big data energy forecasting based on Apache Spark	Ordiano et al. (2018)	Tutkimuksessa verrattiin Apache Sparkin ja R-ympäristön laskentatehoa ja todettiin big datan käyttämisen ennustamiseen saattavan kannattaa. Tuloksessa todettiin Apache Spark -klusterin tuovan huomattavaa etua laskentatehoon, mutta riittävyttä verkon yli tapahtuvaan suorakäyttöön ei voitu tämän perusteella arvioida
Foreign arrivals nowcasting in Italy with Google Trends data	Antolini ja Grassini (2018)	Tuloksena todettiin, että Google Trends -palvelu ennustamisella voidaan ennustaa turistivirtoja, jotka saapuvat Italiaan, kuten myös päiviä, joita he viettävät Italiassa.
The effect of interest in renewable energy on US household electricity consumption: An analysis using Google Trends data	Park ja Kim (2018)	Tutkimuksessa verrattiin sähkön kulutuksen ja "renewable"-hakusanojen välistä suhdetta hyödyntämällä Google Trends -palvelua. Tuloksena havaittiin, että kotitaloussähkönkulutus vähentyi 16,017 miljoonalla kWh:lla jokaista "renewable"-hakusanan yhden yksikön nousua kohden.
Combining official and Google Trends data to forecast the Italian youth unemployment rate	Naccarato, Falorsi, Loriga ja Pierini (2018)	Tutkimuksessa verrattiin tilastollisesti ja Google Trends-palvelun avulla saatuja nuorten työttömien määriä ja todettiin muun muassa Google Trends -palvelun tuottavan tilastollista ennakointia pienempiä virheitä.

2.7 Teknologian kehittymisen aaltoliikkeet

Talouden aaltoliikkeet voivat olla pitkän ajan kuluessa tapahtuvia, joka muuttaa teknologioiden kehittymistä. Suurien talouden aaltoliikkeiden mahdollisen olemassaolon on esittänyt makroekonomiaa tutkinut Nikolai Kondratieff (Kondratieff & Stolper 1935). Kondratieff (1926) aloitti suuria talousaaltoja käsittelevät keskustelut tutkimuksellaan "Die langen Wellen der Konjunktur" (Kondratieff 1926; ref. Kondratieff & Stolper 1935). Aikaisemmin oli esitetty kaksi lyhyempää talouden aaltoliikettä eli Juglarin ja Kitchin aallot (taulukko 10). Stolper käänsi Kondratieffin artikkelin englanniksi. Artikkelissa todetaan Kondratieffin olleen kiinnostunut talouden aaltojen olemassa olosta jo vuosina 1919-1920 (Kondratieff & Stolper 1935). Talousaaltojen vaikutuksia innovaatioiden, patenttien ja teknologian kehittymisen määrään on tutkittu useiden tutkijoiden toimesta. Schumpeter

(1939) tutki suuria talousaaltoja 1787-1938 väliseltä ajalta (Schumpeter 1939; ref. Blackburn 1939; ref. Kuznets 1940). Blackburn (1939) sekä Kuznetz (1940) ovat analysoineet Schumpeterin vuonna 1939 tekemää tutkimusta. Blackburn ja Kuznetz jakoivat yhteisen näkemyksen talouden aaltoliikkeistä (Taulukko 10), jossa aaltoliikkeet on jaettu kolmeen Schumpeterin ja Kondratieffin esittämiin aaltoihin. Schumpeter jakoi talouden aallot 1) pitkiin alun perin Kondratieffin esittämiin aaltoihin, 2) keskipitkiin Juglar ja 3) lyhyisiin Kitchin aaltoliikkeisiin (Blackburn 1939).

Taulukko 10. Eri tutkijoiden näkemyksiä talousaaltojen ajanjaksoista.

	Kondratieff "Die langen Wellen der Konjunktur" (1926)	Juglar "Des crises commerciales" (1862)	Kitchin "Cycles and Trend in Economic Factors" (1923)
Kondratieff (1926)	Noin 50 vuotta	Noin 7-11 vuotta	Noin 42 kuukautta
Blackburn (1939)	Noin 60 vuotta	Noin 10 vuotta	Noin 40 kuukautta
Kuznets (1940)	Noin 50 vuotta	Noin 8-9 vuotta	Noin 40 kuukautta
Grinin, L., Grinin, A. ja Korotayev (2017)	30-60 vuotta (riippuen ajankohdasta)		

(Grinin *et al.* 2017; Kondratieff 1926; ref. Kondratieff & Stolper 1935; Blackburn 1939; Kuznets 1940).

Kondratieffin (1926) tutkimuksessa tutkittiin alueellisesti eri ilmiöitä (Kondratieff 1926; ref. Kondratieff & Stolper 1935). Kuznets (1940) toi tutkimuksessaan esille Schumpeterin (1939) tutkimuksen maantieteellisiä eroja ja totesi talousaaltojen vaikuttaneen osittain alueellisesti. Kuznets (1940) kirjoitti Kondratieffin aallon koostuvan teoriassa kokonaisista Juglar-aalloista ja Juglar-aallot kokonaisista Kitchin-aalloista. Kondratieffin aalto on jaettu neljään vaiheeseen, jota kuvataan taulukon 11 sarakkeissa.

Taulukko 11. Schumpeterin näkemys talouden aaltoliikkeistä (Schumpeter 1939; ref. Kuznets 1940).

	Korkeasuhdanne	Taantuma	Lama	Elpyminen
Teollinen vallankumous (I)	1787-1800	1801-1813	1814-1827	1828-1842
Keskiluokkaistuminen (Bourgeois) (II)	1843-1857	1858-1869	1870-1884/1885	1886-1897
Neo-Mercantillist (III)	1898-1911	1912-1924/1925	1925/1926-1939	(*)

(*) Tutkimus tehtiin vuonna 1940, jolloin neljäs vaihe ei ollut vielä päättynyt.

Uset tutkimukset (Grinin *et al.* 2017; Kondratieff ja Stolper 1935, Blackburn 1939; Kuznets 1940) esittivät uusien teknologioiden kehittyvän voimakkaammin noin 45-70 vuoden

välein kapitalistisissa talouksissa. Kondratieffin aalto tunnetaan myös termeillä supersykli tai *K-wave*. Wilenius ja Kurki (2012) ovat tehneet julkaisun, jossa tutkitaan Kondratieffin kuudennen aallon ilmiötä ja sen vaikutuksia muun muassa digitalisaatioon Suomessa. Kuudennesta aallosta on useita tutkimuksia, joissa aallon vaiheet ja ajankohta vaihtelee. Grinin *et al.* (2017) esittivät, että kuudes aalto liittyy jo 1950-luvulla alkaneeseen tuotantovallankumoukseen, joka edisti tieteen ja teknologian voimakasta kehittymistä. Tutkimuksessaan he ehdottivat aallon kutsumista kyberneettiseksi, koska samaan aaltoon kuuluu vielä seuraavat 50 vuotta. Tutkimustulokseksi he saivat, että kuudennen talousaallon vaiheen olleen 1950-1990. Tällä ajanjaksolla siirryttiin tieteelliseen suunnitteluun, ennusteisiin, markkinointiin, logistiikkaan, tuotantoon, resurssien ja viestien liikkuvuuteen. Ajanjaksolla radikaalein muutos tapahtui tiedon ja informaatioteknologian alalla. Kuudennen aallon viimeinen jakso alkaa Grinin *et al.* (2017) mukaan noin vuonna 2030-2040 ja kestää aina vuoteen 2070 asti. (Grinin *et al.* 2017)

The World Economic Forum esittää (2018), että olemme siirtymässä vasta neljänteen teolliseen vallankumoukseen, joka vaatii erilaisen lähestymistavan suuriin teknologiamuroksiin. Myös Kondratieffin aaltojen ajankohdista on erilaisia näkemyksiä ja mm. Coccia (2017) esittää, että kuudes talousaalto alkaa vasta vuonna 2045 (Coccia 2017). Coccia (2017) esittääkin seitsemännen talousaallon alkavan 2100 sisältäen esimerkiksi valoa nopeampien teknologioiden sekä gravitaatioaaltojen sovelluksia. Coccia (2018) on esittänyt sodilla olleen ajallinen yhteys talousaaltoihin. Tutkimuksessaan hän esittää talousaallon pohjissa ja huipuissa olleen sotia. Kilpavarustelu vallasta kehittää myös yleishyödyllistä teknologiaa, joka osaltaan vaikuttaa talouden ja teknologioiden voimakkaampaan kehittymiseen (Coccia 2018).

Patentteja myönnetään maailmassa noin yksi jokaista miljoonaa asukasta kohden. Korotayev, Zinkina ja Bogevoľnov (2011) tutkivat patenttien määrän suhdetta Kondratieffin aaltoihin. Tutkimuksessaan he esittivät, että myönnettyjen patenttien määrällä oli tasainen nousu Kondratieffin aallon A-vaiheessa (nousuvaihe). Korotayevin *et al.* (2011) tutkimuksessa saadun tuloksen mukaan myönnettyjen patenttien määrän lasku oli melko vähäistä aallon laskuvaiheessa. Korotayev *et al.* (2011) mukaan Kondratieffin, Schumpeterin koulukunnan odottavan keksintötoiminnan vähenevän A-vaiheessa ja kasvavan B-vaiheessa. Tässä tutkimuksessa pyrittiin selittämään näitä eroja. (Korotayev *et al.* 2011)

Bodrožic ja Adler (2018) näkivät suurissa aaltoliikkeissä myös kytkennän johtamismalleihin. Heidän tutkimuksessaan esitettiin, että johtamismallien muuttaminen ja niiden ongelmien korjaaminen toistuu talouden aaltoliikkeiden mukaisesti. Tutkimuksessa hyödynnettiin Perezin vuonna 2002 (taulukko 12) kuvaamia talouden ja teknologian suuria aaltoliikkeitä (Bodrožic & Adler 2018).

Taulukko 12. *Mukaelma tutkimuksessa esitetyistä talouden aalloista ja sisällöstä (Bodrožic & Adler 2018).*

	1. Aalto: vesivoima ja rauta	2. Aalto: Höyryvoima ja rautatie	3. Aalto: Teräs ja sähkövoima	4. Aalto: Autot ja öljy	5. Aalto: Tietokoneet ja tietoliikenne
Itäminen	1750 luku–1770	1790 luku –1829	1850 luku–1875	1880 luku–1908	1950 luku–1971
Asettuminen	1771–1793	1829–1848	1875–1893	1908–1929	1971–2001
Kriisi/käännepaikka	1793–1797	1848–1850	1893–1895	1929–1944	2001/2008
Käyttöönotto	1797–1829	1850–1873	1895–1918	1944–1974	?
Loppuminen	1830–1840 luku	1873–1890 luku	1918–1940 luku	1974–1980 luku	?

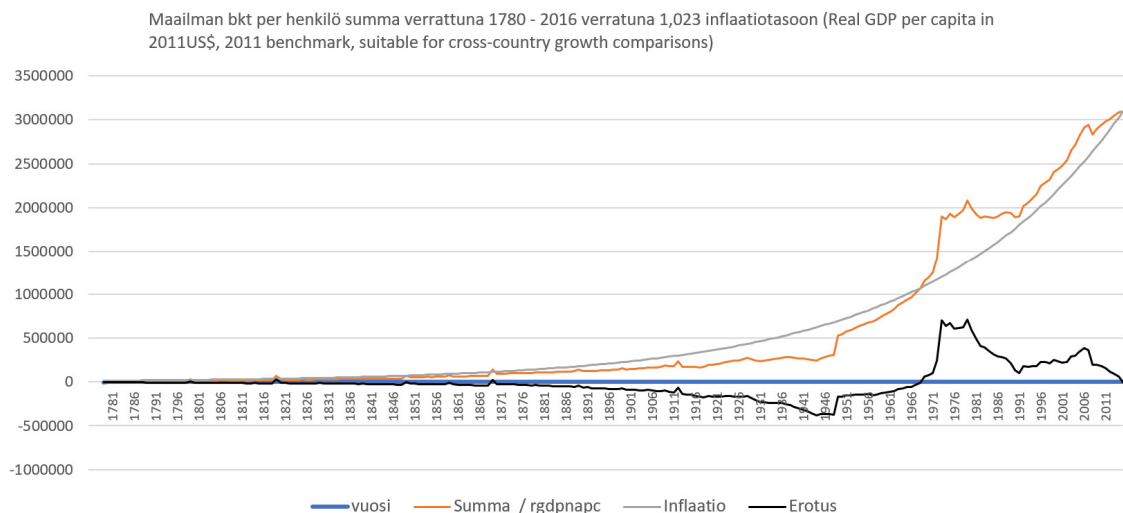
Bodrožic ja Adler (2018) esittävät myös Neo-Schumpeterian talousteorian liitoksen johtamismalleihin eri talousaalloissa. Heidän mukaansa viidennen (nykyisen) talousaallon johtamismalleissa on ominaista järkevöittämistä prosesseja rikkoen sisäisiä ja ulkoisia rajoja. Bodrožicin ja Adlerin näkemys aaltoliikkeistä poikkeaa näkökulmiltaan Grinin *et al.* (2017) näkemyksistä. Taulukossa 13 on esitetty Kondratieffin aaltojen ajakohdille toinen näkökulma. Täysiä ajallisia yhtymäkohtia ei näiden kahden esitetyn taulukon välillä ole helppo löytää.

Taulukko 13. *Mukaelma Grinin et al. (2017) esittämästä näkemyksestä Kondratieffin aalloista.*

Kondratieff aalto	1	2	3	4	5	6
Aikajakso	1780-1840	1840-1890	1890-1940	1940-1980*	1980-2020	2020/2030-2050/2060
Ilmiöitä	Tekstiiliteollisuus	Rautatiet, teräs	Sähkö, kemian teollisuus ja raskas teollisuus	Autoteollisuus, elektroniikka	Mikroelektronikka, tietokoneet	MANBRIC-tekniikat (med-addi-bio-nano-robo-info-cogno)

Aaltojen olemassa oloa voidaan tutkia esimerkiksi maailman pankin tai muun aineiston avulla. Bolt, Inklaar, de Jong ja van Zanden (2018) ovat tuottaneet vapaasti saatavissa

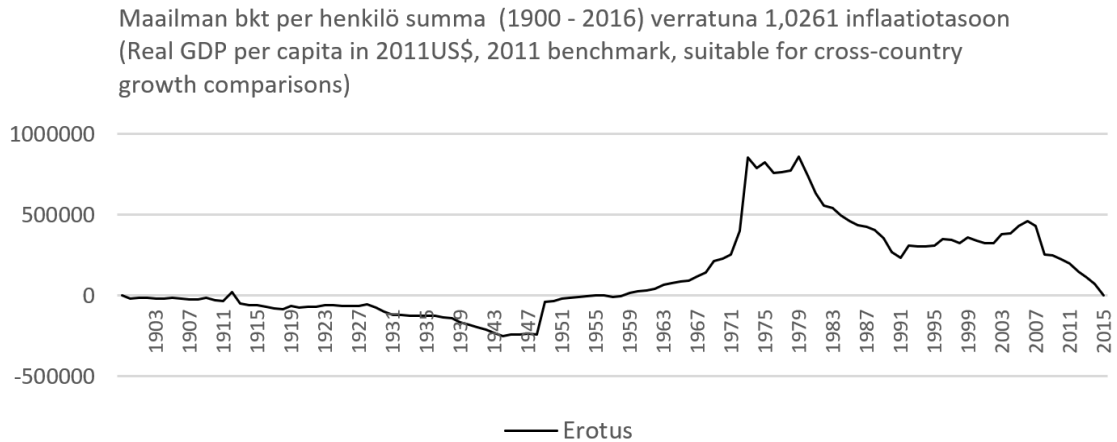
olevan tietokannan Maddison-projektissa, joka mahdollistaa talouden kasvun vertailun pitkällä aikavälillä. Vuoden 2018 tietokanta kattaa 169 maata vuoteen 2016 asti (Bolt *et al.* 2018). Kuvassa 34 on tutkittu Maddison-projektin tuottaman tietokannan aineistoja. Aineistoista on verrattu tietokannassa vuosittain yhteenlaskettua bruttokansantuotetta (BKT, RGDP) kehitystä inflaatioon. Inflaatioprosentti on oletettu olevan vakio pitkällä ajanjaksolla. Laskenta aloitetaan vuoden 1780 lähtötasosta (1780=15701\$) vuoden 2016 tasoon (2016 = 3 092 871\$). Inflaatiotaso saatiin interpoloimalla inflaatiotasoa (BKT 2016 - 1,0228161⁽²⁰¹⁶⁻¹⁷⁸⁰⁾ * 15701\$ = *nolla*) tavoitteena saada vuoden 2016 BKT:n ja 1780 BKT inflaatiokorjatun erotukseksi nolla. Laskentatarkkuus Excel-ohjelmistossa tulee rajoitteeksi, mutta 1780-2016 väliselle ajalle saatiin interpoloiduksi inflaatioksi noin 2,3%. Vakioarvoisen inflaatioon vertaaminen osoittautui mielenkiintoiseksi, koska BTK:n kasvamista voi kuvata kohtuullisen hyvin inflaation avulla; vaikka inflaatio pysyisi vakiona yli 200 vuotta. Inflaatioerotettuna talouden vaihtelut näkyvät mustalla viivalla Kuvassa 34. Vastaavaa avointa dataa maailman taloudesta lyhyemmältä jaksolta on saatavilla mm. Maailmanpankilta <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD> -osoitteessa.



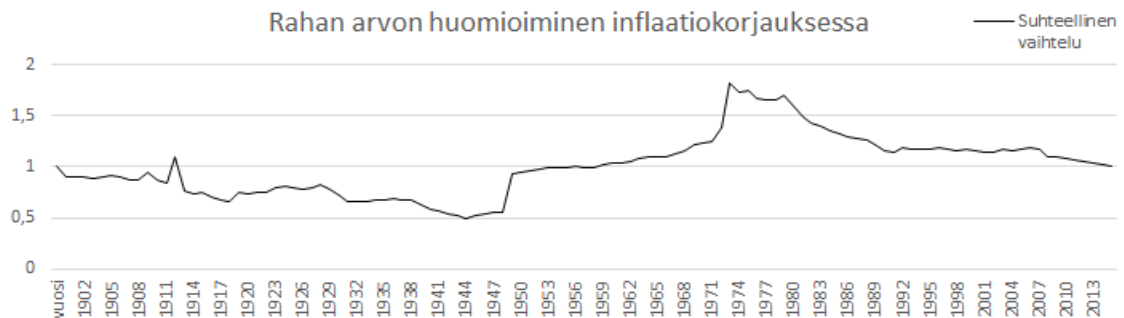
Kuva 34. Maailman reaali bruttokansantuotteen summa jaettuna ihmisten lukumäärällä 1780-2016 verrattuna 2,28161% inflaatiotasoon (Bolt *et al.* 2018).

Vastaava laskenta toistettiin kuvassa 35 vuosille 1900-2016. Kuvan 35 lähtötasoksi asetettiin 1900 = 155675\$, josta laskettiin inflaation kumuloitunut vaikutus vuoteen 2016. Inflaation vaikutus sovitettiin tavoitteena saada laskentatarkkuuden rajoissa sama taso vuoden 2016 reaali bruttokansantuotteen summan kanssa (mahdollisimman pieni ero). Kuvassa 35 viime vuosien vaihtelut ilmenevät korostuneesti ja kuvaajan ensimmäisten vuosien vaihtelut pienempinä. Tämä johtuu inflaation vaikutuksesta rahan arvoon. Kuvassa 36 on korjattu inflaation vaikutus myös rahan arvoon ja vertailtu suhteellista reaali bruttokansantuotteen arvoa kulloinkin olleeseen inflaation kehityskäyrän suhteelliseen arvoon. Laskentatulokset ovat esitetty liitteessä (liite F).

Kuva 36 voi osoittaa pitkän aikavälin vaihteluita taloudessa. Suurien aaltoliikkeiden lukumäärä on vähäinen ja myös muun ilmiön mukainen vaihtelu voi olla mahdollista.



Kuva 35. Maailman reaali bruttokansantuotteen summan ja interpoloidun 2,6% inflaatiotasoon erotus, jossa lähtötaso on asetettu 1900 vuoteen (1900=155675\$). (Bolt et al. 2018).

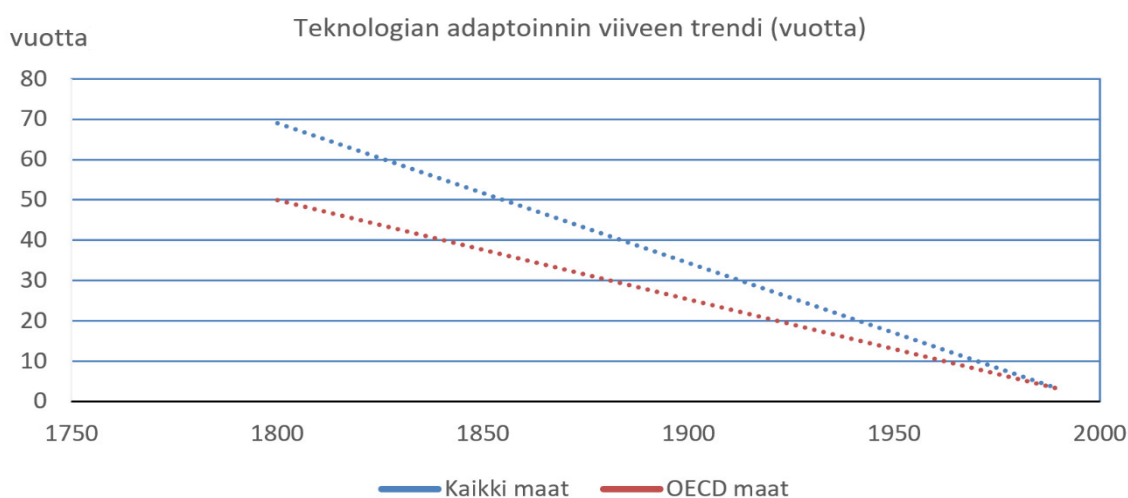


Kuva 36. Rahan arvon (reaali bruttokansantuotteen) esittäminen suhteellisena arvona suhteessa vertailtavaan inflaatioon (liite F).

Tutkijat ovat siis enemmän yhtä mieltä jo tapahtuneista ilmiöistä, joissa hyödynnetään eri katsontakulmia, mutta tulevaisuuden osalta ajankohdissa on vaihtelua riippuen lähteestä. Aaltoliikkeiden ajankohdat ja niiden vaihtelu voivat olla osittain perusteltavissa eri tarkastelunäkökulmilla. Kondratieffin aaltoliikkeiden ja (tietyn alan) patenttien määrän välisen yhteyden osoittaminen ilman eri tutkimusta ei ole mahdollista. Kondratieff (1926) itse kirjoitti tutkimuksessaan, että perustuen saatavilla olevaan tietoon pitkien aaltojen olemassaolo on todennäköistä (Kondratieff 1926; ref. Kondratieff & Stolper 1935). Tämä tutkimus käynnisti edelleen jatkuneet tutkimukset, jotka liittyvät suuriin talousaaltoihin ja useisiin siihen liitettyihin ilmiöihin. Kuvan 34 tuloksista voidaan kuitenkin havaita, että yhä jatkuva inflaatio on yksi merkitsevä tekijä taloudessa pitkällä aikajaksolla.

2.8 Uusien teknologioiden kehittymisen nopeutuminen

Diermeier ja Goecke (2017) esittivät teknologiadiffuusiota käsittelevässä tutkimuksessaan ”Productivity, Technology Diffusion and Digitization”-kuvaajan, jolla indikoidaan teknologian muutosnopeutta ja sen muuttumista ajan funktiona. Kuvassa 37 ”Adoption lags of new technologies” oli jaoteltu tutkimuksessa olleet maat kahteen ryhmään. Ryminä olivat kaikki maat (n=150) sekä Taloudellisen yhteistyön ja kehityksen järjestön (OECD) maat. Tutkimuksessa olleen (kuva 37) kuvaajan mukaan OECD-maiden tekemän yhteistyön merkitys adaptointin viiveessä on pienenevässä (Diermeier & Goecke 2017).



Kuva 37. Tulkinta Diermeier ja Goecke esittämästä ”Adoption lags of new technologies” kuvasta, joka perustuu CHAT -tietokannan tietoihin, Cologne Institution for Economic Research. (Diermeier & Goecke 2017)

Kuvan 37 mukaan teknologiadiffuusioon vaikuttaa yhä enemmän muu kuin virallinen yhteistyö. Uusien teknologioiden adaptointi on nopeutunut. OECD-maiden yhteistyön merkitys voi olla myös globalisaation aiheuttama, jossa yhteistyökumppani voidaan valita kaupallisin perustein ilman valmista yhteistyöverkkoa. Ferraro (2017) on todennut tutkimuksessaan, että uusien teknologioiden käyttäjämäärät kasvavat aikaisempaa nopeammin (Ferraro 2017). Diermeier ja Goecken (2017) esittävät, että uusien palveluiden käyttöönotto on digitaalisuuden myötä voinut muuttunua vuosista ja kuukausista päiviin.

Ulkoisen toimintaympäristö vaikuttaa teknologiadiffuusioon. Olosuhteiden muutosnopeus on kasvanut ja se on vaikuttanut tuotannon edellytyksiin. Dockalikova ja Klozikova (2014) mukaan menestymisen edellytyksenä ei ole enää omistaminen (maa, työvoima ja pääoma). He esittävät, että työntekijöiden ja johtajien tietämys, tiedot ja taidot auttavat reagoimaan muuttuviin olosuhteisiin. Tiedoilla ja taidoilla voidaan tehokkaasti hyödyntää muita tuotantotekijöitä (Dockalikova & Klozikova 2014). Olosuhteita voidaan tarkastella mm. analysoimalla ulkoista toimintaympäristöä, koska ulkoisen toimintaympäristön muutokset vaikuttavat teknologioiden innovointiin ja hyödyntämisen yleistymiseen. (Fertel, Bahn, Vaillancourt & Waaub 2013) tutkimuksessa PESTEL-analyysin avulla voidaan

arvioida yrityksen ulkoista toimintaympäristöä ja sen mahdollisia muutoksia, jotka johtuvat eri tekijöistä (Fertel *et al.* 2013). Ulkoinen toimintaympäristöanalyysi auttaa löytämään yksittäisten osatekijöiden väliset suhteet ja yhteydet ja tunnistamaan mahdollisuudet ja uhat, joihin yrityksen tulisi varautua (Dockalikova & Klozikova 2014). Päätöksenteossa voidaan hyödyntää myös muita analyysimenetelmiä, kuten kustannus-hyötyanalyysi tai monikriteerisiä analyysejä (Ferraro 2017). Useiden analyysimenetelmien lisäksi tulee kiinnittää huomiota analyyseissa annettavien osatekijöiden valinnalle ja painoarvoille, koska riippuen teknologiasta eri tekijät vaikuttavat teknologian yleistymiseen (Dockalikova & Klozikova 2014). Huomattavaa on myös se, että analyysimenetelmän valinta rajaa tutkittavan osa-alueen. Tässä työssä on aiemmin tuotu esille Caplicen ja Sheffin (2006) esittämä rajaamisen vaikutus ennakointiin, joka heikentää ennakkoinnin tuloksia. Dockalikova ja Klozikova (2014) julkaisivat tutkimuksen, jossa he esittivät artikkelin PESTEL-analyysin kriteerien merkittävyydestä. Myös Leite, J., Mello, L., Mello, J., Cruz, E. ja Fontanillas C. (2017) tutkivat PESTEL-analyysin ulkoisten tekijöiden priorisointia parantaakseen sen skenaarioanalyysia suurien rakennushankkeiden yhteydessä.

Leite *et al.* (2017) käyttivät tutkimuksessaan käyttäytymismallinnuksessa hyödynnettävä mittausmenetelmää (MACBETH), jolla pyritään löytämään kriteerejä määritteleviä keskeisiä suhteita. Leite *et al.* (2017) hyödynsivät PESTEL-analyysin (taulukko 14) kehittämisessä Fontanillasin (2016) tekemää osajakoa. Tässä PESTEL-analyysissa tutkittiin öljyasiantuntijoiden näkökulma eri osatekijöiden vaikutuksesta toimitusketjuun.

Taulukko 14. *PESTEL-analyysin jako osatekijöihin. Mukaelma Leite et al. tutkimuksesta. Alkuperäinen lähde: Fontanillas, C.N, väitöskirjasta, Rio de Janeiro, Brasilia, (2016), (Leite et al. 2017).*

Poliittinen	Taloudellinen	Sosiaalinen	Teknologinen	Ympäristö	Laki
Hallituksen va- kaus	BKT	Sosiaalinen liikku- vuus	Tutkimus ja kehitys	Ympäristölaki	Kansainvälinen laki
Yksityistäminen	Korot	Väestönkasvu	Innovaatiot	Arvot	Kansallinen lainsää- däntö
Rahoitus	Vaihtokurssi	Väestöjakauma	Automaatio	Maaperänsuojelulaki	Säätelevät elimet
Kansallinen vai- kuttaminen	Pääoman tarjoa- minen	Elinajanodote	Tietämyksen hal- linta	Ilma	Patenttisuoja
Kansainvälinen vaikuttaminen	Inflaatioindeksi	Syntyvyys	Viestintä	Jätteet	Määräykset
Sota ja konfliktit	Infrastruktuuri	Terveys	Patentit	Energian vähentämi- nen	Työlainsäädäntö
Hallituskausi	Maailmantalous	Työ ammatti	Keksinnöt	Eläimistön suoje- lu	Verolaki
Vaalit	Markkinakerros	Väestön siirtymi- nen	Internet	Eläinten suoje- lu	Kansallinen kauppa
Maakohtaiset suhteet	Kansainvälinen kauppa	Elintaso	Teknologiasiirto	Vaihtoehtoinen ener- gia	Verohelpotukset
Terrorismi	Tuonti	Yrityskulttuuri	Ohjelmisto	Veden kulutus	
Sidosryhmät	Vienti	Koulutus		Kaasun kulutuksen vähentäminen	
	Tuotannon taso				

Dockalikova ja Klozikova (2014) sovelsivat PESTEL-analyysia useiden kriteerien päätöksenteko menetelmässä (MCDM). PESTEL-analyysin taulukon 15 osatekijät valittiin tilanteessa, missä yritys X tarjosi komponentteja liikennevälineisiin ja teleyritys tarjosi palveluita. Tutkimuksessa kartoitettiin ulkoinen toimintaympäristö sekä valittiin priorisoitavat osatekijät syy-yhteyden perusteella (Dockalikova & Klozikova 2014).

Taulukko 15. *PESTEL-analyysin priorisoitavat osatekijät tutkimuksessa (Dockalikova & Klozikova 2014).*

Poliittinen	Taloudellinen	Sosiaalinen	Teknologinen	Ympäristö	Lainsäädäntö
Poliittinen vakaus	Työttömyys	Koulutustaso	Innovaatiot	Jätehuolto	Valtion sääntely
Veropolitiikka	Keskipalkka	Liikkuvuus	Teknologiamuutos	Ympäristön suo- jelu	Lait ja sääntely
Työlainsäädäntö	Ostovoima	Väestökehitys	Hallituksen tuki	Huoli ilmaston- muutoksesta	Oikeusjärjes- telmä
	BKT				

Yhteenvedon Dockalikova ja Klozikova (2014) ja Leite *et al.* (2017) tutkimuksista voi todeta, että

1. PESTEL-analyysin osatekijät valitaan aihekohtaisesti. Tutkimuksissa perusteltiin eri osatekijöiden käyttö, joilla löydetään syy-yhteys muutoksen ja osatekijän välillä.
2. PESTEL-analyysin osatekijöiden käsittelyn tarkkuus voi vaihdella ja muun muassa Leite *et al.* (2017) sekä Dockalikova ja Klozikova (2014) tutkimuksissa on PESTEL-analyysissä käytetty yleisesti samoja vertailtavia osatekijöitä.

Molemmissa tutkimuksissa voidaan hyödyntää aikadimensiota eli miten eri osatekijät muuttuvat strategisella aikajaksolla. Ennustuksen kannalta eri PESTEL-analyysin osaluille saatava tieto tulevista muutoksista vaikuttaa teknologiadiffuusion.

PESTEL-analyysissa on huomioitu esimerkiksi poliittiset päätökset. Poliittisilla päätöksillä voidaan edistää uuden teknologian käyttöönottoa tai vastaavasti hylätä aikaisempi teknologia. Poliittisina perusteina voidaan käyttää esimerkiksi ympäristösyitä, taloudellisuutta tai kestävyyttä. Poliittisella päätöksellä on esimerkiksi Kiina nousemassa tekoälyn kärkimaaksi julistamalla tekoälyn kansalliseksi prioriteetiksi (Ailisto *et al.* 2018).

Taloudellisten tekijöiden vaikutus teknologiadiffuusion on moniulotteista. Globalisaatio, jatkuvat muutokset toimintaympäristössä, kilpailu, talouden volatilitiiteetti ja kvantitatiiviset talouden luvut vaikuttavat päätöksentekoon, joka vaikuttaa teknologiadiffuusion. Ferraro (2017) esitti talouden volatilitiiteetin yhteyden teknologiadiffuusion. Hän kirjoitti talouden volatilitiiteetin kasvun hidastavan teknologiadiffuusiota sekä talouden kasvamista. Volatilitiiteetin ja talouden kasvun negatiivinen yhteys selvisi useissa yhteyksistä, joita hän referoi tutkimuksessaan (Ferraro 2017).

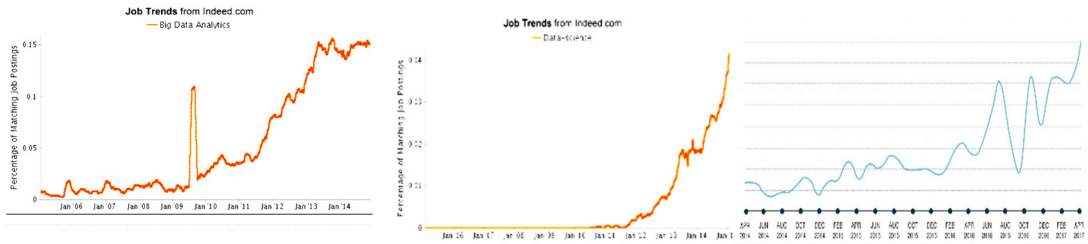
Kiyak ja Pranckevičiūtė (2016) esittivät, että yritysten analysointimenetelmissä arvioidaan yleensä yritysten tulovirtoja, lyhyt- ja pitkäaikaisia velkoja, pääoman määrää ja niiden suhteellisia indikaattoreita. Myös Leite *et al.* (2017) kirjoittivat, että suuri määrä nykyisistä päätöksistä perustuu taloudellisiin tiukasti perusteltuihin kvantitatiivisiin näkökulmiin. Päätösteoria kuitenkin osoittaa, että tulisi kiinnittää huomioita myös laadullisiin tekijöihin. Dockalikova ja Klozikova (2014) ovat todenneet, että globalisaatio on ollut hallitsevaa talouskehityksessä, mikä on vaikuttanut merkittävästi kilpailuun. He esittävät, että jatkuvat muutokset mikro- ja makroympäristössä aiheuttavat yrityksille tarpeen etsiä keinoja menestyäkseen markkinoilla (Dockalikova & Klozikova 2014).

Leite *et al.* (2017) korostivat sosiaalisia tekijöitä, jotka vaikuttavat teknologian hyödyntämisen kautta teknologiadiffuusioon. Sosiaalisina tekijöinä he mainitsivat muun muassa sosiaalisen liikkuvuuden, väestönkasvun, elinajanodotteen, syntyvyyden ja väestön siirtymisen vaikuttavan teknologian kehittymiseen (Leite *et al.* 2017).

Ympäristöystävällisyyden merkitys on viime vuosikymmeninä muuttunut paremmin rahallisesti mitattavaksi, kun ympäristöhaittojen aiheuttamat kustannukset ovat nousseet. Samoin ymmärrys teknologioiden kokonaiskustannuksista on kiinnittänyt huomiota energiatehokkuuteen. Ympäristötekijöitä voidaan tarkastella eri näkökulmista. Näkökulmia ovat esimerkiksi eettinen, ympäristöhaitat ja energian kulutus. Teknologia voi vaikuttaa välillisesti tai välittömästi. Välillisistä haitoista voidaan mainita esimerkiksi hiilen käyttö energiantuotantoon. Energiatehokkaat ja ympäristöystävälliset valinnat ovat strategisia päätöksiä, joiden arvoa ei vielä täysin osata mitata. Ostajien tietoisuus voi nykyaikana aiheuttaa uusia vaatimuksia tuotantotavoille sekä toimitusketjuille. PESTEL-analyysillä voidaan huomioida myös lainsäädännön vaikutuksia. Lainsäädäntö vaikuttaa voimakkaasti teknologian yleistymiseen tai teknologia häviämiseen. Esimerkkinä EU:n hehku-lamppumyyntiä rajoittavat laki, joka edisti uusien valaistusteknologioiden kehittymistä. Taulukossa 14 ja 15 tuotiin esille eri näkökulmia ulkoisen toimintaympäristön vaikuttavista asioista PESTEL-analyysin avulla.

Osaavan työvoiman merkitys teknologialle on merkittävää. Teknologian kehittyminen vaatii tutkimusta, hyödyntäjiä ja teknologian osajia. Davenport (2014) esitti osaavien ihmisten olevan yksi tärkeimmistä asioista big datan hyödyntämisessä. Hänen mukaansa tulevaisuudessa voidaan tarvita vähemmän analyytikkoja, mutta tarkensi, että näin ei ole tapahtunut. Myös Kaushik (2015) kirjoitti Accenture instituutin johtoon kuuluvan Jaenne Harrisin todenneen datan olevan hyödytöntä ilman taitoa käyttää sitä.

Useat tahot tutkivat teknologian kysyntää ja resurssitarpeiden kehittymistä. Sähköisissä palveluissa voidaan resurssitarve kuvata tai raportoida kohtuullisen reaaliaikaisesti esimerkiksi kuukausitasolla (kuva 38).



Kuva 38. Vasemmalla big data -analytiikkaa kuvaavien avoimien työpaikojen määrä vuosina 2006-2015 (Kaushik 2015). Keskellä datatieteiden vastaavien paikkojen määrä yhdysvalloissa (Gleason 2015). Oikealla Analytiikka ja datatieteiden avoimien työpaikkojen määrä 4/2014 - 4/2017 välisenä aikana (Gupta 2015).

Teknologiadiffuusion vaikuttaa osajien ja käyttäjien lisäksi kaupallinen kannattavuus. Aikaisemman teknologian myynnin vähetessä toimittajat luovat uusia arvoa tuottavia ratkaisuja, toimittavat asiakkaalle arvokkaita ratkaisuja ja saavat toimitetuista ratkaisuksista hyötyjä (Davis 2008).

3. TEKNOLOGIADIFFUUSION KASVUN ENNUSTAMINEN

3.1 TeknologiaDIFFUUSION ennustuksen menetelmävalinta

Wang *et al.* (2018) esitti, että *orastavien teknologioiden ennustamisessa* on hyötyä integroida useampia ennustuslähteitä. Tässä työssä huomioidaan Wang *et al.* (2018) esittämä tapa integroida useita tietolähteitä sekä hyödyntää patenttitietokantaa ennustamisessa. Työssä valitaan laskentamalli riippuen tutkittavasta teknologiasta. Tällaisia laskentamalleja tarkasteltiin eri käyttötarkoituksia varten aikaisemmin. Työssä tutkitaan big data -ilmiötä. Kirjallisessa tutkimuksessa todettiin big datan edustavan uutta lähestymistapaa. Uusien tuotteiden osalta Bassin (1969) laskentamallin käyttö on perusteltua, kun teknologia adaptoidaan vain kerran. Tällöin big dataa hyödyntävien loppukäyttäjien määrän arvioimiseen Bassin malli ei välttämättä ole sopiva. Valittujen tietolähteiden lähtötiedot tulee valmistaa valittuun laskentamalliin sopiviksi.

Bassin laskentamallin mukainen teknologiaennustaminen vaatii lähtötietoina kirjallisuustutkimuksessa esitetyt kumuloituvat summat, p ja q -parametrit. Parametrien laskentaa varten tarvitaan otosjoukko, josta lasketaan Bassin mallin mukaiset kumulatiiviset määrät eri ajanjaksoille. Tutkimuksen laskenta koostuu seuraavista vaiheista:

1. otokset valituista tietolähteistä
2. otoksien havaintoyksiköiden määrän laskenta vuosittain
3. kumuloituvien summien esitys vuosittain
4. p ja q -parametrien estimointi otoksista
5. p ja q -parametrien estimointi otoksittain
6. tutkittavien otoksien valinta
7. Bassin mallin mukainen laskenta estimoiduilla parametreilla
8. teknologiaennusteet
9. kriittisten massapisteiden laskenta
10. investointipäätöksen arviointi.

Kvantitatiiviseen ennustamiseen tarvitaan ennustamiseen soveltuvat aineistot. Kirjallisuustutkimuksen avulla voidaan päätellä ennustamisen konteksti ja muut osatekijät sopivien aineistojen valintaa varten sekä tarvittava aineisto-osaaminen. Kirjallisuustutkimuksen perusteella päädyttiin ennustusmenetelmävalintaan, jossa voidaan valita eri lähdeaineistoja ja saavutetaan riippuen lähdeaineistoista erilaisia ennustuksen ominaisuuksia. Kirjallisuudessa ennustusmenetelmät ovat vaihdelleet eri aikajaksoina.

3.2 Teknologian ennustamisen aineistot kvantitatiivisessa tutkimuksessa

Tutkimus voidaan jakaa takautuvaan tiedonhakuun ja tulevien asioiden ennakkointiin (Sassali 2014). Tiedonhaku rajataan kohdealueeseen, jossa tutkitaan takautuvia tapahtumia. Perusteena tutkimuslähteiden valintaan on myös aineiston saatavuus. Otokset poimitaan eri tietolähteistä. Menetelmärajauksena käytetään otsikko, avainsana ja lyhennelmä osiin kohdistuvia tekstihakua. Valittuna ilmiönä tutkitaan big dataa, jota etsitään ”big data”-termiyhdistelmällä. Big data on ilmiönä nopeasti kehittyvää, joka edellyttäisi tutkittaessa uudempien patenttien ja tutkimuksien hyödyntämisen (Sassali 2014). Teknologiadiffuusio ja sen ennustusmallit ovat kuitenkin hitaammin kehittyviä. Otoksiin poimitaan tuloksia (havaintoyksiköitä) koko aineistosta, jonka jälkeen keskitytään teknologian kehittymisen varhaisiin vaiheisiin. Aineiston keräämisessä tulee valita tietolähde tarkoituksen mukaan (Sassali 2014).

Teknologiaennustamisessa valitaan *tutkittava aikajakso*. Uusia teknologioita tutkittaessa on luonnollista, että uusia aineistoja tutkitaan mahdollisuuksien mukaan. Voisi päätellä, että uutta teknologiaa ennustettaessa lähteeksi valitaan reaaliaikaisempi vaihtoehto. Tässä työssä valitaan useampi tietolähde ennalta asetetulla hypoteesilla, että diffuusion ennustaminen reaaliaikaisemmasta tietolähteestä parantaa ennustuksen luotettavuutta. Patenttietokannan hyödyntäminen parantaa ennustustarkkuutta kaupallisesti hyödynnettävien sovellusten osalta. Kirjallisuustutkimuksessa tutkittiin talouden ja teknologian kehittymisen vuorovaikutusta. Makroekonomian osalta toteutettiin talouden suurien aaltojen simulointi Bolt *et al.* (2018) aineistolla. Simuloinnin perusteella vuosien 2006-2016 välillä on ollut talouden laskusuhdanne. Kirjallisuustutkimuksessa todettiin talouden volatilitietin vaikuttavan teknologiadiffuusiioon heikentävästi. Kun tutkitaan nykyisiä ilmiöitä, niin edellä mainittuja asioita voidaan huomioida nykytilanteessa (2018) teknologiadiffuusiota heikentävinä asioina. Toisaalta kirjallisuustutkimuksessa esitettiin teknologiadiffuusion muutosnopeuden kasvamisen sekä adaptioaikojen lyhentyneen jo pidemmän aikaa. Nämä ilmiöt kompensoivat toisiansa 1980-2016 välisenä ajanjaksona (kuva 36 ja kuva 37). Digitalisaatio on mahdollistanut nopean toimitusverkon ja valmiiden palveluiden ”reaaliaikaisen” tuottamisen pienillä kustannuksilla, jolloin *diffuusiiovaihe nopeutuu edelleen*. Cho ja Daim (2016) tutkimuksessa osoitettiin, että teknologian kehittymisen varhaiset mitattavat vaiheet voivat alkaa jo 10 vuotta teknologiadiffuusiota aikaisemmin. Otoksien alkuajankohdan optimiajankohta voidaan selvittää. Tavoitteena ajankohdan selvittämiseksi on saada laadullisesti hyvä otos, jonka koko on rajallinen ja virheellisten havaintojen osuus on pieni. Tässä työssä havainnollistetaan big datan varhaisien vaiheiden indikoitumista artikkeli- ja patenttietokannassa. Arviointia varten toteutetaan otantoja artikkeli- ja patenttietokantaan. Aikajaksoksi artikkelitietokannan osalta valitaan vuoden 1974 alusta vuoden 1994 loppuun (jatkossa vuosien 1974-1994 ajalta). Eri artikkelityypit kuvaavat teknologian kehittymisen eri vaiheita. Artikkelitietokannaksi valitaan Springer Link -artikkelitietokanta. Springer Link -artikkelitietokanta voidaan kohdistaa otokset eri

artikkelityyppeihin. Springer Link -artikkelitietokantaan kohdistettiin kolme eri otantaa (artikkeli, konferenssijulkaisu ja julkaisu -tyyppisiin aineistoihin), joiden tulokset koostettiin taulukkoon 16. Tekstipoiminnan otoskooksi ilman aikarajausta Springer Link -artikkelitietokannasta saatiin 39 799 havaintoyksikköä). Vastaavasti otoskooksi vuosien 1974-1994 ajalta saatiin yhteensä 90 havaintoyksikköä. Otoksen havaintoyksiköt tulkittiin. Tulkinta perustui otsikon sisältöön tai merkitykseen ja tarvittaessa sisältötekstin osittaiseen tulkintaan. Ennen teknologian yleistymistä käsitteet ja termit teknologiasta eivät ole yleistyneet. Käsitteiden puuttuminen johtaa virheellisiin päätelmiin ja aiheuttaa tulkintaan on runsaasti epävarmuutta. Hakutulokset jakautuivat taulukon 16 mukaisesti. Taulukon 16 'epävarma'-kenttä kuvaa virheellistä havaintoja (kpl/vuosi). Tulkinta otoksen yksittäisestä havainnosta (artikkeli tai julkaisu) toteutettiin lukemalla artikkelin otsikko, avainsanat ja tiivistelmä sekä arvioimalla kontekstia, jota kirjoituksessa käsiteltiin.

Taulukko 16. *Springer Link -artikkelitietokannan hakutulosten määrä (vuosien 1974-1994 ajalta).*

Vuosi	Epävarma (kpl)	Artikkeli (kpl)	Konferenssijulkaisu (kpl)	Julkaisu (kpl)	Yhteensä (kpl)
1974	1	1	0	0	2
1975	1	0	0	0	1
1976	1	0	0	0	1
1977	3	0	0	0	3
1978	1	0	0	0	1
1979	1	0	0	0	1
1980	2	0	0	0	2
1981	3	0	0	0	3
1982	3	0	0	0	3
1983	5	0	0	0	5
1984	5	0	0	0	5
1985	6	0	0	1	7
1986	3	0	0	0	3
1987	1	0	0	0	1
1988	7	0	1	0	8
1989	7	2	0	0	9
1990	4	0	0	1	5
1991	7	0	0	0	7
1992	6	0	0	0	6
1993	7	0	0	0	7
1994	10	0	0	0	10

Taulukon 16 mukaan suurien datamäärien keräämistä ja käsittelyä on toteutettu sähköisesti eri muodoissa jo ennen teknologian yleistymistä. Springer Link- tietokantaan toteutetuista otoksista löytyi big datan tarvitsemia osateknologioita sekä yhtymäkohtia nykyisin määriteltyyn big data -teknologiaan 1970 ja 1980 -luvulta. Vastaavia ongelmia oli jo ratkottu big datan, mutta myös muiden tietojen käsittelyn, hallinnan ja varastoinnin vuoksi. Myös erilaisia tilastollisia laskentamalleja on jo käytetty varhaisessa vaiheessa, joita on hyödynnetty big dataan liittyvän teknologian kanssa. Hakutuloksista löytyneiden varhaisien vaiheiden teknologioiden tulkinta osoittautui erittäin haasteelliseksi. Varhaisissa vaiheissa saadussa hakutuloksessa voi olla big dataa käsitteleviä teknologioita, vaikka kaikkia big datan tunnusmerkkejä ei ollut. Big datan saavuttaessa suosiota isojen teknologiayhtiöiden markkinoinnissa, tuotiin big data erittäin selvästi ja tarkasti esille – aineiston tulkinta muuttui tällöin yksiselitteisemmäksi. Tällöin big data oli usein myös erikseen esitetty avainsanoissa, vaikka se ei ollutkaan keskeisin tutkimuskohde.

Springer Link -artikkelitietokannan hakutulosten perusteella vuosien 1974-1994 ajalta (n=90) voidaan todeta, että hakutulokset sisälsivät vain harvoja ilmentymiä määritellystä teknologiasta. Teknologiaa käsitteleviä osateknologioita on kehitetty kyseisenä aikana, mutta kehittäminen ei ole kohdistunut big dataan. Hakutuloksista löytyi big dataan liittyviä teknologioita esimerkiksi suurien datamäärien tallentamisen, korrelaatioiden laskennan ja tilastollisten menetelmien kehittämistä.

Web of science -tietokantaan kohdennettiin otantoja vuoden 1993 alusta vuoden 2007 loppuun (jatkossa vuosien 1993-2017 ajalta). Hakuparametreinä testattiin big data, "big data" ja "big-data" -termiyhdistelmiä. Otoskoko säilyi lähes samana riippumatta hakuparametrien (em. hakutermit) poikkeavuuksista. Arvio populaation todellisesta määrästä vaihteli eri vuosina voimakkaasti. Varhaisissa vaiheissa otokseen tuli vain 0-10 havaintoyksikköä vuodessa ja myöhemmässä vaiheessa useita tuhansia havaintoyksiköitä. Kaikkien havaintoyksiköiden tutkiminen ei ole mielekäästä (25 868 havaintoyksikköä). Otosta voidaan tutkia hyödyntämällä tilastollisia menetelmiä. Tässä työssä käytetään eri otoksille samoja menetelmiä, jotta tulokset ovat menetelmien osalta vertailukelpoisempia. Menetelmänä niissä tapauksissa, jossa hakutulosten määrä on kohtuuttoman suuri tutkittavaksi, poimitaan hakutoksista satunnaisotos. Satunnaisuuden varmistamiseen hyödynnettiin Microsoft Excel 2016 toimintoja ohjelmalla (2).

= *satunnaisluku.väliltä (1; hakutulosten määrä)*

Ohjelma 2. *Käytetty Excel-funktio otoksen keräämiseksi eli tarkastettavien hakutulosten valitsemiseksi vuosittain.*

Microsoftin mukaan Microsoft Excel 2016 käyttää satunnaisuuden ratkaisemiseen random-funktiossa *Mercenne Twister* -algoritmia (Microsoft 2018). Mercenne Twister algoritmi on Matsumoton ja Nishimuran (1998) kehittämä yleistynyt algoritmi pseudosatun-

naislukujen tuottamiseksi. Niissä tapauksissa, jossa otanta oli otoksen osajoukko hakutulosten validisuus tarkistettiin päätelemällä big dataan vertailtavien ominaisuuksien avulla. Ominaisuuksia etsittiin havaintoyksiköistä tutkimalla otsikkoa, hakusanoja ja tiivistelmää. Oikeellisuuden päättely sisältää riskin virheelliseen arvioon. Hakuosumatarkkuuden arviointi oli huomattavasti hitaampaa, mutta helpompaa teknologian alkuvaiheessa. *Hakutulosten tarkkuuden arvioinnin vaihtelu eri ajankohtina vääristää teknologian kehittymisen ajan funktiona.* Teknologian yleistyessä termien käyttäminen oli tarkempaa ja teknologian tulkitseminen sen osalta huomattavan paljon helpompaa. Toisaalta uuden teknologian suosio on saanut useat ihmiset kuvaamaan ratkaisut big dataksi, vaikka se edustaisi aikaisempia suuria tietovarastoratkaisuja (Davenport 2014).

Taulukko 17. *Web of science -hakutulosten oikeellisuuden tarkastaminen. Poiminta on toteutettu vuosien 1993-2017 ajalta. Virheellisten tuloksien osuus on alle 50% vuoden 1999 alusta alkaen. (4/2018; päivitetty 11/2018)*

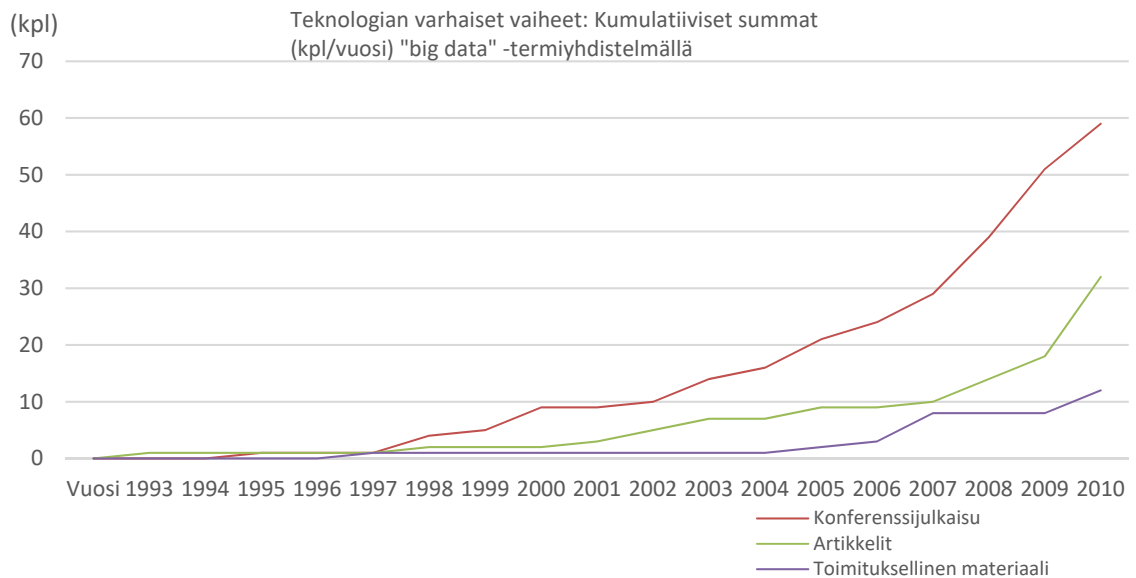
Vuosi	Virheellisten määrä laadullisessa vertailussa	Edustaa teknologiaa	Osumat eli otosyksiköt	Konferenssijulkaisu (proceeding paper)	Artikkelit	Toimituksellinen materiaali	Muu materiaali	Otoksesta laadulliseen vertailuun valittu otanta
1993	1	0	1	0	1	0	0	1
1994	1	0	1	0	1	0	0	1
1995	0	0	0	0	0	0	0	0
1996	1	0	1	1	0	0	0	1
1997	0	0	0	0	0	0	0	0
1998	1	0	1	0	0	1	0	1
1999	1	2	3	3	1	0	0	3
2000	1	0	1	1	0	0	0	1
2001	1	3	4	4	0	0	0	4
2002	1	0	1	0	1	0	0	1
2003	0	3	3	1	2	0	0	3
2004	2	4	6	4	2	0	0	6
2005	0	2	2	2	0	0	0	2
2006	5	2	7	5	2	1	*) Virhe	7
2007	0	4	4	3	0	1	0	4
2008	1	15	16	5	1	5	5	16
2009	3	13	16	10	4	0	2	16
2010	3	14	17	12	4	0	1	17
2011	0	30	30	8	14	4	4	30
2012	1	284	285	195	52	27	11	285
2013	1 (*)		1307	917	262	82	46	300
2014	2 (*)		2825	1825	693	168	139	370
2015	2 (*)		5478	3306	1691	236	226	640
2016	1 (*)		7416	4125	2604	346	326	500
2017	0 (*)		7810	3655	3429	318	349	340
2018		-	4875	647	3569	290	98	-

Taulukon 17 datalla testataan aineiston lisäksi menetelmää sekä teoriaa. Kuvassa 39 on esitetty artikkelitietokannan aineistojen ajallinen kehittyminen, jossa toistetaan Chon ja Daimin (2016) esittämä aineistojen kronologinen järjestys. Konferenssimateriaalin määrä on näissä otannoissa kehittynyt aikaisemmin kuin perustutkimusta kuvaavien artikkeleiden määrä (Kuva 39).

Aineistoista lasketaan kumulatiiviset summat, jolloin hidaskasvu erottuu selvemmin. Kumulatiiviset summat on esitetty taulukossa 18, jotka ovat esitetty kuvassa 39.

Taulukko 18. Kumulatiiviset summat Web of Science -hakutuloksista.

Vuosi	Konferenssijulkaisu	Artikkelit	Toimituksellinen materiaali
1993	0	0	0
1994	0	1	0
1995	0	1	0
1996	1	1	0
1997	1	1	0
1998	1	1	1
1999	4	2	1
2000	5	2	1
2001	9	2	1
2002	9	3	1
2003	10	5	1
2004	14	7	1
2005	16	7	1
2006	21	9	2
2007	24	9	3
2008	29	10	8
2009	39	14	8
2010	51	18	8
2011	59	32	12



Kuva 39. Web of science -tietokannan hakutulosten perusteella "big data" -termiyhdistelmään liittyvien artikkeleiden määrä vuosittain ajan funktiona.

Kumulatiiviset summat taulukossa 18 sekä kuvassa 39 vastaavat kuvaukseltaan Chon ja Daimin (2016) esittämää kronologista järjestystä. Konferenssijulkaisut esiintyvät kronologisesti artikkeleita ja toimituksellista materiaalia aikaisemmin. Kuvan 39 perusteella big data on näiden lähtötietojen perusteella alkanut kehittymään vuonna 1998.

Patenttitietokannasta poimittiin tässä työssä kolme eri otosta. Poimintojen avulla voitiin osoittaa, että patenttitietokannan päivittämisprosessissa on noin 6-8 kk:n viive. Alle 4kk kuluessa poimitusta otoksesta puuttui suurin osa myöhemmin otoksiin tulleista havaintoyksiköitä. Tietoa varsinaisen tutkimuksen tekemisessä hyödynnetään sulkemalla pois uusimmat havainnot. Tässä tutkimuksessa hyödynnetään 2.7.2018 toteutettua poimintaan patenttitietokantaan sekä rajataan otoksen havaintoyksiköt pois 1.1.2018 alkaen laadun parantamiseksi.

Kirjallisuustutkimuksen, artikkelitietokanta- ja patenttitietokantatutkimuksen perusteella teknologiadiffuusion sekä sen varhaisien vaiheiden tarkasteluajankohta alkaen 1998 alusta päättyen 2017 loppuun. Tällöin otoksen koko on ajallisesti riittävä sekä virheitä enintään sisältäneet ajankohdat ovat rajattu pois otannoista. Patenttitietokannan osalta käytettiin oikaistua aineistoa, josta poistettiin virheelliset havainnot myös otoksen alkuvuosilta.

Teknologiadiffuusiota voidaan ennustaa patenttitietokannoista. Patenteja haetaan, kun hakija haluaa suojata tapansa soveltaa keksintöään. Patenttihakemusten mukaan patentit luokitellaan kansainvälisen patenttiluokituksen (IPC) mukaisesti ja siihen pohjautuvaan tarkempaan CPC-patenttiluokitukseen (CPC). IPC-patenttiluokitusjärjestelmällä voidaan kuvata tekniikan alaa, johon keksintö kuuluu. IPC-patenttiluokitukselle on useita versioita, jotka on yleensä ilmoitettu luokituksen yhteydessä. Luokittelu kerrotaan patentin etusivulla esimerkiksi merkinnällä ”Int.Cl. (2015.01)”, jossa suluissa oleva merkintä kertoo tässä tapauksessa IPC-luokitteluversion. Luokittelun ansiosta on patenttien löytäminen helpompaa. (PRH 2018)

Cooperative Patent Classification (CPC) on IPC-luokitukseen perustuva tarkempi jaotteleminen, joka jakaa tekniikan alat 250 000 eri ryhmään. Euroopan ja Yhdysvaltojen patenttivirastot (EPO ja USPTO) pitävät yllä yhdessä tätä järjestelmää, jolloin mm. Patentti ja rekisterihallitus käyttää CPC-järjestelmää luokittelussa. (PRH 2018).

Teknologian menestymistä voidaan ennustaa eri menetelmin. Altuntas, Dereli ja Kusiak (2015) esittivät, että teknologian menestymistä voidaan ennustaa patenttitietokannan avulla. Rajallisin tiedoin ja epävakaa markkinatilanteessa teknologian menestymisen ennakointi on monimutkaista. Tutkimuksessaan hän tarkasteli 1) teknologian elinkaarta, 2) teknologiadiffuusiota, 3) patenttien voimaa ja 4) teknologian kasvupotentiaalia.

Tutkimuksessaan Altuntas *et al.* (2015) keskittyivät kolmeen teknologiaan, joista selvitettiin kaikki tietokannassa olleet kyseisten teknologioiden patentit Yhdysvalloissa. Tutkimus perustui oletukseen, että *patentin vahvuutta ja kasvupotentiaalia pidetään teknologian kattavuusindikaattoreina*. Tutkimuksessa eri kriteereillä kerättyihin tuloksiin sovellettiin algoritmia, jolla tulokset yhdistettiin.

Taulukossa 18 on esitetty ne työvaiheet, joita Altuntas *et al.* (2015) käyttivät heidän tutkimuksessaan.

Taulukko 19. Tutkimuksen työvaiheet (Altuntas *et al.* 2015).

Vaihe	Tehtävä Altuntas <i>et al.</i> (2015) tutkimuksessa
1	Haetaan tiedot patenttitietokannasta.
2	Luokitellaan kohteet osa-alueittain.
3	Selvitetään patenttien määrä vuodessa osa-alueittain.
4	Piirretään S-käyrä teknologian elinkaarelle.
5	Selvitetään patenteihin tehtyjen viittausten määrä vuosittain.
6	Lasketaan teknologian diffuusionopeus.
7	Selvitetään patenttikoodien kokonaismäärä vuosittain.
8	Lasketaan patenttien vahvuus hyödyntäen IPC-koodeja.
9	Lasketaan kokonaismäärä eri patenttikoodille.
10	Lasketaan kasvumahdollisuudet.
11	Suodatetaan data <i>condorcet</i> -metodilla.
12	Tehdään investointiarviointi.

Yleisenä huomiona tutkimusasetelmasta voidaan todeta, että tilastollisesti Altuntas *et al.* (2015) tutkivat kolmen eri teknologian menestymistä. Tutkimustapa oli innovatiivinen, koska se yhdistää neljä eri kriteeriä yhtäaikaisesti ja antaa tutkittavien teknologioiden kehittymisestä paremman kokonaiskuvan. Tuloksissa kuitenkin tulee muistaa konteksti, jota tutkimus koski.

Tutkimusvalintojen osalta päädyttiin ratkaisuun, jossa hyödynnetään *patenttien kasvupotentiaalin arvioimista Bassin laskentamallin mukaisesti* ja rajataan työn laajuudesta pois Altuntas *et al.* (2015) tarkoittama patenttien vahvuus. Ennalta asetettu hypoteesi ”*reaaliaikaisemmasta tietolähteen parantavan ennustuksen luotettavuutta*” ei toteutunut patenttitietokannan osalta ja reaaliaikaisin osuus 1.1.2018 alkaen rajattiin pois tutkimuksesta. Investointiarviointiin valitaan Sundqvist *et al.* (2002) esittämä kriittisen massapisteen laskenta. Teknologian kehittämisessä huomioidaan Jacksonin (2013) esittämien *p* ja *q*-parametrien suhde sekä Lindenin ja Fennin (2003) huomioima teknologian kasvunopeus. Investointiarvioinnissa hyödynnetään patenttitietokantaa. Investointiarvioinnin tulosten tavoitteena on saavuttaa tietoa ennustettavan teknologian kehittämisestä sekä kehittämisen riskeistä.

3.3 Teknologiaennustamisen tietolähteiden valinta

Teknologiaennustamisen *tietolähteiden määrä* valitaan kirjallisuustutkimuksen perusteella. Aikaisemmin tuotiin esille useiden lähteiden käyttö. Muun muassa Wang *et al.* (2018), Cho ja Daim (2016), Lee *et al.* (2018) ja Huang *et al.* (2018) hyödynsivät useiden tietolähteiden menetelmää ennustamiseen. Kirjallisuustutkimuksen perusteella päädyttiin tutkimusmenetelmään, jossa hyödynnetään useita tietolähteitä yhtäaikaisesti. Tässä työssä on koostettu ja järjestetty kronologisesti eri lähteitä. Kronologisesti järjestettyjä lähteitä valitaan tutkimuslähteiksi. Valinnassa huomioidaan Sassalin (2014) esittämä aikaisempia tapahtumia kuvaavien lähteiden valinta. Varsinaiset tietolähteet valitaan kirjallisuustutkimuksessa käytetyistä lähteistä. Tutkimuksessa huomioidaan se, että tietolähteen lisäksi asiakirjatyypit tai muu aineiston tyyppiluokitus vaikuttavat kirjallisuustutkimuksen mukaisesti otokseen. Hauissa huomioidaan eri hakuparametrit sekä asiakirjahaun kohdentaminen populaation ominaisuuksien eri osiin. Asiakirjoissa huomioitavia osia ovat esimerkiksi otsikko, avainsanat ja sisältö. Hakuparametriksi asetettiin ”big data”-termiyhdistelmä, jonka haku kohdennetaan vuosittain. Kirjallisen tutkimuksen taustatyössä saavutettiin aineisto-osaamista tietolähteiden ominaisuuksista. Työssä valitaan lähteitä, jotka ovat kohtuullisella työllä saatavissa sekä mallinnettavissa yhteiselle tutkitavalle aikajaksolle. Perusteluina lähteiden valinnassa käytetään kirjallisessa osiossa esitettyjen teorioiden soveltamista sekä teorioiden kronologisen syy-seuraussuhteiden esittämistä.

Lähdetietojen valinnassa huomioidaan kaupallisuus, tieteelliset artikkelit, toimituksellinen materiaali ja patenttien kehittyminen (kaupallisesti suojattavat ratkaisut). Yhteensä näiden valintojen kautta saatiin 15 tietolähdettä. Valinnoissa hyödynnettiin muun muassa Cho ja Daim (2016) sekä Huang *et al.* (2018) esittämiä tietolähteiden kronologista järjestäytymistä tutkimuksissaan. Suuri osa valituista tietolähteistä perustuu edellä mainittujen tutkimusten tietolähteisiin. Tietokannat valittiin kirjallisuustutkimuksen perusteella seuraavista tietokannoista:

- ProQuest (kaupalliset artikkelit)
- Web of Science (perustutkimus)
- Scopus (perustutkimus)
- patenttietokanta (kaupallisesti suojattavat ratkaisut).

Tietokantojen valinnan perusteena käytettiin kolmen eri kokonaisuuden kuvaamista, jossa on tavoitteena tuoda Huang *et al.* (2018) tutkimuksessa esitetty kaupallisuus ja integroida se Chon ja Daimin (2016) esittämään tutkimukseen. Näillä valinnoilla voidaan selvittää kaupallisen materiaalin ja kaupallisesti suojattavien ratkaisujen eroja (patentit) sekä suhteuttaa edellä mainittuja perustutkimukseen (artikkelit). Valinnoilla etsitään liikeloudellisesti kehittyviä teknologioita.

Kvantitatiiviseen osaan valitaan seuraavia lähteitä kaupallisista, tieteellisistä ja teknisistä julkaisuista seuraavasti:

- kaupalliset artikkelit
- tieteelliset artikkelit
- artikkelit
- toimituksellinen materiaali
- katsaukset
- kokoustiivistelmät
- konferenssijulkaisu
- patenttitietokanta.

Lähteiden valinnassa huomioitiin, että lähteet sijoittuva kronologisesti ennen ennustettavan lähdeaineiston muodostumista. Näillä aineistoilla voitiin varmentaa mahdollisien patenttien määrän kehittymistä.

3.4 Teknologian ennustaminen muilla menetelmillä

Teknologiaa voidaan ennustaa vapaista tekstilähteistä tekemällä ohjattuja tekstihakuja ja tarkastelemalla tekstien yleisyyttä. Kirjoitushetkellä (2018) on yleistymässä automaatio, jonka avulla tekstien hakua ja tekstien luokittelua automatisoidaan sekä saadaan luokitte- luun sisältöä, jota ei määritelty syötteissä. Teknisten ennustamismenetelmien lisäksi suosittuja tapoja teknologian ennustamiseen on myös mm. Delfoi-haastattelumenetelmä. Au- tomaattinen luokittelu ja koneoppiminen ovat teknologioita, jotka tarvitsevat dataa toimi- akseen. Yleisesti näitä ratkaisuja on rakennettu data-altaisiin hyödyntäten big data -rat- kaisuja.

Teknologiaa voidaan ennustaa myös monilla muillakin menetelmillä. Caplice & Sheffi (2006) mukaan tuotteiden yleistymistä voidaan ennustaa esimerkiksi seuraavilla mene- telmillä:

- asiakastutkimukset
- johdon mielipiteet
- myyntihenkilöstön mielipiteet
- näkymien arviointi
- trendikäyrät
- liukuvat keskiarvot
- skenaarioanalyysit
- kokemuskäyrät
- Delphi-metodi
- päätöspuut
- lineaariregressio.

3.5 Big data -teknologian tunnistaminen

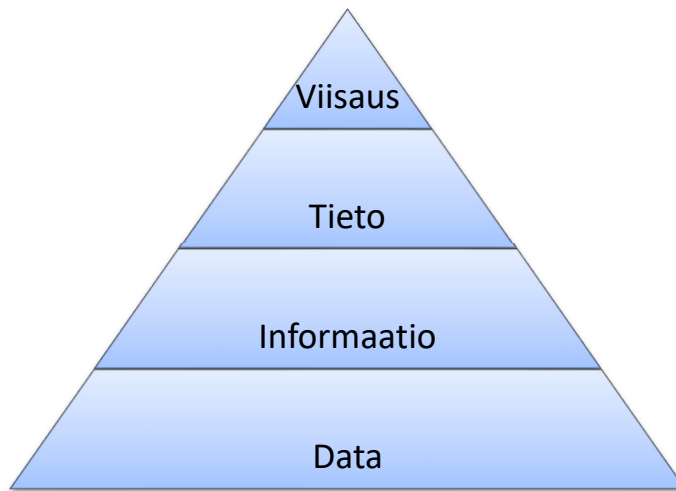
Teknologiadiffuusion mittaaminen vaatii kyseisen teknologian sekä sen osateknologioiden, toimintaympäristön ja kontekstin erittäin hyvän tunnistamisen. Tällä varmistetaan teknologian kehittymisen tunnistaminen. Big data ei ole ilmiönä uusi, vaan kyseessä on uudella tavalla paketoitu teknologiakokonaisuus (Davenport 2014). Davenport toi myös esille, että vuonna 2010 big data -käsite alkoi esiintymään voimakkaasti useassa yhteydessä, vaikka niillä oli Davenportin (2014) muodostaman taulukon 20 mukaan vain vähän yhteistä big datan kanssa.

Taulukko 20. *Big datan ja perinteisen tietovarastoinnin erot (Davenport 2014).*

	Big data	Perinteinen analytiikka
Datan tyyppi	Rakenteeton (unstructured)	Rakenteellinen (structured)
Datan määrä	100 TB – 1 PT	Korkeintaan kymmeniä teratavuja (Tens of terabytes or less)
Data virta	Jatkuva datavirta. (Constant flow of data)	Staattinen datavarasto (Static pool of data)
Analytiikkametodit	Koneoppiminen (Machine learning)	Hypoteesiin pohjautuva (Hypoteesis based)
Ensisijainen tarkoitus	Dataan pohjautuvat tuotteet (Data based products)	Sisäisen päätöksen teon tuki ja palvelut (Internal decision support and services)

Gandomi ja Haider (2015) esittivät big datan olevan laajempi käsite kuin tietomassan suuri koko. He kirjoittivat, että big data on pääosin dataa ilman rakennetta, jossa erilainen data voi olla esimerkiksi teksti-, ääni- tai videomuodossa. Gandomi ja Haider (2015) toivat esille, että yksi keskeinen ominaisuus big datalla on tallennetun tiedon taajuus eli frekvenssi. Laney (2001) esitti jo vuonna 2001, että keskitetyt datavarastot yleistyvät ja datan hallinta käsittää kolme merkittävää osa-aluetta, joita ovat määrä, nopeus/kiertonopeus ja vaihtelu (3D Data Management: Controlling Data Volume, Velocity, and Variety). Julkaisussaan Laney esitti big datalle ominaisen yksittäisen tiedon eli datan (kuva 40) tallentamisen tarpeen ja kirjoitti datan vaihtelevuuden haasteista. Gandomin ja Haiderin (2015) mukaan näitä Laneyn alun perin esittämiä big data 3V's kolmea osa-aluetta on pidetty big dataa määrittävinä ominaisuuksina, jossa 1) *datan määrä*, 2) *datan vaihtelevuus* ja 3) *käsittelyn nopeus ovat suuria*. He toivat esille, että myös muun muassa Gartner ja TechAmerica Foundation määrittelevät big datan näillä kolmella ominaisuudella (Gandomi & Haider 2015).

Big datalla, tallennettavalla tiedolla ja käsiteltävien tietojen välisillä suhteilla on riippuvuuksia, joita voi esittää tiedon pyramidilla (kuva 40). Gandomi ja Haider (2015) esittivät asian myös toisella tavalla toteamalla, että rakenteellinen tieto, joka on käytettävissä muodossa edustaa vain pientä osajoukkoa big datasta. Jifa ja Lingling (2014) tutkivat big datan, tiedon ja tiedon pyramidin keskeisiä suhteita. He esittivät, että tieto koostuu tiedon pyramidimallissa (DIKW) neljästä eri tasosta, jossa jokainen taso tarvitsee edellisten tiedot. Tiedon pyramidin kuvaaminen suomenkielellä edellyttää tarkkuutta. Suomi ei kielellä erottele eri tiedon tasoja tarkasti (data, information, knowledge, wisdom). Tiedon välityksen kuvaaminen sekä tietojen välisien suhteiden esittämistä on haasteellista suomenkielellä.



Kuva 40. Tiedon pyramidin (Jifa & Lingling 2014).

Jifa ja Lingling (2014) mukaan tiedon pyramidissa ensimmäinen taso on data, joka edustaa tiedon perustasoa (kuva 40). He esittivät datan voivan olla käsittelemätöntä ja esiintyvä missä tahansa tai jopa käyttökeltottomassa muodossa. Jifa ja Lingling (2014) mukaan datalla ei itsenäään ole merkitystä eli kyseessä on enemmänkin irrallinen ”data”. He toteavat tiedon pyramidin toisen tason edustavan informaatiota (information), joka koostuu datasta ja on rikastunut dataan liittyvällä kontekstilla eli asiayhteydellä. Heidän mukaansa tiedon olevan rakentuvaa ja kolmannen tason tiedon (knowledge) perustuvan kahteen alempaan tasoon tiedon pyramidissa. He kirjoittivat myös, että tieto (knowledge) sisältää rikastuneen tiedon, *miten tietoa voidaan hyödyntää* ja että viisaus (wisdom) sisältää tiedon *miksi ja koska tietoa kannattaa hyödyntää*. Kuvassa 40 on esitetty tiedon pyramidin (DIKW), jossa data on alimmalla tasolla. Jifa ja Lingling (2014) osoittivat tutkimukseensa, että tiedon pyramidissa datan arvo saadaan jalostamalla palveluita, joilla saavutetaan tiedon pyramidin ylempien tasojen ominaisuuksia.

Big data -ilmiö koostuu useista eri teknologioista. Big data -ilmiö vaatii kehittyäkseen esimerkiksi suurien datamäärien tallentamismenetelmiä, laskentatehon sekä eri kuvaus-

ja ohjelmointikielillä toteutettavia matemaattisia algoritmeja. Päätelminä Jifan ja Liglingin (2014), Davenportin (2014) ja Gandomin ja Haiderin (2015) tutkimuksista voidaan todeta, että tiedon pyramidin kuvaa määrällisesti myös big data -ratkaisuissa tiedon määrää tietotyypin mukaisesti. Keskeisintä big datassa on pienen ja rakenteettoman datan nopea saatavuus, joka on varastoinnin rinnalla infrastruktuurihaaste.

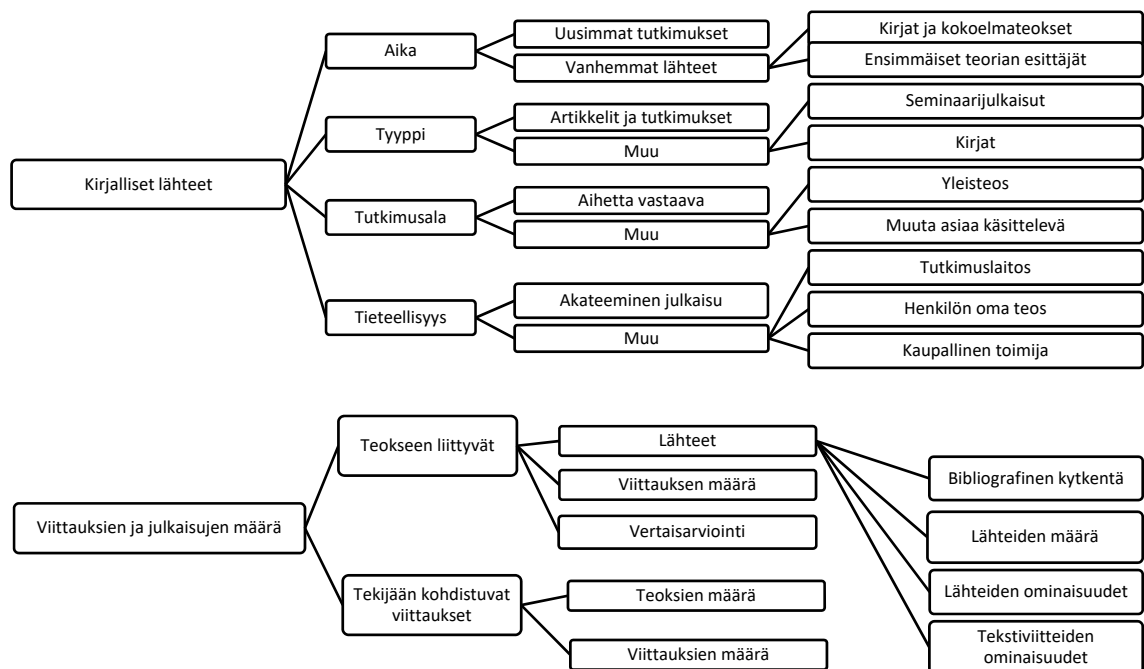
Tiedon pyramidissa (DIKW) *datan* kerääminen ja siirtäminen on helpointa. Siirryttäessä pyramidin ylemmille tasoille siirrettävyys ja käsiteltävyys vaikeutuu. Edellä mainituista syistä big dataa hyödyntäville teknologioille on etua datan keräämisessä *käsittelemättömänä raakadatana*. Tietoa käsittelevät järjestelmät tuottavat dataa eli pientä tietoa: rakenteella tai ilman rakennetta. Big datan perusidea on kerätä dataa eri lähteistä. Rakenteellinen data pyritään tunnistamaan automaattisesti ja rakenteeton koostetaan data-altaksi (Data Lake) ja virroiksi. Dataa voidaan valmistella työkaluohjelmistoilla, jotka hoitavat rakenteen tunnistamisen eri menetelmillä. Big data on teknologioiden alusta, johon voidaan soveltaa esimerkiksi koneoppimista, automaattista luokittelua ja ennustamista matemaattisilla malleilla.

Osateknologioiden kehittyminen ei välttämättä ole ainoastaan riippuvainen kehitettävästä teknologiasta, vaan toinen ongelma tai tarve saattaa kehittää osateknologioita nopeasti. Tietoteknisissä ratkaisuissa datan tallentaminen, käsittely, siirto ja nopea prosessointi on kehittynyt yleisesti useiden eri teknologioiden vaikutuksesta. Teknologian arkipäiväistyminen näkyy jo kulutustuotteissa, kodinkoneissa, liikkumisissa ja asumisessa. Kun teknologian ilmentymisen kehittymistä tutkitaan sanahakujen avulla, vaikuttaa sanaston kehittyminen, hype ja voimakas termeihin liittyvä markkinointi sanojen yleisyyteen. Gandomi ja Haider (2015) ovat tutkineet big data -termin esiintyvyyttä tutkimuksissa ja toivat esille, että useat johtavat teknologiayhtiöt markkinoivat big dataa uutena menetelmänä.

4. TUTKIMUSMENETELMÄT JA AINEISTO

4.1 Kirjallinen tutkimusmenetelmä ja aineisto

Kirjallisuustutkimuksen alkuvaiheessa tutustuttiin tieteellisten julkaisujen arviointiin bibliometrisien menetelmien avulla. Kirjallisuustutkimusta varten rekisteröidyttiin useampaan artikkelitietokantaan. Lähdeluetteloon pyrittiin saamaan jäljitettävyyttä, josta selviää artikkelitietokanta ja viitattu artikkeli sekä sen saatavuus.



Kuva 41. Kirjallisessa tutkimuksessa lähteiden valinnassa tarkasteltuja näkökulmia.

Työssä kiinnitettiin huomiota kirjoittajien ja kirjoittajaryhmien bibliometrisiin sidonnaisuuksiin. Ongelmaa pyrittiin korjaamaan lisäämällä lähteiden määrää käsiteltävästä aiheesta, jotta myös kriittinen näkökulma saadaan tuotua esille. Kuvassa 41 on yleistetty ongelmaa ja kirjoitustyön edetessä havaittiin, että samat kirjoittajat esiintyvät useissa yhteyksissä. Kirjoittamistyön edetessä lähteitä korvattiin työhön paremmin sopivilla. Artikkeleista etsittiin artikkelikohtainen viittausmäärää, tekijän julkaistujen artikkeleiden määrä, vertaisarvioituja tieteellisiä artikkeleita, joita on viitattu muissa tutkimuksissa. Viittausmäärissä havaittiin, että uusimmissa julkaisuissa viittauksia ei juuri ollut, vaikka kirjoittaja oli tunnettu tutkimusalallaan. Viittausten määrästä uusimpien artikkeleiden osalta luovuttiin ja kirjallisuuslähteitä etsittiin kirjoittajilta, jotka toimivat tutkijoina tai professoreina eri yliopistoissa. Artikkeleita löytyi runsaasti, mutta tarkastelunäkökulmaa vastaavien artikkeleiden kirjoittajia löytyi rajatumminkin. Koska työssä keskityttiin uusiin teknologia, artikkelitietokannoista etsittiin uusimpia tutkimuksia. Artikkeleita etsittiin

erityisesti vuosien 2015-2018 ajalta, mutta myös julkaisua odottaviin materiaaleihin tutustuttiin. Olennaisena lähteiden valinnassa pidettiin lähteiden kontekstia, laatua ja sisällön käsittelyä. Bibliometrisien tunnuslukujen määrittelyssä oli haasteita ja tulkinta sisältää virheitä. Taulukoissa 21 olevat tunnusluvut osoittavat kuitenkin suuruusluokat eri lähteistä ja niiden tyypeistä. Osa verkkojulkaisuista ja artikkeleista on esitetty seminaareissa tai osana toista julkaisua, jolloin tyyppi ja julkaisupäivä voivat ilmentyä ristiriitaisena tietona. Tarkkojen osuuksien kuvaaminen ilman tutkimusta antaisi virheellisen arvion osuuksista. Tavoitteeksi asetettujen lähteiden laadulle saatiin seuraavia bibliometrisiä tunnuslukuja

Taulukko 21. *Kolme yleisimmin käytettyä lähteiden tyyppiä kirjallisuustutkimuksessa sekä lähteiden lukumäärät.*

Lähteen laji (tyyppi)	Lähteet (lkm)
artikkeli	62
verkko- tai painettu julkaisu	15
opintomateriaali.	9

Olennaista lähteiden käytössä pidettiin lähteiden arvostettavuutta ja virheettömyyttä. Kirjojen osalta valittiin kirjoittajia, jotka ovat toimineet professoreina yliopistoissa. Artikkeleita valittiin tieteellisistä aikakauslehdistä, joissa julkaistujen artikkeleiden laatu on hyvä. Artikkeleiden osalta keskeisimmäksi lähteeksi valittiin teknologiaennustamiseen erikoistunut tieteellinen aikakauslehti (taulukko 22). Yliopistojen julkaisuissa arvioitiin kirjoittajien tutkimusala ja asemaa organisaatiossa.

Taulukko 22. *Kolme yleisimmin käytettyä lähettä ja niiden lukumäärät.*

Julkaisija (nimi)	Lähteet (lkm)
Technological Forecasting and Social Change	15
Yliopistojen julkaisut	11
kaupallinen toimija (muun muassa Gartner Reseach ja Mathworks).	8

Yliopistoista pyrittiin etsimään lähteitä eri yliopistoista. Kaupallisten toimijoiden osalta kolme käytetystä lähteestä edusti kaupallista tutkimuslaitosta (taulukko 22).

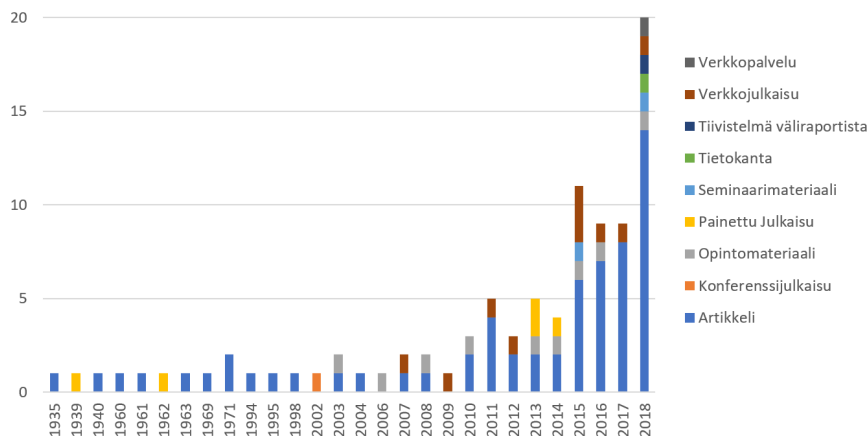
Kirjallisen lähteen vertaisarviota voi arvioida myös kirjoittajajoukon koolla, mikäli tutkimusta on tehnyt isompi ryhmä. Tällöin vertaisarviota voidaan tehdä jatkuvasti tutkimuksen aikana. Taulukossa 23 on esitetty kirjoittajien lukumäärä käytetyissä lähteissä lähde-tyypeittäin.

Käytetyissä lähteissä on bibliometrisiä sidonnaisuuksia. Kapeampi rajaaminen Bassin laskentamalliin lisäsi bibliometrisiä sidonnaisuuksia, joka toisaalta lisäsi tiedon ”yksi-muotoisuutta” sekä kritiikin vähenemistä. Uusimpien tutkimuksien osalta kritiikin löytäminen on haastavampaa.

Taulukko 23. Yleisimmät lähteiden tyypit sekä kirjoittajien lukumäärä lähdetyypeittäin.

Julkaisija (tyyppi)	Yksi (hlö)	Kaksi (hlöä)	Kolme (hlöä)	Yli kolme (hlöä)
Artikkeli	24	18	12	8
Verkkojulkaisu	8	2		
Opintomateriaali	6	3		
Painettu julkaisu	4	1		

Kirjallisuustutkimukseen etsittiin julkaisuja erityisesti vuosilta 2015-2018 painottuen uusimpiin artikkeleihin. Kuvassa 42 on esitetty pylväskuvaajana eri lähteiden tyypit vuosittain.



Kuva 42. Vasemmalla lähteet vuosiluvun perusteella. X-akselilla kaikkia vuosia ei esitetä, koska ennen vuoden 2006 lähteitä käytettiin lähinnä teorian esittäjien osalta.

Tässä työssä keskityttiin kirjallisuuteen, joka tutkii teknologiadiffuusiota ja siihen liittyviä ilmiöitä. Kirjallisuustutkimuksessa tuotiin esille teknologiadiffuusio laajana ilmiönä sekä esiteltiin erilaiset kuvaamiseen sopivia kasvumalleja. Kirjallisuudesta etsittiin tutkimuksissa käytettyjä tietolähteitä, syy-yhteyksiä ja tutkimusmenetelmiä sekä näiden kombinaatioita. Kirjallisuudesta etsittiin tutkimuksia, jotka käsittelevät teknologiaennustamisen eri näkökulmia ja teknologian kehittymisen eri vaiheita. Kirjallisuustutkimuksella etsittiin myös tutkimuksia, joissa oli keskitytty tarkempaan teknologiaennustamiseen. Kir-

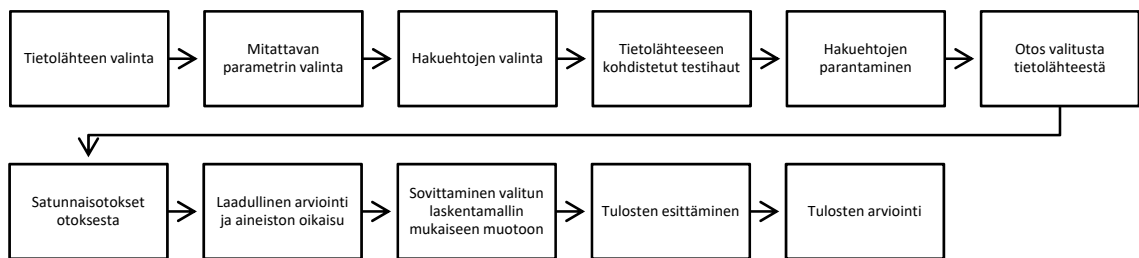
jallisuustutkimuksen perusteella päädyttiin ratkaisuun, jossa keskitytään Bassin laskentamalliin sekä useamman tietolähteen yhtäaikaiseen käyttöön. Bassin laskentamalli tarjoaa suosittuna laskentamallina runsaasti tutkimusta.

Kirjallisuustutkimuksessa asemoitiin kirjallisuudessa esitetyt teknologian kehittymisen vaiheet kronologisesti. Tällöin kronologiset vaiheet voitiin kytkeä osaksi teknologian kehittymistä. Kronologinen kytkentä mahdollisti tutkimuksen kontekstin ja tulosten paremman tulkinnan. Kirjoitustyön edetessä havaittiin, että kronologiset riippuvuudet ja siitä toteutettu luokitus muodostivat uutta tietoa. Samassa kronologisessa vaiheessa tapahtuvat ilmiöt korreloivat keskenään. Korrelointi ei tarkoita, että ilmiöillä olisi suoraa riippuvuutta keskenään.

Teknologian kehitymisellä on kuitenkin useita riippuvaisuuksia ja kirjallisessa osiossa tuodaan esille vain joitakin teknologiaan vaikuttavista asioista, jotka vaikuttavat teknologian yleistymiseen. Kirjallisessa työssä tutkittiin myös eri tietolähteiden käytön vaikutusta teknologin ennustettavuuteen. Teknologiadiffuusio ei noudata mitään kasvumallia, mutta osa malleista voi olla hyvä kuvaamaan teknologian kehittymistä. Valitsemalla kuvaava kasvumalli ja huomioimalla teknologian syy-yhteyksiä voidaan saavuttaa tarkka ennuste teknologian kehitymiselle pitkälle aikajaksolle.

4.2 Kvantitatiivinen tutkimusmenetelmä ja aineisto

Tutkimushypoteesiksi on asetettu kirjallisuustutkimuksen tulosten mukainen tietolähteiden määrien kasvamisen kronologinen riippuvuus. Hypoteesissa viitataan Cho ja Daim (2016) tutkimuksessa esitettyyn tutkimustulokseen, jossa patenttimäärien kasvu toteutuu teknologian kehittyemisessä artikkelimäärän kasvamisen jälkeen. Mittausparametrina käytetään artikkeli- ja patenttitietokantojen osalta hakutulosten lukumäärää ajanjaksolla. Laskettavaan lukumääriin huomioidaan ajanjaksolla julkaistut artikkelit ja patentit. Ensimmäisessä työvaiheessa käytettiin valittuja tietolähteitä (kuva 43).



Kuva 43. Työvaiheet tietolähteen valinnassa ja otantojen valmistelussa

Teknologiadiffuusion tutkimuksen patenttitietojen osalta tietolähteinä käytetään Patentti- ja rekisterihallituksen tarjoamaa Espacenet-tietokantaa. Espacenet on European Patent Office:n tietoihin perustuva tietokanta, jossa on luokiteltu yli 100 miljoonaa patenttia. Espacenet on valittu datalähteeksi avoimuuden perusteella, jotta työssä olevat lähtötiedot ovat saatavissa ja päätelmät ovat helpommin toistettavissa. Espacenet-tietokannassa on tietoja yli 90 maan patenteista ja Patentti- ja rekisterihallituksen mukaan Espacenetissä on suomalaisia patenteja kattavammin, kuin suomalaisessa patenttitietokannassa. Lähde Patentti- ja rekisterihallituksen Patentit-sivu (https://www.prh.fi/fi/patentit/palvelut_ja_tietokannat/maksuttomat_tietokannat/espacenetinhakuohjeita.html).

Espacenet -tietokannan IPC ja CPC -luokitusten perusteella tuloksia saatiin paljon, mutta osumatarkkuus oli heikko. IPC ja CPC -koodistoon on olemassa haku, jolla löytää sopivimmat luokat. Työssä käytettiin mm. https://fi.espacenet.com/classification?locale=fi_FI -hakupalvelua. Ainoastaan big dataan liittyviä patenttiluokituksia ei löydetty.

Koodistohakujen avulla toteutunut heikko saatavuus varmistettiin useilla vapailla tekstihauilla (Big data, ”Big data”) ja todettiin, että vaikka haku toimisikin, niin patenteja ei ole luokiteltu välttämättä big dataksi, vaan se on ainoastaan liitännäinen vaikkakin merkittävä osa teknologiaa.

Haut toistettiin vapaina tekstihakuina, jossa kohdistettiin otsikkokenttään big data-sanayhdistelmiä. Haku noudattaa Boolean algebraa, jossa voidaan lisätä ja poistaa hakutuloksiin (otokseen) haluttuja tuloksia.

4.3 Patenttietokantahaut ja aineistojen käsittely

Hakutulosten luotettavuutta mitattiin toteuttamalla haut tekstihakuina ja tarkastamalla tekstihakuihin saatujen patenttien koodistot. Poiminta poimii vain osajoukon kaikista patenteista, mutta otoskoko on riittävä (n=3735), jotta otoksen perusteella voidaan arvioida koko aineistoa. Taulukossa 24 kuvataan tarkemmin aineistotutkimuksen toimenpiteitä.

Taulukko 24. Patenttietokantaan kohdistuva aineistotutkimuksen toimenpiteet.

Tehtävä	Toimenpide	Tehtävä
1.	Tekstihaku poimii big dataan liittyvät patentit.	Toteutetaan otsikkokenttään kohdistetut haut Espacenet- patenttietokantaan ja tallennetaan tulokset.
2.	Etsitään virheelliset hakutulokset otoksesta	Lasketaan otoksesta big data -teknologiaan kuulumattomat hakutulosten lukumäärä ja arvioidaan poiminnan luotettavuutta.
3.	Poimituilla patenteilla on patenttiluokitus, joka tarkoittaa big dataa IPC-luokituksessa.	Testataan hypoteesia, että poiminnan mukaisella patentilla on IPC-luokitus.
4.	Poimituilla patenteilla on patenttiluokitus, joka tarkoittaa big dataa CPC-luokituksessa.	Lasketaan patenttien CPC-luokituskoodit poiminnasta ja testaan hypoteesia, että poiminnan mukaisella patentilla on CPC-luokitus
5.	Luokitellaan aineisto.	Luokitellaan aineisto tekstihakujen, IPC ja CPC-patenttikoodien perusteella.
6.	Aineistojen jalostaminen luokittelun ja ominaisuuksien perusteella.	Muodostetaan eri aineistot luokittelun perusteella, jossa ryhmitellään aineistoa eri aikadimensioilla (esim. vuosi, kuukausi).
7.	Kuvataan patenttien määrän kehittyminen.	Piirretään laskentamallin mukaiset S-käyrät teknologian elinkaarelle eri luokitelluille aineistoille.
8.	Kuvataan patenttietokannan aineisto-ongelmaa.	Verrataan IPC-tekstihakuun sekä CPC-tekstihakuun ja tekstihakuun tulleiden kooditettujen kasvukäyriä.

Hakutuloksista havaittiin, että virheellisten tulosten suhteellinen osuus oli suurempi teknologian varhaisessa vaiheessa, kuin myöhemmässä vaiheessa. Asiaa selvitettiin tarkastamalla koko otos ja esittämällä löydetty tulokset ilman hypoteesin testaamista.

Poimintatulos edustaa *äärellistä otosavaruutta*, josta tarkastettiin poimitun aineiston otsikkojen tarkoittama sisältö visuaalisesti. Tuloksesta poistettiin ne patentit, joiden kaikki seuraavat hakuparametrit ovat yhtäaikaaisesti voimassa:

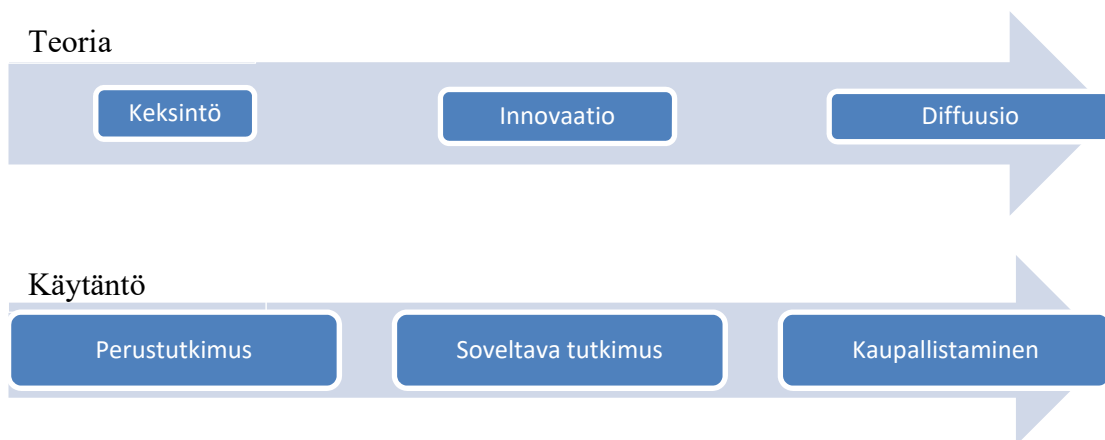
- 1) otsikon sisältö sisältää termin big data
- 2) sanat esiintyvät yhteydessä, jossa ei tarkoiteta big dataa teknologiana
- 3) patenttiluokitus puuttuu (G06F 17/30), jossa CPC-luokituksessa **G 06** tarkoittaa laskentaa ja **F** Sähköinen digitaalinen tietojenkäsittelyä ja siihen liittyvää tiedonhaku (”G06F = electrical digital data processing”, 17/30 = ”information retrieval”).

Suodatetun aineiston (n=3606) otsikointi sisälsi pääosin uniikkeja otsikoita (n=3357) ja otsikon tulkinta sisältää kohtuullisen suuren todennäköisyyden ($P(\text{tulkintavirhe}) > 0.05$) virheelliseen arvioon patentin varsinaisesta sisällöstä, mutta jos virhe on samaan suuntaan systemaattinen, niin sen tuoma virheellinen vaikutus on pienempi ennusteen luomisessa.

5. TULOKSET JA POHDINTA

5.1 Kirjallisuustutkimuksen tuloksia

Tämän työn kirjallisuustutkimuksessa on käyty läpi innovaatioiden vaiheiden kehittymistä. Useista tutkimuksien esittämistä vaiheista tehtiin arvio kronologisesta järjestyksestä, jota tutkimuksessa on käsitelty. Osassa tutkimuksia on esitetty eri teknologian vaiheita ilman teorian ja käytännön erottelua. Jamasb (2007) esitti innovaatiotutkimuksessa teorian ja käytännön vaiheet samassa kuvauksessa, mutta erotteli ne omiksi kokonaisuuksineen.



Kuva 44. Mukaelma Jamasb (2007) esittämästä kuvasta, jossa on kuvattu perustutkimuksen, sovelletun tutkimuksen ja kaupallisuuden vaiheista vastaavien innovaatioteorian vaiheiden yhteydessä.

Kuvauksesta on tässä työssä tehty mukaelma (kuva 44), jossa on pyritty säilyttämään alkuperäisen kuvauksen mukainen järjestys ja mittasuhteet. Kuvauksessa hän esittää teoriaan kuuluvan *keksintö-innovaatio-diffuusio* -vaiheet. Vastaavasti käytäntöön kuvattiin tutkimusnäkökulma (*perustutkimus – sovellettu tutkimus – kaupallistaminen*). Jamasb (2007) esittää perustutkimuksen ja keksinnön välisen yhtäaikaisuuden innovaation kehittymisen varhaisissa vaiheissa.

Tämän työn kirjallisuustutkimuksessa on käsitelty useita tutkimuksia, joissa on esitetty teknologian kehittymisen kronologisia vaiheita. Kirjallisuutta tarkemmin tutkimalla voidaan havaita, että kronologisia vaiheita esitetään teknologiaennustamisessa runsaasti. Kuvaamalla eri ilmiöiden kronologiset vaiheet, voidaan edellä olevilla vaiheilla ennustaa tulevia vaiheita. Esitetyt kronologiset vaiheet koottiin ja kronologisia vaiheita hyödynnetään kvantitatiivisessa tutkimuksessa. Kronologisilla vaiheilla on myös keskinäisiä riippuvuuksia eli samassa vaiheessa voi indikoitua useita ilmiöitä. Tällaisia ilmiöitä voivat olla esimerkiksi tutkimuksen ja keksintöjen välinen positiivinen korrelaatio. Kir-

jallisuustutkimuksessa esitetyt kronologiset järjestykset kuvattiin välivaiheena tauluk-
koon 25. Taulukon 25 tutkimuksissa esitetyt ilmiöt sijoitetaan kronologisesti taulukkoon
26 ja 27, jotka luokitellaan kuuluvaksi uuteen vaiheistukseen.

Taulukko 25. *Arvio eri tutkimuksissa esiintyneistä tietolähteistä, joiden perusteella voidaan ennakoida tulevia teknologioita.*

Tutkija	Esitetty kronologinen järjestys teknologian kehittämisessä
Schumpeter (1934)	1) Keksintö-2) Innovaatio 3) Diffuusio
Ogburn (1937)	1) Keksintö ja sen kehittäminen – 2) Diffuusio ja käyttö – 3) Innovaation vaikutukset
Ogburn ja Gilfillan (1933)	1) Idea 2) kokeiltava laite 3) malli tai suunnitelma 4) ensimmäinen demonstraatio 5) käytännön laite 6) vakiintunut käyttö 7) laaja-alainen
Ogburn (1937b)	1) Idea 2) malli 3) testi 4) kehittäminen 5) markkinointi 6) myynti 7) käyttö 8) vaikutukset
Ogburn (1937a)	1) Idea 2) suunnitelma tai malli 3) design 4) parannukset 5) myynti 6) markkinointi 7) laajan mittakaavan tuotanto
Cho ja Daim (2016)	1) Palkitut tutkimukset 2) konferenssijulkaisut 3) Artikkelit 4) Patentit
Ogburn ja Nimkoff (1940)	1) Idea 2) kehittäminen 3) malli 4) keksintö 5) parannukset 6) markkinointi
Ogburn (1941)	1) Idea 2) suunnitelma 3) konkreettinen muoto 4) parannukset 5) tuotanto 6) markkinointi 7) myynti
Bass (1969), Rogers (2003)	1) Innovaattorit 2) Varhaiset omaksijat 3) Varhainen enemmistö 4) Myöhäinen enemmistö 5) Viimeiset käyttöönottajat
Kaminski (2011)	1) Varhaiset omaksijat 2) kriittinen massapiste 3) Varhainen enemmistö
Bass (1969)	1) Innovaattorit 2) Seuraajat
Xiong <i>et al.</i> (2016)	1) Tieto vertaisvaikutuksessa 2) Kokemukset vertaisvaikutuksessa 3) Ulkoiset vertaisvaikutukset
Mahajan <i>et al.</i> (1995)	1) Ulkoisesta vaikutuksista johtuva adaptio/Mainonta, 2) Sisäisestä vaikutuksista johtuva adaptio/vertaisvaikutus
Forström (2011)	1) "Ulkoisen vaikutus/mainonta 2) Sisäiset vaikutukset/vertaisvaikutukset"
Linden ja Fenn (2003) a	1) Tyypin A yritykset pilotoivat 2) Tyypin B yritykset pilotoivat 3) Tyypin c yritykset pilotoivat
Linden ja Fenn (2003) b	1) Innovaation käynnistys eli teknologian löytyminen 2) Hypehuippu 3) Pettymysten aallonpohja 4) Valoisampi nousu 5) Tasaisempi kannattavuus.

Brescian ja Eppler (2010); Lindenin ja Fenn (2003)	1) Ensimmäinen hypeaalto – 2) Liiketoiminnallinen kypsytys ja toinen hypeaalto – 3) Adaptio eli diffuusio
Mahajan <i>et al.</i> (1995)	1) Tieto vertaisvaikutuksessa 2) kokemukset vertaisvaikutuksessa 3) ulkoiset vaikutukset vertaisvaikutuksessa
Sassali (2014)	1) Web-sivut 2) Sanomalehdet 3) Kokousjulkaisut 4) Aikakauslehdet 5) Hakuteokset, kirjat ja patentit
Cho ja Daim (2016)	1) Perustutkimus 2) Sovellettu tutkimus 3) patentit
Jamasb (2007)	1) Perustutkimus 2) sovellettu tutkimus 3) kaupallistaminen
Jamasb (2007)	1) Technology Push 2) Market Pull
Huang <i>et al.</i> (2018)	1) Aihiot rahoitettaville kohteille 2) Rahoitus ja ympäristön tuki 3) Yliopistot/ Tutkimus 4) "Technology push" 5) Kehitetyt ilmiöt 6) Tuotteistus
Martin (2013)	1) Google Trends 2) Tilastot

Kronologiset vaiheet tunnistettiin tutkimuksien perusteella kuuluvan eri kokonaisuuksiin. Kokonaisuudet tunnistettiin tutkimuskonteksteista eli yksittäisistä tutkimuksista, jotka käsittelevät esimerkiksi rahoitusta, tutkimusta tai innovaatioidiffuusiota. Tutkimuksien eri lähteet kuvaavat teknologian kehittymisen eri ajallisia vaiheita. Taulukon 24 kronologisesti järjestetyt osa-alueet voidaan arvioida kuuluvan neljään eri luokkaan sisällön perusteella. Osa-alueet tässä arvioissa voisivat olla seuraavat:

1. rahoitus
2. tutkimus
3. innovaatio
4. käyttöönotto.

Perusteena esitetylle järjestykselle käytetään taulukossa 24 mainittuja tutkimuksia ja niiden käsittelemiä teorioita ja tuloksia. Myös tutkimuksissa toteutetut kontekstien rajaukset vaikuttivat kyseisen luokittelun valintaan.

Taulukossa 26 on sijoitettu taulukon 25 tutkimuksien osa-alueet neljään luokkaan. Tutkimuksen lähteen sijoittaminen sisältää kohtuullisen suuren virhemahdollisuuden, jonka virhe pienenee toteuttamalla tarkempi tutkimus koostamalla useista eri tutkimuksissa esitetyistä ajallisesti järjestettävistä ilmiöistä.

Taulukko 26. *Arvio tutkimuksien sisällön sijoittumisesta esitettyyn rahoitus, tutkimus, innovaatio ja diffuusio -luokitukseen. Arvio sisältää virhemahdollisuuksia ajallisesti sekä riippuen tutkittavasta kokonaisuudesta.*

Vaiheet	osavaihe 1	osavaihe 2	osavaihe 3	osavaihe 4	osavaihe 5
Vaihe 1: Rahoitus	Tarve, ongelma, kustannukset, haaste tai muu peruste kehittää vaihtoehtoja.	Rahoitusehdotukset	Yritys- ja hallinnollinen tutkimusrahoituspäätös		
Vaihe 2: Tutkimus	Ensimmäisten rahoitusten ja tutkimusten käynnistyminen.	Perustutkimus (artikkelit), keksinnöt	Sovellettu tutkimus, keksinnön kehittäminen	Uudet teknologiat, keksinnän kehittäminen	Viimeiset loppuraportit, (tutkimusrahoitusten loppumisen).
Vaihe 3: Innovaatio	Keksinnön kehittäminen ja soveltaminen	Patenttien haku	Mainonta ja markkinointi, ensimmäinen hypeaalto alkaa	web-sivut	Technology push, innovation trigger, ensimmäinen hypeaalto. Tyypin A yritykset pilotoivat
Vaihe 4: Käyttöön-otto eli diffuusio	Tuotteet ja kaupallisuus, Adaption: innovaattorit, toinen hypeaalto alkaa	Seuraajat: varhaiset omaksijat, patenttien myöntö, (kriittinen massapiste lähestyy loppuvaiheessa).	Seuraajat: aikainen enemmistö, Kriittinen massapiste ylittyi eli kasvu tai laskuvaihe. Kaupallinen kannattavuus on mahdollista	Seuraajat: myöhäinen enemmistö, Kaupallinen kannattavuus heikkenee.	Seuraajat: viimeiset käyttöönottajat. Kaupallinen kannattavuus on heikentynyt. Innovaation vaikutukset, kuten teknologian haittojen ilmeneminen.

Tutkimuksien kronologinen sijoittaminen sisältää virhemahdollisuuden, koska eri tutkimuksien välillä on vain ”semanttisia heikkoja yhteyksiä” ja tutkimukset eivät ole yhteismitallisia tai muuten ole menetelmiltään verrattavia. Myös otoksen koko oli rajallinen.

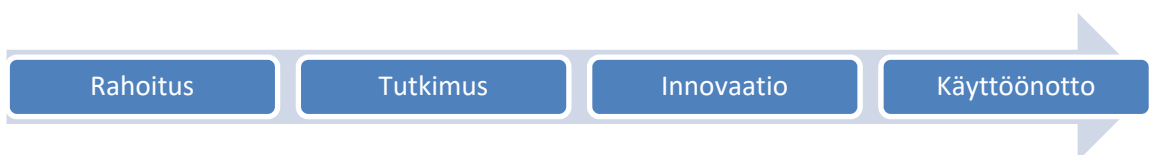
Teknologiaennustuksessa on olennaista selvittää mihin ennustetta tarvitaan ja valita ennustukseen sopiva lähde. Rahoitusehdotuksesta viimeisiin käyttöönottajiin voi kulua vuosia, jolloin vanhoja asioita tutkimalla voi ennustustarkkuus kärsiä. Valitsemalla tietolähteeksi ainoastaan kirjallisuuden tai patenttitietokannan, voi ajallisesti varhaisemman vaiheen ennustaminen tuottaa valmiiksi vanhentuneen ennusteen. Web-sanahakujen hyödyntäminen voi olla reaaliaikaista, jolloin se ei varsinaisesti ennusta nykyhetkeä. Vastavasti tilasto myönnetyistä patenteista kuvaa mennyttä aikaa, mutta toisaalta tilasto tutkimusrahoituksesta voi jo ennustaa patenttien kehittymistä ja kaupallista sovellettavuutta tulevaisuudessa. Aineistojen tunteminen helpottaa oikean aineiston valintaa. Samoin tutkimuskonteksti on helpompi huomioida, jos tuntee aineistot ja niiden ominaisuudet. Valitsemalla tutkimuskontekstiin sopivat ja kronologisesti eroavat aineistot saadaan selkeämpi ennuste, jolla voidaan ennakoita teknologian kehittymisen seuraavia vaiheita.

Taulukossa 27 esitetään arvioita mahdollisesta luokittelusta. Arvion avulla luokitellaan kirjallisuudessa esitettyjen tietolähteiden kronologinen järjestys. Taulukossa 25, 26 ja 26 on esitetty aineistot sekä tässä työssä hyödynnetyt tutkimuksia tulkintoineen. Valitsemalla esimerkiksi tutkimuksesta, innovaatioista ja diffuusiosta tietolähteitä, saavutetaan kronologinen esitys ja tarkempi ennuste orastavan teknologian diffuusiosta.

Taulukko 27. *Tutkimuksien esitettyjen vaiheiden (taulukko 26) sijoittuminen teknologian kehittämisessä. Tutkimuksien (taulukko 25) käsittelemät vaiheet on kronologisesti järjestetty neljään luokkaan.*

	Rahoitus	Tutkimus	Innovaatio	Diffuusio
Schumpeter (1934)				
Ogburn (1937)				
Ogburn ja Giffllan (1933)				
Ogburn (1937b)				
Ogburn (1937a)				
Cho ja Daim (2016)				
Ogurb ja Nimkoff: (1940)				
Ogburn (1941)				
Bass (1969), Rogers (2013)				
Kaminski (2011)				
Bass (1969)				
Xiong <i>et al.</i> (2016)				
Mahajan <i>et al.</i> (1995)				
Forström (2011)				
Linden ja Fenn (2003 a)				
Linden ja Fenn (2003 b)				
Brescian ja Eppler (2010)				
Mahajan (1995)				
Sassali (2014)				
Cho ja Daim (2016)				
Jamas B (2007a)				
Jamas B (2007b)				
Huang <i>et al.</i> (2018)				
Martin (2013)				

Taulukosta 27 havaitaan, että tutkimus (tummennettu alue) rajataan valittuun vaiheeseen ja teknologian kehittyminen voi sisältää useita edeltäviä, mutta myös seuraavia vaiheita.



Kuva 45. *Esitetty luokittelu sekä sen kronologinen järjestys teknologian kehittämiselle.*

Cho ja Daim (2016) hyödynsivät useita tietolähteitä saavuttaakseen tarkkuutta ennusteesseen integroimalla aikaisemman epätarkan ja nykyisemmän tarkan tietolähteen yhtäaikaista. Tutkimuksessaan he pystyivät todentamaan eri tietolähteiden kehittämisellä olevan useamman vuoden ero (Cho & Daim 2016). Useamman lähteen valinnasta saatiin selvää hyötyä myös Lee, Choi ja Sohn (2018) tutkimuksessa, jossa valittiin web-hakuliikenne ja patenttietokanta tietolähteeksi. Tämän tutkimuksen mukaan ennustarkkuus parani olennaisesti, kun molempien tietolähteiden ominaisuuksia voitiin hyödyntää (Lee *et al.* 2018).

Tässä työssä on esitetty, että useamman lähteen käyttö voi parantaa teknologian ennakkoinnin tarkkuutta, joka edellä mainituissa tutkimuksissa on ollut tavoitteena.

5.2 Kvantitatiivinen tutkimus big data -ilmiön kehittämisestä

Kvantitatiivisessa tutkimuksessa hyödynnettiin Science Port-palvelua ja vapaata Espacenet-patenttitietokantaa. Science Port-palvelussa on saatavilla 253 tietokantaa (11/2018). Tutkimukseen valittiin kirjallisuustutkimuksessa esitetyjä tietokantoja. Tietolähteiden valinta perusteltiin tarkemmin kohdassa 4.2. Tarkempaan tutkimukseen valittiin neljä tietokantaa ja 15 eri tietolähdettä, joista neljä tietolähdettä rajattiin pois. Rajaamisen syynä oli työmäärän rajaaminen sekä tietojen yhteismitallisuuden ongelmat aineistoissa. Tietokantoihin toteutettiin erilaisia hakuja. Hakuparametreina kokeiltiin hakuja, jolla mitataan hakutulosten määriä. Hakuja toteutettiin big data -termien yhdistelmillä ja päädyttiin ratkaisuun, jossa ”big data”-termiyhdistelmä rajataan lainausmerkeillä kaikissa hakuparametreissa. Rajaaminen paransi haun tulosten laatua. Erilaisten hakutestien tuloksena Proquest -tietokantaan päädyttiin tekemään kaksi erityyppistä hakuja. Ensimmäinen haku kaupallisiin tietueisiin ja toinen artikkelitietokantaan muilta osin samoilla hakuparametreilla. Kaupalliset tietueet valittiin Huang *et al.* (2018) tutkimuksen perusteella. Tietokantahaut kohdistettiin Proquest tietokantaan ”Any where” -optiolla. Osittain samoilla hakuparametreilla saavutetaan parempi yhteismitallisuus (ProQuest-tietokannan osalta).



Kuva 46. Tekniset työvaiheet kvantitatiivisessa tutkimuksessa ennustusmallissa. Kvantitatiivinen tutkimus aloitetaan ”Kerätään data”-vaiheella.

Kuvassa 46 esitetään selvyyden vuoksi teknisiä työvaiheita tästä työvaiheesta. Taustatyönä on toteutettu datan käsittelyyn sekä laskentaan sopivia ohjelmia sekä prosesseja.

Taulukossa 28 on esitetty kvantitatiivisessa tutkimuksessa käytetyt asiakirjatyypit sekä tietokannat. Haut toteutettiin alkaen vuodesta 1984 ja päättyen 2017 loppuun (jatkossa vuosien 1984-2017 ajalta).

Taulukko 28. *Kvantitatiivisessa tutkimuksessa käytetyt tietolähteet.*

Aineisto	Tietokanta	Tietolähdetyyppi
1	Proquest	Kaupalliset artikkelit
2	Proquest	Tieteelliset artikkelit
3	Web of Science	Tieteelliset julkaisut
4	Web of Science	Artikkelit
5	Web of Science	Toimitukselliset materiaalit
6	Web of Science	Katsaukset
7	Web of Science	Asiakirjatyypinä kokousmuis-
8	Scopus	Artikkelit
9	Scopus	Konferenssi julkaisut
10	Scopus	Konferenssikatsaukset
11	Espacenet	Patentit.

Palveluissa, kuten Scopus ja Web of Science tarjotaan erilaisia toimintoja, joilla tuloksia saa helposti esille. Muun muassa vuosittainen summatason haku tai laskenta onnistui tarjotuissa palveluissa. Summatasojen käyttö palveluissa ei ota kantaa aineiston laatuongelmiin. Tuloksia teknologian ilmenemisestä voidaan saada jo ennen teknologian löytymistä. Aineiston laadullisia haasteita käytiin läpi kohdissa 4.3, 5.3, 5.4, 5.5 ja 5.6. Hakuparametrien vaikutusta otantaan vertailtiin taulukon 29 mukaisesti.

Taulukko 29. *Hakuparametrivaihtoehtojen vaikutuksen testaus tietokantoihin.*

Tietokanta	Termi: big data	Termi: "big data"
Espacenet	6 000 (otsikkoon kohdistuva haku)	5 000 (otsikkoon kohdistuva haku)
Proquest	6 000 000	430 000
Scopus	106 000	54 000
Web of Science	63 000	30 000

Aineistolaatuongelmat vaativat käytännössä aineiston rivi- ja alkiokohtaisen käsittelyn, jolloin aineiston laatua voidaan arvioida esimerkiksi tarkoitukseen sopivalla teknisellä menetelmällä.

Taulukossa 30 on esitetty tietokannat, kohdealueet ja parametrit, joita hyödynnettiin tietokantahauissa. Haun tulokset tallennettiin mahdollisuuksien CSV tai Excel-muotoon jatkokäsittelyä varten.

Taulukko 30. *Kvantitatiivisessa osan tietokantahauissa hyödynnetyt tietokannat, kohdealueet ja hakuparametrit.*

Aineisto	Tietokanta	Kohde ja hakuparametrit
1	Proquest	Missä tahansa hakutuloksessa "big data" -termiyhdistelmä. Dokumenttityyppi: Kaupalliset artikkelit. Aikajakso: 1984-2017.
2	Proquest	Missä tahansa hakutuloksessa "big data" -termiyhdistelmä. Dokumenttityyppi: Tieteelliset artikkelit. Aikajakso: 1984-2017.
3	Web of Science	Otsikko: ("big data"). Dokumenttityyppi: Tieteelliset julkaisut. Aikajakso: 1984-2017: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
4	Web of Science	Otsikko: ("big data"). Dokumenttityyppi: Artikkelit. Aikajakso: 1984-2017: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
5	Web of Science	Otsikko: ("big data"). Dokumenttityyppi: Toimituksellinen materiaali. Aikajakso: 1984-2017: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
6	Web of Science	Otsikko: ("big data"). Dokumenttityyppi: Katsaus. Aikajakso: 1984-2017: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
7	Web of Science	Otsikko: ("big data"). Dokumenttityyppi: Kokoustiivistelmä. Aikajakso: 1984-2017: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
8	Scopus	Otsikossa, tiivistelmä tai avainsanoissa "big data" -termiyhdistelmä, joissa asiakirjatyypinä artikkeli. Aikajakso 1984-2017
9	Scopus	Otsikossa, tiivistelmä tai avainsanoissa "big data" -termiyhdistelmä, joissa asiakirjatyypinä konferenssijulkaisu. Aikajakso 1984-2017
10	Scopus	Otsikossa, tiivistelmä tai avainsanoissa "big data" -termiyhdistelmä, joissa asiakirjatyypinä konferenssijulkaisu. Aikajakso 1984-2017
11	Espacenet	Patenttitietokantahaut, joissa "big data" esiintyy (oikaistu aineisto). Aikajakso 1984-2017.

Hakutulokset eri poiminnoista koottiin taulukkoon 30. Patenttitietokannan osalta hyödynnettiin aineiston tarkempaa tutkimusta. Taulukossa 30 on esitetty poimintatulokset 11 tietolähteen osalta. Patenttitietokannan osalta tulokset ovat oikaistuja tuloksia, jossa on huomioitu ja vältetty kohdassa 5.6 esitettyjä aineistovirheitä.

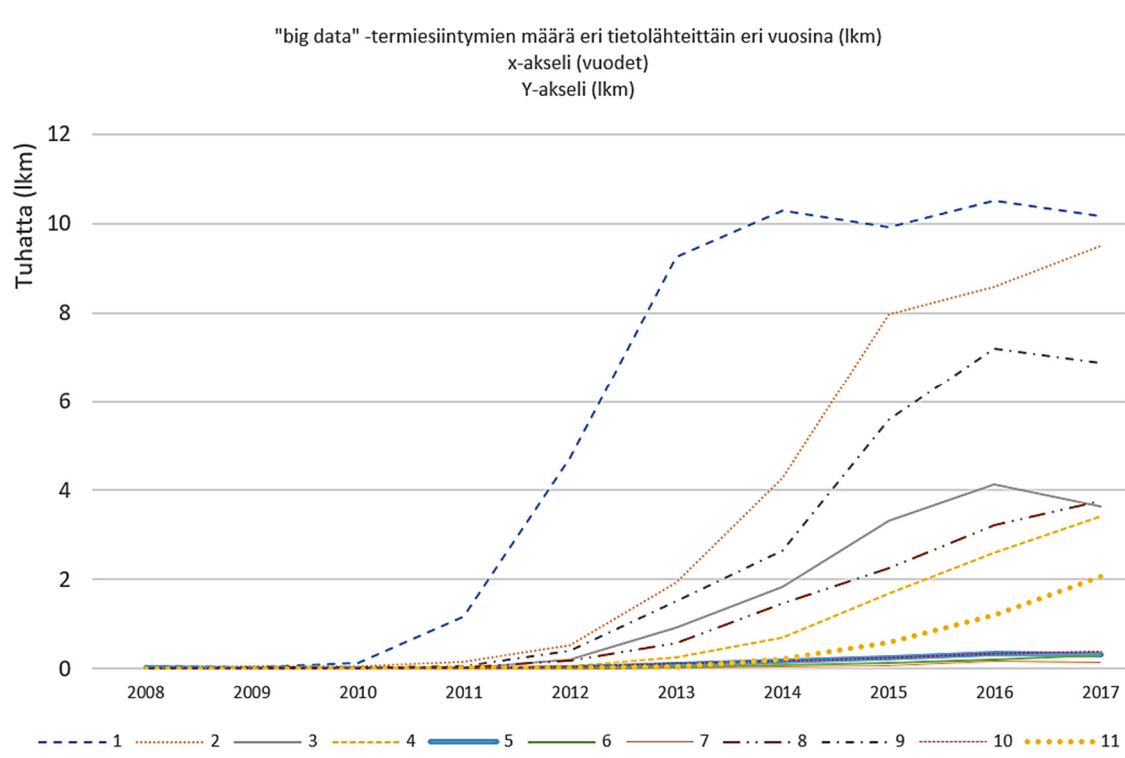
Muiden aineistojen (kymmenen aineistoa) osalta oikaisua ei toteutettu, vaan hyödynnettiin kohdassa 3.2 ”Teknologianennustamisen aineistot kvantitatiivisessa tutkimuksessa” esitettyjä asioita aineiston käyttäytymisestä big dataa tutkittaessa.

Taulukko 31. *Poimintojen välitulokset yhdestätoista eri otoksesta, joita hyödynnetään kvantitatiivisessa tutkimuksessa.*

	Proquest: Kaupalliset artikkelit (Trade Journals), hakuparametri rajoitettu termiyhdistelmään "big data"	Proquest: Scholarly Journals, hakuparametri rajoitettu termiyhdistelmään "big data"	Web of Science: Otsikko: ("big data"), Refined by: DOCUMENT TYPES: (PROCEEDINGS PAPER), Timespan: 1984-2017. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	Web of Science: Otsikko: ("big data"), Refined by: DOCUMENT TYPES: (ARTICLE), Timespan: 1984-2017. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	Web of Science: Otsikko: ("big data"), Refined by: DOCUMENT TYPES: (EDITORIAL MATERIAL), Timespan: 1984-2017. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	Web of Science: Otsikko: ("big data"), Refined by: DOCUMENT TYPES: (REVIEW), Timespan: 1984-2017. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	Web of Science: Asiakirjatyyppinä kokoustilviteilmät (meeting abstract) ; hakuparametri rajoitettu termiyhdistelmään "big data"	Scopus: asiakirjatyyppinä Artikkelit ; hakuparametri rajoitettu termiyhdistelmään "big data"	Scopus: Asiakirjatyyppinä Konferenssi julkaisut (Conference Paper), ; hakuparametri rajoitettu termiyhdistelmään "big data"	Scopus: Asiakirjatyyppinä Conference Review. ; hakuparametri rajoitettu termiyhdistelmään "big data"	Patentit: Espacenet, ; hakuparametri rajoitettu termiyhdistelmään "big data"
Vuodet \ Aineisto	1	2	3	4	5	6	7	8	9	10	11
1984	0	0	0	0	0	0	0	0	0	0	0
1985	0	0	0	0	0	0	0	0	0	0	0
1986	2	1	0	0	0	0	0	0	0	0	0
1987	0	1	0	0	0	0	0	1	0	0	0
1988	1	0	0	0	0	0	0	0	0	0	0
1989	2	0	0	0	0	0	0	0	0	1	0
1990	1	1	0	0	0	0	0	0	0	1	0
1991	5	0	0	0	0	0	0	1	0	1	0
1992	11	2	0	0	0	0	0	0	0	0	0
1993	8	1	0	1	0	0	0	0	0	0	0
1994	8	5	0	1	0	0	0	0	0	0	0
1995	18	4	0	0	0	0	0	0	0	1	0
1996	30	3	1	0	0	0	0	0	1	1	0
1997	35	6	0	0	0	0	0	0	0	1	0
1998	47	6	0	0	1	0	0	1	0	0	0
1999	45	9	3	1	0	0	0	3	0	0	0
2000	46	4	1	0	0	0	0	0	3	0	0
2001	38	5	4	0	0	0	0	0	5	1	0
2002	32	8	0	1	0	0	0	1	1	1	0
2003	23	16	1	2	0	0	1	4	4	2	0
2004	36	16	4	2	0	0	1	4	4	2	0
2005	36	11	2	0	0	0	0	2	6	5	6
2006	38	11	5	2	1	0	1	6	5	3	1
2007	38	15	3	0	1	0	2	3	5	2	2
2008	43	34	5	1	5	0	1	1	7	4	0
2009	35	43	10	4	0	1	2	10	18	3	2
2010	137	63	12	4	0	1	2	7	21	9	2
2011	1161	142	8	14	4	2	0	30	43	9	5
2012	4732	516	195	52	27	2	3	177	410	33	7
2013	9262	1922	917	262	82	18	14	562	1513	118	51
2014	10279	4275	1825	693	168	72	35	1472	2642	165	209
2015	9925	7954	3306	1689	239	133	77	2261	5590	249	575
2016	10500	8588	4125	2602	346	213	157	3213	7191	332	1195
2017	10166	9497	3648	3426	318	299	142	3771	6866	389	2077

Kuvassa 47 on visualisoitu välivaiheena taulukossa 30 esitetyt poimintatulokset, joiden poimintadatasta arvioidaan Bassin mallin mukaiset laskentaparametrit. Laskentapareilla tarkoitetaan tässä kontekstissa innovaattoreiden ja seuraajien osuuksia kuvaavia p

ja q -parametreja, jotka arvioidaan erikseen joka otokselle ($n=11$ kpl). Laskenta tuottaa myös m -parametrin, mutta se ei ole esityksessä relevantti, jos kasvukäyrät suhteutetaan vastaamaan määrällisesti toisiansa. Kuvalla 47 myös visualisoidaan ”big data”-ilmiön varhaisia vaiheita ja kuvan perusteella voidaan todeta, että havaintoyksiköiden määrät ovat vähäisiä kaikissa otoksissa ennen vuotta 2010.



Kuva 47. Poiminnan välitulosten havainnollistaminen. Poimintatuloksesta voidaan huomata markkinoinin (S1) aikaisempi kehittyminen. Eri tietokantoihin toteutettujen otantojen perusteella otoksiin vuosien 1975–1987 ajalta ei saatu kuin yksittäisiä osumia ”big data”-termiyhdistelmillä.

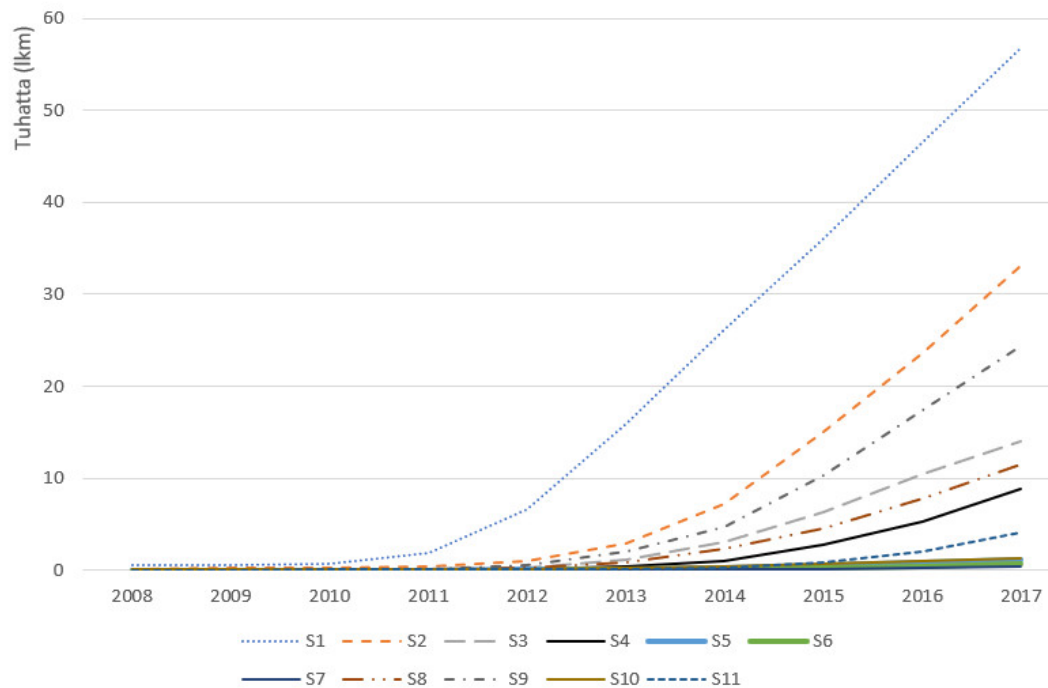


Kuva 48. Työvaiheen esittely ja sijoittuminen kvantitatiivisessa tutkimuksessa.

Työvaiheessa (kuva 48) lasketaan kumulatiiviset summat vuosittaisista havaintoyksiköiden lukumääristä. Laskenta toteutettiin Excel-ohjelmistolla. Kuvassa 49 on laskettu aineistojen kumulatiiviset summat. Summia esitetään symboleilla $S(n)$, missä n vastaa tau-

lukossa 31 ja kuvassa 47 esiintynyttä aineistoa (1-11). Kumulatiivinen summa voi selvittää aineistojen eroja havaintojen määrässä, mikäli kasvu on tapahtunut pitkällä aikajaksolla.

Kaupallisten tietueiden määrä (S1) on Kuvassa 49 varhaisimmassa vaiheessa kasvanut vuonna 2011. Kuvan 49 ja taulukon 32 kumulatiivisten määrien kehittyminen osoittaa teknologiadiffuusion varhaisien vaiheiden oleva tasaisessa kasvussa jo vuonna 2012.



Kuva 49. "big data" -termiyhdistelmän sisältävät hakutulokset kvantitatiivisessa tutkimuksessa, jossa toteuttiin haku 11 eri tietokantaan.

Teknologian kehittymisen varhaisien vaiheiden kehittyminen voi ennakoida teknologiadiffuusion tulevaa kasvua.

Taulukko 32. Lasketut kumulatiiviset määrät valituille ($n=11$) lähdeaineistoille. Ot-sikoissa $S(n)$ kuvaa summaa kaikilta vuosilta.

vuosi	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
1984	0	0	0	0	0	0	0	0	0	0	0
1985	0	0	0	0	0	0	0	0	0	0	0
1986	2	1	0	0	0	0	0	0	0	0	0
1987	2	2	0	0	0	0	0	1	0	0	0
1988	3	2	0	0	0	0	0	1	0	0	0
1989	5	2	0	0	0	0	0	1	0	1	0
1990	6	3	0	0	0	0	0	1	0	2	0
1991	11	3	0	0	0	0	0	2	0	3	0
1992	22	5	0	0	0	0	0	2	0	3	0
1993	30	6	0	1	0	0	0	2	0	3	0
1994	38	11	0	2	0	0	0	2	0	3	0
1995	56	15	0	2	0	0	0	2	0	4	0
1996	86	18	1	2	0	0	0	2	1	5	0
1997	121	24	1	2	0	0	0	2	1	6	0

1998	168	30	1	2	1	0	0	3	1	6	0
1999	213	39	4	3	1	0	0	6	1	6	0
2000	259	43	5	3	1	0	0	6	4	6	0
2001	297	48	9	3	1	0	0	6	9	7	0
2002	329	56	9	4	1	0	0	7	10	8	0
2003	352	72	10	6	1	0	1	11	14	10	0
2004	388	88	14	8	1	0	2	15	18	12	0
2005	424	99	16	8	1	0	2	17	24	17	6
2006	462	110	21	10	2	0	3	23	29	20	7
2007	500	125	24	10	3	0	5	26	34	22	9
2008	543	159	29	11	8	0	6	27	41	26	9
2009	578	202	39	15	8	1	8	37	59	29	11
2010	715	265	51	19	8	2	10	44	80	38	13
2011	1876	407	59	33	12	4	10	74	123	47	18
2012	6608	923	254	85	39	6	13	251	533	80	25
2013	15870	2845	1171	347	121	24	27	813	2046	198	76
2014	26149	7120	2996	1040	289	96	62	2285	4688	363	285
2015	36074	15074	6302	2729	528	229	139	4546	10278	612	860
2016	46574	23662	10427	5331	874	442	296	7759	17469	944	2055
2017	56740	33159	14075	8757	1192	741	438	11530	24335	1333	4132



Kuva 50. Työvaiheen esittely ja sijoittuminen kvantitatiivisessa tutkimuksessa.

Kuvassa 50 on kuvattu työvaihe ja sen riippuvuus edellisiin työvaiheisiin. Ohjelmalla (3) saadaan dataan perustuen ratkaistua p ja q -parametrit. Ohjelmassa (3) hyödynnetään kirjallisuustutkimuksessa esitettyä ratkaisutapaa. Taulukon 32 kumulatiiviset summat on sijoitettu ohjelmaan (3). Ohjelman suoritusten tuloksena saadaan datasta arvioidut innovaattoreita ja seuraajia kuvaavat parametrit p ja q . Parametrit laskettiin kaikkien poimintojen hakutuloksista.

Lähde: Perustuu lähteeseen <https://www.mathworks.com/matlabcentral/answers/410081-errors-in-objective-function>, jossa on useita Mathwors -työtilan käyttäjiä.

```
%data=[0;0;0;37248;58649;85974;132620;200698;266406;325423;383975;443234;498503;558777;612881;659367;710580;749772;794515;847288;910312;96374
6;1016473;1060359;1097173;1135293;1176867;1212260;1239208;1267657];
%data1=[3; 5; 6; 11; 22; 30; 38; 56; 86; 121; 168;213;259;297;329;352;388;424;462; 500;543;578;715;1876;6608;15870;26149;36074;46574;56740];
%data2=[ 2; 2; 3; 3; 5; 6; 11; 15; 18; 24; 30; 39; 43; 48; 56; 72; 88; 99; 110; 125; 159; 202; 265; 407; 923; 2845; 7120; 15074; 23662;
33159];
%data3=[0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 4; 5; 9; 9; 10; 14; 16; 21; 24; 29; 39; 51; 59; 254; 1171; 2996; 6302; 10427; 14075];
%data4=[0; 0; 0; 0; 1; 2; 2; 2; 2; 2; 3; 3; 4; 6; 8; 8; 10; 10; 11; 15; 19; 33; 85; 347; 1040; 2729; 5331; 8757];
%data5=[0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 1; 1; 1; 2; 3; 8; 8; 12; 39; 121; 289; 528; 874; 1192];
%data6=[0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 2; 4; 6; 24; 96; 229; 442; 741];
%data7=[0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 1; 2; 2; 3; 5; 6; 8; 10; 10; 13; 27; 62; 139; 296; 438];
%data8=[1; 1; 1; 2; 2; 2; 2; 2; 2; 3; 6; 6; 7; 11; 15; 17; 23; 26; 27; 37; 44; 74; 251; 813; 2285; 4546; 7759; 11530];
%data9=[0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 1; 4; 9; 10; 14; 18; 24; 29; 34; 41; 59; 80; 123; 533; 2046; 4688; 10278; 17469; 24335];
%data10=[0; 1; 2; 3; 3; 3; 4; 5; 6; 6; 6; 7; 8; 10; 12; 17; 20; 22; 26; 29; 38; 47; 80; 198; 363; 612; 944; 1333];
data=[0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;6;7;9;9;11;13;18;25;76;285;860;2055;4132];
```

```
v=1000;
```

```
x = outer_function(v, data);
disp(x)
```

```
%lsqcurvefit
t=1:30;
ydata=data'
v=1000;
fun=@(x,t)Bassmodel(x,t,v);
lb=[0,0,0,0,0];
ub=[1, Inf, 1, 1, Inf];
x0=[20,1,1,1,300];
options = optimoptions(@lsqcurvefit,'Display','off');
x = lsqcurvefit(fun,x0,t,ydata,lb,ub,options);
y_val = Bassmodel(x,t,v);
```

```
% lsqnonlin
x0=x;
fun = @(x)ydata-(x(1)*x(2)*(1-exp(-(x(1)+x(3))*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3))*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3))*x(4)/(x(4)+(x(3)-x(4))*exp(-
x(5)*v))).*t));
options2 = optimoptions(@lsqnonlin,'Display','off');
x2 = lsqnonlin(fun,x0,lb,ub,options2);
y_val2 = Bassmodel(x2,t,v);
```

```
% plot results
plot(t,ydata,'+')
```

```
hold on
plot(t,y_val,'.')%Solve nonlinear curve-fitting (data-fitting) problems in least-squares sense
```

```
hold on
plot(t,y_val2,'--') %Solve nonlinear least-squares (nonlinear data-fitting) problems
legend('Havainnot','Bass-funktio: epälineaarinen käyräsovitus pienimmällä neliösummalla','Bass-funk-
tio: epälineaarinen pienimmän neliösumman datasoitus','Location','northwest')
hold off
```

```
%show results
disp(['p: ', num2str(x2(1))])
disp(['m: ', num2str(x2(2))])
disp(['q(m):', num2str(x2(3))])
disp(['q_0: ', num2str(x2(4))])
disp(['w: ', num2str(x2(5))])
```

```
%objective function for lsqcurvefit
function [F] = Bassmodel(x,t,v)
F=x(1)*x(2)*(1-exp(-(x(1)+x(3))*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3))*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3))*x(4)/(x(4)+(x(3)-
x(4))*exp(-x(5)*v))).*t));
end
```

```
function j = outer_function(v, data)
fun=@Bass;
nl=@nonlin;
lb=[20,0,0,0,0];
ub=[inf,1,1,1,inf];
x0=[ 0,0,0,0,0];
j=fmincon(fun,x0,[],[],[],[],lb,ub,nl);
```

```
function [F] = Bassmodel(x,t,v)
P = x(1);
m = x(2);
q_m = x(3);
q_0 = x(4);
w = x(5);
F = -(P.*m.*(exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-q_m))))-
1.0)./(P+(q_0.*q_m.*exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-q_m)))))./(q_0-
exp(-v.*w)).*(q_0-q_m)));
end
```

```

function res= Bass(x)
%objective function
F= (- data + x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)))/ (
1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v));
res=sum(F.^2);
end

function [c, ceq]= nonlinc(x)
%constraint
c=[];
ceq=- data + ( x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)))/
( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)));
end

end

```

Ohjelma 3. Laskenta p ja q -parametreille.

Taulukossa 33 on esitetty p , q ja m -parametrien laskennallisia tuloksia. Arvioitujen p ja q -parametrien tuloksien oikeellisuutta vertailtiin muun muassa Lee *et al.* (2018) toteuttamaan tutkimukseen. Vertailulla pyrittiin lähinnä arvioimaan ohjelmien ja menetelmien oikeellisuutta sekä tarkastamaan ovatko tulokset mahdollisia.

Taulukko 33. Laskentatulokset p ja q -parametreille sekä markkinoiden kokoa kuvaavalle parametrille m , vuosien 1998-2017 ajalta.

Aineisto	p	q	m	summa vuosien 1988-2017 ajalta, (lkm)
1	0.00017236	0.39996	5517.8619	56740
2	0.000137	0.30617	13376.9691	33159
3	0.0010871	0.26596	2413.9001	14075
4	0.00012552	0.30123	3446.7935	8757
5	1.6508e-06	0.33324	10714.8093	1192
6	0.0001083	0.28877	406.4523	741
7	8.3965e-09	0.53274	3570.852	438
8	0.0001155	0.28613	6791.4883	11530
9	0.00044165	0.40815	403.1723	24335
10	0.0001067	0.29811	798.7888	1333
11	0.00010321	0.40207	196.1225	4132

Innovaattoreiden ja seuraajien osuuksia voidaan arvioida p ja q -parametrien avulla eri lähdeaineistoissa. Markkinoiden kokoa kuvaava m -parametri ei ole merkitsevä, jos halutaan suhteuttaa eri aineistojen kasvukäyrät yhtä suuriksi kasvaviksi. Taulukon 33 sarakkeessa ”summa” esitetään kunkin aineiston havaintoyksiköiden lukumäärä. Lukumäärä auttaa valitsemaan aineiston, jossa on havaintoyksiköitä riittävästi eri vuosina. Varhaisessa vaiheessa lukumäärät voivat olla erityisen pieniä tai harvoina vuosina olleita yksittäisiä havaintoja.

Taulukossa 33 on esitetty laskentatulokset parametreille p ja q . Laskentaparametrien tulosten jälkeen voidaan valita aineistot laskentamallin mukaiseen ennustamiseen.



Kuva 51. Työvaiheen esittely ja sijoittuminen kvantitatiivisessa tutkimuksessa.

Yhdentoista eri aineiston osalta todettiin, että osalla tutkimukseen otetuista lähdeaineistoista oli taulukon 32 mukaan lähes yhtenevät p ja q -parametrit. Työvaiheessa pyrittiin valitsemaan aineistoja, joiden määrät ovat kehittyneet eriaikaisesti. Laskenta voitaisiin suorittaa kaikille aineistoille, mutta laajuuden rajaamiseksi tässä työssä on keskitytty kolmeen eri aineistoon. Yhdentoista eri aineiston lasketuista parametreista ja vertailusta valitaan mukaan poimintaotokset (aineistot) 1, 2 ja 11. Peruste valinnalla on vahvistaa Cho & Daim (2016) sekä Huang *et al.* (2018) tutkimuksissa osoitetut aineistojen kronologiset järjestykset. Tässä työssä osoitetaan myös kahden eri tutkimuksen aineistojen integrointi. *Integraatiolla pyritään vahvistamaan kaupallisten tietueiden ja patenttitietokannan välinen kronologinen viive.*

Taulukko 34. *Kvantitatiiviseen tutkimukseen valitut aineistot, joilla arvioidaan big data -ilmiön kehittymisen eri vaiheita.*

Aineisto	Aineisto	p-parametri	q-parametri	m
1	Proquest: hakutuloksessa "big data" -termiyhdistelmä. Dokumenttityyppi: Kaupalliset artikkelit. Aikajakso: 1988-2017	0.00017236	0.39996	5517
2	Proquest: hakutuloksessa "big data" -termiyhdistelmä. Dokumenttityyppi: Tieteelliset artikkelit. Aikajakso: 1988-2017	0.000137	0.30617	13376
11	Patenttitietokanta Espacenet (oikaistu aineisto).	0.00010321	0.40207	196

Taulukossa 34 on esitetty valittujen aineistojen laskennassa käytettävät p ja q -parametrit sekä lähdeaineistot, joista parametrit ovat laskettu. Aineistojen käsittelyssä havaittiin, että

havaintoja ennen vuotta 1988 ei juuri ollut (taulukko 18 ja kuva 39). Laskentaa voisi suorittaa eri vaihtoehdoilla ja vertailla saavutettuja tuloksia. Ilman aineiston tarkkaa tuntemista työ ei kuitenkaan ole mielekäästä. Virheet lähtöaineistossa antaisivat vääriä lopputuloksia. Työssä valittiin ajanjaksoksi kohdassa 3.2 esitetyillä perusteilla ajanjakso vuodesta 1998 vuoteen 2017 (sis. edellä mainitut vuodet). Kuvassa 52 esitetään työvaihetta ja sen riippuvuutta aikaisempiin työvaiheisiin.



Kuva 52. Työvaiheen esittely ja sijoittuminen kvantitatiivisessa tutkimuksessa.

Patenttitietokannan osalta valittiin ajanjakso, jolloin patenttien esiintyminen alkoi (2004). Aikajaksoja kokeiltiin, mutta päädyttiin lopputulokseen, että patenttien osalta laskenta aloitetaan vuodesta 2003 ja muiden aineistojen osalta vuodelta 1998. Valinnat perustuvat taulukossa 31 esitettyihin tietoihin sekä patenttitietokannan tarkempaan läpikäyntiin kohdissa 5.3, 5.4, 5.5 ja 5.6. Patenttitietokannan aineisto on oikaistu ja siitä on poistettu varhaisia havaintoja.



Kuva 53. Työvaiheen esittely ja laskennan sijoittuminen työvaiheena kvantitatiivisessa tutkimuksessa.

Menetelmävalintojen ja aineistojen laskentaan tarvittavien asioiden selvittämisen jälkeen voidaan suorittaa laskenta. Kuvassa 53 on esitetty tämä työvaihe sekä sen sijainti laskennassa.

Työvaiheessa hyödynnetään tässä työssä laajasti käsiteltyä Bassin mallin mukaista laskentaa. Laskenta perustuu kaavoissa (1-7) esitettyihin periaatteisiin ja ohjelma 1 mukaiseen algoritmin mukaelmaan. Laskentatulokset visualisoidaan kuvaajiksi (kuva 54).

```
clear variables; close all;

P=100; % saturaatiotaso
P1=5517
P2=13376
P3=196
A1(1)=0;
p1=0.00017236; %p-parametri otoksen yksi osalta
p2=0.000137; %p-parametri otoksen kaksi osalta
p3=0.00010321 %p-parametri otoksen kolme osalta
q1=0.39996; % q-parametri aineisto 1
q2=0.30617; % q-parametri aineisto 2
q3=0.40207; % q-parametri aineisto 2

alkuvuosi1(1)=1998; % lähtövuosi käyrälle 1
alkuvuosi2(1)=1998; %lähtövuosi käyrälle 2
alkuvuosi3(1)=2003; %lähtövuosi käyrälle 3
time_step=0.04
iter = 1000
%P=P1^2;
for i1=2:iter
    s(i1)=(p1 +q1*A1(i1-1)/P)*(P-A1(i1-1))*time_step;
    A1(i1)=s(i1)+A1(i1-1);
    alkuvuosi1(i1)=alkuvuosi1(i1-1)+time_step;
end

%käyrä kahden laskenta
A2(1)=0;
%P=P2^2
for i2=2:iter
    s2(i2)=(p2 +q2*A2(i2-1)/P)*(P-A2(i2-1))*time_step;
    A2(i2)=s2(i2)+A2(i2-1);
    alkuvuosi2(i2)=alkuvuosi2(i2-1)+time_step;
end

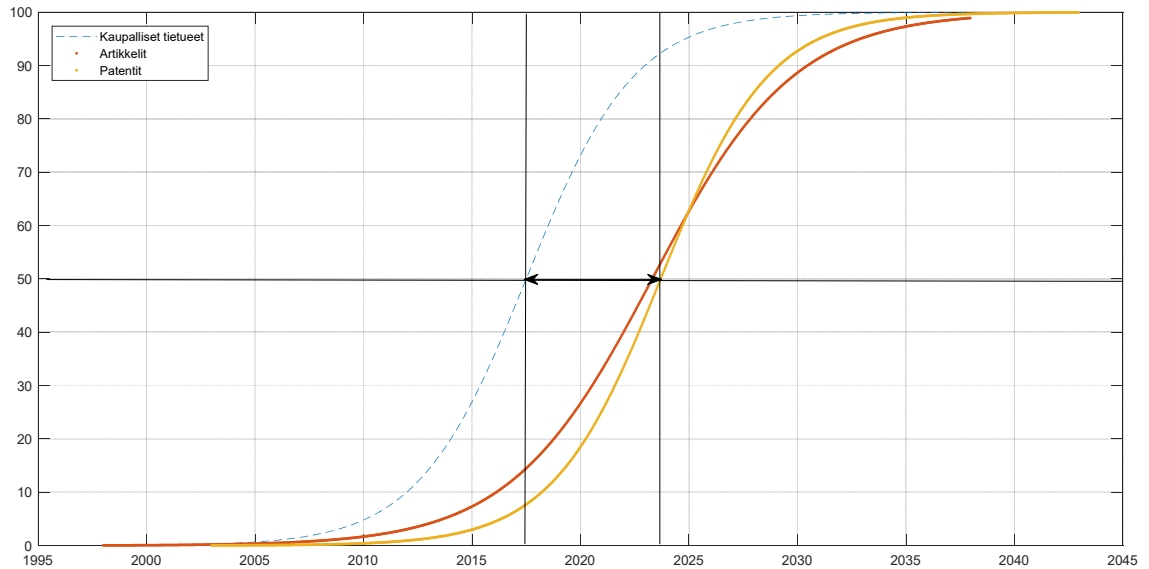
%käyrä kolmen laskenta
A3(1)=0;
%P=P3^2
for i3=2:iter
    s3(i3)=(p3 +q3*A3(i3-1)/P)*(P-A3(i3-1))*time_step;
    A3(i3)=s3(i3)+A3(i3-1);
    alkuvuosi3(i3)=alkuvuosi3(i3-1)+time_step;
end

figure(1)
plot(alkuvuosi1,A1,'--',alkuvuosi2,A2,'.', alkuvuosi3,A3,'.')%plot(x,y1,x,y2,'--
',x,y3,':')
hold on
legend('Kaupalliset tietueet','Artikkelit','Patentit','Location','north-
west','Orientation','vertical')
```

Ohjelma 4. Teknologiadiffuusiolaskenta Bassin mallin mukaan, jossa jokainen aineisto lasketaan erikseen.

Ohjelmassa (4) lasketaan teknologian kehittymistä kuvaavien ilmiöiden vaiheita Bassin laskentamallin avulla. Ohjelmassa (4) lasketaan erikseen kolmelle kasvukäyrälle ennuste.

Ohjelma (4) mahdollistaa tulosten visualisoinnin. Kuvassa 54 on esitetty laskentatulokset, jossa on simuloitu ”big data”-termiyhdistelmän ilmentymistä kolmessa eri lähdeaineistossa vuosien 1998-2045 välisenä ajanjaksona.



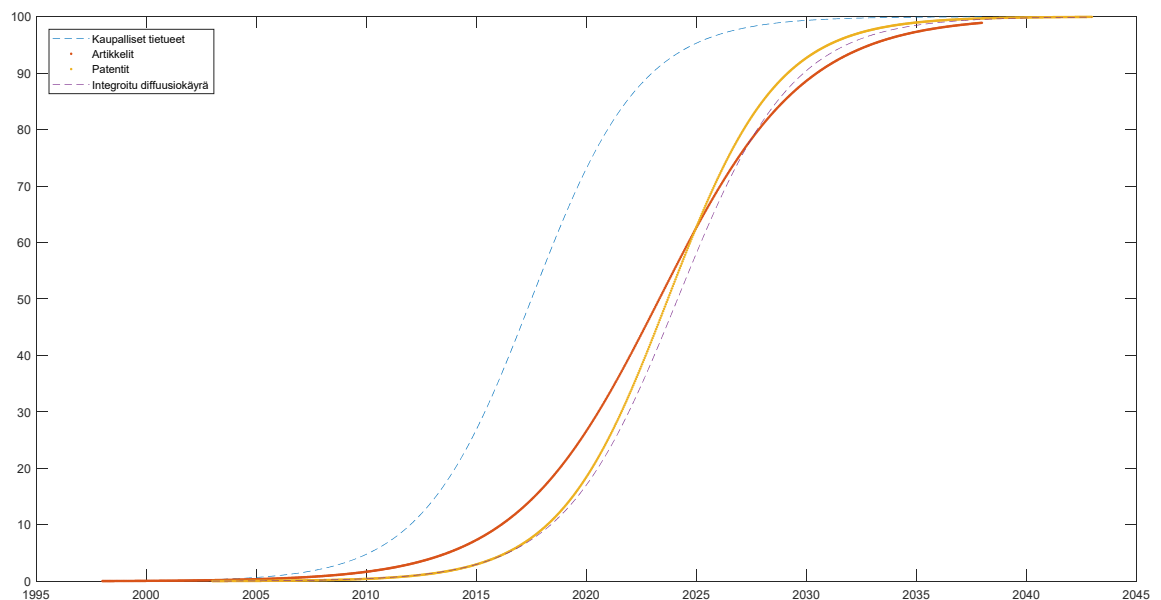
Kuva 54. Aineistojen 1 (kaupallisuus), 2 (tutkimukset) ja 11 (patentit) simulointi Matlab-ohjelmistolla. Viive kaupallisten tietueiden ja kaupallisesti suojattavien patenttien välillä on noin 6-7 vuotta.

Saavutettaessa useamman aineiston laskentatulokset, voidaan aineisto integroida useilla tavoilla. Työssä päädyttiin lopputulokseen, jossa integrointi simuloidaan huomioiden Bassin alkuperäisessä laskentakaavassa esitetyillä joukoilla eli innovaattoreilla ja seuraajilla. Bass (1969) jaotteli teknologiadiffuusiovaiheen adaptoijat innovaattoreihin ja seuraajiin. Esittämässään laskentakaavassa nämä vastaavat p ja q -parametreja. P ja q -parametrien integraatiota kokeiltiin muun muassa neliösummien painotuksilla. Lopputuloksena eri testeistä valittiin painotettujen lukujen summa. Perusteena valintaan oli se, että patenttitietokanta on esitetty muun muassa Lee *et al.* (2018) tutkimuksessa tarkaksi teknologian ennustuslähteeksi.

Painotettujen lukujen summassa painotettuja arvoja on helppo muuttaa esimerkiksi riippuvaksi havaintodatan määristä. Tässä tapauksessa havaintoja oli valituissa aineistoissa riittävästi (taulukko 32).

Taulukko 35. *Aineistojen (1,2 ja 11) integrointi painotetun summa avulla, jossa patenttitietokannan (11) datan käyttäytymiselle ja siitä lasketuille parametreille annetaan painoarvo 60 prosenttia.*

Aineisto	p-parametri	Painotus	p (painotettu)	q-parametri	Painotus	q (painotettu)
1	0,00017236	20 %	0,000034472	0,39996	20 %	0,079992
2	0,000137	20 %	0,0000274	0,30617	20 %	0,061234
11	0,00010321	60 %	0,000061926	0,40207	60 %	0,241242
	Summa: Painotettu p		0,000123798	Summa: Painotettu q		0,382468



Kuva 55. *Kuvassa esitetään kasvukäyrien integrointi, jossa keltaisen patentteja kuvaavan kasvukäyrän vieressä kulkee integroitu diffuusioennustekäyrä.*

Laskennassa on hyödynnetty samoja algoritmeja, kun kaikissa muissakin Bassin mallin ennusteita kuvaavissa malleissa. Varsinainen integraatio eri kasvukäyriä toteutettiin taulukon 35 laskennan avulla. Laskentatuloksia sovellettiin valmiisiin laskenta-algoritmeihin. Investointipäätökselle optimiajotus sijaitsi keskimäärin nousevan vaiheen ja kriittisen massapisteen välillä (riippuen yritystyyppistä A, B ja C). Työssä esitettiin aikaisemmin teoria kriittisestä massapistestä (mm. Sundqvist *et al.* 2002). Kriittisten massapisteiden laskenta voidaan suorittaa p ja q -parametrien avulla. Kriittiset massapisteen laskentaan eri lähteaineistoille. Kriittisten massapisteiden sijainti on merkityksellinen.

Integrointi voitaisiin toteuttaa myös aikaisemmin eli lähtödatassa. Integrointi lähtödatassa toteutetaan vastaavasti suhteuttamalla eri aineistojen funktiot toisiinsa painokertoimilla.

Lähtödatan vastaavaa integrointia ei kokeiltu. Integroitavilla aineistoilla ei havaittu rajoitteita. *Rinnakkainen, vahvistava tai syrjäytyvä teknologia voidaan integroida laskentaan kohtuullisen pienellä työllä esimerkiksi taulukossa 35 esitetyllä menetelmällä.*

Teknologiavalintojen päätöksessä varhaisien vaiheiden kehittyminen on merkityksellistä, mutta kriittinen massapiste (CMP) teknologiadiffuusiassa on ratkaiseva.



Kuva 56. Työvaiheen esittely ja kriittisen massapisteiden laskennan sijoittuminen kvantitatiivisessa tutkimuksessa.

Ohjelmassa (4) lasketaan kriittiset massapisteet aineistoille 1, 2 ja 11.

```

clear variables; close all;
% Sundqvist et al. esittämä tapa: (log((2+ sqrt(3)) * p3/q3)/(- (p3+q3)))
% toinen derivaatta: 1/(p+q) * log(q/p)
CMP1=(log((2+ sqrt(3)) * p1/q1)/(- (p1+q1))) + alkuvuosi1(1)
CMP2=(log((2+ sqrt(3)) * p2/q2)/(- (p2+q2))) + alkuvuosi2(1)
CMP3=(log((2+ sqrt(3)) * p3/q3)/(- (p3+q3))) + alkuvuosi3(1)
%CMPT1 = 1/(p+q) * log(q/p)
%CMPT2 = 1/(p+q) * log(q/p)
%CMPT3 = 1/(p+q) * log(q/p)
  
```

Ohjelma 5. Matlab-koodi kriittisten massapisteiden laskentaan, jossa huomioidaan kaksi tapaa laskea kriittinen massapiste.

Tulokset kriittisten massapisteiden laskennasta

CMP1 (Kaupalliset artikkelit): kriittinen massapiste saavutettiin 2014.

CMP2 (perustutkimus): Kriittinen massapiste saavutetaan 2019.

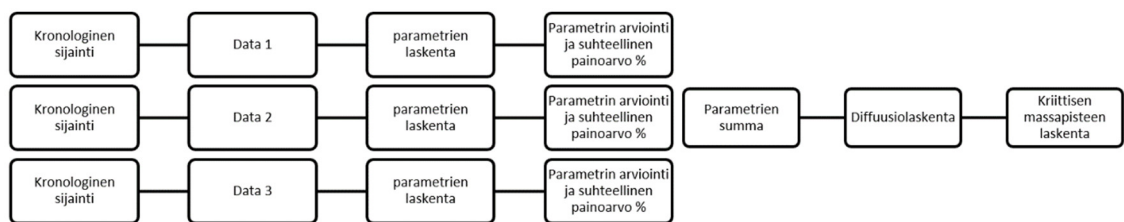
CMP3 (patentit): Kriittinen massapiste saavutetaan 2020.

Kriittinen massapiste integroidulle kasvukäyrälle saavutetaan vuonna 2020 (ei huomioida vahvistavaa verkostovaikutusta). Koska lähtötiedot ovat vuositasoisia, niin tarkempi ennuste vaatisi kuukausikohtaisen datan käsittelyä. Kuukausikohtaisen datan käsittely ei lisää merkittävästi työmäärää, koska dataa käsitellään ohjelmallisesti.

Kriittisen massapisteen saavuttaminen on olennaista, koska saavuttaminen lisää big data -tutkimuksen osalta todennäköisyyttä tutkimuksen voimakkaalle kasvamiselle. Mikäli tutkimuksen osalta kriittistä massapistettä ei saavuteta, niin big data voi ilmiönä heikentyä tutkimuksen kohteena. Vertaamalla Huang *et al.* (2018) tutkimukseen; tällöin rahoitus, keksinnöt, innovaatiot, patenttien määrä sekä kaupallisten tuotteiden jatkokehitys

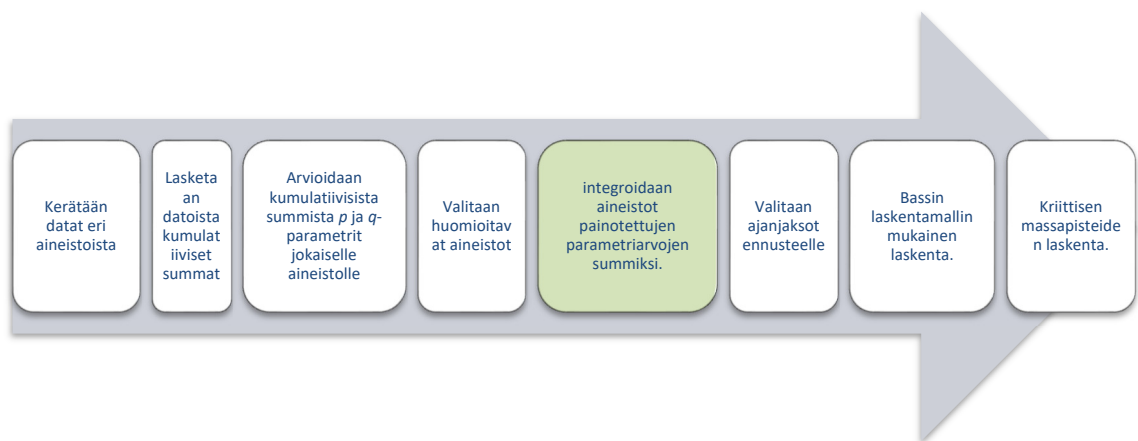
voi olla vähäisempää. Investointipäätöksenä big data on optimiajankohdassa 2018/2019. Huomioiden nopeasti hitaasti ja nopeasti kehittyvät teknologiat, voi big data kuulua ratkaisuna osittain nopeasti kehittyviin teknologioihin, jos ratkaisussa hyödynnetään valmiista infrastruktuuria (Linden ja Fenn 2003). Tällöin big data -ratkaisun perustuessa pilviarkkitehtuuriin saavutetaan nopeampi diffuusion kehittyminen. Tutkimus tulisi toistaa ennen investointipäätöksiä. Artikkeleiden ja patenttien määrä voidaan laskea kuukausitasolla. Tällöin laskenta voidaan suorittaa lähes reaaliaikaisemmin uusimmista aineistoista – huomioiden kuitenkin eri aineistojen ominaisuudet.

Esitetyillä menetelmillä voidaan integroida erilaisia aineistoja, kuten huomioida kilpailevan, syrjäytyvän, tai osa- ja rinnakkaisteknologian integroitu laskentamalli. Laskenta voidaan toteuttaa tuloksissa esitetyllä tavalla.



Kuva 57. Laskennan kuvaus, jossa toteutetaan laskennan integrointi.

Kuvassa 57 toteutettu laskennan integrointi voidaan toteuttaa muillakin tavoilla. Tässä tapauksessa integraatio on yksinkertainen ja lukujen painoarvoilla voidaan vaikuttaa integraatioon helposti. Myös negatiivisia painoarvoja voidaan antaa, jolloin esimerkiksi kilpailevat teknologiat voidaan huomioida helposti.



Kuva 58. Integroidun laskentamallin esittely.

Laskentatavan vahvuus, on skaalautuvuus ja yksinkertaisuus. Integraatio toteutetaan laskennan vaiheessa (kuva 58), jossa laskenta on yksinkertaisimmillaan. Ohjelmassa (6) lasketaan integroiduista p ja q -parametreista integroitu laskenta.

```
clear variables;
close all;
```

```

data1=[3; 5; 6; 11; 22; 30; 38; 56; 86; 121;168;213;259;297;329;352;388;424;
462;500;543;578;715;1876;6608;15870;26149;36074;46574;56740];

P=100; % saturaatiotaso
P1=5517
P2=13376
P3=196
P4=196

A1(1)=0;
p1=0.00017236; %p-parametri otoksen yksi osalta
p2=0.000137; %p-parametri otoksen kaksi osalta
p3=0.00010321 %p-parametri otoksen kolme osalta
p4=0.000123798 %p-parametri integraation osalta
q1=0.39996; % q-parametri aineisto 1
q2=0.30617; % q-parametri aineisto 2
q3=0.40207; % q-parametri aineisto 3
q4=0.382468; % q-parametri aineisto 3

alkuvuosi1(1)=1998; % lähtövuosi käyrälle 1
alkuvuosi2(1)=1998; %lähtövuosi käyrälle 2
alkuvuosi3(1)=2003; %lähtövuosi käyrälle 3
alkuvuosi4(1)=2003; %lähtövuosi integroidulle käyrälle
time_step=0.04
iter = 1000
%P=P1^2;
%käyrä yhden laskenta
for i1=2:iter
    s(i1)=(p1 +q1*A1(i1-1)/P)*(P-A1(i1-1))*time_step;
    A1(i1)=s(i1)+A1(i1-1);
    alkuvuosi1(i1)=alkuvuosi1(i1-1)+time_step;
end

%käyrä kahden laskenta
A2(1)=0;
%P=P2^2
for i2=2:iter
    s2(i2)=(p2 +q2*A2(i2-1)/P)*(P-A2(i2-1))*time_step;
    A2(i2)=s2(i2)+A2(i2-1);
    alkuvuosi2(i2)=alkuvuosi2(i2-1)+time_step;
end

%käyrä kolmen laskenta
A3(1)=0;
%P=P3^2
for i3=2:iter
    s3(i3)=(p3 +q3*A3(i3-1)/P)*(P-A3(i3-1))*time_step;
    A3(i3)=s3(i3)+A3(i3-1);
    alkuvuosi3(i3)=alkuvuosi3(i3-1)+time_step;
end

%käyrä neljän laskenta
A4(1)=0;
%P=P3^2
for i4=2:iter
    s4(i4)=(p4 +q4*A4(i4-1)/P)*(P-A4(i4-1))*time_step;
    A4(i4)=s4(i4)+A4(i4-1);
    alkuvuosi4(i4)=alkuvuosi4(i4-1)+time_step;
end

%lasketaan kaikille käyrille kriittinen massapiste X-akselilla
CMP1=(log((2+ sqrt(3)) * p1/q1)/(-(p1+q1)))+alkuvuosi1(1)%kaupalliset artikkelit
CMP2=(log((2+ sqrt(3)) * p2/q2)/(-(p2+q2)))+alkuvuosi2(1)%tutkimukset
CMP3=(log((2+ sqrt(3)) * p3/q3)/(-(p3+q3)))+alkuvuosi3(1)%patentit
CMP4=(log((2+ sqrt(3)) * p4/q4)/(-(p4+q4)))+alkuvuosi4(1)%integroitu laskenta CMP:lle

figure(1)
plot(alkuvuosi1,A1,'--',alkuvuosi2,A2,'.', alkuvuosi3,A3,'.',alkuvuosi4,A4,'--' )

hold on
legend('Kaupalliset tietueet','Artikkelit','Patentit','Integroitu diffuusiokäyrä','Location','northwest','Orientation','vertical')

```

Ohjelma 6. Integroidun laskentamallin esittely, jossa toteutetaan Bassin mallin mukainen laskenta.

5.3 Arvio tulosten luotettavuudelle otoksien perusteella patenttitietokannasta

Teknologiadiffuusio ei noudata mitään kasvumallia, mutta osa malleista voi olla hyviä kuvaamaan teknologian kehittymistä. Tulosten luotettavuutta voidaan arvioida tilastollisin menetelmin.

Tulosten luotettavuutta voidaan arvioida muun muassa otannan määrän, aineiston todettujen ominaisuuksien ja laskentatavan perusteella. Luotettavuuden arviointi rajautuu tällöin ainoastaan tiedossa oleviin ja valittuihin asioihin. Todellisuudessa tulosten vaihteluväli voi olla ilmaistua suurempi. Tekstihaun otantaan poimittiin useita patenteja, joilla ei ollut yhteyttä big data -ilmiöön. Työssä patentit luokiteltiin vain kahteen tilaan eli patentti kuuluu big data -ilmiöön tai ei kuulu big data -ilmiöön. Kun jako tehdään päätellyn rajan perusteella kahteen eri tilaan, niin koko otokselle voidaan laskea binomijakauman avulla virheellisten esiintymien todennäköisyysfunktio. Otoksen mukaan virheitä esiintyi (0,0345) eli 3,45% tapauksissa. Binomijakauman mukaan, jossa ei huomioida populaatiosta poistettuja tapauksia (129 virhettä 3735 kappaleesta) saadaan todennäköisyys, että poiminnassa saadaan virhe $P(\text{virhe})$ karkeasti välillä 0,02-0,05 (eli 2-5%) välillä huomioiden kuitenkin, että virheiden vaihteluväli on vain ideaalitapauksessa näin pieni (kuva 59 binomitodennäköisyysjakauma otoksesta $n=3735$, jossa saatiin 129 virhettä).

Big dataa määriteltiin Big Data 3v's vaatimuksen täyttämiseksi. Populaatioista poimitujen havaintoyksiköiden otsikoiden, avainsanojen tai tiivistelmien perusteella ei voinut päätellä aukotonta yhteyttä big dataan. Hypessä on tyypillistä korostaa ominaisuuksia, vaikka ominaisuudet eivät täysin täytyisikään. Huolimatta big datan määritelmästä Big Data 3v's vaatimuksesta, jokainen tutkimus, joka sivuaa suuren rakenteettoman tietomasinan käyttöä, toimii kuitenkin big datan osateknologiana. Osateknologiat, jotka kuuluvat otokseen, jos se on lähteessä mainittu big dataan kuuluvaksi esimerkiksi avainsanojen tai sisällön perusteella.

Riippumattomuus:

Laskenta perustuu hypoteesiin, että patenttien hakutulokset ovat toisistaan riippumattomia.

Todennäköisyys ja toistuvuus:

Riippumatta patentin tulkinnasta joukkoon kuuluvaksi todennäköisyys säilyy vakiona tulkittaessa seuraavan patentin kuuluvuutta 'big data' -ilmiöön.

Aineistoon kuuluvuuden ja kuulumattomuuden poissulkevuus:

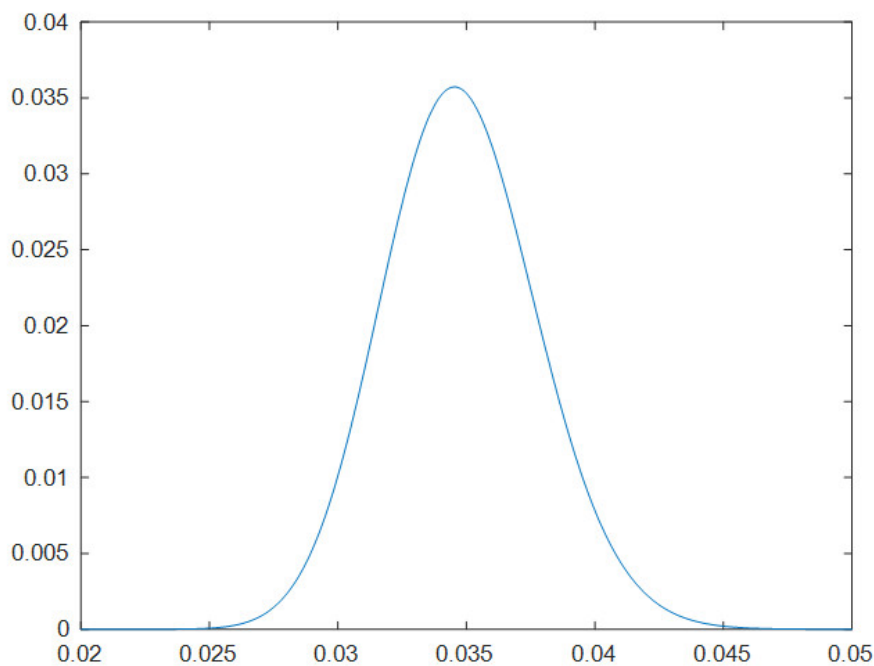
Aineistoon kuulumisen osalta on tehty rajaus, että se joko kuuluu tai ei kuulu tutkittavaan teknologiaan. Tällöin tulkinnassa ei voi olla yhtäaikaisesti voimassa molemmat tilat, niin tapahtumat riippuvat toisistaan ja niiden tulos on toisensa poissulkeva eli todennäköisyys kuuluvuudelle

$$P(\text{kuuluu tutkimusjoukkoon}) = 1 - P(\text{ei kuulu tutkimusjoukkoon})$$

$$P(\text{ei kuulu tutkimusjoukkoon}) = 1 - P(\text{kuuluu tutkimusjoukkoon})$$

eli patenti saa teoriassa tulkinnan jälkeen vain tilat

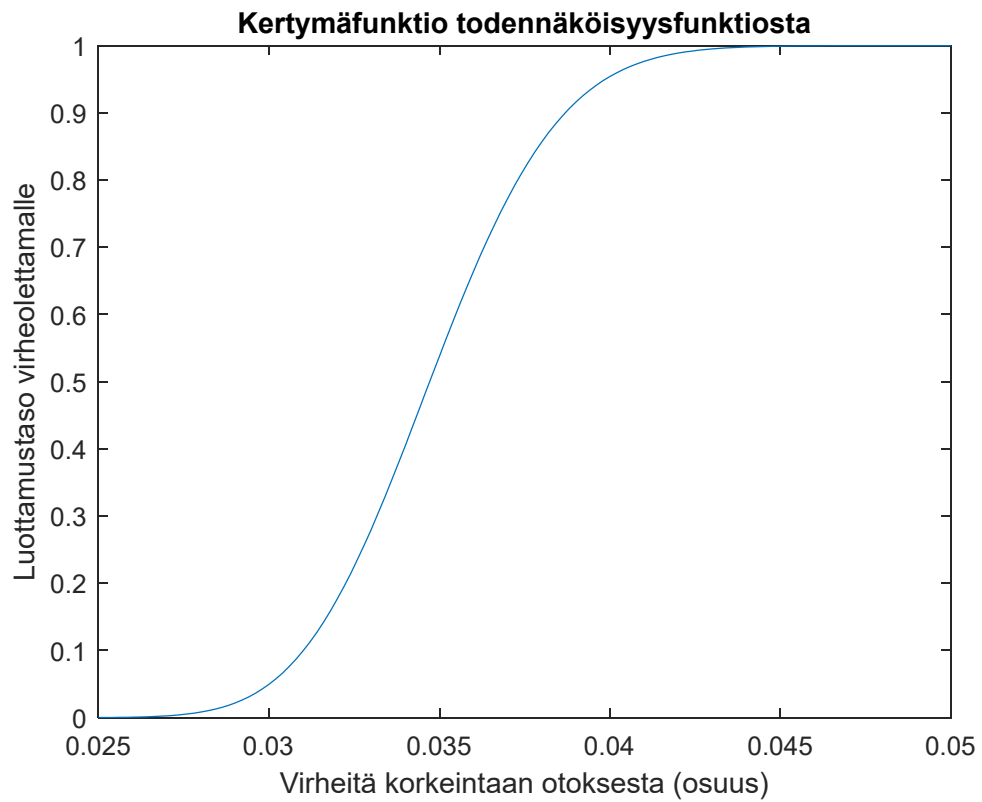
$$S = \{ \text{'kuuluu tutkimukseen'}, \text{'ei kuulu tutkimukseen'} \}.$$



Kuva 59. Patenttitietokannan otoksen poimintavirheiden esiintyvyyden todennäköisyysjakauma, jonka huippuarvo saavutetaan P (virhe = 0,0345) kohdalla, jolloin funktion derivaatta saavuttaa arvon nolla. (Liite G)

Integroimalla virheiden esiintyvyyden todennäköisyysjakauma $\int F(P(\text{virhe}))$ saadaan esitettyä virhetodennäköisyyden kertymäfunktio. Kertymäfunktion avulla voidaan esittää virhetodennäköisyyden luottamustason muuttumista virhetodennäköisyyden oletuksen funktiona. Kuvan 59 ja 60 laskettujen kuvaajien perusteella voidaan kuitenkin todeta, että hyväksymällä lähtöaineiston muut virheet ja 0,05 (5%) virhetaso tulosta voidaan pitää

teknisesti yli 99% luotettavana. Virheiden määrä otoksen kokoisessa populaatiossa on binomijakauman mukaan alle 155 kappaletta otoksesta yli 99% luottamustasolla.



Kuva 60. Kuvassa on esitetty patenttitietokannan otoksen poimintavirheiden esiintyvyyden todennäköisyyden kertymäfunktio. Laskenta on esitetty liitteessä (Liite H).

Kertymäfunktio osoittaa virheiden todennäköisyyden maksiarvon otoksesta. Virheiden todennäköisyyden maksimiarvo lähestyy yhtä jo 0,045 osuuden kohdalla.

5.4 Laskenta hypergeometrisella jakaumalla

Binomijakauman mukainen laskenta *ei huomioi alkuperäisestä populaatiosta pois otettuja tapauksia*, jolloin matemaattisesti tarkempi tulos saadaan hyödyntämällä hypergeometrisella jakaumalla. Laskennassa ei myöskään tiedetä patenttietokannassa olevien patenttien määrää ja arvio tuleekin tehdä kaikille big dataa edustavien patenttien määrälle otoskoosta ($n=3735$) ylöspäin olettaen, että big dataa koskevia patenteja ei ole ääretöntä määrää. Hypergeometrisen jakauman kaava (9)

$$h(x; n, N, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} \quad (9)$$

, jossa

$h(x; n, N, k)$ = hypergeometrisen jakauman esitys, jossa parametrit

k = Big dataa kuvaavat patentit

N = Koko populaatio

n = Poimintaan tulleet

$n-k$ = Poimitaan virheellisesti tulleet patentit

Hypergeometrisen jakauman odotusarvo voidaan laskea kaavalla (10) ja varianssi kaavalla (11)

$$\mu = \frac{nk}{N} \quad (10)$$

, jossa

μ = Odotusarvo

k = Big dataa kuvaavat patentit

N = Koko populaatio

n = Poimintaan tulleet

$$\sigma^2 = \frac{N-n}{N-1} * n * \frac{k}{N} \left(1 - \frac{k}{N}\right) \quad (11)$$

, jossa

σ^2 = varianssi

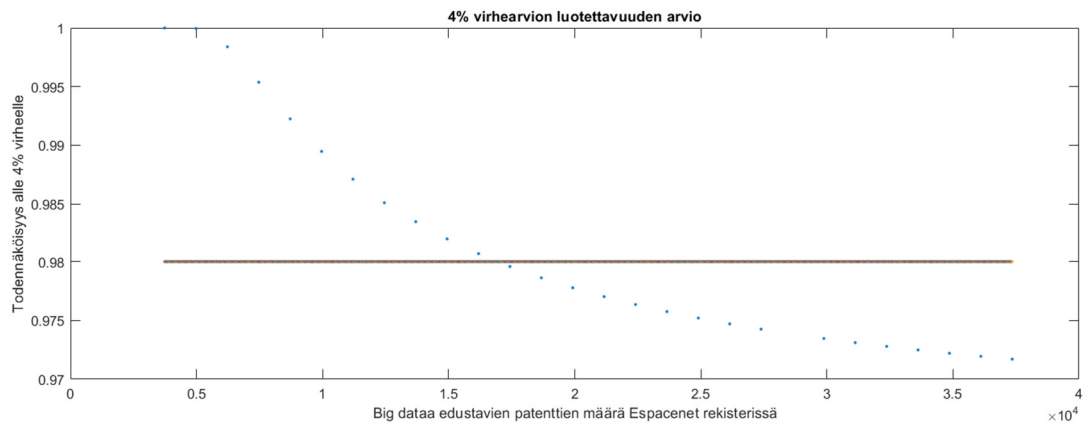
N = Koko populaatio

n = Poimintaan tulleet

k = big dataa kuvaavat patentit

Hypergeometrisen jakaumalla voidaan arvioida koko populaation (big data -patenttien) lukumäärän vaikutus virheellisten patenttien määrään. Tämä laskenta on tehty hypoteesilla, että patenttietokannasta toteutettu poiminta poimii 10-100% patenteista, joissa on

big data edustettuna teknologiana. Kuvassa 61 esitetään koko patenttitietokannan big dataa edustava populaatio x-akselilla, joka saa arvoja $N=[3735...37350]$ välillä. Tuloksen, jossa esitetään 4% määrä virheellisiä tuloksia on asetettu 0,98 luottamustaso.



Kuva 61. Koko populaation (N) funktiona, jossa esitetään populaation koon vaikutus tulosten luotettavuuteen kyseisessä otoksessa, jossa luottamustaso on asetettu 0,98:ään ja testataan 4% virheellisten määrän laskenta (liite I).

Koska poimintajoukko on ollut suuri ($n=3735$), niin laskentatulosten mukaan Espacenet-tietokannassa olleiden ja poimintaan tulemattomien patenttien määrä ei juuri vaikuta virhearvioon. Populaation koon osalta 4% virhearvio saavuttaa tietokannassa olleiden patenttien määrän osalta 98% luotettavuuden, mikäli big data -patenttien määrä on ollut alle 15 000 Espacenet-tietokannassa.

Koodistojen, merkkien ja muiden aineistojen laskennassa hyödynnettiin Matlab R2017b sekä Excel 2016 -ohjelmistoja. Aineiston alkuvalmistelu tehtiin Excel-ohjelmalla, jolla toteutettiin rivimääräinen laskenta ja aineiston muokkaus seuraavasti:

Taulukko 36. Excel-ohjelmistolla muodostettiin otoksesta lähtöaineisto kentätasoisesti esitettynä. Lähtöaineiston muodostamissäännöt on esitetty liitteessä (liite J).

Aineistokentän kuvaus	Muodostamistapa
Numero	Järjestysnumero käsittelyä varten
Vuosi	Muodostettu tieto
Kk	Muodostettu tieto
Paiva	Muodostettu tieto
Vuosi-kk	Muodostettu tieto
Otsikkotieto (Nimitys)	Alkuperäinen tieto
Onko big data -patentti	Muodostettu tulkinnalla
Julkaisunumero	Alkuperäinen tieto
Julkaisupäivä	Alkuperäinen tieto
Kansainvälinen luokitus (IPC)	Alkuperäinen tieto
IPC_lkm	Muodostettu laskennalla
CPC-luokitus	Alkuperäinen tieto
CPC_lkm	Muodostettu laskennalla
Hakemusnumero	Alkuperäinen tieto

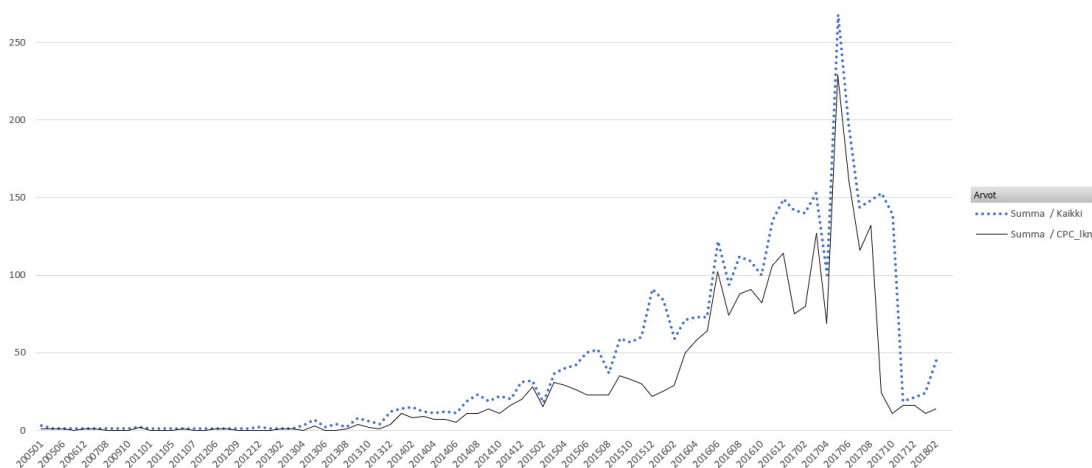
Hakemuksen tekemispäivä	Alkuperäinen tieto
Hakemusvuosi	Muodostettu laskennalla
Hakemuskukausi	Muodostettu laskennalla
Hakemuspäivä	Alkuperäinen tieto
Hakemuspvm	Muodostettu laskennalla
Hakemusviive	Muodostettu laskennalla
Etuoikeusnumero(t)	Alkuperäinen tieto
Tutkimusraportissa siteeratut patentit	Alkuperäinen tieto
Tutkimusraportissa siteerattu kirjallisuus	Alkuperäinen tieto
Patentoitavuustutkimuksessa siteeratut patentit	Alkuperäinen tieto
Patentoitavuustutkimuksessa siteerattu kirjallisuus	Alkuperäinen tieto
Muut patenttiviitteet	Alkuperäinen tieto
Muut kirjallisuusviitteet	Alkuperäinen tieto
Väitteessä siteeratut patentit	Alkuperäinen tieto
Väitteessä siteerattu kirjallisuus	Alkuperäinen tieto
Hakijan siteeraamat patentit	Alkuperäinen tieto
Hakijan siteeraama kirjallisuus	Alkuperäinen tieto
PCT-tutkimuksen viitejulkaisut	Alkuperäinen tieto
PCT-tutkimuksen NPL-viitejulkaisut	Alkuperäinen tieto
PCT-lisätutkimuksen viitejulkaisut	Alkuperäinen tieto
PCT-lisätutkimuksen NPL-viitejulkaisut	Alkuperäinen tieto

Koodistojen läpikäynnin (n=3735) perusteella saatiin (otsikkoon kohdistuvalle ”Big data” -sanayhdistelmälle) koodistohaun tarkkuudelle taulukon 37 ja kuva 62 tuloksia. Patenttikooditus kertoi lähes poikkeuksetta teknologian ylätasoa, mutta tulosjoukkoon poimituissa patenteissa ei usein ollut tarkkaa patentin tarkoitusta. Tarkempi määrittely olisi näkynyt pidempänä IPC-koodina, jonka viimeiset numerot kertovat yksityiskohtaisemmin mitä patentti koskee. (European patent office, <http://www.cooperativepatentclassification.org>, viitattu 26.4.2018)

Taulukko 37. Patenttiluokitukset otoksessa

IPC koodi	Lukumäärä	Euroopan patenttitoimiston määrittelyjen mukainen käyttötarkoitus
G06F17/30	1404	Electrical digital data processing(G06F), data retrieval (17/30)
H04L29/08	368	Transmission of digital information (H04L), Arrangements, apparatus, circuits or systems (29/08)
G06Q10/06	238	[Organising, forecasting, or planning] machines (G06Q10/06)
G06Q30/02	197	Marketing aspects of discounts or coupons
H04L29/06	187	Transmission of digital information, (H04L), Data transfer characterised by protocol
G06Q50/06	163	Electricity, gas or water supply (Monitoring the distribution of energy or determining charges reflecting a consumed quantity of power or energy, e.g. Electricity, gas, therm or water or Determining charges for a dispensed volume or a delivered quantity of a flowing substance, e.g. liquid or gas)
G06F19/00	151	Digital computing or data processing equipment or methods, specially adapted for specific applications
G06K9/00	101	Recognition of data; presentation of data; record carriers; handling record carriers (G06K), Methods or arrangements for reading or recognising printed or written characters or for recognising patterns
G06Q10/04	100	Forecasting or optimisation, e.g. linear programming, "travelling salesman problem" or "cutting stock problem" (data collection)
G06Q50/26	90	Government or public services
G06K9/62	77	Recognition of data; presentation of data; record carriers; handling record carriers (G06K)
H04L12/24	77	Data switching networks (H04L 12/00)

CPC-luokitus löytyi otoksessa 65,3%:sta (kuva 62), joten poimittu aineisto oli siltä osin heikkolaatuista. Kuvassa 62 esitetään CPC- koodituksen esiintyminen patenteissa osuuk-
sina kuukauden aikana. IPC-koodi löytyi lähes kaikista (yli 99,9%) patenteista.



Kuva 62. CPC-koodituksen osuus patenteissa (musta viiva) ja kaikki patentit (sininen katkoviiva).

5.5 Patenttietokannan aineistolliset ominaisuudet ja niiden haasteet

Tekstipoiminnalla toteutetun otoksen virheellisten poimintojen lukumäärä painottui ensimmäisille vuosille patenttietokantaan kohdistuvissa hauissa. Tekstihakuun tulneiden virheellisten tulosten määrä oli 100% ensimmäisillä kuukausilla ja ensimmäisen 5 vuoden aikana niitä oli noin 90% tuloksista. Vaikka virheiden määrä poiminnassa kasvoi, niin niiden suhteellinen osuus pieneni olennaisesti. Vastaavasti teknologian yleistyessä voimakkaasti virheellisten tulkintojen suhteellinen osuus pieneni merkittävästi. Tämä edustaa yhtä hakutulostosta yhdestä valitusta teknologiasta, joten tämän perusteella ei tule tehdä pitkälle meneviä johtopäätöksiä.

Virheiden määrän arvio eri ajankohdille (kuva 63) muuttui seuraavasti:

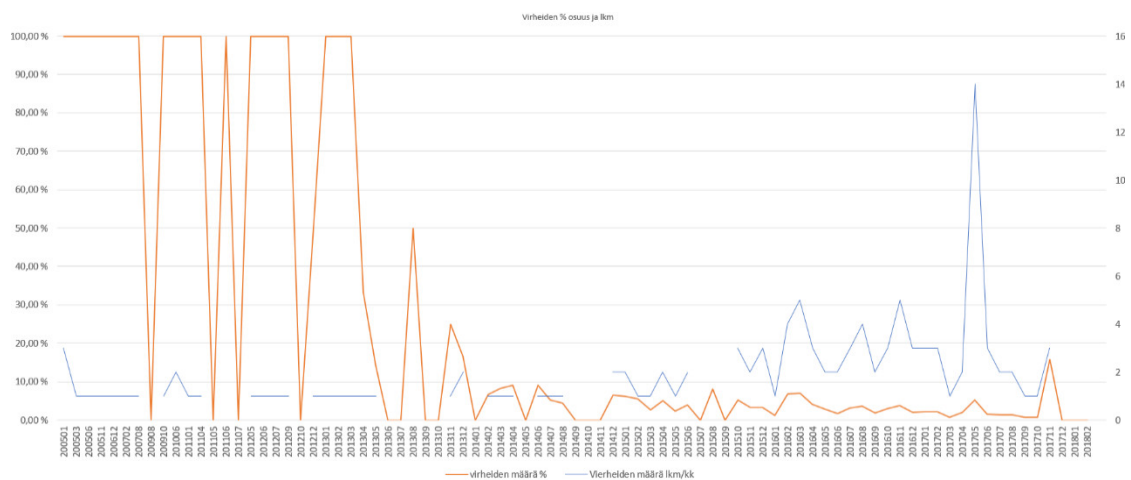
- 1.1.2005 - 31.12.2009, n=11, err= 10, (90% virheellisiä)
- 1.1.2010 - 31.12.2014, n=274, err = 27, (10% virheellisiä)
- 1.1.2015 - eteen päin, n=3450, err=92, (3% virheellisiä),

joissa

err on n-k eli epävalidit havaintoyksiköt

n on otoskoko

k on validit havaintoyksiköt.



Kuva 63. Virheellisten hakutulosten prosentuaalinen osuus sekä absoluuttiset määrät. Ensimmäisinä viitenä vuotena otokseen tuli vähän ilmentymiä. Otoksen ensimmäiset vuodet edustavat heikkolaatuisia havaintoja ja pääosin virheellisiksi tulkittuja tuloksia (yli 90% virheellisiä).

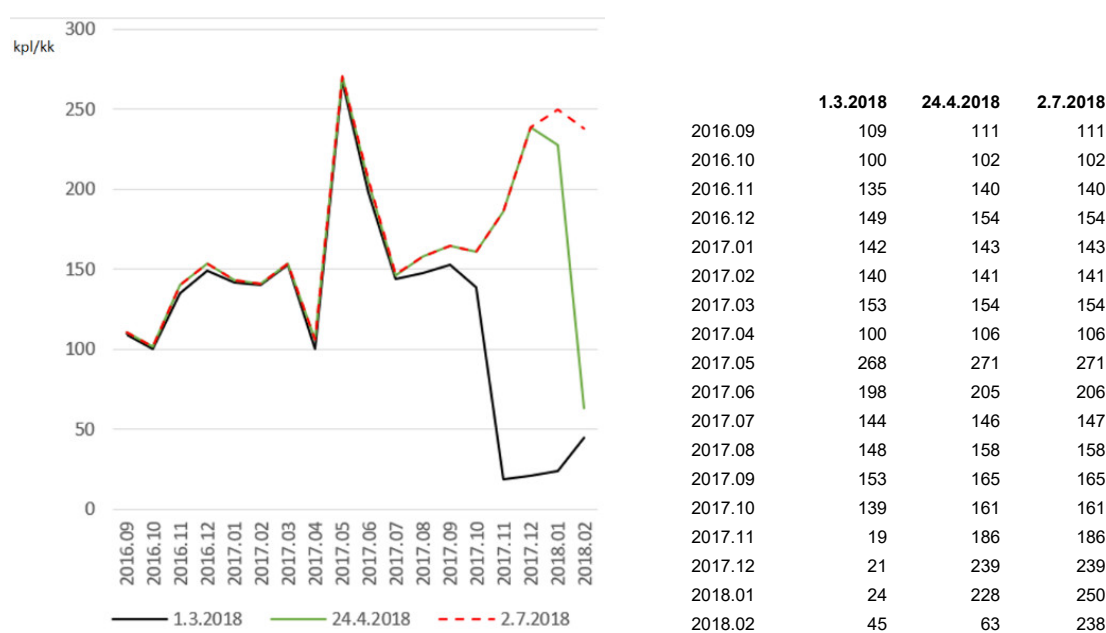
Patenttien tietokantaan tallentamisen viiveestä aiheutuu virheitä, koska patenteja viedään tietokantaan eri aikaisesti. Espacenet on globaali patenttitietokanta. Patenttien tietoja ylläpidetään eri toimijoiden toimesta ja patentin hyväksymisen jälkeen rekisteröintiviive vaihtelee. Viiveen mittaamisessa hyödynnettiin eri tapoja. Viive todennettiin muun muassa hyödyntämällä kolmea eri toteutettua poimintaa, 09/2016-02/2018 väliselle ajalle seuraavasti:

- 1.3.2018 toteutettu otanta ilman aikarajausta
- 24.4.2018 toteutettu otanta 09/2016-02/2018 väliselle ajalle (54vrk ensimmäisestä poiminnasta)
- 2.7.2018 toteutettu otanta 09/2016-02/2018 väliselle ajalle (123 vrk ensimmäisestä poiminnasta).

Ensimmäisen ja toisen poimintahetken välinen aika oli 54 vrk sekä ensimmäisen ja viimeisen poimintahetken välinen aika oli 123 vrk. Tulokset osoittavat patenttitietokantaan

vietyjen tietojen 'viiveen suuruudet' ajan funktiona (kuva 64). Toistamalla vastaavia hakuja patenttitietokantaan voisi viiveiden funktiota sekä sen kehittymistä tutkia lisää omana tutkimuksenaan. Näiden otantojen perusteella ero tuloksissa on erittäin merkittävä ja johtaa virheellisiin päätelmiin alle 6-8kk poimintahetken tuloksien käyttämisessä. Ensimmäisen ja toisen poiminnan välinen ero oli suurimmillaan joulukuussa 2017 eli 2-3kk ensimmäisestä poimintahetkestä (kuva 64). Poiminta-ajankohdasta johtuva ero pieneni merkityksettömäksi patenteista, jotka oli myönnetty 7-8 kuukautta ennen ensimmäistä poimintaa (1.3.2018). Poimintojen ero säilyi nollan ja yhden välillä 09/2016-07/2017 välisellä ajalla.

Ilmiö toistettiin kolmannella poiminnalla, jossa poimintatulosta verrattiin 54 vrk ja 123 vrk aikaisemmin toteutettuun poimintaan. Myös toisen poiminnan virhe oli suurimmillaan 2-3 kk poiminnasta (02/2018) ja pieneni merkityksettömäksi patenttien lukumäärässä, joka oli myönnetty 7-8kk ennen toista poimintaa. Päätelminä voidaan todeta patenttien määrän nousseen 18kk:ssa noin 150%. Yhden poiminnan ja virheellisen tulkinnan perusteella olisi päädytty noin 66% laskuun patenttien määrässä eli *oikaistu tulos on huomattavan erilainen ensimmäiseen patenttitietokantaan kohdistuneeseen poimintaan verrattuna*. Hypoteesi tuoreen aineiston tarkkuudesta ei toteutunut.



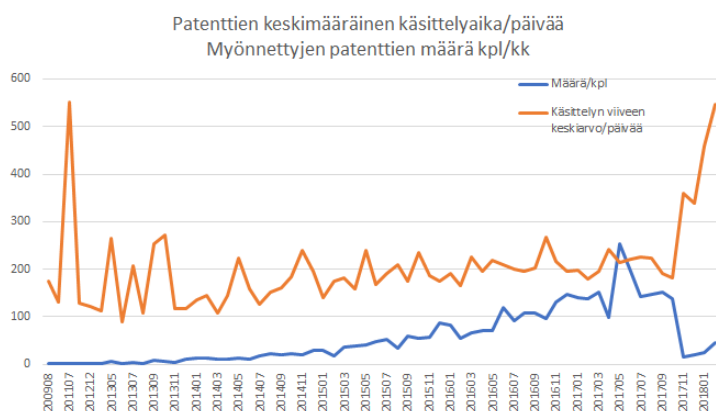
Kuva 64. Tässä työssä tutkitut julkaistujen patenttien tietokantaan viemisen viive, jossa on esitetty kolmen eri otoksen poimintatulokset ajan funktiona samasta populaatiosta kolmena eri aikana.

Patenttitietokantaan kohdistettujen poimintatulosten virheiden löytäminen on teknologiaennustamisessa tärkeää. Suoraviivainen poiminnan tulkinta johtaa väärään päättelyyn.

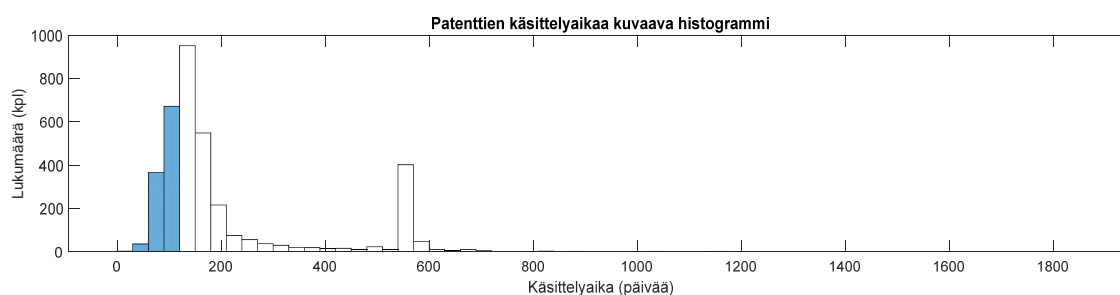
Patenttien käsittelyaikaan vaikuttaa julkaisuviiveen lisäksi käsittelyaika. Patenteilla voi olla yli kahden vuoden käsittelyaika. Patenttijulkaisussa on ilmoitettu patentin hakupäivämäärä. Hakupäivämäärän ja julkaisupäivämäärän perusteella voidaan laskea patentin

käsittelyyn tarvittava aika. Ajantasaisempi tieto antaa mahdollisuuden tarkempaan tulevan teknologian ja sen kehittymisen ennustamiseen. Patenttien keskimääräinen käsittelyaika vaihteli kuvan 65 mukaisesti. Keskimääräisen käsittelyajan tai käsittelyajan mediaanin käyttö tunnuslukuna antaa heikon kuvan kuluneesta ajasta. Keskimääräinen käsittelyaika ja käsittelyajan mediaani eivät ole hyviä tunnuslukuja. Tarkempi tutkiminen osoittaa käsittelyajan koostuvan vähintään kahdesta eri ryhmästä. Kahden eri ryhmän käsittely omina joukkoinaan parantaa käsitystä patenttien käsittelyyn kuluneesta ajasta. Koska patenttietokanta on globaali ja tietoja saadaan useista paikoista, sisältää oletama kahdesta eri ryhmästä useita epävarmuustekijöitä.

Kuvan 65 aineistoa ei ole oikaistu uudemmalla otannalla patenttietokannasta. Toistamalla otanta myöhemmin nousisi patenttien määrä 9/2017 jälkeen noin 250 patenttiin/kk sekä keskimääräinen käsittelyaika putoaisi noin 200 päivään.



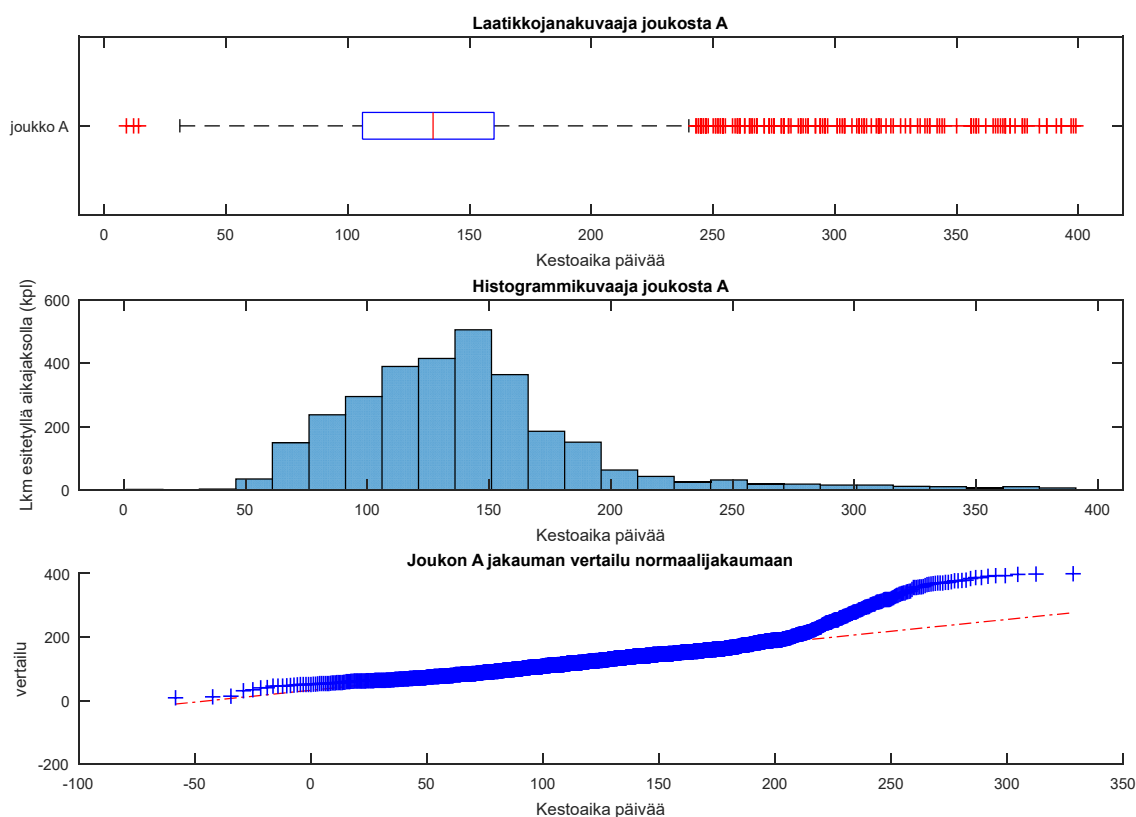
Kuva 65. Keskimääräinen patenttien määrä/kk (Y-akseli) ja käsittelyajan keston keskiarvo ajan funktiona (päivää X-akseli). Kuvaajien lähtödata perustuu 1.3.2018 otantaan. Laskenta on esitetty liitteessä (Liite K).



Kuva 66. Patentin käsittelyaikaa kuvaava histogrammi (liite L).

Patentin käsittely voi kestää 2,5 vuotta. Huolimatta pitkästä käsittelyajasta, hakemispäivä on merkityksellinen, koska myönnetylle patentille annetaan suoja takautuvasti hakemispäivästä alkaen. (lähde PRH, viitattu 24.4.2018 [<https://www.prh.fi/fi/patentit/useinkyytytya.html>]). Kuvassa 66 on esitetty julkaisu- ja hakemispäivämäärän erotus esittävään histogrammiin. Histogrammista voi päätellä, että keskiarvon tai mediaanin laskeminen

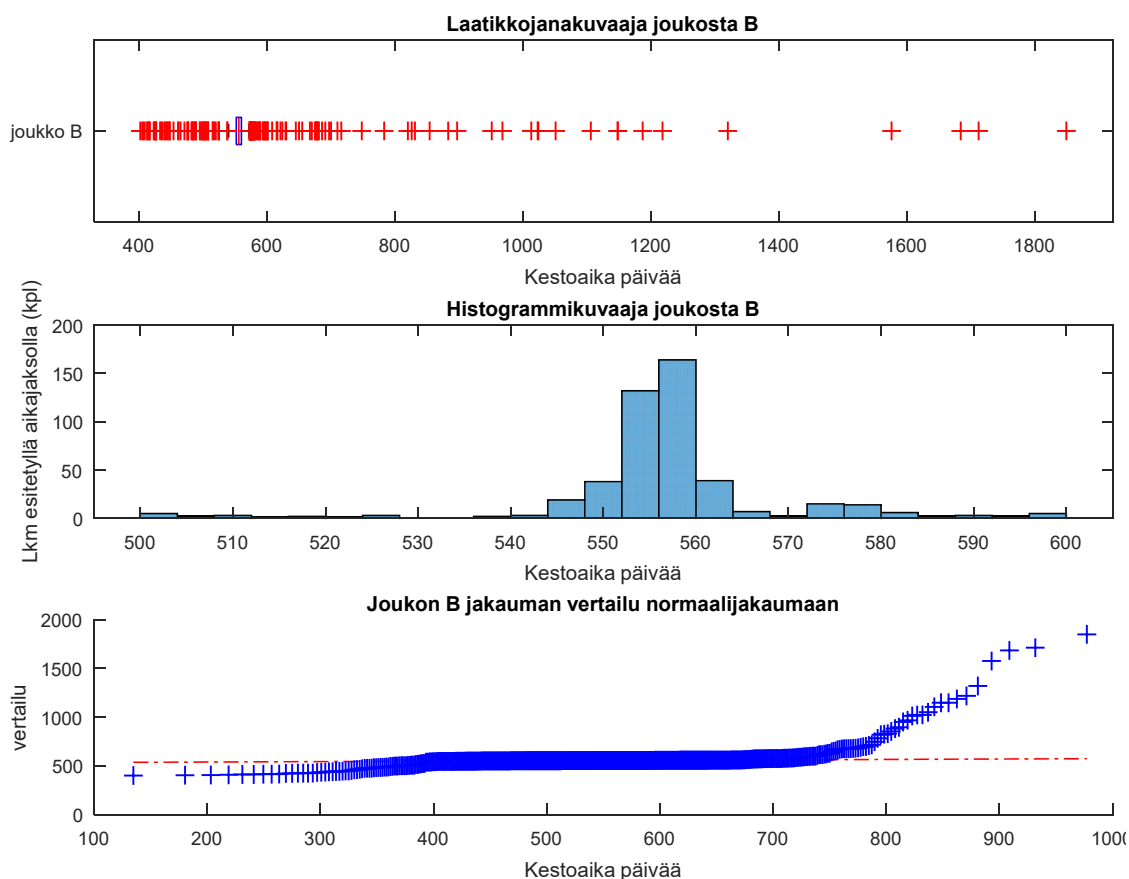
johtavat harhaisiin päätelmiin, koska kestoajassa on selvästi kaksi eri joukkoa. Tämä datajoukko jaettiin joukkoihin A (alle 400 päivää) ja joukkoihin B (vähintään 400 päivää).



Kuva 67. Laatikkojanakuvaaja, histogrammi ja vertailu normaalijakaumaan joukon A osalta (käsittelyaika alle 400 päivää). Laskenta on liitteessä (Liite M).

Laatikkojanakuvaajassa Matlab esittää kvantiilit 0,25, 0,50 ja 0,75. Kvantiileista voidaan todeta jakauman tasapainoisuus käsittelyajan ollessa 30 ja 240 päivän välissä (kuva 67). Mediaania kuvaava viiva (0,50) on asettunut tällöin kohtuullisen keskelle laatikkojanaa. Outlierit kuvan 67 laatikkojanan oikealla puolella on tulkittu jakaumaan kuulumattomaksi. Myös histogrammin perusteella alueen jakauma on vain jonkun verran vino noin 50-220 päivän välillä. Ottamalla useampia otantoja, voisi tuloksista saada tarkempia ja otantojen määrästä johtuva epävarmuustekijän vaikutus jäädä pienemmäksi. Joukon A jakauman vertailu normaalijakaumaan havainnollistaa joukon A datan olevan osittain kuvattavissa normaalijakautumalla pienillä arvoilla (kesto aika alle 240 päivää).

Kuvassa 68 toistettiin Joukon B (kestoaika vähintään 400 päivää) osalta analyysi samojen kuvaajien avulla. Joukon B osalta (vähintään 400 päivää, $n=569$ kpl) oli käsittelyajoista n. 60% välillä 550 ja 560 päivää ($n_{550-560} = 339$ kpl). Muilta osin aineisto jakautui hajanaisesti 400:n ja 2000 päivän välille, jolloin keskittymä (50% otoksesta) 550 päivän ympärillä on havaittavissa vain pienenä sinisenä symbolina laatikkojanakuvaajassa (kuva 69). Joukon B data on jakautunut huomattavan hyvin [400...780] päivän ajanjaksolla.



Kuva 68. Laatikkojanakuvaaja, histogrammi ja vertailu normaalijakaumaan joukon B osalta (vähintään 400 päivää). Laskenta on esitetty liitteessä (liite M).



Kuva 69. Laatikkojanakuvaaja patenttien myöntämisen prosessin kestosta niiden patenttien osalta, joiden käsittely kesti vähintään 400 päivää. Laskenta esitetty liitteessä (Liite N).

Laatikkojanakuvaaja kuvassa 68 sekä kuvassa 69 osoittavat sen, että kesto-aika noin 550-560 sisältää suuremman osan kaikista yli 400 päivän käsittelyajoista.

Haku Espacenet-tietokannasta antaa ensin arvion patenttien määrästä, mutta tarkentaa tulosta hakua jatkettaessa jääden kokonaismäärässä n. 96,5%:iin ensin ilmoitetusta kokonaismäärästä.

Haun tekemisen jälkeen otettiin otoksesta kaksi otantaa (n=518 ja n=317) aineiston laadun tarkastamista varten.

- Patenttipopulaatiosta poimitusta otoksesta alle 5% (129 kpl, n=3735) ei edustanut big dataa. Päätelmä patenttien edustavuudesta toteutettiin tulkitsemalla patentin otsikon merkitys. Otokseen tuli tulkinnan mukaan teknologioita, joissa esiintyi sanat big ja data eri merkityksessä.
- Suuri osa patenteista (61%) oli merkitty vain yhdellä IPC-koodilla (317kpl, v.2005-2015, n=518)
 - Otoksessa esiintyvistä ensisijaisista IPC-koodeista G06F 17/30 (176 kpl, n=317) oli noin 55%:lla patenteista, n=518
 - Otoksessa seuraavaksi yleisin IPC-koodi (H04L29/08, 18kpl, n=317) oli alle 6%:lla patenteista.
 - Suurin osa yhdellä koodilla kooditetuista patenteista sisälsi eri koodin (n=317, jossa hakusana ”Big data”).
- N. 27% patenteista sisälsi kaksi IPC-koodia (141kpl, v.2005-2015, n=518)
 - Koodeissa oli huomattavaa hajontaa, mutta yleisin koodi toisena koodina oli G06F 17/30, joka löytyi toisena koodina 36%:sta patenteista (51kpl, n=141).

Taulukko 38. Kooste patenttietokanta-aineiston IPC-koodi havainnoista (liite O)

IPC -koodien lukumäärä	n=3606
1	1766
2	1103
3	433
4	159
5	72
6	47
7	12
8	5
9	4

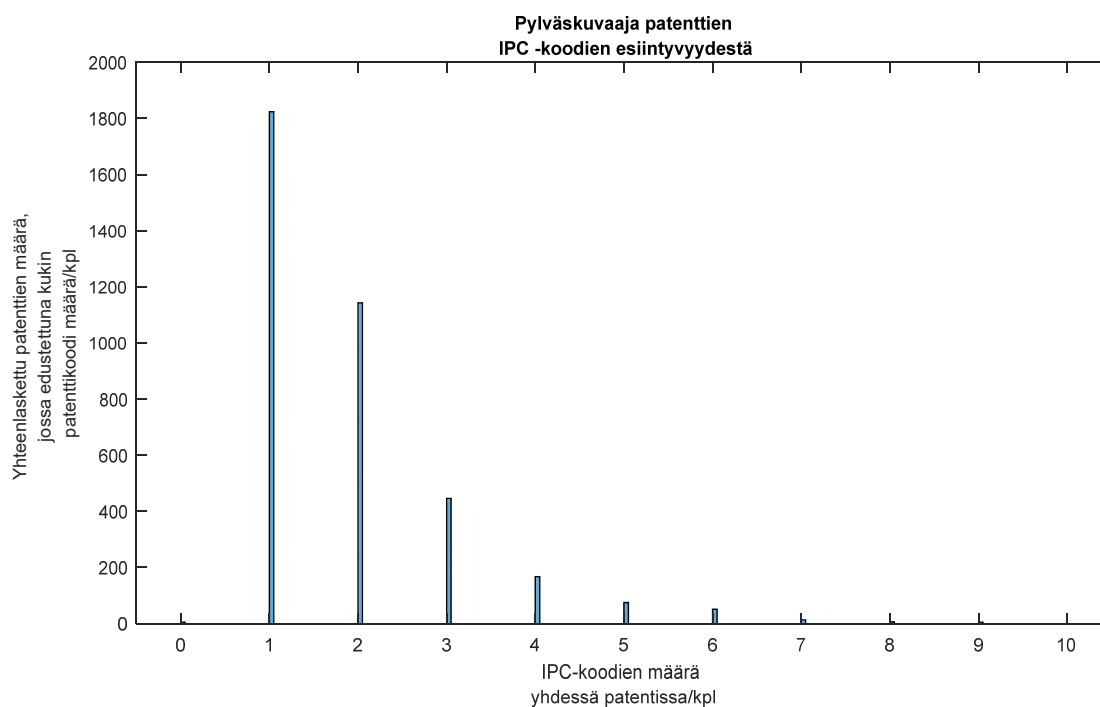
Kuvassa 70 on esitetty patentit, joissa on vain yksi IPC-koodi. Kooditus on tarkentunut ja useampia IPC-koodeja on tullut käyttöön. Aikaisemmin tarkemmin teknologiaa kuvaava koodi on voinut pudota toissijaiseksi koodiksi, koska ensisijainen koodi edustaa yläkategoriaa esimerkiksi datan käsittelyä.

Poimimalla kaikki koodit voidaan löytää etsittyjä teknologioita, mutta tällöin uudemmissa



patenteista löytyisi etsitty teknologia huomattavasti todennäköisemmin.

Kuva 70. Patenttien koodituksen määrän lisääntyminen ja vain yhdellä patenttikoodilla esitettyjen patenttien suhteellisen määrän pienentyminen ajan funktiona. Lähtötiedot esitetty liitteessä (liite O)



Kuva 71. Pylväskuvaaja IPC-koodillisista patenttien määrällisistä osuuksista (kpl) patenteista koodien lukumäärän mukaan (Liite P).

Päätelminä hakutuloksista: Tehtäessä hakuja IPC ja CPC -luokituksella saatiin runsaasti tuloksia, mutta tulosten kohdistaminen tiettyyn teknologiaan ei ole kattavaa. Tämä johtuu

pääosin patenttien niukasta koodituksesta, koska haettu patentti harvoin suoraan liittyy käytettyyn teknologiaan, vaan ainoastaan hyödyntää sitä. Koodien määrä on myös muuttunut ajan funktiona (kuva 71), jolloin koodin käyttäminen ajallisessa vertailussa ei anna yhteismitallisia tuloksia. (Liite O)

Teknologian yleistymisessä on kuitenkin liitännäinen teknologia merkityksellinen, koska se kuvaa teknologian yleistymistä ja hyödyntämistä eri yhteyksissä.

5.6 Yhteenveto patenttitietokantatutkimuksen aineistosta

Patenttitietokantaa tutkittiin valitsemalla tutkittavaksi attribuuteiksi koodistot, julkaisupäivämäärä ja hakupäivämäärä. Syy näiden ominaisuuksien valintaan oli teknologian kehittymisen ajallinen seuranta ja aineiston kattavuus. Syvällisempi arviointi vaatisi patenttien yksityiskohtaisemman läpikäymisen, jotta aineistosta saataisiin yhtä näkökulmaa varten yhteismitallinen. Lähtöhypoteesina on, että patenttien määrä kuvaa teknologian yleistymistä, koska teknologialla nähdään jo kaupallinen suojaamisen tarve.

Taulukko 39. Kooste patenttitietokanta-aineiston ominaisuuksista.

Aineisto-ominaisuus	Havainto
Otoskoko	Patenttitietokannasta saatiin otsikkoon kohdistetulla haullla 3735 patenttia, joka on riittävä määrä erilaisien ennusteiden tekemiseen ajan funktiona (2009-2018).
Viive patentin rekisteröimisessä	Tulosten mukaan patentti rekisteröitiin Espacenet -tietokantaan laskentatavasta riippuen noin 3-6 kuukauden kuluessa julkaisusta. Rekisteröinti-viive haittaa patenttitietokannan käyttämistä uusien teknologioiden enustamisessa. Uusimpien tietojen osalta aineisto ei ole käyttökelpoista ilman korjaavaa laskentaa.
Virheet tekstihakuun perustuvassa poiminnassa	Päätelyn perusteella tekstihakuun perustuvassa poiminnassa oli kohtuullisen vähän virheellisiä tuloksia (n=129). Virheet kuitenkin painottuivat ennen teknologian yleistymistä, jolloin teknologian varhaisessa vaiheessa virheitä oli kohtuuttomasti (yli 90% esiintymistä), joka aiheuttaa jakauman vinouman ja virheellisiä päättelyitä teknologian varhaisessa vaiheessa.
Virheet patenttikoodeihin perustuvassa poiminnassa	Patenttikoodit perustuivat mm. IPC ja CPC-koodituksiin. Patenttien kooditus oli CPC-koodien osalta erittäin puutteellista (n. 60% sisälsi koodin) ja IPC-koodien osalta kooditus koski usein vain yhtä teknologiaa, jolloin big dataa hyödyntävä patentti ei tule mukaan hakutulokseen.
Patenttikoodien tarkentuminen ajan funktiona	Tässä otannassa toissijaisen patenttikoodien määrä lisääntyi ajan funktiona. Vanhempien ja uudempien patenttien vertailu toissijaisten koodien perusteella antaa virheellisen kuvan patenttikoodimäärien kehittymisestä. Toissijaisia koodeja käytettäessä tulee huomioida tämä reunaehto.
Viive patentin käsittelyprosessissa alkuaen hakemisesta julkaisuun.	Patenttien käsittely sisälsi kaksi ajallisesti erillistä joukkoa. Molempien joukkojen keskimääräiset käsittelyajat (n. 130 päivää ja n. 550 päivää) osoittavat patenttitietokannan reagoivan varsin patenttien hakuun.

5.7 Saavutettujen tulosten ja niiden merkityksen esittely

Mitä asioita tulisi huomioida teknologioiden ennustamisessa? Teknologia kehittyy vaiheittain ja yllättävän hitaasti, vaikka teknologiadiffuusiovaihe voi olla erittäin nopea. Teknologiaa voidaan ennakoida ennen teknologiadiffuusion ilmenemistä. Aihealue käsittää monia kokonaisuuksia, joilla on riippuvuuksia keskenään. Ennakointi perustuu aikaisempiin aineistoihin ja ilmiöihin. Teoriassa ja laskentamallissa pyrittiin keskittymään ongelmakentän keskeisiin kysymyksiin. Aineistojen tunteminen auttaa löytämään kronologisesti aikaisemmin sijoittuvat aineistot. Useamman kronologisesti eriaikaisen tietolähteen yhtäaikainen käyttö ennakoinnissa antaa ennustamiseen selkeän edun, parantaa tarkkuutta, varmuutta sekä antaa ennustamiselle runsaasti aikaa. Varhaisemmassa vaiheessa olevan lähteen avulla voidaan ennakoida esimerkiksi patenttien määrän kasvua ennen patenttien määrän kasvun alkamista. Diffuusiovaiheen eli käyttöönottamisen mittaamisesta on suurin hyöty hitaasti kehittyvässä teknologiassa. Diffuusiovaiheen nopeus aiheuttaa sen, että ennustaminen vain diffuusiovaiheesta voi olla haastavaa.

Teknologiaan tulee tutustua riippuvuuksineen ennen lähteiden valintaa. Tutustuminen tulee tehdä, jotta voi ymmärtää kontekstin valitusta teknologiasta. Konteksti vaikuttaa hyvien tietolähteiden valintaan. Kontekstiin tutustumalla voi myös valmistella mittausparametrien valintaa sekä monta muuta asiaa, kuten oikean laskentamallin valintaan.

Mitä vaatimuksia teknologiadiffuusion mitattaville asioille tulee asettaa? Hyvät lähdeaineistovalinnat ovat selvitettävissä kirjallisuustutkimuksien avulla. Kaupallisten tietueiden ja eri tyyppisten artikkeleiden saatavuus mahdollistaa eri kronologisten vaiheiden esittämisen sekä niiden ennakoinnin. Samassa tietolähteessä voi olla runsaasti eri tyyppisiä aineistoja. *Pienet ominaisuuksien erot mittauskohteissa* sijoittavat mittauskohteet kronologisesti eri teknologian kehityksen ajankohtiin. Rahoituksen, ulkoisten ympäristökäyttäjien, rakenteettoman datan, sanahakutulosten ja sosiaalisen median huomioiminen antaa ennakoinnissa perinteisiin malleihin verrattuna etua, joka puoltaa big data -tyyppisen ratkaisun käyttöä ennakoinnissa. Big dataa hyödyntävän teknologian, kuten koneoppimisessa on mahdollisuutena useiden tietolähteiden yhtäaikainen hyödyntäminen. Koneellisesti toteutettu luokittelu mahdollistaa ihmistyönä kohtuuttoman monen luokittelun toteuttamisen yhtäaikaisesti. Tekoälyyn ja datatieteisiin perustuvien korrelaatioiden löytäminen kohtuullisen nopeasti voi olla koneoppimisen keinoin helpompaa, kun dataa on paljon. Tekoälyn kuten koneoppimisen, automaattisen luokittelun ja mm. syväoppimisen hyödyntäminen big data -ratkaisun kanssa on luontevaa. Tällöin dataa koneellisesti yhdistelemällä voidaan löytää uusia näkökulmia, joita ihminen ei järkevällä työmäärällä löydä. Hyödyntäessä sosiaalista mediaa, hakukoneita, sähköpostia ja montaa muuta nykyaikana tietotekniikan peruspalvelua tulee käyttäneensä tai olleensa osa mm. big data -teknologian ratkaisuja.

Miten teknologiadiffuusion menetelmävalinnat vaikuttavat ennustustarkkuuteen? Kirjallisuustutkimuksen tuloksena koottiin yksi näkemys eri tyyppisten aineistojen ja ilmiöiden kronologisessa esiintymisestä teknologian kehittämisessä. Chambers, Mullick ja Smith (1971) esittivät, että pitkän ajan ennustuksissa matemaattiset mallit tuottavat tarkempia ennustuksia. Armstrong (2003) esitti valmiita malleja teknologiaennustamiseen. Hän myös arvosteli terävin sanoin Chambers *et al.* (1971) tutkimusta. Nykykirjallisuus puoltaa useiden ennustusmallien yhdistämistä, jossa laskentaan liitetään useita aineistoja, kirjallisuustutkimus sekä teknologiatiekartta. Konteksti vaikuttaa teknologiadiffuusion laskentamallin ja tietolähteiden valinnassa. Riippuvien ja rinnakkaisteknologioiden kehittyminen vaikuttaa, mutta myös mahdollistaa teknologian kehittymisen. Toimintaympäristö vaikuttaa teknologian kehittymiseen. PESTEL-analyysi on hyvä menetelmä ulkoisen toimintaympäristön arvioimiseen. Esimerkiksi ympäristösytyt, mutta myös lainsäädäntö vaikuttavat voimakkaasti teknologioiden kehittymiseen.

Kirjallisuustutkimuksen tuloksia testattiin kohdassa 3.2. Testauksessa voitiin todeta (taulukko 17 ja kuva 39) ilmiöiden seuraavan toisiansa kirjallisuudessa esitetyllä tavalla. Kuvan 39 esityksessä hyödynnettiin artikkelin ominaisuuksia otannan parametrien valinnassa. Ominaisuuksilla voitiin osoittaa, että ne ilmenevät teknologian kehittymisen eri vaiheessa. Tällaisia ominaisuuksia on esimerkiksi artikkelityyppi. Otoksiin poimitujen aineistojen järjestys vastasi konferenssijulkaisujen ja tieteellisten artikkeleiden osalta kirjallisuudessa esitettyä kronologista järjestystä.

Mittausparametrin valinta ohjaa tuloksia merkittävästi. Teknologiadiffuusiota voidaan mitata lukumääräisesti. Lukumääräinen mittaaminen on valinta, joka ohjaa tuloksia. Teknologiadiffuusion arvioinnissa mitataan yleensä adaptoivien käyttäjien määrää. Käyttäjien määrää voidaan mitata ajan funktiona. Teknologioiden kehittymisen mittaamisessa voidaan myös mitata teknologian hyödyntämistä. Matkapuhelimien osalta voidaan esimerkiksi valita joitakin seuraavista mittauskohteista: puhelinten määrä, puhelujen pituus, puhelujen määrä, käyttäjien määrä, käyttöalueen laajuus tai soittajien ja vastaajien määrä. Usein on etua, jos valitaan useampia mittausparametreja tai aineistoja, jolloin yksittäinen aineisto tai parametri ei täysin ohjaa tuloksia. Mitattavan parametrin valinnassa tulee huomioida parametrin kehittyminen. Korkea teknologia voi olla virtuaalista ja yksi ratkaisu voi maantieteellisesti palvella globaalisti – myös yksi patentti voi palvella globaalisti, jolloin lukumäärän laskeminen ei ole mielekäästä. Teknologia voi olla myös osa toista teknologiaa, jolloin käyttäjä tai palvelun tuottaja tee päätöstä teknologian käyttöönottamisesta, vaan se toi toimia taustalla huomaamattomasti. Uusissa teknologioissa käyttäjät eivät välttämättä tiedä hyödyntävänsä teknologiaa. Tutkimus käytöstä voi tällaisessa tapauksessa antaa myös virheellisen kuvan teknologiadiffuusiosta.

Mittauksien lisäksi tulokset tulee koostaa ja arvioida tulosten laatua. Tässä työssä tutkittiin tarkemmin patenttitietokannan otantoja. Tulokset olivat yllättäviä ja ne saatiin toistettua. Uusimmat aineistot eivät välttämättä ole varmimpia ja vaativat huolellisen ai-

neisto-oikaisun. Huolimaton aineiston käsittely ilman aineisto-oikaisua antaa jopa päinvastaisen kuvan todellisesta tilanteesta. Ilmiötä tutkittiin ja patenttitietokannan osalta saatiin useita havaintoja virheellisten päätelmien mahdollisuuksista.

Työssä käytiin läpi perusteoria ja kehitettiin sen lähdeaineistot integroiva Bassin laskentamalli, jolla ratkaistaan tulevien teknologiadiffuusion kasvun arvio. Varsinaisessa kvantitatiivisessa tutkimuksessa hyödynnettiin kirjallisuustutkimusta sekä tietoa yksityiskohden merkityksestä aineistovalinnoissa. Varsinainen laskentatyö on kohtuullisen nopea suorittaa ja suuri osa ajasta kuluukin aineistorojen selvittämisessä sekä hakujen toistamisessa. Tutkimusongelmaan ja tutkimuskysymyksiin löytyy vastauksia. *Huolellisilla aineistovalinnoilla sekä useamman aineiston integroinnilla* voidaan saavuttaa ennuste teknologiadiffuusiosta ennen sen varsinaista kehittymistä. Laskentamallin valinta toteutettiin kirjallisuustutkimuksen perusteella. Bassin laskentamalli sopii uusien tuotteiden ja esimerkiksi Fisher-Pry sopii paremmin korvaavan teknologian teknologiaennustamiseen. Laskentatulokset tulee visualisoida. Visualisointi auttaa ymmärtämään teknologiadiffuusion eri vaiheiden kehittymistä. Valitsemalla teknologian kehittymisen aikaisempia vaiheita, voidaan teknologian kehittymisen arvioida myös visuaalisesti.

Tämän työn kvantitatiivisessa tutkimuksessa valittiin tutkittavaksi ilmiöksi big data. Big data, kuten monet muutkin nykyilmiöt ovat olleet voimakkaan hypen kohteena. Tämän työn tuloksien perusteella big datan kehittyminen on mahdollista saavuttaen huippunsa arviolta vuonna 2023-2025. Teknologiaennuste vaatii vielä varmistuksen vuosina 2019 ja 2020, koska *investointipäätökselle tärkeää kriittistä massaa ei ole vielä saavutettu patenttitietokannan perusteella*. Tämä voi olla selitys kvantitatiivisen tutkimuksen tutkimuksia kuvaavan q -parametrin heikommasta arvosta (noin 0,3 aineistossa 2), joka kuvaa seuraajien määrää. Myös Gartner pudotti big datan pois omalta hypekäyrältään, joka voi indikoida vaikeaa ennustettavuutta tai liiallista hypeä suhteessa teknologiadiffuusioon.

Investointiajankohta ja monet muut teoriat teknologian kehitymisestä perustuvat Bassin laskentamallin mukaiseen S-muotoiseen käyrään. Laskentamallin käyttäytyminen on olennaista, että tulokset voidaan esittää tai verrata tässä muodossa esimerkiksi *oikean investointiajankohdan tai teknologian hiipumisen arvioimiseksi*. Teknologiaennustamisen varhaisessa vaiheessa voidaan arvioida innovaattoreiden määrää. Kohtuuttoman suuri innovaattoreiden osuus lupaa huonon ennusteen teknologian diffuusion kasvulle. Innovaattoreiden suhteellinen määrä tulee arvioida hyödyntäen esimerkiksi regressioanalyysiä saatavasta diffuusiota kuvaavasta datasta. Tämä voidaan tulkita esimerkiksi p ja q -parametrin arvioidusta suuruudesta.

Käytetyn menetelmän heikkona puolena on monimutkaisuus sekä laskennan ja aineistokäsittelyn vaatimukset. Bassin diffuusiomallin laskentaan sekä erityisesti laskentaparametrien arviointiin tarvittavaa osaamista on rajallisesti ja laskentatulosten sekä sen välivaiheiden varmistaminen on haastavaa. Laskennassa on useita vaiheita, joista yksittäiset virheellisesti suoritettavat vaiheet voivat johtaa vääriin päätelmiin.

Työn laajuus yllätti ja kirjallisuuden sekä kvantitatiivisen tutkimuksen tuloksien oikeellisuutta olisi voitu varmistaa useilla eri tavoilla eri vaiheissa. Lopputuloksena esitellyn integroidun mallin ”viivettä” patenttien, tutkimuksien ja kaupallisten artikkelien välillä vertailtiin taustatyönä. Suuruusluokaltaan päädyttiin noin 4-6 vuoden eroon patentti ja markkinadatan välillä. 11/2018 tarkastettiin, että big data patenttien määrä ovat kasvaneet kuukausittain voimakkaasti 1/2018 ja 6/2018 välisenä aikana. Tämä voi vahventaa työssä esitettyä tulosta, että kriittistä massapistettä ei ole saavutettu 2017 lähdeaineistoon perustuen. Tuloksia voidaan jalostaa ja osoittaa menetelmiä osateknologioiden integroimiseksi samalla laskentamallilla. Tuloksia on tarkastettu esitetyiltä osin, kuten patenttitietokantaan tehdyn aineistotutkimuksen osalta. Työtä olisi voitu jatkaa tutkimalla laskentamallin tarkkuuden arvioinnilla ja laskentamallin painoarvojen optimoinnilla paremman tarkkuuden saavuttamiseksi. Laskentatulosten arviointi parantaisi myös tuloksissa arvioitua tarkkuutta merkittävästi eli sovittamalla toteutuvien patenttien julkaisua ennakoitiin, voidaan iteroiden saavuttaa jatkuvasti tarkempaa ennakointia.

6. YHTEENVETO JA JATKOTUTKIMUKSIEN AIHEET

Kirjallisuustutkimuksessa koottiin useista eri lähteistä teknologian kehittymisen vaiheita. Eri vaiheet ja vaiheissa olevat teknologiaennustamislähteet osoittavat kronologisesti järjestettynä teknologian kehittymisen pitkän aikajänteen. Big datan osalta osoitettiin, että teknologia alkoi kehittymään jo 20 vuotta ennen laajempaa käyttöä. Kirjallisuustutkimuksessa tuotiin esille diffuusiovaiheen jatkuva nopeutuminen erityisesti digitalisaatiota koskevissa tutkimuksissa, kuten mm. Diermeier & Goecke (2017). Tällöin *mahdollisuudet tarkan teknologiaennustuksen tekemisestä teknologiadiffuusiovaiheen perusteella heikkenevät jatkuvasti*, mutta diffuusion ennakkointiin voidaan hyödyntää aikaisempia kronologisia vaiheita.

Diplomityössä keskityttiin tutkimusongelmaan eli uuteen orastavan teknologiaan sopivan ennustusmenetelmän löytämiseen. Työssä huomioitiin ennustamiseen sopivat lähteet, ennustusmallit sekä teoriaa uusien orastavien ennustusmenetelmien hyödyntämiseksi. Kirjallisuustutkimuksen perusteella valittiin ennustettavalle kohteelle sopiva laskentatapa- ja malli. Laskentamallia sovellettiin kvantitatiivisessa tutkimuksessa. Kvantitatiivisessa osuudessa toteutettiin teknologiadiffuusiolaskenta ja -ennuste tutkimuksen kohteena olevaan ilmiöön.

Tutkimustuloksena kehitettiin painotetuilla summilla rikastettu *integroitu Bassin diffuusiolaskentamalli*, joka integroi useita ennustuslähteitä yhtäaikaaisesti. Laskentamallissa integroidaan usean ennustuslähteen datan käyttäytymistä mittaavat parametrit Bassin laskentamallin keskeisiin laskentaparametreihin. Laskentamalli mahdollistaa teknologian ennustamisen useista eri lähteistä yksilöllisillä painoarvoilla. Painoarvot voidaan asettaa parametreittain. Laskentamallilla voidaan arvioida teknologian kehittymistä ja arvioida liiketoimintapäätöksien oikea-aikaisuutta huomioiden nousevat ja hiipuvat teknologiavaliinnat. Työssä keskityttiin laskentamallin laskentaan ja mm. *parametrien laskentaa ei varmistettu tässä työssä*, kuin vertaamalla lukuja vastaaviin parametrien arvoihin. Laskennassa voi olla virheitä. Menetelmän osalta laskennan integrointia tutkittiin ja siihen ei löydetty estettä. Menetelmän toimivuutta tulisi tutkia tarkemmin, parantaako edellisten vaiheiden tai osateknologioiden huomioiminen diffuusiolaskennan tarkkuutta?

Tietolähteet valitaan teknologian käyttötarkoituksiin tutustumisen perusteella. Valinta tehdään kronologisilla perusteilla, jossa huomioidaan tutkimuskonteksti. Tietolähteissä voidaan antaa painoarvoja esimerkiksi talous- ja rahoitusmekanismeille, joka voi edistää tutkimusta ja sen kautta lisätä patenttien määrää ja kaupallisia ratkaisuja. Teknologiaen-

nusteessa tulee huomioida mm. poistuva, syrjäytyvä, rinnakkaisteknologia ja osateknologioiden arviointi. Kehitetyllä laskentamallilla tämä voidaan toteuttaa kohtuullisen helposti.

Ennustukseen käytetty lähde voidaan valita myös teknologian kehittymisen nopeuden mukaan. Työssä tuotiin esille eri teknologioiden erilaiset kehittymisnopeudet, joka voidaan arvioida tunnistamalla teknologia sekä sen riippuvuudet. Osa ennustuslähteistä tuottaa laskennan avulla tarkempia ennustuksia. Mikäli ennustus vaatii tarkkuutta, on ennustukseen käytetyillä lähteillä huomattava merkitys. Teknologian kehittymisnopeudella ei välttämättä ole suoraa yhteyttä teknologian yleistymiseen eli markkinakokoon. Hitaasti kehittyvän teknologian määrä voi kasvaa nopeasti kehittyvää teknologiaa suuremmaksi. Uutta teknologiaa voidaan ennustaa myös aikaisemman teknologian käytön vähenemisellä.

Aineiston poiminnassa huomioitavilla parametreilla on huomattava merkitys. Kirjallisuustutkimus ennen poimintaa mahdollistaa parempilaatuisen poiminnan. Poimittua materiaalia voidaan tutkia mm. tilastollisin menetelmin tai big dataa hyödyntämällä pyrkiä poistamaan heikkolaatuiset aineistot.

Diplomityössä tunnistettiin seuraavat jatkotutkimusaiheet:

1. Äärimmäisen nopeasti kehittyvän teknologian tunnistaminen luokittelun avulla. Diffusiovaihe voi olla äärimmäisen nopea. Työssä on esitetty taulukossa Teknologioiden jakaminen hitaasti ja nopeasti kehittyviin (Linden & Fenn 2003). Digitalisaatio on osoittanut, että käyttäjien määrä uudella palvelulla voi kasvaa erittäin nopeasti. Diermeier & Goecke (2017) esittivät, että Google+ saavutti 10 miljoonaa käyttäjää 16 päivän aikana. Suuret palvelun tarjoajat voivat tarjota aktiivista käytön hyväksyntää, jossa palvelun sisältö ratkaisee palvelun käyttöön ottamisen. Ilmiötä on tutkittu, mutta tutkimusta, jossa diffusiovaiheen kehittymisen riippuvuutta tarjoavan yrityksen käyttäytymiseen ei löytenyt. Aikaisemmat teoriat pureutuvat pääosin käyttäjän valintoihin. Käyttäjät tekevät digitaalisessa maailmassa useita valintoja, joissa heillä varsinaisen palvelun ansaintamekanismista heikot tiedot.

2. Aineisto-ominaisuuksien kehittyminen ja määrien ennustaminen perustuen ominaisuuksiin. Patenttietokannassa havaitut viiveet ja muut ongelmat olivat toistuvia ilmiöitä. Ilmiö korjaantui ajan kuluessa ja noudatti lähes samaa aikajaksoa, jolloin *patenttien määrän kasvua voisi ennakoida tarkemmin*. Patentit ovat vain yksi aineistolähde. Eri lähdeaineistoilla on erilaiset ominaisuudet. Tutkimus lähdeaineistojen ominaisuuksista voisi parantaa teknologian ennustamista laskentamalleilla. Artikkelitietokannoissa viive oli pienempi, mutta toisaalta joissakin artikkelikannoissa näkyi myös tieto tulevista julkistuksista.

3. Big data tai monimuuttujapohjainen ennustaminen *integroidulla Bassin laskentamallilla*. Jatkotutkimuksen kohteena voisi olla ennakoinnin parantaminen hyödyntämällä

useiden erityyppisten asioiden yhdistämisellä big data analytiikkaan (BDA) perustuen. Teknologiaennustuksen laskennan integrointi ja automatisointi. Työssä havaittiin, että laskenta perustuu koodiin, joka on toteuttavissa eri ohjelmointikielillä. Laskentakokonaisuuden voisi automatisoida, jossa esimerkiksi koneoppimisen automaattisella luokittelulla valitaan ennustettavan teknologian lähteet ja ennustusmalli. Laskentamallia kehitettiin tavalla, jossa voidaan erottaa eri lähdeaineistoja muodostavat tekijät sekä vaikuttavat niiden painoarvoon. Kehitettyä laskentamallia voisi kehittää *integroituvaksi ns. monimuuttuja laskentamalliksi*, jossa laskentaan voidaan luokittelun kautta tuoda koneellisesti suuri määrä lähdeaineistoja omilla painoarvoillaan, joiden painoarvot lasketaan esimerkiksi sosiaalisella medialla, myyntimäärillä, Iot-laitedatalla tai PESTEL-analyysin avulla.

4. Innovaatiodiffuusiomalli perustuu innovaattoreiden ja seuraajien osuuksien arviointiin. Laskentamallia voisi kehittää Rogersin ja Bassin viiteen luokkaan perustuen. Bassin malli huomioi innovaattorit ja seuraajat ja viidellä luokalla voidaan parantaa ennustuksen tarkkuutta. Riippuen teknologiasta, voisi teknologia-adaptiota kuvata eri ryhmien käyttäytymisellä paremmin. Verkostovaikutuksessa kriittisen massan määrä oli vaihdellut 1 ja 34 prosentin välillä. Suuri prosentuaalinen vaihtelu voi kuvata muitakin ilmiöitä, kuin yhteen luokkaan kuuluvien henkilöiden käyttäytymistä. Olennaista on Bassin mukaan adaption ajoitus eli kaupallisen kannattavuuden osalta kriittisen massapisteen (CMP) arviointi. Laskentamallissa tulisi huomioida kulttuuri ja alueelliset erot.

5. Talousaaltojen vaikutuksen teknologian kehittymiseen tarkempi tutkiminen. *Kehittyvien talouksien maissa adaptiodaan uusia teknologioita* nopeasti. Tutkimalla talouden suhdannevaihteluita eri alueilla, voidaan kasvattaa teknologiadiffuusiota ja markkinoille pääsyä nopeammin. Suuryritykset hyödyntävät tätä tietoa. Tutkimusalue voisi olla myös kaupallisesti kiinnostava. Uusien tuotteiden vienti voisi kohdentua tarkasti ja onnistua paremmin kehittyvillä alueilla, joissa kyseinen teknologia korjaa ongelman tai tuo erityisen kilpailukyvyn.

6. Teknologioiden keskittyminen ja uusien teknologioiden siirtyessä pilveen teknologiadiffuusio leviää todennäköisesti nopeammin. Sundqvist *et al.* (2002) tutkivat verkostovaikutuksia ja valmiit infrastruktuurit voivat tarjota vastaavia vaikutuksia teknologian kehittymiseen. Eri tyyppisten alustojen yleistyessä teknologioiden kehittymisen muutosnopeus voi kasvaa edelleen.

7. Bassin diffuusiomallissa laskentaparametrien m -parametrin huomioiminen, jolloin voidaan arvioida määrää, johon teknologiadiffuusio kasvaa sekä laskennan integraatiovaihtoehtojen vertailu. Integraatio voidaan toteuttaa jo lähtöluvussa tai muillakin tavoin. Eri tyyppisiä integraatioita voidaan toteuttaa koneoppiviin ratkaisuihin, joissa korrelaatiota etsitään p ja q -parametrien lisäksi lähtödatojen korrelaatioista tai todennäköisyyksistä.

LÄHTEET

- Ailisto, H., Heikkilä, E., Helaakoski, H., Neuvonen, A. & Seppälä, T. 2018. Tekoälyn kokonaisjäsenitys ja kansallinen osaamiskartoitus. VTT Technical Research Centre of Finland Ltd, Helsinki. 25 s. [Tiivistelmä väliraportista]. [Viitattu 23.9.2018]. Saatavissa: <https://www.vtt.fi/Documents/uutiset/DataK%C3%A4sitekartta%20AI.pdf>.
- Al-Sabri, H. & Al-Mashari, M. 2017. Process-Oriented Requirements Engineering: User-Centric LORS Framework. *Journal of Software Engineering and Applications*. [Artikkeli]. Vol. 10:2, S. 95-127. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.4236/jsea.2017.102006.
- Altuntas, S., Dereli, T. & Kusiak, A. 2015. Forecasting technology success based on patent data. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 96, S. 202-214. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1016/j.techfore.2015.03.011.
- Antolini, F. & Grassini, L. 2018. Foreign arrivals nowcasting in Italy with Google Trends data. *Quality & Quantity, International Journal of Methodology*. [Artikkeli]. S. 1-17. [Viitattu 3.7.2018]. Online ISSN: 1573-7845. Saatavissa: DOI: 10.1007/s11135-018-0748-z.
- Armstrong, J.S. 2003. Selecting Forecasting Methods. In *Principles of Forecasting : A Handbook for Researchers and Practitioners* (Ed. J. Scott Armstrong). Kluwer, 2001. [Artikkeli/painettu julkaisu]. University of Pennsylvania. Saatavilla: http://repository.upenn.edu/marketing_papers/14.
- Bass, F. 2004. Comments on 'A New Product Growth for Model Consumer Durables The Bass Model'. *Management Science*. [Artikkeli]. Vol. 50:12, S. 1833-1840. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1287/mnsc.1040.0300. [Perustuu teoksien: DOI: 10.1287/mnsc.15.5.215 ja DOI: 10.1287/mnsc.1040.02649 sisältöön].
- Bass, F. 1969. A new product growth model for model consumer durables. *Management science*. [Artikkeli]. Vol. 15:5, S. 215-227. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1287/mnsc.15.5.215.
- Bass, F. 1963. A Dynamic Model of Market Share and Sales Behavior. *Proceedings, Winter Conference American Marketing Association, Chicago, IL*. [Artikkeli]. S. 263-276. [Viitattu 3.7.2018]. Saatavissa: <http://bass-basement.org/F/N/FMB/Pubs/Bass%201963%20Dynamic%20Model.pdf>.
- Bessant, J. 2015. Innovation Strategy. Homepage of John Bessant. [Verkkójulkaisu], S. 24-27. [Viitattu 7.9.2018]. Saatavissa: <https://johnrbessant.files.wordpress.com/2015/03/jbtidd1-c02-sim.pdf>.
- Blackburn, R. 1939. Business Cycles. *Accounting Review, Business Source Complete*. [Painettu julkaisu], [Artikkeli]. Vol. 14:4. S. 457, [Viitattu 3.10.2018]. Saatavissa: <https://www.jstor.org/stable/239377>.

- Blazquez, D. & Domenech, J. 2018. Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 130, S. 99-113. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1016/j.techfore.2017.07.027.
- Bodrožić, Z. & Adler, P.S. 2018. The Evolution of Management Models: A Neo-Schumpeterian Theory. *Administrative Science Quarterly*. [Artikkeli]. Vol. 63:1, S. 85-129. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1177/0001839217704811.
- Bollinger, B. & Gillingham, K. 2012. Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Yale School of Forestry & Environmental Studies*. [Artikkeli]. Vol. 31:6, S. 873-1025. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1287/mksc.1120.0727.
- Bolt, J., Inklaar, R., de Jong, H. & van Zanden, J.L. 2018. Maddison Project Database, version 2018. Rebasng 'Maddison': new income comparisons and the shape of long-run economic development. *Maddison Project Working Paper*, nr. 10. [Tietokanta]. [Viitattu 23.8.2018]. Saatavissa: www.ggdc.net/maddison.
- Bresciani, S. & Eppler, M. 2010. Gartner's magic quadrant and hype cycle. Case study. *Collaborative Knowledge Visualization Case Study Serie*, Case Nr. 2, 2008 – Updated in 2010. [Artikkeli]. S. 1-18. [Viitattu 3.7.2018]. Saatavissa: <https://www.researchgate.net/publication/265407275>.
- Caplice, C. & Sheffi, Y. (ref. Mahender Singh) 2006. ESD.260J Logistics Systems. *Massachusetts Institute of Technology*. [Opintomateriaali]. S. 22-24. [Viitattu 23.9.2018], Saatavissa: <https://ocw.mit.edu/courses/engineering-systems-division/esd-260j-logistics-systems-fall-2006/lecture-notes/lect5.pdf>.
- Chambers, J., Mullick, S & Smith, D. 1971. How to Choose the Right Forecasting Technique. *Harvard Business Review*. [Artikkeli]. July 1971. Saatavissa: <https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique>
- Cho, Y. & Daim, T. 2016. OLED TV technology forecasting using technology mining and the Fisher-Pry diffusion model. *Foresight: the Journal of Futures Studies, Strategic Thinking and Policy*. [Artikkeli]. Vol. 18:2, S. 117-137. [Viitattu 5.3.2018]. Saatavissa: DOI: 18.10.1108/FS-08-2015-0043.
- Coccia, M. 2017. The source and nature of general purpose technologies for supporting next K-waves: Global leadership and the case study of the U.S. Navy's Mobile User Objective System. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 116, S. 331-339. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1016/j.techfore.2016.05.019.
- Coccia, M. 2018. A Theory of the General Causes of Long Waves: War, General Purpose Technologies, and Economic Change. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 128, S. 287-295. [Viitattu 8.10.2018]. ISSN 0040-1625. Saatavissa: DOI: 10.1016/j.techfore.2017.11.013.

- Davenport, T.H. (ed) 2014. Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press, Boston, Massachusetts. [Painettu julkaisu]. ISBN: 1422168166, 9781422168165, 240 s.
- Davis, J. 2008. 15.912 Technology Strategy. MIT Sloan School of Management [Homepage of MIT Sloan School of Management]. [Opintomateriaali]. [Viitattu 28.9.2018]. Saatavissa: https://ocw.mit.edu/courses/sloan-school-of-management/15-912-technology-strategy-fall-2008/lecture-notes/lec_04.pdf.
- Diermeier, M. & Goecke, H. 2017. Productivity, Technology Diffusion and Digitization. CESifo Forum, Ifo Institute - Leibniz Institute for Economic Research at the University of Munich. [Artikkeli]. Vol. 18:1, S. 26-32. [Viitattu 3.5.2018]. Saatavissa: <http://www.cesifo-group.de/DocDL/CESifo-Forum-2017-1-diermeier-goecke-digitalization-march.pdf>.
- Dockalikova, I. & Klozikova, J. 2014. MCDM Methods in Practice: Determining the Significance of PESTEL Analysis Criteria, 11 2014, Academic Conferences International Limited. [Artikkeli]. S. 418-427. [Viitattu 8.2.2018]. Saatavissa: https://msed.vse.cz/msed_2014/article/362-Dockalikova-Iveta-paper.pdf.
- Ferraro, D. 2017. Volatility and slow technology diffusion. European Economic Review. [Artikkeli]. Vol. 96, S. 18-37. [Viitattu 15.3.2018]. Saatavissa: DOI: 10.1016/j.eurocorev.2017.04.005.
- Fertel, C., Bahn, O., Vaillancourt, K. & Waub, J. 2013. Canadian energy and climate policies: A SWOT analysis in search of federal/provincial coherence. Energy Policy. [Artikkeli]. Vol. 63, S. 1139-1150. [Viitattu 15.3.2018]. Saatavissa: DOI: 10.1016/j.enpol.2013.09.057.
- Fisher, J.C. & Pry, R.H. 1971. A simple substitution model of technological change. Technological Forecasting and Social Change. [Artikkeli]. Vol. 3, S. 75-88. [Viitattu 15.3.2018]. Saatavissa: DOI: 10.1016/S0040-1625(71)80005-7.
- Forsström, J. 2011. VTT:n tutkimus "kotitalousjätteen ennakointimalli", VTT-R-03229-11, VTT, Helsinki. [Verkojulkaisu 25.5.2018]. S. 1-44. [Viitattu 28.9.2018]. Saatavissa: <http://docplayer.fi/6677442-Vtt-r-03229-11-tutkimusraportti-kotitalousjätteen-ennakointimalli-luottamuksellisuus.html#>.
- Fourt, L.A. & Woodlock, J.W. 1960. Early prediction of market success for new grocery products. Journal of Marketing. [Artikkeli]. Vol. 25:2, S. 31-38. [Viitattu 22.8.2018]. Saatavissa: DOI: 10.2307/1248608.
- Gandomi, A. & Haider, M. 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management. [Artikkeli]. Vol. 35:2, S. 137-144. [Viitattu 3.5.2018]. Saatavissa: DOI: 10.1016/j.ijinfomgt.2014.10.007.
- Gartner 2017. Emerging-Technology-Hype-Cycle-for-2017. Gartner inc. [Verkojulkaisu]. [Viitattu 28.9.2018]. Saatavissa: https://blogs.gartner.com/smarterwithgartner/files/2017/08/Emerging-Technology-Hype-Cycle-for-2017_Infographic_R6A.jpg.

- Jackson, M. O., 2013. Social and Economic Networks: Models and Analysis, 5.2: Bass Model. Stanford University. [Opintomateriaali]. [Viitattu 15.11.2018]. Saatavissa: <https://www.coursera.org/lecture/social-economic-networks/5-2-bass-model-nMrkH>.
- Leite, J., Mello, L., Mello, J., Cruz, E. & Fontanillas, C. 2017. Using the MACBETH Method to improve the scenario analysis tool PESTEL in large civil construction projects. Universidad Nacional de Colombia. Dyna. [Artikkeli]. Vol. 84:203, S. 322-327. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.15446/dyna.v84n203.65359.
- Gleason, S. 2015. Workforce needs for next generation big data-analytics. ORAU 70th Annual Meeting Big Data Analytics March 4, 2015, Oak Ridge Tennessee USA , ed. S. Gleason. [Seminaarimateriaali]. S. 1-22. [Viitattu 28.9.2018]. Saatavissa: <https://www.slideshare.net/orau/workforce-needs-for-next-generation-big-data-analytics>.
- Godin, B. 2011. The linear model of innovation: Maurice Holland and the research cycle, Social Science Information. [Artikkeli]. Vol. 50:3-4, S. 569-581. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1177/0539018411411032.
- Godin, B. 2015. Models of Innovation: Why Models of Innovation are Models, or What Work is Being Done in Calling Them Models? Project on the Intellectual History of Innovation Working Paper No. 22. [Artikkeli]. S. 1-62. [Viitattu 8.10.2018]. Saatavissa: <http://www.csiic.ca/PDF/WorkingPaper22.pdf>.
- Godin, B. 2010. Innovation Without the Word: William F. Ogburn's Contribution to the Study of Technological Innovation. Project on the Intellectual History of Innovation Working Paper No. 5. [Artikkeli]. S. 1-56 s. [Viitattu 8.10.2018]. Saatavissa: <http://www.csiic.ca/PDF/IntellectualNo5.pdf>.
- Godin, B. 2008. In the Shadow of Schumpeter: W. Rupert Maclaurin and the Study of Technological Innovation. Project on the Intellectual History of Innovation Working Paper No. 2, [Artikkeli]. S. 1-29. [Viitattu 28.9.2018]. Saatavissa: <http://www.csiic.ca/PDF/IntellectualNo2.pdf>.
- Ordiano, J., Bartschat, A., Ludwig, N., Braun, E., Waczowicz, S., Renkamp, N., Peter, N., Döpmeier, C., Mikut, R., Hagenmeyer, V. 2018. Concept and benchmark results for Big Data energy forecasting based on Apache Spark. Journal of Big Data. [Artikkeli]. Vol. 5:11, S. 1-11. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1186/s40537-018-0119-6.
- Grinin, L., Grinin, A. & Korotayev, A. 2017. Forthcoming Kondratieff wave, Cybernetic Revolution, and global ageing. Technological Forecasting and Social Change. [Artikkeli]. Vol. 115, S. 52-68. [Viitattu 3.7.2018]. Saatavissa: DOI: 10.1016/j.techfore.2016.09.017.
- Gupta, B. 2015. Analytics India Magazine [Homepage of Analytics India Magazine]. [Verkojulkaisu]. [Viitattu: 5.11.2018]. Saatavissa: <https://analytic-sindiamag.com/analytics-and-data-science-india-jobs-study-2017-by-edvancer-aim/>.

- Hassani, H. & Silva, E.S. 2015. Forecasting with Big Data: A Review. *Annals of Data Science*. [Artikkeli]. Vol. 2:1, S. 5-19. [Viitattu 20.8.2018]. Saatavissa: DOI: 10.1007/s40745-015-0029-9.
- Huang, Y., Porter, A.L., Cunningham, S.W., Robinson, D.K.R., Liu, J. & Zhu, D. 2018. A technology delivery system for characterizing the supply side of technology emergence: Illustrated for Big Data & Analytics. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 130, S. 165-176. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1016/j.techfore.2017.09.012.
- Jamasb, T. 2007. Technical Change Theory and Learning Curves: Patterns of Progress in Electricity Generation Technologies. *The Energy Journal*. [Artikkeli]. Vol. 28:3, S. 51-72. [Viitattu: 8.10.2018]. Saatavissa: DOI: 10.5547/ISSN0195-6574-EJ-Vol28-No3-4.
- Jamasb, T. & Köhler, J. 2007. Learning curves for energy technology: a critical assessment. Cambridge University Press, University of Cambridge, [Verkkojulkaisu]. S. 1-21. CWPE 0752 & EPRG 0723. [Viitattu: 28.9.2018]. Saatavissa: <http://www.eprg.group.cam.ac.uk/wp-content/uploads/2008/11/eprg0723.pdf>.
- Jifa, G. & Lingling, Z. 2014. Data, DIKW, Big Data and Data Science, *Procedia Computer Science*. [Artikkeli]. Vol. 31, S. 814-821. [Viitattu 8.10.2018]. ISSN 1877-0509. Saatavissa: DOI: 10.1016/j.procs.2014.05.332.
- Jokela, D. 2016. Innovaatiotoiminnan resurssien johtaminen. Tampereen teknillinen yliopisto. [Artikkeli]. 12/2016. Tampere. 13 s.
- Kaminski, J. 2011. Diffusion of Innovation Theory *Canadian Journal of Nursing Informatics*. Theory in Nursing Informatics Column. [Artikkeli]. Vol. 6:2, S. 1-7. [Viitattu 8.10.2018]. Saatavissa: <http://cjni.net/journal/?p=1444>.
- Kanagal, N.B. 2015. Innovation and product innovation in marketing strategy, Indian Institute of Management. *Journal of Management and Marketing Research*. [Artikkeli]. Vol. 18, S. 1-25. [Viitattu: 28.9.2018]. Saatavissa: <http://www.aabri.com/manuscripts/152209.pdf>.
- Kanagal, N.B. 2017. Strategic Marketing Issues for an Emerging Market Firm to go Global. *International Business Research*. 10th edn, Indian Institute of Management Bangalore. [Artikkeli]. Vol. 10:6, S.35-45. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.5539/ibr.v10n6p35.
- Kaushik, D. 2015. 10 Reasons Why Big Data Analytics is the best career move. [Verkkojulkaisu]. [Viitattu 11.5.2018]. Saatavissa: <https://www.datasciencecentral.com/profiles/blogs/10-reasons-why-big-data-analytics-is-the-best-career-move>.
- Kiyak, D. & Pranckevičiūtė, L. 2016. Determining the relation between the business environment and companies solvency factors in the post - crisis period. *Ekonomika*. [Artikkeli]. Vol. 95:3, S. 64-80. [Viitattu 3.5.2018]. Saatavissa: DOI: 10.15388/Ekon.2016.3.10329.

- Kondratieff, N.D. & Stolper, W.F. 1935. The Long Waves in Economic Life. [Artikkeli]. Vol. 17:6, S. 105-115. [Viitattu 15.4.2018]. Saatavissa: <http://www.jstor.org/stable/1928486>.
- Korotayev, A., Zinkina, J. & Bogevolnov, J. 2011. Kondratieff waves in global invention activity (1900–2008). *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 78:7, S. 1280-1284. [Viitattu 14.3.2018]. Saatavissa: DOI: 10.1016/j.techfore.2011.02.011.
- Kuusi, O. 2018. Rahan merkitys ihmisten arjessa: näkymiä 2030-luvulle. [Seminaarimateriaali]. S. 16. [Viitattu 9.10.2018]. Saatavissa: http://www.tutuseura.fi/wp-content/uploads/2018/04/kesaesem2018rahan-merkitys_Kuusi.pdf.
- Kuznets, S. 1940. Schumpeter's Business Cycles. *American Economic Review*. [Artikkeli]. Vol. 30:2, S. 257. [Viitattu 8.10.2018]. Saatavissa: https://www.jstor.org/stable/1807049?seq=1#page_scan_tab_contents.
- Lee, W.S., Choi, H.S. & Sohn, S.Y. 2018. Forecasting new product diffusion using both patent citation and web search traffic. *PLoS ONE*. [Artikkeli]. Vol. 13:4, S. 1-12. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1371/journal.pone.0194723.
- Linden, A. & Fenn, J. 2003. Understanding Gartner's Hype Cycles. Strategic Analysis Report. Gartner Research. [Verkkojulkaisu]. S.1-12. [Viitattu 28.9.2018]. Saatavissa: <https://www.bus.umich.edu/KresgePublic/Journals/Gartner/research/115200/115274/115274.pdf>.
- Liu, Y., Tseng, F. & Tseng, Y. 2018. Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 130. [Viitattu 8.10.2018]. ISSN 0040-1625, S.123-134. Saatavissa: DOI: 10.1016/j.techfore.2018.01.018.
- Mahajan, V., Muller, E. & Bass, F. 1995. Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Science*. [Artikkeli]. Vol. 14:3, S. 79-88. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1287/mksc.14.3.G79.
- Mansfield, E. 1961. Technical change and the rate of imitation, *Econometrica*. [Artikkeli]. Vol. 29:4, S. 741-766. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.2307/1911817.
- Marinakis, Y. 2012. Forecasting technology diffusion with the Richards model. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 79, S.172-179. [Viitattu]. Saatavissa: DOI: 10.1016/j.techfore.2011.02.010.
- Martin, H. 2013. Big Data for Development: From Information- to Knowledge Societies. University of California. Davis. [Artikkeli]. S.1-39. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.2139/ssrn.2205145.

- Mathworks, I. 2016. Introducing Machine Learning, 92991v0 edn. S. 1-12, [Viitattu: 28.9.2018]. [Opintomateriaali]. Saatavissa: https://www.mathworks.com/content/dam/mathworks/tag-team/Objects/i/88174_92991v00_machine_learning_section1_ebook.pdf.
- Matsumoto, M. & Nishimura, T. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* [Artikkeli]. Vol. 8:1, S. 3-30, Saatavissa: DOI: 10.1145/272991.272995.
- Microsoft 2018. Office tukisivusto [Homepage of Microsoft]. [Verkköjulkaisu]. [Viitattu 8.5.2018], Saatavissa: <https://support.office.com/en-us/article/RAND-function-4CBFA695-8869-4788-8D90-021EA9F5BE73>.
- Naccarato, A., Falorsi, S., Loriga, S. & Pierini, A. 2018. Combining official and Google Trends data to forecast the Italian youth unemployment rate., *Technological Forecasting and Social Change*, [Artikkeli]. Vol. 130, S. 114-122, [Viitattu 8.10.2018]. ISSN 0040-1625. Saatavissa: DOI: 10.1016/j.techfore.2017.11.022.
- National Science Foundation (NSF) 2014. Innovation-Related Indicators of U.S. and Other Major Economies, Industry, Technology, and the Global Marketplace [Homepage of National Science Foundation]. [Verkköjulkaisu]. [Viitattu: 28.9.2018], Saatavissa: <http://www.nsf.gov/statistics/seind14/index.cfm/chapter-6/c6s4.htm>.
- Ostojic, I. 2010. Lecture with Computer Exercises: Modelling and Simulating Social Systems with MATLAB, Bass innovation diffusion model and its application in policy analysis for adoption of renewable energy technologies. [Opintomateriaali]. [Viitattu 8.10.2018]. Saatavissa: http://www.webarchiv.ethz.ch/soms/teaching/MatlabFall10/projects/BassInnovationDiffusionModelApplicationInPolicyAnalysis-ForAdoptionOfRETechnologies_Ostojic.pdf, Zürich.
- Park, S. & Kim, J. 2018. The effect of interest in renewable energy on US household electricity consumption: An analysis using Google Trends data. *Renewable Energy*. [Artikkeli]. S. 1004-1010. [Viitattu 8.10.2018]. ISSN 0960-1481. Saatavissa: DOI: 10.1016/j.renene.2018.05.044.
- Parker, P. 1994. Aggregate diffusion forecasting models in marketing: A critical review *Source International Journal of Forecasting*. [Artikkeli]. Vol. 10:2, S. 353-380. ISSN 0169-2070, [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1016/0169-2070(94)90013-2.
- Ratcliffe, S. 2016, *Oxford Essential Quotations*, Amara, Roy Quotation, eISBN: 9780191826719 [Artikkeli]. [Viitattu 8.10.2018]. Saatavissa: <http://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018679>, 4th edn, Oxford University Press, Oxford.
- Rogers, E. 2003. *Diffusion of Innovations*, Fifth Edition edn, 553 s. [Painettu julkaisu]. ISBN: 0743222091, Free Press, New York.

- Rogers, E. (ed) 1962. Diffusion of innovations, First Edition edn, Free Press of Glencoe, [Painettu julkaisu]. New York. 367 s.
- Roos, T. & Hagström, H. 2018. Elements of AI Mooc-kurssi. [Opintomateriaali]. [Viitattu 3.7.2018], Saatavissa: <https://www.elementsofai.com/>, Aalto yliopisto & Reaktor Innovations Oy, Helsinki.
- Rose, J., Jones, M. & Furneaux, B. 2015. An integrated model of innovation drivers for smaller software firms. *Information & Management*. [Artikkeli]. Vol. 53:3, S. 307-323. [Viitattu 8.10.2018]. ISSN 0378-7206. Saatavissa: DOI: 10.1016/j.im.2015.10.005.
- Roßmann, B., Canzaniello, A., von der Gracht, H. & Hartmann, E. 2018. The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 130, S. 135-149. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1016/j.techfore.2017.10.005.
- Sage Publications 2018, SAGE Research Methods Online (SRMO) [Homepage of Sage Publications]. [Verkkopalvelu]. [Viitattu 7.9.2018]. Saatavissa: <http://methods.sagepub.com/>.
- Sassali, J. 2014. Tiedonhankintakurssi 030005P, Luonnontieteellisten ja teknillisten tieteiden opiskelijoiden Tiedonhankintakurssi, [Opintomateriaali]. [Viitattu 3.9.2018]. Saatavissa: <https://wiki oulu.fi/pages/viewpage.action?pageId=28087528>.
- Schuelke-Leech, B. 2018. A model for understanding the orders of magnitude of disruptive technologies. *Technological Forecasting and Social Change*. [Artikkeli]. Vol. 129, S. 261-274, [Viitattu 8.10.2018]. ISSN 0040-1625. Saatavissa: DOI: 10.1016/j.techfore.2017.09.033.
- Sganzerla, C., Seixas Filho, C. & Conti, A. 2016. Disruptive Innovation in Digital Mining. *Procedia Engineering*. [Artikkeli]. Vol. 138, S. 64-71. [Viitattu 8.10.2018]. ISSN 1877-7058. Saatavissa: 10.1016/j.proeng.2016.02.057.
- Sims, G. 2015. What is machine learning? Android Authority. [Opintomateriaali]. [Viitattu 3.7.2018]. Saatavissa: <https://www.androidauthority.com/what-is-machine-learning-621659/>.
- Sundqvist, S., Frank, L., Puumalainen, K. & Kämäräinen, J. 2002. Forecasting the Critical Mass of Wireless Communications. ANZMAC 2002 Conference Proceedings., [Konferenssijulkaisu]. S. 551-557. [Viitattu 8.10.2018]. Saatavissa: https://www.researchgate.net/publication/266451687_Forecasting_the_Critical_Mass_of_Wireless_Communications.
- Tidd, J. & Bessant, J. (eds) 2013. *Managing innovation: Integrating Technological, Market and Organizational Change.*, 5th edn, John Wiley & Sons, Inc., Chichester. [Painettu julkaisu]. ISBN: 111836063X, 680 s.
- Trafimow, D. 2009. *The Theory of Reasoned Action.*, A Case Study of Falsification in Psychology. *Theory & Psychology - THEOR PSYCHOL*. [Verkkójulkaisu]. Vol.

19:4, S. 501-518, [Viitattu 8.10.2018]. Saatavissa: DOI: 10.1177/0959354309336319.

- Wang, B., Liu, Y., Zhou, Y. & Wen, Z. 2018. Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods. *Nano Energy*. [Artikkeli]. Vol. 46, S. 322-330. [Viitattu 8.10.2018]. ISSN 2211-2855. Saatavissa: DOI: 10.1016/j.nanoen.2018.02.020.
- Wang, Y., Pei, L. & Wang, Z. 2017. The NLS-based grey Bass model for simulating new product diffusion. *International Journal of Market Research*. [Artikkeli]. Vol. 59:5, S. 655-669. [Viitattu 8.10.2018]. Saatavissa: DOI: 10.2501/IJMR-2017-045.
- Wang, Z.-X & Dang, Y.-G & Pei, Lingling 2011. On Greying Bass Model and Its Application. *Journal of Grey System*. [Artikkeli]. Vol. 23:1, S. 7-14. [Viitattu 8.10.2018]. Saatavissa: <http://journals.sagepub.com/doi/abs/10.2501/IJMR-2017-045>.
- Wilenius, M. & Kurki, S. 2012. Surfing the Sixth Wave: Exploring the Next 40 years of Global Change, The 6th Wave and Systemic Innovation for Finland: Success Factor for the Years 2010-2050. [Verkköjulkaisu]. 124 S., [Viitattu 8.10.2018]. ISBN 978-952-249-148-0., Saatavissa: http://www.utu.fi/fi/yksikot/ffrc/julkaisut/e-tutu/Documents/eBook_2012-10.pdf.
- Xiong, H., Payne, D. & Kinsella, S. 2016. Peer effects in the diffusion of innovations: Theory and simulation. *Journal of Behavioral and Experimental Economics*. [Artikkeli]. Vol. 63, S. 1-13. [Viitattu 8.10.2018]. ISSN 2214-8043, Saatavissa: DOI: 10.1016/j.socec.2016.04.017.
- Xiong, H., Puqing, W. & Bobashev, G.V. 2018. Multiple Peer Effects in the Diffusion of Innovations on Social Networks: A Simulation Study. *Journal of Innovation and Entrepreneurship*. [Artikkeli]. [Viitattu 8.10.2018]. ISSN: 2192-5372. Saatavissa: DOI: 10.1186/s13731-018-0082-7.

LIITE A: TEKNOLOGIAN KEHITTÄMISEN ENNUSTUSMALLEJA

Tutkimuksissa (Cho ja Daim 2016; Marinakis 2012) toivat esille, että eri malli sopii erilaisen ilmiön ennustamiseen. Diplomityössä ei käsitellä kaikki malleja, mutta tuodaan esille muun muassa Fisher-Pryn, Gompertz ja Bassin mallit, jotka soveltuvat erilaisen tuotteiden, ratkaisuiden ja tilanteiden ennustamiseen.

Taulukko 40. *Mukaelma teknologian ennustamiseen soveltuvista malleista (Cho & Daim 2016).*

Mallin nimi	Ennustusmalli	Esittäjä
Logistic or Pearl	$Y = \frac{L}{1 + ae^{-bt}}$	Robertson (1923), Griliches (1957)
Lotka–Volterra	$\frac{dM}{dt} = (a_m - b_m M + c_{mn} N)M, \frac{dN}{dt} = (a_n - b_n N + c_{nm} M)N$	Lotka (1925), Volterra (1931), Farrel (1993)
Gompertz	$Y = Le^{-b-kt}$	Winsor (1932)
Mansfield-Blackman	$\ln\left(\frac{Y_t}{L-Y_t}\right) = \beta_0 + \beta_1 t$	Mansfield (1961), Blackman (1972)
BASS	$Y_t = \frac{[1 - e^{-(p+q)t}]}{[1 + e^{(q/p)}]e^{-(p+q)t}}$	Bass (1969)
Fisher-Pry	$Y_t/(1 - Y_t) = e^{2a(t-t_0)}$	Fisher & Pry (1971)
Extended Riccati	$\frac{Y_t}{Y_{t-1}} = \beta_0 + \beta_1 Y_{t-1} + \beta_2 \left(\frac{1}{Y_{t-1}}\right) + \beta_3 \ln(Y_{t-1})$	Levenbach & Reuter (1976)
Weibull	$\ln\left(\ln\left[\frac{L_t}{L - Y_t}\right]\right) = \beta_0 + \beta_1 \ln t$	Sharif & Islam (1980)
NSRL	$\ln Y_t = \beta_0 + \beta_1 \ln(Y_{t-1}) + \beta_2 \ln(L - Y_{t-1})$	Easingwood & al. (1981)
Harvey	$\ln Y_t = \beta_0 + \beta_1 t + \beta_2 \ln(Y_{t-1})$	Harvey (1984)

LIITE B: OTOS GOOGLE TRENDS -PALVELUSTA

Liitteessä B on diplomityössä toteutettu otos Google Trends -palvelusta. Otos perustuu poimintaan ja poiminnan käsittelyyn, esittämiseen ja laskentaa. Poimintapäivämäärä Google Trends palvelusta 20.7.2018.

Otoksen perustiedot ja poimintaehdot:

Palvelu: Google Trends

Url: <https://trends.google.com/trends/?geo=FI>

Selain: Mozilla Firefox Quantum 60

Otoksen ajankohta: 20.7.2018

Asetetut hakuparametrit: Machine learning, Deep learning, Big data, Data science, Data Analytics

Hakualue: Koko maailma

Aikarajaus: 01.08.2008 - nykyhetki

Aiherajaus: Kaikki luokat

Hakuluokka: Verkkohaku

Tietojen tuonti: Export – MultiTimeLine.csv

Tietojen muokkaus: Notepad (tabulointien lisäys ja rakenteen korjaus) ja käsittely Excel-2017 ohjelmistolla.

Taulukko 41. *Otos Google Trends-palvelusta*

Vuosi	Kuukausi	Machine learning	Deep learning	Big data	Data science	Data analytics
2008	08	8	1	5	5	2
2008	09	10	1	5	6	2
2008	10	11	1	4	5	2
2008	11	10	1	4	6	3
2008	12	9	1	3	4	2
2009	01	9	1	4	7	2
2009	02	11	1	4	7	2
2009	03	10	1	4	6	2
2009	04	11	1	3	6	3
2009	05	10	1	4	6	2
2009	06	10	1	4	5	2
2009	07	9	1	4	4	3
2009	08	10	1	4	5	3
2009	09	11	1	5	8	3
2009	10	11	1	5	7	3
2009	11	11	1	4	6	3
2009	12	10	1	4	6	3
2010	01	9	1	4	7	3
2010	02	11	1	4	7	3
2010	03	10	1	5	8	4
2010	04	11	1	5	6	3
2010	05	10	1	4	5	3
2010	06	10	1	5	5	3
2010	07	9	1	4	5	4
2010	08	9	1	5	5	3
2010	09	11	1	4	7	3
2010	10	11	1	5	6	4
2010	11	11	1	4	6	3
2010	12	10	1	5	5	3
2011	01	10	1	6	7	3
2011	02	11	1	5	6	3
2011	03	10	1	5	6	4
2011	04	11	1	6	6	4
2011	05	10	1	7	5	4
2011	06	10	1	7	5	4
2011	07	9	1	7	4	4
2011	08	9	1	7	6	4
2011	09	12	1	9	8	5
2011	10	15	1	10	7	5
2011	11	14	1	11	7	5
2011	12	13	1	10	6	4
2012	01	13	1	12	7	6
2012	02	13	1	16	7	6
2012	03	13	1	17	6	6
2012	04	13	1	20	6	6
2012	05	14	1	21	7	7
2012	06	11	1	20	6	6

2012	07	12	1	21	5	6
2012	08	12	1	22	6	7
2012	09	14	1	26	9	7
2012	10	14	2	31	8	9
2012	11	13	2	31	8	8
2012	12	12	2	28	6	7
2013	01	13	2	33	8	9
2013	02	12	2	36	7	9
2013	03	14	2	43	7	10
2013	04	15	2	45	9	10
2013	05	15	3	46	9	11
2013	06	14	2	45	8	11
2013	07	14	2	45	8	11
2013	08	12	2	44	9	11
2013	09	16	3	47	12	12
2013	10	18	3	52	10	12
2013	11	17	3	52	11	13
2013	12	16	3	44	10	11
2014	01	16	3	53	13	13
2014	02	18	4	57	12	14
2014	03	20	4	61	13	15
2014	04	20	4	60	13	15
2014	05	19	4	60	12	14
2014	06	17	4	57	12	13
2014	07	18	4	62	13	15
2014	08	19	4	59	16	16
2014	09	21	5	70	19	16
2014	10	23	5	69	16	17
2014	11	23	5	69	16	17
2014	12	22	6	60	16	15
2015	01	23	7	66	19	17
2015	02	25	8	71	17	19
2015	03	25	8	76	19	20
2015	04	28	7	73	20	20
2015	05	26	8	64	19	19
2015	06	28	9	69	20	21
2015	07	28	9	63	19	20
2015	08	27	9	62	23	22
2015	09	31	9	73	26	23
2015	10	32	10	72	24	22
2015	11	35	11	72	24	21
2015	12	33	11	60	22	19
2016	01	32	14	64	26	21
2016	02	34	14	74	26	24
2016	03	39	18	73	28	24
2016	04	39	18	72	27	25
2016	05	41	17	72	27	26
2016	06	42	18	71	27	26
2016	07	38	16	64	26	24
2016	08	40	18	63	29	24
2016	09	47	19	73	35	27
2016	10	52	21	72	33	28
2016	11	55	23	77	34	29
2016	12	52	24	64	35	26
2017	01	61	27	72	39	30
2017	02	63	28	79	41	33
2017	03	70	30	80	43	33
2017	04	71	31	76	43	32
2017	05	77	33	74	44	33
2017	06	76	31	70	42	33
2017	07	75	31	67	42	33
2017	08	76	33	66	47	34
2017	09	82	35	74	51	35
2017	10	88	37	78	53	36
2017	11	92	38	80	53	37
2017	12	85	37	63	49	30
2018	01	90	38	66	57	35
2018	02	93	39	73	58	38
2018	03	97	38	76	56	38
2018	04	94	37	74	58	38
2018	05	100	39	70	59	39
2018	06	95	37	64	57	34
2018	07	97	36	63	58	37

LIITE C: SOVELLUS KRIITTISEN MASSAPISTEEN ESITTÄMISTÄ VARTEN (P JA Q -PARAMETREILLA)

Laskentakaavan alkuperäinen lähde: Sundqvist, Frank, Puumalainen ja Kämäräinen (2002), joka perustuu Bassin (1969) esittämään teoriaan. Kuvassa 12 esitetty laskenta Matlab 2017b -ohjelmistolla, jonka algoritmi on esitetty kokonaisuudessa liitteessä C.

```
clear variables, close all;
iteraatiot=500;
laskentavali=0.1;
q=0.15;% q-parametri (kuva 12 Bass laskentamallin q-parametri)
m=1500000;
A(1)=0;
p=0.011218; % p-parametri (kuva 12 Bass laskentamallin p-parametri)
t(1)=0;
CMP=(log((2+sqrt(3))*p/q)/(-(p+q))); %kriittisen massapisteen laskenta
T = 1/(p+q)*log(q/p) %kriittinen massapiste on esitetty olevan toisen derivaatan nollakohdassa.

for i=2:iteraatiot
    s(i)=(p +q*A(i-1)/m)*(m-A(i-1))*laskentavali;
    A(i)=s(i)+A(i-1);
    t(i)=t(i-1)+laskentavali;
end

plot(t,A)
hold on
plot(CMP,265000,'O') %265 000 on arvio käyrällä, jossa laskettu kriittinen massa on saavutettu (tämän laskenta on rajattu pois laskenasta).
hold on
disp(['kriittinen piste:', num2str(CMP)])
```

LIITE D: SOVELLUS KRIITTISEN MASSAPISTEEN LASKENTAA VARTEN, JOSSA RATKAISTAAN P JA Q -PARAMETRIT

Kriittisen massapisteen osalta laskentakaavan alkuperäinen lähde: Sundqvist, Frank, Puumalainen ja Kämäräinen (2002), joka perustuu Bassin esittämään teoriaan ja laskentaan. Muilta osin perustuu (mm. epälineaarisen lsqcurvefit, lsqnonlin ja muiden funktioiden osalta) osittain lähteeseen <https://www.mathworks.com/matlabcentral/answers/410081-errors-in-objective-function>, jossa on useampia kirjoittajia. Osa kirjoittajista toimii nimimerkillä, mutta heidät ja heidän vastauksensa on vertaisarvioitu Matlab answers -palvelussa

Laskentasovelluksessa asetetaan kuvitteellinen data ja huolehditaan, että data on ajan funktiona reilusti ennen laskennallista massapistettä. Massapisteen Y-akselin arvoa ei ole arvioitu tässä Matlab-koodissa. Laskentatulokset esitetään arvona X-akselilla.

Ohjelmisto: Matlab 2017b

```
%Perustuu lähteeseen https://www.mathworks.com/matlabcentral/answers/410081-errors-in-objective-function
clear variables, close all;
data=[0;0;0;0;1;1;1;1;2;2;3;3;2;4;4;4;3;3;4;3;6;4;10;13;15;20;30;45;67;100];

t=1:30 %1991:2020;
v=1000;
x = outer_function(v, data);
disp(x)

%lsqcurvefit
ydata=data'
v=1000;
fun=@(x,t)Bassmodel(x,t,v);
lb=[0,0,0,0,0];
ub=[1, Inf, 1, 1, Inf];
x0=[20,1,1,1,300];
options = optimoptions(@lsqcurvefit,'Display','off');
x = lsqcurvefit(fun,x0,t,ydata,lb,ub,options);
y_val = Bassmodel(x,t,v);

% lsqnonlin
x0=x;

fun = @(x)ydata-(x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v)))*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v)))*t));
options2 = optimoptions(@lsqnonlin,'Display','off');
x2 = lsqnonlin(fun,x0,lb,ub,options2);
y_val2 = Bassmodel(x2,t,v);

%Kriittinen massapiste
p=x2(1);
q=x2(3);
%vertaillaan Sundqvist et al. (2002) ja toisen derivaatan välistä eroa
CMP=(log((2+sqrt(3))*p/q)/(-(p+q)));
T = 1/(p+q)*log(q/p)
disp(['kriittinen piste:', num2str(CMP)])

% plot results
plot(t,ydata,'+')

% hold on
% plot(t,y_val,'.')%Solve nonlinear curve-fitting (data-fitting) problems in least-squares sense
```

```

% hold on
% plot(t,y_val2,'--') %Solve nonlinear least-squares (nonlinear data-fitting) problems

% hold on
% plot(CMP,2,'O') %Kriittinen massapiste

% hold on
% legend('Havaintodata','Bass-funktio: epälineaarinen käyrä-sovitus pienimmällä neliösum-malla','Bass-funktio: epälineaarinen pienimmän neliösumman data-sovitus','Kriittinen massapiste','Location','northwest')
% hold off

%show results
disp(['p: ', num2str(x2(1))])
disp(['m: ', num2str(x2(2))])
disp(['q(m): ', num2str(x2(3))])
disp(['q_0: ', num2str(x2(4))])
disp(['w: ', num2str(x2(5))])

%objective function for lsqcurvefit
function [F] = Bassmodel(x,t,v)

F=x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-
x(4))*exp(-x(5)*v))).*t));
end

function j = outer_function(v, data)

fun=@Bass;
nl=@nonlinc;
lb=[20,0,0,0,0];
ub=[inf,1,1,1,inf];
x0=[ 0,0,0,0,0];
j=fmincon(fun,x0,[],[],[],[],lb,ub,nl);

function [F] = Bassmodel(x,t,v)
p = x(1);
m = x(2);
q_m = x(3);
q=q_m;
q_0 = x(4);
w = x(5);

F = -(p.*m.*(exp(-t.*(p+(q_0.*q_m)./(q_0-exp(-v.*w).*(q_0-q_m)))))-
1.0))./(p+(q_0.*q_m.*exp(-t.*(p+(q_0.*q_m)./(q_0-exp(-v.*w).*(q_0-q_m))))))./(q_0-exp(-
v.*w).*(q_0-q_m)));

end

function res= Bass(x)
%objective function
F= (- data + x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)))/
( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)));
res=sum(F.^2);
end

function [c, ceq]= nonlinc(x)
%constraint
c=[];
ceq=- data + ( x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-
x(5)*v)))/ ( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)));
end
end

```


LIITE E: CHO JA DAIM (2016) TUTKIMUSTULOKSIEN MALLINTAMINEN MATLAB-OHJELMISTOLLA HYÖDYNTÄEN BASS (1969) LASKENTAMALLIA

Liitteessä E esitetään diplomityössä simuloitu Bass -funktion mukainen laskenta. Itermuuttuja kuvaa summien lukumääriä. Laskentaalgoritmi sovittaa iter-muuttujan vastamaan Chon ja Damin (2016) esittämän kuvan vuosille sopivaksi hyödyntämällä kuvissa esitettyjä alku- ja loppuvuosia.

Ohjelmisto: Matlab 2017b

```
clear variables;
close all;

P=100; % saturaatiotaso
A1(1)=0;
p1=0.0012; %p-parametri (innovaattorit) käyrän yksi osalta
p2=0.0012; %p-parametri (innovaattorit) käyrän kaksi osalta
q1=0.4; % q-parametri (seuraajat)
q2=0.4;% q-parametri (seuraajat)

alkuvuosi1(1)=1996; % lähtövuosi käyrälle 1
alkuvuosi2(1)=2002; %lähtövuosi käyrälle 2
loppuvuosi1= 2025 %loppuvuosi
time_step=0.01
iter = (loppuvuosi1(1)-alkuvuosi1)/time_step
for i1=2:iter
    s(i1)=(p1 +q1*A1(i1-1)/P)*(P-A1(i1-1))*time_step;
    A1(i1)=s(i1)+A1(i1-1);
    alkuvuosi1(i1)=alkuvuosi1(i1-1)+time_step;
end

%käyrä kahden laskenta
A2(1)=0;

for i2=2:iter
    s2(i2)=(p2 +q2*A2(i2-1)/P)*(P-A2(i2-1))*time_step;
    A2(i2)=s2(i2)+A2(i2-1);
    alkuvuosi2(i2)=alkuvuosi2(i2-1)+time_step;
end
% plot(t,P-A,'r')
% hold on
figure(1)
plot(alkuvuosi1,A1,'--',alkuvuosi2,A2) %plot(x,y1,x,y2,'--',x,y3,':')
hold on
legend('perustutkimus','patentit','Location','northwest','Orientation','vertical')
```

LIITE F: MADDISON PROJEKTISTA LASKETUT INFLAATIO-KORJAUKSET JA SUHTEELLISEN VAIHTELUN LASKENTATULOKSET

Bolt, J., Inklaar, R., de Jong, H. & van Zanden, J.L. 2018. Maddison Project Database, version 2018. Rebasing ‘Maddison’: new income comparisons and the shape of long-run economic development. Maddison Project Working Paper, nr. 10. [Viitattu 23.8.2018]. Saatavissa: www.ggd.net/maddison.

Otoksen perustiedot ja poimintaehdot:

Palvelu: Maddison Project Database

Url: <https://www.rug.nl/ggd/historicaldevelopment/maddison/releases/maddison-project-database-2018>

Selain: Mozilla Firefox Quantum 60

Otoksen ajankohta: 23.8.2018

Tietojen muokkaus: Notepad (tabulointien lisäys ja rakenteen korjaus) ja käsittely Excel-2017 ohjelmistolla.

Maddison projektissa esitetyn ohjeen mukaiset viitetiedot aineistolle:

Full data: Data in single table

countrycode 3-letter ISO country code

country Country name

year: Year

cgdppc: Real GDP per capita in 2011US\$, multiple benchmarks (suitable for cross-country income comparisons)

rgdnpac: Real GDP per capita in 2011US\$, 2011 benchmark (suitable for cross-country growth comparisons)

pop: Population, mid-year (thousands)

i_cig 0/1/2: observation is extrapolated (0), benchmark (1), or interpolated (2)

i_bm For benchmark observations: 1: ICP PPP estimates, 2: Historical income benchmarks, 3: Real wages and urbanization, 4: Multiple of subsistence, 5: Braithwaite (1968) PPPs

Partial countries: Data for selected sub-national units with long time series

Note: real GDP per capita figures are rounded to the nearest dollar, population figures are rounded to the nearest 1000.

Taulukko 42. Otoksesta muodostetut tulokset Excel-ohjelmiston käsittelyn jälkeen.

vuosi	Rgdnpac/\$	Inflaatio/\$	Erotus/\$	vaihtelu
1900	155675	155675	0	1
1901	143042	159738,1	-16696,1	0,895478
1902	148223	163907,3	-15684,3	0,90431
1903	151717	168185,3	-16468,3	0,902083
1904	153414	172574,9	-19160,9	0,888971
1905	158547	177079,1	-18532,1	0,895346
1906	166076	181700,9	-15624,9	0,914008
1907	167490	186443,3	-18953,3	0,898343
1908	165872	191309,4	-25437,4	0,867035
1909	169981	196302,6	-26321,6	0,865913
1910	189393	201426,1	-12033,1	0,94026
1911	179661	206683,3	-27022,3	0,869257
1912	179697	212077,8	-32380,8	0,847317
1913	238209	217613	20596,01	1,094645
1914	172630	223292,7	-50662,7	0,773111
1915	169715	229120,6	-59405,6	0,740723
1916	176048	235100,7	-59052,7	0,74882
1917	170643	241236,8	-70593,8	0,707367
1918	168532	247533,1	-79001,1	0,680846
1919	168839	253993,7	-85154,7	0,664737
1920	194921	260622,9	-65701,9	0,747904
1921	194992	267425,2	-72433,2	0,729146
1922	204421	274405	-69984	0,744961
1923	211960	281567	-69607	0,752787
1924	229523	288915,9	-59392,9	0,794429
1925	239097	296456,6	-57359,6	0,806516
1926	241850	304194,1	-62344,1	0,795052
1927	245872	312133,5	-66261,5	0,787714
1928	257035	320280,2	-63245,2	0,802532
1929	273963	328639,5	-54676,5	0,833628
1930	262127	337217	-75090	0,777324
1931	247488	346018,4	-98530,4	0,715245
1932	234798	355049,5	-120251	0,661311
1933	242887	364316,3	-121429	0,666693
1934	249118	373824,9	-124707	0,666403
1935	257414	383581,8	-126168	0,67108
1936	267488	393593,2	-126105	0,679605
1937	278984	403866	-124882	0,690784
1938	280098	414406,9	-134309	0,675901
1939	285282	425222,9	-139941	0,6709
1940	272377	436321,3	-163944	0,624258
1941	264952	447709,2	-182757	0,591795
1942	265312	459394,5	-194082	0,577525
1943	257898	471384,7	-213487	0,547107
1944	253758	483687,8	-229930	0,524632
1945	244255	496312	-252057	0,49214
1946	267910	509265,8	-241356	0,526071
1947	283051	522557,6	-239507	0,541665
1948	300057	536196,4	-236139	0,559603
1949	308266	550191,1	-241925	0,560289
1950	527731	564551,1	-36820,1	0,93478
1951	544430	579285,9	-34855,9	0,93983
1952	574392	594405,2	-20013,2	0,966331
1953	595792	609919,2	-14127,2	0,976838
1954	617822	625838,1	-8016,11	0,987191
1955	640465	642172,5	-1707,48	0,997341
1956	658661	658933,2	-272,186	0,999587

1957	677772	676131,3	1640,658	1,002427
1958	684281	693778,4	-9497,37	0,986311
1959	707167	711886	-4718,99	0,993371
1960	746073	730466,2	15606,79	1,021366
1961	774487	749531,4	24955,62	1,033295
1962	801631	769094,1	32536,85	1,042305
1963	831294	789167,5	42126,5	1,053381
1964	875916	809764,8	66151,22	1,081692
1965	909400	830899,6	78500,36	1,094476
1966	941737	852586,1	89150,88	1,104565
1967	966391	874838,6	91552,39	1,104651
1968	1014470	897671,9	116798,1	1,130112
1969	1064948	921101,1	143846,9	1,156168
1970	1157567	945141,9	212425,1	1,224755
1971	1197265	969810,1	227454,9	1,234536
1972	1246623	995122,1	251500,9	1,252734
1973	1419045	1021095	397950,2	1,389729
1974	1901608	1047745	853862,6	1,814952
1975	1864631	1075092	789539,5	1,734393
1976	1928985	1103151	825833,6	1,748613
1977	1889966	1131944	758022,3	1,669664
1978	1927671	1161487	766183,6	1,659657
1979	1966523	1191802	774720,8	1,650041
1980	2082323	1222908	859414,7	1,702763
1981	1996565	1254826	741738,8	1,591109
1982	1921927	1287577	634349,9	1,492669
1983	1879344	1321183	558161,1	1,422471
1984	1895783	1355666	540117,2	1,398415
1985	1889414	1391049	498365,3	1,358266
1986	1885993	1427355	458638	1,32132
1987	1901882	1464609	437273	1,29856

1988	1927046	1502835	424210,7	1,282274
1989	1949221	1542059	407161,7	1,264038
1990	1934741	1582307	352434	1,222734
1991	1890407	1623605	266801,8	1,164327
1992	1898294	1665981	232312,7	1,139445
1993	2018449	1709463	308985,5	1,18075
1994	2057080	1754080	302999,5	1,17274
1995	2103009	1799862	303147	1,168428
1996	2154156	1846838	307317,6	1,166402
1997	2246675	1895041	351634,2	1,185555
1998	2288928	1944501	344426,6	1,177128
1999	2321165	1995253	325912,1	1,163344
2000	2408120	2047329	360791	1,176225
2001	2440066	2100764	339301,7	1,161513
2002	2478100	2155594	322505,8	1,149613
2003	2537578	2211855	325722,8	1,147262
2004	2647858	2269585	378273,3	1,166671
2005	2714418	2328821	385597,2	1,165576
2006	2818167	2389603	428564	1,179345
2007	2912701	2451972	460729,3	1,187902
2008	2944782	2515968	428813,9	1,170437
2009	2836153	2581635	254518,1	1,098588
2010	2897766	2649016	248750,4	1,093903
2011	2939936	2718155	221781,1	1,081593
2012	2986475	2789099	197376,3	1,070767
2013	3006988	2861894	145093,8	1,050699
2014	3048227	2936590	111637,4	1,038016
2015	3084382	3013235	71147,37	1,023612
2016	3092871	3091880	990,9492	1,000321

LIITE G: LASKENTA, JOLLA LASKETAAN PATENTTITietoKANNAN OTOKSEN POIMINTAVIRHEIDEN ESIINTYVYYDEN TODENNÄKÖISYYSJAKAUMA

Patenttitietokannan otoksessa aineistoon kuulumisen osalta on tehty rajausta, että se joko kuuluu tai ei kuulu tutkittavaan teknologiaan, jolloin voidaan hyödyntää binomijakaumaa. Tällöin tulkinnassa ei voi olla yhtäaikaaisesti voimassa molemmat tilat, niin tapahtumat riippuvat toisistaan ja niiden tulos on toisensa poissulkeva eli todennäköisyys kuuluvuudelle.

Ohjelmisto: Matlab 2017b

```
% LIITE G: PATENTTITIE TOKANNAN OTOKSEN POIMINTAVIRHEIDEN ESIINTYVYYDEN TODENNÄKÖISYYSJAKAUMA
```

```
clear variables;  
close all;
```

```
F=[0.02:0.0001:0.05]  
N=[100:1.0:150]  
P=binopdf(129,3735,F)  
figure(1)  
plot(F, P)
```

*) Tällöin ei kuitenkaan huomioida tilannetta, jossa otoksen joukko pienenee, jos virheetön tai virheellinen havainto poistetaan otoksesta.

LIITE H: PATENTTITOKANNAN OTOKSEN POIMINTAVIRHEIDEN ESIINTYVYYDEN TODENNÄKÖISYYDEN KERTYMÄFUNKTIO

Liitteessä G ollutta laskentaa hyödynnetään liitteen H laskennassa. Todennäköisyysjakauma integroidaan. Liitteessä H on integroitu liitteen G virheiden esiintyvyyden todennäköisyysjakauma $\int F(P(\text{virhe}))$. Tuloksena on saatu virhetodennäköisyyden kertymäfunktio.

Kertymäfunktion avulla voidaan esittää virhetodennäköisyyden luottamustason muuttamista virhetodennäköisyyden oletuksen funktiona. Hyväksymällä lähtöaineiston muut virheet ja 0,05 (5%) virhetaso tulosta voidaan pitää teknisesti yli 0,99 (99%) luotettavana. Virheiden määrä otoksen kokoisessa populaatiossa on binomijakauman mukaan alle 155 kappaletta otoksesta yli 99% luottamustasolla.

Ohjelmisto: Matlab 2017b

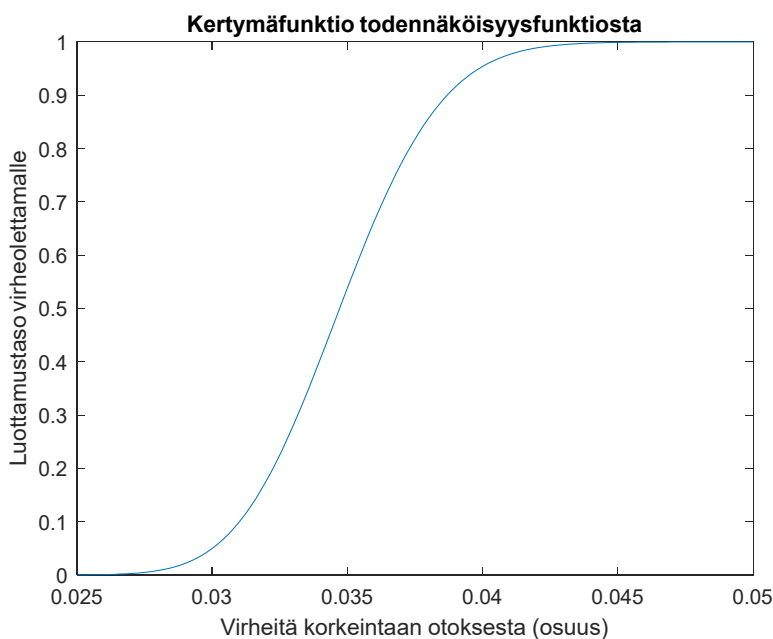
```
clear variables;
close all;

F=[0.025:0.0002:0.05] %virheen funktio, joka saa arvoja 0.025 ja 0.05 välillä

%todennäköisyyden
P_tot=1-binocdf(129,3735,F)

% maksimäärä otoksen perusteella saadulla virnehypoteesilla, joka saavutetaan 99% luottamustasolla
maara=binoinv(0.99,3725,0.0345)

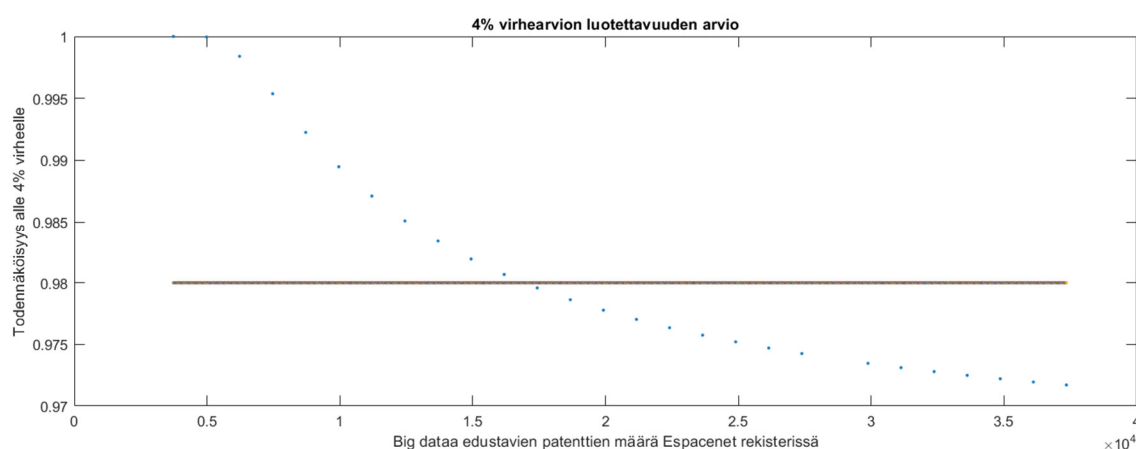
figure(1)
plot(F,P_tot)
```



Kuva 72. Laskentatulokset eli patenttitietokannan otoksen poimintavirheiden esiintyvyyden todennäköisyyden kertymäfunktio.

LIITE I: HYPERGEOMETRISEN JAKAUMAN LASKENTA 4% VIRHEARVIOLLE

Hypergeometrisen jakaumalla voidaan arvioida koko populaation (big data -patenttien) lukumäärän vaikutus virheellisten patenttien määrään. Tässä tapauksessa halutaan tarkastella äärellisen populaation ongelmia. Otoksen jokainen tapaus kuuluu tai ei kuulu tarkasteltavaan joukkoon. Olettama on, että jokainen samankokoinen otos tulee valituksi yhtä suurella todennäköisyydellä. Laskenta diplomityössä on tehty hypoteesilla, että patenttietokannasta toteutettu poiminta poimii 10-100% patenteista, joissa on big data edustava teknologiana. Kuvassa 60 esitetään koko patenttikannan big dataa edustava populaatio x-akselilla, joka saa arvoja $N=[3735...37350]$ välillä. Tuloksen, jossa esitetään 4% määrä virheellisiä tuloksia on asetettu 0,98 luottamustaso.



Kuva 73. Koko populaation (N) funktiona, jossa esitetään populaation koon vaikutus tulosten luotettavuuteen kyseisessä otoksessa, jossa luottamustaso on asetettu 0,98:ään ja testataan 4% virheellisten määrän laskenta (liite I).

```
clear variables
close all
N= [3735:1:37350], n = 3735;
p_def = (129/3735)
k = p_def*N;
P_hyge = hygecdf(0.04*n, N, k, n);
plot(N, P_hyge, '.', N, 0.98, '.')
```

LIITE J: EXCEL -OHJELMISTON HYÖDYNTÄMINEN JA TYÖVAIHEET PATENTTIETOKANNAN TIETOJEN KÄSITTELYSSÄ

- Aineisto kerääminen Espacenet-tietokannasta käyttämällä Export-toimintoa. Espacenet-tietokannasta saa kerättyä poimintatuloksia 500 riviä kerrallaan, jolloin aineiston kerääminen on kohtuullisen nopea toimenpide.
- Aineiston jalostaminen, jossa lisättiin aineistoon uusia kenttiä. Uusia tarvittavia tietoja olivat IPC-koodien lukumäärä, CPC-koodien lukumäärä, vuosi, kuukausi, sekä muokkaus laskentakelpoiseksi (esim. päivämäärämuotojen muotoilu).
- Aineistojen suodattaminen eri ehdoilla: Excel suodatukset eri ehdoilla
- Lukumääräinen laskenta kuukausittain, vuosittain eri ehdoilla: Pivot taulukko
- Aineiston muotoilu Matlab-lähdeaineistoksi, Excel export

Aineiston solujen tekstistä laskettiin eri asioita seuraavasti:

- IPC-koodien lukumäärä laskettiin tunnistamalla soluista uniikki merkki '/', joka toistui vain kerran IPC-koodissa. Laskemalla solun merkit vertaamalla solun merkkien määrää siihen määrään, joka olisi ilman näitä merkkejä saadaan IPC-koodien lukumäärä kyseisessä solussa. Excel-funktiolla "=PITUUS(L2)-PITUUS(VAIHDA(L2;"/";""))" voitiin toteuttaa kyseinen laskentatapa IPC- ja CPC- koodien osalta. Laskentatuloksen oikeellisuus todennettiin manuaalisella tarkastamisella (n=100), että laskenta toimii virheettömästi.
- Kuukausi, pvm ja muiden kenttien muodostaminen toteutettiin muuttamalla Excelin toiminnoilla päivämääräkentät tekstimuotoisiksi ja hyödyntämällä Poimitexti toimintoa esim. "=POIMI.TEKSTI(I4;1;4)" toiminnolla poimittiin sarakkeen 4 ensimmäistä merkkiä, jotka edustivat ("VVVV") vuositietoa päivämääräkentästä.
- Kuukausi, vuosi, päivämäärätietojen yhdistäminen ja muotoilu toteutettiin käyttämällä konkanaatiotoimintoja uusissa kentissä esim. vuosi- ja kuukausikenttä yhdistettiin "=B2&C2" -koodilla
- Patenttien laskenta kuukausittain laskettiin käyttämällä valitsemalla laskettava alue ja hyödyntämällä Pivot toimintoa ja valitsemalla laskettavan kentän ominaisuuksista laskettava asia. Esimerkiksi patenttien keskimääräinen käsittelyaika valmisteltiin Pivot -taulukon valinnoilla
 - Pivot: Onko big data -patentti: on (suodattimet)
 - Pivot: Rivit: Vuosi ja kk
 - Pivot: Arvot: Summa, hakemusten määrä (julkaistut)
 - Pivot: Arvot: Määrä, hakemusten määrä (julkaistut)

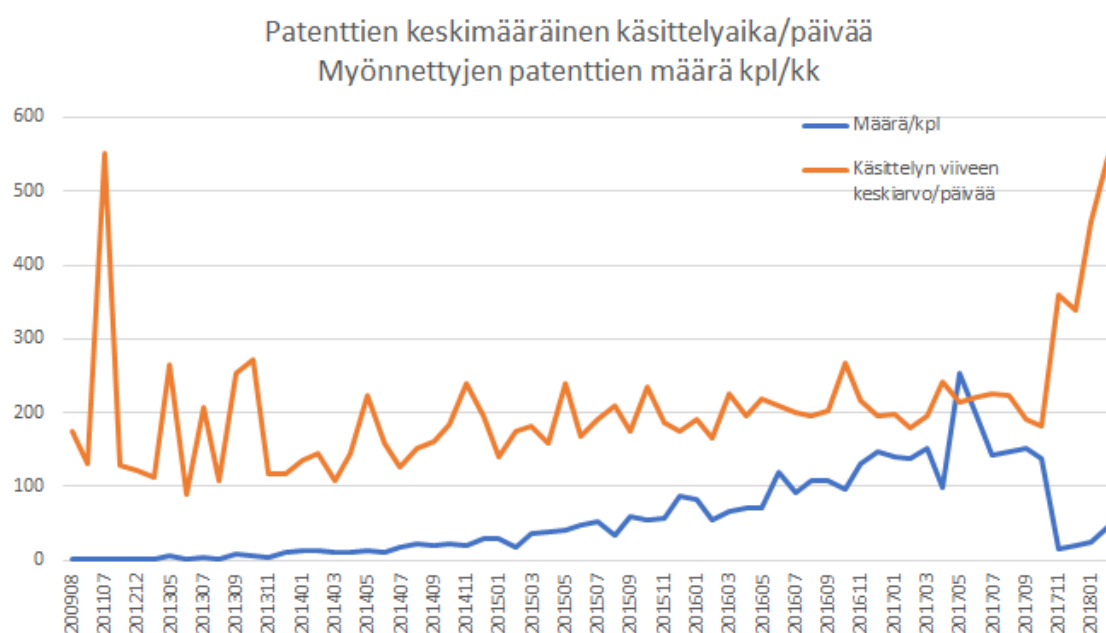
Jolla saatiin lähtötiedot kuukausittain julkaistuista patenteista ja niiden käsittelyn kestoajasta

LIITE K: PATENTTIEN KÄSITTELYAIKOJEN KESKIMÄÄRÄINEN VIIVE AJAN FUNKTIONA

Patenttitietokannan otoksen tarkastelu käsittelyaika ja julkaisuviiveen osalta. Laskenta on toteutettu poimimalla tietokannasta patentin hakupäivä ja julkaisupäivä. Tässä laskennassa ei huomioida, mistä viive koostuu.

Keskimääräisen käsittelyajan tai käsittelyajan mediaanin käyttö tunnuslukuna antaa kuvan kuluneesta ajasta. Keskimääräinen käsittelyaika ja käsittelyajan mediaani eivät ole hyviä tunnuslukuja. Tarkempi tutkiminen osoittaa käsittelyajan koostuvan vähintään kahdesta eri ryhmästä. Kahden eri ryhmän käsittely omina joukkoinaan parantaa käsitystä patenttien käsittelyyn kuluneesta ajasta.

Kuvan 61 aineistoa ei ole oikaistu uudemmalla otannalla patenttitietokannasta. Diplomi-työssä toteutettiin kaksi uudempaa otosta samasta populaatiossa. Toistamalla otanta myöhemmin nousi patenttien määrä 9/2017 jälkeen noin 250:een/kk ja keskimääräinen käsittelyaika putoaisi noin 200 päivään.



Kuva 74. Keskimääräinen patenttien määrä kpl/kk (Y-akseli) ja käsittelyajan keston keskiarvo ajan funktiona (päivää X-akseli). Kuvaajien lähtödata perustuu 1.3.2018 otantaan.

Taulukko 43. Poiminta- ja laskentatulokset perustuen 1.3.2018 otokseen.

Julkaisukuukausi	Yhteenlaskettu käsittelyaika	Määrä	keskiarvo
200908	175	1	175
201105	132	1	132
201107	552	1	552
201210	128	1	128
201212	121	1	121
201304	224	2	112

201305	1593	6	265,5
201306	178	2	89
201307	831	4	207,75
201308	107	1	107
201309	2022	8	252,75
201310	1627	6	271,166667
201311	351	3	117
201312	1165	10	116,5
201401	1896	14	135,428571
201402	2030	14	145
201403	1192	11	108,363636
201404	1458	10	145,8
201405	2668	12	222,333333
201406	1581	10	158,1
201407	2290	18	127,222222
201408	3313	22	150,590909
201409	3073	19	161,736842
201410	4050	22	184,090909
201411	4787	20	239,35
201412	5676	29	195,724138
201501	4206	30	140,2
201502	2964	17	174,352941
201503	6526	36	181,277778
201504	6063	38	159,552632
201505	9775	41	238,414634
201506	8086	48	168,458333
201507	9945	52	191,25
201508	7107	34	209,029412
201509	10359	59	175,576271
201510	12664	54	234,518519
201511	10820	58	186,551724
201512	15347	88	174,397727
201601	15820	83	190,60241
201602	9059	55	164,709091
201603	14905	66	225,833333
201604	13707	70	195,814286
201605	15568	71	219,267606
201606	25245	120	210,375
201607	18212	91	200,131868
201608	21123	108	195,583333
201609	21662	107	202,448598
201610	25884	97	266,845361
201611	28237	130	217,207692
201612	28491	146	195,143836
201701	27503	139	197,863309
201702	24574	137	179,372263
201703	29760	152	195,789474
201704	23739	98	242,234694
201705	54246	254	213,566929
201706	43189	195	221,482051
201707	31923	142	224,809859
201708	32429	146	222,116438
201709	29118	152	191,565789
201710	24973	138	180,963768
201711	5765	16	360,3125
201712	7091	21	337,666667
201801	11009	24	458,708333
201802	24595	45	546,555556

LIITE L: PATENTTIEN KÄSITTELYAIKAA KUVAAVA HISTOGRAMMI

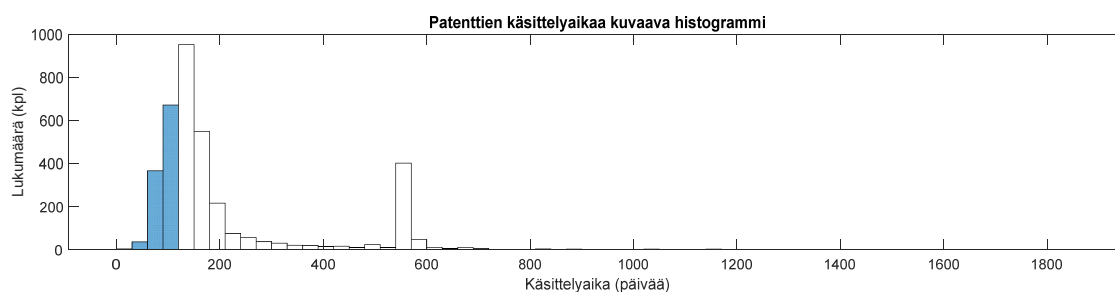
Julkaisu- ja hakemispäivämäärän erotuksen laskenta käsittelyajan funktiona. Laskenta esitetty histogrammina, jossa pyritään esittämään selvä poikkeama monista jakaumista sekä käsittelyaikojen jakautuminen kahteen eri joukkoon.

Toisin sanoen histogrammin perusteella keskiarvon tai mediaanin laskeminen johtavat harhaisiin päätelmiin, koska kesto-aika on jakautunut kahteen joukkoon ja joukot tulee käsitellä erikseen. Tämä datajoukko jaettiin joukkoihin A (alle 400 päivää) ja joukkoihin B (vähintään 400 päivää).

Ohjelmisto: Matlab 2017b

```
clear variables;
close all;
data = load("C:\Users\allu1\OneDrive\Tiedostot\DI-tyo\Patentit\Julkaisuviive\Julkaisuviive_big_data_kylla_2009-2018.dat");

F_size = 8;
width = 3;
bar_w = 0.7;
figure(1)
histogram(data);
title('Patenttien käsittelyaikaa kuvaava histogrammi')
xlabel('Käsittelyaika (päivää)')
ylabel('Lukumäärä (kpl)')
```



LIITE M: MATLAB-KOODI JOUKON A JA B LAS-KENTAAN JA KUVAAMISEEN

Aineiston visuaalisella kuvaamisella voidaan kuvata aineistoa laajasti muutamilla kuvaajilla. Laatikkojanakuvaaja, histogrammi ja vertailu normaalijakaumaan toteutetaan Matlab 2017b -ohjelmiston valmisfunktioilla joukoille A ja B.

Histogrammien perusteella alueiden jakaumat eivät ole symmetriset. Otoksiko on riittävä ja joukossa A on käsittelyaika jakautunut tasaisemmin. Joukon B osalta käsittelyaika on hyvin pienellä alueella.

Laatikkojanakuvaajien perusteella Matlab esittää kvantiilit 0.25, 0.50 ja 0.75. Kvantiileista voidaan todeta jakauman vähäinen vinous joukossa A, koska mediaanin sijaitsee keskellä. Joukon B osalta laatikkojana on äärimmäisen lyhyt, joka kuvaa käsittelyajan keskittymistä lyhyelle aikajaksolle. Joukon B osalta voidaan laatikkojanan koostuvan lähinnä outlieriestä. Matlab etsii outlieriä tarkastelemalla 50% laatikkojanan pituutta, josta se kertoo pituuden 1,5:llä. Matlab etsii outlieriä tarkastelemalla 50% laatikkojanan pituutta, josta se kertoo pituuden 1,5:llä (eli Outlinerit: $x > q_3 + 1,5(q_3 - q_1)$ or $x < q_1 - 1,5(q_3 - q_1)$) ja päättää tällä tuloksella outlierin. Jos kuvaaja ei ole symmetrinen tai otoksessa kuuluu olla huomattavasti hajontaa, voidaan tulkita oikeita tuloksia outlieriksi. Joukon B osalta tämä voi olla todennäköistä. Poistamalla otoksesta 550-560 käsittelypäivää kestäneet patentit, voisi tutkimusta jatkaa.

Normaalijakaumien osalta joukot mukailevat osittain normaalijakaumaa, mutta irtoavat normaalijakaumasta ääriarvoja saavuttaessa, jotka ovat tulkittavissa outlieriksi.

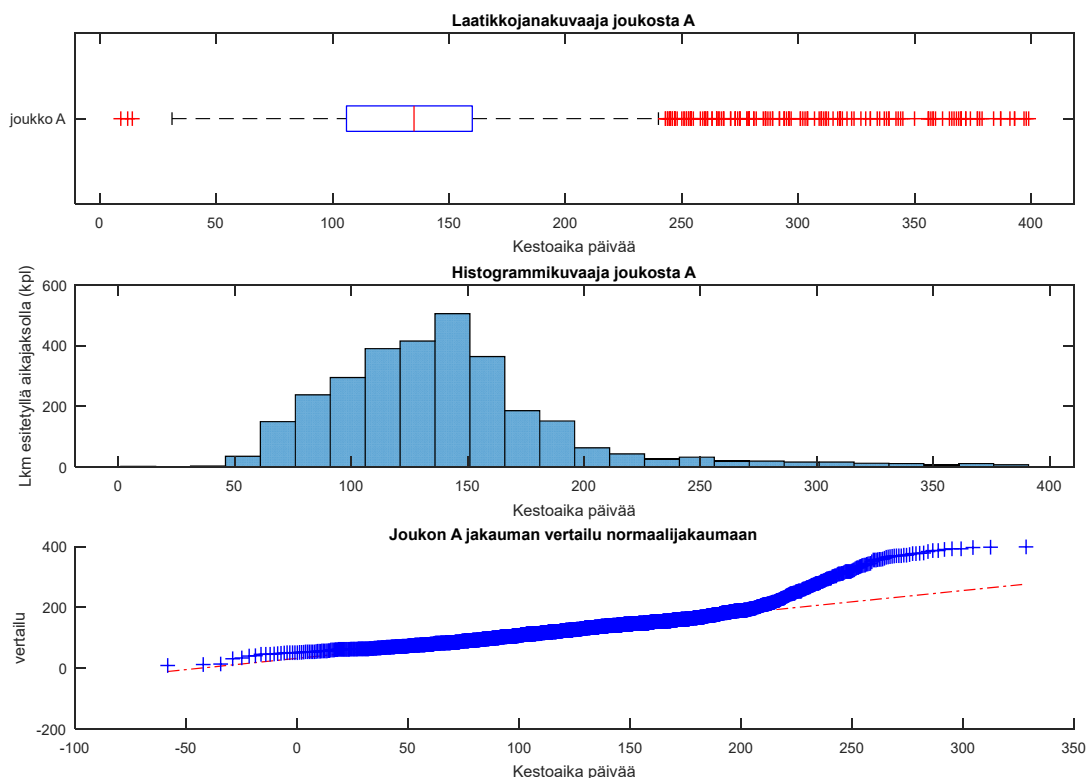
```
clear variables, close all
data = load("C:\Julkaisuviive\Julkaisuviive_big_data_kylla_2009-2018.dat")
joukkoA = data(data < 400)
F_size = 8;
width = 15;
bar_w = 1;
mean_A=mean(joukkoA)
Kuvaajien_X_alaraja=1;
Kuvaajien_X_ylaraja=400;
%NM = 'probability'; %suhteelliset arvot = n1/n
edges = [Kuvaajien_X_alaraja:width:Kuvaajien_X_ylaraja];
pd_ms = makedist('normal','mu',median(joukkoA),'sigma',std(joukkoA));
figure(1)
subplot(3,1,1)
boxplot(joukkoA, 'plotstyle', 'traditional','orientation','horizontal', 'Labels',{'joukko A'})
set(gca, 'FontSize', 6);
title('Laatikkojanakuvaaja joukosta A')
xlabel('Kestoaika päivää')

subplot(3,1,2)

histogram(joukkoA,edges);
set(gca, 'FontSize', 6);
title('Histogrammikuvaaja joukosta A')
xlabel('Kestoaika päivää')
ylabel('Lkm esitetyllä aikajaksolla (kpl)')

subplot(3,1,3)
qqplot(joukkoA, pd_ms) %asetetaan arvot ja jakauma, johon verrataan //
set(gca, 'FontSize', 6);
title('Joukon A jakauman vertailu normaalijakaumaan')
xlabel('Kestoaika päivää')
```

```
ylabel('vertailu')
```

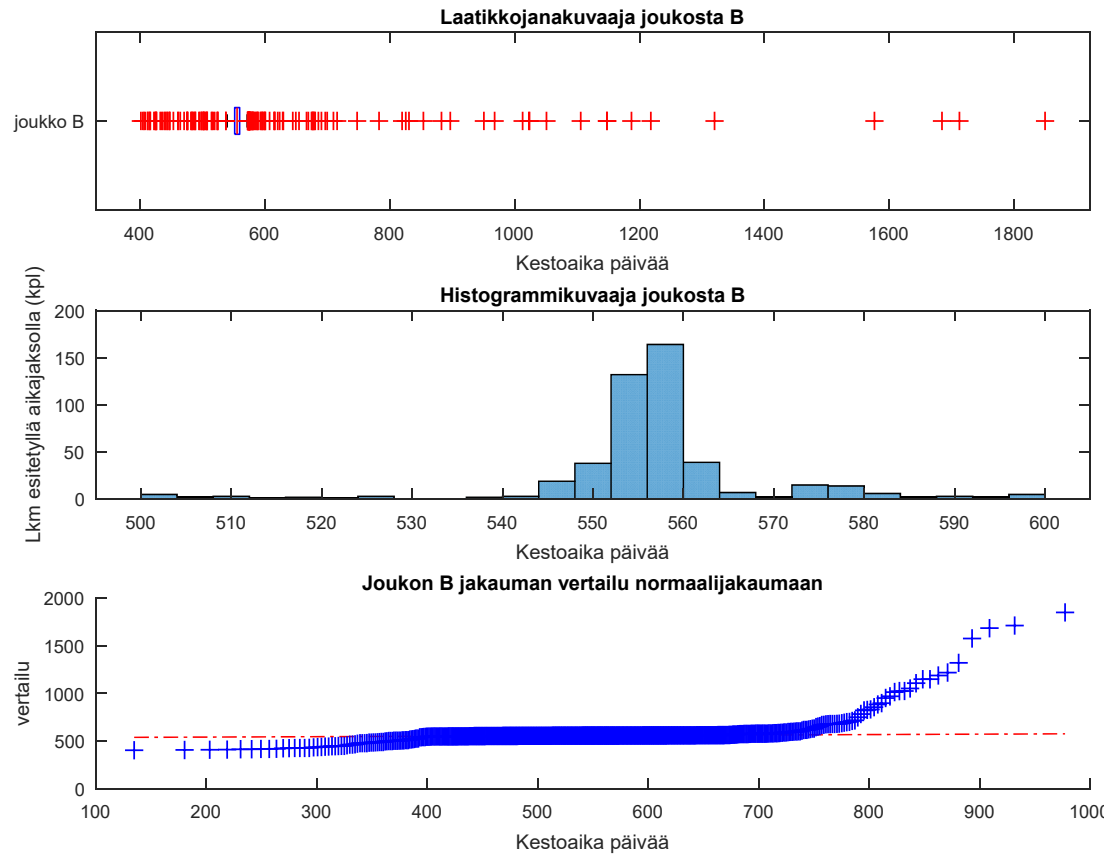


```
clear variables, close all
data = load("C:\Users\allu1\OneDrive\Tiedostot\DI-tyo\Patentit\Julkaissuviive\Julkaissuviive_big_data_kylla_2009-2018.dat")
joukkoB = data(data > 399)
F_size = 8;
width = 4;
bar_w = 1;
mean_A=mean(joukkoB)
Kuvaajien_X_alaraja=500;
Kuvaajien_X_ylaraja=600;
%NM = 'probability'; %suhteelliset arvot = n1/n
edges = [Kuvaajien_X_alaraja:width:Kuvaajien_X_ylaraja]';
pd_ms = makedist('normal', 'mu', median(joukkoB), 'sigma', std(joukkoB));
figure(1)
subplot(3,1,1)
boxplot(joukkoB, 'plotstyle', 'traditional', 'orientation', 'horizontal', 'Labels', {'joukko B'})
set(gca, 'FontSize', 6);
title('Laatikkojanakuvaaja joukosta B')
xlabel('Kesto aika päivää')

subplot(3,1,2)

histogram(joukkoB, edges);
set(gca, 'FontSize', 6);
title('Histogrammikuvaaja joukosta B')
xlabel('Kesto aika päivää')
ylabel('Lkm esitettyllä aikajaksolla (kpl)')

subplot(3,1,3)
qqplot(joukkoB, pd_ms) %asetetaan arvot ja jakauma, johon verrataan //
set(gca, 'FontSize', 6);
title('Joukon B jakauman vertailu normaalijakaumaan')
xlabel('Kesto aika päivää')
ylabel('vertailu')
```



LIITE N: MATLAB-KOODI, JOLLA LASKETAAN JA ESITETÄÄN LAATIKKOJANA KUVAAJA JOUKOSTA, JOILLA PATENTTIEN KÄSITTELYAIKA ON VÄHINTÄÄN 400 PÄIVÄÄ

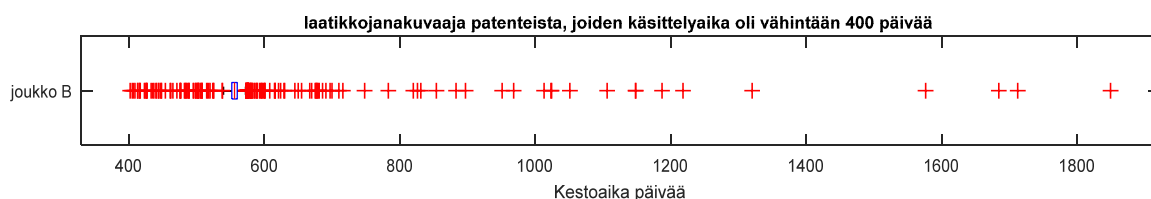
```
clear variables;
close all;
data = load("C:\Users\allu1\OneDrive\Tiedostot\DI-tyo\Patentit\Julkaisuviive\Julkaisuviive_big_data_kylla_2009-2018.dat")
joukko_vahintaan_400_paivaa = data(data > 399)

F_size = 8;
width = 15;
bar_w = 0.7;

mean_B=mean(joukko_vahintaan_400_paivaa)
Kuvaajien_X_alaraja=530;
Kuvaajien_X_ylaraja=590;

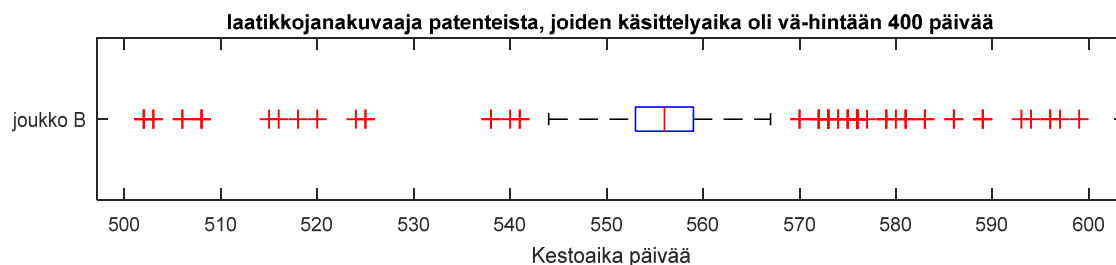
edges = [Kuvaajien_X_alaraja:width:Kuvaajien_X_ylaraja]';
pd_ms = makedist('normal','mu',median(joukko_vahintaan_400_paivaa),'sigma',std(joukko_vahintaan_400_paivaa));
figure(1)

boxplot(joukko_vahintaan_400_paivaa, 'plotstyle', 'traditional','orientation','horizontal', 'Labels',{'joukko B'})
set(gca, 'FontSize', 6);
title('laatikkojanakuvaaja patenteista, joiden käsittelyaika oli vähintään 400 päivää')
xlabel('Kestoaiika päivää')
```



Joukkoa B voidaan tarkemmin tarkastella, joka vahvistaa olettaa, että aineisto on keskittynyt voimakkaasti.

```
clear variables;
close all;
data = load("C:\Julkaisuviive_big_data_kylla_2009-2018.dat")
data = data(data > 499)
data = data(data < 600)
figure(1)
boxplot(data, 'plotstyle', 'traditional','orientation','horizontal', 'Labels',{'joukko B'})
set(gca, 'FontSize', 6);
title('laatikkojanakuvaaja patenteista, joiden käsittelyaika oli vähintään 400 päivää')
xlabel('Kestoaiika päivää')
```



LIITE O: ESPACENET -TIETOKANNASSA ESIINTYNEET IPC-KOODIT BIG DATAAN LIITTYEN KK \ LKM. (08/2009-2/2018)

*) poimintatuloksessa ei käsitelty määrittelemättömiä (Null) esiintymiä

Liite O IPC-koodit Kk \ lkm	0	1	2	3	4	5	6	7	8	9	Yhteensä
200908			1								1
201105		1									1
201107					1						1
201210		1									1
201212		1									1
201304		2									2
201305		5	1								6
201306		2									2
201307		4									4
201308		1									1
201309		2	5	1							8
201310		3	2	1							6
201311		3									3
201312		9	1								10
201401		10	3	1							14
201402		11	1	1	1						14
201403		7		3	1						11
201404		8	2								10
201405		6	5	1							12
201406		5	1	4							10
201407		11	5	2							18
201408		9	9	2	2						22
201409		12	6	1							19
201410		13	7	1	1						22
201411		15	3	2							20
201412		18	8	3							29
201501		17	11	1	1						30
201502		8	8		1						17
201503		21	12	3							36
201504		22	14	2							38
201505		26	13	2							41
201506		30	16	2							48
201507		24	23	4	1						52
201508		21	12	1							34
201509		37	13	9							59
201510		32	17	3	2						54
201511		23	25	9	1						58
201512		54	27	5	2						88
201601	1	37	28	12	4	1					83
201602		29	17	8			1				55
201603		34	24	5	2		1				66
201604		40	21	6	2	1					70
201605	1	39	22	7	1	1					71
201606		67	34	15	4						120
201607		45	32	10	2	1	1				91
201608		63	30	8	6		1				108
201609		56	38	10	1	1	1				107
201610		49	36	8	2	2					97
201611		55	38	19	8	5	2	3			130
201612		74	41	20	7	2	2				146
201701		70	41	16	5	3	4				139
201702		62	37	26	6	3	1		2		137
201703		71	55	12	9	2	2	1			152
201704		49	27	14	4	4					98
201705		113	79	33	10	5	11	2		1	254
201706		70	71	29	9	11	4	1			195
201707		50	35	24	18	7	6	2			142
201708		53	42	31	9	5	3	1	1	1	146
201709		57	43	28	11	10		1		2	152
201710	3	53	47	15	15	1	2	1	1		138
201711		8	2	2	1		3				16
201712		5	5	3	3	3	1		1		21
201801		6	6	6	4	1	1				24
201802		37	1	2	2	3					45
Yhteensä	5	1766	1103	433	159	72	47	12	5	4	3606

LIITE P: PYLVÄSKUVAAJAESITYS IPC-KOODILLISISTA PATENTTIEN MÄÄRÄLLISISTÄ OSUUKSISTA (KPL) PATENTEISTA KOODIEN LUKUMÄÄRÄN MUKAAN

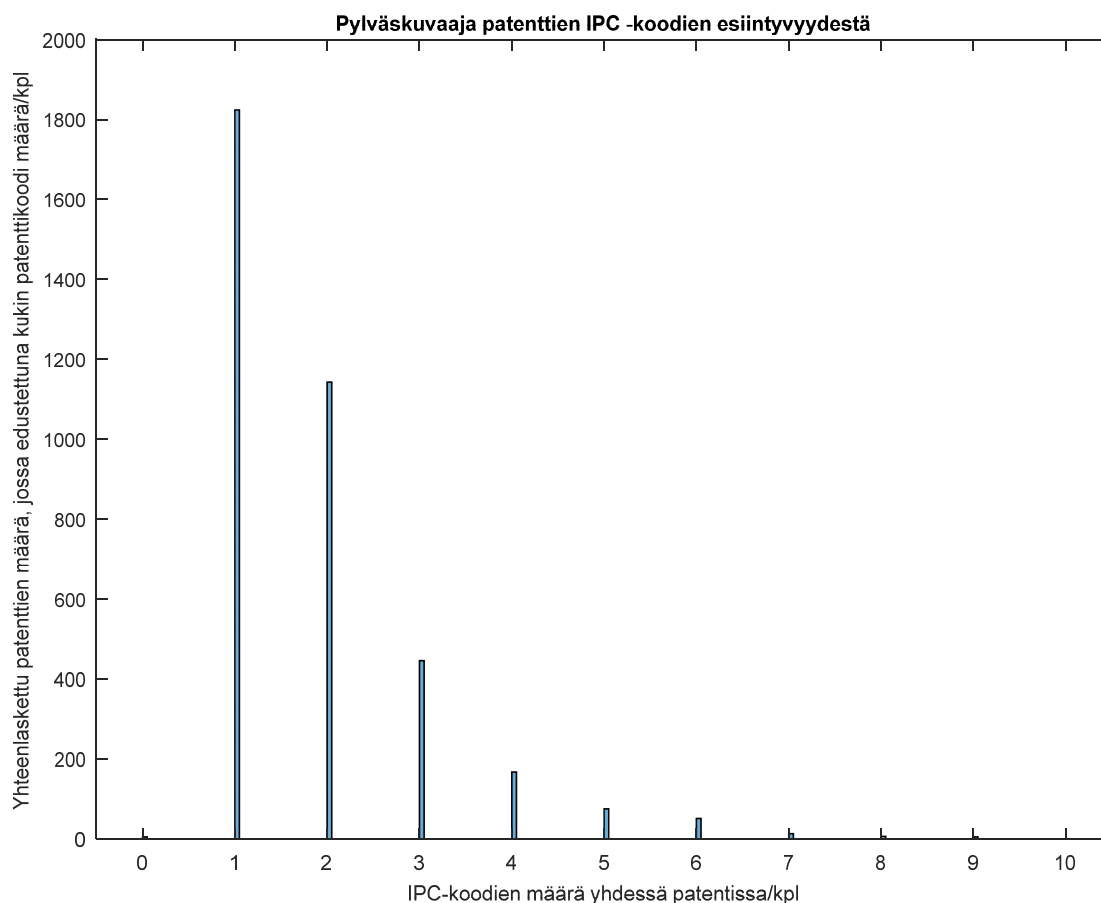
```
clear variables, close all
data = load("C:\Users\allu1\OneDrive\Tiedostot\DI-tyo\Patentit\IPC\ipc-koodi.dat")
data=data(:,1)
F_size = 8;
width = 0.05;
bar_w = 1;

Kuvaajien_X_alaraja=0;
Kuvaajien_X_ylaraja=10;

edges = [Kuvaajien_X_alaraja:width:Kuvaajien_X_ylaraja]';

figure(1)

histogram(data,edges)
set(gca, 'FontSize', 6);
title('Pylväskuvaaja patenttien IPC -koodien esiintyvyydestä')
xlabel('IPC-koodien määrä yhdessä patenttissa/kpl')
ylabel('Yhteenlaskettu patenttien määrä, jossa edustettuna kukin patenttikoodi määrä/kpl')
```



LIITE Q: MATLAB-OHJELMA EPÄLINEAARISEN SYSTEMIN PARAMETRIEN SELVITTÄMISEEN

%Lähde: Perustuu lähteeseen: <https://www.mathworks.com/matlabcentral/answers/410081-errors-in-objective-function> -koodeista, joissa useampi kirjoittaja.

```
clear variables;
close all;

data=[0;0;0;0;1;1;1;1;1;2;2;3;3;2;4;4;4;3;3;4;3;6;4;10;13;15;20;30;45;67;100];
v=100;

x = outer_function(v, data);
disp(x)

%lsqcurvefit
t=1:30;
ydata=[0;0;0;0;1;1;1;1;1;2;2;3;3;2;4;4;4;3;3;4;3;6;4;10;13;15;20;30;45;67;100];
v=1000;
fun=@(x,t)Bassmodel(x,t,v);
lb=[0,0,0,0,0];
ub=[1, Inf, 1, 1, Inf];
x0=[20,1,1,1,300];
options = optimoptions(@lsqcurvefit,'Display','off');
x = lsqcurvefit(fun,x0,t,ydata,lb,ub,options);
y_val = Bassmodel(x,t,v);

% lsqnonlin
t=1:30;
ydata=[0;0;0;0;1;1;1;1;1;2;2;3;3;2;4;4;4;3;3;4;3;6;4;10;13;15;20;30;45;67;100];
v=1000;
lb=[0,0,0,0,0];
ub=[1, Inf, 1, 1, Inf];
x0=x;
fun = @(x)ydata-(x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-
x(4))*exp(-x(5)*v))).*t)));
options2 = optimoptions(@lsqnonlin,'Display','off');
x2 = lsqnonlin(fun,x0,lb,ub,options2);
y_val2 = Bassmodel(x2,t,v);

% plot results
plot(t,ydata,'x')
hold on
plot(t,y_val,'o')
hold on
plot(t,y_val2,'*')
hold off

%show results
disp(['x2(1): ', num2str(x2(1))])
disp(['x2(2): ', num2str(x2(2))])
disp(['x2(3): ', num2str(x2(3))])
disp(['x2(4): ', num2str(x2(4))])
disp(['x2(5): ', num2str(x2(5))])

%objective function for lsqcurvefit
function [F] = Bassmodel(x,t,v)
    F=x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-
x(4))*exp(-x(5)*v))).*t)));
end

function j = outer_function(v, data)
fun=@Bass;
nl=@nonlin;
lb=[20,0,0,0,0];
ub=[inf,1,1,1,inf];
x0=[ 0,0,0,0,0];
j=fmincon(fun,x0,[],[],[],[],lb,ub,nl);

function [F] = Bassmodel(x,t,v)
    P = x(1);
    m = x(2);
```

```

q_m = x(3);
q_0 = x(4);
w = x(5);
F = -(P.*m.*(exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-q_m)))-
1.0))./(P+(q_0.*q_m.*exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-
q_m))))./(q_0-exp(-v.*w)).*(q_0-q_m)));
end

function res= Bass(x)
%objective function
F= (- data + x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-
x(5)*v)))/ ( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)));
res=sum(F.^2);
end

function [c, ceq]= nonlinc(x)
%constraint
c=[];
ceq=- data + ( x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-
x(4))*exp(-x(5)*v)))/ ( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v)));
end

end

```

LIITE R: BASSIN INNOVAATIODIFFUUSIO LASKENTA

%Lähde: perustuu lähteeseen <https://www.mathworks.com/matlabcentral/answers/410081-errors-in-objective-function>

```
clear variables;
close all;

data=[0;0;0;0;1;1;1;1;1;2;2;3;3;2;4;4;4;3;3;4;3;6;4;10;13;15;20;30;45;67;100];t=1:25;
v=1000;

x = outer_function(v, data);
disp(x)

%lsqcurvefit
ydata=data'

v=1000;
fun=@(x,t)Bassmodel(x,t,v);
lb=[0,0,0,0,0];
ub=[1, Inf, 1, 1, Inf];
x0=[20,1,1,1,300];
options = optimoptions(@lsqcurvefit,'Display','off');
x = lsqcurvefit(fun,x0,t,ydata,lb,ub,options);
y_val = Bassmodel(x,t,v);

% lsqnonlin
x0=x;
fun = @(x)ydata-(x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t)));
options2 = optimoptions(@lsqnonlin,'Display','off');
x2 = lsqnonlin(fun,x0,lb,ub,options2);
y_val2 = Bassmodel(x2,t,v);

% plot results
plot(t,ydata,'+')

hold on
plot(t,y_val,'.')%Solve nonlinear curve-fitting (data-fitting) problems in least-squares
sense

hold on
plot(t,y_val2,'--') %Solve nonlinear least-squares (nonlinear data-fitting) problems
legend('Havainnot','Bass-funktio: epälineaarinen käyräsovitus pienimmällä neliösummal-
la','Bass-funktio: epälineaarinen pienimmän neliösumman data-sovitus','Location','north-
west')
hold off

%show results
disp(['p: ', num2str(x2(1))])
disp(['m: ', num2str(x2(2))])
disp(['q(m):', num2str(x2(3))])
disp(['q_0: ', num2str(x2(4))])
disp(['w: ', num2str(x2(5))])

%objective function for lsqcurvefit
function [F] = Bassmodel(x,t,v)
F=x(1)*x(2)*(1-exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t))./...
(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))*exp(-(x(1)+x(3)*x(4)/(x(4)+(x(3)-x(4))*exp(-x(5)*v))).*t)));
end

function j = outer_function(v, data)
fun=@Bass;
nl=@nonlin;
lb=[20,0,0,0,0];
ub=[inf,1,1,1,inf];
x0=[ 0,0,0,0,0];
j=fmincon(fun,x0,[],[],[],[],lb,ub,nl);

function [F] = Bassmodel(x,t,v)
P = x(1);
m = x(2);
q_m = x(3);
```

```

    q_0 = x(4);
    w = x(5);
    F = -(P.*m.*(exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-q_m))))-
    1.0))./(P+(q_0.*q_m.*exp(-t.*(P+(q_0.*q_m)./(q_0-exp(-v.*w)).*(q_0-
    q_m)))))./(q_0-exp(-v.*w)).*(q_0-q_m));
end

function res= Bass(x)
%objective function
F= (- data + x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-x(4))*exp(-
x(5)*v)))/ ( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v));
res=sum(F.^2);
end

function [c, ceq]= nonlinc(x)
%constraint
c=[];
ceq=- data + ( x(1)*x(2)*(1 - exp( x(2) + (( x(3)*x(4))/ x(3)-
x(4))*exp(-x(5)*v)))/ ( 1+ (( x(3)*x(4))/ x(3)-x(4))*exp(-x(5)*v));
end

end
.
```