

Eetu Eloranta

**BIG DATAN JA DATA-ANALYTIIKAN HAASTEET
MARKKINOINNISSA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2021

TIIVISTELMÄ

Eloranta, Eetu

Tutkimusraportin otsikko

Jyväskylä: Jyväskylän yliopisto, 2021, 70 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Taipalus, Toni

Big data ja siihen liittyvä analytiikka on jo pitkään ollut IT-alan ja liiketoiminnan hehkutetuimpia teknologiatrendejä. Big datan mahdollisuudet on laajalti tunnustettu eri toimialoilla ja sen potentiaali on saanut painoarvoa niin tieteellisessä tutkimuksessa, kuin käytännön liiketoiminnassakin. Big data -analytiikka on saanut myös markkinoijat kiinnostumaan sen luomista mahdollisuuksista onnistuneen markkinoinnin arvonluoja. Tutkimuksia big datan haasteista on tehty aiemmin, mutta vähemmän painoarvoa tutkimuksessa on kuitenkin saanut big datan ja analytiikan haasteet markkinoinnissa. Tässä tutkielmassa pyrimme löytämään keskeiset haasteet big datan ja analytiikan hyödyntämiseen markkinoinnin kontekstissa hyödyntämällä integroivaa kirjallisuuskatsausta tutkimusmetodina. Jaamme haasteet teknisiin, volyyymiin liittyviin, nopeuteen ja reaaliaikaiseen hyödyntämiseen liittyviin, monipuolisuuteen liittyviin, kompleksisuuteen liittyviin, arvoon liittyviin ja yksityisyyteen ja omistajuuteen liittyviin haasteisiin. Pyrimme luomaan kokonaisvaltaisen kuvan markkinoinnin toimialalla big datan ja analytiikan yleisistä kompastuskivistä. Keskeisinä löydöksinä havaitsemme tutkielmassa, että big datan ja analytiikan hyödyntämiseen markkinoinnissa liittyy paljon haasteita niin datan hallinnan, prosessoinnin kuin varastoinninkin näkökulmasta. Osa löydetyistä haasteista on toimialasta riippumattomia, osa taas markkinoinnin toimialalle ominaisia ongelmia. Kirjallisuuskatsauksen tulosten pohjalta luomme viitekehyksen big datan ja analytiikan merkittävistä haasteista markkinoinnissa.

Asiasanat: Big data, Big data markkinointi, data-analytiikka, ...

ABSTRACT

Eloranta, Eetu

Challenges of Big data and analytics in marketing

Jyväskylä: University of Jyväskylä, 2021, 70 pp.

Information Systems,

Supervisor: Taipalus, Toni

Big data and analytics have long been one of the most popular technology trends in the IT-industry and business. The potential of big data has been widely recognized in various industries and its potential has gained weight in both scientific research and practical business. Big data-analytics have also made marketers interested in the opportunities it creates as a value creator of successful marketing. Research on the challenges of big data has been done in the past, but less emphasis has been placed on the challenges of big data and analytics in the context of marketing. In this paper, we aim to find key challenges in utilizing big data and analytics in the context of marketing by utilizing integrative literature review as a research method. We divide these challenges to technical, volume-related, speed-related, variety-related, complexity-related, value-related, and privacy and ownership challenges and aim to create a holistic view of the common stumbling blocks of big data and analytics in the marketing industry. As key findings, we find that the utilization of big data and analytics in marketing involves many challenges from the perspective of data management, processing, and storage. Some of the challenges identified are valid in every industry, while others are specific to the marketing industry. Based on the results, we create a framework for significant challenges of big data and analytics in marketing.

Keywords: Big data, Big data marketing, data-analytics, ...

KUVIOT

KUVIO 1 Viitekehys big datan ja analytiikan haasteista markkinoinnissa

SISÄLLYS

| | |
|---|----|
| TIIVISTELMÄ | 2 |
| ABSTRACT | 3 |
| KUVIOT | 4 |
| SISÄLLYS..... | 5 |
| 1 JOHDANTO..... | 6 |
| 2 BIG DATA JA DATA-ANALYTIikka..... | 8 |
| 2.1 Big datan määritelmä | 8 |
| 2.2 Big data- analytiikka..... | 11 |
| 2.2.1 Big data-analytiikan teknologiat | 11 |
| 2.3 Big data ja analytiikka markkinoinnissa | 12 |
| 3 BIG DATAN JA DATA-ANALYTIIKAN HAASTEET MARKKINOINNISSA | 16 |
| 3.1 Big datan haasteet markkinoinnissa | 16 |
| 3.1.1 Big datan ja analytiikan tekniset haasteet..... | 17 |
| 3.1.2 Kompleksisuuteen liittyviä haasteita | 17 |
| 3.1.3 Datan nopeuteen liittyviä haasteita | 18 |
| 3.1.4 Datan volyyymiin liittyviä haasteita | 19 |
| 3.1.5 Datan heterogeenisuuteen liittyviä haasteita..... | 20 |
| 3.1.6 Big datan ja analytiikan arvo markkinoinnissa..... | 22 |
| 3.1.7 Big datan omistajuuteen ja yksityisyyteen liittyviä haasteita.... | 23 |
| 3.1.8 Viitekehys big datan ja analytiikan haasteista markkinoinnissa | 24 |
| JOHTOPÄÄTÖKSET JA JATKOTUTKIMUSAIHEET | 26 |
| LÄHTEET | 28 |

1 Johdanto

Viime vuosina on syntynyt laaja yhteisymmärrys yritysten markkinoijien ja markkinoinnin tutkijoiden välillä siitä, että datan hyödyntämisellä markkinointipäätöksissä on potentiaalia muuttaa näkemyksiä markkinoinnista yrityksen kulueränä yrityksen arvonluojaksi (Kumar, Chattaraman, Neghina, Skiera, Aksoy, Buoye & Henseler, 2013). Big data on ollut jo vuosien ajan yksi keskeisimmistä IT-alan trendeistä. Sen hyödyt ja mahdollisuudet ovat paljon esillä niin tiedeyhteisöissä, kuin käytännön liiketoiminnassakin. Big datan potentiaalista huolimatta sen hyödyntäminen liiketoiminnassa ja päätöksenteossa ei ole ongelmatonta. Columbia Business School ja New York American Marketing Associationin tekemän kyselytutkimuksen mukaan niistä markkinointijohtajista ja markkinointipäälliköistä, joiden yritykset keräävät volyymiltaan suurta määrää dataa, 39 prosenttia ilmaisi, että ei kykene muuntamaan kerättyä dataa hyödynnettäväksi oivalluksiksi (Kumar ym., 2013). Kumar ym. (2013) nostavatkin big datan yhdeksi suurimmista haasteista digitaalisessa markkinoinnissa.

Vaikka big datan hyödyntämisellä liiketoiminnassa ja markkinoinnin päätöksenteossa on epäilemättä potentiaalia, on sen hyödyntämiseen liittyvät haasteet saaneet vähemmän painoarvoa kirjallisuudessa. Big datan ja big data-analytiikan haasteet tiedostamalla voidaan ymmärtää paremmin niihin liittyviä rajoitteita, sekä voidaan kehittää ratkaisuja näiden haasteiden ehkäisemiseksi. Big datan ja analytiikan haasteiden tiedostaminen on tärkeää markkinoijille, sillä analysoitavan tiedon määrä tulee mitä todennäköisemmin tulevaisuudessa ainoastaan kasvamaan. Haasteet tiedostamalla markkinoijat pystyvät hyödyntämään big datan potentiaalinen arvon luojana ja ehkäisemään sen haasteita omalla toiminnallaan. Big datan haasteita on tutkittu kirjallisuudessa vain vähän, mutta markkinoinnin kontekstista big datan haasteita on tutkittu sitäkin vähemmän. Tästä syystä tutkimusaihe on tarpeellinen myös johtuen sen kontribuutiosta tiedeyhteisölle. Tässä tutkielmassa tutkimus tehdään integroivan kirjallisuuskatsauksen muodossa ja pyritään kokoamaan kirjallisuudesta vastaus seuraavaan tutkimuskysymykseen:

Tutkimuskysymys: Mitä haasteita liittyy big datan ja big data-analytiikan hyödyntämiseen markkinoinnissa

Kirjallisuuskatsauksessa tietokantoina on hyödynnetty Google Scholaria, IEEE explorea, sekä Scopusta.

Tutkielman ensimmäisessä luvussa käsitellään tutkielman näkökulmasta keskeistä käsitteistöä. Ensiksi käsittelemme big datan käsitettä, rakentaen pohjan sille, millaisia artikkeleita valikoidaan varsinaiseen kirjallisuuskatsaukseen. Luvussa määrittelemme myös big data-analytiikan ja käymme läpi siihen liittyviä keskeisiä teknologioita. Lopulta käsittelemme big datan ja analytiikan roolia markkinoinnin alalla ja esitämme keskeisiä big data-analytiikan sovelluskohteita markkinoinnissa. Tutkielman toisessa sisältöluvussa pyrimme löytämään tutkimuskysymykseemme vastauksen. Jaamme löydetyt haasteet teknisiin haasteisiin, kompleksisuuteen liittyviin, nopeuteen liittyviin, volyyymiin liittyviin, monipuolisuuteen liittyviin, arvoon liittyviin sekä omistajuuteen ja yksityisyyteen liittyviin haasteisiin. Luvussa luomme kirjallisuuskatsauksen pohjalta viitekehysten haasteiden visualisoimiseksi selkeään muotoon. Tutkielman yhteenvedossa kerrataan tutkimuksen keskeiset tulokset, analysoidaan tutkimuksen rajoitteita ja ehdotetaan jatkotutkimusaiheita kirjallisuuskatsauksessa havaittujen puutteiden pohjalta.

2 Big data ja data-analytiikka

Tässä luvussa ensin paneudumme big datan ja big data-analytiikan määrittelyihin. Sen jälkeen kuvaamme yleiskatsauksen niiden merkityksestä markkinoinnin kontekstissa ja kerromme miten big dataa ja data-analytiikkaa voidaan hyödyntää markkinoinnin alalla.

2.1 Big datan määritelmä

Big dataan liittyvän kirjallisuuden nopea ja kaoottinen kehitys on estänyt yleisesti hyväksytyt ja muodollisen määrittelyn syntymistä big data-termille (Mauro, Greco & Grimaldi, 2016). Mauro Greco & Grimaldi (2015) mukaan termiä käytetään kuvaamaan niin teknologista kykyä tallentaa, koota ja käsitellä tietoja kuin kulttuurin muutosta, joka vallitsee yrityksiä ja yhteiskuntaa. Varhaiset big data-termin määrittelyt liittyvät niin suureen tietomäärään, että sen tallentamista, hallintaa ja prosessointia ei voida tehokkaasti suorittaa perinteisillä tiedonhallinnan menetelmillä ja työkaluilla (Kaisler, Armour, Espinosa & Money, 2013). Yksinkertaisimmillaan big datan määrittelyssä otetaan huomioon ainoastaan datan määrä. Useimmiten korostetaan kuitenkin muita datan ominaisuuksia yhtä keskeisinä tekijöinä termin määrittelyyn.

2.1.1. Big datan ominaisuudet

Laney (2001) esitteli artikkelissaan nykyään jo yleisesti big datan määrittelyssä käytössä olevan niin sanotun 3V:n mallin suurien datamäärien hallinnan ulottuvuuksista, jossa otetaan huomioon datan koon (volume) lisäksi, myös datan nopeus (velocity) ja datan monipuolisuus (variety). Gandomi & Haider (2015) kuvailevat Big datan 3 V:tä seuraavalla tavalla.

- Koolla tarkoitetaan datan suuruutta; big datan kohdalla koko ilmoitetaan useimmiten teratavuina tai petatavuina.

- Monipuolisuudella tarkoitetaan tietojoukon rakenteellista heterogeenisuutta. Big datan tulee olla rakenteeltaan monipuolista, eli se sisältää rakenteellisen datan lisäksi semi-strukturoitua ja strukturoimatonta dataa.
- Nopeus viittaa siihen nopeuteen, jolla data syntyy ja nopeuteen, jolla sitä voidaan analysoida ja hyödyntää.

Gandomi & Haiderin (2015) mukaan on esimerkiksi toimialasta ja datan tyyppistä suhteellista miten suurta, monipuolista ja nopeaa datan täytyy olla, jotta se luokitellaan big dataksi. He tuovat esiin myös, että tallennuskapasiteetin kasvessa se mitä nykyään pidetään big datana, ei välttämättä täytä kriteerejä enää tulevaisuudessa. Lisäksi heidän mukaansa kaksi samankokoista tietoaineistoa voi vaatia erilaisia tiedonhallintatekniikoita riippuen datan tyyppistä. Näin ollen big datalla ei ole perusteltua luoda tarkkoja arvoja datan volyymille, nopeudelle ja monipuolisuudelle, sillä eri konteksteissa voi olla perusteltua käsitellä erilaisista data-aineistosta big datana, kuin jossain toisessa kontekstissa. Gandomi & Haider (2015) perustelee big datan nopeutta keskeisenä attribuuttina määritelmässä, sillä digitaalisten laitteiden leviäminen on johtanut poikkeukselliseen datanluonnin nopeuteen, mikä taas lisää tarvetta reaaliaikaiseen analyysiin ja näyttöön perustuvaan suunnitteluun. Heidän mukaansa perinteiset tiedonhallintajärjestelmät eivät kykene käsittelemään valtavia datasyötteitä reaaliajassa.

Big datan määritelmään on myös ehdotettu edellisten lisäksi muita tekijöitä. Näistä yleisiä ovat esimerkiksi arvo (value), vaihtelevuus (variability), kompleksisuus (complexity) sekä todenmukaisuus (veracity). Esimerkiksi Jin, Wah, Cheng & Wang (2015) nostavat todenmukaisuuden ja arvon big dataa määrittäväksi attribuuteiksi perinteisen 3V:n mallin attribuuttien lisäksi. Kaisler, Armour, Espinosa & Money (2013) ottavat huomioon 3V:n mallin ominaisuuksien lisäksi kompleksisuuden ja arvon. Datan arvolla tarkoitetaan sen käyttökelpoisuutta päätöksenteossa (Kaisler, Armour, Espinosa & Money, 2013). Esimerkiksi pelkkä numeerinen data ilman käyttöarvoa ei ole arvokasta. Datan tulee olla siis sellaista, että sitä voidaan aidosti hyödyntää päätöksenteossa. Vaihtelevuudella tarkoitetaan datan siirtonopeuksien vaihtelevuutta (Gandomi & Haider, 2015). Gandomi & Haider (2015) mukaan datan kompleksisuudella tarkoitetaan, että datalla on lukemattomia lähteitä. Bello-Organ, Jung & Camacho (2016) mukaan datan todenmukaisuudella viitataan informaation oikeellisuuteen ja tarkkuuteen. Big datalle ominaista on, että todenmukaisuuden taso on matala (Jin, Wah, Cheng & Wang, 2015). Tällä tarkoitetaan sitä, että datamassaan liittyy paljon epävarmuutta.

Wu, Zhu, Wu & Ding (2013) esittävät niin sanotun Hacen teoreemaan kuvaamaan big dataa. Hacen teoreeman mukaan big data on volyymiltaan suurta, heterogeenistä, autonomisista hajautetusti valvotuista lähteistä saatavaa dataa, josta pyritään löytämään monimutkaisia ja jatkuvasti kehittyviä suhteita. Big datan heterogeeninen ja monipuolinen luonne johtuu siitä, että eri informaation kerääjät käyttävät omia rakenteitaan datan tallentamiseen ja useiden sovelluksien käyttö johtaa myös monipuoliseen datan esiintymiseen (Wu, Zhu, Wu &

Ding, 2013). Autonomisuutensa takia datalähteiden on mahdollista luoda ja kerätä informaatiota ilman keskitettyä kontrollia (Wu, Zhu, Wu & Ding, 2013). Wu, Zhu, Wu & Ding (2013) mukaan kun big datan volyymi kasvaa, samalla kasvaa myös sen kompleksisuus ja suhteiden määrä datamassan sisällä.

Hacen teoreema tarjoaa erilaisen attribuuttipohjaisen mallin big datan määrittelyyn, jossa on kuitenkin paljon yhtäläisyyksiä edellä mainittujen 3V:n ja sen muunnosten kanssa. Sekä Hacen teoreema, että 3V:n malli pitävät big datan keskeisinä ominaisuuksina volyymia ja datan monipuolisuutta ja heterogeenisuutta. Hacen teoreeman keskeiset eroavaisuudet 3V:n malliin ovat se, että datan nopeutta ei käsitellä osana big datan määrittelyä. Sen sijaan Hacen teoreemassa korostetaan kompleksisuutta, jatkuvaa muutosta ja autonomisuutta. Kompleksisuus ja jatkuva muutos on tunnistettu big datan ominaisuuksiksi myös muissa määrittelyissä lähinnä lisäyksiä puutteelliseksi koettuun 3V:n malliin. Sen sijaan autonomisuuden painottaminen erottaa selkeästi Hacen teoreeman mukaisen määrittelymallin V-mallien kirjosta.

Big datan määrittely on edelleen jossain määrin tulkinnanvarainen. Kuitenkin laaja yhteisymmärrys on syntynyt tutkijoiden välillä siitä, että Big data on volyymiltaan massiivinen määrä dataa, jota voidaan hyödyntää reaaliajassa ja se on rakenteeltaan monipuolista. Lisäksi yhteisymmärrys kirjallisuuden perusteella vaikuttaisi olevan siitä, että Big dataa ei pystytä tehokkaasti käsittelemään perinteisillä datan hallinnan työkaluilla. Laneyn (2001) alun perin esittämää 3V mallia sovelletaan yleisesti big datan määrittelyssä ja tutkijasta riippuen mallia täydennetään muilla attribuuteilla kuten arvo, kompleksisuus, todenmukaisuus ja vaihtelevuus. Hacen Teoreema on 3V:n mallin ja sen muunnoksien tapaan attribuuttipohjainen tapa määrittellä big data. Hacen teoreemassa keskeisiä datan ominaisuuksia ovat datan volyymi, heterogeenisuus, hajautuneisuus, autonomisuus, jatkuva kehittyminen ja kompleksisuus.

2.2 Big data- analytiikka

Big data - analytiikka on kaikkea strukturoimattoman tiedon prosessointia puhelulokista, mobiilipankin transaktioista, online-käyttäjien luomasta sisällöstä, kuten blogikirjoituksista, twiiteistä, verkkohauista ja kuvista, mitkä voidaan muuntaa arvokkaiksi yritystiedoiksi hyödyntäen laskennallisia tekniikoita trendien ja data-aineistojen välille syntyvien kaavojen paljastamiseksi. Big data-analytiikan avulla käsitellään sellaista dataa, joka on liian laajaa, strukturoimattointa ja nopealiikkeistä hallittavaksi perinteisillä metodeilla. (Zakir, Seymour & Berg, 2015).

2.2.1 Big data-analytiikan teknologiat

Del Río, López, Benítez & Herrera (2014) mukaan tietovarastoja on perinteisesti hyödynnetty suurien datamäärien käsittelyyn, mutta jatkuvasti kasvavalla datalle ne eivät kykene toimimaan riittävällä nopeudella. Näin ollen tietovarastot eivät ole big datan käsittelyyn mielekkäitä ja on ollut tarve kehittää uudenlaisia ratkaisuja, mitkä kykenevät käsittelemään big dataa tarpeeksi tehokkaasti. Myös Ducange, Pecori & Mezzina (2017) totesivat, että klassiset ohjelmointi- ja varastointiparadigmat eivät sovellu massiivisten datamäärien käsittelyyn sosiaalisesta mediasta ja sosiaalisista verkostoista. MapReduce on yksi keskeisistä teknologioista joka on kehitetty tähän tarpeeseen. MapReduce on alusta, joka on suunniteltu massiivisten datamäärien prosessointiin rinnakkain sekä skaalautuvien ja vikasietoisien sovellusten kehittämiseen (Del Río ym., 2014). Ducange, Pecori & Mezzinan (2017) mukaan MapReduce onkin yksi keskeisistä teknologioista Big datan käsittelemiseen. Myös Del Río ym., (2014) mukaan MapReduce on yksi suosituimmista alustoista Big datan käsittelemiseen. Hadoop ohjelmointialusta on suosituin avoimen lähdekoodin implementointi MapReducesta, jonka tarkoituksena on helpottaa skaalautuvien ja hajautettujen ratkaisujen kehittämistä (Del Río ym., 2014). Myös Ducange, Pecori & Mezzinan (2017) nostavat sen kaikkein keskeisimmäksi ohjelmointiympäristöksi MapReducen hyödyntämiseen. Ducange, Pecori & Mezzina (2017) nostavat Hadoopin lisäksi myös Sparkin ja NoSQL-tietokannat esimerkeiksi teknologioista, joiden avulla voidaan louhia hyödyllistä tietoa sosiaalisen median big datasta. He kuitenkin toteavat, että kun dataa tuotetaan nopealla tahdilla, se muuttuu nopeasti ja on olemassa tarve reaaliaikaiselle analyysille, Apache Storm ja Apache Samza saattavat olla edellä esitettyjä teknologioita hyödyllisempiä. Apache Storm on hajautettu laskentajärjestelmä reaaliaikaisen datavirran analysointiin (Karunaratne, Karunasekera & Harwood, 2017). Apache Samza on LinkedIn:in kehittämä nykyään avoimen lähdekoodin datavirran prosessointityökalu (Zhuang, Feng, Pan, Ramachandra & Sridharan, 2016). Ducange Pecori & Mezzina (2017) esittävät kappaleessa edellä esiteltyjen teknologioiden olevan relevantteja big datan hallintatyökaluja markkinoinnin toimialalla. Heidän mukaansa edellä mainitut teknologiat yhdessä automaattisen koneoppimi-

sen, luonnollisen kielen käsittelyn, verkkoanalyysin ja tilastojen kanssa mahdollistavat tiedon louhimisen sosiaalisen median big datasta.

2.3 Big data ja analytiikka markkinoinnissa

Big data ja analytiikka on luonut mahdollisuuksia markkinoijille syvällisemmän asiakasymmärryksen saavuttamiseen. Tässä luvussa kuvaamme big datan ja analytiikan roolia markkinoinnin toimialalla. Ensimmäisenä lyhyesti määrittelemme markkinoinnin, sitten käymme läpi markkinoinnin kannalta keskeisiä analytiikan keinoja, ja niiden merkitystä markkinoinnille.

2.3.1. Markkinointi ja markkinointiäly

Markkinoinnissa yleinen tapa tarkastella markkinointipäätöksiä on tarkastella niitä niin sanotun 4 P:n markkinointimixin näkökulmasta. 4 P:n markkinointimix koostuu neljään komponenttiin (tuote, hinta, paikka, markkinointiviestintä) kohdistuvista päätöksistä, jotka yrityksen tulisi ottaa kohdennetusti huomioon ennen tuotteen lanseeraamista (Singh, 2012). Viime aikoina markkinoinnin määritelmää on koettu tarpeen päivittää transaktiokeskeisestä 4P:n mallista asiakassuhteita painottavaan suuntaan. American Marketing Association (2020) määrittelee markkinoinnin organisatorisena toimena ja prosessien joukkona, jonka tarkoituksena on luoda, kommunikoida ja toimittaa arvoa asiakkaille, sekä hallita asiakassuhteita tavoilla, jotka hyödyttävät organisaatiota ja sidosryhmiä.

Markkinointiäly (marketing intelligence) on keskeinen työkalu organisaatioille saada hyödynnettävää tietoa markkinoinnin päätöksenteon tueksi. Fleisher (2008) määrittelee markkinointiällyn systemaattisena, kohdennettuna, ajallisena ja eettisenä yrityksenä kerätä, syntetisoida ja analysoida kilpailua, markkinointia ja ulkoista ympäristöä, tuottaen käytännöllisiä oivalluksia päätöksentekijöille. Fan, Lau & Zhao (2015) määrittelevät markkinointiällyn oivallusten kehittämisenä datasta markkinoinnin päätöksentekoa varten. Esimerkiksi datan louhintatekniikat voivat auttaa markkinoinnin päätöksentekoa edesauttavien hyödyllisten oivallusten saavuttamista eristämällä tai havaitsemalla käyttäytymismalleja tai ennustamalla asiakkaan käyttäytymistä laajojen data-aineistojen pohjalta Fan, Lau & Zhao (2015). Fan, Lau & Zhao (2015) mukaan tyypillinen markkinointiällyn tehtävä on niin sanottu mielipiteiden louhiminen, jota voidaan nykypäivänä toteuttaa laajalti erilaisista datan lähteistä kuten sosiaalisen median datasta, transaktiotiedoista, kyselydatasta ja anturiverkosta saatavasta datasta.

2.3.4 Big data-analytiikan keinot ja rooli markkinoinnissa

Big datan hyödyntämisellä nähdään olevan keskeinen merkitys liiketoiminnassa. Yritykset kykenevät nykyään seuraamaan ja ylläpitämään massiivisia tietokantoja kuluttajien ostokäyttäytymisestä, joissa data on saatavilla reaaliaikaisena ja data on useissa eri muodoissa, kuten numeerisena datana teksti-, ääni - ja videotiedostoina (Chintagunta, Hanssens & Hauser, 2016). Rust & Huang (2014) esittävät viitekehysten, jossa todetaan, että kuluttajilta kerätty big data yhdessä kuluttajien kasvavan kommunikaation kanssa johtaa parempaan ja personoidumpaan palveluun, mikä taas johtaa syvempien asiakassuhteiden muodostumiseen. Tämä johtaa heidän mukaansa yritykselle tuottoisampiin asiakkaisiin ja mahdollistaa näin ollen palvelun laajentamisen. Big datan ollessa reaaliaikaisesti hyödynnettävää, se mahdollistaa markkinoinnin alalla kustomoitujen markkinointityökalujen kehittämisen asiakkaille pohjautuen esimerkiksi asiakkaiden tiedonhakuun, hintavertailuun ja ostoihin (Chintagunta, Hanssens & Hauser, 2016). Näyttääkin siltä, että Big datan avulla voidaan saavuttaa oivalluksia, jotka edesauttavat markkinointipäätöksissä paremmin kohdennetun palvelun muodossa, mikä edesauttaa koko organisaation menestymistä, sekä mahdollistaa palvelutarjoaman laajentamisen. Big data ei ole kuitenkaan arvokasta ilman keinoja käsitellä sitä. Seuraavaksi luomme katsauksen siihen, miten markkinoinnin alalla pyritään big dataa ja analytiikkaa hyödyntämään.

Vuosien varrella poikkeuksellisella nopeudella syntynyt valtava määrä informaatiota, on luonut tarpeen analytiikalle, joka kykenee saavuttamaan mielekkäitä oivalluksia strukturoimattomasta raakadatasta tekstin kuvien ja videoiden muodossa (Shayaa, Jaafar, Bahri, Sulaiman, Wai, Chung & Al-Garadi, 2018). Chong, Ch'ng, Liy & Li (2017) tekemässä tutkimuksessa yhtenä löydöksenä oli, että Big data teknologioiden avulla voidaan tehokkaasti koota suuria data-aineistoja ja avustaa tuotteiden kysynnän ennustamisessa verkkosivuympäristöissä. Myös Kaisler, Armour Espinosa & Money (2013) toteavat että Big datan arvo syntyy kyvystä analysoida datasta käyttökelpoista tietoa. He löysivät tekniseen kirjallisuuteen perustuvan kyselyn pohjalta viisi oleellista tekijää, joiden avulla big data edesauttaa arvonluontia organisaatioissa.

- läpinäkyvyyden luominen tekemällä big data vapaasti saatavaksi liiketoiminnallista ja funktionaalista analyysia varten
- kokeellisen analyysin tukeminen yksittäisissä sijainneissa, joilla pysytään testaamaan päätöksiä ja lähestymistapoja
- Tarkemman markkinasegmentoinnin avustaminen pohjautuen asiakastietoon
- Reaaliaikaisen analyysin ja päätöksenteon tukeminen pohjautuen data-aineistoihin asiakkailta ja sulautetuilta antureilta, jota analysoidaan hienostuneella analytiikalla

- Tietokoneohjattujen innovaatioiden helpottaminen tuotteissa, jotka perustuvat sulautettujen tuotesensoreiden avulla saatavaan informaatioon asiakkaiden vastauksista (responses) ärsykkeisiin

Markkinoinnin alalla hyödynnetään erilaisia analyttisiä keinoja, joilla pyritään saavuttamaan liiketoimintaa edesauttavia oivalluksia big datasta. Näitä ovat esimerkiksi tekstianalytiikka, videoanalytiikka, sosiaalisen median analytiikka, ennustava analytiikka sekä äänianalytiikka. Tekstianalytiikalla tarkoitetaan tekniikoita, joiden avulla erotellaan informaatiota tekstipohjaisesta datasta. Tekstianalytiikan avulla voidaan muuntaa suurista määristä ihmisen luomaa tekstiä merkityksellisiksi yhteenvedoiksi, mitkä tukevat näyttöön perustuvaa päätöksentekoa. Video-analytiikalla tarkoitetaan tekniikoita, joiden avulla monitoroidaan, analysoidaan ja eristetään merkityksellistä informaatiota videostriimeistä. Äänianalytiikalla analysoidaan ja eristetään informaatiota strukturoimattomasta äänidatasta. Ennustavalla analytiikalla tarkoitetaan sellaisia tekniikoita, joiden tarkoituksena on ennustaa tulevaisuuden tapahtumia, pohjautuen historialliseen ja nykyaikaiseen dataan. Ennustavan analytiikan avulla voidaan esimerkiksi ennustaa asiakkaan seuraavia toimia, pohjautuen heidän ostohistoriaansa tai sosiaalisen median käyttäytymiseen. (Gandomi & Haider, 2015). Sosiaalisen median analytiikalla pyritään keräämään, visualisoimaan, seuraamaan, analysoimaan ja kokoamaan sosiaalisen median dataa yleensä kohdesovelluksen spesifien vaatimusten ohjaamana (Zeng, Chen, Lusch & Li, 2010). Ducange, Pecori & Mezzina (2018) mukaan arvokkaan tiedon louhinnasta ja analysoinnista sosiaalisen median suuren datamäärän joukosta on tulossa perusedellytys tehokkaalle ja onnistuneelle markkinointikampanjalle. Markkinointi onkin viime vuosina ollut ensisijainen sosiaalisen median analytiikan hyödyntäjä (Gandomi & Haider, 2015).

Mielipideanalyysi on tekstianalytiikan muoto, jolla on keskeinen merkitys markkinoinnin kannalta. Mielipideanalyysillä (sentiment analysis) pyritään analysoimaan mielipiteitä tekstipohjaisesta materiaalista, tarkoituksena ymmärtää ilmaistuja mielipiteitä ja tunteiden tyyppejä aihealueen eri osia kohtaan (Thet, Na & Khoo, 2010). Lisäksi mielipideanalyysin avulla voidaan arvioida mielipiteitä tuotteista, ihmisistä, tapahtumista, organisaatioista tai aihealueista sekä käyttäjän että käyttäjäryhmän perspektiivistä (Shayaa ym., 2018). Gandomi & Haiderin (2015) mukaan markkinointi onkin keskeinen mielipideanalyysin sovelluskohde. Gandomi & Haider (2015) jakaa mielipideanalyysin kolmeen osaan: dokumenttitason (document-level) analyysiin, lausetason (sentence-level) analyysiin, sekä näkökulmaan perustuvaan (aspect based) analyysiin. Heidän mukaansa dokumenttitason analyysin avulla pyritään ymmärtämään sisältääkö dokumentti negatiivisesti vai positiivisesti latautuneen mielipiteen, lausetason analyysillä yritetään määrittää lauseen latautuneisuus tunteesta kokonaisuudesta ja näkökulmaan perustuvassa analyysissä pyritään tunnistamaan kaikki dokumentissa olevat mielipiteet ja tunnistamaan kokonaisuudessaan näkökohdat, joihin kukin näistä mielipiteistä viittaa.

Big datan ja data-analytiikan rooli markkinoinnissa on ennen kaikkea näyttöön perustuvan päätöksenteon edesauttaminen. Big data ja data-

analytiikan keinot, jolla tätä suurta strukturoimatonta datamassaa käsitellään mahdollistavat markkinoinnin näkökulmasta mahdollisuuden pääsyn käsiksi spesifeihin tietoihin yksilöiden, sekä ryhmien käytöksestä. Parhaimmillaan Big datan ja ja analytiikan avulla markkinoijat pystyvät profiloimaan kuluttajia ja tulkitsemaan kuluttajakäyttäytymistä tavalla, mikä mahdollistaa parempien markkinointipäätöksien tekemisen. Vaikka big data luo suuria mahdollisuuksia, ei sen hyödyntäminen ole kuitenkaan ongelmaton.

3 Big datan ja data-analytiikan haasteet markkinoinnissa

Tässä luvussa käsittelemme kirjallisuuskatsauksen tuloksia vastaamalla tutkimuskysymykseen millaisia haasteita liittyy big datan ja data-analytiikan hyödyntämiseen markkinoinnissa. Kirjallisuuskatsauksessa pyrimme saavuttamaan kokonaisvaltaisen katsauksen siitä, millaisia haasteita big data ja big data-analytiikka luo markkinoinnin kontekstissa. On huomattavaa, että big datan ja siihen liittyvän analytiikan hyödyntämiseen liittyy useita haasteita, jotka ovat oleellista ottaa huomioon riippumatta toimialasta. Näin ollen osa löydetyistä haasteista on relevantteja markkinoinnin lisäksi myös muilla toimialoilla. Pyrimme samalla myös perustelemaan näitä haasteita markkinoinnin näkökulmasta. Lisäksi käymme läpi kirjallisuudesta löydettyjä haasteita, mitkä liittyvät nimenomaan markkinoinnin osa-alueisiin. Kirjallisuudesta olemme rajanneet pois sellaiset artikkelit, joissa Big datan määritelmässä ei olla otettu huomioon datan volyyymia, monipuolisuutta ja nopeutta. Sen sijaan artikkeleita, joissa Big datassa on otettu huomioon myös muita tekijöitä, ei ole rajattu pois.

3.1 Big datan haasteet markkinoinnissa

Big dataan ja siihen liittyvään analytiikkaan liittyy paljon erilaisia haasteita. Kaisler ym., (2013) jakaa Big dataan liittyvät keskeisimmät haasteet varastointiin liittyviin ongelmiin, datan hallintaan liittyviin ongelmiin ja sen prosessointiin liittyviin ongelmiin. Kaisler ym., (2013) arvioivat datan hallinnan olevan haastavin big datan käsittelyn ongelmista, sillä markkinoilla ei ole olemassa tarpeeksi tehokkaita big datan hallinnan ratkaisuja. Myös Jin ym., (2015) toteavat big datan hyödyntämiseen liittyvän merkittäviä haasteita. He toteavat, että osa haasteista johtuu big datan ominaisuuksista, osa nykyisistä analytiikan metodeista ja malleista ja osa nykyisten datan prosessointisyste-

mien rajoitteista. Bello-Orgaz, Jung & Camacho (2016) artikkelissaan löytävät sosiaalisen median big datan haasteita sekä liittyen yksityisyyteen, algoritmeihin, että datan visualisointiin ja fuusioon. Big datan ominaisuuksista erityisesti datan volyyymi ja datan monipuolisuus aiheuttavat kirjallisuuden mukaan haasteita markkinoijille, mutta myös nopeuteen, kompleksisuuteen ja arvoon liittyviä haasteita on löydetty. Seuraavaksi käymme läpi keskeisiä kirjallisuudesta löytyneitä haasteita tarkemmin.

3.1.1 Big datan ja analytiikan tekniset haasteet

Kaisler ym. (2013) mukaan big datan varastointiin liittyen ongelmaksi muodostuu se, että datamäärän kasvaessa ei ole olemassa tarpeeksi hyviä työkaluja massiivisen tiedon varastointiin. Tämän hetkinen levyteknologioiden tallennustila on noin 4 teratavua per levy, joten eksatavu dataa vaatisi noin 25 000 levyä. Vaikka tällainen tietomäärä pystyttäisiin käsittelemään yhdessä tietokonejärjestelmässä, se olisi kykenemätön suoraan liittämään tarvittavan määrän levyjä ja näin ollen ylikuormittaisi nykyiset viestintäverkot. (Kaisler ym., 2013). Datan siirtämiseen liittyvä ongelma, jonka Kaisler ym. (2013) nostavat esiin on, että valtavan datamäärän siirtäminen johtaa tilanteeseen, jossa tietojen lähettäminen varastointi tai -tallennuspisteestä käsittelyyn kestää kauemmin kuin itse datan prosessoiminen. Big datan volyymin kasvaessa sen hyödyntäminen käytännössä vaatii uusia innovaatiota tiedon varastointiin ja siirtämiseen, jotta big datan hyödyt voidaan tehokkaasti valjastaa käytäntöön. Näin ollen jokainen toimiala, joka käsittelee big dataa, törmää näihin datan varastointiin ja siirtämiseen liittyviin haasteisiin. Myös markkinoinnin alalla on syytä olettaa, että samat ongelmat tulevat vastaan massiivisten datamäärien käsittelyssä. Tätä tukee myös se, että sosiaalisen median palvelut tuottavat suuren osan big datasta ja sosiaalisen median analyysiä hyödynnetään paljon markkinoinnin alalla (Ducange, Pecori & Mezzina, 2018 ; Gandomi & Haider, 2015).

3.1.2 Kompleksisuuteen liittyviä haasteita

Jin ym. (2015) mukaan keskeisiä haasteita big datan hyödyntämiselle aiheuttavat datan kompleksisuus, laskennallinen kompleksisuus sekä systeemiin kompleksisuus. Heidän mukaansa big datan kompleksisuudessa haasteena on, että datan kompleksiset tyypit ja rakenteet tekevät sen tulkinnasta, esittämisestä ja ymmärtämisestä huomattavasti haastavampaa, mikä johtaa laskennallisen kompleksisuuden jyrkkään kasvuun verrattuna perinteisiin laskennan malleihin. Toiseksi keskeiseksi ongelmaksi datan kompleksisuudessa he nostavat perinteisten data-analytiikan ja datanlouhinnan tehtävien suorittamisen haastavuuden big dataa käsiteltäessä.

Laskennallisen kompleksisuuden haasteena on, että useista lähteistä saapuvaa, nopeasti muuttuvaa ja volyymitaan suurta dataa ei voida helposti ana-

lysoida, prosessoida ja laskea perinteisillä laskennan menetelmillä, kuten koneoppimisella, informaation haulla tai datan louhimisella (Jin ym., 2015). Myös Liu, Singh & Srinivasan (2016) mukaan liiketoiminnallisten markkinointioivallusten saavuttaminen sosiaalisen median big datasta on erityisen haastavaa, sillä datan suuren volyymin vuoksi normaalit analytiikan tekniikat eivät sovelu laskennallisesti datan käsittelyyn. Lisäksi Jin ym. (2015) toteavat big datan prosessointiin liittyen, että avoimet ja epävarmat suhteet tietojen välillä, riittämättömät otokset ja datan arvon vaihtelevuus aiheuttavat merkittäviä haasteita datan laskettavuudessa sekä uusien laskentaparadigmojen kehittämisessä. L'heureux, Grolinger, Elyamany & Capretz (2017) nostavat big datan ominaisuuksista erityisesti sen volyymin ongelmalliseksi, sillä datan volyymin kasvaessa myös datan laskennallinen kompleksisuus kasvaa, tehden mahdollisesti yksinkertaisistakin operaatioista aikavaativuudeltaan kalliita. Bello-Orgaz, Jung & Camacho (2016) tukevat tätä väitettä. Heidän mukaansa erityisiä haasteita aiheuttavat riittävän hyvien datan louhinnan työkalujen löytäminen valtavan online-datamäärään käsittelyyn.

Automaattisesti rakennettuja tuoteontologioita voidaan hyödyntää tuotteen maineen hallinnassa ja muissa markkinointiälyn tehtävissä (Fan, Lau & Zhao, 2015). Markkinoinnin näkökulmasta Fan, Lau & Zhao (2015) esittävät että laskennallinen kompleksisuus automatisoitujen tuoteontologioiden eristämisessä sosiaalisesta mediasta vaatii uusien laskennallisten metodien kehittämistä datan volyymiin, nopeuteen ja monipuolisuuteen liittyvien ongelmien takia sosiaalisen median big datassa. Big datan laskennallinen kompleksisuus näin ollen on keskeinen haaste myös markkinoijien näkökulmasta, sillä markkinointiälyä hyödyntävät organisaatiot tulevat väistämättä törmäämään big datan laskennallisen kompleksisuuden haasteisiin sekä markkinointiälyn toiminnallisuuksiin, että big datan louhintaan, prosessointiin ja laskentaan liittyen. Fan, Lau & Zhao (2015) toteavat että laskennallinen kompleksisuus aiheuttaa ongelmia lisäksi nykyisten suosittelujärjestelmien skaalautumiselle big dataa käsiteltäessä sekä suositusten luomiselle potentiaalisille asiakkaille reaaliajassa kyseisissä järjestelmissä.

Systemin kompleksisuus muodostuu haasteeksi järjestelmäarkkitehtuurien, laskennallisten viitekehyksien ja prosessointisysteemien suunnittelussa sekä asettaa tiukkoja rajoituksia järjestelmän tehokkuudelle ja energian kulukselle. Edellä mainittujen suunnittelemisen tehokkaille ja energiatehokkaille big datan prosessointialustoille on keskeinen järjestelmän kompleksisuuteen liittyvä haaste. (Jin ym., 2015).

3.1.3 Datan nopeuteen liittyviä haasteita

Markkinoinnin näkökulmasta reaaliaikainen datan analysointi ja hyödyntäminen on tärkeää. Ducange, Pecori & Mezzina (2017) mukaan datan nopeuteen liittyvä haaste on trendien ja mahdollisuuksien tunnistaminen ennen kilpailijoita kilpailuedun saavuttamiseksi. Heidän mukaansa jo kaksi minuuttia myöhemmin tehty datan analysointi voi olla vaarallista korkean kilpailutason

markkinoilla. Fan, Lau & Zhao (2015) nostivatkin datan lähteiden valinnan keskeiseksi puutteeksi big datan tutkimuksessa markkinoinnin kentällä, sillä nykyisillä teknologioilla ei pystytä ajallisesti käsittelemään kaikkea saatavilla olevaa dataa tarpeeksi tehokkaasti. Datan valinta on heidän mukaansa keskeinen päätös markkinointiälyn hallinnassa. Haasteeksi muodostuu data-aineiston valinta niin että se parhaiten tukee päätöksentekoa markkinoinnissa, sillä eri tilanteissa voi olla tarpeen valita erilainen data-aineisto, jotta se tukisi parhaiten reaaliaikaista päätöksentekoa. Myös Kunz ym. (2017) nostavat keskeiseksi tutkimusaiheeksi big datan oikeiden kanavien sekä asiakkaiden sitouttamista edesauttavien lähteiden valitsemisen sosiaalisen median kanavien laajasta tarjonnasta.

3.1.4 Datan volyyymiin liittyviä haasteita

Big datan massiivinen kooltaan jopa zetataavujen kokoinen volyyymi on omiaan aiheuttamaan ongelmia markkinoijille. Ducange, Pecori & Mezzina (2017) nostavat big datan volyymin suureksi haasteeksi massiivisen informaation hallinnan, varastoinnin ja hankkimisen. Lisäksi haasteita tulevaisuudessa luo se, että se mitä nykyään pidetään volyyymiltään big datana, ei välttämättä ole sitä enää tulevaisuudessa (Fan, Lau & Chao, 2015; Gandomi & Haider, 2015). Fan, Lau & Chao (2015) mukaan onkin tärkeää jatkuvasti kehittää tekniikoita, viitekehyksiä ja metodeja, jotta pystytään vastaamaan tulevaisuuden haasteisiin big datan hallinnassa. Datamäärien kasvaessa markkinoijien ja muiden big datan hyödyntäjien on pystyttävä vastaamaan uusiin haasteisiin, joita datan kasvu aiheuttaa, sen lisäksi, että jo nykyisellään big datan volyyymi tuottaa haasteita sen tehokkaaseen hyödyntämiseen.

Big dataa hyödyntäessä haasteeksi muodostuu myös se, että massiivisten datamäärien tehokas prosessointi vaatii laajamittaista rinnakkaista prosessointia, sekä uusia analytiikan algoritmeja ajantasaisen ja hyödynnettävän informaation saavuttamiseksi datasta (Kaisler ym., 2013). Markkinoijien tulisi ottaa huomioon myös virheellisen korrelaation (spurious correlation) mahdollisuus käsiteltäessä volyyymiltään suurta määrää dataa. Virheellisellä korrelaatiolla tarkoitetaan tilannetta, jossa ei-korreloivien muuttujien välillä havaitaan korrelaatiota johtuen data-aineiston suuresta koosta (Gandomi & Haider, 2015). Datamäärän kasvaessa todennäköisyys haetun yhteyden löytämiselle kasvaa riippumatta siitä onko löydös validi (L'heureux ym., 2017). Big datan ollessa volyyymiltään suurta, sen hyödyntäjät ovat alttiita havaitsemaan data-aineistosta virheellisesti korreloivia tapahtumia. Markkinoinnin kontekstissa tämä voi johtaa esimerkiksi väärin päätöksien tekemiseen virheellisten löydösten pohjalta.

Ducange, Pecori & Mezzina (2017) mukaan datan visualisointi suuresta määrästä dataa helposti ymmärrettävään ja tulkittavaan muotoon on haastavaa, mutta erityisen tärkeää käyttäjäkokemuksen näkökulmasta. Myös Jagadish, Gehrke, Labrinidis, Papakonstantinou, Patel, Ramakrishnan & Shahabi (2014) ja Bello-Orgaz, Jung & Camacho (2016) painottavat, että datan visualisointi on keskeinen haaste big data-analyysissä. Jagadish ym. (2014) mukaan on keskeistä

huomioida, että datan loppukäyttäjät kykenevät havainnoimaan data-analyysin tuloksia ilman, että massiivinen määrä dataa tekee tulkinnasta vaikeaa. On huomattavaa, että vaikka laskennallisessa analyysissä on saavutettu merkittäviä edistysaskelia viime aikoina, monissa tilanteissa, ihminen havaitsee sellaisia yhteyksiä helposti, mitä laskennallisten algoritmien on vaikeaa löytää (Jagadish ym., 2014). Tätä tukee myös se, että Kaisler ym. (2013) toteavat, että reaaliajassa syntynyttä dataa ei tunneta ja ymmärretä hyvin, mikä luo haasteita suunnittelijoille datan visualisoinnissa loppukäyttäjien hyödynnettäväksi. Kuten edellisessä luvussa käsitelimme, markkinoinnin alalla big data- analytiikan avulla saavutetulla informaatiolla on keskeinen merkitys esimerkiksi näyttöön perustuvassa päätöksenteossa. Näin ollen niin kauan kuin ihmiset tulkitsevat analytiikasta saatua informaatiota, minkä pohjalta tärkeitä päätöksiä tehdään, on perusteltua nostaa datan visualisointi yhdeksi big datan hyödyntämisen haasteeksi markkinoinnin toimialalla.

3.1.5 Datan heterogeenisuuden liittyviä haasteita

Big datan heterogeeniset datan lähteet, sekä strukturoimattoman datan suuri osuus luovat erityisiä haasteita markkinoijille. Fan, Lau & Chao (2015) nostavat tutkimusta vaativaksi haasteeksi datan lähteiden heterogeenisuuden, kun tutkitaan samaa markkinoinnin ongelmaa käyttäen eri datan lähteitä. Esimerkiksi kun tutkitaan sosiaalisen median käytön vaikutusta ostokäyttäytymiseen, voi olla tarpeen yhdistää sosiaalisen median data transaktiotietoihin. (Fan, Lau & Chao, 2015). Jotta markkinoijat pystyvät hyödyntämään big datan koko potentiaalisen markkinointipäätöksien tukena, tulisi kehittää ratkaisuja datalähteiden tehokkaaseen integraatioon. Myös Bello-Organ, Jung & Camacho, (2016) nostavat datalähteiden fuusion olennaiseksi haasteeksi sosiaalisen median big datan yhteydessä.

Esimerkiksi segmentointi on erityisen haastavaa Big data ympäristössä. Segmentointia vaikeuttaa esimerkiksi se, että dataa on usein tarpeellista käsitellä erillään toisistaan (Fan, Lau & Zhao, 2015). Myös Kunz ym. (2017) nostavat tutkimusta vaativaksi aihealueeksi big data- ratkaisujen kehittämisen kohdistumaan oikeisiin asiakasryhmiin tavalla, joka kasvattaa asiakkaiden sitouttamista (customer engagement). Liu, Singh & Srinivasan (2016) toteavat, että yhtenä keskeisenä haasteena liiketoiminnallisten markkinointioivalluksien saavuttamiseen Big datasta, on strukturoimattoman datan kuten tekstin, videon, kuvien ja äänen käsittely. Myös Ducange, Pecori & Mezzina (2017) mukaan perinteisillä teknologioilla ei pystytä tehokkaasti käsittelemään strukturoitua, semi-strukturoitua ja strukturoimatonta dataa ja näin ollen uusia big datan hallinnan ratkaisuja tarvitaan. Kunz ym. (2017) toteavat tutkimuksessaan, että myös asiakkaiden sitouttamisen (customer engagement) näkökulmasta haasteeksi muodostuu asiakasdatan kerääminen ja yhdistäminen monipuolisista lähteistä, kuten asiakkaan identiteettiin, profiiliin, sitoutumishistoriaan, suositteluihin, päätöksentekoon ja kuluttajakäyttäytymiseen liittyvistä datan lähteistä.

Kuten edellisessä luvussa käsitelimme Big data-analytiikkaa hyödynnettään usein markkinoinnissa mielipideanalyysin muodossa. Mielipideanalyysi on kuitenkin haastavaa toteuttaa Big datalla (Fan, Lau & Chao, 2015). Shayaa ym. (2018) nostavat mielipideanalyysiin liittyviä haasteita big data- ympäristössä sen toteuttamiseen ja oikeanlaisten mielipideanalyysin tekniikoiden kehittämiseen liittyen. Myös mielipideanalyysin tekniikoiden kehittämiseen liittyen ongelmaksi muodostuu datan heterogeeniset ominaisuudet kuten datan rakenteen heterogeenisuus ja datan lähteiden diversiteetti (Shayaa ym., 2018). Perinteisesti mielipideanalyysin luokittelukeinot ovat pohjautuneet dataan yhdestä lähteestä, siinä missä Big datan kontekstissa mielipideanalyysin luokittelukeinon tulee kyetä käsittelemään lähteiltään ja rakenteeltaan heterogeenista dataa (Shayaa ym., 2018). Toisekseen ongelmia mielipideanalyysissä aiheuttaa epävarman, puutteellisen ja hajanaisen datan analysointi. Hajanainen ja epävarma data heikentää mielipiteiden luokittelun tarkkuutta (Shayaa ym., 2018). Tätä tukee myös Ducange, Pecori & Mezzina (2017), joiden mukaan datan todennäköisyyteen liittyvät ongelmat johtuvat usein datasta saatavien oivallusten laatuun ja luotettavuuteen liittyvistä puutteista. Huomattavaa on myös, että big datasta saatavassa tiedossa on puutteellisuutta johtuen yksityisyysrajoituksista (Shayaa ym., 2018). Tämä johtaa siihen, että kaikkea tarpeellista tietoa ei voida välttämättä hyödyntää mielipideanalyysissä vaikuttaen datan hyödynnettävyyteen oleellisesti. Kolmanneksi kun mielipideanalyysia toteutetaan heterogeenisillä datan lähteillä kuten useasta eri sosiaalisen median palvelusta haasteeksi muodostuu mielipideassosiaatioiden muodostaminen olemassa olevien mallien pohjalta (Shayaa ym., 2018). Tällä tarkoitetaan sitä, että datan ollessa hajanaisista lähteistä on siitä selkeiden kokonaiskuvien muodostaminen haastavaa. Shayaa ym. (2018) nostavat esiin tutkimuksessaan mielipideanalyysin hyödyntämiseen liittyviksi haasteiksi päätöksenteon haastavuuden mielipideanalyysistä saatujen tietojen pohjalta, vaikutusten arvioinnin haastavuuden, sekä sosiaalisten bottien vaikutuksen. He nostavat botteihin liittyen kaksi keskeistä haastetta, jotka ovat mielipideanalyysin tulosten vääristyminen ja bottien kyky vaikuttaa mielipiteisiin. Markkinoinnissa on oleellista saada aidosti tietoa yksilöiden mielipiteistä, jotta saatua tietoa voidaan onnistuneesti hyödyntää markkinoinnin päätöksenteossa. Näin ollen bottien vääristämät tulokset voidaan nähdä keskeisenä ongelmana mielipideanalyysissä markkinoinnin alalla. Shayaa ym. (2018) tuovat esiin mielipideanalyysiin liittyen ongelman siitä, että jokainen mielipide ei ole vaikutuksiltaan samanlainen. He nostavat esimerkiksi tilanteen, jossa sosiaalisessa mediassa esitetyn henkilön mielipiteen vaikutuspiiri on suurempi kuin toisen yksilön. Tilanteessa, jossa yhdellä henkilöllä on enemmän vaikutusta muihin ihmisiin kuin toisella, ei heidän mielipiteidensä kokonaisvaikutus ole sama muiden yksilöiden mielipiteisiin. Markkinoinnissa kun pyritään esimerkiksi selvittämään mielipideanalyysin keinoilla ihmisten mielipidettä markkinoitavasta tuotteesta, voi mielipiteiden pitäminen samanarvoisina vääristää todellista käsitystä tuotteen maineesta. Shayaa ym. (2018) tuovat lisäksi esiin ongelman siitä, että vaikka mielipideanalyysistä voidaan saavuttaa tietoa yksilöiden ja joukkojen mielipiteistä, ei se automaattisesti edesauta

oikeiden päätöksien tekemistä. Vaikka mielipideanalyysistä markkinoija saa käsiinsä oleellista tietoa päätöksenteon tueksi, ei se tarkoita, että markkinoija tekisi tulosten pohjalta strategisesti oikeita päätöksiä.

Myös tuotteen maineen hallinta on haastavaa Big dataa käsiteltäessä. Tähän vaikuttaa oleellisesti esimerkiksi online-ympäristössä olevien kuvien suuri koko yhdistettynä kuvien suureen määrään, sekä muun data-aineiston heterogeenisyys, mikä tekee sen analysoinnista haastavaa Fan, Lau & Zhao (2015). Tuotteen maineen hallinnalla pyritään saavuttamaan tärkeitä informaatiota tuotteen maineen tilasta ja analysoinnin haastavuutta voidaan pitää näin ollen yhtenä keskeisenä markkinoinnin haasteena Big data-ympäristössä.

3.1.6 Big datan ja analytiikan arvo markkinoinnissa

Hofacker, Malthouse & Sultan (2016) huomioivat, että Big datan ollessa pääosin historiallista dataa, se vaatii tuekseen teorioita ja malleja, jotta sen avulla voidaan tehokkaasti saavuttaa hyödynnettäviä oivalluksia markkinoinnin päätöksenteossa. Cappa, Oriani, Peruffo & McCarthy (2021) tekivät tuoreessa tutkimuksessaan mielenkiintoisen löydöksen siitä, että hyödynnettävän big datan volyymi itsessään ottamatta huomioon muita big datan ulottuvuuksia vaikuttaa negatiivisesti yrityksen suoriutumiseen, johtuen kustannuksista liittyen datan varastointiin, hallintaan, analytiikkaan, ja vakuutuksiin. Sen sijaan volyymiltaan suuri data, joka on myös datan lähteiltään monipuolista, on heidän mukaansa nettovaikutuksiltaan yrityksen kannalta voitollista. Kyseessä on verrattain tuore tutkimus, joten tutkimustuloksien vahvistus vaatii vielä toistettavuutta löydösten varmistamiseksi. Tutkimustulokset antavat kuitenkin mielenkiintoista näkökulmaa seuraavaksi käsiteltyihin haasteisiin big datan arvoon liittyen. Esimerkiksi Kaisler ym. (2013) ehdottavat, että Big datan hyödyntäjä voi kaiken datan sijaan käsitellä suurta määrää laadukasta dataa, josta edelleen voidaan tehdä tarkkoja ja arvokkaita johtopäätöksiä. Datan laatuun liittyen Kaisler ym. (2013) nostavat haasteiksi esimerkiksi relevantin data-aineiston valinnan, datan luotettavuuden ja tarkkuuden varmistamisen sekä arvokkaan datan määrittämisen vaikeuden. Hofacker, Malthouse & Sultan (2016) käsittelevät big datan haasteita markkinoinnissa erityisesti kuluttajakäyttäytymisen näkökulmasta; he nostivat tutkimuksessaan useita datan arvoon liittyviä haasteita big dataa käsiteltäessä. Datan luotettavuuteen liittyvinä ongelmina he tuovat esiin puolueellisen datan esiintymisen mahdollisuuden. Lisäksi he toteavat, että datan arvo on helposti altis muuttumaan. He nostavat esimerkiksi tilanteen, jossa datasta saadaan tieto, että asiakas on lähellä yrityksen liiketilaa. Tieto muuttuu melko arvottomaksi tilanteessa, jossa asiakkaan sijainti muuttuu kauemmaksi. Lisäksi heidän mukaansa big data ei ole aina yleistettävissä. Vaikka jollakin ajanjaksolla data olisi esimerkiksi mittausvirheetöntä ja ilman otannallisia virheitä, ei löydöksistä voida olettaa, että näin olisi kaikilla ajanjaksoilla (Hofacker, Malthouse & Sultan 2016). He myös toteavat, että kaikki kuluttajan päätöksiin vaikuttavat muuttujat eivät sisälly kerättyihin Big data-aineistoihin.

Näin ollen Big data-analyysistä puuttuu mahdollisesti oleellisia muuttujia esimerkiksi kuluttajan ostopäätökseen vaikuttavista tekijöistä. Hofacker, Malthouse & Sultan (2016) tuovat esiin haasteen siitä, että Big data voi sisältää vanhentunutta tietoa. Tämä haaste on keskeinen sillä, jos datan laatu on liian heikkoa vanhentuneen informaation takia, myöskään datasta saatuihin markkinointioivalluksiin ei voida välttämättä luottaa. Hofacker, Malthouse & Sultan (2016) nostavat big datan erääksi ongelmaksi sen, että Big datasta saamme informaatiota yksilöiden käyttäytymisestä, mutta ei niinkään käyttäytymisen taustalla olevista asenteista ja motivaatiosta. Motivaatioiden ja asenteiden kokonaisvaltainen ymmärtäminen edesauttaisi markkinoijien näyttöön perustuvaa päätöksentekoa oleellisesti kasvattaen hyödynnettävästä datasta saatavaa arvoa.

Fan, Lau & Chao (2015) mukaan tulevaisuudessa on keskeistä tutkia oikeiden data-analytiikan metodien valintaa kuhunkin markkinoinnin tarpeeseen. Heidän mukaansa big data-analytiikkaa hankaloittaa muun muassa epätasapainoinen tiedon jakautuminen, sekä suuri määrä muuttujia data-aineistossa. Datan louhintametodit nykyisellään eivät kykene käsittelemään tarpeeksi tehokkaasti tällaista dataa (Fan, Lau & Chao, 2015). Datan louhinnan ollessa myös markkinoinnin alalla keskeinen työkalu markkinointioivallusten saavuttamiseen, tulisi markkinoijien kiinnittää tähän haasteeseen erityistä huomiota.

Markkinointiällyn tullessa keskeiseksi kilpailueduksi asiakasymmärryksen ja tuotesuunnittelun alalla, keskeinen tutkimushaaste markkinoinnin organisatiolle on rahallisten investointien tasapainottaminen markkinointiällyn tekniikoihin (Fan, Lau & Zhao, 2015). Markkinointiällyn ja siihen liittyvän big datan hyödyntäminen ei ole edullista, mutta niiden potentiaali markkinoinnin osalualueilla kilpailuedun luojana tulisi ottaa huomioon investointipäätöksiä tehdessä.

3.1.7 Big datan omistajuuteen ja yksityisyyteen liittyviä haasteita

Big datan hyödyntämisen yksi keskeinen markkinoinnin ongelma on se, että big dataa hyödyntävät organisaatiot eivät välttämättä omista itse hyödynnettävää dataa. Big data usein kerätään ulkoisilta alustoilta ja osapuolilta, kuten Googlelta, Facebookilta ja pilvipalveluihin pohjautuvista sähköpostijärjestelmistä. Haasteeksi näin ollen muodostuu Big datan saatavuus, kerääminen ja sen integroiminen yrityksen asiakkuudenhallintajärjestelmän kanssa muodostaessa käsitystä asiakkaasta. (Kunz ym., 2017). Kaisler ym. (2013) mukaan datan laadun varmistaminen perustuu useimmiten enemmän puuttuvaan dataan kuin jokaisen datasta saadun yksikön validoimiseen, johtuen big datan suuresta volyyymistä. Heidän mukaansa uusia laadun varmistamisen ja validoimisen metodeja tarvitaan big datan käsittelyssä. He toteavat, että tulevaisuudessa datan ja informaation alkuperä tulee olemaan keskeinen haaste big dataan liittyen. Lisäksi Big datan omistajuuteen liittyy ongelmia niin eettisestä kuin laillisestakin näkökulmasta (Kunz ym., 2017). Eettisistä ongelmista Hofacker, Malthouse & Sultan (2016) nostavat esimerkkinä kuluttajien epätietoisuuden datan keruusta.

Datan lähteet voivat sisältää esimerkiksi dataa online-käyttäytymisestä, sosiaalisen median osallistumisesta, sijaintidataa, mobiilimastoista saatavaa dataa ja esineiden internetin avulla kerättyä intiimiä dataa, mikä nostaa riskiä yksityisyyteen liittyviin huoliin. (Hofacker, Malthouse & Sultan, 2016). Bello-Organ, Jung & Camacho (2016) toteavat, että sosiaalisen median big datassa havaittuja yksityisyyteen liittyviä haasteita ovat käyttäjän yksityisyyden suojeleminen, sosiaaliseen identifiointiin liittyvät ongelmat, ja yksityisyyden säilyttävien tekniikoiden testaaminen ja arviointi. Sosiaalisessa identifiointissa heidän mukaansa ongelmallista on datan yhdistäminen saatavilla olevista lähteistä ilman henkilön tunnistamista.

3.1.8 Viitekehys big datan ja analytiikan haasteista markkinoinnissa

Kirjallisuuskatsauksen tuloksen pohjalta luomme viitekehyksen markkinoijille big datan ja analytiikan haasteista markkinoinnin alalla (kuvio 1). Viitekehyksessä haasteet on jaettu neljään alakategoriaan, mitkä ovat markkinoinnin toimiin suoraan liittyvät haasteet, big datan ominaisuuksiin liittyvät yleiset haasteet, yksityisyyteen ja omistajuuteen liittyvät haasteet, sekä tekniset yleiset haasteet. Markkinoinnin toimiin suoraan liittyviin haasteisiin kokosimme sellaisia haasteita, jotka olennaisesti liittyvät johonkin markkinoinnin osa-alueeseen. Big datan ominaisuuksiin liittyviin yleisiin haasteisiin kokosimme markkinoinnin alalla relevantteja big datan volyyymiin, monipuolisuuteen, arvoon, nopeuteen ja kompleksisuuteen liittyviä haasteita. Yksityisyyteen ja omistajuuteen liittyviin haasteisiin liittyen nostimme keskeisiä haasteita, jotka liittyvät big datan ulkoiseen omistukseen sekä kuluttajan yksityisyyteen. Teknisistä yleisistä haasteista nostimme viitekehyksessä esille big datan hyödyntämistä vaikeuttavia teknisiä haasteita. Viitekehyksessä pyrimme luomaan läpileikkauksen niistä haasteista, joita markkinoinnin alalla tulee ottaa huomioon big dataan ja big data-analytiikkaan liittyen.

| Markkinoinnin toimiin suoraan liittyvät haasteet | Big datan ominaisuuksiin liittyvät yleiset haasteet | Tekniset yleiset haasteet | Yksityisyyteen ja omistajuuteen liittyvät haasteet |
|---|---|---------------------------------------|---|
| Tuoteontologioiden eristäminen Big datasta | Virheellinen korrelaatio | Datan hallinta | Ulkoisista lähteistä tulevan datan saatavuus, kerääminen ja integrointi CRM-järjestelmän kanssa |
| Suosittelujen luominen reaaliajassa suositelujärjestelmissä | Algoritmien kehittäminen | Datan siirtäminen | Yksilöiden huolet yksityisen datan käsittelyssä |
| Big data ei sisällä kaikkia asiakkaan ostopäätökseen vaikuttavia tekijöitä | Laskennallinen kompleksisuus ja laskentaparadigmojen kehittäminen | Datan varastointi | Kuluttajan yksityisyyden suojeleminen data-analysissä |
| Data-analytiikan metodien valinta markkinoinnin tarpeeseen | Arvokkaan datan määrittäminen | Big datan prosessointi ja analysointi | Datan yhdistäminen eri lähteistä ilman yksilön tunnistamista |
| Suosittelujärjestelmien skaalautuminen | Systeemin kompleksisuus | | Yksityisyyden säilyttävien teknikoiden testaaminen ja arviointi |
| Markkinointilähteen datan lähteiden valinta | Puolueellinen ja vanhentunut data | | |
| Investointien tasapainottaminen markkinointilähteen tekniikoihin | Heterogeenisen datan käsittely | | |
| Trendien ja mahdollisuuksien tunnistaminen Big datasta ennen kilpaillijoita | Datan volyymin kasvu tulevaisuudessa | | |
| Tuotteen maineen hallinta Big datalla | Datan visualisointi | | |
| Segmentointi Big data-ympäristössä | Datan luotettavuuden arviointi ja tarkkuuden arviointi | | |
| Liiketoiminta-avoillustusten saavuttaminen heterogeenisestä datasta | Datan arvon suhteellisuus | | |
| Mielipideanalyysi Big datalla | | | |
| Mielipideanalyysin teknikoiden kehittäminen | | | |
| Datalähteiden heterogeenisyys tutkittaessa samaa markkinointiongelmää käyttäen eri datan lähteitä | | | |
| Asiakkaiden sitouttaminen monipuolisilla datan lähteillä | | | |

Kuvio 1 : Viitekehys big datan ja analytiikan haasteista markkinoinnissa.

Johtopäätökset ja jatkotutkimusaiheet

Tässä tutkimuksessa pyrittiin kirjallisuuskatsauksen pohjalta löytämään haasteita liittyen big datan ja analytiikan hyödyntämiseen markkinoinnin alalla. Kirjallisuuskatsauksessa havaittiin myös, että big datan ja analytiikan haasteet markkinoinnissa ovat tutkimusalueena vielä melko vähän tutkittu. Kirjallisuudesta löytyy paljon tutkimusta big datan haasteista yleisesti, mutta markkinoinnin kontekstissa tutkimusaihe ei ole saavuttanut vielä kovin suurta suosiota. Jotta big datan ja analytiikan potentiaali voidaan valjastaa luomaan arvoa markkinoinnin organisaatioille mahdollisimman tehokkaasti, tulisi aihealuetta tutkia vielä lisää.

Ensimmäisessä sisältöluvussa määriteltiin kirjallisuuskatsauksen kannalta keskeisiä käsitteitä ja aihepiirejä. Kävimme läpi yleisimpiä big datan määritelmiä ja niiden yhtäläisyyksiä, josta saimme myös täsmennystä varsinaisen kirjallisuuskatsauksen kirjallisuuden valintaan. Big datan ollessa monitulkintainen käsite, jouduimme jättämään kirjallisuuskatsauksen ulkopuolelle artikkeleita, joissa big data on käsitelty liian suppeasti tarkoitusperiimme. Lisäksi käsitelimme data-analytiikan teknologioita ja keinoja, joilla big dataa hyödynnetään. Käsitelimme luvussa myös markkinoinnin ja markkinointiälyn käsitteitä ja loimme katsauksen siihen, miten big dataa ja analytiikkaa hyödynnetään markkinoinnin alalla.

Toisessa ja viimeisessä sisältöluvussa pyrimme kirjallisuuskatsauksen avulla löytämään oleelliset haasteet liittyen big dataan ja analytiikkaan markkinoinnin toimialalla. Kirjallisuudesta kävi ilmi, että big dataan ja siihen liittyvään analytiikkaan liittyy paljon toimialasta riippumattomia haasteita. Nämä haasteet vaativat vielä lisää tutkimustyötä, jotta big dataa voitaisiin mahdollisimman tehokkaasti hyödyntää organisaation arvonluojana. Lisäksi kirjallisuudesta löytyi haasteita nimenomaan markkinoinnin kontekstissa, joihin markkinointien kannattaa kiinnittää erityistä huomiota. Keskeisiä haasteita löytyi niin big datan hallintaan, ominaisuuksiin, prosessointiin kuin varastointiinkin liittyen. Lisäksi loimme kirjallisuuskatsauksen tulosten pohjalta lisäksi viitekehyksen, jossa kootaan löydetty haasteet visuaaliseen ja selkeään muotoon.

Tutkimuksen tuloksista on hyötyä sekä akateemisesta, että käytännön näkökulmasta. Esittämiemme haasteiden avulla tutkijat saavat kattavan yleiskäsityksen big datan haasteista markkinoinnin alalla luoden näin ollen oivallisen pohjan jatkotutkimukselle esittelemiemme haasteiden pohjalta. Markkinoijille tutkimuksen tulokset edesauttavat big datan ja analytiikan haasteiden ymmärtämistä käytännössä. Markkinoijille selviää keskeisimmät big datan- ja analytiikan haasteista markkinoinnin kontekstissa, mikä mahdollistaa haasteiden ennaltaehkäisyn, ymmärtämisen sekä ratkaisujen keksimisen haasteiden ehkäisemiseksi. Lisäksi tulokset auttavat ymmärtämään big datan ja siihen liittyvän analytiikan rajoitteita.

Kirjallisuuskatsauksemme sisältää myös rajoitteita, jotka tulee ottaa huomioon. Ensiksi kirjallisuuskatsausta tehdessä havaitsimme, että markkinoinnin kontekstissa big datan haasteita on tutkittu vain vähän. Markkinoinnin ollessa laaja-alainen tutkimusalue on näin ollen syytä olettaa, että tutkimuksemme ei kata suurta osaa markkinoinnin spesifeihin osa-alueisiin liittyvistä haasteista ja jatkotutkimusta on syytä tehdä aihealueen kokonaisvaltaiseksi ymmärtämiseksi. Toiseksi rajoitteeksi nostamme tutkimusaiheen laajuuden, sillä kuten tutkimustuloksista huomataan, voi big datan ja analytiikan haasteita tarkastella monesta eri perspektiivistä. Tämän kirjallisuuskatsauksen ensisijainen tarkoitus onkin luoda katsaus yleisiin haasteisiin big datan ja analytiikan hyödyntämisessä markkinoinnissa. Jatkotutkimusaiheena olisi keskeistä paneutua spesifimmin big datan ja analytiikan vaikutuksiin markkinoinnin eri osa-alueilla. Jatkotutkimusaiheiksi ehdottaisin tutkittavaksi big datan haasteita segmentoinnissa ja big datan haasteita kokonaisvaltaisen asiakasymmärryksen saavuttamisessa. Mieli-pideanalyysin liittyen olisi mielenkiintoista tutkia yleisimpiä tekijöitä, jotka vääristävät mieli-pideanalyysin tuloksia big dataa hyödyntäessä.

LÄHTEET

- American Marketing Association (2021, 16.3). Definitions of Marketing. Retrieved March 16, 2021, from <https://www.ama.org/the-definition-of-marketing-what-is-marketing/>
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45-59.
- Cappa, F., Oriani, R., Peruffo, E., & McCarthy, I. (2021). Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance. *Journal of Product Innovation Management, 38*(1), 49-67.
- Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Marketing science and big data.
- Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research, 55*(17), 5142-5156.
- Del Río, S., López, V., Benítez, J. M., & Herrera, F. (2014). On the use of mapreduce for imbalanced big data using random forest. *Information Sciences, 285*, 112-137.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*.
- De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.
- Ducange, P., Pecori, R., & Mezzina, P. (2018). A glimpse on big data analytics in the framework of marketing strategies. *Soft Computing, 22*(1), 325-342.
- Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research, 2*(1), 28-32.
- Fleisher, C. S. (2008). Using open source data in developing competitive and marketing intelligence. *European journal of marketing*.

- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big data and consumer behavior: Imminent opportunities. *Journal of consumer marketing*.
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, 2(2), 59-64.
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *2013 46th Hawaii international conference on system sciences* (pp. 995-1004). IEEE.
- Karunaratne, P., Karunasekera, S., & Harwood, A. (2017). Distributed stream clustering using micro-clusters on Apache Storm. *Journal of Parallel and Distributed Computing*, 108, 74-84.
- Kumar, V., Chattaraman, V., Neghina, C., Skiera, B., Aksoy, L., Buoye, A., & Henseler, J. (2013). Data-driven services marketing in a connected world. *Journal of Service Management*, 24(3), 330-352.
- Kunz, W., Aksoy, L., Bart, Y., Heinonen, K., Kabadayi, S., Ordenes, F. V., ... & Theodoulidis, B. (2017). Customer engagement in a big data world. *Journal of Services Marketing*.
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety* (). META Group .
- L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *Ieee Access*, 5, 7776-7797.
- Liu, X., Singh, P. V., & Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science*, 35(3), 363-388.
- Rust, R. T., & Huang, M. H. (2014). The service revolution and the transformation of marketing science. *Marketing Science*, 33(2), 206-221.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information systems*, 47, 98-115.
- Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., ... & Al-Garadi, M. A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, 6, 37807-37827.

- Singh, M. (2012). Marketing mix of 4P's for competitive advantage. *IOSR Journal of Business and Management*, 3(6), 40-45.
- Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), 97-107.
- Zakir, J., Seymour, T., & Berg, K. (2015). Big Data Analytics. *Issues in Information Systems*, 16(2).
- Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, 25(6), 13-16
- Zhuang, Z., Feng, T., Pan, Y., Ramachandra, H., & Sridharan, B. (2016, June). Effective multi-stream joining in apache samza framework. In *2016 IEEE International Congress on Big Data (BigData Congress)* (pp. 267-274). IEEE.