

Anni Koivunen

BIG DATAN HYÖDYNTÄMISEN HAASTEET JA MAHDOLLISUUDET FINANSSISEKTORILLA



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2024

TIIVISTELMÄ

Koivunen, Anni

Big Datan hyödyntämisen haasteet ja mahdollisuudet finanssisektorilla

Jyväskylä: Jyväskylän yliopisto, 2024, 30 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja: Mehtälä, Saana

Digitalisaation seurauksena käyttöön otettujen digitaalisten laitteiden ja järjestelmien määrä on kasvanut räjähdysmäisesti. Uusia sovelluksia syntyy jatkuvasti erinäisiin käyttötarkoituksiin eri aloille. Nämä laitteet, järjestelmät sekä sovellukset tarvitsevat jonkinlaista dataa toimiakseen. Lisäksi edellä mainitut asiat luovat paljon uutta dataa, jota analysoimalla voidaan tehdä uusia johtopäätöksiä sekä tulkintoja. Suuri määrä dataa ei ole itsessään vielä tehokas työkalu. Se vaatii ympärilleen osaamista sekä ymmärrystä, jolloin siitä voidaan tulkita kuvioita ja kaavoja. Data-analyysin avulla datasta saadaan muokattua hyvinkin toimiva työkalu moniin erilaisiin tarpeisiin. Datan oikeanlainen analysointi sekä hyödyntäminen mahdollistavat yrityksille merkittävän kilpailuedun toisiinsa nähden. Datan hyödyntämisestä rahoitusmarkkinoilla voidaan saada suuria hyötyjä niin kansallisella kuin kansainväliselläkin tasolla. Parhaillaan tutkimuksissa on enustettu oikeanlaisen datan valitsemisen, keräämisen sekä analysoinnin ehkäisevän jopa talouskriisejä. Big Datan hyödyntämisestä ja analysoinnista rahoitusmarkkinoiden kontekstissa löytyy verrattain vain vähän tutkimusta. Täten tässä kirjallisuuskatsauksessa pyritään kasaamaan yhteen tutkimusta Big Datan haasteista sekä mahdollisuuksista finanssisektorin kontekstissa sekä tuomaan esiin ja käsittelemään näiden tutkimusten tuloksia. Tämä kandidaatin työ on toteutettu kirjallisuuskatsauksena ja se kokoaa yhteen tutkimuksia aiheesta. Kandidaatin työn pohjana käytettiin tieteellisiä tutkimuksia aiheen ympärillä. Keskeisimmät tutkimustulokset Big Datan hyödyntämisen mahdollisuuksiin finanssialalla ovat muun muassa tehokkaampi sijoitustoiminta sekä asiakaskokemuksen parantaminen. Big Dataan liittyviä uhkia ovat puolestaan tietoturvaohut ja sääntelyn ja lainsäädännön vaikutus Big Datan hyödynnettävyyteen.

Asiasanat: Big Data, finanssisektori, data-analytiikka

ABSTRACT

Koivunen, Anni

Challenges and Possibilities of Utilizing Big Data in Financial Industry

Jyväskylä: University of Jyväskylä, 2024, 30 pp.

Information System Science, Bachelor's Thesis

Supervisor: Mehtälä, Saana

Digitalisation has led to an explosion in the number of digital devices and systems deployed. New applications are constantly emerging for a variety of uses in different sectors. These devices, systems and applications need some form of data to function. In addition to the above, a lot of new data is being created, which can be analyzed to make new management decisions and interpretations. A large amount of data is not in itself an effective working fish. It requires knowledge and understanding around it, so that patterns and graphs can be interpreted from it. Data analysis can be used to turn data into a highly functional tool for a wide range of needs. The correct analysis and use of data can give companies a significant competitive advantage over others. The use of data in financial markets can bring major benefits at both national and international level. Studies have predicted that the right choice, collection and analysis of data can even prevent financial crises. There is relatively little research on the use and analysis of Big Data in the context of financial markets. Thus, this literature review aims to bring together research on the challenges and opportunities of Big Data in the context of the financial sector and to highlight and discuss the results of these studies. This bachelor thesis has been conducted as a literature review and it brings together studies on the topic. The thesis was based on scientific studies around the topic. The main research findings on the potential of using Big Data in the financial sector include more efficient investment activities and improved customer experience. Threats associated with Big Data include security threats and the impact of regulation and legislation on the usability of Big Data.

Keywords: Big Data, Finance Industry, Big Data in Finance

TAULUKOT

TAULUKKO 1 Tavun monikerrat kymmenjärjestelmässä	9
TAULUKKO 2 Big Datan lähteet terveydenhuollossa (Ranjan, 2019).....	13
TAULUKKO 3 Big Datan lähteet finanssipalveluissa (Ranjan, 2019).....	13
TAULUKKO 4 Big Datan lähteet lentoyhtiöillä (Ranjan, 2019).....	13

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT JA TAULUKOT

1	JOHDANTO.....	6
2	BIG DATA.....	8
2.1	Big Datan määritelmä.....	8
2.2	Big Datan 3V-malli versus 10V-malli.....	10
2.3	Big Datan kanavat.....	12
2.4	Big Data-analytiikka.....	14
3	FINANSSISEKTORI.....	16
3.1	Finanssisektorin toimijat ja niiden operaatiot.....	17
3.2	Moderni rahoitusjärjestelmä.....	18
4	BIG DATA FINANSSISEKTORILLA.....	20
4.1	Big Dataan liittyvät haasteet finanssisektorilla.....	22
4.2	Big Dataan liittyvät mahdollisuudet finanssialalla.....	23
5	YHTEENVETO.....	25
	LÄHTEET.....	28

1 JOHDANTO

Datan määrä jatkaa kasvamistaan eksponentiaalisesti. Goldstein ym (2021) valaisivat artikkelissaan, että kahdessa päivässä syntyvä datamäärä vastaa ihmiskunnan koko ennen vuotta 2003 luomaa datamäärää. Datan määrän suuri kasvu on osaltaan inspiroinut innovaatioita aiheen ympärillä sekä herättänyt yritykset ymmärtämään datan analysoinnin hyödyt sekä tärkeyden. Nämä innovaatiot ovat johtaneet monipuolisiin mahdollisuuksiin hyödyntää dataa eri aloilla alan tarpeisiin nähden. Jatkuvan datan lisääntymisen myötä dataa voidaan valjastaa jatkuvasti uusiin tarpeisiin. Data ja sen oikeanlainen käsittely on luonut paljon mahdollisuuksia esimerkiksi liiketoiminnan tehostamiseksi, lisäarvon luomiseksi asiakkaille, kuluttajakäyttäytymisen ymmärtämiseksi sekä terveydenhuollon tukemiseksi.

Tämä kandidaatin työ käsittelee Big Datan hyödyntämistä finanssisektorin kontekstissa sekä siihen liittyviä uhkia ja mahdollisuuksia. Datan määrän lisääntyessä ja monipuolistuessa dataa on opittu käyttämään uusilla tavoilla myös rahoitusmaailmassa. Rahoitusmarkkinoilla on pitkät juuret datan hyödyntämisessä liiketoiminnan kasvattamiseksi. Pankkisektori on hyödyntänyt dataa toiminnassaan jossain määrin jo 1980-luvulta lähtien. (Gasser ym, 2017)

Pitkistä juurista huolimatta Big Datasta ja sen hyödyntämisestä finanssialalla on alkanut löytymään enemmän tutkimusta vasta lähivuosina. Tämä on huomattu ja tuotu esille myös muissa kirjallisuuskatsauksissa, jotka käsittelevät kyseenomaista aihetta. Tämä kirjallisuuskatsaus kasaa yhteen tutkimuksia, jotka yhdistävät Big Datan hyödyntämisen sekä finanssisektorin. Aiheesta tarvitaan lisää tieteellistä tietoa, sillä Big Datan hyödyntämisellä finanssisektorilla voi olla hyvinkin suuret vaikutukset talouteen, sen edistämiseen sekä talouden riskien havaitsemiseen. Hyötyjen lisäksi Big Datan hyödyntämiseen finanssisektorilla liittyy myös uhkia sekä riskejä. Täten tässä kandidaatin tutkielmassa perehdytään Big Datan hyödyntämisen haasteisiin ja mahdollisuuksiin finanssisektorilla. Tämä kirjallisuuskatsauksen tarkoitus on koota kasaan aiheesta löytyviä tutkimuksia sekä niiden tuloksia ja helpottaa aiheen ympärillä jatkossa tehtäviä tutkimuksia. Tässä kandidaatin työssä vastataan seuraaviin tutkimuskysymyksiin:

- Mitä haasteita Big Datan hyödyntämisessä on finanssisektorilla?
- Mitä mahdollisuuksia Big Datan hyödyntämisellä on finanssisektorilla?

Tämä kandidaatin työ on koostettu kirjallisuuskatsauksena. Tutkielmaan on etsitty lähteitä pääosin Google Scholarista, JYKDOK:sta, IEEE Xplorerista, Science Directistä, ProQuestista sekä Jyväskylän yliopiston kirjastosta. Hakusanoina on käytetty esimerkiksi seuraavia sanoja: "Big Data", "Big Data in Finance", "Big Data analytics", "FinTech", "Big Data in Banking" sekä "finance industry". Lähteiden laatua on arvioitu Julkaisuforumin julkaisukanavahauulla, joka arvioi artikkeleiden julkaisijan luotettavuutta. Kirjallisuuskatsauksessa on pyritty käyttämään lähteitä, joiden julkaisija on saanut julkaisukanava haussa luotettavan arvon 2 tai 3. Joukossa on myös muutamia tason 1 lähteitä, joista otettuja viitteitä on arvioitu hyvin kriittisesti. Tason 1 lähteitä on käytetty vain muutamisiin kohtiin tukemaan väitteitä, joiden voidaan olettaa pitävän paikkaansa myös aiemmin yliopisto-opinnoissa opittujen asioiden perusteella. Lähteitä valittaessa on pyritty valitsemaan mahdollisimman hyvän arvioinnin saaneita artikkelien ja tutkimuksien julkaisijoita.

Edellä mainitun lisäksi julkaisuforumista löytämättömiä lähteiden luotettavuutta on arvioitu kriittisesti julkaisupaikan, julkaisun kirjoittajan ja lähteiden päivämäärän, viittausmäärän sekä tutkimuksen rakenteen ja metodien perusteella. Tutkielman tietoperustaa on pyritty rakentamaan systemaattisesti käyttäen tuoreita lähteitä. Tällä pyritään varmistamaan, että tutkielman perustana käytettävät tiedot ja teoriat ovat mahdollisimman ajantasaisia. Joukossa on myös muutamia vanhempia lähteitä, jotka edustavat edelleen alansa merkittäviä ja relevantteja viitekehysjä, rakenteita ja malleja. Lähteiden luotettavuuden arviointiin vaikuttaa myös viittausten määrät. Useita satoja tai tuhansia viittauksia saaneiden julkaisuiden viittausmäärät indikoivat luotettavuutta ja vaikutusvaltaisuutta omalla alallaan.

Tutkielman rakenne on seuraavanlainen. Tutkielman luvussa kaksi määritellään käsitettä Big Data. Big Datan lisäksi käsitellään myös muita käsitteitä, jotka liittyvät läheisesti Big Dataan ja auttavat ymmärtämään siihen liittyviä ominaisuuksia, malleja sekä prosesseja ja kokonaiskuvaa Big Datasta. Luvussa kolme käsitellään sekä määritellään nykyaikaista finanssisektoria. Kyseinen käsite ja sen sisältö on valtava, jonka vuoksi luvussa keskitytään käsittelemään finanssisektorin toimijoita, niille ominaispiirteisiä prosesseja, finanssisektorin roolia rahoituksen välittäjänä sekä nykyaikaisen rahoitusjärjestelmän ominaispiirteitä. Kun molemmat tutkimuksen kannalta oleelliset kokonaisuudet on määritelty, siirrytään neljännessä luvussa käsittelemään Big Datan hyödyntämistä finanssisektorilla. Kyseisessä luvussa vastataan tutkielman tutkimuskysymyksiin ja tuodaan esille Big Datan hyödyntämisen haasteet sekä mahdollisuudet finanssialan kontekstissa. Viimeisessä luvussa tiivistetään tutkielman sisältö sekä keskeiset asiat.

2 BIG DATA

Big Data -termissä sana "Big" saattaa olla harhaan johtava, sillä se tuo mieleen jotakin suurikokoista. Suuren datan koon sijaa termillä viitataan suureen datan määrään sekä määrältään massiivisiin datasetteihin. Tässä luvussa käsitellään ja määritellään Big Dataa sekä siihen läheisesti liittyviä käsitteitä. Tässä tutkielmassa Big Datan määrittelyyn käytetään Langley'n (2001) 3V-mallia, jossa Big Dataa lähestytään kolmesta eri näkökulmasta. Big Datan 3V-mallia on myöhemmin jatkojalostettu ja sen pohjalta on tehty paljon myös muita monimutkaisempia Big Dataa käsitteleviä malleja. Nämä jatkojalostetut mallit Big Datasta ovat kehittyneet Big Datan yleistymisen sekä datan arvon ymmärryksen kehittyessä. Samalla Big Datan käsitteen tarve on kasvanut. Luvussa otetaan kantaa siihen, miksi Big Datan 3V-malli on sopivin tämän kirjallisuuskatsauksen näkökulmaan. Big Datan mallien tutkimisen jälkeen keskitytään Big Datan lähteisiin eli siihen, mistä kaikkialta voimme saada Big Dataa. Kun Big Data on saatu kerättyä lähteistään, sitä tulee analysoida, jotta siitä saadaan mahdollisimman suuri hyöty käyttäjälleen. Täten luvun viimeisessä alaluvussa pureudutaan Big Data -analyysiin.

2.1 Big Datan määritelmä

Big Datalle löytyy tieteellisistä artikkeleista useampia hieman erilaisia määritelmiä. Suurimmassa osassa artikkeleilta Big Data määritellään suurin piirtein seuraavasti: Big Data tarkoittaa valtavia tietomassoja, jotka ovat liian suuria ja monimutkaisia perinteisten tietojenkäsittelymenetelmien ja tietokantajärjestelmien käsiteltäväksi (Sagiroglu & Sinanc, 2013). Hung (2016) puolestaan sanoo kirjassaan Big Datan olevan valtava määrä dataa, jota voidaan generoida monista lähteistä ja oikein analysoituna ja visualisoituna käyttää tukemaan päätöksentekoa. Näitä tietomassoja eli Big Dataa generoidaan jatkuvasti useista lähteistä, kuten esimerkiksi sosiaalisesta mediasta, älypuhelinsovelluksista, verkkosivuista, pilvipalveluista ja sensortechnologiasta (George, Haas & Pentland, 2014). Lähes kaikki mitä teemme käyttäen teknologiaa, jättää jälkeensä Big Dataa.

Konkreettisia esimerkkejä Big Datasta edellä mainittujen lisäksi on klikkaukset internetissä, mobiilitransaktiot, lokitiedot ja käyttäjien luoma sisältö eri sosiaalisen median alustoilla. Big Dataa ei voida käsitellä perinteisillä tietokantajärjestelmillä, vaan sen käsittelyyn tarvitaan erikoistuneita tietojenkäsittelymenetelmiä, joita kehitetään jatkuvasti.

Yksinkertaisin malli Big Datan määrittelemiseen sekä ymmärtämiseen on Langley'n (2001) kehittämä 3V-malli. Mallin mukaan Big Datan rakenne voidaan määrittellä kolmiosaiseksi sisältäen seuraavat ominaisuudet: **volyyymi** (volume), **vauhti** (velocity) ja **monipuolisuus** (variety). 3V-mallin julkaisemisen jälkeen ja yhteiskunnan sekä liiketoimintamaailman muututtua dataohjautuvammaksi teknologian kehityksen myötä, Big Datan ympärille on keksitty paljon uusia entistä mallia monimutkaisempia malleja. Näitä malleja on kehitetty 3V-mallin lisäksi aina 4V-mallista jopa 10V-malliin saakka. Eri malleja ja niiden relevanttia käyttöä tämän kirjallisuuskatsauksen näkökulmasta käsitellään luvussa 2.2.

3V-mallin yhteydessä sana **volyyymi** (volume) viittaa suuriin datamassoihin. Big Datan määrä on niin massiivinen, että sitä on vaikeaa ymmärtää. Vertailun vuoksi Sagiroglu ja Sinanc (2013) toteavat, että yhdessä yksityishenkilön tietokoneessa on noin 500 gigatavua. Oussous ym. (2018) puolestaan kertovat tutkimuksessaan, että vuoteen 2015 mennessä digitaalisen datan määrä oli kasvanut 8 zettatavuun (Zettabyte). Samassa tutkimuksessa arvioidaan datan määrän saavutettavan 20 zettatavua ja kasvavan 400-kertaiseksi vuodesta 2016 vuoteen 2020 mennessä. Tavujen suuruusluokat on avattu alla olevassa taulukossa.

TAULUKKO 1 Tavun monikerrat kymmenjärjestelmässä

Nimi	Lyhenne	Arvo
kilotavu	kB	10^3
megatavu	MB	10^6
gigatavu	GB	10^9
teratavu	TB	10^{12}
petatavu	PB	10^{15}
eksatavu	EB	10^{18}
zettatavu	ZB	10^{21}
yottatavu	YB	10^{24}

Mallissa olevalla **vauhdilla** (velocity) viitataan niin nopeaan datan generoitumisen tahtiin, kuin myös nopeuteen, jolla sitä pitäisi pystyä analysoimaan (Gandomi & Haider, 2015). Oussous ym (2015) argumentoivat tutkimuksessaan Big Datan mahdollisimman nopean käsittelyn puolesta. Nopealla ja reaaliaikaisella datan käsittelyllä maksimoidaan mahdollisuudet saada datasta hyödyllistä ja relevanttia tietoa.

Monipuolisuudella (variety) viitataan siihen, että Big Dataa generoidaan hyvin paljon toisistaan poikkeavista tietolähteistä. Tämän vuoksi Big Data on rakenteeltaan hyvin heterogeenista (Gandomi & Haider, 2015) ja se voi olla esimerkiksi kuvia, sähköposteja, sensoridataa, pdf-tiedostoja, videoita tai äänitiedostoja.

Rakenteen puolesta Big Data jaetaan yleisesti strukturoimattomaan, puolistrukturoituun ja strukturoituun dataan (George, Haas & Pentland, 2014).

Strukturoituun dataan luokitellaan sellainen data, joka on ryhmitelty relationaaliseen tietokantaan riveille ja sarakkeille (Bedeley & Iyer, 2014) ja on helposti lajiteltavissa. Vain noin 5 % kaikesta olemassa olevasta datasta on strukturoitua (Gandomi & Haider, 2015). Strukturoimaton data puolestaan on strukturoidun datan vastakohta. Se voi olla esimerkiksi tekstiä, kuvia, ääntä tai videoita (Gandomi & Haider, 2015). Yleensä strukturoimattomalla datalla viitataan dataan, joka on sellaisessa muodossa, että sitä ei voida helposti indeksoida relationaalsiin tauluihin analyysia tai tietokantakyselyä varten (Bedeley & Iyer, 2014). Kahden edellä mainitun väliin jäävällä puolistrukturoidulla datalla ei puolestaan ole tiukkaa standardien mukaista määritelmää (Gandomi & Haider, 2015). Puolistrukturoidulla datalla on usein sekä strukturoidun sekä strukturoimattoman datan piirteitä. Puolistrukturoidusta datasta puhuttaessa viitataan yleensä sellaiseen dataan, jolla on relaationaalisia rakenteita tai attribuutteja, vaikka se ei olisi täydellistä tai säännöllistä (Bedeley & Iyer, 2014). Big Datan rakenteella on tärkeä merkitys datan tehokkaan analysoinnin ja jatkokäytön kannalta. Tätä käsitellään laajemmin kappaleessa 2.3.

Big Datasta ja sen hyödyntämisestä on tullut merkittävä tekijä useilla aloilla, kuten terveydenhuollossa, pankkitoiminnassa, vakuutusyhtiöissä ja markkinoinnissa (Hassani, Huang & Silva, 2018). Esimerkiksi pankkitoiminnassa Big Dataa voidaan käyttää asiakaskäyttäytymisen analysointiin, riskienhallintaan ja petosten torjuntaan (Hassani, Huang & Silva, 2018a). Terveydenhuollossa Big Datasta voidaan saada hyötyä sairauksien ennaltaehkäisyssä, hoidon yksilöllistämässä ja lääketieteellisen tutkimuksen edistämässä (George, Haas & Pentland, 2014).

Big Dataa on myös yhdistetty muihin teknologioihin, kuten lohkoketjuihin, tekoälyyn ja koneoppimiseen (Hassani, Huang & Silva, 2018a). Esimerkiksi pankkitoiminnassa voidaan käyttää lohkoketjua ja Big Dataa yhdessä asiakkaan henkilöllisyyden todentamiseksi ja rahanpesun torjumiseksi (Hassani, Huang & Silva, 2018a). Tekoäly ja koneoppiminen voivat puolestaan auttaa Big Datan käsittelyssä ja analysoinnissa (George, Haas & Pentland, 2014).

Yhteenvetona voidaan todeta, että Big Data on käsite, joka kuvaa massiivisia ja monimutkaisia tietomassoja, joita ei voida käsitellä perinteisillä tietojenkäsittelymenetelmillä. Big Datasta voi olla hyötyä monilla aloilla, ja se on yhdistetty muihin teknologioihin, kuten lohkoketjuihin, tekoälyyn ja koneoppimiseen, jotta sitä voidaan käsitellä ja analysoida paremmin.

2.2 Big Datan 3V-malli versus 10V-malli

Big Datan määrän kasvu maailmassa sekä datan analysoinnin hyödyntämisen mahdollisuuksien ymmärtäminen ovat lisänneet niin kiinnostusta kuin

tutkimusta Big Datan ja sen hyödyntämisen ympärillä. Big Datan ympärillä lisääntyneiden tutkimuksien ohessa Big Datasta on tunnistettu lisää ominaispiirteitä, jonka seurauksena Big Datasta on kehitetty uusia ja tarkempia malleja. Yksinkertaisin ja tunnetuin malli on Big Datan 3V malli, jota käsiteltiin aikaisemmassa luvussa. Sen mukaan Big Datan kolmen keskeisintä ominaisuutta ovat määrä (volume), nopeus (velocity) ja monipuolisuus (variety) (Laney, 2001). Laneyn kehittämä malli on käytetyin Big Datan malli tieteellisissä tutkimuksissa sekä tieteellisessä kontekstissa.

Tutkimuksien myötä Big Datasta on tunnistettu jopa kymmenen ominaispiirrettä, jonka perusteella Big Datasta on tehty malleja aina 3V-mallista 10V-malliin saakka. 3V-mallin ja 10V-mallin välistä yleisissä tutkimuksissa esiintyvä malli 3V-mallin lisäksi on 5V-malli. 10V-mallissa käsiteltävät Big Datan ominaisuudet vaihtelevat lähteiden välillä. Kyseenomaisella mallilla ei ole virallista viitekehystä, joten aihealueen tutkijat käyttävät tutkimuksissaan erilaisia oman tutkimuksen kontekstiin parhaiten sopivia Big Datan ominaisuuksia. Esimerkiksi Ranjan (2019) tuo esille tutkimuksessaan seuraavat kymmenen Big Datan ominaisuutta: määrä (volume), nopeus (velocity) ja monipuolisuus (variety), todenmukaisuus (veracity), vaihtelevuus (variability), arvo (value), pätevyys (validity), volatilitteetti (volatility), haavoittuvuus (vulnerability), visualisoitavuus (visualization). Lyhyet selitykset uudelle seitsemälle Big Datan ominaisuudelle on lisätty alle:

- **Todenmukaisuus** (veracity) kuvaa datan luotettavuutta sen lähteiden ja kontekstin näkökulmasta. Todenmukaisuudella viitataan myös siihen, miten merkityksellistä data on analyysille, johon se perustuu. (Ranjan, 2019)
- **Vaihtelevuus** (variability) kuvaa ajan mittaan saadun datan johdonmukaisuutta. Big Datan kontekstissa tämä voi tarkoittaa esimerkiksi Big Datan epäjohdonmukaista latausnopeutta tietokantaan, epäjohdonmukaisuutta itse datassa tai useita dataulottuvuuksia, jotka saavat aikaan useita erilaisia datan muotoja ja lähteitä. (Ranjan, 2019)
- **voimassaolo/pätevyys** (validity) viittaa datan laatuun sekä tarkkuuteen sen tarkoitettuun käyttötarkoitukseen nähden. (Ranjan, 2019)
- **Volatilitteetilla** (volatility) viitataan siihen, kuinka kauan kerättyä dataa tulee tai kuinka kauan sitä kannattaa säilyttää. Tarpeet datan säilytyspi-tuudelle vaihtelevat hyvin paljon alojen välillä. (Ranjan, 2019)
- **Haavoittuvuus** (vulnerability) kuvaa Big Datan säilytykseen liittyviä tietoturva-vaatimuksia. Big Dataa tulisi säilyttää tarpeeksi turvallisessa paikassa asiakkaan toiveiden sekä voimassa olevien säännösten mukaan. (Ranjan, 2019)
- **Visualisoitavuus** (visualization) viittaa Big Datan havainnollistamiseen. Tähän tarvitaan hyviä työkaluja. Visualisoinnin vaativuus vaihtelee toimialoittain riippuen usein datan rakenteesta sekä monikanavaisuudesta. (Ranjan, 2019) Esimerkiksi strukturoimatonta dataa voi olla haastavaa visualisoida.

- **Arvo** (value) viittaa siihen, kuinka paljon Big Datalla pystytään tuottamaan arvoa sen käyttäjälle. (Ranjan, 2019) Jos data on alkuperäisessä muodossa, jossa se on saatu lähteestä, sen arvo on käyttäjälle usein alhainen. Kun massiivisia määriä kerättyä dataa analysoidaan oikein, datalla voidaan tuottaa suurta arvoa sen käyttäjälle. (Gandomi & Haider, 2015)

Big Datan 10V-mallin avulla dataa voidaan tutkia tarkemmin eri näkökulmista. Tämän mallin pohjalta tarkasteltuna käsiteltävästä datasta voidaan saada syvällisempi ymmärrys verrattaessa esimerkiksi 3V-malliin. 10V-mallin käyttö antaa käyttäjälleen enemmän arvoa, kun tutkimuksen kohteena on jokin spesifi käyttötarkoitus tai operaatio. Tässä kirjallisuuskatsauksessa käsiteltävä aihe Big Datan yhteydessä on niin laaja, että tarkasteluun sopii paremmin vähemmän ominaisuuksia sisältävä malli. Tämän vuoksi käytämme tässä kirjallisuuskatsauksessa Big Datan 3V-mallia niiltä osin, kun on tarpeellista käsitellä Big Datan ominaisuuksia. Syvällisemmälle Big Datan näkökulmien tarkastelulle ei ole tässä tutkimuksessa tarvetta.

2.3 Big Datan kanavat

Big Datan lähteitä kategorisoidaan hyvin eri tavoilla riippuen tutkimuksesta ja sen kontekstista. Tämä on ymmärrettävä ilmiö, sillä eri toimialoilla Big Datan lähteet sekä sen käyttö eroavat toisistaan suuresti. Big Datan hyödyntäminen tulee aina suunnitella sen käyttökohteen mukaan. Big Datan käytettävissä olevien lähteiden määrä on hyvin toimialakohtaista.

Yksinkertaisimmillaan useissa tutkimuksissa lähdetään lähtökohdasta, jonka mukaan Big Datan lähteet ovat joko sisäisiä tai ulkoisia. (Gandomi & Haider, 2015) Sisäisillä lähteillä tarkoitetaan yrityksien tai organisaatioiden jo olemassa olevissa tietojärjestelmissä sijaitsevaa tai niiden kautta luotavaa dataa. Ulkoinen data puolestaan pitää sisällään datan, joka saadaan ulkoiselta datan tarjoajalta, vapaasti käytettävissä olevista lähteistä kuten sosiaalisesta mediasta tai suoraan asiakkaalta (Ranjan, 2019).

Tutkimuksissa on havaittavissa eroja sisäisten ja ulkoisten lähteiden määrällä eri toimialojen välillä. Tämä selittyy toimialojen erilaisilla ominaispiirteillä sekä sillä, että eri toimialat hyötyvät hyvin erilaisesta datasta. Erilaisia datan tarpeita eri toimialoilla on avattu alla olevissa taulukoissa. Taulukoissa on esitetty toimialoittain erilaisia sisäisen ja ulkoisen datan lähteitä jaoteltuna strukturoituun, puolistrukturoituun sekä strukturoimattomaan dataan.

TAULUKKO 2 Big Datan lähteet terveydenhuollossa (Ranjan, 2019)

	Sisäinen data	Ulkoinen data
Strukturoitu	Kliiniset tietojärjestelmät, potilas-tietojärjestelmät	Kliinisten tutkimusten data, julki-set rekisterit, valtion tietokannat, vakuutusilmoitukset
Puolistrukturoitu	Sähköiset terveystiedot	Geneettinen data, Internet haut
Strukturoimaton	Magneettikuvat, muut kuvanta-miset, käsin kirjoitetut muistiinpa-not	Sensoridata laitteista

TAULUKKO 3 Big Datan lähteet finanssipalveluissa (Ranjan, 2019)

	Sisäinen data	Ulkoinen data
Strukturoitu	Nostoautomaatit, mobiilipankki-data, pankkikorttien transaktiot, asiakkuudenhallintajärjestelmät	Osakkeiden kaupankäyntidata, luottoluokitusdata
Puolistrukturoitu		Internet haut, analyttikkojen ja toimialan raportit
Strukturoimaton	Puhelinkeskustiedot	Sosiaalinen media, uutiset, kes-kustelufoorumit internetissä

TAULUKKO 4 Big Datan lähteet lentoyhtiöillä (Ranjan, 2019)

	Sisäinen data	Ulkoinen data
Strukturoitu	Laitteisto- ja varallisuus data	Asiakkaan ostohistoria, maksutie-dot, asiakkaan sijaintitiedot (las-keutumis- ja lähtöajat sekä kau-punki)
Puolistrukturoitu	Lennon lokitiedot, sensorit	Internet haut
Strukturoimaton	Puhelinkeskustiedot	Sää- ja paikkatieto, sosiaalinen me-dia

Sisäisen ja ulkoisen datan jaottelun lisäksi datan lähteiden kategorisointiin löytyy muitakin tapoja riippuen toimialasta. George ym (2014) ovat kategorisoineet Big Datan lähteet viiteen pääkategoriaan. Georgen ym (2014) mukaan kategoriat ja niiden selitteet ovat seuraavanlaiset:

- **Julkinen data** (public data) on julkishallinnon, julkishallinnon organisaatioiden ja paikallisten yhteisöjen keräämää dataa, jota voidaan hyödyntää liiketoiminnassa ja johtamisessa.
- **Yksityinen data** (private data) on yksityisten yritysten, organisaatioiden sekä yksityishenkilöiden hallinnoimaa dataa, jotka eivät ole julkista dataa.
- **Datajäte** (data exhaust) viittaa dataan, joka ei ole keskeistä kerääjälleen ja jota on kerätty ympäristöstä passiivisesti. Tällä datalla on omistajalleen vain vähän tai ei yhtään arvoa.
- **Yhteisötieto** (community data) on strukturoimatonta dataa, joka on lähtöisin yhteisön jäsenten tai käyttäjien tuottamasta datasta. Esimerkki yhteisödatasta on kuluttajien arviot tuotteista.
- **Itsensä mittaamiseen perustuva data** (self-quantification data) on dataa, joita yksilö voi paljastaa mittaamalla henkilökohtaista käyttäytymistään sekä toimiaan. Tästä hyvä esimerkki on aktiivisuusranneke, joka mittaa unta, askelia ja liikuntasuorituksia.

Teknologian kehittyessä, uusien sovelluksien syntyessä ja digitalisaation jatkaessa kasvamistaan myös Big Datan monimuotoisuus sekä määrä kasvaa. Datan määrä jatkaa kasvua eksponentiaalisesti, sillä myös sen lähteiden määrä kasvaa jatkuvasti. (Bedeley, 2014)

2.4 Big Data-analytiikka

Yksinään Big Data ei takaa hallitsijalleen automaattista kilpailuetua tai merkittävää hyötyä alan muihin kilpailijoihin nähden. Kilpailuedun saavuttamiseksi dataa täytyy osata käsitellä oikein. Erinäisillä toimijoilla voi olla hallussaan massiivisiakin määriä Big Dataa, mutta sen täysi potentiaali realisoituu vasta data-analyysin kautta. Data-analyysissä on tärkeää löytää datasta kuvioita ja malleja, joiden avulla voidaan esimerkiksi tehdä johtopäätöksiä analysoitavasta kohteesta ja saada arvokasta tietoa (Sagirlou & Sinanc, 2013). Näin ollen analyysin avulla voidaan paljastaa piileviä yhteyksiä, havaita trendejä ja ennustaa jopa tulevia tapahtumia. Hyvällä ja perusteellisella Big Datan analysoinnilla voidaan saavuttaa kilpailuetua ja avustaa päätöksentekoa (Zakir ym, 2015).

Perinteiset analysoinnin keinot eivät välttämättä toimi täysin Big Datan analysoinnin kohdalla (Zakir ym, 2015). Big Dataa tulee jatkuvasti suurina määrinä useista eri lähteistä, jonka lisäksi se on usein strukturoimatonta. Nämä ominaisuudet luovat haasteita Big datan analysointiin. Big Datan analysointi vaatii selvästi kehittyneitä algoritmeja ja nopeaa datan käsittelyä, sillä dataa syntyy

jatkuvasti. Hyvänä esimerkkinä datan suuresta määrästä toimii hakukone Google, joka vastaa kuukaudessa yli 100 miljardiin kyselyyn. Jopa yhden päivän aikana syntyneen datan analysoiminen on täysin mahdotonta tavanomaisella tietokannalla. (Varian, 2014)

Big Data-analyysissä on erityisen tärkeää löytää oikea data, jota lähdetään analysoimaan. Kaikkea dataa ei voida analysoida, sillä dataa generoituu niin massiivisia määriä, että kaiken datan analysoiminen olisi resurssien hukkaamista. Täten yksi Big Data -analyysin tärkeä vaihe on valita mahdollisimman paljon arvoa tuottava data ja analysoida sitä.

Edellisessä kappaleessa mainittu lähteiden eksponentiaalinen kasvaminen tulee luomaan omat haasteensa Big Data-analytiikkaan. Kun lähteiden määrä kasvaa, Big Data on entistä monimuotoisempaa sekä sitä generoituu jatkuvasti suurempia määriä. Tämä vuoksi Big Data-analyysin täytyy pysyä mukana kehityksessä. Big Datan analysoimiseen tarvitaan tehokkaita algoritmeja, sillä datamassat ovat suuria. Suorituskyvyn parantaminen ja kilpailukyky saadaan analytiikkamalleista, joiden avulla johtajat voivat ennustaa ja optimoida tuloksia. (Bedeley, 2014)

3 FINANSSISEKTORI

Finanssisektorin moderni malli juontaa juurensa 1950-luvulle, jolloin kehitettiin paljon vielä nykypäivänäkin käytössä olevia rahoitukseen liittyviä ydinteorioita sekä malleja. Neljäkymmenen vuoden kuluessa finanssisektorin modernin mallin syntymisestä alasta on tullut yksi eniten opiskelluista aloista kauppatieteiden sisällä opiskelijamäärien ja rahoituksen kurssien lukumäärää muihin kauppatieteiden opintosuuntauksiin verrattaessa. (Miller, 1999) Vaikka finanssisektorin rakenteet ovat vielä tallella, finanssiala on kehittynyt isoissa harppauksissa etenkin uusien teknologioiden implementoinnin, digitalisaation sekä dataohjautuvan yhteiskunnan tuloksena.

Terve sekä elinvoimainen talous vaatii taustalleen rahoitusjärjestelmän, joka on vakaa ja allokoii varoja säästäviltä henkilöiltä sellaisille henkilöille, joilla on tuottavia investointimahdollisuuksia. (Mishkin ym, 2013) Finanssisektorin ytimessä on rahoitusprosessit, jotka pitävät sisällään rahan luontia, hallinnointia sekä investointia (Qin ym, 2021). Vakaa rahoitusjärjestelmä lisää taloudellista tehokkuutta, johon rahoitusjärjestelmämme rakenne vahvasti pyrkii. (Mishkin ym, 2013). Varojen kanavoiminen säästäjiltä varojen käyttäjille ja investoijille on rahoitusjärjestelmän perusta ja ehdottomasti tärkein tehtävä. Ilman nykyaikaista rahoitusjärjestelmää taloutemme olisi varsin primitiivistä, sillä jokaisen yksilön sekä entiteetin tulisi kyetä rahoittamaan itse itsensä ilman, että rahoitusmarkkinoihin turvautuminen olisi mahdollista. (Qin ym, 2021)

Finanssialan toimijoita ja organisaatioita yhdistää yksi asia: raha. Rahoitusjärjestelmän rakenne on hyvin monimutkainen ja se pitää sisällään hyvin monenlaisia yksityisen sektorin rahoitusinstituutioita sekä organisaatioita, joita ovat pankit, vakuutusyhtiöt, rahoitusyhtiöt sekä sijoituspankit. Kaikki edellä mainitut instituutiot ovat tarkasti valtioiden säätelemiä ja valvomia. (Mishkin, 2016) Rahoitusinstituutiot käsittelevät erilaisia rahoitusinstrumentteja, kuten osakkeita ja lainoja, joiden avulla varallisuutta pystytään siirtämään sitä tarvitseville tai haluaville. Paperisilla sekä virtuaalisilla rahoitusinstrumenteilla käydään kauppaa rahoitusmarkkinoilla. Seuraavissa alaluvuissa käsitellään tarkemmin finanssisektorin toimijoita, finanssisektorin roolia rahoituksen välittäjänä sekä

nykyaikaisen rahoitusjärjestelmän ominaispiirteitä. Rahoitusmarkkinat luovat likviditeettiä tuodessaan yhteen myyjiä sekä ostajia.

3.1 Finanssisektorin toimijat ja niiden operaatiot

Finanssisektori on suuri kokonaisuus, joka koostuu monesta pienemmästä osasta, toimijasta sekä prosessista. Käsite itsessään on hyvin laaja, joten helpoin tapa määritellä finanssisektoria on aloittaa tarkastelemalla finanssisektorin toimijoita ja niille ominaispiirteisiä operaatioita. Tässä alaluvussa käsitellään vakuutusyhtiöitä, pankkeja sekä rahoitusyhtiöitä, joiden kautta määritellään myös rahoitusmarkkinoiden toimintaa.

Suomalaiset yksityiset vakuutusyhtiöt hoitavat lakisääteisiä sekä vapaaehtoisesti hankittavia vakuutuksia. Suomessa vakuutustoiminnan harjoittaminen vaatii toimiluvan valtioneuvostolta sekä muut toimiluvat finanssivalvonnalta. Vakuutusala on hyvin tarkasti säädeltyä ja siitä on vastuussa finanssivalvonta (Sosiaali- ja terveysministeriö, 2023). Syy vakuutusalan tarkkaan säätelyyn voidaan nähdä olevan sen suuri vaikutus talouteen. Vain lakisääteisistä vakuutuksista kerääntyvät maksut tekevät suurimman osan koko alan maksutuloista (Finanssiala.fi, 2023). Vakuutukset ovat suuri tekijä suomalaisen sosiaaliturvan kehittämiseksi ja ylläpitämiseksi, sillä huomattava osa sosiaaliturvastamme koostuu lakisääteisistä vakuutuksista. Lakisääteisen vakuutusten lisäksi turvaa ovat täydentävässä vapaaehtoinen vakuutusurva (Finanssialalle, 2023). Vakuutusmarkkinoiden sääntely vakauttaa vakuutusmarkkinoita ja siellä liikkuvia suuria rahamääriä, joilla on myös oleellinen vaikutus talouteen. Vakuutusten perusajatus on riskin tasaaminen jakamalla se suuren ryhmän kesken, jolloin kenenkään maksukyky ei vaarannu. Vakuutusmaksuista kertyneet varat voidaan käyttää vahinkojen korvaamiseen. (Finanssialalle, 2023)

Pankit ovat yksi finanssisektorin keskeisimmistä ja tutuimmista toimijoista. Pankkien tärkein tarkoitus on tallettaa asiakkaiden varoja, antaa luottoja sekä siirtää varoja toimeksiantojen mukaisesti. Edellä mainittujen lisäksi pankit hoitavat myös sijoitustoimintaa sekä varallisuudenhoitopalveluita. (Finanssialalle, 2023) Yksityishenkilöiden pankkiasioiden hoitamisen lisäksi pankit hoitavat myös yritysten, organisaatioiden sekä valtioiden varoja. Pankit ovat hyvin tärkeä osa varojen liikuttamista säästäjältä potentiaaliin investointimahdollisuuksiin talouden kiertokulun ylläpitämiseksi. Rahan tallettamisen, luottojen antamisen sekä varojen siirtämisen vuoksi pankit ovat elintärkeä osa rahoitusjärjestelmäämme sekä rahoitusmarkkinoita. Tämän lisäksi pankit välittävät, ostava ja myyvät asiakkailleen rahoitusmarkkinoiden rahoitusinstrumentteja.

Edellä olevissa kappaleissa mainitut finanssialan toimijat ovat rahoituksen välittäjiä. Heidän kauttaan raha liikkuu paikasta toiseen, joka mahdollistaa osaltaan eläväisen talouden jatkumisen. Näin saadaan liikutettua varoja rahaa säästäviltä ihmisiltä tai ihmisiltä, joiden tulot ylittävät menot, ihmisille, jotka tarvitsevat rahoitusta. Tämänkaltaisen toiminta edistää talouden kiertokulkua ja

toteutettua rahoitusjärjestelmän tärkeintä tehtävää (Mishkin, 2013). Varojen ohjaaminen oikeaan paikkaan mahdollistaa talouden kasvamisen maksimoinnin.

3.2 Moderni rahoitusjärjestelmä

Edellisessä alaluvussa käsiteltiin finanssisektorin toimijoita. Seuraavaksi perehdytään paremmin siihen, minkälainen rooli näillä toimijoilla on modernissa rahoitusjärjestelmässä ja millainen sen rakenne on. Hyvin toimiva rahoitusmarkkinat ovat edellytys taloudelliselle kasvulle ja huonosti toimivat rahoitusmarkkinat ovat syy köyhien valtioiden vaurastumattomuudelle. (Mishkin, 2013)

Rahoitusjärjestelmämme on rakenteeltaan hyvin keskitetty. Euroalue ja sen pankkien kytkeytyminen toisiinsa on hyvä esimerkki keskitetystä rahoitusjärjestelmästä. Keskitetty rahoitusjärjestelmän rakenne esimerkiksi euroalueella johtuu siitä, että jokaisella maalla on oma keskuspankki, joka toimii kyseisessä maassa toimivien pankkien pankkina. Esimerkiksi Suomen keskuspankki on Suomen Pankki. Keskuspankkien tehtävä on vakaasta rahan arvosta ja luotettavasti toimivasta rahoitusjärjestelmästä huolehtiminen. Edellä mainittujen tehtävien lisäksi keskuspankit huolehtivat vakaasta hintatasosta ja turvallisista maksujärjestelmistä. Nämä kaikki tehtävät auttavat edistämään kestävästä talouskasvusta, työllisyyttä sekä maan kansalaisten hyvinvointia. Kaikkien euroalueeseen liittyneiden maiden keskuspankit ovat yhdistäneet oman maansa keskuspankin osaksi eurojärjestelmää. Yksityishenkilöt eivät voi asioida keskuspankissa, mutta esimerkiksi Suomessa Suomen muut pankit voivat tallettaa varojaan tai ottaa lainaa Suomen Pankista. (Suomen Pankki, 2023) Keskitetty rahoitusjärjestelmä ei päde pelkästään Euroopan sisällä, vaan samankaltaisia rakenteita on havaittavissa ympäri maailmaa. Tämän lisäksi Euroopan pankit ovat vahvasti kytköksissä muiden alueiden pankkeihin. Isot pankit ovat vahvasti kytköksissä toisiinsa esimerkiksi lainaamalla paljon myös toisilleen.

Vaikka keskitetyllä rahoitusjärjestelmän rakenteella on tuhansien vuosien takainen historia aina Mesopotamian aikoihin asti, (Qin ym, 2021) keskitetty rakenne itsessään luo järjestelmään haavoittuvuuksia. Edellä mainittu euroalue on hyvä esimerkki siitä, kuinka vahvasti pankit ovat kytköksissä toisiinsa. Esimerkiksi yhteen keskeiseen pankkiin kohdistettu massiivinen kyberhyökkäys voi saada aikaan monen muun pankin tai jopa suuremman alueen lamaantumisen. Yhden pankin toiminnan lamaantuessa sen vaikutukset leviävät koko pankkien verkostoon. Tästä hyvä esimerkki on esimerkiksi vuoden 2008 finanssikriisi, joka lähti liikkeelle Yhdysvalloista ja vaikutti lopulta koko maailman talouteen.

Vuoden 2008 finanssikriisi sai alan asiantuntijat pohtimaan tarkemmin rahoitusjärjestelmän rakennetta sekä sen haavoittuvuuksia. Keskitetyn rahoituksen haavoittuvuuksien tunnistaminen on johtanut uusiin teknologioihin sekä säädöksiin finanssisektorilla. Alan asiantuntijat pyrkivät jatkuvasti keksimään uusia keinoja, jolla rahoitusjärjestelmän rakenteesta saataisiin entistä kestävämpi ja vielä modernimpi. Eräs usein esille nouseva vaihtoehto on hajautettu rahoitus.

Tieteellisissä artikkeleissa käydään paljon keskustelua keskitetystä rahoituksesta verrattuna hajautettuun rahoitukseen. Artikkeleissa pohditaan paljon siirtymistä keskitetystä rahoituksesta hajautettuun rahoitukseen sekä näiden kahden rahoitusmallin rinnakkaisesta käytöstä rahoitusjärjestelmän vakauttamiseksi. Tässä kandidaatin tutkielmassa ei keskitytä sen tarkemmin hajautettuun rahoitukseen.

Mishkin ym (2013) luettelevat kirjassaan kahdeksan kohtaa, jotka auttavat ymmärtämään rahoitusjärjestelmän toiminnan. Nämä kahdeksan kohtaa ovat seuraavat:

1. Osakkeet eivät ole yritysten tärkein ulkoisen rahoituksen lähde.
2. Osakkeiden liikkeeseenlasku ei ole ainoa jälkimarkkinakelpoinen arvopaperityyppi, jolla yritykset rahoittavat toimintaansa.
3. Epäsuora rahoitus, johon sisältyy rahoituksen välittäjien toimintaa, on monta kertaa tärkeämpää kuin suora rahoitus, jossa yritykset hankkivat varoja suoraan lainanantajilta rahoitusmarkkinoilta.
4. Rahoituksen välittäjät, erityisesti pankit, ovat tärkein yritysten rahoittamiseen käytettävien ulkoisten varojen lähde.
5. Rahoitusjärjestelmä on yksi tiukimmin säännellyistä talouden sektoreista.
6. Vain suurilla, vakiintuneilla yrityksillä on helppo pääsy arvopaperimarkkinoille rahoittaakseen toimintaansa.
7. Vakuudet ovat yleinen piirre sekä kotitalouksien että yritysten velkasopimuksissa.
8. Velkasopimukset ovat yleensä erittäin monimutkaisia oikeudellisia asiakirjoja, jotka asettavat lainanottajan käyttäytymiselle huomattavia rajoituksia.

4 BIG DATA FINANSSISEKTORILLA

Finanssisektorin dynamiikka on muuttunut huomattavasti viimeisten vuosikymmenten aikana. Yksi keskeisimmistä tekijöistä dynamiikan muutoksissa on datan roolin muuttuminen. Koko yhteiskunta on muuttunut datavetoisemmaksi tietokoneiden sekä internetin yleistyessä, datan määrien noustessa sekä digitalisaation kiihtyessä. Dataa ja tietovarastoja on hyödynnetty finanssialalla päätöksenteon tukena ja paremman liiketoiminnan ymmärtämiseksi sekä toteuttamiseksi jo aina 1980-luvulta lähtien (Gasser ym, 2017). Vaikka datan hyödyntäminen ja analysointi on paljon esillä nykypäivänä niin tieteellisessä kuin yritysmaailman keskustelussa, voidaan todeta, että datan hyödyntäminen itsessään ei ole kovin uusi ilmiö. Vaikka kyseessä ei ole uusi ilmiö, datan hyödyntäminen on kehittynyt valtavasti viimeisen 40 vuoden aikana. Teknologia on ottanut harppauksia ja tietokoneiden laskentateho on kasvanut huomattavasti, joka on osaltaan vauhdittanut tätä kehitystä. Datan hyödyntäminen ja sen määrän historiallinen kehitys ovat oleellinen pohjustus ymmärtääksemme Big Datan nykyistä asemaa sekä vaikutusta finanssisektoriin.

Useimmiten Big Datasta puhuessa nostetaan esiin kolme Big Datan 3V-mallin mukaista ominaisuutta: määrä (volume), nopeus (velocity) sekä monipuolisuus (variety). Nämä kolme ominaisuutta ovat varsin relevantteja kun puhutaan Big Datasta tietotekniikan kontekstissa, mutta Goldstein ym (2021) pohtivat artikkelissaan, pystytäänkö näiden ominaisuuksien kautta tutkimaan Big Dataa riittävästi finanssialan kontekstissa. Edellä mainittu nosto on tärkeä, sillä konteksti vaikuttaa paljon siihen, miten tutkittavaa asiaa on hyvä lähestyä.

Goldstein ym (2021) esittävät artikkelissaan toisenlaisen näkökulman Big Datan lähestymiseen finanssisektorin kontekstissa. Erona Langleyn (2001) perinteiseen Big Datan 3V-malliin Goldstein ym (2021) esittävät yhdeksi ominaisuudeksi Big Datan laajat ulottuvuudet perinteisen 3V-mallin nopeuden tilalle. Tämä tarkoittaa käytännössä sitä, että finanssisektorilla käsiteltävässä Big Datassa on paljon muuttujia suhteessa otoskokoon. Muuttujien suuri määrä luo haasteita datan analysointiin. Goldstein ym (2021) esittävät ratkaisuksi koneoppimisen hyödyntämisen laajat ulottuvuudet omaavan datan analysoinnissa. Koneoppimista hyödynnetään jo paljon finanssialan

tutkimuksessa ja tämä tehostaa Big Datan analysointia entisestään. Stojkovin ja Warinin (2021) mukaan koneoppimisen soveltaminen finanssialalla on ollut merkittävä metodologinen askel eteenpäin. Hyödyntämällä dataa sekä koneoppimista voidaan ennustaa finanssimarkkinoiden trendejä.

Big Data sekä sen lähteet finanssisektorilla eroavat muilla aloilla olevasta Big Datasta sekä Big Datan lähteistä. Finanssisektorilla alakohtaista Big Dataa ovat muun muassa transaktiodata sekä pörssitiedot, jotka pitävät sisällään osakekurssit, osakekaupan ja sen volyymin sekä korot. (Hasan ym, 2020) Näiden lähteiden lisäksi alaspesifin datan lähteinä toimivat pankkiautomaatit, mobiilipankit, asiakkuudenhallintajärjestelmät, luottoluokitusdata sekä analyttikkojen ja toimialan raportit (Ranjan, 2019). Alaspesifien lähteiden lisäksi finanssisektorilla pystytään myös hyödyntämään lähteitä, joita käytetään datan lähteenä monella muullakin alalla. Näitä lähteitä ovat esimerkiksi seuraavat: sosiaalinen media, kuten Facebook, X (ent. Twitter), sanomalehdet, mainokset, televisio ja niin edelleen. (Hasan ym, 2020) Eräs yleinen lähde monella alalla toiminnan kehittämiseksi on asiakaspalvelun kirjaukset. (Bedeley, 2014) Voidaan todeta, että finanssisektorilla syntyy hyvin paljon erilaista Big Dataa. Finanssisektori pystyy hyödyntämään toiminnassaan monenlaisia Big Datan lähteitä.

Big Datan erilaiset lähteet poikkeavat toisistaan paljon, josta johtuen erilaisista Big Datan lähteistä syntyy hyvin paljon tyypiltään sekä rakenteeltaan erilaista dataa. Esimerkiksi uutistista saatavaa tietoa kerätään, käsitellään, analysoidaan ja säilötään hyvin eri tavoin, kuin transaktiodataa mobiilipankista. Suurin osa olemassa olevasta Big Datasta on strukturoimatonta dataa, joka vaikeuttaa samalla Big Datan analysointia sekä hyödyntämistä. Jopa 95 % Big Datasta on strukturoimatonta dataa. (Tanwar ym, 2015) Strukturoimaton data monimutkaistaa datan analysointia, sillä useimmat tietokoneet vaativat datan käsittelyyn sekä analysointiin rakenteellista datan organisointia. (Tanwar ym, 2015)

Big Datan tehokas hyödyntäminen vaatii määrittelyä siitä, mitä datalla halutaan saada aikaan. Jokainen Big Dataa hyödyntävä organisaatio tarvitsee koherentin datastrategian, joka ottaa kantaa datan järjestämiseen, hallitsemiseen, analysointiin sekä käyttämiseen (DalleMule & Davenport, 2017). Tämä pätee myös finanssisektorilla. Kuten aiemmin on tullut esille, Big Dataa generoituu erilaisista lähteistä valtavia määriä. Big Datan hyödyntämisessä ei ole kyse ainoastaan siitä, kuinka paljon dataa pystytään käsittelemään, vaan pikemminkin siitä, kuinka älykäästä kyseinen data on (George, ym, 2014). Kaiken Big Datan käsittely vaatisi valtavan määrän erilaisia prosesseja eri kanaville, erilaisia paikkoja tallettaa eri muodoissa olevaa dataa, erilaisia prosesseja erilaisen datan käsittelyyn sekä valtavan määrän resursseja. Tämän vuoksi organisaatio tarvitsee datastrategian tämän älykkään ja tarpeellisen datan määrittelyyn sekä sen käsittelyn tueksi.

Seuraavissa alaluvuissa pureudutaan tarkemmin Big Datan hyödyntämiseen liittyviin uhkiin sekä mahdollisuuksiin finanssisektorin kontekstissa. Alaluvuissa nousee esille niin yleiseen julkisista lähteistä saatavaan

Big Dataan, kuin myös alaspesifiin Big Dataan liittyviä huomioita. Big Datan luomien haasteiden ja mahdollisuuksien tarkastelu auttaa hahmottamaan näiden valtavien tietomassojen vaikutuksen finanssisektoriin sekä sen toimijoihin.

4.1 Big Dataan liittyvät haasteet finanssisektorilla

Vaikka Big Data on tuonut mukanaan merkittävää kasvua sekä edistysaskeleita finanssisektorille, sen käyttöön liittyy kiistatta myös haasteita. Näitä haasteita ovat esimerkiksi tietoturva, eettiset kysymykset, laatu ja luotettavuusongelmat, analytiikan haasteet, sääntely ja lainsäädäntö, toimialan muutos ja kilpailu. tiedon kerääminen, anonymisointi, käsittely, reaaliaikainen data, säilöminen, hyödyntäminen sekä tietoturva. Big Datan eri ominaisuudet voidaan nähdä finanssisektorin kontekstissa niin mahdollisuuksia luovina ominaisuuksina kuin myös haasteita luovina ominaisuuksia. Tässä alaluvussa on nostettu esille muutamia useimmiten artikkeleissa listattuja Big Dataan liittyviä haasteita finanssialalla.

Heikko Big Datan laatu nousee esille finanssisektoriin liittyvänä haasteena niin Ravin & Kamaruddin (2017) artikkelissa, kuin Hasan ym (2020) artikkelissa. Big Datalle ominaiset erilaiset rakenteet aiheuttavat haasteita datan laadussa ja oikeellisuudessa. (Ravi & Kamaruddin, 2017) Erilaiset Big Datan rakenteet tekevät Big Datasta vaikeammin kontrolloitavaa. Big Datan erilaiset rakenteet heijastavat Big Datan laatuun negatiivisesti. Ravi & Kamaruddin (2017) kertovat Big Datan huonon laadun olevan yksi suurimmista esteistä Big Data -analytiikan käyttöönotossa. Jos huonolaatuista Big Dataa analysoidaan ja käytetään finanssisektorin toimijoiden päätöksenteon tukena, datan huonolla laadulla voi olla kauaskantoiset seuraukset erityisesti hyödyntäessä Big Dataa isojen päätösten tukena.

Toinen usein artikkeleissa esille nouseva haaste on tietosuojan ja yksityisyyteen liittyvät asiat käsiteltäessä Big Dataa finanssialalla. Finanssisektorin kontekstissa ihmiset antavat itsestään esimerkiksi pankeille ja rahoituslaitoksille hyvin henkilökohtaista tietoa. Pankit pääsevät käsiksi jokaiseen yksilön tai esimerkiksi organisaation transaktioon sekä näkevät jatkuvasti kaikkien asiakkaidensa taloudellisen tilanteen omassa pankissaan. Käsiteltäessä näin henkilökohtaisia tietoja tietosuojan ja yksityisyyden suojan tulee olla kunnossa. Yksityisyyteen ja tietosuojan liittyvät asiat tulee ottaa huomioon monessa eri vaiheessa käsiteltäessä Big Dataa. Tietosuoja ja yksityisyys täytyy ottaa huomioon tietoja kerätessä, tallettaessa ja pohtiessa kuinka henkilötietoja voidaan käyttää kaupallisiin tarkoituksiin (Belhaj & Hachaïchim 2021).

Hyödynnettäessä Big Dataa tulee ottaa erityisen tarkasti huomioon hyödynnettävän Big Datan anonymisointi tietosuojan sekä yksityisyyden säilyttämiseksi. Käytettäessä Big Dataa voidaan käyttää anonymisointialgoritmeja, joiden tarkoitus on suojella yksilöiden yksityisyyttä (DeMauro ym, 2016). Näin Big Dataa voidaan hyödyntää esimerkiksi pankkien päätöksenteossa. Big Datan hyödyntäminen koneoppimismalleissa on huomattu haasteelliseksi siitä syystä, että

monissa malleissa käytetään nimenomaan henkilökohtaisia tietoja (DeMauro ym, 2016).

Henkilötietojen ja suurien datamassojen käsittelystä sekä tietoturvasta on olemassa paljon erilaisia säännöksiä sekä lakeja. Koska finanssisektorilla käsitellään paljon näitä tietoja, finanssisektorin toimijoiden tulee olla hyvin perehtynyt erilaisiin lakeihin sekä säännöksiin, jotka koskevat Big Datan käsittelyä, käyttöä sekä varastointia. Tämä vaatii paljon osaamista ja asiantuntemusta liittyen Big Datan käsittelyyn liittyviin lakeihin sekä säännöksiin. Yksi tunnetuimpia näistä säännöksistä on GDPR (General Data Protection Regulation). Säännökset ja lait aiheuttavat finanssialan toimijoille vaateen toimia useiden sääntelyiden mukaisesti, joka vaatii paljon alan asiantuntijoita.

Big Datan hyödyntäminen vaatii paljon datan käsittelyyn sekä analyysiin liittyvää osaamista. Finanssisektorilla on vielä puutetta pätevistä korkeasti koulutetuista henkilöistä, jotka osaisivat hyödyntää Big Dataa (Ravi & Kamaruddin, 2017). Hyödyntämisen lisäksi Big Dataa syntyy valtavia massoja, jolloin se tarvitsee paljon säilytystilaa. Suuret datamäärät, jotka sisältävät paljon rakenteeltaan erilaista dataa vaativat myös paljon erilaisia tapoja analysoida dataa. Tämän haasteen taklaaminen vaatii paljon alan osaajia työskentelemään Big Datan ja sen hyödyntämisen parissa. Konkreettisesti haasteet Big Datan parissa liittyvät eri tasoihin, kuten tietojen tallentamiseen, varastointiin, hakuun, jakamiseen, analysointiin, hallintaan ja visualisointiin (Ossous ym, 2018).

Lisäksi Big Datan 3V mallissa esitetty ominaisuus nopeus luo omat haasteensa Big Datan hyödyntämiselle. Jokainen finanssisektorin toimija saa päivittäin miljardeittain dataa, mutta kaikkea ei voida käyttää kerrallaan. Kaiken Big Datan analysoiminen ei olisi kustannustehokasta.

4.2 Big Dataan liittyvät mahdollisuudet finanssialalla

Hyvin suunniteltuna Big Datasta voi saada irti paljon mahdollisuuksia sekä kilpailuetua muihin alan toimijoihin nähden. Big Data toimii tekoälyn sekä koneoppimismallien ytimessä. Yhdistämällä Big Dataa edellä mainittuihin asioihin Big Data voi esimerkiksi lisätä tehokkuutta, alentaa tuotteiden ja palveluiden kustannuksia sekä auttaa finanssisektorin toimijoita saavuttamaan parempi asiakastytyväisyys palveluiden tai tuotteiden kautta (Belhaj & Hachaïchim 2021). Tässä alaluvussa avataan muutamia Big Datan käyttöön liittyviä mahdollisuuksia finanssisektorilla, jotka nousevat esille tieteellisissä julkaisuissa.

Big Datan hyötyjä finanssisektorilla käsittelevissä artikkeleissa nousee usein esille dataan pohjautuvien riskien tunnistaminen ja vähentäminen. Puhuttaessa riskeistä finanssisektorin kontekstissa, riskit voivat tarkoittaa esimerkiksi luottoriskejä, markkinariskejä tai operationaalisia riskejä (Ravi & Kamaruddin,

2017, s7) Vertaamalla vanhaa dataa tietyistä tilanteista ja riskeistä reaaliaikaiseen uuteen dataan voidaan tunnistaa kaavoja, jotka indikoivat tiettyjä riskejä. Monia finanssisektorille tyypillisiä riskejä voidaan mitata ja ennustaa analyttisten mallien ja suuren datamäärän kautta. (Ravi & Kamaruddin, 2017, s7) Hasanin ym (2020) mukaan riskien analysointia voidaan pitää tärkeimpänä tekijänä yrityksen voiton maksimoinnissa. Myös petosten havaitseminen Big Dataa hyödyntämällä nousee esille useammissa artikkeleissa.

Riskien hallinnan lisäksi Big Datasta voidaan saada selville myös paljon muita kilpailuetua antavia asioita. Esimerkiksi pankit saavat dataa päivittäisistä transaktioista, asiakaspalvelu muistioista ja sosiaalisesta mediasta. Mitä enemmän pankilla on dataa omasta liiketoiminnastaan, sitä parempi kilpailuetu sillä on kilpailijoita vastaan, jotka eivät osaa hyödyntää tätä ainutlaatuista dataa. (Bedeley, 2014) Datan kerääminen omasta toiminnasta mahdollistaa organisaatioille oivalluksia omasta liiketoiminnasta, parantaa tehokkuutta sekä luo innovaatioita prosessien parantamiseksi (Eberendu, 2016).

Alan tutkimuksissa sanotaan Big Datan mullistaneen finanssisektorin erityisesti reaaliaikaisilla pörssitiedoilla. Nämä reaaliaikaiset tiedot ovat muuttaneet kauppaa sekä investointeja, petosten havaitsemista ja ehkäisyä sekä riskianalyysiä. (Hasan ym, 2020) Reaaliaikaiset tiedot nopeuttavat kaupankäyntiä ja antavat finanssisektorin toimijoille kattavat tiedot, joiden pohjalta on mahdollista tehdä mahdollisimman hyviä ostoja ja myyntejä. Reaaliaikaiset tiedot tehostavat sijoitustoimintaa yleisesti. Finanssisektorin toimijoiden on helpompaa tehdä hyviä päätöksiä, sillä heillä on paljon tietoa, jonka lisäksi he pystyvät analysoimaan ja ennustamaan markkinoiden kehitystä.

Omaa liiketoimintaa dataa tutkimalla on mahdollista tehostaa niin organisaation tai yrityksen toimintamalleja, kuin myös luoda tehokkaampaa asiakasvuorovaikutusta. Onnistuessaan molemmat edellä mainitut asiat mahdollistavat entistä suuremmat tulot kyseiseen organisaatioon tai yritykseen. (Gasser ym, 2017) Big Datan avulla voidaan myös saada tarkempaa tietoa asiakkaiden käyttäytymisestä. Rahoituslalla taritaan paljon tietoa, jota tulee jalostaa ennen kuin sitä käytetään. Tämä tekee Big Datasta ja sen hyödyntämisestä erityisen tärkeää finanssisektorille. (Sun ym, 2020)

Big Datan keräämisellä ja oikeanlaisella analysoimisella voidaan mahdollistaa entistä parempi päätöksenteko. Päätöksenteon onnistuminen Big Datan pohjalta vaatii kuitenkin sen, että data on laadultaan hyvää, hyvin analysoitua ja mittaa oikeaa kohdealuetta. Päätöksenteon tukena käytetään enenevissä määrin myös tekoälyä sekä koneoppimista. Tekoälyn ja koneoppimisen pohjana käytetään Big Dataa. Finanssisektoriin ja Big Dataan liittyvissä tutkimuksissa on alettu tutkimaan myös sitä, missä määrin koneet voivat tehdä jatkossa päätöksiä koneoppimisen pohjalta. Yritykset, jotka ovat nopeampia kuin useimmat perinteiset rahastot, mutta hitaampia kuin korkean taajuuden kauppiat, ovat vielä tois- taiseksi akateemisen tutkimuksen ulkopuolella. (Goldstein ym, 2021)

5 YHTEENVETO

Tämän kirjallisuuskatsauksen aihe on Big Datan hyödyntämisen haasteet ja mahdollisuudet finanssisektorilla. Tutkielmassa käsiteltiin Big Dataa, finanssisektoria sekä sitä, millaisia mahdollisuuksia ja haasteita Big Datan hyödyntämisellä on finanssisektorin kontekstissa. Kirjallisuuskatsauksen tutkimuskysymykset ovat seuraavat :

- Mitä haasteita Big Datan hyödyntämisessä on finanssisektorilla?
- Mitä mahdollisuuksia Big Datan hyödyntämisellä on finanssisektorilla?

Tutkielman toisessa pääluvussa perehdyttiin Big Datan määritelmään, ominaisuuksiin, malleihin, sekä kanaviin. Big Datasta pyrittiin antamaan mahdollisimman kattava käsitys, jotta muun tutkielman lukeminen ja ymmärtäminen olisivat mahdollisimman helppoa. Tutkielman kolmannessa pääluvussa avattiin finanssisektorin toimintaa, toimijoita sekä näiden operaatioita. Lopulta neljännessä pääluvussa käsiteltiin Big Dataa finanssisektorin kontekstissa ja tuotiin esiin erilaisia haasteita sekä mahdollisuuksia Big Datan hyödyntämiseen liittyen finanssisektorilla. Vastaukset tutkielman tutkimuskysymyksiin tulevat esiin neljännessä pääluvussa. Useammassa lähteessä mainittiin Big Datan tutkimuksen finanssisektorin kontekstissa olevan vähäistä. Vähäisen tutkimuksen määrä toimi motivaationa tämän tutkielman kirjoittamiselle. Aiheesta tarvitaan vielä lisää tieteellistä tutkimusta, sillä suuri osa luotettavista aihepiiriä käsittelevistä lähteistä on saattanut vanhentua aihealueen kehittyessä valtavien harppauksien teknologioiden mukana. Datan hyödyntäminen erityisesti koneoppimista sekä tekoälyä hyödyntämällä ovat kehittyneet paljon viimeisen muutaman vuoden aikana. Tämän vuoksi on täysin ymmärrettävää, että aiheesta ei ole vielä kerennyt tulla useita laadukkaita tutkimuksia, joiden julkaisemisessa saattaa kestää jopa vuosia.

Kirjallisuuskatsauksen tuloksena voidaan todeta, että Big Datan hyödyntäminen luo finanssisektorille paljon mahdollisuuksia sekä auttaa

ennustamaan mahdollisia skenaarioita, joita finanssialalla tapahtuu. Big Data-analyysin tuloksilla voidaan myös tukea finanssisektorin päätöksentekoa ja ohjata toimintaa niin, että siitä saadaan mahdollisimman suuri hyöty niin asiakkaalle, kuin pankille tai muulle rahoitusinstituutiolle.

Big Datan hyödyntämisen prosessien täytyy olla kunnossa, jotta Big Datasta saadaan mahdollisimman paljon höytyä irti. Hyödyllisen datan seulominen ja analysointi on ensisijaisen tärkeää, jotta voidaan saada kohteeseen nähden hyödyllistä tietoa. Suurista datamassoista tulee kyetä tunnistamaan datasesitit, joita analysoimalla voidaan höytyä parhaiten. Datasettien pohjalta voidaan muodostaa erilaisia malleja, joita voidaan hyödyntää myös jatkossa. Toisaalta tärkeää on myös ymmärtää ja varmistaa datan oikeellisuus sekä laatu. Jos hyödynnettävä data ei ole luotettavaa, data-analyysin tulokset voivat vääristää todellisuutta ja ohjata päätösten tekijöitä tekemään vääränlaisia ratkaisuja vallitsevaan tilanteeseen nähden. Pahimmillaan tämä voi johtaa suuriinkin taloudellisiin tappioihin yrityksessä, organisaatiossa tai jopa yksilötasolla.

Dataa analysoitaessa on hyvä ottaa huomioon Big Dataa kuvaavat viitekehykset. Tarkastelun kohteena olevan viitekehyksen valinta tulee tehdä käyttökohteen mukaan. Monimutkaisemmissa Big Datan hyödyntämisen käyttökohteissa on hyvä käyttää apuna viitekehystä, jossa käsitellään usemapa Big Datan ominaisuutta.

Big Datan hyödyntämiseen liittyy myös haasteita. Haasteet heijastuvat hyvin paljon Big Datan 3V-mallin mukaisiin ominaisuuksiin. Big Datan syntymisen määrä aiheuttaa omat haasteensa Big Datan käsittelylle. On erityisen tärkeää osata tunnistaa tärkeä data, jotta organisaatioissa ja yrityksissä ei käytetä aikaa sekä resursseja sellaisen datan käsittelyyn, joka ei anna tarpeellista tietoa halutusta kohteesta. Määrän lisäksi Big Datan syntymisen nopeus vaatii huomiota, kun alanysoidaan Big Dataa. Analysointiprosessien tulee olla tarpeeksi nopeita, jotta Big Data analyysistä saatava tieto on relevanttia ja käyttökelpoista. Liiketoiminta on muuttunut äärimmäisen nopeaksi digitalisaation myötä, jolloin myös Big Datan analysointiprosessien täytyy olla nopeita. Nopea vauhti ja suuri Big Datan määrä aiheuttavat paljon vaatimuksia Big Datan käsittelylle. Big Datan 3V-mallin mukainen kolmas ominaisuus vaatii myös huomiota, jotta analysoitu Big Data vastaa tarpeisiin. Heterogeenisen rakenteen vuoksi Big Data analyysi tarvitsee paljon erilaisia analysointikeinoja. Ominaisuuksiin liittyvien haasteiden lisäksi Big Dataa käsitellessä tulee ottaa huomioon Big Datan anonymisointi, tietoturva sekä yksityisyys. Big Dataa käsitellessä tulee ottaa huomioon paljon erilaisia säännöksiä ja lakeja, joka vaatii Big Dataa käsittelevän yritykseen tai organisaatioon paljon osaamista. Tietoturvan lisäksi on erityisen tärkeää ottaa huomioon käytettävän Big Datan laatu. Big Datalla pyritään tukemaan päätöksentekoa ja heikkolaatuinen data saattaa vääristää asioita. Jos heikkolaatuisen Big Datan pohjalta lähdetään tekemään päätöksiä, Big Data ohjaa päätöksentekoa väärään suuntaan.

Aiheesta löytyy paljon jatkotutkittavaa erityisesti Big Datan hyödyntämisen uusine innovaatioiden sekä alan nopean kehittymisen vuoksi. Big Datan konkreettista hyödyntämistä voisi tutkia enemmän niin, että aiheesta

saataisiin lisää luotettavai tutkimuksia. Lisäksi aihetta tutkiessa löytyi viitteitä siitä, että Big Datan analysointiin ja hyödyntämiseen on käytetty koneoppimista sekä tekoälyä. Ottaen huomioon tekoälyn ja koneoppimisen viimeaikaisen kehittymisen sekä lisääntymisen niin yritys- kuin tutkimusmaailmassakin, olisi hienoa nähdä lisää laadukkaita tutkimuksia, jossa Big Dataa käsitellään tekoälyn ja koneoppimisen keinoin. Näitä keinoja on todennäköisesti kehitetty viimeisien vuosien aikana, mutta se ei näkynyt vielä aihetta käsittelevissä luotettavissa tutkimuksissa.

Aiheen laadukkaan ja laajempaan tutkimuksen puutteesta huolimatta kirjallisuuskatsauksen tulokset osoittautuivat sellaisiksi, joita niiden ajateltiin olevan. Tutkielman lopputulos auttaa ymmärtämään mihin asioihin Big Dataa hyödyntäessä tulee kiinnittää erityistä huomiota ja mitä tarjottavaa Big Datan analysoinnilla on organisaatiolle sekä yrityksille. Tulokset auttavat myös ymmärtämään sen, kuinka monimutkaisesta aiheesta lopulta on kyse. Aiheesta voisi tehdä laajempaa tutkimusta käyttämällä Big Datan 3V-mallin sijaan esimerkiksi Big Data 10V-mallia ja täten ottamalla huomioon enemmän Big Dataan liittyviä ominaisuuksia. Tämä aihe tulee todennäköisesti olemaan näkyvissä alan tutkimuksessa myös jatkossa, sillä se on kehittynyt lähivuosien aikana suuresti.

LÄHTEET

- Bedeley, R. (2014). Big Data opportunities and challenges: the case of banking industry.
- Belhaj, M., & Hachaïchi, Y. (2021). Artificial intelligence, machine learning and big data in finance opportunities, challenges, and implications for policy makers.
- Cockcroft, S., & Russell, M. (2018). Big data opportunities for accounting and finance practice and research. *Australian Accounting Review*, 28(3), 323-333.
- DalleMule, L., & Davenport, T. H. (2017). What's your data strategy. *Harvard business review*, 95(3), 112-121.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library review*, 65(3), 122-135.
- Eberendu, A. C. (2016). Unstructured Data: an overview of the data of Big Data. *International Journal of Computer Trends and Technology*, 38(1), 46-50.
- Finanssialalle.fi, 2023
<https://www.finanssialalle.fi/opintomateriaalit/finanssialanperusteet/finanssialalla-toimiminen/finanssimarkkinoiden-toimijat/vakuutuksen-periaate.html>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Gasser, U., Gassmann, O., Hens, T., Leifer, L., Puschmann, T., & Zhao, L. (2017). Digital banking 2025.
- George, Gerard, Martine R. Haas, and Alex Pentland. "Big data and management." *Academy of management Journal* 57.2 (2014): 321-326.
- Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, 34(7), 3213-3225.
- Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, 7(1), 1-17.
- Hassani, H., Huang, X., & Silva, E. (2018). Banking with blockchain-ed big data. *Journal of Management Analytics*, 5(4), 256-275.
- Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. *Big Data and Cognitive Computing*, 2(3), 18.
- Hung, P. C. (Ed.). (2016). *Big data applications and use cases*. Springer.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.

- Miller, M. H. (1999). The history of finance. *Journal of Portfolio Management*, 25, 95-101.
- Mishkin, F. S. (2016). *The economics of money, banking and financial markets* (11th ed., global ed.). Pearson Education.
- Mishkin, F. S. , Matthews, K., Giuliadori, M. (2013). *The economics of money, banking and financial markets European Edition*. Pearson Education.
- Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4), 431-448.
- Ranjan, J. (2019). The 10 Vs of Big Data framework in the Context of 5 Industry Verticals. *Productivity*, 59(4).
- Ravi, V., & Kamaruddin, S. (2017). Big data analytics enabled smart financial services: opportunities and challenges. In *Big Data Analytics: 5th International Conference, BDA 2017, Hyderabad, India, December 12-15, 2017, Proceedings 5* (pp. 15-39). Springer International Publishing.
- Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE.
- Sosiaali- ja terveystieteiden ministeriö, 2023 [Luettu: 25.5.2023]
<https://stm.fi/vakuutuslaitokset>
- Sun, H., Rabbani, M. R., Sial, M. S., Yu, S., Filipe, J. A., & Cherian, J. (2020). Identifying big data's opportunities, challenges, and implications in finance. *Mathematics*, 8(10), 1738.
- Tanwar, M., Duggal, R., & Khatri, S. K. (2015, September). Unravelling unstructured data: A wealth of information in big data. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.
- Zakir, J., Seymour, T., & Berg, K. (2015). Big Data Analytics. *Issues in Information Systems*, 16(2).
- Warin, T., & Stojkov, A. (2021). Machine learning in finance: a metadata-based systematic review of the literature. *Journal of Risk and Financial Management*, 14(7), 302.
- Qin, K., Zhou, L., Afonin, Y., Lazzaretti, L., & Gervais, A. (2021). CeFi vs. DeFi-- Comparing Centralized to Decentralized Finance. *arXiv preprint arXiv:2106.08157*.