

Timo Koskimäki

DATA-ANALYTIikka RAHOITUSMARKKINOILLA



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2021

TIIVISTELMÄ

Koskimäki, Timo

Data-analytiikka rahoitusmarkkinoilla

Jyväskylä: Jyväskylän yliopisto, 2021, 28 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Taipalus, Toni

Tässä tutkielmassa tarkastellaan data-analytiikkaa rahoitusmarkkinoilla. Tutkielman tavoitteena on kertoa, miten data-analytiikkaa hyödynnetään rahoitusmarkkinoilla. Tutkielman aihe on erityisen tärkeää, sillä digitalisaation mukanaan tuoma teknologiamurros on muokannut myös rahoitusmarkkinoita suuresti. Rahoitusmarkkinoiden siirryttyä toimimaan sähköisille alustoille, syntyy valtavia määriä dataa. Tätä dataa hyödyntämällä voidaan tehdä päätöksiä, jotka ovat tehokkaampia, tuottavampia ja joita ei välttämättä olisi syntynyt ilman tätä dataa. Tutkimusaihe on alitutkittu ja onkin ensiarvoisen tärkeää tehdä aiheesta lisää tutkimusta. Tutkimuksessa käydään läpi aiheeseen liittyvää terministöä, joka selitetään aikaisempien tutkimusten pohjalta ja näihin viitaten. Tutkimuksessa selvisi, että data-analytiikkaa hyödynnetään monin tavoin finanssialalla. Tästä esimerkkinä korkeataajuinen kaupankäynti, taloudellisten tietojen hallinta ja viitetietojen hallinta, sääntely, niiden toteuttaminen ja niihin valmistautuminen, riskianalyysi, kaupankäynnin analyysi, luottoluokitus ja ennakoiva analyysi, sekä seuranta ja raportointi. Tästä tutkielmasta on hyötyä niin aiheesta kiinnostuneelle, kuin rahoitusmarkkinoilla toimivalle taholle, joka käyttää tai harkitsee aloittavansa data-analytiikan käytön päätöksenteossa rahoitusmarkkinoilla. Tutkielma antaa yleiskuvan siitä, miten data-analytiikkaa hyödynnetään rahoitusmarkkinoilla.

Asiasanat: big data, data-analytiikkaa, rahoitusmarkkinat

ABSTRACT

Koskimäki, Timo

Data analytics in financial markets

Jyväskylä: University of Jyväskylä, 2018, 28 p.

Information System Science, Bachelor's Thesis

Supervisor: Taipalus, Toni

This thesis examines data analytics in financial markets. The aim of the thesis is to explain how data analytics is utilized in the financial markets. The topic of the thesis is particularly important, as the technological revolution brought about by digitalization has also greatly shaped the financial markets. As financial markets move to electronic platforms, huge amounts of data will be generated. Utilizing this data can make decisions that are more efficient, productive and that might not have occurred without this data. The research topic is under-researched, and it is important to do more research on the topic. The study reviews related terminology, which is explained based on previous studies. The study found that data analytics is utilized in many ways in the financial sector. Examples include financial information management and reference data management, implementation and preparation, regulation, risk analysis, trading analysis, credit rating or proactive analysis, and monitoring and reporting. This thesis is useful for both those interested in the topic and those operating in the financial markets who use or are considering starting to use data analytics in decision-making in the financial field. The thesis provides an overview of how data analytics is utilized in the financial markets.

Keywords: big data, data-analytics, financial markets

KUVIOT

Kuvio 1 Big datan 7V:n malli	11
Kuvio 2 Konseptimalli big data-korkeataajuuskaupassa	21

TAULUKOT

Taulukko 1 Tiedonlouhintaprosessin vaiheet	9
Taulukko 2 Alueet, joilla pääomamarkkinoilla hyödynnetään big dataa	18

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
KUVIOT	4
TAULUKOT	4
SISÄLLYS.....	5
1 JOHDANTO.....	6
2 MÄÄRITELMIÄ.....	8
2.1 Data-analytiikka.....	8
2.2 Big data.....	9
2.3 Rahoitusmarkkinat	11
3 DATA-ANALYTIKKA RAHOITUSMARKKINOILLA	14
3.1 Rahoitusinnovaatiot	14
3.2 Big data -analytiikka rahoitusmarkkinoilla	16
3.3 Korkeataajuuskaupan vaikutukset rahoitusmarkkinoihin.....	20
3.4 Digitalisaation mukanaan tuomat riskit rahoitusmarkkinoille	23
4 YHTEENVETO	24
LÄHTEET	26

1 JOHDANTO

”Et voi hallita sitä mitä et mittaa” sanonnassa on paljon viisautta. Yritykset ja organisaatiot voivat suurten tietomassojen avulla mitata ja siten tietää enemmän liiketoiminnastaan ja kääntää tiedon suoraan päätöksenteon ja suorituskyvyn parantamiseksi. Voimme hallita ja mitata tarkemmin kuin koskaan ennen. Voimme tehdä parempia ennusteita ja älykkäämpiä päätöksiä. Voimme kohdistaa voimavaroja tehokkaampiin toimenpiteisiin, jotka perustuvat tietoon, eivätkä intuitioon. (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012)

Digitalisaation myötä niin yksilöiden, yritysten, pankkien kuin kaikkien rahoitusmarkkinoilla toimivien elinten toiminta on siirtynyt paperilta digilaitteille. Tätä kutsutaan dataperusteiseksi teknologiamurrokseksi. (Kanniainen, & Pekkola, 2016) Rahoitusmarkkinoilla dataa kerätään jokapäiväisistä toiminnoista, kuten markkinoiden muillakin osa-alueilla. Tässä kuvailevassa kirjallisuuskatsauksessa pyritään lähestymään aihetta tutkimuskysymyksellä ”Miten rahoitusmarkkinoilla hyödynnetään data-analytiikkaa?”.

Tutkimuslähteitä kirjallisuuskatsaukseen haetaan pääosin Jykdok tietokannasta, mutta myös IEEEExplore sekä Google Scholar hakupalveluista. Hakuosoina käytetään lähinnä seuraavia termejä: *”data analytics”*, *”big data”*, *”finance”*, *”financial markets”* *”data analytics in finance”*, *”data-analytiikka rahoitusmarkkinoilla”* ja *”data analytics in financial markets”*. Lähteitä arvioidaan kriittisesti hyödyntäen Julkaisuforumin tasoluokituksia. Tutkimuksessa käytetään ainoastaan tason 1 ja sitä korkeammiksi luokiteltuja lähteitä. Mikäli Julkaisuforumi ei ole arvioinut lähteen julkaisualustaa, voidaan sitä arvioida itse siteerausmäärien, tutkimuksen tekijöiden ja tutkimuksessa viitattaviin lähteisiin perustuen, lähteen iän, sekä lähteen totuudellisuuden ja puolueettomuuden mukaisesti.

Tutkielman rakenne muodostuu siten, että tätä johdantolukua seuraa luku, jossa määritellään aiheeseen liittyviä termejä aikaisempien akateemisten määritelmien mukaisesti. Tätä lukua seuraa itse kirjallisuuskatsauksen pääaihe, eli *”Data-analytiikka rahoitusmarkkinoilla”*. Tässä luvussa käydään kattavasta aiheen tutkimuksia läpi ja vastataan tutkimuskysymykseen, *”Miten data-analytiikkaa hyödynnetään rahoitusmarkkinoilla?”*. Tutkimus osoittaa, että data-analytiikkaa hyödynnetään monin eri tavoin rahoitusmarkkinoilla. Suurin

data-analytiikkaa hyödyntävä menetelmä on tutkimuksen mukaan korkeataajainen kaupankäynti, joka hyödyntää algoritmeja päätöksenteossa ja kaupankäynnissä. Muita data-analytiikkaa hyödyntäviä rahoituksen osa-alueita ovat muun muassa riskienhallinta, taloudellisten tietojen hallinta, sääntely, kaupankäynnin ja ennakkoinnin analyysi, sekä seuranta ja raportointi. Tätä lukua seuraa vielä yhteenveto, joka kattaa motivaation tutkimukseen, tutkimusmenetelmän, tärkeimmät tulokset ja niistä tehdyt johtopäätökset ja niiden merkityksen, sekä jatkotutkimusaiheet.

2 MÄÄRITELMIÄ

Informaatio maailmassa tuplaantuu joka 20. kuukausi. Tärkeimmät data-lähteet ovat liiketoiminta- ja teollisuusprosessit, tekstit ja strukturoidut tietokannat, kuvat ja videot, sekä fyysiset- ja biolääketieteelliset-tiedot. Data-analytiikan avulla voidaan löytää asiaankuuluvia tietoja, rakenteita ja malleja, saada uusia oivalluksia, tunnistaa syyt ja seuraukset, ennustaa tulevaa kehitystä tai ehdottaa optimaalisia päätöksiä. (Runkler, 2016) Tässä luvussa käydään läpi aiheen "Data-analytiikka rahoitusmarkkinoilla" sisältämiä termejä ja niiden yleisiä määritelmiä.

2.1 Data-analytiikka

Data-analytiikka (data-analytics) määritellään tietojärjestelmien soveltamisena suurten tietojoukkojen analysointiin päätösten tueksi. Data-analytiikka on hyvin monialainen ala, joka on ottanut huomioon muita tieteenaloja, kuten tilastot, koneoppimisen, mallintunnistuksen, systeemiteorian, operaatiotutkimuksen ja tekoälyn. Tyypilliset data-analyysiprojektit voidaan jakaa useisiin vaiheisiin. Tiedot arvioidaan ja valitaan, puhdistetaan ja suodatetaan, visualisoidaan ja analysoidaan ja analyysitulokset lopulta tulkitaan ja arvioidaan. (Runkler, 2016)

Data-analytiikkaan liittyy myös monia muita asiaan liittyviä sivutermejä. Datamelu (noisy data) on Xiong, Pandey, Steinbach ja Kumar (2006) mukaisesti irrelevanttia eli ei asiaan liittyvää tai merkityksetöntä dataa. Lisäksi he toteavat, että tiedonpuhdistusmenetelmät (data cleaning methods) keskittyvät pääosin melun havaitsemiseen ja poistamiseen. Tarve melun poistamiselle on selvä, sillä se on haitallista melkein pä minkä tahansa tyyppiselle data-analyysille. Melu johtuu tiedonkeruuprosessin puutteista ja melu voi koostua asiaankuulumattomista tai heikosti merkityksellisistä dataobjekteista. Parempi melunpoisto tuo tarkempaa data-analyysia. (Xiong ym., 2006)

Data-analytiikan ongelmat (errors) ovat hyvin samankaltaisia kuin missä tahansa todellisissa sovelluksissa ilmenevät ongelmat. Data sisältää virheitä ja melua (noisy data) ja ne on skaalattava ja muunnettava tai ne on kerättävä erilaisista ja mahdollisesti monimuotoisista tietolähteistä. Poikkeamat (outliers), jotka määritellään yksittäisiksi tiedoiksi, joilla on suuret poikkeamat normaalista, on tunnistettava ja poistettava tai korjattava. Poikkeuksia tai melua voidaan vähentää suodattamalla. Muun tyyppisiä virheitä tai ongelmia ovat väärin kaavojen käyttö johdettujen tietojen laskennassa tai mittausvirheet, jotka johtuvat väärästä kalibroinnista, väärästä skaalauksesta tai anturin kulkeutumisesta. (Runkler, 2016)

Tiedonlouhinnan (data mining) ja tiedon hankinnan (knowledge discovery in databases) tavoitteena on poimia tieto tiedosta. Tässä yhteydessä tieto määritellään mielenkiintoisiksi malleiksi, jotka ovat yleensä päteviä, uusia, hyödyllisiä ja ymmärrettäviä ihmisille. (Runkler, 2016) Termiä tiedonlouhinta (data mining) käytetään yleensä operatiiviseen menettelyyn uusien oivallusten saamiseksi, kun taas tutkimusyhteisö käyttää termiä tietonhankinta tietokannoista (knowledge discovery in databases) uusia tekniikoita kehittäessä (Chung & Gray, 1999).

Tiedonlouhintaprosessi (data mining process) sisältää Runkler'n (2016) mukaan kuusi vaihetta: valinta, esikäsittely, muunnos, tiedonlouhinta, tulkinta ja arviointi. Chung ja Gray (1999) määrittelevät sen hieman laajemmin. He jakavat tiedonlouhintaprosessin yhdeksään vaiheeseen (taulukko 1).

Taulukko 1 Tiedonlouhintaprosessin vaiheet (Chung & Gray, 1999, s. 12)

Seuraavat vaiheet kuuluvat yleensä tiedonlouhintaprosessiin. Vaiheet ovat iteratiivisia, jolloin prosessi voi siirtyä taaksepäin aina tarvittaessa.
1. Ymmärrä käytettävää sovellusta, aiempaa asiaankuuluvaa tietoa ja loppukäyttäjän tavoitteita.
2. Luo kohdetietojoukko, jota käytetään tiedonlouhintaan.
3. Puhdista ja esikäsittele tiedot
4. Pienennä muuttujien lukumäärää ja etsi datan muuttumattomia esityksiä, jos mahdollista.
5. Valitse tiedonlouhintatehtävä (luokittelu, regressio, klusterointi jne.).
6. Valitse tiedonlouhinnan algoritmi.
7. Etsi kiinnostavia malleja (datanlouhintavaihe)
8. Tulkitse louhittuja tietoja. Toista tarvittaessa vaiheet 1-7.
9. Vahvista löydetty tieto ja laadi raportti

2.2 Big data

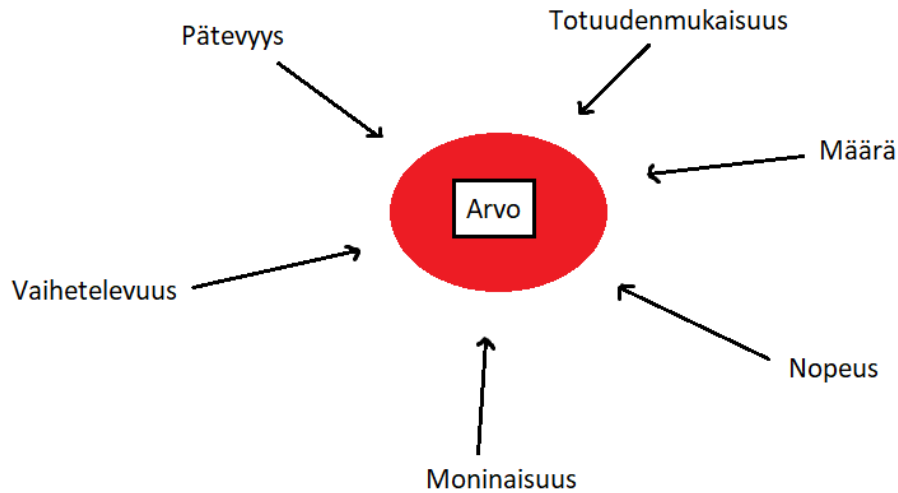
Big data on digitaalisella vuosisadalla laajentanut rajansa tieteellisistä tiedoista esineiden internetissä oleviin anturitietoihin. Tämä tuo lisää ominaisuuksia: tunnetua määrää (volume), nopeutta (velocity), vaihtelevuutta (variety), to-

denmukaisuutta (veracity) ja arvoa (value) Samalla tavalla laajennus tuo uusia haasteita big datan käsittelyyn ja analyysiin. Big data -teknologiat pyrkivät käsittelemään suuria määriä nopeita ja hyvin erilaisia tietoja, kuten teksti-, kuva-, ääni- ja videotiedostoja, aiotun data-arvon keräämiseksi ja alkuperäisten tietojen korkean oikeellisuuden ja saadun tiedon varmistamiseksi. Tämä vaatii kustannustehokasta ja innovatiivisia datan ja tiedon käsittelyn muotoja paremman näkemyksen, päätöksenteon ja prosessien hallinnan parantamiseksi. Kaikki nämä vaativat uusia tietomalleja, datatiloja ja uusia infrastruktuuripalveluja ja -työkaluja, jotka mahdollistavat myös tietojen hankkimisen useista lähteistä ja tietojen toimittamisen eri muodoissa erilaisille datan ja tiedon kuluttajille ja laitteille. (Hsu, H., Chang, C. & Hsu, C., 2017)

Big datan määritelmistä on erilaisia mielipiteitä. Laney'n (2001) kolmen V:n määritelmä sisältää kolme ulottuvuutta (määrä, nopeus, vaihtelu), jotka ovat yhteisiä useimmille määritelmille. Näiden kolmen V:n lisäksi moniin määritelmiin sisältyy teknisiä osia ja komponentteja, jotka liittyvät tietojen aiotuun käyttöön, kuten analyysiin tai päätöksentekoon. Monet määritelmät ovat loogisesti epä johdonmukaisia, mikä on yksi syy termin, "big data", epämääräisyyteen. Tyypillinen puute on sisällyttää määritelmään tiedot ja niiden aiottu käyttö. Nämä tulisivat erottaa toisistaan. (Ylijoki & Porras, 2016)

Tässä tutkielmassa termin "big data" määritelmänä ei kuitenkaan käytetä Ylijoki & Portaan (2016, s. 73-74) mukaisesti Laney'n (2001) 3V:n mallia pitäen määritelmän yksinkertaisena ja helpommin ymmärrettävänä, vaan otamme rahoitusmarkkinoilla yleisesti käytetyn 7V -mallin (Van Rijmenam, 2013) Seddon ja Currie'n (2016) tavoin käyttöön. 7V malli Ali-Ud-Din Khan, Uddin ja Gupta:n (2014) mukaisesti datan arvo (value) on keskeisin asia big datan määrittelyn kannalta. Arvon lisäksi malli sisältää pätevyyden (validity), totuudenmukaisuuden (veracity), määrän (volume), nopeuden (velocity), moninaisuuden (variety) ja vaihtelevuuden (volatility).

Big data sai alkunsa sen suuresta määrästä, mutta se mitä big datan avulla pyritään saamaan aikaan, on arvokasta tietoa jostain asiasta. Tästä syystä arvo on noussut big datan määrittelyn keskiöön. (Ali-Ud-Din Khan ym., 2014) Seuraavaksi 7V:n malli visuaalisesti esitettynä Ali-Ud-Din Khan ym. (2014) mukaisesti (kuvio 1).



Kuvio 1 Big datan 7V:n malli (Ali-Ud-Din Khan, Uddin & Gupta, 2014, s.2)

2.3 Rahoitusmarkkinat

Rahoitusmarkkinoilla (financial markets) on keskeinen rooli yksityisomistukseen perustuvissa talouksissa, eli kapitalistisissa talouksissa. Markkinatalous edellyttää aina rahoitusjärjestelmää, jonka kautta investoinnit ja säästöt välitetään. Tämä mahdollistaa riskien siirtämisen osapuolelta toiselle. Markkinoilla toimiva elin voi laskea liikkeelle arvo-osuuksia omasta liiketoiminnastaan ja näin ollen saa vastineeksi pääomaa liiketoiminnalleen. Tavallinen markkinoilla toimiva elin voi olla yksilö tai perhe, yritys, rahoitusinstituutio tai hallitus. Lisäksi markkinoilla toimijat kohtaavat erilaisia riskejä, joita he haluavat vakuuttaa esimerkiksi ostamalla johdannaisopimuksia. Rahoitusmarkkinat siis tarjoavat keinon hinnoittaa erilaisia arvopapereita tavalla, joka käytännössä kertoo talouden pääoman jakautumisesta. (Antelo & Peon, 2011) Kovács ja Kajtor-Wieland (2017) mukaan rahoitusmarkkinoilla rahoitusvälineiden myyjät ja ostajat kohtaavat. Tämä auttaa heidän mukaansa investoijia löytämään rahoitusta toiminnalleen (Kovács & Kajtor-Wieland, 2017).

Rahoitus on osa taloustieteitä ja yhteiskuntatieteitä, jotka tutkivat, miten ihmiset jakavat niukat resurssit, joilla on vaihtoehtoisia käyttötarkoituksia, kilpailevien tavoitteiden joukossa. Resurssien riittämättömyys kaikkien tavoitteiden saavuttamiseksi sekä mahdolliset vaihtoehtoiset käyttötavat ovat välttämättömiä taloudellisten ongelmien ainesosia. Rahoitus tutkii tällaisia ongelmia vaihtoehtojen kanssa, joihin liittyy rahaa, riskiä ja aikaa. Taloudelliset ongelmat voivat koskea yritystoimintaa, jolloin puhumme yritysrahoituksesta, mutta se voi koskea myös yksityishenkilöitä (henkilökohtainen talous), hallituksia (julkinen talous) ja muita organisaatioita. Taloudelliset valinnat voidaan tehdä suo-

raan tai edustajien välityksellä, kuten osakkeenomistajien puolesta toimivien yritysjohtajien tai sijoittajien puolesta toimivien rahastojen hoitajien kautta. Rahoitusmarkkinat helpottavat, yksinkertaistavat ja lisäävät mahdollisuuksia valita. (Van Der Wijst, 2013)

Rahoitusmarkkinoiden yhteyteen on hyvä liittää myös muutama relevantti termi, jotka koskettavat niin rahoitusmarkkinoita, kuin tulevaa kirjallisuuskatsausta aiheesta. Kvantitatiivinen (quantitative finance) eli määrällinen rahoitus on matemaattisten mallien ja erittäin suurten aineistojen käyttöä rahoitusmarkkinoiden ja arvopapereiden hinnoitteluun ja analysointiin, sekä riskien mittaamiseen (Corporate Finance Institute, 2021).

Korkeataajuuden kauppiat (KTK) terminä on suhteellisen uusi, eikä sillä vielä ole virallista määritelmää (U.S. Securities and Exchange Commission, 2010). Korkeataajuuden kauppiat ovat ammattimaisia kauppiaita, jotka harjoittavat strategioita, jotka tuottavat suuren määrän kauppvoja päivittäin. Korkeataajuuden kauppiat yleensä käyttävät erittäin nopeita ja pitkälle kehitettyjä tietokoneohjelmia tilausten tuottamiseen reitittämiseen ja toteuttamiseen. Heillä on erittäin lyhyet aikataulut positioden muodostamiseksi ja selvittämiseksi, eli kyse voi olla sekunneista, jona aikana tilaus on voimassa. Lisäksi tyypillistä on esittää lukuisia lyhytaikaisia tilauksia, jotka umpeutuvat ennen kaupantekoa, eli tilaukset peruuntuvat ennen tilauksen vahvistumista. Korkeataajuuden kauppiat päättävät kaupankäyntipäivän mahdollisimman lähellä pörssin sulkeutumista, jotta yön aikana olisi mahdollisimman vähän suojaamattomia positioita. (Chordia, Goyal, Lehmann, Saar, 2013) Korkeataajuuden kauppioiden limiittitilausten keskimääräinen voimassaoloaika on alle 10.5 sekuntia (Li, Cooper & Van Vliet, 2018). Commodities Futures Trading Commission (CFTC) (2010) määritelmä korkeataajuuden kauppiaille ottaa hieman erilaisen lähestymisperspektiivin. Heidän mukaansa korkeataajuuden kaupan harjoittajat käyttävät algoritmeja päätöksentekoon, transaktion tekemiseen, luomiseen, reitittämiseen ja jokainen tilaus tehdään ilman ihmisen väliintuloa. Algoritmi tässä yhteydessä määritellään joukoksi ohjeita, jotka käsittelevät markkinatietoja reaaliajassa ja toimittavat tilauksia yhdelle tai useammalle markkinapaikalle (Gomber & Haferkorn, 2013). CFTC:n määritelmässä jätetään tarkoituksella transaktioiden paljous ja salkun suuri vaihtuvuus huomioimatta. Tässä tutkimuksessa käytämme korkeataajuuden kauppiaista CFTC:n määritelmää.

Matalataajuuden kauppiat (MTK) ovat kauppiaita, joiden toimeksiannoilla ei ole rajoitteita, eli he myyvät ja ostavat markkinahinnalla ilman aikaa tai muihin tekijöihin perustuvia rajoituksia. Matalan taajuuden kauppiaisiin kuuluvat myös ne, joiden limiittitilausten keskimääräinen voimassaolo on yli 10.5 sekuntia. Eli ne, joiden toimeksiannot eivät sulkeudu tai toteudu keskimäärin alle 10.5 sekunnissa. (Li ym., 2018) Eli käytännössä matalataajuuden kauppiaisiin kuuluvat pääasiassa yksityishenkilöt, jotka säästävät ja sijoittavat omaisuuttaan.

Volatiliteetti (volatility) mittaa aropaperin tuoton keskihajontaa. Eli kuinka paljon arvopaperin tuotto elää ja täten myös arvopaperin hinta. Tämä on samalla myös yksi kokonaisriskin mittari sijoituskohteelle. Likviditeetti (liquidi-

ty) puolestaan mittaa arvopaperin muutettavuutta rahaksi. Eli kuinka tehokkaasti ja nopeasti arvopaperi kykenee välittämään kauppvoja. (Vaihekoski, 2005)

Flash Crash oli toukokuun 6. päivänä vuonna 2010 Yhdysvaltain markkinoilla tapahtunut järjestelmällinen päivänsisäinen tapahtuma, jossa epätavallisen korkean volatilitiitin ja likviditeetin ohenemisen taustalla oli suuri peruskaupan toimija, joka aloitti myymään yhtä arvopaperia, arvoltaan noin 4,1 miljardia dollaria. Edellä mainittu suuri kauppiaas valitsi automaattisen myyntialgoritmin suorittamaan myyntitoimeksiantoja E-minimarkkinoille, rajarvonaan 9 % edellisen minuutin aikana lasketusta kaupankäyntimäärästä, ilman hinnan ja ajan määritelmää. Tästä syntyi ketjureaktio, jossa myytiin suuri määrä kyseistä arvopaperia hyvin lyhyessä ajassa. Tämän myyntialgoritmin toteuttaminen aiheutti suuren notkahduksen markkinoille muutamien minuuttien ajaksi. (Kirilenko, Kyle, Samadi, Tuzun, 2017)

3 DATA-ANALYTIikka RAHOITUSMARKKINOIL- LA

Tässä tutkielman pääluvussa toteutetaan kuvaileva kirjallisuuskatsaus aiheesta "Data-analytiikka rahoitusmarkkinoilla". Luku jaetaan neljään alalukuun: rahoitusinnovaatiot, big data -analytiikka rahoitusmarkkinoilla, korkeataajuuskaupan vaikutukset rahoitusmarkkinoihin, sekä digitalisaation mukanaan tuomat riskit rahoitusmarkkinoille.

Nykyään kiinnostus big data -alaan ulottuu lukuisille aloille, kuten teollisuus, rahoitus ja terveydenhuolto. Tämä tuo tutkijoille uusia haasteita ja mahdollisuuksia. (Seddon & Currie, 2017). Liiketoiminta- ja tietojärjestelmätutkimuksella on pitkä historia tutkia data-analytiikkaa, jossa keskuskoneiden, henkilökohtaisten tietokoneiden ja viime aikoina Internetin teknologiset muutokset asettavat yritysjohtajille uusia tiedonhallinnan haasteita (Yoo, 2015).

3.1 Rahoitusinnovaatiot

Teknologian luoma globalisaatio ja rahoitusmarkkinoiden suuri korrelaatio maailman kehitykseen on muokannut viime vuosikymmenten aikana rahoitusmarkkinoita suuresti. Tässä luvussa käsitellään digitalisaation mukanaan tuomaa kehitystä rahoitusmarkkinoilla.

Tietojen lisääntyminen on muuttanut rahoitusala dramaattisesti paitsi salkkuanalyyysissä ja riskienhallinnassa, myös vähittäispankkitoiminnassa ja luottopisteytyksessä. Talousdatan jatkuvasti kasvavan määrän, nopeuden ja monimuotoisuuden ohella pääomayritykset ovat tutkineet tapoja tehdä big data hallittavammaksi ja tiivistää valtava määrä tietoa toimiviksi oivalluksiksi pitääkseen kilpailuetunsa muihin yrityksiin nähden. (Fang & Zhang, 2016)

Rahoitusmarkkinat ovat entistä pirstoutuneimpia ja monimutkaisempia, globalisaation ja teknologiamurroksen takia (Funk & Hirschman, 2014) eikä pörseillä ole enää monopoliasemaa arvopapereiden kaupassa (Seddon & Cur-

rie, 2017). Rahoitusmarkkinoiden kolme tärkeintä toimintaa ovat rahastojen hallinta pitkäaikaisille sijoittajille, perinteisten välittäjien matalataajuinen kaupankäynti ja korkeataajuuden kaupankäynti (Blocher, Cooper, Seddon, & Van Vliet, 2016). Biais'n ja Woolley'n (2011) mukaisesti rahoitusmarkkinoiden kaksi päätehtävää on välittää hintasignaaleja talouteen ja antaa markkinatoimijoiden hyötyä tästä tilanteesta. Rahoitusmarkkinoiden kaksi tärkeintä tehtävää omaisuuden hinnoittelussa on likviditeetti ja hintatutkimustietojen sisällyttäminen hintoihin (O'Hara, 2003). Rahoitusmarkkinat perinteisesti turvaavat nämä toiminnot välittäjien avustuksella (Brogaard, Hendershott & Riordan, 2014)

Rahoitusinnovaatiot ovat hyödyllisiä siinä määrin kuin ne tehostavat markkinoiden palvelemissa toimintoja. Korkeataajuuskappale on yksi suurimmista viimeaikaisista rahoitusalan innovaatioista markkinoilla. Korkeataajuuskauppa käyttää kehittyneitä tietokoneohjelmia analysoimaan markkinatietoja ja etsimään kaupankäyntimahdollisuuksia. Tietokoneet sitten kartoittavat nämä tiedot kaupankäyntistrategioihin ja reitittää tilaukset markkinapaikkoihin, ilman ihmisen suoraa väliintuloa. Tietojen saapuminen tietokoneelle ja tilauksen toteutus on luokkaa millisekunteja, paljon nopeammin kuin ihmiset pystyvät rekisteröimään alkuperäiset tiedot. Korkeataajuuden kauppiat kilpailevat nopeudesta, päivittämällä tietokoneita yhä tehokkaammiksi, parantamalla yhteyksiä verkkoon ja hankkimalla parempia ohjelmia. (Biais & Woolley, 2011)

Kvantitatiivinen rahoitus, eli määrällinen rahoitus, on alue, jolla data on elintärkeää tietoa kaikilta osin. Johtavat rahoituslaitokset ja yritykset ottavat käyttöön edistyneitä big data -teknologioita saadakseen käytännöllisiä oivalluksia massiivisista markkinatiedoista, standardoivat rahoitustietoja useista lähteistä, lyhentävät reaaliaikaisiin tietovirtoihin vasteaikaa, sekä parantavat algoritmien ja ohjelmistopinojen skaalautuvuutta uusiin arkkitehtuureihin. Nykyään suuret voitot ajavat finanssialan ammattilaisia kehittämään ja ottamaan käyttöön finanssituotteiden big data -ratkaisuja aina toimiston algoritmisesta kaupasta back-office-tiedonhallintaan ja analytiikkaan. Algoritmisen kaupan ja tutkimuksen helpottamiseksi käytetään laajasti paitsi monilähdedatan keräämistä ja puhdistamista, suuritehoisten datavirtojen tehokasta visualisointia ja nopeaa ohjelmitavuutta massiivisesti rinnakkaisilla prosessointiarkkitehtuureilla. (Fang & Zhang, 2016)

Ala on käyttänyt päivittäisessä käytössä paitsi jäseneltyjä tietoja useista pörssistä, pankeista ja tiedon toimittajilta, myös strukturoimattomia tietoja uutisista ja jopa sosiaalisesta mediasta. Näitä käytetään erilaisiin tarkoituksiin, kuten sijoituspäätösten tekemiseksi, tuotteiden räätälöimiseksi asiakaskohtaiseksi vähittäispankkitoiminnassa kokonaisvaltaisen kuvan saamiseksi yksilön luottokelpoisuudesta ja niin edelleen. (Fang & Zhang, 2016)

Vuosikymmen sitten, osakkeiden pörssihinnat ilmoitettiin seuraavana päivänä. Nykyisillä rahoitusmarkkinoilla osakkeella voi olla noin 500 noterausmuutosta ja noin 150 kauppaa 1 millisekunnin aikana. Eri toimijat ovat suunnitelleet ja ottaneet käyttöönsä erilaisia infrastruktuureja, laitteisto- ja ohjelmistotekniikoita, kuten kymmeniä markkinoita ja paikkoja yhdistävän korkeataajuuden datan saamiseksi ja tilausten jättämiseksi kansallisesti tai

jopa maailmanlaajuisesti erittäin matalalla viiveellä. Yksinkertainen syy sille, että alalla toimivat yritykset, etenkin korkeataajuden kauppaa hyödyntävät yritykset ovat olleet halukkaita investoimaan näihin infrastruktuureihin tai tekniikoihin kymmeniä miljoonia dollareita saadakseen pienempää viivettä toiminnalleen on se, että alalla ”millisekunnit tarkoittavat miljoonia”. (Fang & Zhang, 2016)

Kuten yritysmaailmassa, niin myös rahoitusmarkkinoilla pärjätäkseen täytyy investoida ja innovoida. Kilpailuetujen tavoittelun seurauksena on syntynyt hyvin paljon erilaisia innovaatioita, joita nykypäivänä hyödynnetään jokapäiväisessä kaupankäynnissä rahoitusmarkkinoilla. Edempänä tässä luvussa mainitut innovaatiot ovat hyviä esimerkkejä näistä menestyneistä innovaatioista.

3.2 Big data -analytiikka rahoitusmarkkinoilla

Miten big data -teknologioita ja analytiikkaa hyödynnetään rahoitusmarkkinoilla. Tämä luku käsittelee data-analytiikkaa rahoitusmarkkinoilla nimenomaan big datan näkökulmasta.

Big datan käsite finanssialalla on erilainen kuin se tieteellisessä tai vähittäiskaupan yhteydessä olisi. Esimerkiksi vähittäiskaupan yrityksissä asiakkaiden profiloinnin analyysi sisältää lähinnä sosiaalisen median lähteistä peräisin olevan strukturoimattoman datan analyysin. Rahoitusmarkkinat käsittelevät kuitenkin ensisijaisesti strukturoitua dataa, joka on kerätty rajoitetuista lähteistä, kuten vaihdoista ja tiedon toimittajista. Vaikka yritysten kanssa on käytetty jäsentämättömiä tietojoukkoja mielipiteiden analysointiin ja kaupankäyntiin, nämä eivät ole perinteisesti olleet liiketoiminnalle ensisijaisen tärkeitä aineistoja. (Singh, 2014) Strukturoitu, eli jäsenneilty tieto on tietoa, jolla on kiinteä rakenne ja pituus. Rahoituslalla suurin osa jäsenneilyistä tiedoista on aikasarjoja. Markkinoilla on erilaisia jäsenneiltyjä tietoja. Instrumenttityyppien perusteella markkinoilla on osakkeita, futuureja, optioita, ETF:a ja OTC-osakkeita. Eri markkinoilla ja tapahtumapaikoilla on yleensä erilaiset muodot jopa samalle instrumentille. Strukturoimaton data, eli jäsenneilemätön data, on järjestämätöntä tietoa, joka ei kuulu ennalta määritettyyn malliin. Tämä sisältää tietoja, jotka on kerätty sosiaalisen median lähteistä, kuten artikkeleista, säätiedoista, twiiteistä, sähköposteista tai jopa ääni- ja videotiedoista. (Fang & Zhang, 2016)

Rahoitusmarkkinoilla big data -ongelmien ei katsota edustavan Laney'n 3 V:n mallin mitään V:tä yksin. Taloudellisessa kontekstissa big data haasteissa viitataan yleensä hankkeisiin, joihin liittyy useita tekijöitä. Nykyään reaaliaikainen suoratoistodata on laajasti saatavilla. Tietojen lisääntyminen muuttaa merkittävästi finanssiyritysten liiketoimintamalleja, joko markkina-

takauksessa tai pitkäaikaisessa salkunhoidossa. Jopa pitkäaikaissijoitusten salkunhoitajat lisäävät nykyään tietopohjaisten signaalien näyttöjä salkunvalintamalleihinsa epävakauden ja melun tiivistämiseksi ja puhtaiden tuottojen saamiseksi sijoittajille. Salkunhoitajat, jotka jättävät huomiotta tai tutkivat liian vähän saatavilla olevaa tietoa, lisäävät huomattavan riskin sijoitussalkkuihinsa. (Fang & Zhang, 2016)

Yhdistämällä tiedonhallinnan perinteisiin siiloihin rahoitusyhtiöt pystyvät hallitsemaan salkkuja, analysoimaan riskialttiutta, suorittamaan yritystason analyysia ja noudattamaan säännöksiä kokonaisvaltaisemmasta näkökulmasta. Jatkuvasti kasvavat tietomäärät ja vaatimukset niiden nopeasta saatavuudesta, yhdistämisestä, analysoinnista ja toimimisesta rajoitetussa ajassa tekevät perinteisten tekniikoiden, kuten relaatiotietokannan hallintajärjestelmien, suoriutumisen näissä kehittyneissä analyyseissä mahdottomaksi useimmissa tapauksissa. Uusista bid data -tekniikoista tulee kuitenkin korvaamaton kyky vastata nopeasti muuttuviin vaatimuksiin. Ne voivat viime kädessä auttaa yrityksiä löytämään monia innovatiivisia ja strategisia ohjeita, joita yritykset eivät voineet saada aikaisemmin. (Fang & Zhang, 2016)

Seddon ja Currie'n vuonna 2017 tekemä kirjallisuuskatsaus kertoo, että data-analytiikka ja erityisesti big data -analytiikka ovat jääneet vähäiselle teoreettiselle teoriapohjalle ja ovat empiirisesti aliedustettuja yritystutkimuksessa. Big data rahoitusmarkkinoilla kattaa monia asiaankuuluvia alueita, kuten sääntelyn ja niiden noudattamisen, maailmanlaajuiset kaupankäyntistrategiat ja infrastruktuurit, verkkojen, laitosten ja yritysten väliset liiketoimet, sekä kaupankäynnin ja riskienhallinnan algoritmit (MacKenzie & Millo, 2003). Kolmen viime vuosikymmenen aikana rahoitusmarkkinoiden vauhdin, määrän ja alkupeuran muutos on ollut ennennäkemättömän suurta (Wheatley, 2014). Tietokoneistaminen on ollut näiden muutosten eturintamassa korkeataajuuksisen kaupankäynnin, digitaalisen rahansiirron, maksuteknologian, vertaisrahoituksen ja salkkuanalyysin myötä, jotka kaikki riippuvat tiedonkulun nopeudesta ja tarkkuudesta yhä verkottuneemmalla ja kansainvälisemmällä finanssialalla (MacKenzie, 2008).

Big dataa voidaan hyödyntää pääomamarkkinoilla monin tavoin. Taloudellisen tiedon hallinta ja viitetietojen hallinta, tietojen tallentaminen historiallista kauppaa varten, sisäisen tiedon hallinta ja viitetietojen kokonaisvalvonta ovat isot tehtävät finanssialalla. Organisaatioiden on sekoitettava historialliset ja uudet tiedot saadakseen oivalluksia parempien, kustannustehokkaampien ratkaisujen toimittamisesta. Tämän strategisen tarpeen edistäminen luo entistä enemmän kysyntää tehokkaalle, tarkalle viitetietojen hallinnalle. Voi olla hankalaa ylläpitää, käsitellä ja prosessoida tietoja eri omaisuusluokista, jotka ovat peräisin useilta toimittajilta. Big data tarjoaa tähän ongelmaan apua. Big dataa hyödynnetään myös sääntelyssä. Rahoituspalvelut ovat kokeneet paljon sääntelyn muutoksia viimeisen vuosikymmen aikana. Pääomamarkkinat kohtaavat uusia säännöksiä ja valmistautuvat tuleviin säännöksiin. Myös riskianalyysissa hyödynnetään big dataa. Riskianalyysi sisältää petosten lieventämistä, rahanpesun torjuntaa, huijauskauppaa ja

omaisuuden suorituskyvyn jatkuvaa seurantaa ja pääoman sijoituksen sellaisille liiketoiminnan alueille, jotka tuottavat korkeamman riskisopeutetun pääoman tuoton. Kaupankäynnin analyysi myös hyödyntää big dataa. Tämä sisältää useita analyysialueita kaupankäynnissä kuten korkeataajuuskauppa, kauppaa edeltävä päätöksenteko, tukeva analyysi usein käytäville kaupoille, mielipiteiden mittaus ja ajallinen analyysi. Niin myös luottoluokituksessa hyödynnetään big dataa. Verrattuna vanhempiin malleihin, joita käytetään yrityksen luottokelpoisuuden ennustamiseen, voidaan big datan avulla tätä analysoida entistä paremmin, koska big data tarjoaa entistä laajemman valikoiman tietoja tähänkin toimintaan. Nämä big datan avulla saadut muuttujat voidaan integroida perinteisempiin luottoluokitusmalleihin. Tämä voi johtaa parempaan ennakkointikykyyn luottoluokituksen suhteen. Tämä hybridimalli antaisi tarkemman tavan soveltaa luottoluokituksia yrityksille ja hallituksille. Edellä mainittujen toimintojen lisäksi tietojen merkintään voidaan hyödyntää big dataa. Yrityksen seuranta- ja raportointialueella on hyvin vaikeaa sovittaa yhteen kauppvoja erilaisista järjestelmistä, jotka on rakennettu erilaisille tekniikan standardeille. Tämä saattaa johtaa päällekkäisiin, virheellisiin ja tapahtumattomiin kauppvoihin. Datan merkitseminen voi ratkaista sen helposti seuraamalla kaikkia kauppvoja ja tapahtumia. (Singh, 2014) Seuraavaksi esitetään sama vielä yksinkertaistettuna taulukossa (Taulukko 2).

Taulukko 2 Alueet, joilla pääomamarkkinoilla hyödynnetään big dataa (Singh, 2014, s.43-44)

1. Taloudellisten tietojen hallinta ja viitetietojen hallinta
2. Sääntely, niiden toteuttaminen ja niihin valmistautuminen
3. Riskianalyysi
4. Kaupankäynnin analyysi
5. Luottoluokitus ja ennakoiva analyysi
6. Seuranta ja raportointi

Big datan pääominaisuudet, mukaan lukien määrä, vaihtelevuus ja nopeus (Laney'n 3V), on sisällytetty kaikkiin taloudellisiin tietoihin ja markkinoihin. Tietomäärä rahoitusmarkkinoilla on kasvanut valtavasti. Kun algoritmista kaupasta tuli Wall Street'n valtavirta viimeisen vuosikymmenen aikana, pääomamarkkinat siirtyivät myös big data -aikakauteen. Esimerkiksi korkeataajuuden kauppa, joka on kvantitatiivisen kaupankäynnin ensisijainen muoto, käyttää tietokoneiden omia kaupankäyntistrategioita tekemään toimeksiantoja, sekuntien tai sekuntien murto-osien aikana. (Fang & Zhang, 2016)

Tarvittaessa pääsy oikeisiin tietoihin ja kyky analysoida sitä reaaliajassa ja oikeaan aikaan ovat välttämättömiä älykkäille päätöksentekoinvestoinneille (decision-making investment). Yksi tärkeimmistä hyödyistä, mitä big data mukanaan tuo, on sen silmukointikyky reaaliaikaisen palautteen käsittelyssä, reagoida tähän, analysoida tuloksia ja jatkaa optimointia silmukoinnissa. Tämä palautussilmukointi mahdollistaa suurten tietomäärien testaamisen

reaaliajassa, käyttämällä erilaisia tietoja useista lähteistä, lyhyemmässä ajassa. Pääomamarkkinoilla nopeus, jolla tietoja voidaan testata, analysoida ja käyttää, antaa todennäköisen kilpailuedun etenkin tulevan ennustamiseen. (Singh, 2014)

Yksi suurimmista syistä big data -projektien soveltamisen epäonnistumisista on ollut yhteensopivuusongelmat pääomamarkkinoiden liiketoimintavaatimuksissa. big data -projektien tarkoitusta on myös harkittava huolellisesti. Jos liiketoiminnalle merkityksellistä kohdetta ei tunnisteta, big data -strategiat eivät tuota tuottoa tehdyistä sijoituksista. Tästä syystä analyysi voi johtaa ennalta-arvaamattomiin tietovajeihin tai asiayhteyteen liittyvien tietojen puutteeseen. Projektista vastaavien henkilöiden riittämätön tekninen taito ja tieto voi olla myös merkittävä syy projektin epäonnistumiseen. Big data projekti korostaakin riittävän ammattitaitoisen henkilöstön tarvetta tällaisen strategian toteutuksessa. Suurin tekijä tämän suhteen on oikeiden henkilöiden löytäminen tai konsulttien hyödyntäminen, jotka ymmärtävät liiketoiminnan ja tekniikan. Myös tietosuoja on suuri huolenaihe big data projekteissa, etenkin rahoitusmarkkinoilla. Tämä on haaste, joka on yleistä pilvi-tekniikan yhteydessä. Organisaatioiden tulisi huolehtia, että kaupallisesti arkaluonteisten tietojen suojan on oltava riittävä lain ja säännösten mukaisesti. Nämä säännökset voivat lisäksi vaihdella alueittain. (Singh, 2014)

Ammattisijoittajien toiminnan lisäksi nopeuden kasvua tapahtuu myös jokaisen ihmisen tavallisessa elämässä. Verkko- ja mobiililaitteiden käyttö on lisännyt dramaattisesti tapahtumien nopeutta ja tiheyttä kaikille. Ihmiset tilaavat ruokaa, kirjautuvat sisään ja ulos, tekevät talletuksia ja maksavat laskuja mobiilisovellusten avulla muutamissa sekunneissa. Kaikki nämä ovat haasteita ja mahdollisuuksia rahoituslaitoksille. Analytiikka ja kyky hyödyntää tehokkaasti big data -teknologiaa, edistyksellistä tilastomallinnusta ja ennakoivaa analytiikkaa reaaliaikaisen päätöksenteon tueksi liiketoimintakanavien ja toimintojen välillä erottavat epävarmoilla markkinoilla kukoistavat yritykset epäonnistuneista yrityksistä. (Fang & Zhang, 2016)

Big datan käyttöönotto pääomamarkkinoilla on vielä kaukana valmiista. Sen on vielä saavutettava vakaa piste huipputason myyntiyhtiöissä. Vaikka investoinneille onkin monia toteuttamiskelpoisia syitä, ne keskittyvät yleensä erityisiin käyttötapauksiin kattavat big data -strategian sijaan. Ne, jotka ovat tällä hetkellä tietoisia big datan eduista ja ovat onnistuneesti toteuttaneet osan näistä sisäisistä esteistä huolimatta, haluavat varmasti laajentaa näiden strategioiden ja tekniikoiden käyttöä. Nämä toteutukset ovat kuitenkin todennäköisesti kasvuvaiheessa lähitulevaisuudessa. (Singh, 2014)

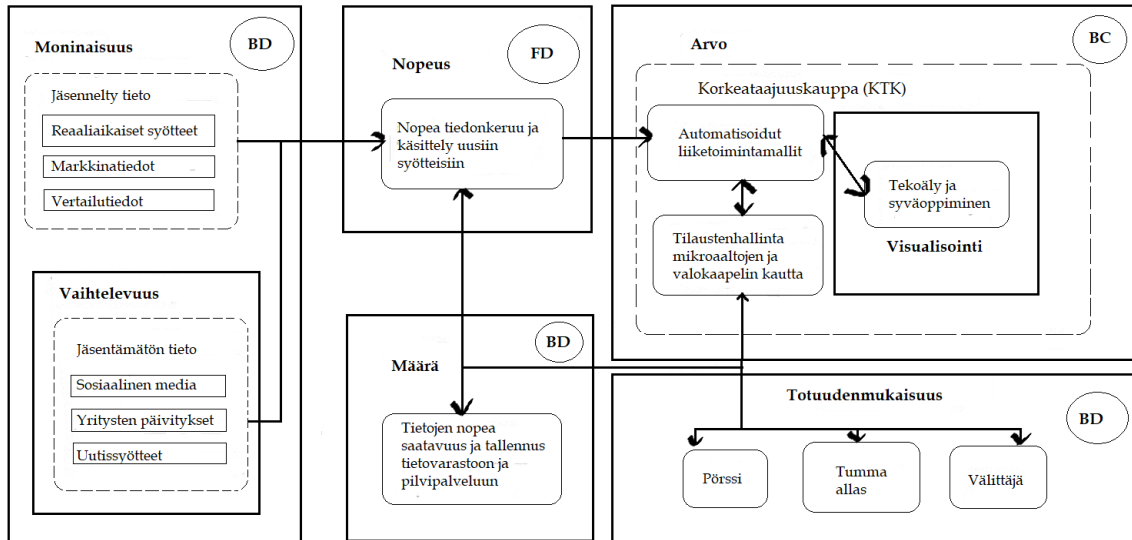
Taloudellisen big datan (financial big data) yksinkertaisin mittaus on tiettyjen osakenimien mainitsemiskertojen määrä, tiettyjen avainsanojen hakutiheys, sekä sisällöstä poimittu mielipide. Wysocki (1998) oli ensimmäinen, joka dokumentoi Yahoo! -yhtiössä listattujen yli 3000 osakkeen viestien lähetysmäärän välisen suhteen. Wysocki havaitsi, että yön yli tapahtuvaa viestien lähetysmäärää voitiin käyttää ennustamaan muutoksia seuraavan päivän osaketuloissa, mikä johti päätökseen, että viestien lähettämisen määrä

liittyy kohdeyritysten perustekijöihin. Da, Engelberg ja Gao (2011) havaitsivat, että sijoittajien huomion kasvu voi ennustaa korkeampia osaketuottoja seuraavien kahden viikon aikana ja että sijoittajien huomio voi myös selittää suuren arvopaperin suuren debyyttipäivän markkinoilla ja myös pitkäaikaisen alituloksen. (Shen, & Chen, 2018)

3.3 Korkeataajuuskaupan vaikutukset rahoitusmarkkinoihin

Miten korkeataajuuskauppa vaikuttaa rahoitusmarkkinoihin? Tähän vastauksen tuo aiheen tieteelliset tutkimukset, joita tässä luvussa käydään läpi. Korkeataajuuskauppa, myöhemmin KTK, hallitsee kaupankäyntien määrällä mitattuna rahoitusmarkkinoita selkeästi (Li ym., 2017). Merkittävä, suurelta osin negatiivinen mediajulkisuus korkeataajuisesta kaupankäynnistä ja "Flash Crash" -pörssiromahduksesta 2010 herätti merkittävää huolta ja kiinnostusta markkinoiden oikeudenmukaisuudesta ja korkeataajuisen kaupankäynnin roolista vakaudessa ja hintatehokkuudessa. Kuitenkin Nasdaq-pörssin tutkijoille luovutettujen tietojen perusteella KTK kaupankäynnillä on suotuinen rooli hintojen löytämisprosessissa ja se pienentää hinnoitteluvirheitä. Tähän suhtaudutaan yleisesti myönteisesti, sillä informatiivisemmat osakekurssit voivat johtaa resursien tehokkaampaan käyttöön markkinoilla. KTK:n tarjoama likviditeetti on suurempaa, kuin sen käyttämä likviditeetti, joten KTK ei aseta haitallisia nettovalintoja muille makrotaloudessa. KTK:n kokonaisvirran negatiivinen yhteys hinnoitteluvirheisiin osoittaa, että KTK ei yleensä liity hintojen manipulointikäyttäytymiseen. Kuitenkin KTK:n yhteydestä markkinoiden vakauteen voidaan olla montaa mieltä. Ne käyvät kauppaa siirtymäkauden hinnoitteluvirheiden vähentämiseksi, sekä keskimääräisinä päivinä, että kaikkein epävakaimpina päivinä. Se, että KTK:n takia likviditeettitoimittajille aiheutuvat valintakustannukset voivat johtaa siihen, että muut kuin KTK:n maksuvalmiuden toimittajat vetäytyvät markkinoilta, kuten Biais, Foucault ja Moinas (2012) ovat esittäneet. Tämä voi epäsuorasti johtaa KTK:n heikentämään markkinoiden vakautta, huolimatta siitä, että KTK:n likviditeettitoimittajat pysyvät aktiivisina. (Brogaard ym., 2014)

Van Rijmenam (2013) 7V:n mallin mukaisesti seuraavaksi avataan Seddon ja Currie'n (2016) mukaisesti big datan selventäminen KTK-yrityksissä. 7V malli on jaettu kolmeen luokkaan, Suuri laskentakyky (Big Computing, BC) suurten tietomäärien käsittelyyn, big data (BD) tietojen tallentamiseen ja käsittelyyn ja kolmantena Nopea tiedonsiirto (Fast Data, FD) reaaliaikaista kaupankäyntiä varten. Teorian mukaan KTK-yritykset priorisoivat nämä ominaisuudet eri tavalla omien markkinastrategioidensa ja käytäntöjensä perusteella. Seuraavaksi kuvio konseptimallista big data korkeataajuuskaupassa Seddon ja Currie (2016) mukaisesti (Kuvio 2):



Kuvio 2 Konseptimalli big data-korkeataajuuskaupassa (Seddon & Currie, 2016, s. 302)

KTK on tekninen keino toteuttaa vakiintuneita kaupankäyntistrategioita. KTK ei ole kaupankäyntistrategia sinänsä, mutta se soveltaa uusinta teknologista kehitystä markkinoille pääsyssä, markkinatietoihin pääsyssä ja toimeksiantojen reititys vakiintuneen kaupankäyntistrategioiden tuoton maksimoimiseksi. Siksi arvioinnissa ja sääntelykeskustelussa KTK:sta tulisi keskittyä strategioihin eikä KTK:aan sellaisenaan. KTK on arvopaperimarkkinoiden luonnollinen kehitys täysin uuden ilmiön sijaan. Uuden tekniikan käyttöönotossa on selkeä evoluutioprosessi kilpailu, innovaatio ja sääntely. Kuten kaikki muutkin tekniikat, algoritmikauppa ja KTK mahdollistavat kehittyneiden markkinaosapuolten saavuttamaan lailliset edut investoinneistaan, etenkin teknologiaan, ja korvauksen markkinoista, vastapuolista ja operatiivisista riskeistä. Monet KTK:aan liittyvät ongelmat juurtuvat Yhdysvaltain markkinarakenteeseen. Flash Crash ja pikatilauksia koskevat keskustelut liittyvät Yhdysvaltain osakemarkkinoihin ja verkkohallintajärjestelmään. Euroopassa, jossa käytetään joustavampaa parhaan toteutuksen järjestelmää ja osakehotainen volatilitteettisuojausjärjestelmä on ollut käytössä jo kahden vuosikymmenen ajan, ei markkinoiden laatuongelmia KTK:aan liittyen ole toistaiseksi dokumentoitu. (Gomber, Arndt, Lutat & Uhle, 2011)

Suurin osa KTK pohjaisista strategioista edistää markkinoiden likviditeettiä (markkinatakausta) hintojen löytämiseen ja markkinoiden tehokkuuteen (arbitraasi strategiat). Estämällä nämä strategiat puutteellisella sääntelyllä tai heikentämällä taustalla olevia liiketoimintamalleja liiallisen taakan kautta, voi aiheuttaa haitallisia ja odottamattomia vaikutuksia markkinoiden laatuun. Valvontaviranomaisten on kuitenkin torjuttava tehokkaasti markkinoiden eheyden vastaista väärinkäyttöä. (Gomber ym., 2011)

Suurin osa akateemisissa kirjallisuudessa esitetystä KTK-pohjaisten strategioiden vaikutuksista markkinoiden laatuun on positiivisia. Enemmistö KTK:aan keskittyvistä asiakirjoista ei löydä todisteita kielteisestä vaikutuksesta markkinoiden laatuun. Päinvastoin suurin osa näistä dokumenteista väittää,

että KTK yleensä myötävaikuttaa markkinoiden laatuun ja hintojen muodostumiseen ja havaitsee positiivisia vaikutuksia likviditeettiin ja lyhyen aikavälin volatilitettiin. Vain yksi artikkeli huomauttaa, että tietyissä olosuhteissa KTK saattaa lisätä haitallisia valintaongelmia. (Gomber ym., 2011)

KTK toimii dynaamisilla ja nopeasti kehittyvillä markkinoilla, joilla yhden yrityksen teknologianinfrastruktuuri kilpailee toisen kanssa (Menkveld, 2013). Tälle toiminnalle ei kuitenkaan ole yleistä määritelmää, koska niiden noudattamilla strategioilla on erilaiset tietovaatimukset, mikä tekee kaiken yleistämisen KTK-yrityksissä vaikeaa (Seddon & Currie, 2017).

KTK:n merkityksellisyys markkinoilla edellyttää valvontaa, mutta myös avoimuutta ja avointa viestintää luottamuksen varmistamiseksi arvopaperimarkkinoilla. Kun otetaan huomioon ihmisten arkuus rahoitusalan innovaatioille kriisin jälkeen, KTK:aa soveltavien yhteisöjen vastuulla on ennakoivasti kommunikoida sisäisistä suojatoimenpiteistä ja riskienhallintamekanismeista. (Gomber ym., 2011)

Jotkin KTK:n taktiikoista on suunniteltu peittämään sopimuksia ja estämään muita markkinaosapuolia löytämästä ja hyödyntämästä kaupankäyntitarkoituksiaan. Siinä määrin, kun suhteellisen pieni määrä korkeataajuuden kauppiaita kattaa merkittävän osan liikevaihdosta markkinoilla, voivat he myös manipuloida markkinoita. Kolme tällaista manipuloinnin muotoa ovat täyte, savustaminen ja huijaaminen. Täyte tarkoittaa korkeataajuuden kauppiaita, jotka toimittavat kohtuuttoman määrän tilauksia markkinoille. Tämä aiheuttaa ruuhkia. Näissä olosuhteissa hitaiden kauppiaiden pääsy markkinoille on heikentynyt. Heillä ei ole selkeää kuvaa kaupankäynnin nykytilasta ja se on syy, miksi heidän on hankala toteuttaa kauppiaita. Samaan aikaan pikakauppiaat, jotka ymmärtävät paremmin tapahtumia ja joilla on ylivoimainen pääsy markkinoiden moottoriin, pystyvät toteuttamaan kannattavia kauppiaita hitaiden kauppiaiden kustannuksella. Savustamista harjoittavat korkeataajuuden kauppiat lähettävät ensin houkuttelevat limiittitilaukset hitaille kauppiaille. Sitten he tarkistavat nämä tilaukset nopeasti vähemmän anteliaisiin ehtoihin toivoen, että joku hitaista kauppiaista tarttuu tähän muokattuun tarjoukseen. Huijausta harjoittavat korkeataajuuden kauppiat, joiden todellinen tarkoitus on ostaa, asettaa alun perin myyntitilauksia tilauskantaan. Näitä tilauksia ei ole tarkoitettu toteutettavaksi, vaan ne asetetaan parhaan tarjouksen yläpuolelle. Ja koska korkeataajuuden kauppias on nopeampi kuin muut markkinaosapuolet, hän voi olla varma, että hänellä on aikaa peruuttaa myyntitilaukset ennen kuin ne toteutetaan. Eli tämä varmuus mielessään korkeataajuuden kauppias asettaa sarjan myyntitilauksia parhaan kysynnän yläpuolelle, mahdollisesti erittäin suurella määrällä. Tavoite on pelotella markkinoita ja saada joku naiivista markkinatoimijoista myymään parhaan tarjouksen ali. (Biais & Woolley, 2011).

Haitallinen valikoituminen koskee niitä markkinaosapuolia, joilla on varhainen ja etuoikeutettu pääsy asiaankuuluvaan tietoon ja kyky käydä sillä kauppaa muiden kanssa. Siksi korkeataajuiset kauppiat ovat valmiita investoimaan kalliisiin laitteisiin ja perimään yhteissijaintimaksuja vaihdossa saa-

dakseen tämän edun. Näiden investointien ansiosta KTK-tilaukset kätkevät nopeasti kaikki käytettävissä olevat tiedot. (Biais & Woolley, 2011).

Koneperusteinen kaupankäynti, joka on vaikuttanut suurempiin osakkeisiin, on saattanut heikentää kaupankäynnin kiinnostusta satojen pienten pörsyhtiön osakkeisiin, mistä on negatiivinen vaikutus likviditeettiin. (Friederich & Payne, 2011). KTK toiminta parantaa sekä matalantaajuuden limiittitilausten likviditeettiä, että tilausten toteutuksen laatua (Li ym., 2018). KTK lisää matalantaajuuden tilausten kaupankäyntitiheyttä ja kauppamäärää, lyhentää odotusaikaa ja parantaa matalantaajuuden rajoitustoimeksiantojen toteutuksen todennäköisyyttä (Li ym., 2018).

Korkeataajuisen kaupankäynnin hyödyt rahoitusmarkkinoilla ovat siis selkeät ja haitat ovat hyvinkin kiistellyt. Rehellisesti ja markkinoiden lakeja kunnioittaen korkeataajuuskauppa saa aikaan hyvin vähän negatiivisia vaikutuksia markkinoihin, kuten edellä käytiin läpi. Pohdittavaa kuitenkin jää korkeataajuuskaupan moraaliseen ulottuvuuteen. Onko oikein hyödyntää ja hyväksikäyttää niin sanottujen hitaiden kauppioiden vanhentuneita hintamäärittäjiä? Ovatko korkeataajuuden kauppiat etuoikeutettuja saadessaan tiedon nopeampaa kuin muu markkina ja huomaavatko he siten asioita, mitä muut markkinoilla toimijat eivät huomaa. Vai ovatko korkeataajuuden kauppiat tehneet vain enemmän töitä kuin muut, saadakseen informaatiota, joka hyödyntää heitä. Näihin kysymyksiin on harvemmin suoraa vastauksia.

3.4 Digitalisaation mukanaan tuomat riskit rahoitusmarkkinoille

Martin Wheatley, Financial Conduct Authorityn pääjohtaja (Riippumaton finanssialan sääntelyelin Britanniassa), vuonna 2014 pitämässään puheessa nostaa esiin teknisiä riskejä, joita rahoitusmarkkinat kohtaavat ja tulevat kohtaamaan: järjestelmien tai tietojen fyysinen menetys, sääntelyviranomaisten lainkäyttövaltaan liittyvät kysymykset, petokset, henkilötietojen mahdollinen väärinkäyttö, sekä tietoverkkorikollisuus (cybercrime). Hänen mukaansa jo puolet maailman pörseistä ovat kokeneet kyberhyökkäyksiä. Yhdeksän kymmenestä yrityksestä on kärsinyt turvallisuusrikkomuksista vuoden 2013 aikana ja 70 % prosenttia pääjohtajista kokee turvallisuusrikkomukset keskeisenä kasvuriskinä. (Wheatley, 2014)

4 YHTEENVETO

Tässä tutkielmassa tarkasteltiin data-analytiikkaa rahoitusmarkkinoilla kuvailevan kirjallisuuskatsauksen mukaisesti. Aluksi tutkielmassa määriteltiin aiheeseen liittyviä termejä, alan akateemisten tutkimusten mukaisesti. Näistä tärkeimpinä Runkler'n (2016) mukaisesti data-analytiikka määritellään tietojärjestelmien soveltamisena suurten tietojoukkojen analysointiin päätösten tueksi ja big data määritellään Seddon ja Currie'n (2016) tavoin Van Rijmenam'n (2013) esittämän 7V-mallin mukaisesti. Tutkielmassa lähestytään 7V -mallia Ali-Ud-Din Khan ym. (2014) mukaisesti datan arvo (value) keskeisimpänä tekijänä big datan määrityksessä. Arvon lisäksi malli sisältää pätevyyden (validity), totuudenmukaisuuden (veracity), määrän (volume), nopeuden (velocity), moninaisuuden (variety) ja vaihtelevuuden (volatility).

Tutkielmassa kerrotaan, että data-analytiikkaa hyödynnetään rahoitusmarkkinoilla hyvin monipuolisesti. Käytetyin ja suosituin tämän hetken data-analytiikan muoto rahoitusmarkkinoilla on korkeataajuinen kaupankäynti, joka tarkoittaa automatisoitujen algoritmien ajamista kaupankäynnin päätöksenteossa. Muita data-analytiikkaa hyödyntäviä menetelmiä rahoitusmarkkinoilla on muun muassa riskianalyysin, luottoluokituksen, ennakoivan analyysin, sekä taloudellisten tietojen hallinnan yhteydessä käytettävä datananalyysi.

Seddon ja Currie'n vuonna 2017 tekemä kirjallisuuskatsaus kertoo, että Data-analytiikka ja erityisesti big data-analytiikka ovat aliteoreetoituja ja empiirisesti aliedustettuja yritystutkimuksessa. Toisaalla Yoo (2015) kertoo tutkimuksessaan data-analytiikan tutkimuksen pitkästä historiasta liiketoiminta- ja tietojärjestelmälalla. Tutkimusmateriaaleja tähän tutkielmaan etsiessäni, itse huomasin nimenomaan "data-analytiikka rahoitusmarkkinoilla" - käsitteen alle mahtuvan tieteellisen materiaalin vähyyden. Aihe on minusta todella alituttu ja vaatii kipeästi laadukasta ja kattavaa tutkimusta.

Big Data rahoitusmarkkinoilla kattaa monia asiaankuuluvia alueita, kuten sääntelyn ja niiden noudattamisen, maailmanlaajuiset kaupankäyntistrategiat ja infrastruktuurit, verkkojen, laitosten ja yritysten väliset liiketoimet, sekä kaupankäynnin ja riskienhallinnan algoritmit (MacKenzie & Millo, 2003).

Tutkimuksen motiivina toimi henkilökohtainen kiinnostus rahoitusmarkkinoita ja data-analytiikkaa kohtaan, sekä big datan jatkuvasti merkittävämpi läsnäolo niin rahoitusmarkkinoilla, kuin myös muilla aloilla. Data-analytiikka tulee olemaan ja jo osittain onkin markkinoilla pärjäämisen elin ehto niin yritysjohdon, kuin rahoitusmarkkinoiden puolella.

Tälle tutkimukselle tarvitaan jatkotutkimusta ja sen tulisi kohdistua rahoitusmarkkinoilla tapahtuvan kaupankäynnin data-analytiikkaan. Miten data-analytiikkaa hyödynnetään käytännössä rahoitus ja investointipäätöksenteossa? Miten sitä voitaisiin hyödyntää entistä tehokkaammin? Miten voidaan tutkia sijoituskohteen kannattavuutta entistä tarkemmin, ennustettavammin ja nopeammin? Muun muassa näihin kysymyksiin tarvittaisiin vastauksia jatkotutkimuksen muodossa. Tämä tuottaisi tarpeellista tietoa niin rahoituslaitoksille, investointipäätöksiä tekeville yrityksille ja organisaatioille, kuin myös yksilöille, jotka säästävät ja sijoittavat omaisuuttaan.

LÄHTEET

Ali-Ud-Din Khan M., Uddin M. F. & Gupta, N. (2014). *Seven V's of Big Data understanding Big Data to extract value*

Antelo M. & Peon D. (2011). *Financial markets: A guided tour*. Nova Science Publishers.

Biais B., Foucault T. & Moinas S. (2012). Equilibrium High-Frequency Trading. *SSRN Electronic Journal*

Biais B. & Woolley P. (2011). High frequency trading. *Manuscript, Toulouse University, IDEI*.

Blocher J., Cooper R., Seddon J. & Van Vliet B. (2016). Phantom liquidity and high-frequency quoting. *The Journal of Trading*, 11(3), 6–15.

Brogaard J., Hendershott T. & Riordan R. (2014). High-Frequency Trading and Price Discovery. *The Review of Financial Studies*, 27(8), 2267–2306.

Chung H. M. & Gray P. (1999). Data mining. *Journal of management information systems*, 16(1), 11-16.

Chordia T., Goyal A., Lehmann B. N. & Saar G (2013). High-Frequency Trading. *Johnson School Research Paper Series No. #20-2013*

Corporate Finance Institute (2021). Haettu 1.5.2021 osoitteesta <https://corporatefinanceinstitute.com/resources/knowledge/finance/quantitative-finance/>

Da Z., Engelberg J. & Gao P. (2011). In search of attention. *Journal of Finance*, 66(5), 1461–1499.

Fang B. & Zhang P. (2016) Big Data in Finance. In: Yu S., Guo S. (eds) *Big Data Concepts, Theories, and Applications*

Friederich S. & Payne R. (2011). Computer based trading, liquidity, and trading costs.

Funk R. J. & Hirschman D. (2014). Derivatives and Deregulation: Financial Innovation and the Demise of Glass-Steagall. *Administrative science quarterly*, 59(4), 669–704.

Gomber P., Arndt B., Lutat M. & Uhle T (2011) High-Frequency Trading. House of Finance, Goethe University Frankfurt

Gomber P. & Haferkorn M. (2013). High-Frequency-Trading. *Business & Information Systems Engineering*, 5(2), 97-99.

Hsu H., Chang C. & Hsu C. (2017). *Big data analytics for sensor-network collected intelligence*. Academic Press, an imprint of Elsevier.

Kanniainen, J., & Pekkola, S. (2016). Big data rahoitusmarkkinoilla. *Futura* 35 (2016): 2.

Kirilenko A., Kyle A., Samadi M., Tuzun T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market. *Journal of Finance*, 72(3), 967-998

Kovács L. & Kajtor-Wieland I. (2017). The differentiation and definition of money and capital markets. *Forum on Economics and Business*, 20(133), 21.

Laney D., 2001. 3D data management: Controlling data volume, velocity, and variety. META Group Research Note 6, 70.

Li K., Cooper R. & Van Vliet B. (2018). How Does High-Frequency Trading Affect Low-Frequency Trading? *The Journal of Behavioral Finance*, 19(2), 235-248

McAfee A., Brynjolfsson E., Davenport T. H., Patil D. J. & Barton D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

MacKenzie D. (2008). *An engine, not a camera: How financial models shape markets*. Mit Press.

MacKenzie D., & Millo Y. (2003). Constructing a market, performing theory: The historical sociology of a financial derivatives exchange. *American journal of sociology*, 109(1), 107-145.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of financial Markets*, 16(4), 712-740.

O'Hara M. 2003. Liquidity and price discovery. *Journal of Finance* 58: 1335-54.

Runkler T. A. k. (2016). *Data Analytics: Models and Algorithms for Intelligent Data Analysis* (2nd ed. 2016.). Springer Fachmedien Wiesbaden.

Seddon J. & Currie W. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, 70, 300.

Shen D., & Chen S. H. (2018). Big data finance and financial markets. In *Big Data in Computational Social Science and Humanities* (pp. 235-248). Springer, Cham.

Singh M. (2014). Big Data in Capital Markets. *International Journal of Computer Applications*, 107(5).

U.S. Commodities Futures Trading Commission and the U.S. Securities and Exchange Commission, 2010. Preliminary findings regarding the market events of May 6, 2010. <https://www.sec.gov/sec-cftc-prelimreport.pdf>

U.S. Securities and Exchange Commission, 2010, Concept release on equity market structure 34- 61358. <https://www.sec.gov/rules/concept/2010/34-61358.pdf>

Vaihekoski M. (2005). Rahoitusalan sanasto - Glossary of Financial terms in finnish. https://users.utu.fi/moovai/mv_sanasto.html

Van Der Wijst N. (2013). *Finance: a quantitative introduction*. Cambridge University Press.

Van Rijmenam A. (2013). Why the 3 V's are not sufficient to describe big data

Wheatley M. (2014). The technology challenge. *Financial conduct authority (speech, 10/06/2014)*

Wysocki P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, (98025).

Xiong H., Pandey G., Steinbach M. & Kumar V. (2006). Enhancing data analysis with noise removal. *IEEE Transactions on Knowledge and Data Engineering*, 18(3), 304–319.

Ylijoki O. & Porras J. (2016). Perspectives to Definition of Big Data: A Mapping Study and Discussion. *Journal of Innovation Management*, 4(1), 69–91. https://doi.org/10.24840/2183-0606_004.001_0006

Yoo Y. (2015). It is not about size: a further thought on big data. *Journal of Information Technology*, 30(1), 63–65.