

Iita Suomala

# DATA-ANALYTIKKA RISKIENHALLINASSA



JYVÄSKYLÄN YLIOPISTO  
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA  
2021

## TIIVISTELMÄ

Suomala, Iita

Data-analytiikka riskienhallinnassa

Jyväskylä: Jyväskylän yliopisto, 2021, 30 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja(t): Seppänen, Ville

Tämä tutkielma on kirjallisuuskatsaus, jossa pyritään selvittämään data-analytiikan hyödyntämistä riskienhallinnassa. Riskienhallinta on prosessi, jossa riskit tunnistetaan, arvioidaan ja niiden vaikutuksia pyritään lieventämään. Tässä tutkielmassa tarkastellaan miten data-analytiikan keinoja, kuten massadataa ja tiedonlouhintaa, hyödynnetään riskienhallinnan eri vaiheissa ja erilaisissa riskeissä. Tulosten perusteella data-analytiikkaa hyödynnetään monipuolisesti ja sillä todettiin olevan useita hyötyjä riskienhallinnassa. Erityisesti operatiivisessa riskienhallinnassa data-analytiikan käyttökohteita löytyi useita.

Asiasanat: data-analytiikka, riskienhallinta, massadata, tiedonlouhinta

## **ABSTRACT**

Suomala, Iita

Data analytics in risk management

Jyväskylä: University of Jyväskylä, 2021, 30 pp.

Information Systems Science, Bachelor's Thesis

Supervisor(s): Seppänen, Ville

This paper is a literary review that aims to examine the utilizing of data analytics in risk management. Risk management is a process in which risks are identified, estimated and their effects are minimized. In this paper methods of data analytics, such as big data and data mining, are examined on how they are used in different phases of risk management and on different risks. The results of this paper found that data analytics are used diversely and usage of data analytics in risk management were found to have many benefits. Especially operative risk management has many uses for data analytics.

Keywords: data analytics, risk management, big data, data mining

## TAULUKOT

Taulukko 1 Riskien lähteet ja esimerkkejä riskeistä .....	11
Taulukko 2 Riskeille altistuneiden resurssien luokittelu ja esimerkit .....	11
Taulukko 3 Operatiivisen riskienhallinnan luokittelukategoriat, tarkoitusperät, käytetyt metodit ja työkalut .....	20

# SISÄLLYS

TIIVISTELMÄ  
ABSTRACT  
TAULUKOT

1	JOHDANTO.....	6
2	RISKIENHALLINTA.....	8
	2.1 Riskien tunnistus.....	10
	2.2 Riskien arviointi.....	12
	2.3 Riskien lievennys .....	12
3	DATA-ANALYTIikka.....	14
	3.1 Massadata .....	15
	3.2 Tiedonlouhinta ja tiedon hankinta.....	16
4	DATA-ANALYTIikka RISKIENHALLINNASSA.....	18
	4.1 Data-analytiikka ja riskienhallinnan eri teemat .....	18
	4.2 Operatiivinen riskienhallinta .....	19
5	YHTEENVETO .....	23
	LÄHTEET.....	26

# 1 JOHDANTO

Vuonna 2014 oli arvioitu, että jopa 90 % maailman datasta oli syntynyt viimeisen kahden vuoden aikana (Morton, 2014, s. 1). Runklerin (2012, s. 5) mukaan informaatio maailmassa tuplaantuu 20 kuukauden välein. Informaatioteknologiassa tapahtuneet edistysaskeleet ovat mahdollistaneet ennennäkemättömän kasvun niin datan määrässä kuin sen monimuotoisuudessaakin. Uusien innovaatioiden ansiosta on helpompaa luoda, kerätä ja varastoida suuria datajoukkoja (Virender & Cherry, 2018). Datan määrää viime vuosina ovat lisänneet muun muassa verkkomyynti, älypuhelimet ja sosiaalinen media. Data-analytiikka hyötyy tästä datan määrän kasvusta, sillä tilastollinen luotettavuus paranee otoskoon noustessa. (Duan, 2014) Toisaalta datan määrän lisääntyessä todennäköisesti lisääntyy myös hyödyntämiskelvottoman datan määrä (Virender & Cherry, 2018).

Kuuselan ja Ollikaisen (2005, s. 15–16) mukaan riskien hallitsemiseen on monenlaisia keinoja, kuten esimerkiksi riskialttiin toiminnan välttäminen, riskin kanssa eläminen, tuuriin luottaminen, vahingon rajoittaminen riskin toteutuessa sekä riskin siirtäminen toiselle taholle vakuutusten avulla. Tänä päivänä organisaatiot kohtaavat toimintaympäristössään nopeita muutoksia, ja Tchankovan (2002) mukaan riskienhallinnasta on tullut organisaatioiden yksi päätehtävistä. Nopeasti muuttuvassa ympäristössä ajankohtaisen ja hyödyllisen tiedon hankkiminen ei ole helppo tehtävä. Tämän hetken hyödyllinen ja oikeellinen tieto voi olla vanhentunutta jo seuraavana päivänä. Siksi onkin tärkeää nähdä datan hyödyntäminen jatkuvana prosessina, jonka tuloksia arvioidaan ja näiden tulosten perusteella tehdään tarvittavia muutoksia toiminnassa. Muutoin organisaatiossa tapahtuva päätöksenteko voi perustua virheelliselle, mikä voi aiheuttaa epätoivottuja seurauksia. (Bumblauskas, Nold, Bumblauskas & Igou, 2017)

Data-analytiikasta löytyy huomattavia määriä tieteellisiä julkaisuja ja kirjoja, kuin myös riskienhallinnasta. Data-analytiikasta riskienhallinnassa löytyy myös lukuisia julkaisuja, joista suurin osa keskittyy tiettyyn riskienhallinnan osa-alueeseen, kuten toimitusketjujen riskienhallintaan tai rakennusprojektien riskienhallintaan. Monet artikkelit ovat keskittyneet johonkin tarkempaan teemaan tai tapaustutkimuksiin. Julkaisuja, joissa luodaan yleiskatsaus data-analytiikan hyödyntämisestä riskienhallinnan eri osa-alueilla, on edellä mainittuja julkaisuja

vähemmän. Tämän tutkielman aihe on hyvinkin laaja, ja luultavasti juuri aihepiirin laajuuden vuoksi suurin osa julkaisuista on rajannut tarkastelukulmaa tarkemmaksi. Tästä huolimatta julkaisuille, jotka luovat ajankohtaisen ja laajan yleiskatsauksen data-analytiikasta riskienhallinnassa, lienee tarvetta.

Tämä tutkielma toteutettiin kirjallisuuskatsauksena, jonka primäärilähteet haettiin lähinnä seuraavista sähköisistä tietokannoista: Google Scholar, IEEE Xplore Digital Library, ProQuest, ScienceDirect sekä Wiley Online Library. Osa aineistosta oli hieman vanhempia, 20–30 vuotta vanhoja. Kuitenkin suuri osa kerätyistä lähteistä oli julkaistu viimeisen kymmenen vuoden sisällä, ja hyvin moni lähteistä oli vain muutaman vuoden vanha. Aineiston etsinnässä hakusanoina hyödynnettiin useimmiten "data analytics", "data mining", "big data analytics", "data analytics in risk management", "risk management" sekä "operational risk management".

Tutkielma tarkastelee data-analytiikkaa riskienhallinnassa melko yleisellä tasolla, ja se pyrkiiin vastaamaan asetettuun tutkimuskysymykseen "miten data-analytiikkaa hyödynnetään riskienhallinnassa?". Tarkoituksena on tarkastella data-analytiikan hyödyntämistä monipuolisesti riskienhallinnan eri osa-alueilla, vaikkakin operatiivisen riskienhallinnan näkökulmaa tutkitaan hieman tarkemmin.

Johdannon jälkeen tutkielmassa käsitellään riskienhallintaa. Kyseisessä luvussa on tarkoitus selventää riskienhallintaa käsitteenä, ilmiönä ja prosessina sekä esitellä aiheeseen liittyviä oleellisia termejä. Kolmannessa luvussa käydään läpi data-analytiikkaa. Tässä luvussa määritellään data-analytiikkaa sekä tuodaan ilmi sen taustoja, hyödyntämistarkoituksia sekä aloja, joissa data-analytiikkaa hyödynnetään. Myös aiheeseen liittyviä tärkeitä termejä, kuten massadata ja tiedonlounhint, esitellään ja määritellään. Tutkielman neljäs luku, data-analytiikka riskienhallinnassa, pyrkii vastaamaan esitettyyn tutkimuskysymykseen "miten data-analytiikkaa hyödynnetään riskienhallinnassa?". Luvussa esitellään todettuja hyötyjä ja konkreettisempia tapausesimerkkejä data-analytiikan käytöstä riskienhallinnassa. Tarkemmin tutustutaan operatiivisen riskienhallinnan näkökulmaan, ja määritellään mitä operatiivisilla riskeillä tarkoitetaan. Loput luvusta keskittyy tarkastelemaan data-analytiikan hyötyjä ja käyttöä operatiivisen riskienhallinnan eri soveltamisaloilla. Lopuksi viimeisessä luvussa tehdään tutkielmasta yhteenveto sekä esitetään mahdollisia jatkotutkimusmahdollisuuksia.

## 2 RISKIENHALLINTA

Tässä tutkielman toisessa luvussa määritellään termit riski ja riskienhallinta sekä kuvataan tarkemmin riskienhallinnan prosessia. Luvussa tarkastellaan vielä erikseen operatiivista riskienhallintaa sekä käydään läpi riskienhallinnan prosessivaiheiden tarkoituksia ja päämääriä.

Termiä ”riski” käytetään, kun halutaan kuvata vaaraa, joka liittyy onnettomuuden mahdollisuuteen (Kuusela & Ollikainen, 2005, s. 16). Riski ja epävarmuus käsitteinä voidaan erottaa sen suhteen, ovatko todennäköisyydet tunnettuja. Riskiksi sanotaan tilannetta, jossa päätöksentekoon liittyvät todennäköisyydet ja tekijät ovat tunnettuja ennalta. Toisaalta jos todennäköisyydet ei ole ennalta tiedossa, kyseessä on enemmänkin epävarmuus. (Kreps, 1990, s. 117) Mikäli riski toteutuu, voi seurauksena olla mikä tahansa arvon menetys. Rahallisen arvon lisäksi menetettävänä voi olla esimerkiksi terveydellinen ja yhteiskunnallinen arvo. Yritysten näkökulmasta uhkia aiheuttavat monet tekijät. Tällaisia tekijöitä ovat esimerkiksi markkinoiden muutokset sekä yrityksen sisäiset virhearviot. Jotta yritys menestyisi kilpailussa, sen on välttämätöntä ottaa riskejä. (Kuusela & Ollikainen, 2005, s. 16–17)

Riskienhallinnalla tarkoitetaan prosessia, jonka avulla pyritään tunnistamaan ja arvioimaan riskejä. Sen avulla halutaan myös valita toimenpiteitä, joiden avulla voidaan vähentää riskien aiheuttamia seurauksia. (Kuusela & Ollikainen, 2005, s. 35) Riskienhallinnalla ei ole virallista vaiheluokittelua, ja se voidaan jakaa eri vaiheisiin eri tavoin. Usein riskienhallinnan näkee jaettavan kolmeen kriittiseen vaiheeseen: tunnistukseen, arviointiin ja riskien lievennykseen (Fan & Stevenson, 2018). Toisinaan neljänneksi vaiheeksi esitellään riskien valvontaa (Houj & Zhao, 2021). Esimerkiksi Gaudenzi ja Borghesi (2006) jaottelivat riskienhallinnan neljään seuraavaan vaiheeseen: riskien arviointi, riskien raportointi ja päätös, riskien hoito sekä riskien valvonta. Vaikka riskienhallinnan vaiheiden jaottelusta ei olla yksimielisiä, kukin vaihe on silti merkityksellinen osa riskienhallintaa. Edellinen vaihe täytyy hoitaa onnistuneesti, jotta seuraava vaihe on mahdollista suorittaa menestyksekkäästi. Jos organisaatiota uhkaavia riskejä ei osata tunnistaa, niitä ei ole myöskään mahdollista arvioida. Kun riskejä ei kyetä onnistuneesti arvioimaan, niiden aiheuttamia epätoivottuja seurauksia on mahdotonta



lievittää. (Kern ym., 2012; Craighead ym., 2007) Puri ja Sharma (2020) tuovatkin esille sen, että riskienhallinta erottelemalla ja kontrolloimalla riskejä on paras lähestymistapa käsitellä riskejä.

Tchankova (2002) kuvailee riskienhallintaa jatkuvaksi prosessiksi, joka on riippuvainen organisaation sisäisen ja ulkoisen ympäristön muutoksista. Hänen mukaansa riskienhallinnan päätarkoitus on helpottaa muita hallintatehtäviä saavuttamaan organisaation tavoitteita, ja riskienhallinnasta onkin tullut organisaatioiden yksi päätehtävistä. (Tchankova, 2002) Riskienhallinnan mieltäminen nimenomaan jatkuvana prosessina on tärkeää. Organisaatioiden ympäristöissä tapahtuu jatkuvasti muutoksia, minkä johdosta organisaatioiden on syytä jatkuvasti skannata ja analysoida mahdollisia riskejä. Kun organisaation ympäristöä tarkkaillaan huolellisesti, mahdolliset relevantit riskit tunnistetaan ajoissa ja niiden lieventämistoimet voidaan aloittaa. (Craighead ym., 2007; Tchankova, 2002)

Riskienhallinta tuottaa organisaatiolle kustannuksia ja resurssien menetyksiä. Käytettyjä resursseja ovat esimerkiksi käytetty aika riskienhallintaan liittyviin tapaamisiin ja henkilöstö, jotka kehittävät riskienhallintasuunnitelmia. (Hall, 1999) Resurssien menetyksistä huolimatta riskienhallinta on olennainen osa organisaation toimintaa, ja sen merkitystä ei pidä väheksyä (Tchankova, 2002). Riskienhallinnalla on huomattu olevan seuraavanlaisia positiivisia vaikutuksia organisaatiossa:

- Tietoisuuden lisääntyminen
- Riskin mahdollisuuden ja vaikutuksen vähentäminen
- Yllätysten vähentyminen
- Tarkemmat arviot
- Paremmat mahdollisuudet onnistumiseen
- Lisääntynyt itsevarmuus projektien tavoitteiden saavuttamiseen. (Ellegaard, 2008)

Operatiivisella riskillä tarkoitetaan riskejä, jotka aiheutuvat ihmisten toimesta, sisäisistä prosesseista, järjestelmistä ja ulkoisista tapahtumista (Weeserik & Spruit, 2018). Näistä aiheutuu haittoja esimerkiksi tietokoneiden järjestelmissä sekä sisäisessä valvonnassa ja kontrollissa (Culp, 2002, s. 22). Beroggi ja Wallace (1994) määrittelevät operatiivisen riskienhallinnan odottamattomien tapahtumien tunnistamiseksi, niiden seurausten arvioimiseksi sekä päätöksenteoksi muuttaa suunniteltua toimintalinjaa. Operatiiviset riskit voidaan jakaa useisiin eri kategorioihin:

- Sisäiset riskit eli arvon menetys aiheutuu jonkun sisäisen sidosryhmän toimesta, esimerkiksi lakien tai määräysten kiertäminen johdon toimesta.
- Ulkoiset riskit eli arvon menetys johtuu ulkoisesta tapahtumasta tai toimesta, esimerkiksi tietomurto.
- Työturvallisuus ja työpaikan käytännöt, eli arvon menetys johtuu esimerkiksi turvallisuuteen tai työoikeuteen liittyvien toimien epäjohtonmukaisuuteen

- Asiakkaat, tuotteet ja liiketoimintaan liittyvät käytännöt, eli arvon menetys voi johtua tuotteen huonosta laadusta, mikä vähentää asiakkaita.
- Järjestelmien vikatilat voivat aiheuttaa arvon menetyksiä.
- Kuljetus ja prosessijohtaminen, eli arvon menetys voi johtua huonosti toteutetusta prosessijohtamisesta tai kuljetukseen liittyvistä haasteista. (Weeserik & Spruit, 2018)

## 2.1 Riskien tunnistus

Riskien tunnistus on riskienhallinnassa ensimmäinen vaihe (Tchankova, 2002; Houj & Zhao, 2021; Kleindorfer & Saad, 2005), jossa pyritään määrittämään tietyn aihealueen kaikki mahdolliset riskit (Panjehfouladgaran & Lim, 2020; Kern ym., 2012). Onnistunut riskien tunnistus on tärkeää, sillä se vaikuttaa merkittävästi riskienhallinnan tehokkuuteen: jos riskiä ei tunnisteta, sitä ei silloin voi myöskään hallita. (Tchankova, 2002) Riskien tunnistamisella on iso vaikutus riskienhallinnan myöhempisiin vaiheisiin, arviointiin ja riskien lievennykseen. Onnistuneella riskien tunnistamisella mahdollistetaan se, että organisaation päättäjät voivat kehittää riskien lieventämisstrategioita. (Panjehfouladgaran & Lim, 2020) Tchankova (2002) tuo artikkelissaan ilmi, että riskien tunnistamisella organisaatio voi löytää toimintoja, joissa organisaation resurssit ovat alttiita riskeille. Hän kuvailee riskien tunnistamista neljällä osatekijällä:

- riskien lähteet
- uhkatekijät
- vaarat
- riskeille altistuneet resurssit.

Riskien lähteillä Tchankova (2002) tarkoittaa organisaation ympäristössä olevia tekijöitä, joiden seurauksena voi olla positiivisia tai negatiivisia lopputuloksia. Esimerkkinä hän mainitsee markkinatilanteen, joka vaikuttaa suuresti organisaation päätökseen alkaa tuottamaan uutta tuotetta. Näin ollen markkinatilanne, johon liittyy esimerkiksi kilpailijoiden määrä, on organisaatiolle riskin lähde. Riskien lähteet voidaan ylimalkaisesti luokitella fyysisiin, sosiaalisiin ja taloudellisiin lähteisiin. Tchankova huomauttaa, että riskien tunnistamisen syvempi tarkastelu voi tarvita laajempaa ja tarkempaa luokittelua. Hän luokittelee riskien lähteet seitsemään eri ympäristöluokkaan (taulukko 1).

Taulukko 1 Riskien lähteet ja esimerkkejä riskeistä (Tchankova, 2002)

<b>Riskien lähteet</b>	
Fyysinen ympäristö	<ul style="list-style-type: none"> <li>• Luonnonkatastrofit, kuten myrskyt ja maanjäristykset</li> </ul>
Sosiaalinen ympäristö	<ul style="list-style-type: none"> <li>• Muutokset ihmisten käytöksessä, arvoissa ja sosiaalisissa rakenteissa</li> <li>• Levottomuudet, kuten mellakat ja lakot</li> <li>• Erot sosiaalisissa arvoissa ja kulttuurissa</li> </ul>
Poliittinen ympäristö	<ul style="list-style-type: none"> <li>• Poliittiset päätökset</li> </ul>
Operationaalinen ympäristö	<ul style="list-style-type: none"> <li>• Yhtiön operatiiviset toimet, kuten henkilöstön irtisanominen ja palkkaus, tuotantoprosessit ja työolosuhteet</li> </ul>
Taloudellinen ympäristö	<ul style="list-style-type: none"> <li>• Globalisaatio, taloudelliset taantumakaudet ja lamat</li> </ul>
Juridinen ympäristö	<ul style="list-style-type: none"> <li>• Muuttuvat lait, eri valtioilla on erilaiset oikeusjärjestelmät</li> </ul>
Kognitiivinen ympäristö	<ul style="list-style-type: none"> <li>• Inhimilliset tekijät</li> </ul>

Uhkatekijä-käsitteellä Tchankova (2002) viittaa olosuhteeseen tai tilaan, joka lisää menetyksen mahdollisuutta ja vakavuutta, esimerkiksi inhimillinen erehdys. Vaara on lähellä riskin käsitettä, mutta vaaralla on ainoastaan negatiivinen ja epätuottava lopputulos. Riskeille altistuneet resurssit kokevat mahdollisesti positiivisia tai negatiivisia lopputuloksia. Mikäli riski toteutuu, kohdistuu se näihin resursseihin. Riskeille altistuneiden resurssien analysointi onkin tärkeää, sillä riskienhallinta keskittyy lähinnä sellaisiin riskien lähteisiin, jotka voivat aiheuttaa vaaraa organisaatiolle ja sen resursseille. Tchankova jaottelee riskeille altistuneet resurssit kolmeen luokkaan: fyysisiin, taloudellisiin ja henkilöstöön liittyviin resursseihin (taulukko 2).

Taulukko 2 Riskeille altistuneiden resurssien luokittelu ja esimerkit (Tchankova, 2002)

<b>Riskeille altistuneet resurssit</b>	
Fyysisten resurssien altistumiset	<ul style="list-style-type: none"> <li>• Välineet, rakennukset ja koneet</li> </ul>
Henkilöstöresurssien altistumiset	<ul style="list-style-type: none"> <li>• Fyysiset ja psyykkiset terveydelliset haasteet, kuolemat</li> <li>• Uuden teknologian käyttöönotto</li> <li>• Työntekijöiden uskollisuus</li> </ul>
Taloudellisten resurssien altistumiset	<ul style="list-style-type: none"> <li>• Varat ja lainat</li> </ul>

## 2.2 Riskien arviointi

Riskien arviointi on riskienhallinnan toinen vaihe. Kyseessä on systemaattinen prosessi, jossa arvioidaan relevantteja riskitekijöitä, niiden tapahtumisen todennäköisyyttä sekä niiden merkittävyyttä. (Ramamoorti ym., 1999; Kern ym., 2012) Riskien arvioinnissa johtavat tahot tekevät päätöksiä, kuinka reagoivat tunnistettuihin riskeihin (Panjehfouladgaran & Lim, 2020; Fan & Stevenson, 2018). Joissain artikkeleissa riskien tunnistus sisällytetään riskien arviointi -vaiheeseen. Tästä huolimatta ne eroavat toisistaan. Siinä missä riskien tunnistamisessa pyritään löytämään kaikki mahdolliset relevantit riskit, riskien arvioinnin tavoitteena on arvioida riskiä yksilöllisesti ja ymmärtää sen relevanttiutta. Riskien arvioinnissa käydään siis kukin löydetty riski yksityiskohtaisemmin läpi, arvioidaan riskin mahdollisia laukaisijoita, aiheuttajia ja vaikutuksia. (Kern ym., 2012)

Riskien arviointi -vaiheen lopputulemana tulisi olla jonkinlainen luokittelu tunnistetuille riskeille (Kern ym., 2012). Riskien luokittelussa riskit lajitellaan yhtenäisiin ryhmiin riskien myöhempää tutkimusta sekä riskien lievennysstrategioita varten (Panjehfouladgaran, 2020). Yksi tapa luokitella riskejä on tehdä graafinen kuvitus, jossa käy ilmi mahdollisen riskin tapahtuma-aika, -paikka sekä tapahtumisen todennäköisyys (Ritchie & Brindley, 2007).

## 2.3 Riskien lievennys

Riskien lievennys -vaiheessa usein hyödynnetään riskienhallinnan aiemmista vaiheista syntynyttä dataa. Datan avulla pyritään osoittamaan mahdolliset riskit sekä niihin kohdistuvat vastatoimet. (Kern ym., 2012) Riskien lievennyksessä organisaatio kehittää riskien lievennysstrategioita, joiden tarkoituksena on vähentää sekä riskien esiintymistodennäköisyyttä, että riskien aiheuttamia epätoivotuja vaikutuksia. (Chang, Ellinger & Blackhurst, 2015; Tomlin, 2006)

Riskien lievennys voidaan jakaa kahteen teemaan: strategiat ennen riskin tapahtumista ja suunnitelmat riskin tapahtumisen jälkeen (Kern ym., 2015). Kleindorfer ja Saad (2005) huomauttavat, että ennaltaehkäisy olisi parempi strategia kuin riskien aiheuttamien vaikutusten parantaminen.

Chang, Ellingerr ja Blackhurst (2015) käsittelevät artikkelissaan toimitusketjun riskienhallintaa riskien lievennyksen näkökulmasta. Heidän mukaansa toimitusketjun riskien lievennysstrategiat voidaan jakaa kahteen teemaan: ylimäärään ja joustavuuteen. Jotkut käyttävät myös termejä ennakoiva ja reaktiivinen (Panjehfouladgaran & Lim, 2020). Ylimäärään nojautuvaa strategiaa on esimerkiksi lisätä tuotteiden saatavuutta pitämällä varastossa ylimääräisiä resursseja. Myös kapasiteetin lisääminen, usean toimittajasuhteen ylläpitäminen sekä varastokapasiteetin lisääminen ovat esimerkkejä ylimäärään pohjautuvista strategioista. Sen sijaan joustavuuteen pohjautuvat lievennysstrategiat nojaavat siihen, että organisaatio parantaa kapasiteettiaan havaita uhat ja kykenee

reagoimaan niihin nopeasti (Zsidisin & Wagner, 2010). Tätä strategiaa voidaan toteuttaa esimerkiksi yhteistyöllä muiden organisaatioiden kanssa, jakamalla tietoa ja kommunikoimalla (Zsidisin & Wagner, 2010; Chang ym., 2015).

### 3 DATA-ANALYTIikka

Tutkielman kolmannessa luvussa tarkastellaan data-analytiikan määritelmää sekä esitellään aiheeseen liittyviä oleellisia termejä. Tarkempaan tarkasteluun otetaan termit massadata sekä tiedonlouhinta.

Runklerin (2012, s.2) mukaan data-analytiikka tarkoittaa tietojärjestelmien hyödyntämistä suurten datamäärien analysointiin päätöksenteon tueksi. Data-analytiikka on luonteeltaan hyvin monimuotoista ja se on ottanut vaikutteita lukuisilta eri tieteenaloilta; muun muassa tilastotiede, signaaliteoria, hahmontunnistus, koneoppiminen ja laskennalliset tieteet ovat jättäneet vaikutuksensa data-analytiikkaan. (Runkler, 2012, s. 2; Tsai, Lai, Chao & Vasilakos, 2015) Data-analytiikan alkuaikoina tilastollisia metodeja käytettiin datan, esimerkiksi julkisten mielipidemittausten, analysointiin (Tsai ym., 2015).

Datan keräämisessä ja analysoinnissa päämääränä on hyödyntää sitä päätöstenteossa ja sitä kautta luoda organisaatiolle arvoa. Data-analytiikka usein mahdollistaa ennustavia analyyssejä, mikä usein juuri vaikuttaa organisaatioissa tehtyihin päätöksiin laatua parantavasti. (Bumblauskas ym., 2017) Organisaatioilla on kuitenkin usein haasteita hyödyntää keräämäänsä dataa, ja onkin arvioitu, että organisaatiot ovat analysoineet vain 0,5 % olemassa olevasta datasta (Sommer, 2015). Haasteista huolimatta organisaatioilla on halu hyödyntää dataa. Esimerkiksi Schoenherr ja Speier-Pero (2015) huomasivat, että toimitusketjuihin erikoistuneista alan ammattilaista puolet joko hyödynsivät massadatan analytiikkaa, tai suunnittelivat sen aloittaista lähitulevaisuudessa. Päätöksenteon parantaminen nähtiin massadatan analysoinnin yhtenä tärkeimpänä hyötynä.

Data-analytiikka on ottanut vaikutteita monilta tieteenaloilta, mutta toisaalta sen hyödyntäminen on myös levinnyt hyvin laajalle. Data-analytiikka on ottanut roolia esimerkiksi :

- terveydenhuollossa (Felot, Howard & Osgood, 2018)
- organisaatioiden toimitusketjuissa (Rajpopat, Jamar & Lekhrjani, 2017)
- yritysten myynnin, liikevaihdon ja tuotannon ennustamisessa (Bumblauskas ym., 2017)

- pankkitoiminnassa (Araz, Tsan-Ming, Olson, & Salman, 2020)
- urheilussa (Kaur & Jagdev, 2020)
- markkinoinnissa (Johnson, Muzellec, Sihi, & Zahay, 2019; Svilar, Chakraborty & Kanioura, 2013)
- rakennusalalla (Jiwat, Numan & Khawar, 2019; Aibinu, Koch & Ng, 2019).
- Energia-alalla
- Autoteollisuudessa
- Ilmailualalla
- Vähittäiskaupassa. (Goel, Datta & Mannan, 2017)

Tyypilliset data-analyysiprojektit jakautuvat useaan vaiheeseen. Data valitaan ja arvioidaan, minkä jälkeen data suodatetaan ja puhdistetaan. Suodatettu data visualisoidaan ja analysoidaan, ja näitä analyysien tuotoksia lopulta tulkitaan ja arvostellaan. (Runkler, 2012, s. 2) Datan käsittely ja arviointi ennen itse analyysia on ensiarvoisen tärkeää, sillä datan oikeellisuus on usein suurikin haaste analytiikassa (Jean-Baptiste, Qiu, Gai & Tao, 2015 ; Virender & Cherry, 2018). Suuret määrät dataa sisältävät myöskin suuria määriä turhaa ja ylimääräistä tietoa, eli datamelua (Ying-Kui & Qian-Mao, 2012). Melua poistamalla voidaan parantaa datan laatua, mikä puolestaan parantaa analyysia (Xiong, Pandey, Steinbach & Kumar, 2006).

### 3.1 Massadata

Dataa luodaan jatkuvasti ja aina vain kiihtyvällä tahdilla. Uudet teknologiset ratkaisut, kuten älypuhelimet, koneet, IoT-laitteet ja sensorit luovat jatkuvasti suuria määriä dataa, joka pitää säilyttää ja josta toisaalta pitäisi saada jotain hyötyä irti. Datan määrä itsessään tuottaa haasteita hyödyllisten kaavojen ja informaation löytämisessä. Kuitenkin suurempi haaste on valtaviin datamääriin analysointi, etenkin kun data on rakenteeltaan hyvin vaihtelevaa ja sitä luodaan suurella vauhdilla. (EMC Education Services, 2015, s. 2) Toisaalta data ilman objektiivista analyysia tuottaa hyvin rajallisesti hyötyä organisaatiolle (Bumblauskas ym., 2017).

Massadata on moniulotteinen termi, jolla ei ole yhtä vakiintunutta määritelmää. Termiin usein liitetään suurten datamäärien kerääminen ja analysoiminen. (Bumblauskas ym., 2017) Yksi suosituimmista määritelmistä on niin sanottu kolmen V :n määritelmä. Nämä kolme V :tä kuvaavat massadatalle tyypillistä valtavaa määrää (volume), nopeutta (velocity) ja vaihtelevuutta (variety). Määrällä tarkoitetaan luodun ja varastoidun datan määrää ja nopeus puolestaan ilmentää datan luomisnopeutta. Vaihtelevuudella viitataan siihen,

että dataa kerätään useista eri lähteistä, ja se on muodoltaan vaihtelevaa eli kerätty data voi olla rakenteellista, osarakenteellista ja rakenteetonta. (Emani, Cullot & Nicolle, 2018; Loyd & Kannan, 2017; EMC Education Services, 2017, s. 2).

Tutkimusnäyttöä on, että perinteinen kolmen V:n malli massadatan määrittelylle ei palvele tämän päivän massadatan kuvausta (Tsai ym., 2015). Uudemmissakin määritelmässä on vaihteluita, mutta erityisesti termit totuudenmukaisuus, arvo ja muuttuvuus toistuu useissa artikkeleissa. Totuudenmukaisuudella tarkoitetaan datan oikeellisuutta ja luotettavuutta, ja muuttuvuudella taas datan epäjohdonmukaisuutta ja datan muutoksia. Arvo merkitsee massadatatista saatua hyötyä. (Emani ym., 2018; Bumblauskas ym., 2017; Goel ym., 2017)

Massadatan luonteeseen kuuluu myös se, että sen tehokkaaseen analysointiin ei riitä enää perinteiset menetelmät ja tietokannat. Koostaan ja rakenteestaan johtuen massadata vaatii uudenlaisia teknologioita sen varastointiin, hallintaan ja hyödyn realisointiin. Näillä uusilla teknologioilla on mahdollista luoda ja hallita suuria datajoukkoja ja niiden varastointiympäristöjä. (EMC Education Services, 2017, s. 2)

Massadata-analytiikalla on huomattu olevan lukuisia positiivisia vaikutuksia organisaation toimintaan. Se mahdollistaa kaikilla sidosryhmillä laadukkaampia päätöksiä organisaatioiden eri osa-alueilla, kuten asiakassuhteiden hallinnassa, operaatioiden optimoinnissa, uhkien ja huijausten ennaltaehkäisyssä sekä uusien tulonlähteiden hyödyntämisessä. (Loyd & Kannan, 2017). Goel, Datta ja Mannan (2017) listasivat massadata-analytiikan hyödyiksi oikea-aikaiset ja tehokkaat päätökset liittyen kustannuksiin, aikaan, tuotekehitykseen ja optimointiin.

### 3.2 Tiedonlouhinta ja tiedon hankinta

Tiedonlouhinnalla tarkoitetaan prosessia, jossa datasta pyritään saamaan hyödyllistä tietoa irti. Feyaad (1996) erottelee tarkemmin termit tiedonlouhinta (data mining) ja tiedon hankinta (knowledge discovery in databases). Hänen mukaansa tiedonlouhinta on enemmän matemaattisten algoritmien hyödyntämistä, jotta datasta voidaan löytää hyödyllisiä malleja. Tiedon hankinta sisältää useita prosessivaiheita ennen ja jälkeen mallien etsimisoperaatioita, eli tiedonlouhinta on osavaihe koko tiedon hankintaprosessissa.

Choi, Chan ja Yue (2017) määrittelevät tiedonlouhinnan prosessiksi, jossa datasta etsitään hyödyllisiä suhteita, kaavoja ja tekijöitä. Tiedonlouhinta voi ajatella vertauskuvallisesti kullanhuuhtomisena, jossa maa-aineksestä löydetään eri metodeilla kultaa (Tsai ym., 2015). Varman (2014) mukaan tiedonlouhinnassa yhdistyy erilaisia data-analyysitekniikoita, kuten tilastollista analyysia ja mallintamista. Nisbet, Miner, Yale, Elder ja Peterson (2018, s. 22) tiivistävät tiedonlouhinnan prosessiksi, jossa käytetään koneoppimisalgoritmeja, jotka etsivät sekavista



ja suurista datajoukoista malleja ja suhteita. Nämä löydökset sitten vaikuttavat toimiin, joiden tarkoituksena on tuottaa jonkinlaista hyötyä, kuten diagnoosit ja tuotot.

Fayaad (1996) jakaa tiedon hankinnan kuuteen eri vaiheeseen: valintaan, esikäsitteilyyn, muunnokseen, tiedonlouhintaan, tulkintaan ja arviointiin. Näillä kaikilla vaiheilla voidaan suorittaa data-analytiikkaprosessi, jossa data ensin kerätään, sitten siitä löydetään tietoa ja lopulta tieto näytetään käyttäjälle (Tsai ym., 2015). Fayaadin (1996) mukaan tiedon hankintaa hyödynnetään monilla eri aloilla, kuten rahoitus- ja pankkialalla, vähittäiskaupoissa, tuotannossa, valvonnassa ja diagnostiikassa, markkinoinnissa ja terveydenhuollossa. Tiedonlouhinnalla ja tiedonhankinnalla on havaittu olevan paljon hyötyjä organisaatioiden näkökulmasta. Tiedonlouhinta auttaa organisaatioita muun muassa poikkeamien havainnoinnissa, ennustavien mallien teossa ja sosiaalisten verkostojen analysoinnissa. (Varma, 2014) Kaikki nämä hyödyt auttavat organisaatioita saavuttamaan kilpailuetua markkinoilla (Gan, Luo & Lin, 2009)

## 4 DATA-ANALYTIikka RISKIENHALLINNASSA

Kirjallisuuskatsauksen pääluku keskittyy vastaamaan asetettuun tutkimuskysymykseen: miten data-analytiikkaa hyödynnetään riskienhallinnassa? Luvussa esitellään data-analytiikan yleisiä hyötyjä riskienhallinnassa sekä tuodaan esiin konkreettisempia käyttötapauksia data-analytiikan keinoista riskienhallinnan apuna. Data-analytiikan hyötyjä tarkastellaan erityisesti operatiivisen riskienhallinnan näkökulmasta.

Luvussa hyödynnetään viitekehyksenä Arazin ja kollegoiden (2020) esittämää operatiivisen riskienhallinnan luokittelua kuuteen eri soveltamisalaan: toimitusketjujen riskienhallinta, katastrofihallinta, julkinen terveys, elintarviketurvallisuus, kuljetus ja hyvinvointi.

### 4.1 Data-analytiikka ja riskienhallinnan eri teemat

Alati muuttuvassa yritysmaailmassa dataa täytyy jatkuvasti uudelleenarvioida, jotta sillä voidaan tuottaa arvoa johtamiseen (Bumblauskas ym., 2017). Organisaatiot joutuvat kohtaamaan aina vain enemmän vaatimuksia, monimutkaisuutta ja riskejä. Tietojärjestelmien on havaittu tuovan monia etuja, kun niitä hyödynnetään riskienhallintaan. (Weeserik & Spruit, 2018) Riskienhallintajärjestelmät voivat hyödyntää ennustavaa analytiikkaa sekä tilastollista analyysiä, ja järjestelmissä saattaa olla tarvetta reaaliaikaiselle analyysille. Riskienhallintajärjestelmät hyötyvätkin suuresti teknologian kehityksestä, sillä uusille analytiikan mahdollisuuksille on tarvetta. Massadata-analytiikan käyttöönotto riskienhallinnan eri prosesseissa mahdollistaa reaaliaikaisen analytiikan sekä tarjoaa parhaan tavan hallita riskejä. (Loyd & Kannan, 2017)

Data-analytiikan eri keinoja hyödynnetään monipuolisesti riskienhallinnassa eri aloilla. Goel, Dattan ja Mannan (2017) totesivat tutkimuksessaan, että massadata-analytiikan käyttö prosessiturvallisuudessa ja riskienhallinnassa kehittyi. Heidän mukaansa massadata-analytiikan käyttö tarjoaa mahdollisuuden

tiedostavampaan strategisointiin sekä operatiivisiin riskeihin liittyvään päätöksentekoon, ja he mainitsivat muun muassa seuraavanlaisia hyötyjä massadatan käytöstä prosessiturvallisuuksessa:

- Analytiikka mahdollistaa dynaamisen riskiprofiilien arvioinnin.
- Data-analytiikka mahdollistaa optimaalisen huoltoaikataulun, mikä vähentää suunnittelemattomia sulkemisia ja pysähdyksiä. Tämä parantaa toimintojen turvallisuutta sekä luotettavuutta.
- Tietoa hyödyntämällä riskit voidaan sijoittaa järjestykseen, mikä mahdollistaa resurssien jakamisen riskien ennaltaehkäisyyn ja lieventämiseen.
- Uudet tai parannellut mittausmenetelmät parantavat valvontaa.
- Yksityiskohtaisten analyysien teko kehittää tarkastuksia, onnettomuustutkintoja sekä uhkien arviointia.

Hormozi ja Giles (2006) huomauttavat, että tiedonlouhintaa voidaan käyttää monenlaisten riskien hallinnassa. Esimerkkeinä he mainitsevat kilpailusta tulevat riskit, tuotteiden keho laatu ja asiakkaiden menetys. Asiakkaiden menetyksiä voidaan heidän mukaansa ennaltaehkäistä kehittämällä malli asiakkaille, jotka todennäköisesti siirtyvät kilpailijalle asiakkaaksi.

Data-analytiikan keinot ovat löytäneet tiensä myös muunlaisten taloudellisten riskien torjuntaan. Tiedonlouhintaa hyödynnetään muun muassa vakuutuslalla riskien analysoinnissa (Choi, Chan & Yue, 2017) sekä talouspetosten löytämisessä (Jean-Baptiste ym., 2015). Gan, Luo ja Lin (2019) kehittivät tiedonlouhintatekniikkaa hyödyntämällä mallin, joka tunnistaa sellaiset lainansaajat, jotka eivät luultavasti pysty maksamaan lainojansa. Mallia kehitettiin käyttämällä oikeaa dataa kiinalaisesta pankista, ja malli pystyikin melko hyvällä menestyksellä tunnistamaan pankille tappiota aiheuttavat asiakkaat. Koh, Tan ja Goh (2004) hyödynsivät tiedonlouhintaa kehittäessään mallia luottoluokitukselle. Heidän mukaansa luottoluokitusmallin etuna on objektiivinen analyysi asiakkaan luottokelpoisuudesta, mikä vähensi syrjintää asiakkaita kohtaan.

## 4.2 Operatiivinen riskienhallinta

Operatiiviset riskit nähdään monien viime vuosikymmenten suurten taloudellisten epäonnistumisten aiheuttajina (Weeserik & Spruit, 2018). Operatiivinen riskienhallinta onkin kriittinen osa-alue mille tahansa organisaatiolle. Erityisesti massadatan aikakautena erilaiset analytiikan työkalut ovat kehittyneet nopeasti. (Gan, Luo & Lin, 2019). Shi, Wang ja Alwan (2020) totesivat tutkimuksessaan datalähtöisen analytiikan löytävän lisää jalansijaa operatiivisessa riskienhallinnassa, ja alan tutkijat ovatkin kiinnostuneita ratkaisemaan operatiivisia haasteita hyödyntämällä dataa ja koneoppimistekniikoita. Araz ja kollegat (2020) toteavatkin,

että huomattava määrä kehittyviä analytiikkamenetelmiä on otettu käyttöön nykypäivän operatiivisen johtamisen ongelmienratkaisuisissa.

Koyuncugil ja Ozgulbas (2012) tutkivat, miten tiedonlouhintaa voidaan hyödyntää riskien löytämisessä. He kehittivät tiedonlouhinnan avulla finanssialalle ennakkovaroitusjärjestelmän, joka löysi 15 eri riskitekijää, jotka vaikuttavat taloudellisiin vaikeuksiin. Lisäksi ennakkovaroitusjärjestelmän avulla kehitettiin neljä eri toimintamallia riskien ennaltaehkäisyyn sekä taloudellisen suorittamisen parantamiseen. Hailemariam, Hill ja Demissie (2012) puolestaan selvittivät tiedonlouhinnan hyötyjä asiakasuskollisuuden sekä velkojen takaisinmaksun laiminlyöntien ennustamiseksi. Heidän mukaansa erityisesti J48 -niminen algoritmi menestyi hyvin tapausten luokittelussa. Deshmukh ja Talluru (1998) tutkivat tiedonlouhinnan ja massadatan hyödyntämistä määriteltäessä riskejä, jotka tulevat johtotason petoksista. He totesivat tiedonlouhinnan olevan kilpailukykyinen työkalu operatiivisessa riskienhallinnassa.

Shi ja kollegat (2020) pyrkivät tutkimuksessaan luomaan datalähtöisen lähestymistavan inventaarionhallintaan. He yhdistivät sekä ennustavaa että ohjailevaa analytiikkaa ja hyödynsivät muun muassa koneoppimista. Ennustavalla analytiikalla riskiä vähennettiin tuottamalla luotettavia ennustuksia, ja ohjailevalla analytiikalla puolestaan perustettiin malleja inventaarion optimointiin. Shi ja kollegat huomasivat tutkimuksessaan, että ennustavalla analytiikalla voitiin vähentää keskimäärin 20 % kustannuksista välttämällä epätuottavien tavaroiden lähettämistä ulkomailla sijaitseviin varastoihin.

Araz ja kollegat (2020) tarkastelevat artikkelissaan operatiivista riskienhallintaa eri soveltamisalojen kautta (Taulukko 3) ja he ottivat lähempään tarkasteluun seuraavat teemat: toimitusketjun riskienhallinta, katastrofihallinta, julkinen terveys, elintarviketurvallisuus, kuljetus sekä hyvinvointi. Analytiikkatekniikat he jakavat neljään eri tyyppiin: kuvailevaan, diagnostiseen, ennustavaan ja ohjailevaan analytiikkaan. Lisäksi tutkimusryhmä esittelee data-analytiikan työkalujen applikaatioille kehittämiä strategisia tarkoituksia: tietojärjestelmien hallinta sekä tiedonhallinta.

Taulukko 3 Operatiivisen riskienhallinnan luokittelukategoriat, tarkoitusperät, käytetyt menetöt ja työkalut (Araz ym., 2020, s. 5)

Luokittelukategoriat	Tarkoitusperä	Käytetyt menetöt ja työkalut
Soveltamisala	<ul style="list-style-type: none"> <li>• Toimitusketjun riskienhallinta</li> <li>• Katastrofihallinta</li> <li>• Julkinen terveys</li> <li>• Elintarviketurvallisuus</li> <li>• Kuljetus</li> <li>• Hyvinvointi</li> </ul>	<ul style="list-style-type: none"> <li>• Riskien arviointi ja hallinta</li> <li>• Riskien lievennys</li> <li>• Valmius ja reagointi</li> <li>• Järjestelmien luotettavuus ja turvallisuus</li> </ul>

Analytiikkatekniikat	<ul style="list-style-type: none"> <li>• Kuvaileva</li> <li>• Diagnostinen</li> <li>• Ennustava</li> <li>• Ohjaileva</li> </ul>	<ul style="list-style-type: none"> <li>• Tilastotiede</li> <li>• Visualisointi</li> <li>• Simulointi</li> <li>• Koneoppiminen</li> <li>• Tiedonlouhinta</li> <li>• Optimointi</li> </ul>
Analytiikkastrategiat	<ul style="list-style-type: none"> <li>• Tietojärjestelmien hallinta</li> <li>• Tiedonhallinta</li> </ul>	<ul style="list-style-type: none"> <li>• Toiminnanohjausjärjestelmät</li> <li>• Ketjulohkoteknologia</li> <li>• Massadatan työkalut</li> </ul>

Wieland, Handfield ja Durach (2016) totesivat artikkelissaan, että massadata ja sen analytiikka sekä riskienhallinta ovat kasvavia trendejä toimitusketjujen hallinnassa. Arazin ja kollegoiden (2020) mukaan data-analytiikasta toimitusketjujen riskienhallinnassa löytyy lukuisia tutkimusartikkeleita. Esimerkiksi Jamshidi ja kollegat (2017) tutkivat massadata-analytiikan hyödyntämistä rautatiealan riskien määrittelyssä. Ban ja Rudin (2018) puolestaan selvittivät tiedonlouhinnan hyötyjä inventaariopäätöksissä, tarkoituksena minimoida liian täysiä tai tyhjiä varastoja. Kara, Firat ja Ghadge (2020) kehittivät tiedonlouhintaa hyödyntämällä viitekehysten toimitusketjujen riskienhallintaan. He muun muassa tunnistivat riskitekijöitä, keräsivät ja säilyttivät riskiin liittyvää dataa sekä analysoivat kyseistä dataa tiedonlouhinnan algoritmeilla ja kehittivät riskien lievennysstrategioita. Massadataa ja tiedonlouhintaa on myös käytetty toimittajien luotettavuuden arvioimiseen (Zage, Glass, Colbaugh, 2013).

Araz ja kollegat (2020) mainitsevat katastrofihallinnan yhtenä operatiivisen riskienhallinnan soveltamisalana. Data-analytiikkaa hyödynnetään muun muassa riskien arviointiin, niiden lieventämiseen sekä valmiuden ja reagoinnin parantamiseen. Zagorecki, Johnson ja Ristvej (2014) tutkivat tiedonlouhinnan sekä koneoppimisen hyödyntämistä katastrofihallinnan päätöksenteon tukena. Peng ja Chang (2020) hyödynsivät massadata-analytiikkaa kehittäessään energiatehokkaan reititystekniikan. Reititystekniikalla on tarkoitus ratkaista ongelmia liittyen pakettien lähetykseen kriisialueilla. Massadata-analytiikkaa hyödynnetään myös tulvien esiintyvyyden ennustamisessa (Yusoff, Din, Yusof & Khan, 2015). Shah, Seker, Hameed, ja Draheim (2019) pohtivat artikkelissaan massadata-analytiikan kasvavaa roolia katastrofihallinnassa. He mainitsevat muun muassa seuraavia hyötyjä massadata-analytiikan sekä IoT-laitteiden käytöstä katastrofihallintajärjestelmissä:

- Reaaliaikainen analytiikka, joka koostuu nopeasta datan analytiikasta, reaaliaikaisesta kyselyihin vastaamisesta ja nopeasta tulosten luomisesta
- Useat datalähteet, mikä mahdollistaa uusia näkökulmia
- Datavarastot, jotka toimivat reaaliaikaisesti ja pitävät sisällään rakenteellisesti vaihtelevaa dataa.

Avvenuti, Cresci, Vigna, Fagni sekä Tesconi (2018) kehittivät CrisMap -nimisen sosiaalisesta mediasta kerättyyn massadataan perustuvan karttajärjestelmän. CrisMap kerää Twitteristä potentiaalisia kriisiin liittyviä twiittejä, joissa on sijaintitunniste. Näitä twiittejä käytetään luomaan karttoja. Karttojen avulla voidaan jakaa resursseja paremmin alueille ja ihmisjoukoille, joihin kriisi on eniten vaikuttanut. CrisMap auttaa myös hahmottamaan kriisin aiheuttaman vahingon kokonaistilannetta paremmin.

Data-analytiikkaa on mahdollista hyödyntää myös ruuan turvallisuuteen liittyvässä riskienhallinnassa. Biffis ja Chavez (2017) käyttävät tiedonlouhinta-menetelmiä satelliittitietoihin, ja näin luovat sääindeksejä. Kuumuuteen ja sademäärään perustuvat indeksit ovat hyödyllisiä riskiprofiilien luomisessa. Biffis ja Chavez (2017) argumentoivat, että heidän luoman mallin perusteella vakuutus-kuluja koskien sadon menettämistä voidaan vähentää jopa 30 prosenttia. Data-analytiikka tuo hyötyjä myös elintarvikkeiden kylmäketjujen riskienhallintaan. Elintarvikkeiden prosessointivaiheessa analytiikka parantaa laaturiskien tunnistamista. Jakeluvaiheessa data-analytiikkaa voidaan hyödyntää jakelureittien optimoinnissa, mikä vähentää tuotteiden laadun heikkenemistä sekä kuljetuskustannuksia. (Chaudhuri, Dukovska-Popovska, Subramanian, Chan & Bai, 2018) Sosiaalisen median data-analytiikkaa on hyödynnetty ruokateollisuuden toimitusketjujen riskien hallinnassa. Singh, Shukla ja Mishra (2018) analysoivat Twitteristä pihviteollisuuden asiakkaiden kommentteja kyseisen sosiaalisen median alustalla. Twiittauksia analysoimalla kävi ilmi, että ruuan turvallisuus oli yksi suurimmista huolen lähteistä. Tulosten perusteella tutkijat kehittivät suosituksia kuluttajakeskeiseen toimitusketjuun.

## 5 YHTEENVETO

Tämä tutkielma keskittyy tutkimaan data-analytiikan hyödyntämistä riskienhallinnassa. Aluksi lukijaa johdatellaan aiheeseen ja esitetään tutkimuskysymys. Johdantoluvussa tuodaan esiin tutkielman taustoja sekä esitellään käytettyjä tutkimusmenetelmiä. Tässä luvussa myös selvennetään lukijalle tutkielman rakennetta ja näkökulmaa.

Tutkielman toisessa luvussa tarkastellaan riskienhallintaa. Luvussa kuvailtiin riskienhallintaa prosessina sekä esiteltiin määritelmiä. Riskienhallinta on prosessi, jonka päätarkoituksena on helpottaa muita hallintatehtäviä saavuttamaan organisaation asettamat tavoitteet. (Tchankova 2002) Prosessi jaetaan usein kolmeen vaiheeseen: tunnistukseen, arviointiin ja riskien lievennykseen. (Fan & Stevenson, 2018) Riskien tunnistus on riskienhallinnassa ensimmäinen vaihe, jossa pyritään tunnistamaan ja määrittelemään tietyn aihealueen kaikki riskit (Panjehfouladgaran & Lim, 2020). Riskien arviointi on tunnistuksen jälkeinen vaihe, jonka tavoitteena on yksityiskohtaisesti arvioida relevantteja riskitekijöitä, niiden tapahtumisen todennäköisyyttä sekä niiden merkittävyyttä. Lievennysvaiheessa hyödynnetään usein aiemmista vaiheista syntyneitä dataa ja kehitetään riskien lievennysstrategioita. (Kern ym., 2012). Lievennysstrategioiden päämääränä on vähentää riskien esiintymistodennäköisyyttä sekä riskien aiheuttamia menetyksiä (Tomlin, 2006). Riskienhallinnalla on huomattu olevan lukuisia positiivisia vaikutuksia organisaation toimintaan. Esimerkkinä Ellegaard (2008) mainitsee muun muassa yllätysten vähentymisen, arvioiden tarkentumisen sekä riskin mahdollisuuden ja vaikutuksen vähentämisen.

Tutkielman kolmannen luvun aiheena on data-analytiikka, sen määritelmät, päämäärä ja aiheeseen liittyvät termit. Näiden lisäksi esitellään data-analytiikan hyödyntämiskohteita. Runkler (2012, s. 2) määrittelee data-analytiikan tietojärjestelmien hyödyntämisenä suurten datamäärien analysoinnissa päätöksenteon tukena. Data-analytiikka on ottanut vaikutteita monilta eri tieteenaloilta ja sitä toisaalta hyödynnetään monilla eri aloilla aina urheilusta (Kaur & Jagdev, 2020) rakennusalaan (Jiwat, Numan & Khawar, 2019).

Termit massadata sekä tiedonlouhinta ovat merkittäviä data-analytiikan kontekstissa. Massadatalla tarkoitetaan valtavia, muodoltaan vaihtelevia ja

nopeasti syntyviä datajoukkoja, joiden tehokkaaseen analysointiin ei riitä perinteiset menetelmät ja tietokannat (EMC Education Services, 2017, s. 2). Perinteisen massadatan kolmen V:n määrittelylle on tullut kilpailevia määritelmiä, sillä kolmen V:n malli ei pysty täysin kuvaamaan tämän päivän massadataa (Tsai ym., 2015). Määritelmään usein lisätään termit totuudenmukaisuus, arvo sekä muuttuvuus. Tiedonloughintaa puolestaan kuvaillaan prosessina, jonka tarkoituksena on saada datasta hyödyllistä tietoa, kuten suhteita, kaavoja ja tekijöitä, irti (Choi, Chan ja Yue, 2017). Feyaad (1996) argumentoi, että tiedonloughinta on matemaattisia algoritmeja hyödyntävä vaihe tiedon hankintaprosessissa.

Tutkielman neljännen luvun aiheena on data-analytiikka riskienhallinnassa. Tässä luvussa pyritään vastaamaan tutkielmassa esitettyyn tutkimuskysymykseen: miten data-analytiikkaa hyödynnetään riskienhallinnassa? Data-analytiikan hyötyjä riskienhallinnan näkökulmasta esitellään sekä yleisellä tasolla että konkreettisempien tapausesimerkkien avulla. Luvussa esitellään tarkemmin data-analytiikkaa erityisesti operatiivisen riskienhallinnan näkökulmasta.

Data-analytiikan hyödyntämistä operatiivisessa riskienhallinnassa tarkastellaan Arazin ja kollegoiden (2020) tarjoaman viitekehyksen näkökulmasta. He ehdottivat operatiiviseen riskienhallintaan kuusi eri soveltamisalaa: toimitusketjun riskienhallinta, katastrofihallinta, julkinen terveys, elintarviketurvallisuus, kuljetus ja hyvinvointi. Tutkielmassa tutustuttiin tarkemmin toimitusketjun riskienhallintaan, katastrofihallintaan ja elintarviketurvallisuuteen. Arazin ja kollegoiden mainitsemista analytiikkatekniikoiden ja -strategioiden metodeista ja työkaluista (taulukko 3) korostuivat tutkielmassa koneoppiminen, tiedonloughinta sekä massadatan työkalut.

Tietojärjestelmien ja data-analytiikan keinojen hyödyntämisellä riskienhallinnassa on todettu olevan positiivisia vaikutuksia. Data-analytiikka mahdollistaa reaaliaikaisen analytiikan, mikä tuo tukea päätöksentekoon ja auttaa hallitsemaan riskejä (Loyd & Kannan, 2017). Data-analytiikan keinoja hyödynnetäänkin riskienhallinnan eri alueilla, kuten esimerkiksi prosessiturvallisuudessa ja toimitusketjujen riskienhallinnassa. Tutkielmassa kävi ilmi, että hyvin monenlaiset riskit, kuten petokset, tuotteiden laatu ja luottokelpoisuus, olivat data-analytiikan hyödyntämisen kohteena. Tutkielman perusteella on havaittavissa, että data-analytiikkaa hyödynnetään kaikissa aiemmin määritellyissä riskienhallinnan vaiheissa: tunnistuksessa, arvioinnissa ja riskien lievennyksessä. Esimerkiksi massadata-analytiikan avulla tunnistettiin ja määriteltiin rautatiealan riskejä (Jamshidi ym., 2017). Tiedonloughintaa on käytetty vakuutuslalla riskien arvioinnissa (Choi, Chan & Yue, 2017) sekä pankkialalla riskien lieventämisessä (Gan, Luo ja Lin, 2019). Pankkialalla tiedonloughinnan hyötyä lainanmyöntämisen päätöksentekoon on tutkittu melko paljon (Gan, Luo & Lin, 2009).

Tässä tutkimuksessa pyrittiin luomaan yleiskuva data-analytiikan hyödyntämisestä riskienhallinnassa. Aiheesta löytyy paljon tieteellisiä artikkeleita, ja erityisen suosittuja vaikuttavat olevan data-analytiikan käyttö operatiivisessa riskienhallinnassa, terveydenhuollossa, energia- ja rahoituslalla sekä toimitusketjujen ja inventaarioiden riskienhallinnassa (Choi, Chan & Yue, 2017; Araz ym., 2020). Tulevaisuudessa voisi olla mielekästä tutkia data-analytiikan



hyödyntämistä voittoa tavoittelemattomien organisaatioiden riskienhallinnassa. Kiinnostava tutkimusaihe voisi olla esimerkiksi data-analytiikan hyötyjen selvittäminen verkossa tapahtuvien äänestysten riskienhallinnassa. Sähköinen äänestys voi olla tulevaisuudessa ajankohtainen teema ja siihen liittyy paljon huolia muun muassa tietoturvariskeistä. Etenkin sähköinen vaaliäänestys verkossa olisi sen verran laaja kokonaisuus, että data-analytiikan käyttö riskienhallinnassa vaikuttaa perustellulta

## LÄHTEET

- Kuusela, H. & Ollikainen, R. (2005). *Riskit ja riskienhallinta*. Tampere : Tampere University Press 2005.
- Fan, Y. & Stevenson, M. (2018). *A review of supply chain risk management: definition, theory, and research agenda*". International Journal of Physical Distribution & Logistics Management, Vol. 48 No. 3, s. 205-230
- Kern, D. Moser, R. Hartmann, E. Moder, M. (2012). *Supply risk management: model development and empirical analysis*. International Journal of Physical Distribution & Logistics Management, Vol. 42 No. 1, s. 60-82.
- Tchankova, L. (2002). *Risk identification – basic stage in risk management*. Environmental Management and Health, Vol. 13 No. 3, s. 290-297.
- Panjehfouladgaran, H. Lim, S.F.W.T. (2020). *Reverse logistics risk management: identification, clustering and risk mitigation strategies*. Management Decision, Vol. 58 No. 7, s. 1449-1474.
- Hou, J. and Zhao, X. (2021), "Toward a supply chain risk identification and filtering framework using systems theory", Asia Pacific Journal of Marketing and Logistics, Vol. 33 No. 6, s. 1482-1497.
- Ramamoorti, S. Bailey, A. Traver, R. (1999). *Risk assessment in internal auditing: a neural network approach*. Intelligent Systemns in Accounting, Finance and Management, Vol. 8, No. 3, s. 159-180.
- Gaudenzi, B. Borghesi, A. (2006). *Managing risks in the supply chain using the AHP method*. The International Journal of Logistics Management, Vol. 17 No. 1, s. 114-136.
- Craighead, C. Blackhurst, J. Rungtusanatham, M. Handfield, R.B. (2007). *The severity of supply chain disruptions: design characteristics and mitigation capabilities*. Decision Sciences, Vol. 38, No. 1, s. 131-156.
- Kleindorfer, P. Saad, G. (2005). *Managing disruption risks in supply chains*. Production and Operations Management, Vol. 14, No. 1, s. 53-68.
- Ritchien B. Brindley, C. (2007). *Supply chain risk management and performance: a guiding framework for future development*. International Journal of Operations & Productions Management, Vol. 27, No. 3, s. 303- 322
- Puri, A. Sharma, S. (2020). *Risk Management in Software Engineering Using Big Data*. International Conference on Intelligent Engineering and Management
- Chang, W. Ellinger, A. Blackhurst, J. (2015). *A contextual approach to supply chain risk mitigation*. The International Journal of Logistics Management, Vol. 26 No. 3, s. 642-656.

- Zsidisin, G. Wagner, S. (2010). *Do perceptions become reality? The moderating role of supply chain resiliency on disruption occurrence*. *Journal of Business Logistics*, Vol 31, No. 2, s. 1-20
- Tomlin, B. (2006). *On the value of mitigation and contingency strategies for managing supply chain disruption risks*. *Management Science*, Vol. 52, No. 5, s. 639-657
- Hall, E. (1999). *Risk management return on investment*. John Wiley & Sons.
- Ellegaard, C. 2008. *Supply risk management in a small company perspective*. *Supply Chain Management: An International Journal* 13/6
- Runkler, T. (2012). *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Vieweg+Teubner Verlag.
- Tsai, CW. Lai, CF. Chao, HC. Vasilakos, A. (2015). *Big data analytics: a survey*. *Journal of Big Data*
- Felot, K. Howard, C. Osgood, E. (2018). *The evolving role of analytics to determine hospital risk management staffing*. *Journal on Healthcare Risk Management*, Vol. 38, No 2, s. 12-18
- Araz, O. Tsan-Ming, C. Olson, D. Salman, F. (2020). *Role of Analytics for Operational Risk Management in the Era of Big Data*. *Decision Sciences*, Atlanta. Vol 51, No 6.
- Rajpopat, J. Jamar, R. Lekhrajani, S. (2017). *Data Analytics in Banking*. *International Journal of Advanced Research in Computer Science*. Vol 8, No. 9.
- Kaur, G. Jagdev, G. (2020). *Analyzing and Exploring the Impact of Big Data Analytics in Sports Science*. *Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*. S. 218-224
- Johnson, D. Muzellec, L. Sihi, D. Zahay, D. (2019). *The marketing organization's journey to become data-driven*. *Journal of Research in Interactive Marketing; Brandford*. Vol. 13, No 2, s. 162-178.
- Svilar, M. Chakraborty, A. Kanioura, A. (2013). *Big data analytics in marketing*. *OR-MS Today; Linthicum*. Vol 40, No 5.
- Jiwat, R. Numan, K. Khawar, A. (2019). *Adoption of Big Data analytics in construction: development of a conceptual model*. *Built Environment Project and Asset Management; Bingley*. Vol 9, No 4, s. 564-579.
- Aibinu, A. Koch, F. Ng, S. (2019). *Data analytics and big data in construction project and asset management*. *Built Environment Project and Asset Management; Bingley*. Vol 9, No 4, s. 474-475.
- Jean-Baptiste, H. Qiu, M. Gai, K. Tao, L. (2015). *Meta Meta-Analytics for Risk Forecast Using Big Data Meta-Regression in Financial Industry*. *IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015*, pp. 272-277

- Bumblauskas, D., Nold, H., Bumblauskas, P., & Igou, A. (2017). Big data analytics: Transforming data to action. *Business Process Management Journal*, 23(3), 703-720.
- Ying-Kui, W, Qian-Mao, T. (2012). A Query Keywords Basen Approach for Noisy Data Elimination. IEEE.
- Virender, K. Cherry, K. (2018). *Data Cleaning-A Thorough Analysis and Survey on Unstructured Data*. IEEE International Conference on Cloud Computing, Data Science & Engineering, 11-12.1.2018.
- Xiong, H. Pandey G. Steinbach M. Kumar V. (2006). *Enhancing data analysis with noise removal*. IEEE Transactions on Knowledge and Data Engineering, Vol 8, No 3, s. 304-319.
- Emani, C. Cullot, N. Nicolle, C. (2018) *Understandable Big Data: A survey*. Computer Science Review, Vol 17, s. 70-81.
- Loyd, B. Kannan, D. (2017). *Identifying Desing Patterns for Risk Management system using Big Data Analytics*. International Conference on Trends in Electronics and Informatics (ICEI), 2017, s. 305-312.
- Goel, P. Datta, A. Mannan, M. (2017). *Application of big data analytics in process safety and risk management*. 2017 IEEE International Conference on Big Data (Big Data), s. 1143-1152.
- EMC Education Services. (2015). *Data science & big data analytics: discovering, analyzing, visualizing, and presenting data*. John Wiley and Sons.
- Sommer, B. (2015). *The big impact of Big Data on accounting*. Presentation in AAA Annual meeting, American Accounting Association, Chicago. Elokkuu, 8-12.
- Schoenherr, t. Speier-Pero, C. (2015). *Data science, predictive analytics, and big data in supply chain manahement: current state and future potential*. Journal on Business Logistics, Vol 36, No 1, s. 120-132.
- Feyyad, U. (1996). *Data mining and knowledge discovery: making sense out of data*. IEEE Expert, Vol 11, No 5, s. 20-25.
- Choi, T. Chan, H. K. Yue, X. (2017) *Recent Development in Big Data Analytics for Business Operations and Risk Management*. in IEEE Transactions on Cybernetics, vol. 47, no. 1, pp. 81-92
- Varma, T. 2014. *Information Technology in Supply Chain Management*. Journal of Supply Chain Management
- Nisbet, R., Miner, G., Yale, K., Elder, J. & Peterson, A. (2018). *Handbook of statistical analysis and data mining applications* (Second edition.). Academic Press.,
- Gan, Q. Luo, B. Lin, Z. (2019). *Risk Management of Residential Mortgage in China Using Date Mining A Case Study*. 2009 International Conference on New Trends in Information and Service Science, s. 1378-1383.

- Beroggi, G. Wallace, W. (1994). *Operational risk management: A new paradigm for decision making*. IEEE Trans. Syst. Man Cybern. Syst., vol. 24, no. 10, s. 1450-1457.
- Weeserik, P. Spruit, M. (2018). *Improving Operational Risk Management using Business Performance Management technologies*. Sustainability, Vol 10, No 3.
- Culp, C. (2002) *The risk management process: Business strategy and Tactics*. John Wiley and Sons.
- Shi, Y. Wang, T. Alwan, L. (2020). *Analytics for Cross-Border E-Commerce: Inventory risk management of an online fashion retailer*. Decision Sciences, Vol 51, No 6.
- Koyuncugil, A. OZgulbas, N. (2012). *Financial early warning system model and data mining application for risk detection*. Expert Systems with Applications, Vol 39, No 6, s. 6234-6253.
- Hailemariam, G. Hill, S. Demissie, S. (2012). *Exploring data mining techniques and algorithms for predicting customer loyalty and loan default risk scenarios at wisdom microfinance, Addis Ababa, Ethiopia*. Proceedings of the International Conference on Management of Emergent Digital Ecosystems. 10/2012, s. 183-184.
- Deshmukh, A. Talluru, L. (1998). *An application of a data mining technique for assessing the risk of management fraud*. Review of business information systems, Vol 2, No 1, s. 1-16.
- Wieland, A. Handfield, R. Durach, C. (2016). *Mapping the landscape of future research themes in supply chain management*. Journal of business logistics, Vol 37, No 3, s. 205-212.
- Jamshidi, A. Faghih-Roohi, S. Hajizadeh, S. Núñez, A. Babuska, R. Dollevoet, R. Li, Z. Schutter, B. (2017). *A big data analysis approach for rail failure risk assessment*. Risk Analysis an international journal, Vol 37, No 8, s. 1495-1507.
- Ban, G.Y. Rudin, C. (2018). *The big data newsvendor: practical insights from machine learning*. Operations research, Vol 67, No 1, s. 2-4.
- Kara, M. Firat, S. Ghadge, A. (2020). *A data mining-based framework for supply chain risk management*. Computers & Industrial Engineering, Vol 139, No 1.
- Zage, D. Glass, K. Colbaugh, R. (2013). *Improving supply chain security using big data*. 2013 IEEE International Conference on Intelligence and Security Informatics. 4-7.6.2013.
- Zagorecki, A. Johnson, D. Ristvej, J. (2014). *Data mining and machine learning in the context of disaster and crisis management*. International Journal of Emergency Management, Vol 9, No 4.
- Koh, H. Tan, W. Goh, C. (2004). *Credit scoring using data mining techniques*. Singapore Management Review, Vol 26, No 2.

- Hormozi, A. Giles, S. (2006). *Data mining: a competitive weapon for banking and retail industries*. Information Systems Management. Vol 21, No 2, s. 62-71.
- Biffis, A. Chavez, E. (2017). Satellite data and machine learning for weather risk management and food security. Risk analysis an international journal, Vol 37, No 8, s. 1508-1521.
- Peng, X. Chang, Y. (2020). *Energy-efficient routing technique for reliable data transmission under the background of big data for disaster region*. Computational Intelligence, Vol 36, No 4, s. 1593-1608.
- Yusoff, A. Din, N. Yussof, S. Khan, S. (2015). Big data analytics for Flood Information Management in Kelantan Malaysia. *IEEE Student Conference on Research and Development (SCOReD)*, s. 311-316
- Shah, S. Seker, D. Hameed, S. Draheim, D. (2019). *The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects*. IEEE Access, vol. 7, s.. 54595-54614
- Avvenuti, M. Cresci, S. Vigna, F. Fagni, T. Tesconi, M. (2018). CrisMap: a big data crisis mapping system based on damage detection and geoparsing. Information Systems Frontiers, Vol 20, No 5.
- Chaudhuri, A. Dukovska-Popovska, I. Subramanian, N. Chan, H. Bai, R. (2018). *Decision-making in cold chain logistics using data analytics: a literature review*. The international journal of logistics management, Vol 29, No 3.
- Singh, A. Shukla, N. Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries. Transportation Research, Vol 112, s. 298-415.
- Virender, K. Cherry, K. (2018). *Data Cleaning-A Thorough Analysis and Survey on Unstructured Data*. IEEE International Conference on Cloud Computing, Data Science & Engineering
- Morton, J. (2014). Big data: opportunities and challenges. Swindon: BCS, The Chartered Institute for IT.
- Duan, L. (2014). *Big data analytics and business analytics*. Journal of Management Analytics, Vol 2, No 1, s. 1-21.