RISKIT BIG DATAN HYÖDYNTÄMISESSÄ VALMISTAVASSA TEOLLISUUDESSA

Risk assessment of utilizing big data in manufacturing industry

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tuotantotalouden kandidaatintyö

2023

Eero Hämäläinen

Tarkastaja: Dosentti Kalle Elfvengren

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tuotantotalous

Eero Hämäläinen

Riskit big datan hyödyntämisessä valmistavassa teollisuudessa

Tuotantotalouden kandidaatintyö

2023

38 sivua, 10 kuvaa, 4 taulukkoa

Tarkastaja: Dosentti Kalle Elfvengren

Avainsanat: big data -analytiikka, datariski, teollinen big data, riskienhallinta

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

Digitalisaation ja nopean teknologian kehityksen myötä big data eli massiivisen kokoiset datamassat ovat aloittaneet uuden digitaalisen vallankumouksen aallon valmistavassa teollisuudessa. Valmistavassa teollisuudessa big datan hyötyjä on onnistuttu todistamaan monella osa-alueella, mutta siihen liittyviin haasteisiin ja riskeihin yritysten on kuitenkin varauduttava. Tässä työssä perehdytään big datan hyödyntämiseen liittyviin haasteisiin ja riskeihin yritysten näkökulmasta. Työ on toteutettu kirjallisuuskatsauksena ja se sisältää lisäksi lyhyen case-osuuden.

Työn alussa perehdytään johdannon jälkeen ensin yleisesti big datan teoriaan ja tarkemmin teollisen big datan hyödyntämiseen ja lähteisiin. Tämän jälkeen käsitellään lyhyesti yrityksen riskienhallinnan teoriaa ja riskien luokittelua ja käydään läpi ISO 31000 -standardin mukainen riskienhallintaprosessi. Varsinaisen teoriaosuuden jälkeen toteutetaan riskienarviointiprosessi big dataan liittyvistä riskeistä, missä ensin tunnistetaan olennaiset riskit, tehdään riskianalyysi ja lopuksi arvioidaan riskien merkitystä. Case-osuudessa käydään läpi erään suomalaisen PK-yrityksen tunnistamia riskejä.

Työssä havaittiin, että big datan tuomien hyötyjen takana on paljon merkittäviä riskejä, joihin yritysten on varauduttava riskienhallinnan näkökulmasta. Big datan hyödyntämiseen liittyy paljon epävarmuutta usealla osa-alueella, mikä luo haasteita myös eri riskitekijöiden tunnistamiselle. Merkittävimmät riskit liittyvät epävarmuuteen, tietoturvaan sekä datan laadunhallintaan.

Sisällysluettelo

1 Johdanto

Digitalisaation ja jatkuvan teknologian kehityksen myötä maailmanlaajuinen datan määrä on lähtenyt lähes eksponentiaaliseen kasvuun. Esimerkiksi viimeisen kymmenen vuoden aikana maailmassa liikkuvan datan määrän on arvioitu yli kymmenkertaistuneen. (Taylor 2022) Datan määrän kasvun lisäksi samalla datan varastointi sekä yleiset ominaisuudet ovat monimutkaistuneet, luoden uusia haasteita datan hallintaan sekä analysointiin. Nykyaikaa voidaankin kutsua jo niin sanotuksi ”big datan” -aikakaudeksi. Big data on kasvava trendi, jonka erilaisista mahdollisuuksista ja sovelluskohteista on kirjoitettu paljon, mutta kuitenkin hieman varjoon ovat jääneet vielä sen hyödyntämiseen liittyvät haasteet ja uhat. (Clarke 2016)

Etenkin valmistavan teollisuuden toimialalla big datan kerääminen on aloittanut uuden digitaalisen vallankumouksen. Nykyaikaisten edullisempien sensorien ja kehittyneiden tietoteknisten järjestelmien myötä valmistavan teollisuuden toimiala tuottaa enemmän dataa kuin mikään muu toimiala (Choo & Dehghantanha 2020, 39). Teollinen big data sisältää paljon potentiaalia luoda lisäarvoa tuotannon arvoketjun eri osa-alueilla, minkä myötä teollisen big data -analytiikan osaajista on syntynyt merkittävää kilpailua yritysten välillä. Teollisen big datan hallinta sisältää kuitenkin merkittäviä haasteita, joihin yritysten tulisi varautua riskienhallinnan keinoin big data -analytiikan suosion kasvaessa. (Wang, Zhang, Shi, Duan & Liu 2018)

Big data teknologioiden käyttöönoton onnistumisen kannalta kaikkien mahdollisten riskitekijöiden tunnistaminen on avainasemassa. Big dataan liittyvä monimutkaisuus ja epävarmuus hankaloittaa myös riskitekijöiden tunnistamista, minkä takia riskeihin on varauduttava inhimillisestä näkökulmasta riskienhallinnan keinoin liiketoimintaprosessien jatkuvuuden takaamiseksi (Clarke 2016).

1.1 Tavoite ja tutkimuskysymykset

Tämän kandidaatintyön tavoitteena on selvittää, minkälaisia riskejä liittyy big datan hyödyntämiseen valmistavan teollisuuden yritysten näkökulmasta ja kuinka nämä riskit tulisi huomioida yrityksen riskienhallinnassa. Big datan mahdollisuuksia ja käyttökohteita on tutkittu paljon, mutta kirjallisuutta sen käyttöön liittyvistä ongelmista ja näiden tuomista riskeistä on tarjolla paljon suppeammin. Työn varsinainen tavoite on siis muodostaa kattava riskikartoitus big datan luomista riskeistä ja haasteista valmistavan teollisuuden toimialalla. Työn päätutkimuskysymykset ovat:

”Mitkä ovat merkittävimpiä big datan hyödyntämiseen liittyviä riskejä?”

”Millaisia vaikutuksia tunnistetuilla riskeillä on?”

Päätutkimuskysymyksen ja sen tulosten ymmärtämiseksi on asetettu vielä lisäksi osatutkimuskysymys:

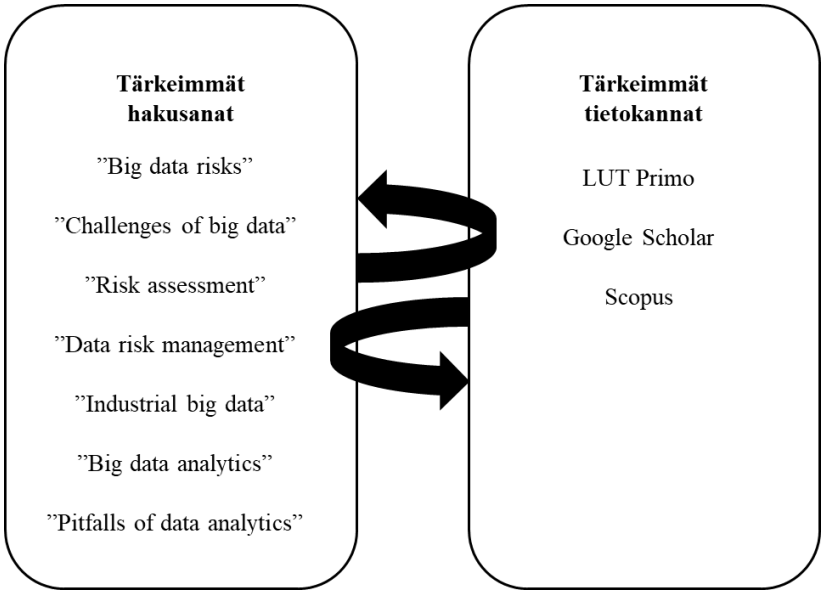
”Kuinka valmistavan teollisuuden yritykset hyödyntävät teollista big dataa?”

Tutkimuskysymysten vastausten perusteella pyritään luomaan lukijalle selvä kuva merkittävimmistä riskeistä big datan hyödyntämiseen liittyen ja näiden vaikutuksista. Tämän kandidaatintyön tulokset voivat olla hyödyksi etenkin valmistavan teollisuuden pienemmille yrityksille, jotka suunnittelevat big data teknologioiden käyttöönottoa ja haluavat tehostaa omaa datan hallintaa. Lisäksi työ voi herättää mielenkiintoa yleisesti big datasta ja sen hyödyntämisestä kiinnostuneissa henkilöissä.

1.2 Teoriatausta ja rajaus

Tämä kandidaatintyö rakentuu kahden suuremman aihekokonaisuuden pohjalle, jotka ovat riskienhallinta ja data-analytiikka. Työ on toteutettu kirjallisuuskatsauksena ja työn teoriatausta on rakennettu keräämällä tietoa molemmista aihekokonaisuuksista monipuolisesti useista eri tieteellisistä lähteistä. Lisäksi työn aikana on toteutettu yksi haastattelu, josta on muodostettu lyhyt case-osuus työn loppuun. Haastattelu toteutettiin erääseen suomalaiseen PK-yritykseen, joka toimii valmistavan teollisuuden toimialalla. Haastattelun avulla kartoitettiin case-yrityksen tunnistamia haasteita ja riskejä big datan keräykseen liittyen.

Työtä on rajattu data-analytiikan näkökulmasta big datan hyödyntämiseen valmistavan teollisuuden yrityksissä. Työssä keskitytään siis big datan tuomiin uusiin haasteisiin eikä käsitellä tavalliseen datan käsittelyyn liittyviä riskejä ja ongelmia. Riskienhallinnan näkökulmasta työtä on rajattu puolestaan ISO 31000 -standardin mukaiseen riskien arviointiprosessiin (risk assessment) kokonaisvaltaisen riskienhallinnan sijasta. Työssä pyritään siis tunnistamaan, analysoimaan ja arvioimaan erilaisia big dataan liittyviä riskejä. Työssä ei siis varsinaisesti pohdita erilaisia hallintakeinoja tunnistetuille riskeille tai perehdytä riskien seurantaan. Kuvassa 1 on vielä esitettynä tärkeimmät tiedonhaussa käytetyt hakusanat sekä tietokannat.



Kuva 1: Tiedonkeruussa käytettyjä hakusanoja ja tietokantoja

1.3 Työn rakenne

Tämä kandidaatintyö koostuu kokonaisuudessaan viidestä eri pääluvusta. Johdannon jälkeen on vuorossa kaksi teorialukua, jotka muodostavat lukijalle kokonaisvaltaisen pohjan työn tulosten ymmärtämiseksi. Työn toisessa pääluvussa käsitellään yleisesti big dataa ja analytiikkaa sekä perehdytään tarkemmin teolliseen big dataan ja sen hyödyntämiseen. Kolmannessa pääluvussa siirrytään yrityksen riskienhallinnan teorian esittelyyn, jonka yhteydessä luokitellaan erilaisia riskejä ja käydään läpi ISO 31000 -standardin mukainen riskienhallintaprosessi. Työn neljännessä pääluvussa siirrytään työn varsinaiseen aiheeseen eli big datan

luomiin riskeihin ja niiden arviointiin. Neljännessä luvussa yhdistetään aikaisemmin läpikäytyä teoriaa oman analyysin kanssa, minkä pohjalta tunnistettuja riskejä analysoidaan. Viidennessä pääluvussa käydään läpi case-osuus suomalaisen PK-yrityksen tunnistamista riskeistä big datan keräyksessä. Kuudennessa eli työn viimeisessä pääluvussa käydään läpi työn keskeisimmät tulokset ja niiden pohjalta tehdyt johtopäätökset.

2 Big data valmistavassa teollisuudessa

Big data on yksi tärkeimmistä piirteistä nykypäivän digitaalisessa ja kasvavassa tietoteknisessä yhteiskunnassa. Yleisesti termillä viitataan suurien datamäärien varastointiin, hallitsemiseen ja analysointiin. (Schintler & McNeely 2022, 79) Ominaisuuksiltaan big data eroaa merkittävästi tavallisesta datasta ja sen käsittelyyn eivät sovellu enää tavanomaiset metodit tai ohjelmistot (Fisher, DeLine, Czerwinski & Drucker 2012). Seuraavissa kappaleissa käsitellään tarkemmin big datan eri ominaisuuksia, big data -analytiikkaa ja teollisen big datan lähteitä ja käyttökohteita.

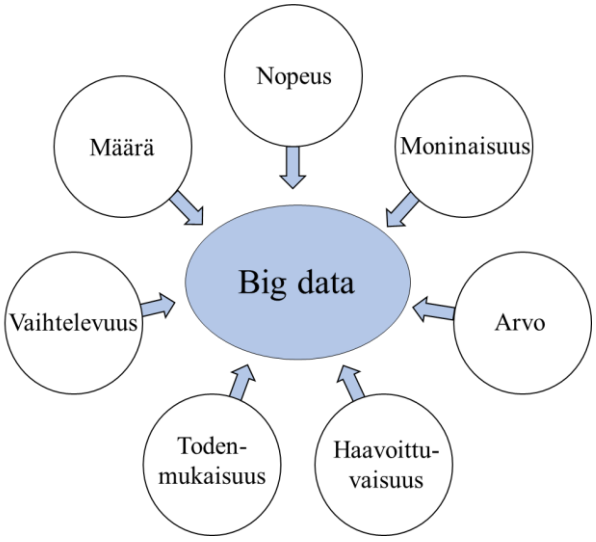
2.1 Big datan yleinen määritelmä

Big data on laaja ja kompleksi käsite, jolle ei ole sen vuoksi olemassa suoraan yksiselitteistä määritelmää. Termin määritelmä elää jatkuvasti ja on varsin tulkinnanvarainen riippuen, mistä näkökulmasta sitä lähestyy, esimerkiksi aihe voidaan tuoda esille eri tavalla eri sidosryhmille (Schintler & McNeely 2022, 79). Big dataa määritellään yleensä sen eri ominaisuuksien perusteella, joiden avulla voidaan luoda yleinen malli datasta. Big datan alkuperäinen ja ehkä tunnetuin versio, kolmen V:n malli tulee sanoista määrä (volume), nopeus (velocity) ja moninaisuus (variety) (McAfee & Brynjolfsson 2012, 62-63). Nämä kolme ulottuvuutta perustavat yhdessä käsitteellisen mallin (conceptual model) big datan kuvailemiselle (Schintler & McNeely 2022, 80).

Gandomin ja Haiderin (2015) mukaan big datan määrä ominaisuutena on varsin suhteellinen, eikä pelkän määrän perusteella voi vetää rajaa mikä lasketaan big dataksi ja mikä ei. Määrällä tarkoitetaan datamassan kokoa ja sitä kuinka nopeasti se kasvaa jatkuvasti (Hariri, Fredericks & Bowers 2019). Datan muoto ja ajankohtaisuus ovat datan määrän kannalta merkittävimmät tekijät. Esimerkiksi kaksi saman kokoista datamassaa voivat vaatia täysin erilaiset hallintamenetelmät riippuen datan muodosta. Vastaavasti datan ajankohtaisuus vaikuttaa merkittävästi sen määrän arviointiin johtuen teknologian nopeasti kehityksestä. Vaikka jokin datamassa luokiteltaisiin tällä hetkellä big dataksi, se ei todennäköisesti täytä määrältään kriteerejä enää tulevaisuudessa. Datan moninaisuudella tarkoitetaan datajoukon

rakenteellista heterogeenisyyttä eli datajoukko koostuu rakenteeltaan erilaisesta datasta. Data voi olla rakenteeltaan strukturoitua, osittain strukturoitua eli semistrukturoitua tai strukturoimatonta. Strukturoidulla datalla tarkoitetaan dataa, joka on saatavilla relaatiotietokannoista ja käsiteltävissä taulukkolaskentaohjelmilla. Strukturoimatonta dataa ovat puolestaan esimerkiksi erilaiset multimediatiedostot, jotka eivät ole tarpeeksi jäsenneltyjä koneoppimisen (machine learning) menetelmien hyödynnettäviksi. On arvioitu, että noin 80 - 90 prosenttia kaikesta maailman datasta on strukturoimatonta (Dialani 2020). Osittain strukturoitu data tarkoittaa nimensä mukaisesti sekalaista dataa, joka kuitenkin sisältää piirteitä strukturoidusta datasta. (Gandomi & Haider 2015)

Big datan kolmen V:n mallin jälkeen määritelmästä on nähty lukuisia eri versioita, mutta tällä hetkellä ajankohtaisimpana voidaan pitää Schintlerin ja McNeelyn (2022, 80) esittämää seitsemän V:n mallia (Kuva *2*). Lisäyksiä kolmen V:n mallin ominaisuuksiin ovat vaihtelevuus (variability), todenmukaisuus (veracity), haavoittuvaisuus (vulnerability) ja arvo (value). Vaihtelevuudella tarkoitetaan erilaisia epäjohdonmukaisuuksia datavirrassa, jotka luovat vaihtelua ja monimutkaisuutta (complexity). Todenmukaisuudella tarkoitetaan datan laadun ja luotettavuuden arviointia. Big data voi sisältää useita eri puutteellisuuksia, jotka vaikuttavat merkittävällä tavalla sen todenmukaisuuteen. Haavoittuvaisuus linkittyy olennaisesti esimerkiksi big datan käsittelyyn ja jakeluun liittyviin haasteisiin turvallisuuden ja yksityisyyden kannalta. Viimeinen ominaisuus eli arvo kuvaa kapasiteettiä kuinka paljon hyödyllistä arvoa big datasta saadaan puristettua ulos. (Schintler & McNeely 2022, 80)



Kuva 2: Big datan 7 V:n malli (mukaillen Schintler & McNeely 2022, 80)

Big datan erilaiset mallit ovat aiheuttaneet jonkin verran ristiriitaisuutta, siitä milloin big datan kriteerit varsinaisesti täyttyvät. Kitchin ja McArdle (2016) totesivat, että suurimmasta osasta tunnetuimmista malleista puuttuu ontologinen selkeys eli mallit ovat epämääräisiä ja luokittelevat big datan liian laajasti. Heidän tekemän tutkimuksen perusteella esimerkiksi vain pieni osa big dataksi luokitelluista datakokoelmista täyttävät kaikki yleisten big datan mallien ominaisuudet (Kitchin & McArdle 2016).

2.2 Big datan elinkaari

Datan elinkaarella tarkoitetaan koko datan hallinnan prosessia alusta loppuun saakka. Datan elinkaari alkaa datan keräämisestä ja päättyy datan tuhoamiseen. Elinkaari siis määrittelee datan virran yritysten järjestelmissä. (Rahul & Banyal 2020) Big datan elinkaari voidaan tyypillisesti jakaa viiteen eri vaiheeseen, jotka ovat datan keräys, varastointi, analytiikka, hyödyntäminen ja tuhoaminen (Koo, Kang & Kim 2020). Tässä työssä big datan elinkaari on kuitenkin jaettu neljään eri vaiheeseen, joissa datan analytiikka ja hyödyntäminen on liitetty samaan vaiheeseen tiivistäen hieman edellistä määritelmää (Kuva *3*).

Kuva 3: Big datan elinkaaren vaiheet (mukaillen Koo et al. 2020)

Big datan elinkaari alkaa datan keräyksestä, jossa dataa kerätään useista erilaista lähteistä eri muodoissa. Seuraavassa vaiheessa kerätty data varastoidaan talteen myöhempää käyttöä varten. Datan varastointi on yksi elinkaaren tärkeimmistä vaiheista, jotta dataa voidaan hyödyntää arvon luomiseksi. Analysointi- ja hyödyntämisvaiheessa dataa käsitellään erilaisilla kehittyneillä tekniikoilla, joiden avulla datasta saadaan kaivettua irti hyödylliset tiedot. Datan käsittelystä saatua uutta tietoa kyetään hyödyntämään sen jälkeen päätöksenteossa.

Elinkaaren viimeisessä vaiheessa eli datan tuhoamisessa analyysissä käytetty data poistetaan. Varsinkin yksityinen data on syytä tuhota ilman viiveitä tietoturvariskien välttämiseksi. (Koo et al. 2020)

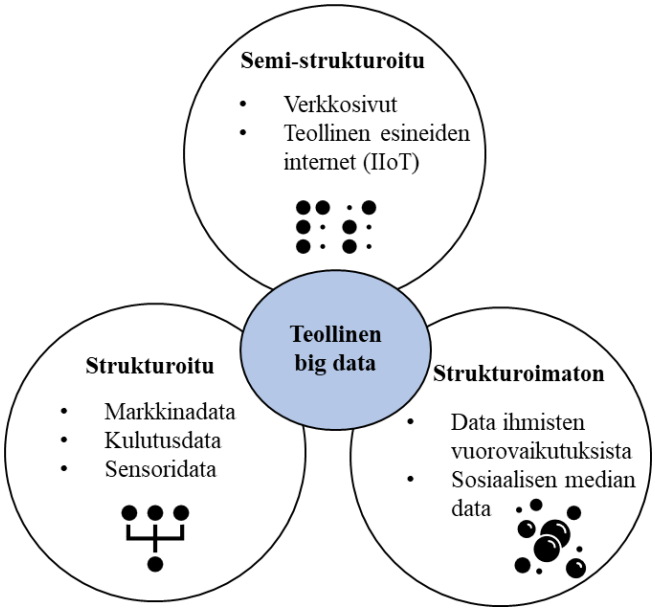
2.3 Valmistava teollisuus ja big datan lähteet

Digitalisaatio on ajanut valmistavan teollisuuden toimialan murrokseen kohti älykkäämpää tuottamista. Älykkäämpiä tehtaita rakennetaan jatkuvasti ja uusia kehittyneitä teknologioita otetaan käyttöön tuotannon tehostamiseksi ja uusien kilpailukykytekijöiden kehittämiseksi. Valmistusteknologioiden kehitys ja uudet digitaaliset liiketoimintamallit luovat paljon uusia mahdollisuuksia sekä tuovat lisäarvoa koko valmistavan teollisuuden arvoketjuille (VTT 2023). Big data -analyysiin pohjautuva päätöksenteko on yksi merkittävimmistä tekoälyn mahdollistamista teknologioista, joka parantaa kilpailukykyä koko toimialalla ja auttaa yritysten johtoa tekemään päätöksiä entistä laajemmalla skaalalla haastavissa ja monimuotoisissa valmistusympäristöissä. (Li, Chen & Shang 2022)

Teollisuusyritykset tuottavat tänä päivänä valtavia määriä monimuotoisia datamassoja. Näiden suurien datamassojen potentiaalin hyödyntäminen on kuitenkin osoittautunut haastavaksi, jotta dataa pystyttäisiin käsitellä sujuvasti uudessa toimintaympäristössä. Voidaan sanoa siis, että big data on aloittanut valmistavassa teollisuudessa digitaalisen vallankumouksen aallon. Erikokoiset teollisuusyritykset ympäri maailman pyrkivät kohti älykkäämpää tuottamista uusien digitaalisten strategioiden avulla kilpailukykytekijöiden saavuttamiseksi. Älykäs tuotanto kattaa tavallisen valmistavan teollisuuden tunnuspiirteet, mihin lisäksi yhdistetään näkökulmia informaatioteknologiasta. Älykkäällä tuotannolla pyritään hyödyntämään tuotteen koko elinkaaren eri vaiheista kerättyä dataa tuomaan lisäarvoa tuotantoon. (Li et al. 2022)

Teollisella big datalla tarkoitetaan dataa, jota kerätään valmistavassa teollisuudessa tuotteen eri elinkaaren vaiheista. Dataa kerätään esimerkiksi tuotteen suunnittelusta, tuotannosta, toimitusketjuista, markkinoinnista ja asiakaspalautteista. Datan lähteen perusteella teollinen big data voidaan jakaa karkeasti kahteen eri luokkaan, jotka ovat järjestelmädata ja IoT-data. Järjestelmädataa kerätään erilaisista yritysten omista järjestelmistä, jonka myötä se on myös yleisesti hyvin strukturoitua. IoT-dataa kerätään puolestaan erilaisilla sensoreilla ja RFID-lukijoilla (radio frequency identification), minkä takia data on vain osittain strukturoitua.

Lisäksi dataa kerätään internetistä kaupallisilta ja sosiaalisen median alustoilta ihmisten vuorovaikutuksista. Internetdata on usein strukturoimatonta, joka vaikeuttaa huomattavasti sen hyödyntämistä suuressa skaalassa. (Li et al. 2022) Kuvaan 4 on jaoteltu teollisen big datan lähteet datan rakenteen perusteella.



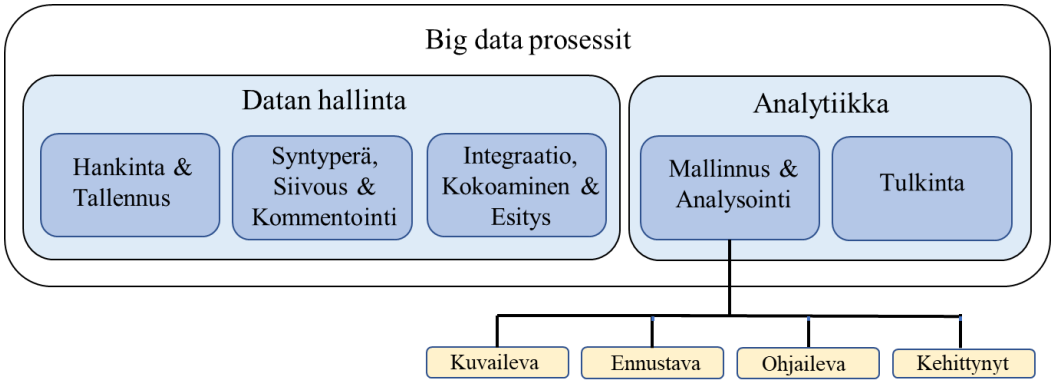
Kuva 4: Teollisen big datan lähteet (mukaillen Li et al. 2022)

Vaikka valmistavan teollisuuden toimintaympäristöissä tuotantoprosessit luovat koko ajan entistä enemmän yrityksille vaurasta dataa hyödynnettäväksi, on se silti vain pieni osa koko datamassasta. Yleensä suurin osa teollisesta raakadatasta on kokonaan hyödyntämättömissä tai jää hyödyntämättä, koska sille ei ole varsinaista käyttötarkoitusta (Li et al. 2022). Teollisen big datan analysointi ja hyödyntäminen älykkään tuotannon konseptissa on mahdollista vain, kun pystytään siirtymään perinteisestä tarkasti strukturoidun ja staattisen datan käsittelystä monipuolisempaan monirakenteisen ja dynaamisen datan hallintaan (Mourtzis, Vlachou & Milas 2016). Monipuolinen datan hallinta vaatii kehittyneitä analytiikan menetelmiä, jotta datasta saadaan varsinainen lisäarvo irti. Älykkään tuotannon konseptissa tarvittava tietotaito tuotannon kehittämiseen saadaan big data -analytiikan menetelmistä, jotka vahvistavat tuotannon kilpailukykyä globaaleilla markkinoilla. (Alcácer & Cruz-Machado 2019)

2.4 Big data -analytiikka

Perinteisellä data-analytiikalla tarkoitetaan datan käsittelyä, analysointia ja raportointia erilaisilla tietoteknisillä sovelluksilla, mikä tuo lisäarvoa päätöksenteon tueksi (Runkler 2020, 2). Perinteiset data-analytiikan työkalut eivät kuitenkaan kykene suoraan käsittelemään big datan tuomia valtavia datamassoja, vaan avuksi tarvitaan kehittyneempiä ohjelmistoja (Tsai, Lai, Chao & Vasilakos 2015).

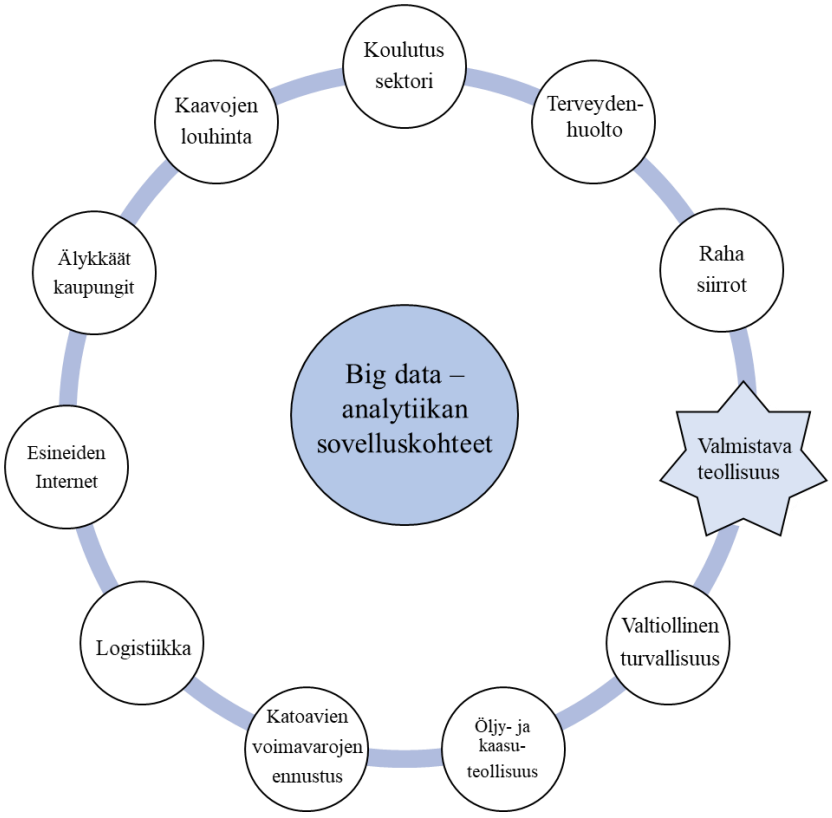
Big data -analytiikka termiä käytetään yleisesti kehittyneiden analytiikka menetelmien hyödyntämisestä, joita ovat esimerkiksi datan louhinta, tilastollinen analyysi, ennakoiva analytiikka ja koneoppiminen. Näiden menetelmien avulla pystytään havaitsemaan valtavista datamassoista esimerkiksi erilaisia piileviä kaavamaisuuksia ja korrelaatioita, trendejä sekä muuta liiketoiminnan kannalta arvokasta tietoa. Koko big datan käsittelyprosessin voi jakaa viiteen eri vaiheeseen, jotka ovat datan integrointi, datan hallinta, datan esikäsittely, datan louhinta ja tiedon esittäminen. Varsinainen big data -analytiikka voidaan lisäksi jakaa tarkemmin neljään eri luokkiin riippuen mistä näkökulmasta dataa tutkitaan (Kuva *5*). Kuvaileva analytiikka tarkastelee mitä datalle on tähän asti tehty, ennustava analytiikka pyrkii varautumaan ja ennakoimaan mitä datalle tullaan tekemään ja ohjaileva analytiikka keskittyy datan nykytilan analysointiin. Kehittynyt analytiikka on yhdistelmä ennustavaa ja ohjailevaa analytiikka, millä pyritään luomaan kattava yleiskuva tulevaisuus painotteisesti hyödyntämällä monipuolisia kehittyneitä menetelmiä. (Hassanien & Darwish 2021, 30;41)



Kuva 5: Big data prosessien luokittelu (mukaillen Gandomi & Haider 2015: Hassanien & Darwish 2021, 30)

Gandomi ja Haider (2015) puolestaan kuvailevat big data -analytiikkaa prosessiksi, jossa suurista ja nopeasti kehittyvistä datamassoista louhitaan irti hyödyllisiä oivalluksia. He jakavat prosessin viiteen eri vaiheeseen (Kuva *5*), jotka yhdessä muodostavat kaksi alaprosessia; datan hallinta ja analytiikka. Datan hallintaan kuuluu datan hankinta ja varastointi, datan siivous ja käsittely sekä datan kokoaminen analysointia varten. Analytiikka puolestaan pitää sisällään erilaisten analysointitekniikoiden hyödyntämistä datasta saatavan hyödyn louhimiseksi sekä havaintojen tulkitsemisen ja esittämisen. (Gandomi & Haider 2015)

Big data -analytiikan tuomia hyötyjä ja mahdollisuuksia on tutkittu paljon ja sitä on onnistuttu soveltamaan monenlaisiin käyttökohteisiin. Yleisimpiä sovelluskohteita ovat tähän mennessä olleet koulutussektori, terveydenhuolto, rahoitustapahtumat, turvallisuusvirasto, öljy- ja kaasuteollisuus, logistiikka, katoavien voimavarojen ennustaminen, esineiden internet (IoT), älykkäät kaupungit, kaavamaisuuksien louhinta ja valmistava teollisuus (Kuva *6*). (Ikegwu, Nweke, Anikwe, Alo & Okonkwo 2022) Big data -analytiikkaa kuitenkin sovelletaan jatkuvasti uusiin käyttötarkoituksiin yritysten tuottaessa entistä enemmän dataa ja omaksuessa yhä enemmän kehittyneitä analysointi menetelmiä.



Kuva 6: Big data -analytiikan sovelluskohteita (mukaillen Ikegwu et al. 2022)

Valmistavassa teollisuudessa big datasta saatavia hyötyjä on havaittu jo usealla eri osa-alueella. Big data -analytiikan avulla saatuja kehittyneitä ennusteita on käytetty hyödyksi esimerkiksi eri toimitusketjun vaiheissa kuten kysynnän suunnittelussa, hankintatoimessa, tuotannossa, varaston hallinnassa ja logistiikassa (Wang, Gunasekaran, Ngai & Papadopoulos 2016). Yleisesti big data -analytiikka auttaa tuotantojärjestelmien kehittämisessä sekä yrityksiä tekemään järkevämpiä päätöksiä esimerkiksi tuotteen ennustamisessa, yrityksen tehokkuuden johtamisessa, tuotesuunnittelussa ja asiakaspalvelussa. Varsinkin hieman vanhahtavissa tuotantojärjestelmissä työntehokkuutta voidaan tehostaa merkittävästi big datan avulla (Li et al. 2022). Big data -analytiikka hyödyntää useita erilaisia kehittyneitä teknologioita, kuten esimerkiksi koneoppimista, syväoppimista (deep learning) ja ennustavaa mallinnusta. (Shukla, Tiwari & Beydoun 2019) Varsinkin syväoppimisen algoritmit ovat tehokkaita käsittelemään hallitsemattomia datamassoja ja esittämään dataa (Najafabadi, Villanustre, Khoshgoftaar, Seliya, Wald & Muharemagic 2015). Kyvykkyys analysoida teollista big dataa on yrityksille siis arvokas resurssi, jolla voidaan luoda lisäarvoa.

Big data -analytiikka kokonaisuudessaan sisältää kuitenkin vielä paljon erilaisia pullonkauloja sekä varsinkin erilaisten big data teknologioiden omaksuminen sisältää omat haasteensa. Suurimmat haasteet liittyvät etenkin datan säilömiseen sekä epävarmuuteen datan analysoinnissa. Datan varastointiin liittyvät salassapito ja tietoturvallisuus ongelmat herättävät jatkuvasti enemmän huomiota ja niihin pyritään keksimään uusia ratkaisuja. Datan analysointiin liittyvä epävarmuus johtuu puolestaan datan laadusta. Datan laatuun vaikuttavat merkittävästi datan hankintaprosessi sekä analysointi menetelmät. Jotta dataa saadaan hyödynnettyä päätöksenteon tukena, on datamassoista löydettävä vain oikeat ja olennaiset havainnot, jolloin dataan perustuva analyysi ei ole harhaanjohtavaa. (Li et al. 2022) Big data -analytiikkaan liittyviä haasteita ja niiden luomia riskejä käsitellään tarkemmin neljännessä pääluvussa.

3 Yrityksen riskienhallintaprosessi

Tänä päivänä jatkuvasti kasvavassa ja muuttuvassa maailmassa yritykset altistuvat entistä enemmän erilaisille riskitekijöille. Lähes kaikki yritysten toiminta linkittyy epävarmuuteen tulevaisuuden kehitysaskelista, jotka voivat ilmetä yrityksen kannalta joko uhkina tai mahdollisuuksina. Yrityksille ei ole enää hyväksyttävää joutua kuitenkaan tilanteisiin, joissa ennalta arvaamattomat tapahtumat aiheuttavat taloudellisia menetyksiä, keskeytyksiä tavalliseen toimintaan tai vahinkoa yrityksen maineelle. Etenkin erilaiset sidosryhmät olettavat yritysten varautuvan kokonaisvaltaisesti kaikkiin yrityksen toimintaan vaikuttaviin riskitekijöihin. Tämän seurauksena riskienhallinnan rooli on korostunut, ja siitä on muodostunut pakollinen osa-alue yritysten toiminnassa. (Hopkin 2022, 21; Hunziker 2021, 2)

Yrityksen riskienhallinta on jatkuva prosessi, jonka myötä yritykset kykenevät varautumaan erilaisiin sisäisiin ja ulkoisiin haasteisiin. Hunzikerin (2021, 12) määritelmän mukaan yrityksen riskienhallinta on koko yrityksen laajuinen prosessi tunnistaa, arvioida ja hallita yrityksen toimintaa koskevia riskejä, minkä avulla luodaan arvoa kaikille eri sidosryhmille. Riskienhallinta on myös tärkeässä asemassa yrityksen strategian turvaamisessa (Kamensky 2014, 334).

3.1 Riskien luokittelu

Riskeillä voi olla niin positiivisia kuin negatiivisiakin vaikutuksia tai niiden lopputulema voi perustua epätietoisuuteen. Riskit voidaan yhdistää siis mahdollisuuksiin, menetyksiin tai epävarmuuden läsnäoloon. Näiden vaikutusten perusteella riskit voidaan lajitella kolmeen eri luokkaan; vahinkoriskit (hazard), epävarmuusriskit (uncertainty) ja mahdollisuusriskit (opportunity). Yleisesti yritykset pyrkivät lieventämään vahinkoriskejä, hallitsemaan epävarmuusriskejä ja omaksumaan mahdollisuusriskejä. (Hopkin 2014, 15)

Vahinkoriskit heikentävät riskin kohdetta, ja riskin vaikutusta mitataan sen merkittävyydellä. Vahinkoriskeillä on siis ainoastaan negatiivisia vaikutuksia yritysten toimintaan. Vahinkoriskit ovat tyypillisesti yhteydessä liiketoiminnan riippuvaisuuksiin ja tietoteknisiin järjestelmiin. Esimerkiksi tietotekniset järjestelmät voivat altistua helposti erilaisille

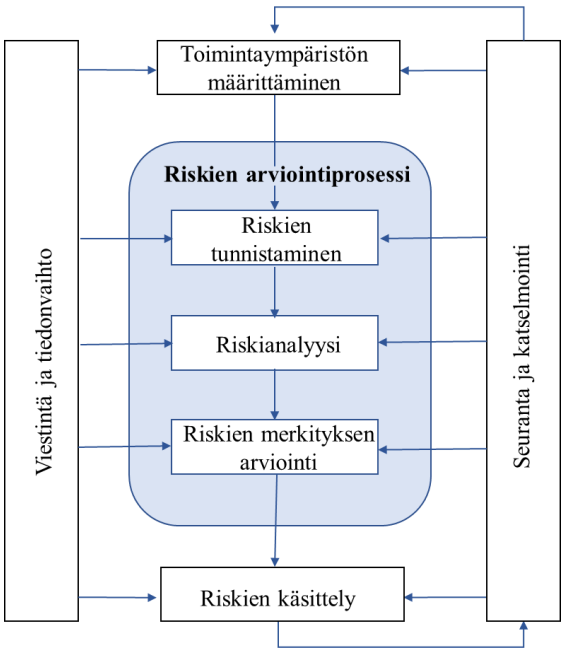
vahinkoriskeille, mikäli kyberturvallisuus asiat eivät ole ajan tasalla tai mahdollisten laiterikkojen sattuessa. Yleisesti varkaudet ja petokset ovat merkittävimpiä vahinkoriskejä suurimmalle osalle yrityksistä. (Hopkin 2014, 22)

Epävarmuusriskeihin liittyvä tietämättömyys ja epävarmuus tekee riskeistä haastavasti tunnistettavia ja määriteltäviä. Näille riskeille on tunnusomaista esimerkiksi, että riskin seuraamukset ovat jollain tapaa tiedossa, mutta riskin todellista vaikutusta ja ajankohtaa on vaikea ennakoida ja hallita. Epävarmuusriskien hallinta perustuu siis pitkälti riskin seurausten arviointiin ja ennustamiseen. Epävarmuusriskit ovat tyypillisiä etenkin projektin hallinnassa. (Hopkin 2014, 15;27;32)

Mahdollisuusriskit ovat riskejä, joita yritykset ottavat tietoisesti mahdollisten positiivisten hyötyjen saavuttamiseksi. Mahdollisuusriskien yhteydessä yritysten on suhteutettava tarkasti riskin realisoituessa syntyvät seuraukset riskistä saataviin potentiaalisiin hyötyihin. Mahdollisuusriskejä on tyypillisesti kahdentyyppisiä, riskit voivat liittyä joko mahdollisuuden ottamiseen tai mahdollisuuden käyttämättä jättämiseen. Tyypillisesti mahdollisuusriskit ovat taloudellisia ja hyvä esimerkki mahdollisuusriskeistä on investointiriskit. (Hopkin 2014, 16)

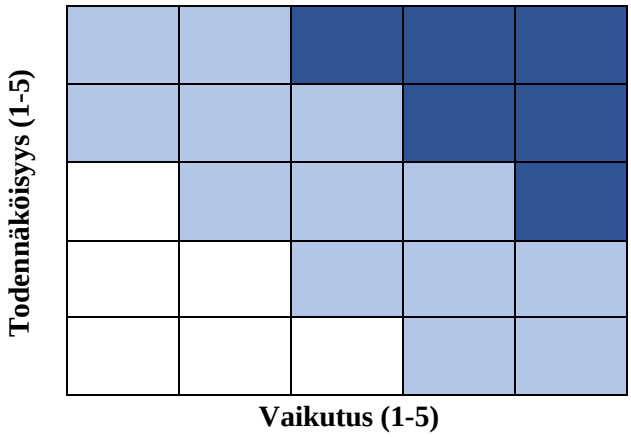
3.2 ISO 31000 mukainen riskienhallinta

ISO 31000 on yksi suosituimmista maailmalla käytetyistä riskienhallinnan standardeista. Standardi on kansainvälisen ISO (International Organization for Standardization) organisaation tuottama ja ylläpitämä. Standardissa esitetään yleiset suuntaviivat riskienhallinnan pohjalle ja ohjeita riskienhallinnan toimintapuitteiden ja prosessien kehittämiselle. Standardi koostuu kolmesta peruselementistä, jotka ovat riskienhallinnan periaatteet, organisaation huomioitavat puitteet ja riskienhallintaprosessin kuvaus (Kuva *7*). (Riskikompassi 2023)



Kuva 7: Riskienhallintaprosessin kuvaus (mukaillen SFS-ISO 2018, 13)

Standardin mukainen riskienhallintaprosessi alkaa toimintaympäristön määrittämisellä. Yrityksen tulee määritellä riskienhallintaprosessin kohde ja laajuus ja linjata nämä koko organisaation tavoitteiden mukaisesti. Riskienhallinnan lähestymistapaa suunniteltaessa huomioitavaa on etenkin yrityksen tavoitteet ja riskienhallintaprosessin odotettu lopputulema, sopivat riskienarviointi työkalut, saatavilla olevat resurssit ja vastuusta huolehtiminen sekä prosessin suhteet yrityksen muun toiminnan kanssa. Yrityksen tulee myös määritellä heidän riskinkantokykynsä, eli minkälaisia riskejä he kykenevät ottamaan saavuttaakseen tavoitteensa. (SFS-ISO 2018, 14-15) Riskinkantokyvyn määrittämiseksi yrityksen on pystyttävä arvioimaan riskejä niiden todennäköisyyden sekä vaikutusten perusteella. Yleinen työkalu riskien arviointiin on riskimatriisi, johon riskit voidaan sijoittaa niille lasketun todennäköisyyden ja vaikutusten perusteella. Riskimatriisi skaalataan todennäköisuus- ja vaikutusakseleilla, kuten kuvassa 8 on esimerkiksi esitetty. Todelliset riskitasot voivat kuitenkin muuttua jatkuvasti riippuen riskin ajankohdasta ja usean riskin yhteisvaikutuksesta. (Hopkin 2018, 124)



Kuva 8: Esimerkki riskimatriisista

Seuraava vaihe riskienhallintaprosessissa on riskienarviointiprosessi. Riskienarviointiprosessi koostuu kolmesta vaiheesta, jotka ovat riskien tunnistaminen, riskianalyysi ja riskien merkittävyyden arviointi. Riskienarviointiprosessin onnistumisen kannalta tärkeää on toimia systemaattisesti ja lähestyä riskejä kaikkien eri sidosryhmien näkökulmista. Riskien tunnistamisessa tavoitteena on tunnistaa kaikki mahdolliset yrityksen toimintaan vaikuttavat riskitekijät ja kuvailla näiden riskien piirteitä. Riskianalyysin tarkoitus on pohtia syitä tunnistettujen riskien taustalla sekä näiden seurauksia. Riskeillä voi olla useita erilaisia vaikutuksia, minkä takia on syytä varautua useisiin eri skenaarioihin. Riskianalyysissä olennaista on tunnistettujen riskien luokittelu todennäköisyyden ja vaikutuksien perusteella. Riskienarviointiprosessin viimeisessä vaiheessa arvioidaan riskien merkitystä vertaamalla riskien todellista tasoa yrityksen määrittämään riskienkantokykyyn. Vertailun perusteella kyetään tekemään päätöksiä riskien käsittelyn jatkotoimenpiteistä. (SFS-ISO 2018, 16-17)

Riskien käsittelyvaiheessa valitaan ja toteutetaan riskeille parhaiten soveltuvat hallintakeinot. Riskien käsittely on jatkuva iteratiivinen prosessi, joka koostuu siis riskin käsittelyn suunnittelusta, toteuttamisesta ja arvioimisesta. Mikäli riski ei ole käsittelyn jälkeen vieläkään hyväksyttävissä otetaan se uudelleen jatkokäsittelyyn. Yleisiä riskienkäsittely keinoja ovat riskialttiin toiminnan kokonaan välttäminen tai lopettaminen, riskin hallittu vastaanottaminen siitä saatavan mahdollisen hyödyn vuoksi, riskilähteen eliminointi, riskin todennäköisyyden tai seurausten muuttaminen, riskin jakaminen ja riskin hyväksyminen. Riskien hallintamenetelmien valinnassa on huomioitava erityisesti organisaation tavoitteet, riskinkantokyky, saatavilla olevat resurssit sekä eri sidosryhmät. (SFS-ISO 2018, 17-18)

4 Big datan luomien riskien arviointi

Valmistava teollisuus on muovautumassa kohti älykkäämpää tuotantoa big datan käynnistämän digitaalisen vallankumouksen myötä. Big dataan pohjautuva analyysi ja päätöksenteko on kuitenkin vielä vasta varhaisessa asemassa osana teollista esineiden internettiä, jonka myötä uusia ratkaisuja on vielä kehitettävä. Perinteisen valmistavan teollisuuden ja nykypäivän kehittyneen tietotekniikan välille on rakennettava siltoja, jotka helpottavat siirtymistä kohti älykkäämpää tuotantoa. Big datan tuomia hyötyjä lukuisilla eri osa-alueilla on pystytty todistamaan, mutta vastaavasti big data pohjaisten teknologioiden kehitys ja käyttöönotto sisältää merkittäviä haasteita.

Big datan luomat haasteet keskittyvät etenkin ongelmiin datan varastoinnissa, siirrossa, turvallisuudessa, yksityisyydessä ja laskennallisessa monimutkaisuudessa. Jatkuvasti kasvavat datamassat vaativat nopeasti latautuvia varastoratkaisuja, jotka pystyvät skaalautumaan tarvittaessa. Esimerkiksi suuri osa tänä päivänä kerätystä datasta joudutaan sivuuttamaan tai poistamaan, koska suurille data määrille ei ole tarpeeksi varastointitilaa. Erilaiset pilvipalvelut ovat nousseet suosituiksi ratkaisuiksi datan varastoinnissa, kuitenkin sisältäen omat riskinsä esimerkiksi turvallisuuteen ja varastointikuluihin liittyen. Datan varastointiin kytkeytyy myös olennaisesti ongelmat datan siirrossa. Kasvavien datamassojen siirtely tuottaa ongelmia kaikissa eri datan elinkaaren vaiheissa. Jotta dataa saadaan siirreltyä tehokkaasti paikallisten keskusten ja pilven välillä, edellyttää se älykästä datan esikäsittelyä ja algoritmeja datan koon pakkaamiseksi. Datan käsittely vaatii kuitenkin kehittyneitä informaatioteknologia resursseja, jotka johtavat uusiin haasteisiin laskennallisuuden monimutkaisuudessa. (Ramya, Sakthi Devi, Senthil Pandian, Suguna, Suganya & Manimozhi 2023)

Datan turvallisuuteen liittyvät haavoittuvaisuudet korostuvat big datan ja pilvipalveluiden integraation myötä. Tavanomaisissa tietokannoissa datan käsittelyä koskevat turvallisuus käytännöt eivät toimi yhtä tehokkaasti monimuotoisen datan käsittelyssä pilviympäristöissä. Etenkin datan luottamuksellisuutta, yhtenäisyyttä ja saatavuutta on turvattava, jotta uusia tehokkaampia datan hallintajärjestelmiä ja varastoratkaisuja voidaan kehittää. Lisäksi datan yksityisyyden suojaaminen on toinen huomioitava tekijä turvallisuuden ohella. (Ramya et al. 2023)

Big data teknologioiden omaksumiseen liittyy siis paljon epävarmuutta eri osa-alueilla, mikä luo erilaisia riskejä yrityksille. Epävarmuuden eri olomuodot muun muassa vaikuttavat negatiivisesti big data -analytiikan tehokkuuteen ja tarkkuuteen (Hariri et al. 2019). Calvard ja Jeske (2018) puolestaan toteavat big data järjestelmien monimutkaisuuden ja epävarmuuden hankaloittavan oleellisten riskitekijöiden tunnistamista. Heidän mukaan dataan liittyvät vahingot saavat usein alkunsa monen pienemmän, yksinään vaarattomamman tekijän yhteisvaikutuksesta, jolloin ratkaisevaa riskitekijää on vaikea tunnistaa. Yrityksen riskienhallinnan rooli on kuitenkin tunnistaa nämä riskitekijät ja luoda strategia, jonka avulla yritys kykenee varautumaan eri riskeihin ja täten pääsemään asettamiinsa tavoitteisiin (Hopkin 2014, 4). Riskien tunnistamisen ja ymmärtämisen apuna tässä kappaleessa hyödynnetään ISO 31000-standardin määrittelemää riskienarviointiprosessia. ISO 31000 -standardi on yksi yleisimmistä standardeista, joita sovelletaan big data arkkitehtuurin yhteydessä (Malik & Singh 2019).

4.1 Riskien tunnistaminen

Riskien tunnistaminen on ensimmäinen vaihe ISO 31000 -standardin mukaisesta riskien arviointiprosessista. Riskien tunnistamisprosessissa pyritään löytämään kaikki yrityksen toimintaan liittyvät riskit eri lähteistä, sekä kuvailemaan ja ymmärtämään tunnistettuja riskejä (de Oliveira, Marins, Rocha & Salomon 2017). Tässä luvussa tavoitteena on siis tunnistaa mahdollisimman kattavasti big datan hyödyntämiseen liittyviä riskejä valmistavan teollisuuden yritysten näkökulmasta. Riskit ovat jaettu neljään eri osa-alueeseen perustuen datan elinkaaren eri vaiheisiin. Datan elinkaari on jaettu neljään vaiheeseen, jotka ovat datan keräys, datan varastointi, datan analysointi ja käyttäminen ja datan tuhoaminen.

4.1.1 Datan keräyksen riskit

Datan keräysvaiheeseen liittyvät riskit ovat yritysten kannalta merkittävimpiä, sillä niiden realisoituessa syntyvät ongelmat luovat uusia riskejä myöhempiin datan elinkaaren vaiheisiin. Suurimpina riskeinä datan keräyksessä voidaan pitää epävarmuutta kerättävästä datasta ja tietomallin puutteellisuutta. Tietomalli tarkoittaa yleisesti mallinnusta yrityksen

keräämästä ja tuottamasta datasta ja niihin liittyvistä suhteista (Microsoft 2023). Mikäli yritys kerää liikaa ylimääräistä dataa tai jättää hyödyllistä dataa keräämättä aiheuttaa se ongelmia datan hyödyntämisessä. Kerätty data tulee myös puhdistaa varastointia varten ja etenkin datan laadun ja jäljitettävyyden on oltava kunnossa. Mikäli data ei ole jäljitettävissä tai se on puutteellista, dataan perustuvien päätösten epätarkkuus kasvaa merkittävästi. Varsinkin liiallinen datan keräys vaikeuttaa datan puhdistamista ja lisää epävarmuutta sen käsittelyssä. Dataan liittyvällä epävarmuudella ja datamassan koon välillä on havaittu positiivinen korrelaatio, eli datamassojen kasvaessa niihin liittyvä epävarmuus kasvaa vastaavasti (Hariri et al. 2019).

Suurten datamassojen kerääminen vaatii paljon resursseja, minkä vuoksi investoinnit big data teknologioiden käyttöönottoon sisältää taloudellisia riskejä. Etenkin puute erilaisista informaatiojärjestelmistä ja infrastruktuurin tuesta ovat yleisiä ongelmia (Raguseo 2018). Lisäksi sopivan ohjelmistokehyksen löytäminen ja sovittaminen yrityksen omaan toimintaympäristöön voi olla haastava ja monimutkainen prosessi, joka vaatii paljon resursseja etenkin, jos yrityksen on kehitettävä oma ohjelmistokehys (Li et al. 2022). Investointiriskit tulisi suhteuttaa niistä saataviin hyötyihin, mutta big datan ympärillä vellova epävarmuus hankaloittaa myös datan potentiaalisten hyötyjen arviointia.

Taulukko 1: Riskit datan keräyksessä

Datan keräyksen riskit	Riskityyppi
Epävarmuus kerättävästä datasta	Epävarmuus
Puutteellinen tietomalli	Vahinko
Sopivan viitekehyksen sovittaminen toimintaympäristöön	Epävarmuus
Puute informaatiojärjestelmistä ja infrastruktuurin tuesta	Vahinko
Datan laatu ja validointi puutteellista	Epävarmuus
Datan puhdistuksen työläys	Epävarmuus
Investointiriskit	Mahdollisuus
Epävarmuus big datan potentiaalisten hyötyjen arvioinnissa	Epävarmuus
Datan jäljitettävyys	Epävarmuus
Liikaa kerättävää dataa	Epävarmuus
Tietoturvariskit	Vahinko

4.1.2 Datan varastoinnin riskit

Yritykset hyödyntävät datan varastoinnissa tyypillisesti kahdenlaisia eri ratkaisuja eli pilvipalveluja sekä sisäisiä tietovarastoja (on-premise). Big datan aikakaudella pilvipalveluiden suosio on kuitenkin noussut reilusti laajemman varastointitilan ja alempien kustannusten takia. Pilvipalvelut ovat kuitenkin tietoturvallisuuden kannalta riskialttiimpia kuin tavalliset yritysten sisäiset tietovarastot. Lisäksi suuria ja kasvavia datamassoja tallennettaessa syntyy myös epävarmuutta lopullisten kustannusten arviointiin ja varastojen skaalautuvuuteen. (Maglaras, Janickle & Amine Ferrag 2022, 7; Ramya et al. 2023)

Data varastojen skaalautuvuuden puute hidastaa eri lähteistä kerättävän datan käsittelyä, mikä laskee koko prosessin tehokkuutta. Varsinainen ongelma datan varastoinnissa on siis kehittää skaalautuva varastointiteknologia, joka kykenee varastoimaan tehokkaasti monimuotoista dataa eri puolilta maailmaa. (Lv, Song, Basanta-Val, Steed & Jo 2017) Riskinä big dataa käsiteltäessä on, että kaikkea kerättyä dataa ei saada tallennettua myöhempää käyttöä varten. Lisäksi kerättäessä monipuolista dataa useista eri lähteistä ongelmaksi voi nousta erimuotoisten datojen yhdistäminen yhteensopivaan muotoon varastointia varten. Myös ongelmat varastossa olevan datan saatavuudesta ja jaettavuudesta ovat huomioitavia riskejä. (Ikegwu et al. 2022)

Datan varastoinnin yhteydessä merkittävimmät riskit koskevat datan tietoturvaa ja yksityisyyden suojaa. Li et al. (2021) totesivat datan turvallisuuden hallinnan olevan merkittävin big data teknologioita koskeva riski datan varastoinnissa. Sopimattoman turvallisuuden hallinnan seurauksena yritykset voivat joutua katastrofaalisten vahinkojen ja menetysten uhriksi (Li et al. 2021).

Taulukko 2: Riskit datan varastoinnissa

Datan varastoinnin riskit	Riskityyppi
Epävarmuus kulujen arvioinnissa	Epävarmuus
Varastojen skaalautuvuus	Vahinko
Tietoturvariskit	Vahinko
Yksityisyyden suoja	Vahinko
Datan yhdistäminen (integraatio)	Epävarmuus
Datan jakaminen	Epävarmuus
Datan saatavuus	Epävarmuus

4.1.3 Datan analysoinnin ja käytön riskit

Datan varsinaiseen hyödyntämiseen eli analysointiin ja raportointiin liittyvät riskit koskevat pääosin datan pohjalta tehtyjen päätöksien epätarkkuutta. Virheelliset tai harhaanjohtavat johtopäätökset ovat yleisin riski datan analysoinnissa. Big dataan pohjautuvien päätöksien paikkansapitävyyttä voi olla välillä vaikea arvioida ja usein päätöksille annetaan enemmän painoarvoa, kuin mitä se todellisuudessa on. Tämä johtaa helposti esimerkiksi resurssien epätarkkaan kohdentamiseen, mikä ennen pitkään näkyy negatiivisena vaikutuksena yrityksen liiketoiminnassa. (Clarke 2016)

Epätarkkojen päätöksien taustalla on yleensä työntekijöiden teknisen osaamisen puute tai puutteellinen datamassa (Raguseo 2018). Analysoinnin yhteydessä käytettyjen algoritmien skaalautuvuus ja monimutkaisuus sekä mahdollisesti vanhentuneiden työkalujen hyödyntäminen ovat myös huomioitavia riskejä. Datamassan laatua voidaan itsessään analysoida big datan eri tunnuspiirteiden kautta. Esimerkiksi datan luotettavuus, ajankohtaisuus ja monimuotoisuus ovat merkittäviä tekijöitä datan laadun kannalta. (Najafabadi et al. 2015) Lisäksi datan visualisointi voi osoittautua myös riskiksi, jolloin datasta saatavaa hyötyä ei kyetä esittämään. Etenkin heterogeenisten datamassojen analysointi ja visualisointi voi olla varsin haasteellista, koska ne sisältävät kriittisiä eroavaisuuksia tiedon rakenteessa (Wang et al. 2018). Analyysin ja raportoinnin ymmärrettävyys on siis merkittävä riski, jottei datan käsittelyyn käytetyt resurssit mene hukkaan.

Taulukko 3: Riskit datan analysoinnissa ja käytössä

Datan analysoinnin ja käytön riskit	Riskityyppi
Algoritmien skaalautuvuus	Epävarmuus
Algoritmien monimutkaisuus	Epävarmuus
Tekninen epävarmuus	Epävarmuus
Teknisen osaamisen puute	Epävarmuus
Datan laatu	Vahinko
Datan luotettavuus	Epävarmuus
Datan ajankohtaisuus	Epävarmuus
Datan monimuotoisuus	Epävarmuus
Analyysin ja raportoinnin ymmärrettävyys	Epävarmuus
Vanhentuneet työkalut ja teknologiat	Vahinko
Virheelliset johtopäätökset	Epävarmuus
Datan visualisointi	Epävarmuus
Tietoturvariskit	Vahinko

4.1.4 Datan tuhoamisen riskit

Datan tuhoamiseen liittyvät riskit korostuvat merkittävästi ulkoisten datan varastointi ratkaisujen kuten pilvipalveluiden yhteydessä. Riskiksi nousee tiedon lopullinen tuhoaminen eli miten yritys voi varmistua siitä, että heidän poistamansa tiedot ovat lopullisesti tuhottuja. Esimerkiksi käytettäessä ulkoista pilvipalvelua yrityksen on pystyttävä varmistamaan, että tiedot eivät ole enää saatavilla tai palautettavissa palveluntarjoajalle. Koska big datan yhteydessä on puhe valtavista datamääristä, on dataa yleensä tallennettuna myös useissa eri järjestelmissä. Tämä nostaa riskiksi sen, että dataa ei välttämättä saada poistettua samanaikaisesti pois kaikista eri järjestelmistä. Datan puutteellisen tuhoamisen seurauksena yritys voi altistua siis uusille tietoturvariskeille, jotka voivat ilmetä esimerkiksi datan väärinkäyttönä ja tietovuotoina. (Anbar, Abdullah & Manickam 2021)

Datan tuhoamisen yhteydessä on myös riski, että inhimillisen virheen seurauksena poistetaan vahingossa oleellista dataa. Suuria monimuotoisia datamassoja käsitellessä inhimillisten virheiden mahdollisuus kasvaa myös huomattavasti ja niiden taustalla voi olla esimerkiksi tekninen epävarmuus ja osaamisen puute. Vaikka poistettu data olisi palautettavissa, ei sen roolia välttämättä tunnisteta puuttuvaksi enää jälkeenpäin.

Taulukko 4: Riskit datan tuhoamisessa

Datan tuhoamisen riskit	Riskityyppi
Arvokkaan/oleellisten tietojen poistaminen vahingossa	Epävarmuus
Datan poistaminen kaikista järjestelmistä	Epävarmuus
Dataa ei saada tuhottua kokonaan	Epävarmuus
Datan väärinkäyttö	Vahinko
Tietoturvariskit	Vahinko

4.2 Riskianalyysi

Riskianalyysissä pyritään luomaan ymmärrys tunnistettujen riskien luonteesta ja kuinka ne vaikuttavat yrityksen päätöksentekoon (de Oliveira et al. 2017). Pohtimalla riskien syy-seuraussuhteita kyetään myös arvioimaan niiden vaikutusta ja todennäköisyyttä. Tunnistetuille big data riskeille ei voida kuitenkaan laskea yleistä numeraalista riskitasoa todennäköisyyden ja vaikutuksen perusteella, sillä ne ovat yrityskohtaisia. Vaikka yritykset kohtaavat

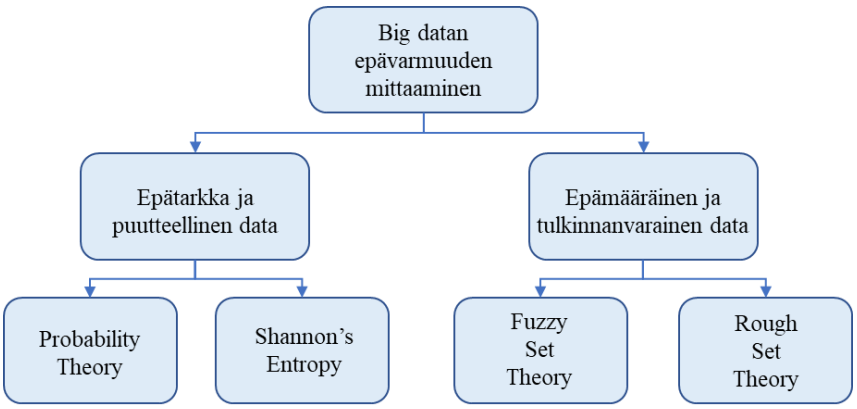
samoja riskejä on niiden todellinen vaikutus ja toteutumisen todennäköisyys erilainen riippuen yrityksen riskinkantokyvystä (Hopkin 2014, 19). Tunnistettujen riskien vaikutuksia kuitenkin vertaillaan keskenään ja niiden todennäköisyyksiä arvioidaan karkeasti.

Kaikki tunnistetuista riskeistä ovat luonteeltaan joko vahinko- tai epävarmuusriskejä eli niiden vaikutus yrityksen toimintaan on negatiivinen. Tässä kappaleessa riskien analysointi on jaettu kahteen osaan, joista ensimmäisessä käsitellään epävarmuusriskejä ja toisessa osassa vahinkoriskejä. Vahinkoriskit koostuvat pääosin tietoturvariskeistä.

4.2.1 Big datan epävarmuuden käsittely

Big datan tuoma epävarmuus on yksi merkittävimmistä riskitekijöistä, joka johtaa useisiin pienempiin riskeihin. Epävarmuuden takia näitä riskejä voi olla hankala arvioida tai jopa kokonaan tunnistaa. Epävarmuuden käsittelyllä on siis merkittävä vaikutus myös riskien käsittelyn onnistumisen kannalta. Epävarmuus datan keräyksessä, käsittelyssä ja analysoinnissa aiheuttavat yrityksille usein etenkin ongelmia resurssien kohdentamiseen, ongelmia tuotantoon sekä taloudellisia menetyksiä. Näiden riskien vaikutukset ovat varteenotettavia, mutta kuitenkin harvemmin kriittisiä laajemmassa skaalassa koko yrityksen toiminnan kannalta.

Arvioimalla big datan epävarmuutta kyetään paremmin hahmottamaan ja arvioimaan epävarmuuden aiheuttamia riskejä. (Hariri et al. 2019) Hariri et al. (2019) esittävät big datan epävarmuuden mittaamiselle neljä erilaista teoriaa, joita voi soveltaa riippuen mitattavan datan luonteesta (Kuva *9*). Seuraavassa kappaleessa käsitellään lyhyesti yksi näistä teorioista, Shannonin entropia.



Kuva 9: Big datan epävarmuuden mittaaminen (mukaillen Hariri et al. 2019)

Shannonin entropia on kaava systeemien epäjärjestyksen mittaamiselle, mitä voidaan soveltaa myös big datan epävarmuuden mittaukseen. Metodi soveltuu hyvin etenkin sattumanvaraisuuden ja subjektiivisen epävarmuuden mittaamiseen, joita esiintyy monimutkaisissa datamassoissa. Shannonin entropia on matemaattinen kaava, joka perustuu tiedon määrittämiseen muuttujaan, jonka perusteella voidaan havaita keskimäärin puuttuvan tiedon määrä eri lähteistä. (Hariri et al. 2019) Kaavan avulla laskettu entropian arvo kuvaa epävarmuutta lasketussa systeemissä. Mitä enemmän systeemi sisältää pieniä harvakseltaan esiintyviä muuttujia, sitä korkeampi entropian arvo on. (Bi, Jin, Maropoulos, Zhang & Wang 2021) Teorian avulla voidaan esimerkiksi laskea karkeasti kuinka paljon datamassat sisältävät ylimääräistä tai turhaa tietoa ja onko dataa kadonnut. Epävarmuuden mittaaminen auttaa siis etenkin datan laadun ja eheyden arvioinnissa.

Suurten datamassojen epävarmuuden vähentämisessä voidaan hyödyntää erilaisia kehittyneitä koneoppimisen menetelmiä. Koneoppimisen menetelmistä syväoppimisen ja aktiivisen oppimisen algoritmien on todettu soveltuvan parhaiten epävarmuuden hillitsemiseen. Syväoppimisen algoritmit soveltuvat etenkin puutteellisten ja epäjohdonmukaisten datamassojen käsittelyyn. Algoritmien avulla voidaan luoda monimutkaisia ja tiivistettyjä esityksiä suurista datamassoista (Najafabadi et al. 2015). Syväoppimisen malli edellyttää kuitenkin suuria laskennallisia resursseja, joista voi aiheutua merkittäviä kustannuksia. (Hariri et al. 2019)

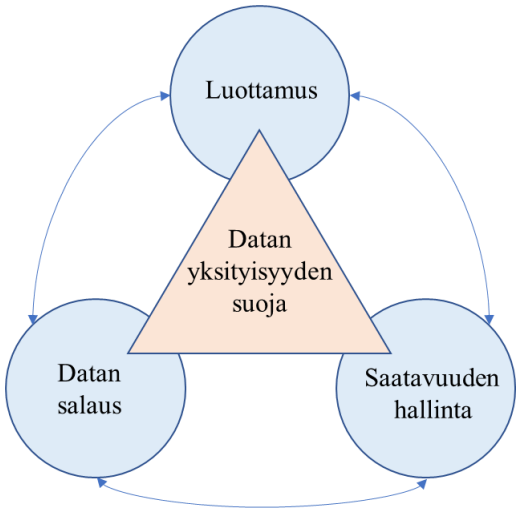
4.2.2 Tietoturvariskien vaikutus

Tietoturvariskit ovat läsnä koko datan elinkaaren ajan, mikä tekee niiden tunnistamisesta ja niihin varautumisesta äärimmäisen tärkeää. Tietoturvariskien vaikutukset voivat koitua todella merkittäviksi ja esimerkiksi Maglaras et al. (2022, 1) toteavat kyberhyökkäyksien olevan yksi kaikista merkittävimmistä yrityksiä uhkaavista riskeistä tämänpäiväisessä maailmassa. Kyberhyökkäyksistä voi aiheutua muun muassa taloudellisia tappioita, tietomurtoja ja pahimmassa tapauksessa liiketoiminnan keskeytyksiä. Varsinkin vakavista tietomurroista yrityksille voi jäädä pysyviä mainehaittoja. (Sheehan, Murphy, Mullins & Ryan 2019)

Etenkin viimeisen vuosikymmenen aikana IoT-laitteisiin kohdistuneiden kyberhyökkäysten määrä on noussut huomattavasti. IoT-laitteiden toimintaympäristön haavoittuvaisuudet liittyvät epäturvallisiin pilvipalveluiden käyttöliittymiin, puutteelliseen tunnistautumiseen ja

yksityisyyden suojaan sekä epäturvallisiin verkkoliittymiin ja ohjelmiin. Jatkuvasti yleistyneet pilvipalvelut ja kolmansien osapuolien tarjoamat uudet teknologiat hankaloittavat entisestään yrityksien datajärjestelmien turvaamista. (Chhabra, Singh & Singh 2020; Maglaras et al. 2022, 1)

IoT-laitteilla kerättävä data ja pilvipalveluiden hyödyntäminen datan varastoinnissa ovat merkittäviä komponentteja valmistavan teollisuuden digitalisaatiossa. Yritysten kohtaamat tietoturvariskit ovat siis väistämättömiä, jolloin riskitekijöiden pienentäminen on avainasemassa riskienhallinnassa. Riippuen yritysten erityistarpeista, heidän käyttämät pilvipalvelumallit voivat hieman poiketa toisistaan, mikä vaikuttaa myös tietoturvariskien luonteeseen (Chhabra et al. 2020). Kuitenkin yleisimpiä uhkia ovat tietovuodot, palvelunestohyökkäykset ja APT-hyökkäykset (Advanced Persistent Threat). APT-hyökkäyksillä tarkoitetaan hyvin organisoituja kyberhyökkäyksiä, joiden takana on tyypillisesti vahvoilla resursseilla varustettuja hakkeriryhmiä. Hyökkäyksille tyypillistä on toistuvuus ja huomaamattomuus. (Chen, Desmet & Huygens 2014, 64)



Kuva 10: Datan yksityisyyden suojaan vaikuttavat tekijät (mukaillen Maglaras et al. 2022, 13)

Datan yksityisyyden riskejä voidaan tarkastella useasta eri näkökulmasta, kuten esimerkiksi saatavuuden hallinnan, pilvijärjestelmien, asiakkaiden tai tallennetun tiedon kannalta. Yritysten kannalta on tärkeä ymmärtää datan yksityisyyttä ja siihen liittyviä periaatteita, jotta uhkiin on helpompi varautua. Datan yksityisyyden hallintaan vaikuttaa merkittävästi kolme tekijää, jotka ovat luottamus, saatavuuden hallinta ja datan salaus (Kuva *10*). Luottamus on keskeisessä roolissa etenkin asiakkaan näkökulmasta tietovuotojen uhkan vähentämisessä.

Asiakkaat huolestuvat tietovuoto mahdollisuuksista ja he voivat solmia yritysten välillä erilaisia luottamusstandardeja. Saatavuuden hallinta ja pilvipalvelut sisältävät ongelmia etenkin luvattomien henkilöiden pääsyssä dataan, mikäli saatavuuden hallintaa ei olla käsitelty huolellisesti. Datan saatavuuden hallinnassa on määriteltävä tarkasti henkilöt, joilla on pääsy dataan, mitkä tiedostot ovat saatavilla kenellekin ja mihin aikaan tietty data on saatavilla. Viimeiseksi datan vahvalla salauksella (encryption) varmistetaan datan yksityisyyden suoja ja, että asiakkaan tiedot ovat turvattuna. (Maglaras et al. 2022, 13)

4.3 Riskien merkityksen arviointi

Riskianalyysin perusteella kaikista tunnistetuista riskeistä tietoturvariskien voidaan todeta olevan kaikista merkittävimpiä vaikutuksiltaan yritysten toiminnan kannalta. Raguseo (2018) päätyi niin ikään samaan lopputulokseen hänen tekemänsä empiirisen tutkimuksen perusteella. Hänen tutkimuksensa perusteella yksityisyys- ja turvallisuusongelmat olivat kaikista merkittävimpiä valmistavan teollisuuden yritysten tunnistamia riskejä big data teknologioiden käytössä. Tutkimuksen mukaan tietoturvariskien jälkeen seuraavaksi merkittävimpiä riskejä olivat puute informaatiojärjestelmistä ja infrastruktuurin tuesta, pääoma investoinnit ilman taetta tuotoista, vähäinen tietotekninen kokemus ja tekninen epävarmuus. Tutkimukseen oli osallistunut yhteensä 200 ranskalaista yritystä, joista valtaosa (86,5 %) oli PK-yrityksiä. (Raguseo 2018)

Yleisellä tasolla voidaan todeta big datan hyödyntämiseen liittyvien riskien olevan merkittäviä, luoden huomattavia haasteita yrityksille big datan tuomien hyötyjen saavuttamiseksi. Varsinkin vakavimmat riskit voivat olla realisoituessaan kriittisiä koko yrityksen liiketoiminnan kannalta. Riskien jatkokäsittelyn osalta yritysten tulee suhteuttaa tunnistetut riskit heidän omaan riskinkantokykyyn, minkä perusteella voidaan tehdä päätelmät ovatko riskit hyväksyttäviä vai eivät.

5 Big datan keräysprosessin haasteet ja riskit suomalaisessa yrityksessä

Lyhyessä case-osuudessa tarkastellaan erään suomalaisen PK-yrityksen tunnistamia haasteita ja riskejä big datan keräämisprosessissa. Case-yritys on uusi ja kasvava toimija valmistavassa teollisuudessa, jonka tuotteita hyödynnetään etenkin tekstiiliteollisuudessa. Datan kerääminen ja hyödyntäminen on ollut merkittävässä roolissa yrityksen tutkimus- ja kehitystoiminnassa, jonka pohjalle yrityksen tuotantoteknologia rakentuu. Case-osuudessa pyritään selvittämään minkälaista dataa yritys kerää tänä päivänä, kuinka dataa varastoidaan ja hyödynnetään ja millaisia haasteita ja riskejä yritys on tunnistanut. Case-yrityksestä haastateltavana oli tietoratkaisujen päällikkö.

Yrityksen keräämä ja hyödyntämä data koostuu pitkälti prosessidatasta. Prosessidataa kerätään erilaisista tuotannon kenttälaitteista ja instrumenteista automaation kautta sekä laboratoriolaitteista. Tuotannosta datan kerääminen tapahtuu erilaisten sensorien sekä IoT-laitteiden avulla. Tuotannosta kerätty data on osittain strukturoitua eli semistrukturoitua ja sitä tallennetaan yrityksen aikasarjatietokantaan. Laboratoriolaitteista kerätty data on strukturoitua ja sitä tallennetaan puolestaan relaatiotietokantaan. Strukturoimattoman datan käyttö on puolestaan vähäistä, eikä sitä erikseen kerätä. Strukturoimattoman datan muodostumisen ehkäisyksi koneiden ja ihmisten rajapintaan on kehitetty käyttöliittymiä, jotka muokkaavat esimerkiksi ihmisten interaktioista syntyvän datan suoraan haluttuun muotoon.

Datan varastoinnissa yritys hyödyntää pilveä sekä erilaisia on-premise ratkaisuja. On-premise ratkaisu tarkoittaa, että data on tallessa niin sanotusti talon sisällä eli riippumattomana internetistä. Yrityksen aikasarja- ja relaatiotietokannat ovat kuitenkin säilössä pilvessä. Toistaiseksi datan varastointi ei ole vielä tuottanut yrityksellä ongelmia, sillä datan määrä on pysynyt hallittavissa. Kuitenkin tulevaisuudessa teknologian skaalautuessa ja uusien tuotantolaitoksien myötä hallittavat datamäärät tulevat moninkertaistumaan ja varastointi ratkaisuihin on kiinnitettävä enemmän huomiota.

Datanelinkaaren vaiheista merkittävimmät haasteet tähän mennessä ovat syntyneet elinkaaren alkupäässä eli datan keräysvaiheessa. Datan keräämisen määrittelyyn sisältyy paljon epävarmuutta, sillä yritys on vielä uusi toimija sekä heidän teknologiansa poikkeaa toimialan

valtavirrasta. Datan keräykselle ei siis ole valmista tietomallia tai viitekehystä. Oman tietomallin kehitys on vaatinut yritykseltä paljon resursseja ja sen tarkka määrittely yrityksen toiminnan alkuvaiheessa on kriittisessä asemassa tulevaisuuden kasvun kannalta. Datan keräyksen mallinnuksessa on tärkeä huomioida nykyisen tuotannon lisäksi tulevaisuuden suunnitelmat, jotta olemassa olevaa tietomallia voitaisiin hyödyntää skaalattuna myös tulevissa tuotantolaitoksissa. Tämän takia mallin määrittelyssä on huomioitava mahdolliset riskit, joita puutteellisen mallin käyttämisestä voi realisoitua tulevaisuudessa. Esimerkiksi rakennettaessa uutta aikaisempaa suurempaa tuotantolaitosta, datan keräysprosessiin syntyy uusia muuttuvia tekijöitä, jotka voivat aiheuttaa merkittäviäkin ongelmia. Mikäli uusiin ongelmakohtiin ei olla varauduttu, voivat ne aiheuttaa suuria taloudellisia tappioita.

Datan keräykseen liittyvät haasteet koskevat lähinnä kerättävän datan määrittelyä eli mitä kaikkea dataa on syytä kerätä ja missä vaiheessa datan hallinta on otettava mukaan prosessiin. Datan hallinnalla tarkoitetaan datan keräämistä yhteisen standardin mukaan. Datan hallintaa pyritään tuomaan mukaan mahdollisimman aikaisessa vaiheessa, mutta kuitenkin liian tarkka hallinta saattaa johtaa siihen, että oikeita arvoja rajautuu pois keräyksestä. Datan keräyksessä tapahtuneet virheet ja huolimattomuudet kantautuvat suoraan datan analysointi- ja raportointivaiheeseen saakka. Puutteellisen datan pohjalta tehdyt päätökset ovat epätarkkoja ja pahimmassa tapauksessa datan jatkojalostus voi olla täysin mahdotonta.

Yrityksen datan keräysprosessi kokonaisuudessaan koostuu useasta eri kerroksesta. Ensimmäisessä eli alimmassa kerroksessa on automaatio eli sensorit, jotka keräävät dataa tuotannosta. Toisessa kerroksessa tuotannon sensorit yhdistetään OPC UA -serverille, jotta data saadaan kerättyä talteen. Kolmannessa kerroksessa serveriltä otetaan yhteys automaatioon ja täten data saadaan säilöttyä tietovarastoon. Tietovarastoon voidaan olla yhteydessä eri paikoista ja dataa pystytään jalostamaan eteenpäin, jolloin puhutaan neljännestä kerroksesta. Kerroksellisuuden haasteena on tiedonkulun varmistaminen, eli kuinka voidaan varmistua, että kaikki kerrokset päivittyvät, kun jonnekin tehdään muutoksia ja kuinka metadata säilyy hypittäessä kerrosten välillä. Lisäksi kerrosten välillä työskentely edellyttää useita eri yhteyshenkilöitä ja sidosryhmiä, mikä hidastaa toimintaa ja korostaa mahdollisuutta inhimillisiin virheisiin.

Yrityksen tunnistamista riskeistä yleisin riskitekijä on puutteellinen tai epätarkka data, joka ilmenee väärinä johtopäätöksinä. Jos data ei päivity eri kerrosten välillä tai sitä tarkastellaan suoraan väärällä tasolla, syntyy virheellisiä arvioita helposti. Esimerkiksi laitteiden

huoltotarpeita ei välttämättä huomioida ajoissa epätarkkuuksien takia, mikä voi ilmetä ongelmina tuotannossa kuten laiterikkoina tai heikompana laatuna. Laatuongelmat voivat aiheuttaa taloudellisia menetyksiä. Lisäksi erilaiset tietoturvariskit ovat merkittäviä datan käsittelyn yhteydessä, ja yritys onkin panostanut etenkin kyberturvallisuuteen riskienhallinnassaan. Yrityksen liiketoiminnan kannalta vakavin riski olisi tuotantoteknologiaan liittyvien immateriaalioikeuksien vuotaminen, minkä seurauksena yritys voisi menettää kilpailuetunsa markkinoilla. Tietoturva ja kyberturvallisuus ovat siis merkittävässä asemassa vakavien riskien myötä. Yritys on muun muassa hyödyntänyt tuotantolaitosten automaatioverkkojen suunnittelussa teollisen kyberturvallisuuden standardia minimoidakseen riskitekijöitä.

6 Johtopäätökset

Tämän kandidaatintyön tarkoituksena oli selvittää millaisia riskejä big datan hyödyntämiseen liittyy valmistavan teollisuuden yritysten näkökulmasta. Ennen kuin varsinaiseen aiheeseen pureuduttiin, työssä käsiteltiin ensiksi teoriataustaa aiheen pohjaksi muutamalla kappaleella. Big datan teoriakappaleilla johdateltiin lukijalle yleinen ymmärrys big datasta ja etenkin sen roolista valmistavan teollisuuden digitaalisessa murroksessa. Riskienhallinnan teoriakappaleissa puolestaan käsiteltiin tiiviisti yrityksen riskienhallinnan merkitys osana yrityksen liiketoimintaa sekä luotiin pohja riskienarviointiprosessille. Koko työn pohjalta voidaan esittää johdannossa esitettyihin tutkimuskysymyksiin seuraavat vastaukset:

”Mitkä ovat merkittävimpiä big datan hyödyntämiseen liittyviä riskejä?”

Big data teknologioiden käyttöönotto vaatii yrityksiltä merkittäviä resursseja päästäkseen käsikseen niiden tuomiin potentiaalisiin hyötyihin. Käyttöönottoa voidaan pitää siis yhtenä isona investointiriskinä, joka sisältää mahdollisuutensa ja samalla riskinsä. Big datan hyödyntämiseen liittyy monenlaisia eri riskejä, joita esiintyy kaikissa eri datan elinkaaren vaiheissa. Varsinkin turvallisuuteen on kiinnitettävä erityshuomioita koko elinkaaren ajan. Kuitenkin työssä havaittiin, että elinkaaren alkupään eli datan keräykseen liittyvät riskit ovat merkittävimpiä, sillä niiden realisoituessa syntyvät ongelmat kulkeutuvat myös seuraaviin elinkaaren vaiheisiin.

Big datan tuomista riskeistä voidaan tunnistaa kolme merkittävintä riskitekijää, jotka ovat datan epävarmuus, laadunhallinta ja tietoturvallisuus. Nämä riskitekijät luovat suurenosan kaikista tunnistetuista riskeistä. Big datan hyödyntämiseen liittyvää epävarmuutta esiintyy niin ikään kaikissa datan elinkaaren vaiheissa. Merkittäviä epävarmuuden luomia riskejä ovat esimerkiksi kerättävän datan määrittely, potentiaalisten hyötyjen ja kulujen arviointi, tekninen epävarmuus uusista teknologioista ja johtopäätösten tarkkuus. Datan laadunhallinnan riskit liittyvät datan puutteellisuuteen ja etenkin epätarkan datan perusteella tehtyihin päätöksiin. Big data sisältää sen määritelmän mukaisesti useita haastavia ominaisuuksia, jotka on kaikki huomioitava datan laadunhallinnassa. Merkittävimmät riskit laadunhallinnan kannalta ovat datan jäljitettävyys, ajankohtaisuus, määrä, monimuotoisuus ja vaihtelevuus.

Big datan haavoittuvaisuus ja sen yhteydessä käytetyt pilvipalvelut nostavat tietoturvallisuuden kaikista merkittävimmäksi riskitekijäksi. Tietoturvariskejä on monenlaisia ja ne ovat läsnä koko ajan. Suurimpia tietoturvariskejä ovat erilaiset kyberhyökkäykset, tietovuodot, palvelunestohyökkäykset, APT-hyökkäykset sekä riskit yksityisyyden suojassa.

”Millaisia vaikutuksia tunnistetuilla riskeillä on?”

Lähes kaikki tunnistetuista riskeistä ovat joko vahinko- tai epävarmuusriskejä eli ne aiheuttavat jollain tapaa negatiivisia vaikutuksia yrityksille. Kuitenkin riskien kohteilla ja merkittävyyksillä on selviä eroja. Suurin osa riskeistä aiheuttaa realisoituessaan yrityksille taloudellisia menetyksiä, ongelmia tuotantoon tai johtavat resurssien epätarkkaan kohdentamiseen. Suurimpien riskien toteutuessa myös vaikutukset voivat olla kuitenkin koko yrityksen liiketoiminnan kannalta kriittiset. Esimerkiksi vakavien tietoturvariskien myötä voi seurauksena olla yritykselle tärkeiden immateriaalioikeuksien ja asiakkaiden tietojen vuotaminen. Nämä esimerkit voivat johtaa yrityksen toiminnan kannalta kriittisiin mainehaittoihin ja jopa kilpailukyvyn menettämiseen markkinoilla.

”Kuinka valmistavan teollisuuden yritykset hyödyntävät teollista big dataa?”

Osatutkimuskysymyksen tarkoituksena oli luoda lukijalle ymmärrys teollisen big datan hyödyistä eli minkä takia yritykset ovat valmiina ottamaan siihen liittyviä riskejä. Valmistavan teollisuuden yritykset keräävät dataa monipuolisesti tuotteiden elinkaaren eri vaiheista ja erilaisista lähteistä. Dataa kerätään etenkin yritysten omista järjestelmistä ja erilaisilla IoT-laitteilla ja sensoreilla tuotannosta. Analysoimalla suuria datamassoja yritykset kykenevät tekemään entistä järkevämpiä päätöksiä suuremmalla skaalalla. Positiivisia tuloksia on havaittu jo lähes kaikilla tuotannon arvoketjun eri osa-alueilla.

Yleisellä tasolla big data -analytiikka auttaa yrityksiä tuotantojärjestelmien kehittämisessä ja optimoinnissa ja etenkin hieman vanhahtavissa tuotantoympäristöissä saatu hyöty on merkittävä työtehokkuuden kannalta. Big data -analytiikan menetelmien avulla laadittuja ennusteita kyetään hyödyntämään myös toimitusketjun hallinnan tukena, esimerkiksi kysynnän suunnittelussa, hankintatoimessa, varaston hallinnassa ja logistiikassa. Big datan hyväksikäyttö edistää siis kokonaisvaltaisesti koko yrityksen toiminnan tehokkuuden johtamista.

Tässä työssä keskityttiin vain big datan hyödyntämiseen liittyvien riskien tunnistamiseen ja analysointiin. Kiinnostava jatkotutkimusaihe työlle, olisi kokonaisvaltaisen riskienhallintaprosessin toteuttaminen, eli lisäyksenä vielä varsinainen riskien käsittely ja riskien

seuraaminen. Jatkotutkimuksen toteuttamiseksi tutkimus tulisi suorittaa suoraan tietyssä yrityksessä, jolloin riskien tarkat vaikutukset kyettäisiin laskemaan. Lisäksi kiinnostavaa olisi tutkia big datan riskejä myös muilla toimialoilla ja selvittää onko merkittävimpien riskitekijöiden joukossa eroavaisuuksia.

Lähteet

Alcácer, V., Cruz-Machado, V., 2019. Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. Engineering Science and Technology, an International Journal, Vol. 22, s. 899–919.

Anbar, M., Abdullah, N., Manickam, S., 2021. Towards Understanding the Challenges of Data Remanence in Cloud Computing: A Review, in: Advances in Cyber Security, Communications in Computer and Information Science. Springer Singapore Pte. Limited, Singapore, s. 495–507.

Bi, Z., Jin, Y., Maropoulos, P., Zhang, W.-J., Wang, L., 2021. Internet of things (IoT) and big data analytics (BDA) for digital manufacturing (DM). International Journal of Production Research.

Chen, P., Desmet, L., Huygens, C., 2014. A Study on Advanced Persistent Threats, in: De Decker, B., Zúquete, A. (Eds.), Communications and Multimedia Security, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, s. 63–72.

Chhabra, G.S., Singh, V.P., Singh, M., 2020. Cyber forensics framework for big data analytics in IoT environment using machine learning. Multimed Tools Appl, Vol. 79, s. 15881–15900.

Choo, K.-K.R., Dehghantanha, A., 2020. Handbook of Big Data Privacy. Springer International Publishing, Cham.

Clarke, R., 2016. Big data, big risks. Information Systems Journal, Vol. 26, s. 77–90.

de Oliveira, U.R., Marins, F.A.S., Rocha, H.M., Salomon, V.A.P., 2017. The ISO 31000 standard in supply chain risk management. Journal of Cleaner Production, Vol. 151, s. 616–633.

Dialani, P., 2020. The Future of Data Revolution will be Unstructured Data. Analytics Insight. [WWW-dokumentti]. [Viitattu 28.2.2023]. Saatavilla: https://www.analyticsinsight.net/the-future-of-data-revolution-will-be-unstructured-data/

Fisher, D., DeLine, R., Czerwinski, M., Drucker, S., 2012. Interactions with big data analytics. Interactions, Vol. 19, s. 50–59.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, Vol. 35, s. 137–144.

Hariri, R.H., Fredericks, E.M., Bowers, K.M., 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data, Vol. 6, s. 44.

Hassanien, A.E., Darwish, A., 2021. Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges, Studies in Big Data. Springer International Publishing, Cham.

Hopkin, P., 2018. Fundamentals of Risk Management: Evaluating and Implementing Effective Risk Management. 5th Edition. The institute of Risk Management, London

Hopkin, P., 2014. Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management. Third edition. Kogan Page, London

Hunziker, S., 2021. Enterprise Risk Management: Modern Approaches to Balancing Risk and Reward. Springer Fachmedien, Wiesbaden.

Ikegwu, A.C., Nweke, H.F., Anikwe, C.V., Alo, U.R., Okonkwo, O.R., 2022. Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. Cluster Comput, Vol. 25, s. 3343–3387.

Kamensky, M., 2014. Strateginen johtaminen: menestyksen timantti, 4., tarkistettu painos. ed. Talentum, Helsinki.

Kitchin, R., McArdle, G., 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, Vol. 3, No. 1.

Koo, J., Kang, G., Kim, Y.-G., 2020. Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges. Sustainability, Vol 12, 10571.

Li, C., Chen, Y., Shang, Y., 2022. A review of industrial big data for decision making in intelligent manufacturing. Engineering Science and Technology, an International Journal, Vol. 29, 101021.

Li, Z., Xu, W., Shi, H., Zhang, Y., Yan, Y., 2021. Security and Privacy Risk Assessment of Energy Big Data in Cloud Environment. Computational Intelligence and Neuroscience, Vol. 2021, 2398460.

Lv, Z., Song, H., Basanta-Val, P., Steed, A., Jo, M., 2017. Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. IEEE Transactions on Industrial Informatics, Vol. 13, s. 1891–1899.

Maglaras, L., Janickle, H., Amine Ferrag, M., 2022. Cybersecurity of Critical Infrastructures: Challenges and Solutions. Sensors, Vol. 22, 5105.

Malik, V., Singh, S., 2019. Cloud, Big Data & IoT: Risk Management. International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), s. 258–262.

McAfee, A., Brynjolfsson, E., 2012. Big Data: The Management Revolution. Harvard business review, Vol. 90, s. 6-60, 68, 128.

Microsoft, 2023. Mitä tietojen mallinnus tarkoittaa? [WWW-dokumentti]. [Viitattu 10.4.2023]. Saatavilla: https://powerbi.microsoft.com/fi-fi/what-is-data-modeling/

Mourtzis, D., Vlachou, E., Milas, N., 2016. Industrial Big Data as a Result of IoT Adoption in Manufacturing. Procedia CIRP, 5th CIRP Global Web Conference - Research and Innovation for Future Production (CIRPe 2016), Vol. 55, s. 290–295.

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data, Vol. 2, No. 1.

Raguseo, E., 2018. Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. International Journal of Information Management, Vol. 38, s. 187–195.

Rahul, K., Banyal, R.K., 2020. Data Life Cycle Management in Big Data Analytics. Procedia Computer Science, International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020, Vol. 173, s. 364–371.

Ramya, S., Sakthi Devi, R., Senthil Pandian, P., Suguna, G., Suganya, R., Manimozhi, N., 2023. Analyzing Big Data challenges and security issues in data privacy. International Research Journal of Modernization in Engineering Technology and Science, Vol 5, s. 421-428.

Riskikompassi, 2023. ISO 31000 peruselementit. [WWW-dokumentti]. [Viitattu 25.4.2023]. Saatavilla: https://riskikompassi.fi/johtaminen/viitekehyksia/iso-31000-peruselementit/

Runkler, T.A., 2020. Data Analytics: Models and Algorithms for Intelligent Data Analysis. Springer Fachmedien Wiesbaden, Wiesbaden.

Schintler, L.A., McNeely, C.L., 2022. Encyclopedia of Big Data. Springer International Publishing, Cham.

SFS-ISO, 2018. SFS-ISO 31000:2018 Riskienhallinta. Ohjeet. 2. painos. Suomen Standardisoimisliitto SFS ry.

Sheehan, B., Murphy, F., Mullins, M., Ryan, C., 2019. Connected and autonomous vehicles: A cyber-risk classification framework. Transportation Research Part A: Policy and Practice, Vol 124, s. 523–536.

Shukla, N., Tiwari, M.K., Beydoun, G., 2019. Next generation smart manufacturing and service systems using big data analytics. Computers & Industrial Engineering, Vol. 128, s. 905–910.

Taylor, P., 2022. Total data volume worldwide 2010-2025. Statista. [WWW-dokumentti]. [Viitattu 27.2.2023]. Saatavilla: https://www.statista.com/statistics/871513/world-wide-data-created/

Tsai, C.-W., Lai, C.-F., Chao, H.-C., Vasilakos, A.V., 2015. Big data analytics: a survey. Journal of Big Data, Vol 2, No. 21.

VTT, 2023. Valmistava teollisuus. VTT. [WWW-dokumentti]. [Viitattu 15.3.2023]. Saatavilla: https://www.vttresearch.com/fi/toimialat/valmistava-teollisuus

Wang, G., Gunasekaran, A., Ngai, E.W.T., Papadopoulos, T., 2016. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, Vol. 176, s. 98–110.

Wang, J., Zhang, W., Shi, Y., Duan, S., Liu, J., 2018. Industrial Big Data Analytics: Challenges, Methodologies, and Applications.