

Data-analytiikan käyttömahdollisuudet tämän päivän työelämässä

Henrik Manninen

Tekijä(t) Henrik Manninen	
Koulutusohjelma Liiketalouden koulutusohjelma	
Opinnäytetyön nimi Data-analytiikan käyttömahdollisuudet tämän päivän työelämässä	Sivu- ja liitesivumäärä 34 + 0
<p>Opinnäytetyön tarkoituksena oli käydä läpi dataa, data-analytiikkaa ja miten data-analytiikkaa hyödynnetään tämän päivän työelämässä. Työ alkaa tietoperustan läpikäymisellä, jossa käymme läpi aiheen ymmärtämisen kannalta tärkeää teoriaa ja taustatietoa. Selvitämme mitä on data, data-analytiikka sekä minkälaisia ominaisuuksia näillä aiheilla on.</p> <p>Työn tutkimusosa toteutettiin case-tutkimuksena. Lähdin case-tutkimuksen avulla hakemaan vastausta pääkysymykseen, miten data-analytiikkaa hyödynnetään työelämässä. Valitsin neljä yritystä eri aloilta, jotta olisi mahdollista saada mahdollisimman laaja käsitys data-analytiikan monikäyttöisyydestä. Yritykset, jotka valikoituivat työhön toimivat logistiikan, hissitoiminnan, lentoliikenteen sekä lääketieteen parissa.</p> <p>Materiaali opinnäytetyöhön kerääntyi alan kirjallisuudesta, artikkeleista sekä alan asiantuntijoiden blogikirjoituksista.</p> <p>Sain tämän työn avulla haluamani vastauksen kysymykseen, minkälainen rooli data-analytiikalla on ja miten sitä tänä päivänä hyödynnetään erilaisissa yrityksissä. Aiheeseen aikaisemmin perehtymättömänä, opin paljon uutta mielenkiintoisesta aiheesta ja opin käyttämään tätä informaatiota tutkimuksessa.</p> <p>Tutkimalla esimerkkiryhmiä ja niiden tapoja käyttää data-analytiikkaa, tällä työllä saatiin vastauksia siihen, mitä lähdettiin hakemaan.</p>	
Asiasanat Data, big data, data-analytiikka	

Sisällys

1	Johdanto	1
1.1	Tutkimuskysymykset ja menetelmä	1
1.2	Opinnäytetyön rakenne	2
2	Data ja Big Data	3
2.1	Data	3
2.2	Big data	4
2.2.1	Big datan käyttö	5
2.2.2	Big data ja COVID-19	7
2.2.3	Big datan haasteet ja riskit	7
3	Data-analytiikka	9
3.1	Koneoppiminen ja tekoäly	10
3.2	Analytiikka ennen	11
3.3	Data-analytiikka tänään	12
3.3.1	Isot vs. pienet yritykset	14
3.4	Miltä näyttää tulevaisuus?	16
3.5	Internet of Things	16
3.6	Data-analytiikka alustat	17
3.6.1	SAS Advanced Analytics	18
3.6.2	Microsoft Power BI	19
4	Tutkimusmenetelmät	20
5	Data-analytiikan käyttö yrityksissä	21
5.1	Volvo ja Mack kuorma-autot	21
5.2	Data-analytiikka ja lentoliikenne – Heathrow'n lentoasema	22
5.3	KONE	24
5.4	Data-analytiikka taistelussa syöpää vastaan	25
6	Yhteenveto	27
7	Pohdinta	29

1 Johdanto

Opinnäytetyö. Se on ollut asia, jota olen enemmän tai vähemmän kammoksunut siitä asti, kun ensimmäisessä moduulissa käytiin läpi raporttipohjaa, johon se pitää noin kolmen vuoden päästä kirjoittaa. Miten opiskelija, jonka suurin ongelma läpi henkilökohtaisen opiskeluhistorian on ollut tekstin tuottaminen pitkiä raportteja varten, selviytyy urakasta ilman kyyneleitä ja kuhmuja? Vastaus tähän on aihe. Aihe, joka kiinnostaa tarpeeksi, että saa puskettua itsensä tekemään työtä eteenpäin. Minulle vastaus tähän kysymykseen löytyi kolmannen vuoden opintojen aikana: data-analytiikka.

Data-analytiikalla tarkoitetaan prosessia, menetelmää tai menetelmiä, jossa olemassa oleva data tai informaatio muutetaan hyödyllisempään ja helposti luettavampaan muotoon (AWAcademy 2021). Tätä dataa analysoidaan erilaisilla data-analytiikan ohjelmilla ja tilastollisilla menetelmillä, jonka avulla informaatiosta tehdään asiaan sopivia graafeja, malleja ja ennustuksia (Sedkaoui 2018, 44–45).

Ilman minkäänlaista aikaisempaa käytännön kokemusta data-analytiikasta, aihe oli kiinnostava alusta asti. Ison numeerisen datan muuttaminen helposti luettavaan muotoon oli aluksi haastavaa, mutta tuntui palkitsevalta ja melkein jopa terapeuttiselta. Juuri data-analytiikan kurssin aikana tämä aihe alkoi kiinnostaa enemmän ja mietin aiheen valitsemista myös opinnäytetyöni aiheeksi.

Tämän opinnäytetyön tarkoituksena on tutustua tarkemmin data-analytiikkaan, dataan, ajankohtaisiin käsitteisiin, terminologiaan, historiaan, tulevaisuuden trendeihin ja käytännön käyttötarkoituksiin erilaisin esimerkein. Aiheen kiinnostavuuden lisäksi aihe on mielestäni ajankohtainen muun muassa viimeaikaisten uutisten perusteella. Esimerkiksi vallitsevan maailmantilanteen vuoksi Suomessa kehitetty Koronavilkku -sovellus hyödyntää data-analytiikkaa (Terveyden ja hyvinvoinnin laitos 2020).

1.1 Tutkimuskysymykset ja menetelmä

Pääkysymykset, joihin lähden tällä opinnäytetyöllä vastaamaan, ovat:

- Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksissä?
- Mitä data-analytiikan ohjelmistoa yrityksissä käytetään?

Lähden hakemaan vastauksia näihin kysymyksiin tutkimalla neljää eri yritystä ja sitä, miten ne käyttävät data-analytiikkaa päivittäisessä liiketoiminnassaan. Käyn käytännön esimerkkejä läpi case-periaatteella. Opinnäytetyö pitää sisällään neljä erilaista casea eri aloilta, joiden avulla saamme lisää informaatiota data-analytiikan monipuolisuudesta ja sen moninaisista käyttötarkoituksista.

1.2 Opinnäytetyön rakenne

Opinnäytetyö lähtee liikkeelle datan ja data-analytiikan määritelmistä. Mitä ne itseasiassa tarkoittavat? Käsitteet voivat kuulostaa tutuilta, mutta mitä ne oikeasti ovat? Käyn läpi, minkälainen vaikutus ja rooli datalla ja data-analytiikalla on ollut ja miten ne ovat muuttaneet jokaisen meistä elämää. Tutkin, onko muutos vaikuttanut erilaisiin yrityksiin niiden koosta riippumatta, minkälaisia riskejä alalla on havaittavissa ja minkälaisia trendejä on odotettavissa data-analytiikan osalta. Lähden hakemaan vastauksia näihin kysymyksiin lukemalla alan kirjallisuutta, artikkeleita ja alan ammattilaisten blogikirjoituksia sekä kuuntelemalla heidän kommenttejaan aiheista. Aihe on erittäin ajankohtainen, joten kirjallisuutta ja materiaalia on hyvin tarjolla aiheesta kiinnostuneille.

Työn loppuun kerron omat johtopäätökseni ja pohdintani aiheesta ja mitä koko urakasta jäi käteen. Pohdin, sainko omasta mielestäni vastauksen niihin kysymyksiin, mihin lähdin vastauksia hakemaan. Avaan myös sitä, oliko tutkimus omasta mielestäni onnistunut ja käyn läpi erilaisia haasteita, mitä työn tekeminen aiheutti.

2 Data ja Big Data

Digitaalinen data ja informaatio ympäröivät meitä. Jokainen meistä tuottaa dataa päivittäisellä toiminnallaan ja internetkäyttötymisellään. Tekemällä sitten ostoksia internetissä, pelaamalla videopelejä, sijoittamalla osakkeisiin tai ihan vain kantamalla mukanaan älypuhelinta GPS-ominaisuudella, tuottaa ihminen dataa yritysten käytettäväksi. Itse datan määrä on räjähtänyt kuluneiden vuosien aikana. Jos perehtyy aiheeseen ja lukee artikkeleita datasta, vastaan tulee nopeasti lause ”90 prosenttia kaikesta maailman datasta on syntynyt viimeisen kahden vuoden ajalta”. Tämä luku on alun perin vuonna 2011 julkaisusta artikkelista IBM Systems Journalista. Toteamus antaa hyvän kuvan siitä, kuinka suhteellisen uudesta asiasta vielä puhutaan.

Datan ja informaation määrä ja rooli ovat kasvaneet kuluneiden vuosien aikana, eikä hidastumisen merkkejä ole vielä havaittavissa. Tämä informaation kasvu on entisestään vaikeuttanut yritysten päätöksentekoa. Informaation ja datan määrän nopean kasvun takia organisaatioiden on pitänyt kehittyä mukana. Käyttöön on otettu erilaisia analytiikkatyökaluja, jotka ovat helpottaneet keskittymistä olennaiseen. Tässä kappaleessa tutustumme dataan, data-analytiikkaan ja pelottavalta kuulostavaan ”big dataan”.

2.1 Data

Mitä on data? Datalla yleensä tarkoitetaan faktoja, havaintoja, mittaustuloksia ja raakaa informaatiota eli raakadataa. Data antaa lisäinformaatiota yksilöstä, asiasta tai havainnosta. (Sedkaoui 2018, 4.) Data itsenään ei vielä hyödytä organisaatiota päätöksenteossa. Yrityksellä voi esimerkiksi olla bittikaupalla dataa, mutta jos sitä ei osaa prosessoida oikein, sen hyöty jää vielä minimaaliseksi.

Datan määrä on kasvanut eksponentiaalisesti kuluneiden vuosien aikana ja tämän datan määrän kasvun on mahdollistanut digitaalinen kehitys (Sedkaoui 2018, 7–8). Monet isot yritykset ovat raportoineet, että heidän hallussaan oleva datan määrä jatkaa kasvuaan kasvavalla vauhdilla (Sedkaoui 2018, 8). On arvioitu, että datan määrä tuplaantuu kahden vuoden välein ja vuonna 2018 Sedkaoui (2018, 6–7) arvioi, että vuonna 2020 datan määrä olisi 44 000 miljardia gigabittiä, tämä on yli 10 kertaa enemmän kuin vuonna 2013. Pelkästään vuoden 2011 aikana syntyi enemmän dataa kuin koko kuluneen ihmiskunnan historian aikana. 2011 julkaistussa IBM-artikkelissa sanotaan, että maailmassa syntyy päivittäin 2,5 kvintiljoonaa tavua päivässä. Viimeisen kymmenen vuoden aikana datan ja informaation määrä on kasvanut huomattavasti ja nopeasti. (Sedkaoui 2018, 8.)

Kun puhutaan luvuista näin isossa mittakaavassa, voi olla vaikea käsittää minkälaisella skaalalla liikumme. Käytännössä tämä tarkoittaa, että datan määrä on kasvanut ja jatkaa kasvuaan eksponentiaalisesti. On arvioitu, että datan määrä maailmassa kymmenkertaistuu joka viides vuosi. Esimerkkinä voimme käyttää Facebookia: vuonna 2013 Facebook säilöi 100 petabittiä dataa ja vuonna 2014 tämä luku oli jo 300 petabittiä. Facebook triplasi vuosittaisen datan määrän yhden vuoden aikana vuonna 2014. (Sedkaoui 2018, 8.)

Aiheeseen perehtymättömälle puheet petabiteistä tai terabiteistä, ei välttämättä vielä anna tarvittavaa kuvaa aiheen laajuudesta ja skaalasta. Jotta näitä mittasuhteita olisi helpompi ymmärtää, olen tehnyt alle taulukon, joka antaa parempaa kuvaa tavujen mittasuhteista ja lyhenteistä. Tämän taulukon tarkoituksena on tuoda paremmin esille, kuinka suurista määristä dataa ja informaatiota puhumme.

Taulukko 1. Tavujen mittasuhteet, mukautettu Sedkaoui 2018

Nimi	Lyhenne	Arvo
1 tavu	-	8 bittiä
1 kilotavu	KB	10^3 tavua
1 megatavu	MB	10^6 tavua
1 gigatavu	GB	10^9 tavua
1 teratavu	TB	10^{12} tavua
1 petatavu	PB	10^{15} tavua
1 eksatavu	EB	10^{18} tavua
1 tsettatavu	ZB	10^{21} tavua
1 jottatavu	YB	10^{24} tavua

2.2 Big data

”Big data ei ole yksiselitteinen käsite” (Tilastokeskus 2018). Näin alkaa Tilastokeskuksen selitys big datasta. Big dataa syntyy koneiden välisessä kommunikaatiossa. Esimerkiksi dataa, jota syntyy erilaisissa tuotantoprosesseissa tai kun käytetään sosiaalista mediaa, kutsutaan big dataksi. Määriteltäessä big dataa Tilastokeskus tuo esiin kolme big datan tyyppillistä ominaisuutta. Ensimmäisenä, niin kuin nimestäkin voi jo päätellä, on datan hyvin suuri määrä. Toisena ominaisuutena on datan muodon vaihtelu. Koska dataa tulee monesta eri lähteestä todella isoja määriä, se ei aina ole samassa muodossa, vaan se voi olla rakenteellista tai ei-rakenteellista. Tämä tarkoittaa, että datan muoto voi vaihdella

esimerkiksi lokitiedoista kuviin. Kolmantena ominaisuutena on datan nopea saatavuus, kertyminen ja muuttuminen. (Tilastokeskus 2018.)

Tiedon Big Data -liiketoiminnan kehittämisestä vastaava johtaja kirjoitti blogikirjoituksen Talouselämä -lehteen, jossa hän mainitsi big datan muuttavan maailmaa pysyvästi. Vuonna 2013 ennustettiin, että vain kolmen vuoden kuluessa, vuoteen 2016 mennessä, puolet kaikesta sähköisestä tiedosta olisi liitoksissa big dataan. Jo vuonna 2013 asiantuntijat arvioivat, että big data on melkein 100 miljardin business. (Talouselämä 2013.) Vuonna 2020 julkaistussa ”market research report” raportissa, big datan markkinoiden arvioitiin olevan 138,9 miljardin dollarin arvoinen ja tämän luvun arvioidaan nousevan vuoteen 2025 mennessä 229,4 miljardiin dollariin (Markets And Markets 2020).

Big datasta tekee mielenkiintoista se, että dataa on ollut olemassa jo pitkään. Data itsessään ei siis ole uusi asia. On tärkeää huomata, minkälaisessa muodossa dataa on tuotettu viime aikoina. Big dataa tuotetaan ja kerätään automaattisesti ja digitaalisesti samaan aikaan. On myös tärkeää kiinnittää huomiota, että ennen big dataa ei ollut big data -teknologiaa. Tämä nopea kasvu on raivannut tietä niin sanottuun kolmen V:n malliin. Kolmen V:n mallia käytetään, kun kuvaillaan big dataa kolmen ominaisuuden avulla: määrä (volume), vauhti (velocity) ja moninaisuus (variety). Määrä tarkoittaa sananmukaisesti itse datan määrää tai kokoa, vauhdilla tarkoitetaan tiedonsiirtonopeutta ja kuinka nopeasti on tarpeellista reagoida ja moninaisuudella tarkoitetaan datan heterogeenisyyttä. (Sedkaoui 2018, 11.)

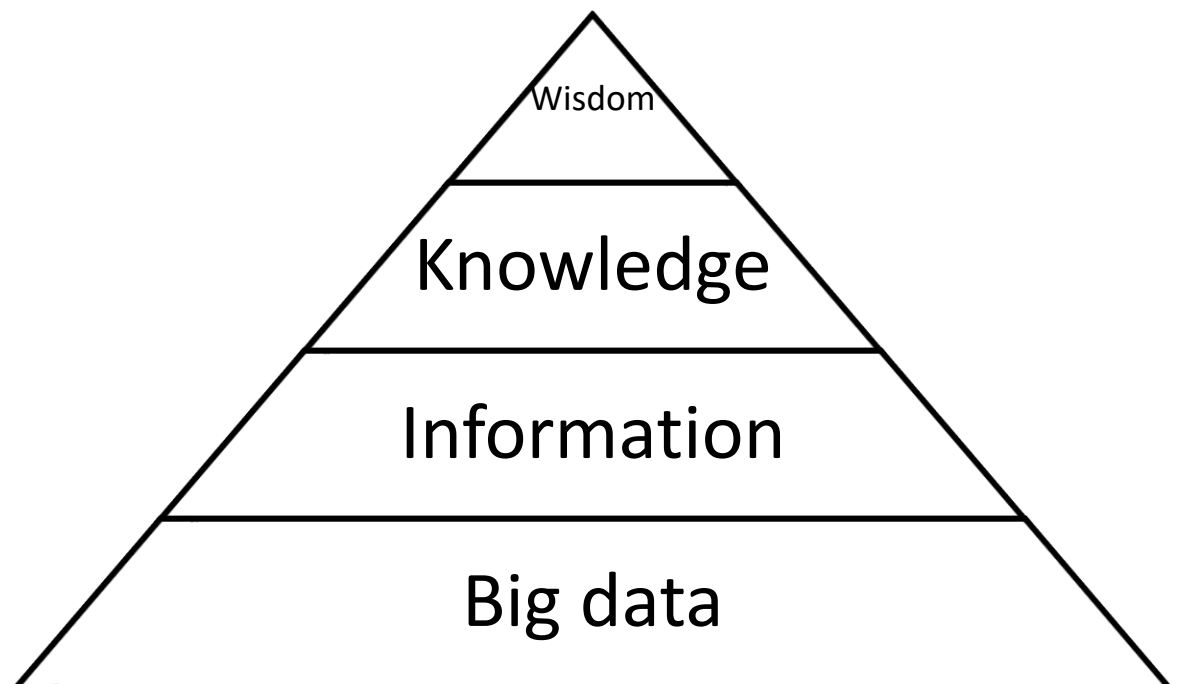
2.2.1 Big datan käyttö

Englantilainen filosofi Sir Francis Baconin on raportoitu sanoneen: ”Tieto on valtaa”. Big datan hyödyntäminen toimii pitkälti samalla periaatteella: suuri tiedon määrän kasvu mahdollistaa datan käytön täysin uudella tavalla, mikä ei ole ollut mahdollista aikaisemmin. Mitä enemmän tietoa henkilöllä on asiasta tai tilanteesta, sitä parempia valintoja hän osaa tehdä hetkessä ja sitä paremmin hän myös pystyy ennustamaan tulevaisuuden trendejä. Vertaamalla dataa, alkaa huomaamaan niiden välisiä syy-seuraussuhteita. Pystymme hallussa olevan datan avulla luomaan malleja ja simulaatioita, joiden avulla voimme mitata, minkälaisia vaikutuksia tekemillämme muutoksilla voi olla. Näiden työkalujen avulla yksittäinen toimija tai organisaatio voi tehdä omaa toimintaansa edistäviä päätöksiä tehokkaammin kuin koskaan aikaisemmin. (Marr 2020.)

Maailman vaikutusvaltaisimpiin businessvaikuttajiin kuuluva Bernard Marr tuo esille, että big data on avannut ovia muuallakin kuin pelkässä businessmaailmassa ja näistä hän nostaa esille kolme alaa. Terveystieteissä big data auttaa tunnistamaan sairauksia aikaisemmin analysoimalla potilaiden tietoja (kuten kuvia ja numerodataa). Toisena mainitaan luonnonkatastrofien ja ihmisten aiheuttamien katastrofien ennustaminen: voidaan esimerkiksi ennustaa, missä seuraava maanjäristys todennäköisimmin iskee. Kolmantena Marr mainitsee poliisityön ja rikosten estämisen. (Marr 2020.)

Kun puhutaan vaikeasti ymmärrettävästä sekä mittakaavaltaan isosta asiasta, käytännön esimerkit helpottavat asian ymmärtämistä. Jay Liebowitz (2013, 122–123) tuo kirjassaan esille käytännön esimerkkejä big datan käytöstä ja hyödyntämisestä. Yhtenä esimerkkinä on amerikkalainen vähittäistavarakauppajätti Walmart. Walmart käsittelee yli miljoona asiakastapahtumaa joka tunti. Tästä muodostuva data säilytetään Walmartin ”data-basessa” eli tietokannassa, joka sisältää yli 2,5 petatavua dataa. Tämä on yli 167 kertaa enemmän informaatiota, kun maailman suurimman kirjaston, USA:n kongressin kirjaston, kirjoissa yhteensä. (Liebowitz 2013, 122–123.)

Datan muuttaminen informaatioksi ja tämän informaation muuttaminen tiedoksi ja viisaudeksi on avaintekijä yrityksen menestymiselle (Sedkaoui 2018, 15). Tätä voidaan kuvata Russel Ackoffin ”Knowledge pyramid” (kuva 1) kuviolla (Sedkaoui 2018, 26).



Kuva 1. Knowledge pyramid, mukautettu Russell Ackoff 1989

Jotta kaikesta tästä datasta ja informaation määrästä saa kaiken irti, täytyy apuna käyttää tietokoneita ja analytiikkatyökaluja. Tietokoneet ovat ihmisiä nopeampia käsittelemään isoja määriä dataa ja tunnistamaan erilaisia trendejä ja yhteyksiä. Big datan hyödyntäminen kulkee siis vahvasti yhdessä data-analytiikan, koneoppimisen ja tekoälyn kanssa. Palaamme näihin käsitteisiin myöhemmin vielä tarkemmin luvussa kolme.

2.2.2 Big data ja COVID-19

Zoom-palaverit, kasvomaskin käyttö ja etätyöskentely ovat tulleet jokaiselle meistä tutuksi viimeisen vuoden aikana. Ympäri maapalloa levinnyt koronavirus on pakottanut maat suuriin toimenpiteisiin taistelussa pandemiaa vastaan. Big datan käyttö on ollut suuressa roolissa koronaviruksen selättämisessä ja informaatiota jaetaan maiden ja organisaatioiden välillä, jotta voidaan keskittyä siihen mikä on todistetusti toiminut. Big datan avulla on muun muassa voitu luoda reaaliaikaisia ennusteita viruksen leviämisestä sekä sen vaikutuksista. (Marr 2020.)

Hyvä käytännön esimerkki big datan käytöstä koronavirusta vastaan on Suomen Terveysten ja hyvinvoinnin laitoksen Koronavilkku -sovellus. Koronavilkku on mobiilisovellus, jossa ihmiset voivat nimettömästi ilmoittaa omasta tartunnasta ja saada ilmoituksia mahdollisista altistumisista. Sovellus antaa jokaiselle puhelimelle tunnuskoodin ja tätä koodia puhelimet jakavat toisilleen. Kun jollakin sovellusta käyttävällä todetaan koronavirustartunta ja hän raportoi tartuntansa sovelluksen kautta, saavat tämän henkilön (puhelimien) kanssa lähikontaktissa olleet henkilöt myös tiedon mahdollisesta altistumisesta sovelluksen kautta. Kaikki tämä on mahdollista datan ja sen käytön ansiosta. (Terveysten ja hyvinvoinnin laitos 2020.)

2.2.3 Big datan haasteet ja riskit

Kaikkien mahdollisuuksien rinnalla, big data tuo mukanaan myös haasteita ja riskejä. On tärkeää, että organisaatiot ja yksilöt osaavat toimia vastuullisesti ja ammattitaitoisesti, koska väärinkäytöksistä voi seurata vakavia taloudellisia ja rikosoikeudellisia seurauksia.

Niin kuin jo monta kertaa on tullut esille, käytössä olevaa dataa on paljon ja se on monimutkaista ja myös eri laatuista. Organisaation on osattava keskittyä heille ajankohtaiseen informaatioon. Soraya Sedkaoui (2018, 32) sanoo kirjassaan: ”Enemmän dataa ei aina tarkoita parempaa dataa”.

Alan ammattilaiset ovat pitkälti yhtä mieltä siitä, että suurimmat haasteet tulevat tiedon turvallisuuden ja yksityisyyden suojaamisen kanssa. Jokainen meistä tuottaa paljon informaatiota, jonka haluamme pitää yksityisenä. Bernard Marr (2020) kuvaa tätä artikkelissaan hyvin tasapainotelluksi: meidän kaikkien tulee löytää tasapaino sovellusten tuoman mukavuuden ja luovuttamamme tiedon määrän välille. Myös organisaatioiden on osattava käsitellä haltuunsa saamaa tietoa oikein ja lainmukaisesti (Marr 2020).

Tästä tiedon väärinkäytöstä on tullut esiin valitettavia esimerkkejä kuluneiden vuosien aikana. Esimerkiksi vuonna 2018 Facebook jäi kiinni, kun heidän epäonnistuneen datan käsittelyn takia, data-analytiikkayritys Cambridge Analytica pääsi käsiksi yli 50 miljoonan Facebook -käyttäjän tileihin. Loppujen lopuksi yli 87 miljoonaa Facebook profiilia joutui uhriksi. Facebookin perustaja Mark Zuckerberg joutui todistamaan kongressin edessä ja Facebook sai 5 miljardin dollarin sakon. (Ma & Gilbert, Business Insider 2019.)

Suomessa vastaava tapaus tuli julki marraskuussa 2020, kun yli 30 000 henkilön potilastietoja ja kertomuksia alettiin julkaisemaan hakkerin toimesta. Hakkeri käytti näitä tietoja kiristääkseen rahaa Psykoterapiakeskus Vastaamolta sekä sen asiakkailta. Oikeuden asiakirjoista selvisi, että Vastaamon potilaiden potilastiedot olivat olleet verkossa ilman asiaankuuluvaa suojausta yli vuoden, uutisoi YLE. (Hämäläinen, YLE 2021.)

Koska käsiteltävä data voi olla arkaluontoista ja sen määrä on kasvanut huomattavasti, turvallisuuteen kiinnitetään jatkuvasti enemmän huomiota. Ala muuttuu nopeasti ja toimijoiden täytyy pysyä ajan tasalla niin suojauksen kuin itse datan käsittelyn kanssa. Turvallisuuden käytettävät ponnistukset todistavat, että datan parissa työskentelevät organisaatiot luottavat koko ajan kehittyvään infrastruktuuriin. (Sedkaoui 2018, 33–34.)

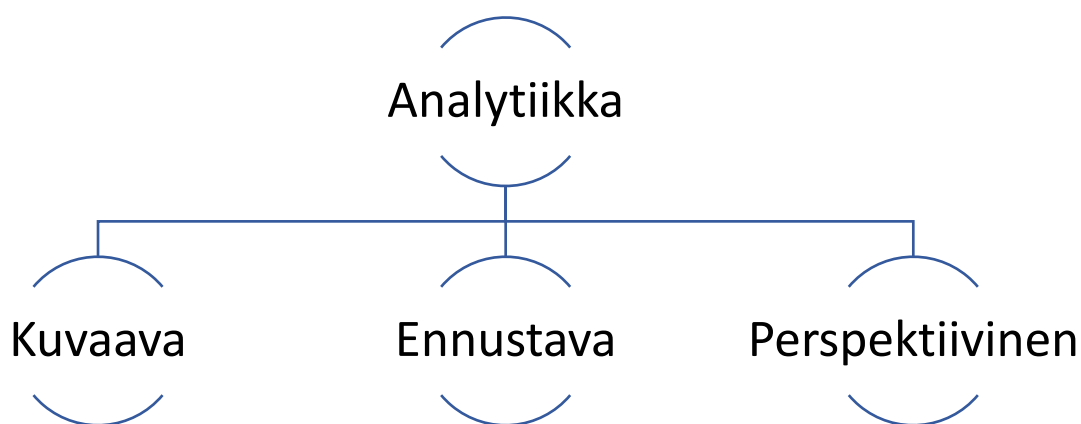
3 Data-analytiikka

Olemme tähän asti puhuneet datasta ja kuinka se voi oikein käytettynä avata uusia ovia niin organisaation kuin yhteiskunnan tasolla. Pelkällä datalla ei vielä suoraan saada haluttuja tuloksia, vaan tarvitaan myös osaamista sen käsittelyyn ja analysoimiseen. Tässä esiin astuu data-analytiikka. Puhumme aiheesta käsitellessä myös ohimenneen koneoppimisesta ja tekoälystä. Näiden termien läpikäyminen syvällisesti vaatisi oman päättötyön, mutta käymme niitä läpi, jotta näiden yhteys data-analytiikan kanssa tulisi selväksi.

Ymmärtämisen kannalta on tärkeää tuoda esiin, että alan kirjallisuudessa data-analytiikka, business analytics ja business intelligence (BI) ovat vahvasti yhteydessä toisiinsa. Data-analyysi käytännössä tarkoittaa menetelmiä, joilla raakadatasta saadaan muodostettua informaatiota, josta on apua päätöksenteossa. (AWAcademy 2021.)

Davenportin ja Harrisin (2007, 7) määritelmä analytiikalle on datan, tilastitiikan, kvantitatiivisen analyysin, erilaisten mallien sekä faktoihin perustuvan johtamisen käyttöä johtamisen ja päätöksenteon apuna.

Soraya Sedkaoui (2018, 45) nostaa kirjassaan esille Delenin ja Demirkanin, joiden mukaan yritykset yleensä tekevät kahta pääanalyysiä (kuva 2), kuvaavaa (descriptive) ja ennustavaa (predictive). Kuvaava analyysi keskittyy siihen, mitä on tapahtunut aikaisemmin, kun taas ennustava käyttää nimensä mukaan tuotettua dataa antaakseen ennustuksia tulevaisuuden tapahtumista. Delen ja Demirkan (2013) korostavat, että big data mahdollistaa myös kolmannen tavan analysoida dataa: perspektiivinen (perspective). Tämä yhdistää kuvaavan ja ennustavan analyysin tuottavan datan ulkoisen datan kanssa, ennustaakseen toimenpiteitä.



Kuva 2. Erilaisia analyysitapoja, mukautettu Delen & Demirkan 2013

3.1 Koneoppiminen ja tekoäly

Niin kuin aikaisemmin jo mainittiin, en käy näitä termejä läpi yksityiskohtaisesti. Koneoppimisen ja tekoälyn ymmärtäminen on kuitenkin tärkeää data-analytiikan kannalta, sillä näiden yhteys tulevaisuudessa tulee kasvamaan. Tämän takia termien avaaminen ja niiden välisten suhteiden läpikäyminen on haluttuun lopputulokseen pääsemisen kannalta tärkeää.

Tänä päivänä asiat ovat koko ajan enemmän yhteydessä toisiinsa, ja tämä lisää jo valmiiksi massiivista datan määrää entisestään. Tämä puolestaan lisää tarvetta työkaluille, jotka kykenevät lukemaan ja analysoimaan kasvavan datan määrän. On tullut selväksi, että tietokoneet kykenevät prosessoimaan dataa ja tunnistamaan yhteyksiä huomattavasti nopeammin ja isommassa mittakaavassa kuin yksikään ihminen. (Sedkaoui 2018, 103.)

Tekoälyn määritelmä SAS:in (SAS 2021) mukaan on tiede, jossa kone koulutetaan toimimaan ihmisen tavoin ja tekemään ihmisen tehtäviä. Ilman tekoälyä ei olisi koneoppimista, sillä koneoppimisessa hyödynnetään tekoälyä. Tekoäly on siis laajempi käsite, joka pitää sisällään koneoppimisen. Koneoppiminen taas on tapa analysoida dataa automaattisesti erilaisten mallien avulla. Koneoppiminen on osa tekoälyä ja se perustuu idealle, että järjestelmä osaa tehdä päätöksiä datan perusteella ilman suurempaa ihmisen osallistumista.

Järjestelmä oppii lukemalla dataa ja tunnistaa sen välisiä yhteyksiä ja tekee omat johtopäätöksensä näiden avulla. (SAS 2021.)

Aikaisemmassa kappaleessa mainittiin Delenin ja Demirkanin analyysitavat kuvaava, ennustava ja perspektiivinen. Näistä ennustava ja perspektiivinen eivät olisi mahdollisia ilman koneoppimista ja tekoälyä. Datan ja teknologian kehityksen myötä yritysten on mahdollista ennustaa asiakkaiden käyttäytymistä ja reaktioita. Tässä käytännössä analysoidaan aikaisempaa dataa ja informaatiota ja luodaan niiden perusteella ennuste tulevasta. (Sedkaoui 2018, 104–105.)

Tällä hetkellä koneoppimista hyödynnetään kaikkialla, missä käsitellään suuria määriä dataa, esimerkiksi rahoitussektorista vähittäistavarakauppaan ja terveydenhuoltoon (SAS 2021). Rahoitussektorilla koneoppimista käytetään esimerkiksi petosten estämiseen tunnistamalla korkean riskin profiileja, sekä auttamaan sijoittajia tekemään sijoittamisen kannalta oikeita päätöksiä, milloin ostaa ja milloin myydä. Vähittäistavarakaupassa koneoppimisen avulla voidaan yrittää personalisoida asiakkaan ostokokemus. Käyttämällä asiakkaan aikaisempia ostotottumuksia ja mielenkiinnon kohteita voidaan asiakkaalle tarjota häntä kiinnostavia asioita. (Sedkaoui 2018, 108.) Terveydenhuollossa koneoppiminen mahdollistaa asiakkaan tilan monitoroimisen reaaliajassa. Tämä teknologia auttaa terveydenhuollon ammattilaisia diagnosoimissa sekä hoitojen suunnittelussa. (Sedkaoui 2018, 108)

3.2 Analytiikka ennen

Vaikka data-analytiikka on moderni työkalu, on analytiikkaa ja tiedon analysointia tehty jo usean vuosikymmenen ajan. Tällä hetkellä suurin osa informaatiosta on sähköisessä muodossa ja niiden analysointiin käytetään sähköistä infrastruktuuria. Ennen tietotekniikan vallankumousta, vastaava informaatio oli paperisessa muodossa. Ihminen ei pysty prosessoimaan informaatiota samalla nopeudella kuin tietokone, joten data-analytiikka ja sen potentiaali ovat vahvasti sidoksissa tietotekniikan kehitykseen. (Sedkaoui 2018, 47.)

Internetin suosion kasvun ja teknologian kehityksen myötä data-analytiikka on ottanut isoja kehitysaskelleita. Vaikka analytiikkaa käytettiin jossakin muodossa jo toisen maailmansodan aikana, yritysmaailmassa data-analytiikan hyöty päätöksenteossa huomattiin 1960-luvun aikana, kun tutkijat alkoivat puhua tiedon ja informaation säilyttämisestä tietokannassa. 1970-luvulla ryhmä tutkijoita IBM:ltä alkoivat käyttää matemaattisia periaatteita erilaisten mallien ja ennustusten tekemiseen dataa hyväksi käyttäen. Tämä sai muutkin

kiinnostumaan asiasta ja tämän jälkeen olemassa olevaa informaatiota on sen kehityksen myötä käytetty datan hyödyntämisessä. (Sedkaoui 2018, 47.)

Data-analytiikan hyöty oli tiedossa, mutta 1970- ja 1980-luvulla tietokoneet olivat vielä isoja ja kalliita, joten vain isot yritykset pystyivät käyttämään niitä analyysin työkaluina. 1970-luvulla kehitettiin myös tietokantojen käsittelyyn käytettävä kieli SQL, joka toimii standardina vielä tänäkin päivänä. (Sedkaoui 2018, 48–49.)

3.3 Data-analytiikka tänään

Tämän työn aikana on jo monta kertaa tullut puheeksi, kuinka suureen rooliin data-analytiikka on noussut kuluneiden vuosien aikana. Pelkästään saatavilla olevan datan ja informaation määrä on mahdollistanut aivan uudenlaisen kasvun ja avannut uusia ovia yrityksille. Kilpailun kiristyessä yritykset keksivät luovia ratkaisuja hallussa olevan datan avulla pystyäkseen kilpailemaan markkinoilla. Teemme päivittäin asiakkaina toimenpiteitä, joita yritykset voivat hyödyntää omassa päätöksenteossaan ja tämä kaikki liittyy data-analytiikkaan. (EMC Education Services 2015, 2–3.)

Helpon data-analytiikan roolia tänä päivänä voi selittää sosiaalisen median avulla. Joka kerta, kun ihmiset esimerkiksi vierailevat internetsivulla tai katsovat YouTube -videon, jää tästä digitaalinen jalanjälki. Internetkäyttäjymisen avulla yritykset voivat kohdistaa ihmisille heitä kiinnostavia mainoksia. Eteen on voinut tulla tilanne esimerkiksi, kun internetissä ostoksia tehdessä, tai vain vieraillessa jollakin nettisivulla, tämän nettisivun mainoksia tulee eteen tulevaisuudessa myös muilla internetsivuilla. Tämä on esimerkki, miten yritykset voivat kohdistaa mainontaa heille relevanttiin asiakasryhmään heidän internetkäyttäjymisensä avulla. Vierailtaessa internetsivuilla tai verkkokaupoissa sivu kysyy lupaa käyttää tietoja myös tulevaisuudessa. (Mimaroglu 2015.)

Toisena esimerkkinä data-analytiikasta on tieto- ja tietoliikennetekniikkajätti Siemens. Yhtenä liiketoiminta-alueena Siemens valmistaa erilaisia osia junaliikenteeseen. He sanovat pystyvänsä sensoreiden tuottamalla datalla ennustamaan mahdolliset viat ja rikkoutumiset viikkoja etukäteen. Sen sijaan, että he reagoisivat tapahtuneeseen, Siemens käyttää data-analytiikkaa ja koneoppimista ennustaakseen mahdollisia ongelmatilanteita. (Sharda, De- len & Turban 2017, 444–445.)

Toimenpiteiden tekemistä käytössä olevan datan avulla - tätä data-analytiikka on käytännössä. Soraya Sedkaoui (2018, 61) nostaa esille neljä tapaa, miten data-analytiikka voidaan hyödyntää vahvistaakseen omaa asemaa kilpailijoihin nähden.

Ensimmäisenä näistä mainitaan data-analytiikka päätöksenteon tukena. Yrityksillä on massiiviset määrät informaatiota käytössään, mutta tämä tieto on kuitenkin pääosin hyödyttöä, jos sitä ei tulkita oikein tai se ei ole oikeassa muodossa. Yritys voi seuloa hallussa olevan datan, jotta voidaan keskittyä juuri olennaiseen ja tehdä sen avulla heille parhaita toimenpiteitä. Tämä tekee yrityksen toiminnasta tehokkaampaa, koska voidaan keskittyä olennaiseen, jolloin myös esimerkiksi resurssit ovat parhaiten käytössä. (Sedkaoui 2018, 61.)

Toiseksi, operatiivisten kustannusten alentamista pidetään yhtenä yritysten suurimmista haasteista. Data-analytiikan avulla voidaan tehdä tarkempia toimenpiteitä ja keskittyä ongelma-alueisiin tuhlaamatta resursseja. (Sedkaoui 2018, 61.)

Kolmantena, on myös tärkeää, että yritys tuntee asiakkaansa, tai ainakin heidän halunsa, preferenssit ja mielenkiinnon kohteet. Näiden informaatioiden avulla yritykset voivat suunnitella tuotteensa ja palvelunsa vastaamaan näitä. Informaation saamisessa sosiaalinen media on isossa roolissa. (Sedkaoui 2018, 61.)

Neljäntenä Sedkaoui (2018, 61) puhuu avoimen datan käytöstä. Avoimella datalla tarkoitetaan dataa, joka on nimensä mukaisesti vapaasti käytettävissä, uudelleenkäytettävissä. Esimerkiksi informaatio, jonka kaupallinen käyttö on kielletty, ei kuulu tähän. Kuluneiden vuosien aikana avoimen datana käyttö on lisääntynyt tuotteiden ja palvelujen kehittämisessä. (Sedkaoui 2018, 61.)

Suomessa avoimeen dataan kuuluu esimerkiksi Tilastokeskuksen ylläpitämät tietokantaineistot sekä Ilmatieteenlaitoksen avoin data. Tilastokeskuksesta saatavaan avoimeen dataan kuuluu esimerkiksi kuntien ja kuntayhtymien raportoimat taloustiedot, liikenneonnettomuuksien tietokanta sekä verohallinnon tietokanta. (Tilastokeskus 2021.) Ilmatieteenlaitos tarjoaa avointa dataa esimerkiksi auringon säteilyhavainnoista, ilmanlaatuhavainnoista, säähavainnoista sekä meriveden korkeushavainnoista (Ilmatieteen laitos 2021). Digi- ja väestötietovirasto ylläpitää internetsivua avoindata.fi, johon on kerätty kaikki Suomen avoin data yhteen paikkaan (Digi- ja väestötietovirasto 2021).

3.3.1 Isot vs. pienet yritykset

Data-analytiikka on nykypäivää ja jos yritys haluaa olla kilpailukykyinen omalla alallaan, täytyy heidän ottaa tämä huomioon. Miten tämä näkyy pienissä ja isoissa yrityksissä? Ovatko erikokoiset toimijat eri lähtökohdissa? Voiko pienempi toimija myös hyötyä data-analytiikasta?

Data-analytiikka ei ole vain isojen yritysten työkalu, vaan myös pienemmät yritykset ja yksittäiset toimijat voivat hyödyntää data-analytiikkaa päivittäisessä toiminnassaan. Se on kuitenkin selvää, että isoilla yrityksillä on käytössään enemmän itse tuotettua dataa ja informaatiota. Esimeriksi yritykset kuten Google ja Facebook tuottavat massiivisen määrän dataa itselleen pelkästään heidän omalla toiminnallaan, mutta se ei tarkoita, että data-analytiikan hyödyt olisivat pienempien toimijoiden ulottumattomissa. (Marr 2020.)

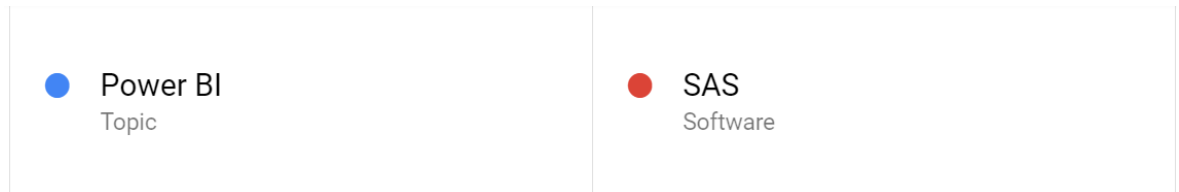
Yrityksen kasvun myötä tulevia etuja ei voi väheksyä ja koko tuokin mukanaan omia etujaan, kuten resurssien määrän. Alan johtaviin ammattilaisiin kuuluva Bernard Marr (2020) sanoo kirjassaan, että omalla tavallaan data-analytiikasta voi oikein käytettynä olla enemmän hyötyä pienille toimijoille heidän ketteryytensä vuoksi. Hän korostaa asiakkaiden tuntemisen tärkeyttä, trendien tunnistamista ja kilpailijoiden silmälläpitämistä (Marr 2020).

Niin kuin monesti jo tämän työn aikana on tullut esille, informaatio on data-analytiikan tärkein polttoaine. Informaatiota on tarjolla yrityksille koosta riippumatta ja vaikka datan määrä vaihtelee yrityksen koon mukaan, on periaate kuitenkin sama. Analysoimalla omia myyntinumeroita ja asiakaskäyttäytymistä, voi tehdä liiketoimintaa edistäviä toimenpiteitä. Tässäkin sosiaalinen media on jälleen isossa roolissa: sosiaalinen media on tehnyt asiakasanalyysistä helpompaa ja halvempaa. (Marr 2020.)

Trendejä tulee ja menee ja niiden ennustaminen on hyvin vaikeaa. Olemalla oikeassa paikassa oikeaan aikaan, voi tehdä hyvän tilin. Bernard Marrin (2020) mukaan trendien ennustaminen liittyi aikaisemmin vahvasti ”gut instincts” eli vaistoihin ja arvailuihin. Data-analytiikka auttaa tässä pienentämällä arvaamisen roolia trendien ennustamisessa. Facebook ja Twitter näyttävät jatkuvasti mitkä aiheet ovat ”trendeamassa” milläkin hetkellä ja tämä informaatio on kaikkien käytössä ilmaiseksi. (Marr 2020.)

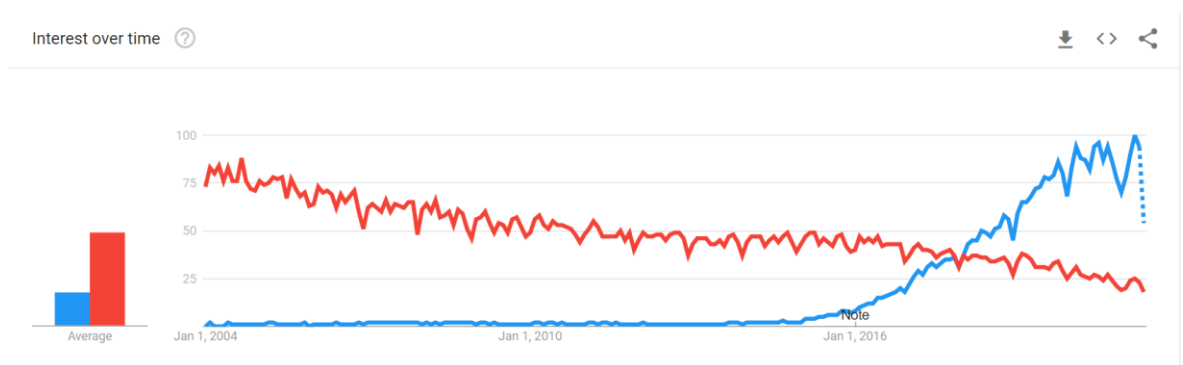
Kilpailijoiden seuraaminen ei koskaan ole ollut näin helppoa. Internetin avulla yritykset voivat vertailla itseään kilpailijoihin ja katsoa millaisessa asetelmassa he ovat toisiin yrityksiin nähden. Google Trends, Youtube analytics, ja sosiaalisen median analyysit voidaan tehdä

omalta kotikoneelta. On tärkeää muistaa, että kaikki tämä informaatio on myös kilpailijan käytössä ja he tietävät sinusta yhtä paljon, kun sinä heistä. (Marr 2020.)



Kuva 3. Power BI vs. SAS, Google Trends vertailu

Google Trends on nopea ja helppo työkalu antamaan yleiskuvan tämänhetkisestä tilanteesta, olettaen, että hakee oikeita asioita. Google Trendsin avulla voi vertailla esimerkiksi kahta vastaavaa data-analytiikan ohjelmaa ja katsoa minkälaisessa asetelmassa ne ovat toisiaan vasten. Voi esimerkiksi vertailla, kumpi on suosituampi hakutulosten perusteella, joko maakohtaisesti tai maailmanlaajuisesti. Samalla näkee missä maissa ja millä maanosalla haetaan kumpaakin enemmän ja miten näiden hakutulosten määrä on kehittynyt vuosien saatossa. Tarkasteltavan ajan pituutta voidaan myös muuttaa, jos halutaan tarkempaa tietoa lyhyemmältä ajalta. Google Trends antaa myös ehdotuksia muista aiheeseen liittyvistä hakukohteista. Tein tätä kappaletta varten nopean Google Trends vertailun kahden suosituksen data-analytiikka ohjelmiston Power BI:n ja SAS:in välille. Power BI on merkattu tässä vertailussa sinisellä ja SAS punaisella värillä (Kuva 3). Power BI tuli julkisille markkinoille vuonna 2015 ja niin kuin kuvasta voi huomata (Kuva 4) sen trendikäyrä on vahvasti ylöspäin.



Kuva 4. Power BI ja SAS trendikehitys 6.5.2021

3.4 Miltä näyttää tulevaisuus?

Data-analytiikan rooli ja tarve työelämässä on kasvanut datan määrän lisääntyessä. Kulu-
neiden vuosien aikana on nähty massiivista kasvua datan määrässä, mutta voiko sama
kasvu jatkua vielä tulevaisuudessakin? Miten data-analytiikka aikoo vastata tämänhetki-
seen kasvuun? Jatkaako data-analytiikka vahvaa kasvuaan ja millaisia trendejä alalla on
nähtävissä?

Voiko datan määrä jatkaa kasvuaan vielä tulevaisuudessakin? Lyhyesti ja ytimekkäästi
”kyllä”. Tällä hetkellä datan määrä jatkaa kasvuaan, eikä näytä hidastumisen merkkejä.
Kasvuvauhti on niin nopeaa, että yritysten on täytynyt keksiä uusia tapoja datan käsitte-
lyyn, koska vanhanaikaisemmat systeemit eivät pärjää informaation määrän kanssa. Da-
tamarketin vuosittainen kasvu on ollut noin seitsemän kertaa nopeampaa kuin koko tietö-
ja kommunikaatiomarketin. Tällä hetkellä suuri osa alan investoinneista menee software-
kehitykseen ja data mining -teknologiaan. (Liebowitz 2013, 143.)

Bernard Marr (2021) tuo esille Forbesissa julkaistussa artikkelissaan data-analytiikan tren-
dejä tulevaisuudelle. Hän korostaa muun muassa tekoälyn roolia tulevaisuudessa sano-
malla, että tekoälyllä voidaan saavuttaa asioita, joista vielä vuosia sitten on voinut vain
haaveilla. Toisena trendinä mainitaan datan visualisointi ja sen tulkinta. Datan tulkinta on
vaikeaa ja kaikki eivät saa saman tien selkoa koneen tuottamasta informaatiosta. Tämän
kanssa on tehty läpimurtoja, jotka mahdollistavat kysymysten esittämisen datasta tietoko-
neelta. Tämä helpottaa datan ymmärtämistä huomattavasti. (Forbes, Marr 2021.)

3.5 Internet of Things

Internet of Things tarkoittaa ”normaalin” fyysisen elämän yhdistämistä internetin kanssa.
Tässä laitteet ovat yhteydessä erilaisiin sensoreihin, jotka kommunikoivat keskenään ja
jakavat dataa. Tässä esimerkiksi joku laite on yhteydessä sensoreihin, joka kerää infor-
maatiota ja dataa laitteen toiminnasta, sijainnista ja kunnosta. Tätä kautta tuotettua infor-
maatiota voidaan tämän jälkeen analysoida ja käyttää haluttuun tarkoitukseen. Esimerkit
aiheesta vaihtelevat futuristiseen itse ajavaan autoon, jääkaappiin, joka itse tilaa maitoa
sen ollessa vähissä tai jo olemassa oleviin urheilusuoritusta seuraaviin urheilu- tai älykel-
loihin. (Sharda, Delen & Turban 2017, 445.)

Tämä trendi on vahvassa kasvussa, koska tämä teknologia on jatkuvasti halvempaa ja
kannettavassa muodossa. Jokaisella meistä on älypuhelin ja suurella osalla väestöstä on

äly- tai urheilukello. Nämä laitteet tuottavat dataa, oli se sitten urheilusuorituksen aikana tai kauppareissulla. (Sharda ym. 2017, 445–446.)

Sharda, Delen ja Turban (2017, 446) mainitsevat kirjassaan kolme pääsyytä sille, miksi Internet of Things on vahvassa kasvussa. Ensimmäisenä syynä on, niin kuin aikaisemmassa kappaleessa jo mainitsinkin, että teknologia mahtuu pienempään tilaan, on halvempaa valmistaa ja tehokkaampaa kuin aikaisemmin. Datan käsittelyn sekä mobiililaitteiden kustannukset ovat pienentyneet 97 % viime vuosikymmenestä. (Sharda ym. 2017, 446.) Toisena syynä on analytiikkatyökalujen määrän lisääntyminen, niiden saatavuuden helpottuminen sekä niiden teknologian kehittyminen. Yritykset voivat tänä päivänä valita useasta eri toimijasta itselleen sopivimman ratkaisun. Kolmantena mainitaan, kuinka nopeaa tämä kehitys on ollut ja uusia käyttömahdollisuuksia tulee koko ajan lisää. Voimme tulevaisuudessa käyttää tätä teknologiaa asiaan, jota ei tällä hetkellä vielä edes ole olemassa. (Sharda ym. 2017, 446.)

Juniper Researchin mukaan vuoteen 2020 mennessä yli 38 miljardia laitetta on yhteydessä internetiin. Tämä olisi 285 % kasvu vuodesta 2015. (Juniper Research 2016.)

3.6 Data-analytiikka alustat

Olen tähän asti kirjoittanut paljon datasta, sen määrästä ja sen analysoinnista. Tässä kappaleessa käyn läpi erilaisia data-analytiikkaan suunniteltuja sovelluksia. Koska opinnäytetyön sivumäärä on rajattu, joudun karsimaan, mitä sovelluksia ja kuinka syvällisesti voimme niitä käydä läpi. Valitsen tarkasteluun ohjelmia, jotka alan ammattilaiset ja kirjallisuus nostaa esille tulevaisuuden trendeinä ja toimijoina. Koska data-analytiikkaohjelmat ovat suurelta osin maksullisia, tämä luku perustuu pitkälti saatavilla olevaan materiaaliin sekä ammattilaisten kommentteihin, mielipiteisiin ja arvosteluihin.

Miettiessäni, mitkä alustat otan tarkasteluun, luin läpi paljon erilaisia listoja aiheesta. Tarjoituksenani oli löytää, mitkä alustat ja ohjelmistot ovat alalla työskentelevien mielestä vahvoilla tulevien vuosien aikana. Valitsin vertailuun alustat, joista alan toimijat olivat samaa mieltä. Alan suurimpiin vaikuttajiin kuuluva Bernard Marr, sekä tietotekniikkajätti Gartner antoivat omat mielipiteensä ja valitsin alustat näiden julkaisujen perusteella. Halusin myös, että valitsen sellaisia alustoja, jotka ovat käytössä case-tutkimukseeni kuuluvilla yrityksillä. Neljästä case-yrityksestäni kolme käyttävät joko SAS:in data-analytiikka ohjelmia tai Microsoftin Power BI:tä. Tämän takia valitsin nämä kaksi alustaa tarkempaan tarkasteluun.

3.6.1 SAS Advanced Analytics

SAS jatkaa rooliaan yhtenä johtavana analytiikkakehittelijänä. Gartner valitsee vuosittain alan johtavan myyjän ja SAS on valittu tässä vertailussa parhaiden joukkoon kahdeksan vuotta peräkkäin. SAS on myös alan ylivoimaisesti suurin toimija. IDC Researchin mukaan SAS on johtavassa markkina-asemassa analytiikan parissa hieman vajaan 28 % markkinaosuudella vuonna 2018. Jotta tämä luku tarkoittaisi jotain, se on yli kaksi kertaa suurempi, kun alan toiseksi suurin. (SAS 2019.)

Tuhannet yrityksen maailman ympäri käyttävät SAS:in alustoja. Näihin kuuluu muun muassa Volvo kuorma-autot, NHL joukkue Los Angeles Kings, Lockheed Martin ja WWF. Vuoden 2019 Fortune 500 listalla 91 yritystä 100:sta suurimmasta yrityksestä oli SAS:n asiakas. (SAS 2021.) SAS on kehittänyt ohjelmiston visuaalista kykenevyyttä ja sen automaatiota. Se on suunniteltu, jotta asiakkaat voisivat pitää koko analytiikka tietovirran yhden alustan sisällä. (Marr 2020.)

Tässä luvussa on tarkoitus lyhyesti kertoa, mitä SAS:in data-analytiikka alustalla on mahdollista tehdä. SAS listaa omilla sivuillaan data-analytiikka alustojensa käyttömahdollisuuksia ja pääominaisuuksina nostetaan tiedonlouhinta, tilastollinen analyysi, ennustaminen, tekstin analytiikka sekä optimointi ja simulointi.

Tiedonlouhinnalla voidaan tunnistaa asioiden välisiä suhteita erilaisten graafien ja taulukoiden avulla. Voidaan luoda uusia malleja halutuilla muuttujilla. Tilastollista analyysiä voidaan käyttää esimerkiksi myyntinumeroiden tai asiakasdatan analysoimiseen. Tämän avulla voidaan verrata menneitä lukuja tämänhetkiseen tilanteeseen ja antaa ennustuksia tulevaisuuden luvuista. Ennustamisella ohjelma luo automaattisesti olemassa olevan informaation avulla ennustuksia tulevaisuudesta. Tämän avulla voidaan tehdä suunnitelmia ja testejä tulevaisuutta varten. Tekstin analytiikan avulla voidaan analysoida isoja määriä tekstiä nopeasti. Voidaan myös automatisoida aikaa vievät manuaaliset toimenpiteet ja tunnistaa nopeammin tärkeät asiat suuresta tekstin määrästä. Optimointi ja simulointi auttavat tunnistamaan toimenpiteitä, joilla pääsee parhaaseen tai haluttuun lopputulokseen. Tämä helpottaa syy- ja seuraussuhteen tunnistamista. Voidaan myös muun muassa yrittää ennustaa, minkälainen vaikutus jollakin toimenpiteellä on. (SAS 2021.)

3.6.2 Microsoft Power BI

Niin kuin SAS, myös Microsoftin Power BI on menestynyt Gartnerin analytiikka-alustojen vertailussa. Vuoden 2021 julkaisussa Microsoft Power BI valittiin analytiikka-alustojen johtajan paikalle. (Microsoft 2021.) Myös data-analytiikka -alan johtaviin vaikuttajiin kuuluva Bernard Marr (2020) on nostanut Power BI:n tulevaisuuden data-analytiikkaohjelmaksi kuululle vuodelle. Hänen mukaansa Power BI:n ehdoton etu on siinä, että se on tutun tuntuinen suurelle osalle alan ihmisistä, koska se kuuluu Office 365 ekosysteemiin (Marr 2020).

Power BI:n avulla voidaan muuttaa olevassa olevan datan helposti luettavampaan muotoon, kuten taulukoiksi ja graafeiksi. Tämä helpottaa ja nopeuttaa informaation lukemista. Power BI ohjelmistoa voidaan käyttää suoraan esimerkiksi Google Analytics- tai esimerkiksi Excel -tiedostojen käsittelemiseen. Power BI toimii myös muiden Office -pakettiin kuuluvien sovellusten kanssa, kuten Teamsin tai SharePointin. Tämän yhteyden avulla dataa voidaan jakaa organisaation muiden jäsenten kanssa, joka tekee työntekijöiden välisestä kommunikaatiosta ja yhteistyöstä helpompaa. (Microsoft 2021.)

Power BI:tä voidaan käyttää paikan päällä tai pilvessä. Power BI:n avulla käyttäjä voi valita haluamansa datan lähteen, kuten esimerkiksi Excelin tai Azuren. Power BI:n avulla voidaan analysoida hallussa olevaa dataa ja luoda tämän informaation avulla ennustuksia tulevaisuudesta. Tämän avulla voidaan suunnitella paras mahdollinen tapa saavuttaa halutut tavoitteet. (Microsoft 2021.)

Power BI voi hyödyntää myös tekoälyä datan analysoimisessa. Ohjelmalla voidaan analysoida tekoälyn avulla esimerkiksi numeerisen datan lisäksi myös tekstiä ja jopa kuvia. Tämä tekee esimerkiksi asiakaspalautteiden ja sosiaalisen median analysoinnista huomattavasti helpompaa. Power BI:ssä on myös ominaisuus, jossa voidaan esittää kysymyksiä aineistolta ja tämän jälkeen Power BI etsii datasta haettua asiaa ja esittää nämä hakutulokset. (Microsoft 2021.)

4 Tutkimusmenetelmät

Tämän opinnäytetyön tarkoituksena on saada yleinen käsitys data-analytiikasta ja lähteä käytännön esimerkein selvittämään, miten sitä käytetään yrityksissä. Lähdin hakemaan vastausta tähän kahdella pääkysymyksellä:

- Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksissä?
- Mitä data-analytiikan ohjelmistoa yrityksissä käytetään?

Alun perin tarkoituksena oli lähteä hakemaan vastausta näihin kysymyksiin kahdella tavalla. Ensiksi ajattelin etsiä kirjallisuudesta esimerkkejä data-analytiikan eri käyttömahdollisuuksista ja miten yritykset sitä käytännön tasolla hyödyntävät. Tämän lisäksi halusin esittää alan parissa työskenteleville kysymyksiä, miten heillä data-analytiikka näkyy päivittäisessä toiminnassa. Tämän jälkeen tarkoituksena oli vertailla näistä kahdesta lähteestä saatuja tuloksia. Ajan edetessä, kun vastauksia ei alkanut kuulua, jouduin kuitenkin muuttamaan alkuperäistä suunnitelmaa.

Alkuperäisen suunnitelman sijasta päädyin tekemään case-tutkimuksen neljästä eri yrityksestä ja niiden data-analytiikan käytöstä. Näissä case-tutkimuksissa oli tarkoituksena käydä esimerkkien avulla, mitä ohjelmistoa ja mihin valittua data-analytiikan ohjelmaa käytetään.

Yritykset valikoituivat mielenkiinnon, erilaisuuden sekä materiaalin saatavuuden mukaan. Halusin tutkittavaksi erilaisia yrityksiä eri aloilta, jotta saisin paremmin kuvaa data-analytiikan monikäyttöisyydestä. Yritysten valinnassa oli tärkeää, että valitsee yrityksiä, joista on saatavilla tarpeeksi materiaalia, jotta pääkysymyksiin saa vastattua tarpeeksi kattavasti. Suurimmilla ja suosituimmilla data-analytiikka-alustojen tarjoajilla on omilla sivuillaan listoja asiakkaistaan, jotka käyttävät heidän data-analytiikkaohjelmaansa. Näiltä sivuilta valitsin kiinnostavia yrityksiä, joista löytyi materiaalia työtä varten. Myös nämä yritykset julkaisivat omaa materiaaliaan, jota pystyi käyttämään työssä hyödyksi. Jokaisesta case-yrityksestä löytyi omia esimerkkitapauksia data-analytiikan käytöstä ja sen hyödyntämisestä, mutta joidenkin yritysten kohdalla materiaalia oli rajallisempi määrä kuin toisilla.

Informaation ja materiaalin keräämisen jälkeen, kokosin saamani tiedot yrityskohtaisiin kappaleisiin.

5 Data-analytiikan käyttö yrityksissä

Miten data-analytiikkaa hyödynnetään yrityksissä tänä päivänä? Olen tähän asti kirjoittanut paljon datasta, sen määrästä ja sen analysoinnista teoriasolla. Tässä kappaleessa käyn läpi erilaisia käytännön esimerkkejä, miten data-analytiikkaa hyödynnetään eri alojen yrityksissä. Tämä antaa parempaa kuvaa data-analytiikan monipuolisuudesta. Jotta saataisiin mahdollisimman monipuolinen kuva, valitsen eri data-analytiikka-alustoja ja eri toimialojen yrityksiä. Valitsen tarkasteluun ohjelmia, jotka alan ammattilaiset ja kirjallisuus nostaa esille tulevaisuuden trendeinä ja toimijoina. Jokaisen casen jälkeen on tehty taulukko, jossa vedetään yhteen casen pääasiat.

5.1 Volvo ja Mack kuorma-autot

Volvo ja Mack kuorma-autot käyttävät data-analytiikkaohjelmistonaan SAS:in tuottamaa alustaa. SAS jatkaa rooliaan yhtenä johtavana analytiikkakehittelijänä ja tämän lisäksi SAS on myös alan ylivoimaisesti suurin toimija. IDC Researchin mukaan SAS on johtavassa markkina-asemassa analytiikan parissa hieman vajaan 31 % markkinaosuudella. Jotta tämä luku tarkoittaisi jotain, se on yli kaksi kertaa suurempi, kun alan toiseksi suurin kilpailija. (SAS 2021.)

Ruotsalaisen Volvo AB:n tytäryhtiöt Volvo Trucks ja Mack ovat ymmärtäneet data-analytiikan tuoman hyödyn ja ovat alkaneet varustaa rekkojaan data-analytiikkaominaisuuksilla ja -ohjelmistoilla (SAS 2020).

Jokapäiväinen elämämme on riippuvainen tavara- ja rekkaliikenteestä. Päivittäin miljoonat rekat kuljettavat tarvittavia elintarvikkeita ja raaka-aineita paikasta toiseen. Jos ajoneuvoon tulee yllättäen vika, joka estää ajoa, voi tämä tulla kalliiksi niin kuljetusfirmalle kuin asiakkaalle. Vian tullessa ei enää pystytä suunnitellussa aikataulussa, jolla on suora vaikutus yrityksen päivittäiselle toiminnalle. (SAS 2020.)

Volvo ja Mack ovat molemmat alkaneet asentaa ajoneuvoihinsa erilaista analytiikkaa ja tekoälyä omaavaa teknologiaa. Näiden teknologioiden avulla voidaan pitää huoli siitä, että ajoneuvot viettävät mahdollisimman paljon aikaa tien päällä ja mahdollisimman vähän turhissa huolloissa. Käytössä olevan diagnostiikan avulla ajoneuvoja voidaan huoltaa tehokkaammin, tarkemmin sekä etukäteen ennen isompia ongelmia. (SAS 2020.)

Conal Deedy (connected vehicle services johtaja) Volvo Trucksilta sanoo, että osa ajoneuvojen ongelmista liittyy ajoneuvon omaan ohjelmistoon. Data-analytiikka mahdollistaa näiden ongelmien hoitamisen, analysoimisen ja päivittämisen etänä. Ohjelmisto-ongelmalle voidaan ajaa etänä tarkistus, josta selviää, vaatiiko vika huoltoa, vai voiko ohjelmiston päivittää tien päällä ilman huoltotarvetta. Huoltoa vaatiessa huolto saa siitä tiedon etukäteen. Tämän ansiosta he tietävät etukäteen mitä osia tarvitaan ja mikä ajoneuvossa on vikana. (Deedy C. Volvo trucks, SAS 2020.)

Tällä hetkellä Volvolla on yli 175 000 rekkaa, joihin voidaan tehdä etänä toimenpiteitä. Näiden ohjelmistojen ansiosta Volvo on pystynyt pienentämään diagnostiikkaan käytettävää aikaa 70 % ja huoltoihin käytettävää aikaa 25 % sanoo Conal Deedy. (Deedy C. Volvo trucks, SAS 2020.)

Näiden lisäksi Volvo on kiinnostunut myös Lithium-ion-akuista ja miten niiden käyttöikää voisi mahdollisesti pidentää rekkaliikennettä varten. Rikard Vinge työskentelee joukossa data-analytikoita, jotka analysoivat dataa sähkö- ja hybridajoneuvoista. Näistä saadun informaation avulla voidaan ennustaa, koska on paras aika vaihtaa akku ja mistä syystä se kannattaa vaihtaa. Datalla sähköakkujen kanssa on Rikard Vingen mukaan isoa potentiaalia, mutta vielä ollaan alkutekijöissä. (Vinge, R., Volvo Group 2020.)

Taulukko 2. Yhteenveto Volvo & Mack Trucks

Yhteenveto	
Data-analytiikka ohjelma	Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksessä?
SAS Advanced Analytics	<ul style="list-style-type: none"> - Ajoneuvojen tilan seurana etänä - Mahdollisten vikojen korjaus etänä - Huoltojen suunnittelu ja niihin valmistautuminen datan avulla - Datan analysointia akkujen käyttöä pidentämiseksi

5.2 Data-analytiikka ja lentoliikenne – Heathrow’n lentoasema

Microsoft Power BI on valittu Heathrow’n lentokentällä apuvälineeksi hektistä arkea vastaan. Tämän data-analytiikka-alustan suurin etu on, kuinka tutulta se tuntuu, koska se

pelaa yhteen Office 365 ohjelmistojen kanssa (Bernard Marr 2020). Se on myös Microsoft Azuren kanssa työkalu, jota käytetään yhdellä maailman eniten liikennöidyistä lentokentistä, Lontoon Heathrow'n lentokentällä.

Yhteen maailman kiireisimmistä lentokentistä kuuluvalla Heathrow'n lentoasemalla käy vuosittain yli 80 miljoonaa asiakasta matkallaan ympäri maailmaa. Kaikkien näiden asiakkaiden lentomatkustamisesta täytyy tehdä mahdollisimman mukavaa, sillä lentomatkustaminen on jo valmiiksi tarpeeksi stressaavaa. Tällainen asiakasmäärä tuottaa myös massiivisen määrän dataa. Miten Heathrow hyödyntää data-analytiikkaa tehdäkseen lentämisestä mahdollisimman mukavaa asiakkaille ja mahdollisimman helppoa johtaa heille itselleen?

Heathrow'n lentoasema käyttää Azurea kerätäkseen taustatoiminnoista tärkeää tietoa, jonka jälkeen se työntää tämän tiedon eteenpäin Power BI:hin. Power BI mahdollistaa, että työntekijöillä on kyky seurata reaaliajassa, miten matkustajaliikenne etenee ja valmistautua tulevaan asianmukaisesti. Lentokentillä täytyy ottaa huomioon monia asioita. Heathrow'n lentoasema käyttää data-analytiikkaa kerätäkseen informaatiota kaikesta, mikä voi millään tavalla vaikuttaa lentoliikenteeseen. Näitä ovat esimerkiksi sää, lentojen aikataulut, reaaliaikainen lentoliikenne, matkatavara- ja rahtiliikenne, turvatarkastus sekä matkustajien liikkeet lentokentällä. Tämä informaatio on luettavissa sitä tarvitsevalla henkilökunnalla. Jos esimerkiksi useampi lento on myöhässä samaan aikaan ja 6000 asiakasta laskeutuu kentälle yhtäaikaisesti, ollaan tästä data-analytiikan avulla tietoisia etukäteen. Data-analytiikan ansiosta tähän on voitu varautua etukäteen lentokentällä laittamalla resursseja sinne, missä niitä tarvitaan. (Microsoft 2018.)

Stuart Birrell, joka toimii tietojenhallintajohtajana Heathrow'n lentokentällä sanoo, että he kokevat päivittäin isoja muutoksia ja vaihteluja. Käytössä olevien ohjelmistojen ansiosta näihin muutoksiin voidaan reagoida pitämällä asiakkaiden ja työntekijöiden stressitasot mahdollisimman alhaisena. Hänen mukaansa he käyttävät analytiikkaa ja tekoälyä luodakseen asiakkaille parhaan mahdollisen lentokenttäelämyksen. (Birrell, S. 2018, Microsoft.)

Taulukko 3. Yhteenveto Heathrow'n lentokenttä

Yhteenveto	
Data-analytiikka ohjelma	Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksessä?
Microsoft Power BI	- Matkustajaliikenteen seuraaminen ja ennakoiminen

	<ul style="list-style-type: none"> - Lentojen aikataulujen seuraaminen - Reaaliaikaisesti lentoliikenteen seuraaminen - Säätietojen seuraaminen ja ennakointi
--	--

5.3 KONE

Kaikki suomalaiset tietävät KONEen. KONE oli vuoden 2020 markkina-arvoltaan mitattuna Suomen suurin yritys yli 31 miljardin markkina-arvolla. Jos olet Suomessa käyttänyt hissiä tai liukuportaita, on sinulla hyvin todennäköisesti omakohtaista kokemusta KONEen tuotteista. KONE kuuluu maailman johtaviin hissi- ja liukuporrasvalmistajiin, joka toimii yli 60:ssa maassa ja liikuttaa yli miljardia ihmistä päivästä toiseen. (KONE 2021.)

KONE arvioi, että maailmanlaajuisesti yli 200 000 ihmistä muuttaa kaupunkeihin joka päivä. Jotta tähän kasvavaan tarpeeseen kyetään vastaamaan myös tulevaisuudessa, KONE on integroinut analytiikkaa ja tekoälyä omiin systeemeihinsä yhdistämällä Salesforce Einstein ja IBM Watsonin. (Salesforce.)

Reaaliaikainen data tekee laitteiston huoltamisesta ja ylläpitämisestä helpompaa ja tätä kautta parantaa asiakaskokemusta. Michael Williams (Head of Service Field Solutions Development) sanoo, että aikaisemmin oli vaikeuksia vastata asiakkaiden kysymyksiin mahdollisista ongelmatilanteista, koska heillä ei ollut tarvittavaa informaatiota. Data-analytiikan avulla kalusto on liitetty ohjelmistoon, jonka avulla esimerkiksi hissien toimintaa voi seurata reaaliajassa. Koska kaikki elementit ovat liitetty yhteen, on ongelmanratkaisusta tehty tehokkaampaa. (Williams, M. KONE.)

Sensoreiden antamaa dataa seuraamalla KONE voi monitoroida laitteiden toimintaa, analysoida laitteiston tuottamaa dataa ja tämän avulla ennustaa mahdolliset riskit ja ongelmatilanteet ennen niiden tapahtumista. Tämän teknologian avulla voidaan minimoida aika, jonka aikana koneet olisivat täysin käyttämättä, sanoo koneen pääarkkitehti Fareed Ahmed. (Ahmed, F. 2018 IBM.)

Michael Williamsin mukaan tämän teknologian ansiosta asiakaspalvelupäivystäjät voivat etänä katsoa miltä tilanne näyttää asiakkaan puolella. Päivystäjä näkee myös missä on lähin huoltohenkilö, jolla on vaadittavat työkalut ja osaaminen ongelman korjaamiseen.

Kun huoltohenkilö pääsee paikan päälle hän voi järjestelmästä katsoa, mikä laitteessa on todennäköisesti vikana. (Williams, M. KONE.)

Koneen toimitusjohtaja Henrik Ehrnrooth sanoo, että tämän teknologian avulla KONE voi palvella asiakkaitaan entistä paremmin, tarjota yksilöllistä palvelua ja tehdä kaiken tämän laadukkaasti ja luotettavasti. (Ehrnrooth, H. KONE.)

Taulukko 4. Yhteenveto KONE

Yhteenveto	
Data-analytiikka ohjelma	Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksessä?
Salesforce Einstein	<ul style="list-style-type: none">- Kaluston kunnan reaaliaikainen seuranta- Datan avulla huoltoihin valmistautuminen- Huoltojen suunnittelulla minimoidaan hissien ja liukuportaiden seisomista

5.4 Data-analytiikka taistelussa syöpää vastaan

Amsterdam UMC on yksi Euroopan suurimmista syöpäsairaaloista. Amsterdam UMC on alkanut hyödyntää data-analytiikkaa ja tekoälyä, jotta kasvaimet saadaan tunnistettua ja arvioitua mahdollisimman tarkasti ja nopeasti. Näitä teknologioita käytetään luomalla malleja ja analyyseja, joiden avulla voidaan tunnistaa kandidaatteja hoitoja varten. Niin kuin kaikessa muussakin data-analytiikkaan liittyvässä, kaikki lähtee liikkeelle datan määrästä. Tämä pitää paikkansa myös Amsterdam UMC:n tapauksessa. Jokaisesta potilaasta kerätään isot määrät dataa esimerkiksi biomarkkereita, genomeja ja DNA:ta. (SAS 2020.)

”Tekoäly tulee auttamaan meitä pelastamaan henkiä. Olen aivan varma tästä.” Näin sanoo Amsterdam UMC:n kirurgisen syöpäosaston johtaja Tohtori Geert Kazemier. Hänen mukaansa he ovat jo tällä hetkellä kykeneviä ennustamaan, minkälainen reaktio hoidoilla saadaan aikaan. Tämä prosessi on Geert Kazemierin mukaan nopeampi ja tarkempi, kuin

jos se olisi ihmisen tekemä. Geert Kazemier näkee isoa potentiaalia ja mahdollisuuksia analytiikassa lääketieteissä. (Kazemier, G. UMC Amsterdam, SAS.)

Tekoälyn ja analytiikan avulla voidaan muun muassa tunnistaa kasvaimia ja niiden muutoksia nopeammin ja tarkemmin. Amsterdam UMC käyttää erilaisia tekoäly- ja koneoppimisen malleja ennustaakseen, kuinka nopeasti ja tarkasti kemoterapia vaikuttaa potilaaseen ja minkälaisia vaikutuksia kemoterapialla tulee olemaan. Tämänlainen teknologia auttaa lääkäreitä tunnistamaan nopeasti muutoksia kasvaimen koossa ja muodossa ja tekemään nopeammin mahdollisesti henkiä pelastavia päätöksiä potilaan tulevia hoitoja varten. Analytiikan ja tekoälyn avulla voidaan luoda 3D -malleja, jotka auttavat lääkäreitä tutkimaan kasvainta vielä entistäkin tarkemmin. (Kazemier, G. UMC Amsterdam, SAS.)

Amsterdam UMC:n tutkijat käyttävät data-analytiikkaa myös vahvistaakseen omia tutkimuksiaan. He pystyvät analytiikan avulla tunnistamaan trendejä, analysoimaan tutkimustuloksia sekä luomaan ennustavia malleja. Analytiikkaa käytetään myös erilaisten lääketieteellisten kuvien analysoimiseen: käyttämällä analytiikkaa kuvien analysoimiseen pienennetään riskiä tehdä virheitä, esimerkiksi väärin positiivisten ja väärin negatiivisten määrää on saatu vähennettyä. (SAS 2020.)

Taulukko 5. Yhteenveto Amsterdam UMC

Yhteenveto	
Data-analytiikka ohjelma	Miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksessä?
SAS Advanced Analytics	<ul style="list-style-type: none">- Mallien ja analyysien luominen tarkempia diagnooseja varten- Ennustavat datan avulla hoidon toimivuutta- Luodaan 3D-malleja tarkempaa tarkastelua varten- Kuvien analysoimiseen

6 Yhteenveto

Tämän opinnäytetyön tarkoituksena oli perehtyä syvällisemmin dataan ja data-analytiikkaan ja tutkia miten data-analytiikkaa hyödynnetään eri yrityksissä tänä päivänä. Tarkastelin tätä aihetta siltä osin, kun se oli ilman käytännön omakohtaista kokemusta mahdollista, vastaamalla pääkysymyksiin, miten data-analytiikkaa hyödynnetään tällä hetkellä yrityksissä sekä mitä data-analytiikan ohjelmistoa yrityksissä käytetään. Materiaali näihin case-tutkimuksiin löydettiin yritysten omille internetsivuille tehdyistä julkaisuista, esitteilyvideoista sekä data-analytiikkaohjelmistojen omilta sivuilta.

Opinnäytetyössä on tarkasteltu data-analytiikkaa käytännön esimerkein neljän case-tutkimuksen kautta. Tarkasteluun valikoitui suuria yrityksiä eri aloilta, jotka eri tavoin hyödyntävät data-analytiikkaa päivittäisessä toiminnassaan. Volvo ja Mack trucks hyödyntävät data-analytiikkaa rekkaliikenteessä (SAS 2020) ja akkujen kapasiteetin kehittämisessä (Vinge, R., Volvo Group 2020). Heathrow'n lentoasema oli mielestäni hyvä esimerkkitaupaus, koska se käyttää data-analytiikkaa moneen eri tarkoitukseen, kuten esimerkiksi ihmismäärän "hallitsemiseen" ja lentoliikenteen ohjaamiseen (Microsoft 2018). Kotimaiseksi yritykseksi valitsin KONE:en, sillä sen toiminta näkyy meidän kaikkien suomalaisten elämässä. KONE hyödyntää data-analytiikkaa hissien päivittäisen toiminnan ylläpitämisessä (Salesforce). Neljänneksi esimerkiksi valitsin UMC Amsterdamin, joka on ottanut data-analytiikan käyttöön taistelussa syöpää vastaan (SAS 2020).

Dataan ja data-analytiikkaan painottuneessa aineistossa korostuu jatkuvasti, kuinka datan määrä on lisääntynyt ja moninkertaistunut vuosien saatossa. Tällä hetkellä trendit eivät näytä hidastumisen merkkejä, vaan datan määrä jatkaa eksponentiaalista kasvuaan. (Sedkaoui 2018, 8.) Jokainen meistä tuottaa päivittäin massiivisen määrän dataa esimerkiksi käyttämällä sosiaalista mediaa, älypuhelimia tai muuten vaan asioimalla päivittäistavarakaupoissa. Näistä muodostunutta informaatiota yritykset voivat käyttää omiin tarkoituksiinsa, kuten esimerkiksi oman toimintansa kehittämiseen ja asiakaskokemuksen parantamiseen.

Tutkimusta tehdessä huomasin, että case-yritysten data-analytiikan käytössä sekä tietopuolesta läpi käydyssä informaatioissa, on havaittavissa samoja asioita. Mainitsin tietopuolesta, kuinka paljon nopeammin tietokoneet pystyvät käsittelemään dataa kuin ihmiset. (Sedkaoui 2018, 103.) Tästä saatiin monia esimerkkejä case-tutkimusten aikana, kuten hybridi- ja sähköajoneuvojen datan analysointi, nopeampien ja tarkempien lääke diagnostiikoiden tekeminen ja lentokenttätoiminnan tuottaman datan analysointi. Mainitsin, kuinka data-analytiikka auttaa tekemään yrityksen toiminnasta tehokkaampaa, kun pidetään

resurssit töissä siellä missä ne ovat tehokkaimmillaan. (Sedkaoui 2018, 61.) Samalla kävin läpi, kuinka erilaisten sensoreiden tuottaman datan analysointi auttaa yrityksiä reagoimaan nopeasti ja jopa ennakoimaan ongelmatilanteita (Sharda ym. 2017, 444–445). Volvo ja KONE ovat huomanneet tämän myös omassa hissi- sekä ajoneuvotoiminnassaan. He seuraavat kalustonsa kuntoa ja ovat valmiina huoltamaan ja ennakoimaan suurempia ongelmatilanteita. (SAS 2020; Vinge, R., Volvo Group 2020; Salesforce).

7 Pohdinta

Tutkimuksen aiheena oli data-analytiikka ja miten sitä voidaan hyödyntää tämän päivän yritystoiminnassa. Tarkoituksena oli selvittää, minkälaisessa roolissa data-analytiikka on tällä hetkellä ja miten sitä käytetään tämän päivän yrityksissä. Tutkimuksesta selvisi, että yritykset pyrkivät data-analytiikan avulla tehostamaan toimintaansa esimerkiksi lyhentämällä ja ennakoimalla huolloista aiheutuneita pysähtymisiä. Data-analytiikka on myös osallisena tarkemmissa ja nopeammassa syöpädiagnooseissa. (SAS 2020.)

Data-analytiikkaan liittyvä aihe valikoitui aiheeksi, sillä se jäi koulun erikoistumisopinnoista kiinnostavimpana mieleen. Kävimme kurssilla muun muassa läpi, että data-analytiikan rooli on kasvanut työpaikoilla. Jos siis aihe kiinnostaa ja hakee itselleen oikeaa osaamista ja ammattitaitoa, voivat tulevaisuuden uramahdollisuudet olla valoisat. Koska opinnäytetyön tekeminen on suhteellisen iso urakka, ajattelin, että sen tekeminen aiheesta, joka edes jollakin tapaa kiinnostaa, helpottaisi urakan etenemistä. Oma osaaminen data-analytiikan kurssilta on vielä alkutekijöissä, joten ajattelin, että opinnäytetyön tekeminen aiheesta antaisi paremman kuvan itselle, kuinka kiinnostunut olen aiheesta. Kaksi karpästä yhdellä iskulla.

Tutkimusmenetelmäksi valikoitui case-tutkimus. Koin, että case-tutkimuksen avulla saa mahdollisimman selkeästi annettua käytännön esimerkkejä, jolloin aiheen sisäistäminen olisi mahdollisimman yksinkertaista ilman aikaisempaa käytännön kokemusta. Alun perin case-tutkimusten lisäksi oli tarkoituksena esittää pari kysymystä alan parissa työskenteleville. Näitä kahta olisi voinut sitten verrata toisiinsa. Vastauksien puuttumisen myötä jouduin kuitenkin jättämään tämän pois ja keskittymään vain kirjallisuudesta löytyvään informaatioon.

Tutkimus oli luotettava ja vastaukset pääkysymyksiin saatiin. Tutkimuksesta olisi kuitenkin saanut luotettavamman ja laajemman, jos tarkasteluun olisi valittu useampi yritys, sekä useampi data-analytiikkaohjelma. Jos työhön olisi saanut myös lisättyä haastattelukysymykset alalla työskenteleviltä, niin kuin olin alkuperäisesti suunnitellut, olisi työstä tullut vielä luotettavampi. Haastattelujen avulla olisi saanut enemmän käytännön näkökulmia data-analytiikan käytöstä.

Ennen työn aloittamista oma osaaminen aiheesta oli aivan alkutekijöissä. Koska aihe on suhteellisen monimutkainen, odotin kirjallisuuden olevan vaikeasti luettavaa. Olen tätä mieltä nyt myös työn loppuvaiheessa. Aiheesta löytyy suhteellisen paljon kirjallisuutta,

mutta suuri osa siitä on ihmisille, joilla on aikaisempaa kokemusta ja ammattitaitoa aiheesta. Aloittelijana jouduin käymään läpi paljon materiaalia, jotta löysin kirjallisuutta, jota ymmärsin riittävän hyvin kirjoittaakseni aiheesta. Kirjallisuus oli myös hyvin isolla painolla englanninkielistä. Tämä ei itsessään olisi ongelma, mutta vaikea ja uusi aihe vieraalla kielellä vaati työtunteja. Ainoa suomenkielinen materiaali, jota käytin työssä oli muutamat artikkelit ja viranomaisten tekemät julkaisut. Aihealueen vaikeus aiheutti kirjallisuuden kanssa suurimmat haasteet työn tekemiselle.

Mielestäni tällä työllä saa yleisen käsityksen siitä, millainen rooli data-analytiikalla on ja miten sitä hyödynnetään erilaisissa yrityksissä. Jos aihetta haluaa lähteä tutkimaan enemmän, syvemmän ja yksityiskohtaisemman ymmärryksen saavuttamiseksi, vaatii se data-analytiikkaohjelmien käyttämistä. Kirjallisuuden perusteella voi saada ympäröivään kuvan ja ymmärryksen aiheesta, mutta asian syvällisempi ymmärtäminen vaatii mielestäni omakohtaista kokemusta ja ammattitaitoa.

Yhteys tietoperustan ja tutkimuksen välillä on havaittavissa ja sillä saavutettiin se, mitä alun perin lähdettiin hakemaan.

Lähteet

Ahmed, F. 2018. How KONE Uses Data Analytics with Event-Driven Compute. IBM. Luettavissa: <https://www.ibm.com/cloud/blog/how-kone-uses-data-analytics-and-event-driven-maintenance-to-keep-cities-in-motion> Luettu 15.4.2021

AWAcademy 2021. Mitä data-analytiikka on ja miksi data-analytiikkakoulutus kannattaa? Luettavissa: <https://www.awacademy.fi/stories-insights/alanvaihto/ura-it-alalla-mita-on-data-analytiikka> Luettu 1.4.2021

Davenport, T & Harris, J. 2007. Competing on Analytics, The New Science of Winning. Harvard Business Review Press

Digi- ja väestövirasto. 2021. Avoindata.fi Luettavissa: <https://www.avoindata.fi/fi> Luettu 6.5.2021

EMC Education Services. 2015. Data Science and Big Data Analytics : Discovering, Analyzing, Visualizing and Presenting Data. John Wiley & Sons, Incorporated

Hämäläinen, V-P. 2021. Ehkä jopa 32 000 Vastaamon potilaan tiedot ilmestyivät viime yönä Tor-verkkoon – poliisi: ”Emme tiedä monenko käsissä tietokanta on.” Luettavissa: <https://yle.fi/uutiset/3-11757676> Luettu 6.5.2021

Hämäläinen, V-P. 2021. Uudet tiedot: Vastaamon potilaiden tiedot olivat ehkä jopa vuosia suojaamatta netissä – tietoturva-asiantuntija: ”Älyvapaata” Luettavissa: <https://yle.fi/uutiset/3-11750220> Luettu 6.5.2021

Ilmatieteen laitos. 2021. Avoin data avattavat aineistot. Luettavissa: <https://www.ilmatieteenlaitos.fi/avoin-data-avattavat-aineistot> Luettu 6.5.2021

Juniper Research. 2016. 'Internet of Things' connected devices to almost triple to over 38 billion units by 2020 Luettavissa: <https://www.juniperresearch.com/press/iot-connected-devices-to-triple-to-38-bn-by-2020> Luettu 10.4.2021

KONE. 2021. KONE in brief. Luettavissa: <https://www.kone.com/en/company/> Luettu 15.4.2021

Library Of Congress. 2021. About the Library Luettavissa: <https://www.loc.gov/about/> Luettu 23.3.2021

Liebowitz J. 2013. Business Analytics : An Introduction. Auerbach Publishers, Incorporated.

Ma, A & Gilbert, B. 2019. Facebook understood how dangerous the Trump-linked data firm Cambridge Analytica could be much earlier than it previously said. Here's everything that's happened up until now. Luettavissa: <https://www.businessinsider.com/cambridge-analytica-a-guide-to-the-trump-linked-data-firm-that-harvested-50-million-facebook-profiles-2018-3?r=US&IR=T> Luettu 23.3.2021

Markets And Markets 2020. Big Data Market. Luettavissa: <https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html> Luettu 6.5.2021

Marr B. 2016. Big Data For Small Business For Dummies. John Wiley & Sons, Incorporated. New York

Marr B. 2021. The Biggest Trends In Big Data Analytics Right For 2021 Luettavissa: <https://www.bernardmarr.com/default.asp?contentID=2161> Luettu 1.4.2021

Marr B. 2020. What Is Big Data? Luettavissa: <https://bernardmarr.com/default.asp?contentID=766> Luettu 23.3.2021

Marr B. 2020. The 10 Best Data Analytics And BI Platforms And Tools In 2020. Luettavissa: <https://www.bernardmarr.com/default.asp?contentID=2057> Luettu 1.4.2021

Marr B. 2020. How Can Small Businesses Use Big Data? Here Are 6 Practical Examples Luettavissa: <https://www.bernardmarr.com/default.asp?contentID=1442> Luettu 1.4.2021

Marr B. 2020. The Vital Role Of Big Data In The Fight Against Coronavirus Luettavissa: <https://www.bernardmarr.com/default.asp?contentID=2034> Luettu 1.4.2021

Marr B. 2021. The Biggest Trends in Big Data And Analytics Right For 2021. Forbes. Luettavissa; <https://www.forbes.com/sites/bernardmarr/2021/02/22/the-4-biggest-trends-in-big-data-and-analytics-right-for-2021> Luettu 5.4.2021

Microsoft. 2018. Heathrow prepares rather than reacts: uses data to deliver airport calm. Luettavissa: <https://customers.microsoft.com/en-us/story/heathrow-airport-travel-transportation-powerbi-azure> Luettu 10.4.2021

Microsoft 2021. Why Power BI. Luettavissa: <https://powerbi.microsoft.com/en-us/why-power-bi/> Luettu 6.5.2021

Mimaroglu, A 2015. Optimize Your Social Media Ad Spend With Advanced Targeting Options. NeilPatel, blog Luettavissa: <https://neilpatel.com/blog/optimize-your-social-media-ad-spend-with-advanced-targeting-options/> Luettu 6.5.2021

Salesforce. KONE drives growth, innovation, and proactive service with Salesforce insights. Luettavissa: <https://www.salesforce.com/eu/customer-success-stories/kone/> Luettu 15.4.2021

SAS 2021. Advanced Analytics Software, IDC Research Luettavissa: https://www.sas.com/fi_fi/solutions/analytics.html Luettu 10.4.2021

SAS 2021. Big Data Analytics – What it is and why it matters Luettavissa: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html Luettu 1.4.2021

SAS 2021. Machine Learning – What it is and why it matters Luettavissa: https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention. Luettu 5.4.2021

SAS 2020. UMC Amsterdam Luettavissa: https://www.sas.com/en_gb/customers/amsterdam-umc.html Luettu 15.4.2021

SAS 2020. Volvo trucks and Mack trucks Luettavissa : https://www.sas.com/en_gb/customers/volvo-trucks-mack-trucks.html Luettu 10.4.2021

Sedkaoui S. 2018. Data Analytics and Big Data. John Wiley & Sons, Incorporated

Sharda, R., Delen, D. & Turban, E. 2017. Business Intelligence, analytics, and data science : a managerial perspective. Harlow, England : Pearson

Talouselämä 2013. Big Data muuttaa maailmaa Luettavissa: <https://www.talouselama.fi/kumppaniblogit/big-data-muuttaa-maailmaa/6e3988d0-e07e-35ea-b52c-dc3e31a91394>. Luettu 23.3.2021

Terveyden ja hyvinvoinnin laitos 2020. Koronavilkku Luettavissa: <https://koronavilkku.fi/> Luettu 23.3.2021

Terveyden ja hyvinvoinnin laitos 2020. Koronavilkku – Usein kysytyä Luettavissa: <https://koronavilkku.fi/ukk/> Luettu 23.3.2021

Tilastokeskus 2018. 5. Big Data. Luettavissa: https://www.stat.fi/til/icte/2018/icte_2018_2018-11-30_kat_005_fi.html. Luettu: 23.3.2021
Tilastokeskus 2021. Avoimet tietokanta-aineistot. Luettavissa: <https://www.stat.fi/org/avoindata/pxweb.html> Luettu 6.5.2021

Vinge Rikard. 2020. Using data to optimize batteries for electric trucks. Volvo Group. Luettavissa: <https://www.volvogroup.com/en/news-and-media/news/2020/dec/optimizing-electric-vehicles-with-data-analytics.html> Luettu 10.4.2021