

Jenni Kääriäinen

**BIG DATAN KÄYTTÖ TYÖNTEKIJÖIDEN
SEURANNASSA JA REKRYTOINNISSA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2019

TIIVISTELMÄ

Kääriäinen, Jenni

Big datan käyttö työntekijöiden seurannassa ja rekrytoinnissa

Jyväskylä: Jyväskylän yliopisto, 2019, 36 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja(t): Makkonen, Pekka

Big data on yksi IT-alan suosituimpia termejä tällä hetkellä ja monet yritykset haluavat löytää big datassa piilevän hyödyn. Tavalliseen data-analyysiin verrattuna big dataa kerätään ja analysoidaan suurempia datamääriä sekä rakenteetontakin dataa. Tämän tutkielman tarkoituksena on tutkia big datan käyttöä työntekijöiden seurannassa ja rekrytoinnissa, erityisesti hyötyjä joita big datan käyttö voi tuoda henkilöstönhallinnalle. Työntekijöiden seuranta on yrityksissä tehty vuosikausia ja monet nykyaikaiset muutokset työnteossa kuten etätyöt sekä teknologian kehittymisen tuomat uudet haasteet työssä vaativat seurannallekin uusia askeleita. Rekrytointi on aina ollut yritykseen suuresti vaikuttava ja haastava tehtävä, johon yritykset mielellään ottavat käyttöön tehokkaampia keinoja.

Big datan käsittelemiseen liittyy sekä sen hallinta että analysointi ja työkaluja näihin molempiin löytyy useita. Kirjallisuutta löytyy paljon sekä big datasta että henkilöstönhallinnasta, mutta kirjallisuutta big datan hyödyntämisestä seurannassa ja rekrytoinnissa löytyy vielä suhteellisen vähän.

Tutkimuksessa huomattiin, että big data hyödyttää yrityksiä henkilöstönhallintaa eniten päätöksenteossa, kustannusten vähentämisessä, tyytyväisyyden parantamisessa sekä sopivien työntekijöiden valinnassa. Erityisesti seurannassa big datan avulla voidaan suorittaa tarkempaa analyysia suurilla datamäärillä, seurata työntekijöiden terveyttä, suoriutumista sekä tyytyväisyyttä. Rekrytoinnissa big datan käyttö voi helpottaa sopivien työntekijöiden löytämistä ja valitsemista, parantaa yrityksen brändiä, tehostamalla hakemusten läpikäyntiä sekä parantaa rekrytointiprosessia.

Yrityksille voi siis olla runsaasti hyötyjä big datan käytöstä. Kun otetaan huomioon datan tietoturva, työntekijöiden oikeudet sekä yritystä velvoittavat lait sekä big dataa käsittelevän henkilökunnan ammattimaisuus, on mahdollisuus parantaa sekä työntekijöiden seuranta että rekrytointia big datan avulla.

Asiasanat: Big data, Henkilöstönhallinta, Rekrytointi, Seuranta

ABSTRACT

Kääriäinen, Jenni

Big data use in employee monitoring and recruiting

Jyväskylä: University of Jyväskylä, 2019, 36 pp.

Information Systems, Bachelor's Thesis

Supervisor(s): Makkonen, Pekka

Big data is one of the most popular terms in IT right now and multiple companies are wanting to find the benefits hidden in big data. Compared to usual data-analysis big data is about collecting big amounts of data, storing and analyzing it and even collecting unstructured data. The purpose of this study is to examine the use of big data in monitoring and recruiting employees, especially the benefits that the use of big data can bring to the context of Human resources. Companies have monitored employees for decades and many recent changes in the way we work such as remote working and the progress of technologies have brought new requirements at work that require new steps to be taken in monitoring. Recruitment has always been one of the most influential and challenging missions of a company. That is why many companies would like to find more efficient ways to recruit.

Data management and data analysis relate to the use of big data and tools to these are plenty. There is a lot of literature relating to both big data and HR but there are still relatively few studies to the use of big data in employee monitoring and recruitment.

In this study it was discovered that big data can benefit HR in decision making, lowering costs, improving employee satisfaction and choosing the best applicants. Especially in monitoring big data can help with getting more precise analysis with big volumes, monitoring employee health, performance and satisfaction. In recruiting use of big data can help in finding and choosing more suitable employees, improving company brand, making going through applications more efficient and improving recruitment process.

Companies can benefit a lot from the use of big data and when we factor in information security, employee rights and laws that bind the company and the professionalism of the big data employees, there is a chance to improve employee monitoring and recruitment with big data.

Keywords: Big data, Human resources, Recruitment, Monitoring

TAULUKOT

TAULUKKO 1: Big datan määritelmä	10
TAULUKKO 2 : Big datan käytön hyödyt	19
TAULUKKO 3: Hyödyt datan käsittelyn vaiheen mukaan	28
TAULUKKO 4 : Hyödyt henkilöstöhallinnon alueiden mukaan	28
TAULUKKO 5: Hyödyt arvon tuoton perusteella viitekehyksellä	29
TAULUKKO 6: Hyödyt arvon kautta.....	29

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
TAULUKOT	4
1 JOHDANTO.....	6
2 BIG DATA JA SEN TEKNOLOGIAT	9
2.1 Big datan määritelmä ja ominaisuudet.....	9
2.2 Big datan tekniikat ja teknologiat.....	10
2.2.1 Big datan hallinta	11
2.2.2 Big datan analyysi	13
3 BIG DATAN KÄYTÖSTÄ YRITYKSISSÄ	16
3.1 Big datan käyttö liiketoimessa	16
3.2 Big datan käytöstä yrityksen sisällä	17
3.3 Huomioitavaa big datan käytössä.....	17
3.4 Big datan hyödyistä.....	18
4 BIG DATAN KÄYTTÖ TYÖNTEKIJÖIDEN SEURANNASSA JA REKRYTOINNISSA.....	20
4.1 Työntekijöiden seuranta ja rekrytointi	20
4.2 Big datan käyttö työntekijöiden seurannassa ja rekrytoinnissa.....	22
4.2.1 Työntekijöiden seuranta big datan avulla	22
4.2.2 Rekrytointi big datan avulla	25
4.2.3 Big datan hyödyt työntekijöiden seurannassa ja rekrytoinnissa	27
5 YHTEENVETO	30
LÄHTEET	33

1 JOHDANTO

Big data on yksi merkittävimmistä teknologioista mitä tänä päivänä löytyy. Vuonna 2010 McKinsey Quarterly listasi big datan kymmenen parhaan joukkoon tärkeimpien businessstrendien joukossa. Vuotta myöhemmin McKinsey julkaisikin raportin, jossa kuvaili big datan tuomia hyötyjä ja teknologioita (Manyika ym., 2011). Raportissa big data kuvataan tulevaisuuden kilpailun ja kasvun tekijäksi yrityksissä. Big data tuo läpinäkyvyyttä, mahdollistaa kokeilua, auttaa kustomoinnissa, päätöksenteossa ja innovoinnissa.

Yritykset haluavat käyttää big dataa kilpailuedukseen ja ovat huomanneet big datan mahdollisuudet. IDC:n eli International data corporationin (2014) mukaan dataa oli vuoteen 2013 mennessä 4.4 tsettatavua ja vuoteen 2020 mennessä dataa on ennustettu olevan 44 tsettatavua (IDC, 2014). Big datan markkinat vuonna 2012 olivat noin 28 miljardin dollarin kokoiset ja tänä päivänä sen markkinat ovat luultavasti kasvaneet paljon suuremmiksi (Simon, 2013). Yritykset ovat huomanneet, että tieto tuo valtaa, kuten myös monet tilastoja tekevät yritykset. Esimerkiksi Statistan premium-käyttäjänä pääsee tutkimaan yli miljoonaa Statistan tekemää tilastoa, mutta premium-tili maksaa jopa 50 dollaria kuussa. Data voi siis olla erittäin arvokasta ja monet yritykset eivät tahdo hukata mitään dataa, josta voisi mahdollisesti olla hyötyä.

Big data ei Bain & Companyn (2013) tutkimuksen mukaan kuitenkaan ollut vielä 2013 kovin monella yrityksellä johdon työvälineenä käytössä, mutta ne yritykset, joilla big datan analytiikkaa oli käytössä, olivat keskimääräistä tyytyväisempiä big dataan kuin muihin tutkimuksen työkaluihin, kuten strategiseen suunnitteluun (Rigby & Bilodeau, 2013). Tällöin onkin odotettavissa, että big datan suosio johdon työvälineenä kasvaa.

Työntekijöiden seurantaan ja rekrytointiin liittyy paljon haasteita, joiden kanssa yritysten henkilöstöhallinta sekä johto painivat. Seurannasta voidaan saada monia hyötyjä kuten parantaa työntekijöiden turvallisuuskäyttäytymistä (Rivera, 2015). Gartner arvioikin 50% parannuksen turvallisuuskäyttäytymisessä yrityksillä, jotka seurasivat ainakin kolmea työntekijöidensä turvallisuuskäyttäytymiseen vaikuttavaa tekijää. Työntekijöiden tehokkuutta on vaikeampi arvioida, kun IWG tutkimuksen mukaan 70% ammattilaisista tekee töitä kotonaan ainakin

kerran viikossa (Browne, 2018). Etätyöskentelystä seuraa useita haasteita kuten työntekijöiden sitoutuneisuuden varmistaminen sekä tietoturvan varmistaminen, jos työntekijä käyttää omia laitteitaan työnteossa. Monessa yrityksessä kehoitetaan käyttämään yrityksen laitteistoa etätöissäkin. Seuranta voi olla ainoa keino varmistaa yrityksen työntekijöiden sitoutuneisuus työntekoon.

Myös uudet teknologiat kuten puettavat teknologiat sekä älylaitteet mahdollistavat uudenlaisia seurantakeinoja, joilla voidaan parantaa seurannan tehokkuutta sekä kehittää uusia seurantakeinoja. Tilastoja keräävän Statistan mukaan verkkoon yhdistettyjä puettavia laitteita oli vuonna 2016 325 miljoonaa ja arvion mukaan vuonna 2021 niitä olisi jo 929 miljoonaa. Näistä laitteista kerättävän datan määrä on valtava, mutta niiden datasta voidaan big datan keinoin löytää hyödyllistä tietoa (Statista, 2019). Älypuhelimilla on maailmanlaajuisesti arvioitu olevan 3 miljardia käyttäjää (Takahashi, 2018).

Yrityksillä kuluu rekrytointiin paljon resursseja. Yhden uuden työntekijän rekrytointi aiheuttaa keskimäärin 4000 dollarin kustannukset yritykselle Yhdysvalloissa (Mariotti, Robinson & Esen, 2017) ja Suomessa kolme työpaikkailmoitusta verkossa maksaa keskimäärin 1550 euroa (Eilakaisla, 2019). Monet yritykset ilmoittavat verkossa hakevansa työntekijää sekä käyttävät verkkohakemuksia hakemusten vastaanottamiseen. Verkkorekrytointi onkin helpottanut hakemista entisestään. Myös Suomen työttömyysetuuden vastuut osaltaan vaikuttavat siihen, että ihmiset saattavat hakea työpaikkoihin, joihin eivät oikeasti halua tai joihin eivät ole päteviä. Tämä vaikeuttaa sopivan työntekijän löytämistä.

Tutkielmani tarkoitus on tutkia kirjallisuuden avulla tapoja, joilla työntekijöistä tai rekrytoitavista voidaan kerätä ja käsitellä dataa, millaista dataa sekä mitä hyötyjä tästä datasta ja sen analysoinnista on. Erityisesti keskityn big datan teknologioihin ja niistä seuraaviin hyötyihin erityisesti yrityksen työntekijöihin liittyen.

Tutkimusongelmiksi muodostui:

- "Millä tavoin big dataa voidaan hyödyntää työntekijöiden rekrytoinnissa ja seurannassa?"
- "Mitä teknologiaa käytetään datan hallintaan ja analysoimiseen työntekijöiltä?"

Tutkielmassa tutkin, miten big dataa voidaan käyttää työntekijöiden seurannassa ja rekrytoinnissa hyödyksi, sekä big dataan liittyviä teknologioita. Analysoin teknologioita ja tekniikoita big dataan liittyen sekä big datan potentiaalia työntekijöiden seurannan ja rekrytoinnin helpottamiseksi. Kirjallisuutta etsin useiden eri tietokantojen avulla. Tietokantoja, joista kirjallisuutta etsin oli muun muassa JYKDOK, IEEE, Google Scholar sekä Scopus. Hakusanoina käytin muun muassa "big data", "big data AND employee", "big data AND workforce", "big data AND human resources" sekä "workplace monitoring".

Seuraavaksi määrittelen big datan ja käyn läpi tärkeimmät big dataan liittyvät tekniikat ja teknologiat joiden avulla big dataa hallintaan ja analysoidaan. Kolmas luku käsittelee big datan käyttöä sekä hyötyjä, joita yritys voi saada big datan käytöstä. Käsittelem myös suppeasti lakeja muita seikkoja, jotka voivat

vaikuttaa big datan käytössä. Käsittelen myös, miten big datasta saatavia hyötyjä voidaan lajitella.

Neljännessä kappaleessa käsittelen nykytilaa työntekijöiden seurannassa ja rekrytoinnissa sekä miten nämä big datan teknologiat ja tekniikat voivat tuottaa yritykselle hyötyä näissä alueissa. Viimeisenä tiivistän johtopäätökset ja tulokset yhteenvedossa.

2 BIG DATA JA SEN TEKNOLOGIAT

Datan määrä kasvaa maailmassa kiihtyvästi ja sen käyttöön liittyy paljon kiinnostusta. Gandomin ja Haiderin (2015) mukaan termi tuli yleisesti käyttöön 2011. Datan analysoimisesta on puhuttu jo kauan yritysmaailmassa, jonka myötä myös big data on käsitteenä kiinnostanut tutkijoita ja yrityksiä jo muutamia vuosia. Silti IDC:n (2014) mukaan olemassa olevasta valtavasta datan määrästä hyödynnetään vain 0.5% (Regalado, 2013). Tähän syynä vaikuttaa olevan datan määrän valtava kasvu, dataa käsittelevien teknologioiden rajoitukset sekä se, että suuri osa datasta on rakenteetonta. Myös jotta yritykset voivat saada täyden potentiaalinen käytetyksi big datasta, on saatava houkutelua oikeat asiantuntijat yritykseen sekä osattava käyttää sopivia teknologioita.

Kuitenkaan yritysten johtajat eivät ole yhteisymmärryksessä big datan käsitteen määritelmästä. Harris Interactiven (2012) kyselyssä suurin osa johtajista kuvasi big datan liiketoiminnan datan määrän suureksi kasvuksi, seuraavaksi eniten big data määriteltiin sen määrää, monipuolisuutta ja nopeutta huomioivien teknologioiden kautta. Osa johtajista ymmärsi big datan tiedon tallennukseen liittyviksi vaatimuksiksi ja osa uusiksi datan lähteiksi (SAP, 2012). Tieteellisessä tutkimuksessaakin big datan määritelmä on usein epäselvä ja siksi seuraavassa luvussa määrittelen termin tähän tutkimukseen kirjallisuuden kautta.

2.1 Big datan määritelmä ja ominaisuudet

Big datan määritelmä ei ole aina kovin selkeä. Joskus big data määritellään suuriksi datamassoiksi, joiden käsittely tavallisilla tietokantatyökaluilla on mahdotonta (Manyika ym., 2011) Tämä määritelmä jättää vielä paljon tilaa tulkinnalle, sillä se ei osoita mitään tiettyä määrää dataa tai määrittele mitä tavalliset työkalut ovat. Työkalujen kehittyessä ja datan määrän kasvaessa on mahdotonta itsessään määritellä aina tarkasti mitä big data on.

Toisaalta big data voidaan määritellä myös olevan suuria määriä rakenteetonta dataa (Simon, 2013). Rakenteeton data on rakenteellisen, yleensä tietokantoihin järjestetyn, datan vastakohta. Sillä tarkoitetaan dataa, jolla ei ole rakennetta tai jota ei voida jäsentää. On olemassa myös semi-rakenteellista dataa, jota on esimerkiksi XML eli Extensible Markup Language. Rakenteeton data on riippumatonta, sekalaista, tekstintäyteistä ja Simonin (2013) mukaan eniten big dataa muistuttavaa. Se voi olla siis tekstiä, kuvia tai videota, jota löytyy esimerkiksi sosiaalisesta mediasta tai kovalevyiltä yksittäisinä tiedostoina.

Kuitenkin yleensä big datan yhteydessä otetaan esille sen kolme ominaisuutta: määrä, monipuolisuus sekä nopeus (Laney, 2001). Englanniksi näitä kuvaillaan kolmeksi V:ksi eli volume, velocity ja variety. Useimmat big dataan keskittyvät tutkimukset viittaavat yhä näihin Laneyn vuosituhannen alussa kuvaamiin big datan ominaisuuksiin ja nämä ominaisuudet hyvin määrittelevät mitä

on big data. Big data on suurta, monipuolista sekä nopeasti kasvavaa. Suuruudella tarkoitetaan datan suurta määrää, yleensä siis puhutaan tera- tai petatavuista (Gandomi & Haider, 2015). Tämänkin tosin on nopeasti muuttuva määritelmä, koska tallennuskyky kehittyy koko ajan ja suuruudella tarkoitetaan tulevaisuudessa isompia datan määriä kuin nyt. Monipuolisuus viittaa eri tyyppisendatan käyttöön, eli jo mainittuihin rakenteettomaan, että rakenteelliseen ja semi-rakenteelliseen dataan. Varsinkin rakenteetonta dataa syntyy paljon ja sen käsitteleminen on haasteellisempaa kuin rakenteellisen. Datan luomisen ja käsittelyn nopeus on yksi big datan ominaisuuksista (Gandomi & Haider, 2015). Datan syntymistä on nopeuttanut älypuhelinien ja erilaisten sensoreiden yleistyminen. Taas erityisesti nämä tekijät ovat vaikuttamassa rakenteettoman datan räjähtävään määrän kasvuun.

Big data -termin suomenkielisenä käännöksenä käytetään joskus termiä "massadata", mutta yleisesti tutkimuksessa ja informaatioteknologian ammattilaisten keskuudessa suositaan englanninkielistä termiä. Tämän takia käytän englanninkielistä termiä. Big datan määritelmäksi tässä tutkimuksessa voidaan siis ymmärtää, että big dataa ei voida käsitellä tavanomaisilla työkaluilla, se on useimmiten rakenteetonta ja sillä on Laneyn (2001) kuvaamat kolme ominaisuutta (määrä, monipuolisuus ja nopeus).

TAULUKKO 1: Big datan määritelmä

Lähde	Määritelmä
Manyika ym., 2011	Big datan käsittely on tavallisilla tietokantatyökaluilla mahdotonta.
Simon, 2013	Big data on usein myös rakenteetonta dataa.
Gandomi & Haider, 2015	Big data on suurta, sekä rakenteellista että rakenteetonta ja sitä syntyy nopeasti.
Laney, 2001	Määrä, monipuolisuus ja nopeus (3 V:tä eli volume, variety and velocity).

2.2 Big datan tekniikat ja teknologiat

Jotta big datan todellinen hyöty saadaan esille, on datasta saatava esiin päätöksentekoon oleellista tietoa. Tämä prosessi voidaan jakaa kahteen alempaan prosessiin: Datan hallintaan ja analyysiin (Gandomi & Haider, 2015). Datan hallintaan kuuluu datan keräämistä, tallentamista ja valmistelua analysointia varten ja analyysiin kuuluu analyysi, mallintaminen ja tulkitseminen.

Big datan hallintaan ja analyysiin on olemassa lukuisia teknologioita ja tekniikoita. Dataa käsiteltäessä tilastotieteen menetelmät ovat aina isossa roolissa, mutta jätän tässä tutkielmassa tilastollisten menetelmien käsittelyn vain maininnan tasolle. Keskityn informaatioteknologian mahdollistamiin teknologioihin ja tekniikoihin. Näitä ovat muun muassa tiedonlouhinta, luonnollisen kielen prosessointi, uudenlaiset tietokannat kuten NoSQL sekä big datan käsittelyyn

liittyvät ohjelmistot kuten Hadoop sekä koneoppiminen. Seuraavat alaluvut avaavat näitä datan hallinnan sekä datan analyysin keinoja.

2.2.1 Big datan hallinta

Big datan hyödyntäminen lähtee datan keräämisestä ja esivalmistelusta. HR-tietojärjestelmissä löytyy paljon dataa sekä työntekijöistä, heidän suoriutumistaan, palkoistaan että myös yrityksen työtehtäviin hakeneista (Angrave, Charwood, Kirkpatrick, Lawrence & Stuart, 2016). Dataa löytyy muun muassa työntekijöiden työhistoriasta, taidoista, koulutuksesta, demografisista tiedoista, työtunneista, palkasta, myynneistä ja asiakastunneista.

Simon (2013) toteaa sensoreiden ja nanoteknologian olevan yksi tärkeimmistä big datan keräysteknologioista. Dataa kerätään sensoreiden lisäksi myös älypuhelimilla, ohjelmistojen logeista ja mikrofoneista (Philip Chen & Zhang, 2014). Älypuhelimien jokapäiväistyminen on lisännyt paljon datan määrää ja tehnyt datan käsittelystä yhä ajankohtaisempaa. Kirjallisuudessa käsitellään myös puettavien teknologioiden käyttöä datan keräämiseen (Sedighi-Maman, Alamdari Yazdi, Cavuoto & Megahed, 2017). Puettavilla teknologioilla saadaan varsinkin terveydentilaan liittyvää dataa. CCTV eli valvontakameroiden dataa voidaan kerätä suuria määriä ja analysoida seuraavassa kappaleessa kuvatuin tavoin (Gandomi & Haider, 2015). Monissa yritysten tiloissa on valvontakamerat ja niiden keräämästä videodatasta voisi olla hyötyä muutenkin kuin valvonnassa. Technode (2017) uutisoinnin mukaan Kiinassa on kolmen vuoden sisällä jo 626 miljoonaa valvontakameraa ja Beijing ilmoitti 2015, että kaupunki on sataprosenttisesti katettu kameroilla.

Chen ja Zhang (2014) mukaan big datan käyttöä hankaloittaa usein huono data, joka voi väärentää analyysin tulosta. Tämän takia dataa joskus ”esivalmistellaan” muun muassa dataa puhdistamalla, integroimalla, muodonmuutoksella tai vähentämällä. Saman tutkimuksen mukaan big datan tallentamiseen liittyy läheisesti ongelma, että prosessointi on nopeaa mutta I/O on köyhää (Philip Chen & Zhang, 2014). Tämä tarkoittaa sitä, että Mooren lain mukaan prosessorien tehokkuus parantuu kaksinkertaiseksi joka 18 kuukausi ja siten datan määräänkin lisääntyy suurella vauhdilla, mutta datan prosessointi ei seuraa perässä yhtä nopeasti. Tähän puolijohdelevyt eli SSD:t tai vaiheenmuutosmuisti eli PCM saattavat olla avuksi (Philip Chen & Zhang, 2014).

Big dataan liittyvistä teknologioista puhuttaessa ensimmäisenä mieleen tulee Apachen Hadoop. Hadoop on tunnettu avoin ohjelmistoprojekti, joka kehittää big datan käsittelyyn liittyviä ohjelmistoja (Apache, 2019a). Hadoopin komponentit auttavat prosessoimaan ja säilömään isoja datamääriä (Simon, 2013). Hadoop toimii siis alustana big datan käsittelemiselle. Simon (2013) mainitsee sen hyväksi puoliksi joustavuuden, kyvyn käsitellä monenlaista dataa, helpon skaalautuvuuden, virheiden sietämisen sekä avoimen lähdekoodin. Hadoop-projekti koostuu monista komponenteista kuten Map/Reducesta ja Hadoopin hajatetusta tiedostojärjestelmästä eli HDFS:stä (Simon, 2013). Hadoopin ohjelmointiviitekehys Map/Reduce on suurien datamäärien keräämisen ja

prosessoinnin malli sekä toteutus. Se hajottaa isot ongelmat pienempiin ongelmiin kunnes ne ovat ratkaistavissa ja lopulta yhdistettävissä ison ongelman ratkaisuksi. Hapoodin osa HDFS taas on hajautettu tiedostojärjestelmä, joka säilyttää Hadoopin eri osien tuottamaa suurta määrää dataa.

NoSQL on tällä vuosituhannella suosiota kerännyt tietokantamalli. NoSQL ei ole perinteisellä tavalla relationaalinen, mutta ei myöskään täysin välttele SQL:ää (Philip Chen & Zhang, 2014). Nimi tulee sanoista "Not only SQL". NoSQL:än tarkoitus on vähentää big datan käyttöä rajoittavia ominaisuuksia. Neljä yleistä NoSQL tyyppiä ovat Avain-arvo, kolumnisukuun kuuluvat, dokumentti- ja graafitietokannat (Simon, 2013). HBase on Apachen suosittu NoSQL tietokanta (Philip Chen & Zhang, 2014). Apachen HBasen sivut kuvaavat kuitenkin sitä enemmän tietovarastoksi kuin tietokannaksi, koska siitä puuttuu monia relationaalisen tietokannanhallintasynteesin ominaisuuksia (Apache, 2019b). Tietovarasto sopii Apachen HDFS:än kanssa ja on kolumnisuku-tyyppinen (Simon, 2013). Muita Apachen projekteja on muun muassa Pig, Hive-infrastruktuuri ja Mahout-koneoppimisen algoritmit (Apache, 2019) ja muita NoSQL-tietokantoja on monia (Simon, 2013). Kun NoSQL:ssä huomattiin rajoitteita, kehitettiin NewSQL. NewSQL:n oli tarkoitus mahdollistaa NoSQL skaalautuvuus, mutta NoSQL:stä puuttuvilla SQL:n ominaisuuksilla.

Datajärvet ovat datan säiliöitä, joissa rakenteetontakin dataa voi säilyttää sellaisenaan ja suurellakin skaalalla (Amazon, 2019). Dataa voidaan analysoida järvessä esimerkiksi koneoppimisen ja SQL kyselyiden avulla. Tavallisiin tietovarastoihin verrattuna datajärvien etuna on, ettei ole tarvetta datan rakenteen määrittelylle, siirtelemiselle ja kalliille varastoratkaisuille (Miloslavskaya & Tolstoy, 2016). Datajärvet säilyttävät siis raakaa dataa ja analysoivat dynaamisilla keinoilla dataa. Datajärvet ovat tasaisia arkkitehtuuriltaan eli sen rajat tiedetään, mutta rakenne luodaan vasta luettaessa ja data erotetaan toisistaan identifioijilla ja saapumisjärjestyksellä. Datajärveen saapuvalla datalla määritetään identifioija sekä metadatatägit. Metadata on siis tietoa datasta, jota tallennetaan järveen, esimerkiksi saapumisjärjestys tai lisäysaika. Datajärvet voivat hyödyntää sekä NoSQL että SQL:ää.

Nopeasta datasta eli Fast datasta taas puhutaan, kuten kerätään reaaliaikaista dataa, jota olisi syytä analysoida heti (Miloslavskaya & Tolstoy, 2016). Nopean datan tarve voi syntyä jos esimerkiksi datan analysoinnin tulisi olla mahdollisimman reaaliaikaista, ettei analysoinnista saatava tieto vanhene. Tieto vanhenee varsin nopeasti muun muassa sijoitusmarkkinoilla. Nopea data voi liittyä massadataan läheisesti, mutta koetaan eri asiaksi kuin big data. Nopean datan käsittelemisessä tulee useimmiten ottaa huomioon prosessointi ja miten data säilytetään, eli mikä data tietokannoissa pitkän aikaa vai analysoidaanko dataa heti, kun sitä kerätään. Tällöin verrataan eräajo- ja virtaus-prosessointia (Miloslavskaya & Tolstoy, 2016).

Uusia teknologioita big datan hallintaan kehitetään koko ajan. Chen ja Zheng (2014) ehdottavat mahdollisiksi tuleviksi big datan teknologioiksi kvanttietokoneita, pilvilaskentaa sekä biolaskentaa.

2.2.2 Big datan analyysi

McKinsey (2011) luetteli ja kuvaili raportissaan tärkeimpiä teknologioita ja tekniikoita, joilla hallita ja analysoida big dataa. Myös Gandomi ja Haider (2015) kuvailivat tutkimuksessaan näitä teknologioita ja tekniikoita. Molemmista tutkimuksista löytyi samoja tärkeitä analyysin teknologioita. Analyysin keinoja on valtavasti, joten kaikkia ei kannata tässä tutkimuksessa käydä läpi. Käynkin eniten läpi useissa tutkimuksissa esiin tulleita edelleen käytettäviä teknologioita ja tekniikoita (Manyika ym., 2011) (Philip Chen & Zhang, 2014)(Gandomi & Haider, 2015).

Tiedonlouhinta on varmaan yhtä suosittu termi yritysmaailmassa kuin big data. Tiedonlouhinnalla tarkoitetaan menetelmiä, joilla suuresta määrästä dataa yritetään löytää merkityksellistä tietoa (Manyika ym., 2011). Siinä käytetään koneoppimisen ja tilastotieteen menetelmiä kuten assosiaatioiden sääntöjen oppimista, luokittelusta, regressiota ja klusterointia. Assosiaatioiden sääntöjen oppimisessa etsitään assosiaatioita muuttujien välillä ja mahdollista sääntöä assosiaatioon. Näin voidaan esimerkiksi löytää tuotteet, joita yleensä ostetaan yhdessä. Luokittelussa yritetään luokitella kategorioita testikategorioiden avulla. Regressio on tilastollinen tekniikka, jossa tutkitaan miten jotkin muuttujat voivat vaikuttaa tutkittavaan muuttujaan, esimerkiksi miten markkinat voivat vaikuttaa myyntimääriin. Regressiota käytetään usein ennustamaan tutkittavien muuttujien käyttäytymistä. Klusteroinnissa isompaa ryhmää jaetaan tietyn ominaisuuden mukaan pienempiin ryhmiin ja siten ”klustereihin”.

Muita tilastollisia tekniikoita big datan analyysiin on muun muassa A/B-testaus, datan fuusio sekä integrointi. A/B-testaus on tilastotieteen tekniikka, jolla verrataan testiryhmää useisiin muihin ryhmiin, jotta löydetään ryhmään vaikuttavia muuttujia (Manyika ym., 2011). Datan fuusiossa ja integraatiossa käytetään useita datalähteitä, jotta saadaan tarkempia tuloksia. Myös biologiaa käytetään datan analyysissä hyväksi, esimerkiksi geneettiset algoritmit käyttävät luonnonvalintaa mallinaan algoritmien suunnitteluun ja parhaan ratkaisun valintaan.

Koneoppimista käytetään big datan analysoimisessa (Simon, 2013) erityisesti ennakoimaan tulevaa (Gandomi & Haider, 2015). Tarkoituksena koneoppimisella on opettaa konetta automaattisesti havaitsemaan rakenteita ja tekemään hyviä päätöksiä (Manyika ym., 2011). Simonin (2013) mukaan esimerkiksi Huffington Post käyttää koneoppimista kommenttien sopivuuden arvioimiseen. Ennakoimiseen voidaan käyttää koneoppimisen lisäksi esimerkiksi simulaatioita eli kohdealueen mallinnusta tai aikasarja-analyysia. Luonnollisen kielen prosessointi eli Natural language processing (NLP) on kasa tekniikoita, joissa käytetään algoritmeja ihmiskielen analysoimiseen (Manyika ym., 2011). Luonnollisen kielen prosessointi voidaan toteuttaa koneoppimisen keinoin. Luonnollisen kielen prosessoinnin avulla voidaan etsiä mielipideanalyysin keinoin tekstistä mielipiteitä ja tunteita. Koneoppimiseen liittyy läheisesti neuroverkot, joiden avulla voidaan löytää kuvioita ja rakenteita, joiden avulla voidaan esimerkiksi optimoida prosesseja (Manyika ym., 2011). BI eli business intelligence-työkaluja käytetään

datan analysoimiseen, raportointiin ja esittämiseen (Manyika ym., 2011). Näissä työkaluissa käytetään useita aiemmin mainittuja tekniikoita.

Tekstidataa syntyy valtavasti esimerkiksi sosiaalisessa mediassa, sähköpostina, blogeissa, uutisissa. Näistä saadaan eroteltua tietoa tiedon erottamisella, joka tunnetaan paremmin englannin kielisellä termillä information extraction (IE) (Gandomi & Haider, 2015). Tässä tekniikassa käytetään algoritmeja, jotka erottelevat rakenteettomasta tekstidatasta oleellisia tietoja kuten esimerkiksi uutisesta ajan, paikan, henkilön tai yrityksen, johon uutinen kohdistuu. Ohjelmoimalla ja koneoppimisen avulla voidaan mahdollistaa tekstin yhteenvedot. Yhteenvedo voidaan luoda ohjelmoimalla alkuperäistä tekstiä käyttäen tai tekstistä voidaan luoda tiivistelmä luonnollisen kielen prosessoinnilla, jolloin luodaan myös uutta tekstiä. Jälkimmäisessä tekniikassa tarvitsee syvemmin selvittää tekstin tarkoitus, joten alkuperäisen tekstin käyttäminen voi olla monelle helpompi tapa. Question answering (QA) eli kysymyksiin vastaaminen on teknologia, jota käytetään monien tuntemassa Sirissä sekä IBM:n Watsonissa. Vastaamiseen liittyy myös luonnollisen kielen prosessointi ja kysymyksiin vastaamisen tekniikat voidaan jakaa kolmeen kategoriaan: information retrieval (IR) eli informaatioon pohjautuvaan, knowledge eli tietoon pohjautuvaan sekä hybridiin.

Äänianalyysissa rakenteettomista äänitiedostoista etsitään tärkeä tieto esille. Äänianalyysin alle kuuluu esimerkiksi puheanalyysi. Puhetta analysoidessa on käytössä suuren sanaston jatkuva puheen tunnistus eli SSJPT (englanniksi LVCSR largevocabulary continuous speech recognition) tai foneettinen systeemi (Gandomi & Haider, 2015). SSJPT-systeemissä puhe muunnetaan kirjalliseen muotoon automaattisilla puheentunnistuksen (ASR) algoritmeilla ja sanakirjan avulla. Foneettisessa systeemissä puheesta löydetään foneemit ja käännetään foneemit suoraan tekstiksi. Ero foneettisen ja SSJPT välillä on siinä, tehdäänkö tekstiksi sarja sanoja vai foneemeja. Puhetta voidaan myös analysoida ja automaattisesti käsitellä vuorovaikutteista vastausta eli Interactive Voice Response (IVR)-teknologiaa käyttäen, jota käytetään usein esimerkiksi automaattisissa puhelinpalvelukeskuksissa.

Videon analysoinnin haaste on videodatan suuruus verrattuna esimerkiksi tekstidataan (Gandomi & Haider, 2015). Silti videoiden analysointi yleistyy kasvavissa määrin ja big datalle yhä kehittyvät teknologiat auttavat helpottamaan analyysia. Algoritmeilla videoista saadaan poimittua mahdollisesti tärkeää tietoa, kuten asiakkaan ikä, sukupuoli sekä missä ja miten he liikkuvat. Palvelinarkkitehtuurissa videota lähetetään pakattuna palvelimelle, joka analysoi videon. Videoiden analysointiin voidaan käyttää puheen ja tekstinanalyysissä käytettyjä keinoja. Reuna-arkkitehtuurissa video analysoidaan suoraan videodatasta, mutta se on kallista ja melko tehotonta (Gandomi & Haider, 2015).

Sosiaalisesta mediasta syntyy sekä rakenteellista, että rakenteetontakin dataa ja tätä dataa analysoidaan sisällön tai rakenteen mukaan. Sosiaalisessa verkostanalyysissa analysoidaan yksöiden välisiä suhteita (Manyika ym., 2011). Sosiaalisen median sisällön suhteen käytetään samoja tekstin-, äänen- ja videonanalysointikeinoja, mitä aiemmin käsiteltiin, mutta rakenteen analysointiin tarvitaan muita tekniikoita ja teknologioita (Gandomi & Haider, 2015).

Esimerkiksi verkostosta voidaan etsiä sisäisiä yhteisöjä tai käyttäjiä, jotka kommunikoivat erityisen paljon keskenään. Sekä Gandomi ja Haider (2015) että Chen ja Zhang (2014) mainitsevat sosiaalisen verkostanalyysin tärkeäksi analyysin keinoksi.

Chenin ja Zhangin (2014) mukaan visualisointimenetelmät ovat tärkeitä big datan esitysmenetelmiä. Myös McKinseyn raportti (2011) mainitsee visualisointimenetelmiä kuten tagitaulun, klusterointitaulun, historian virtaus-kuvion ja avaruudellisen tiedonvirtauksen. Tagitaulussa useasti mainitut sanat on merkitty isompina kuin harvemmin mainitut sanat, klusterointitaulussa näytetään klusteroinnin avulla löydetyt ryhmät, historian virtaus-taulussa nähdään raporttien muokkaukset ajallisessa järjestyksessä ja avaruudellisessa tiedonvirtauksessa kuvataan tiedon virtausta paikan suhteen. Visualisoinnin menetelmillä voidaan kuvata analyysistä saadut tiedot helposti ymmärrettävästi ja näin tiedot on helpompi esitellä esimerkiksi yrityksen päätöksentekijöille.

3 BIG DATAN KÄYTÖSTÄ YRITYKSISSÄ

Big datan elämäkaaren voi jakaa neljään vaiheeseen: keräämiseen, yhdistämiseen, analyysiin ja käyttöön (Federal State Commission, 2016). Tässä luvussa keskitymme siihen mihin ja miten big dataa käytetään, mitä seuraamuksia big datan käytöllä voi olla, mitä käytössä on otettava huomioon sekä mitä hyötyjä käytöstä on.

Big dataa voidaan käyttää hyödyksi yrityksissä muun muassa markkinoinnissa, hinnoittelussa sekä tulosten arvioinnissa (Philip Chen & Zhang, 2014). Niin yksityiset yrityksetkin, kuin julkisen puolen sekä terveydenhuollon laitokset käyttävät big dataa hyväkseen (Manyika ym., 2011). Muun muassa Yhdysvalloissa menestyvä marketti Wal-Mart käyttää big datan teknologioita analysoidakseen liiketoimen dataa (Philip Chen & Zhang, 2014).

Big dataa hyödyntämällä yritykset saavat huomattavaa kilpailuetua (Manyika ym., 2011). Kilpailuetua voidaan synnyttää analysoimalla asiakasdataa sekä yrityksen työntekijöiden dataa, jotta voidaan arvioida paremmin millaista mainontaa halutaan, mihin asiakkaat ovat tyytyväisiä tai tyytymättömiä sekä löytämään täysin uusia liiketoimintamahdollisuuksia. Big datan käyttöön liittyy myös useita onnistumiseen vaikuttavia seikkoja, jotka on otettava huomioon. Näitä ovat esimerkiksi tietoturva, lait ja säädökset sekä henkilökunnan ammattimaisuus.

3.1 Big datan käyttö liiketoimessa

Big datalla on mahdollisuus mullistaa monia toimialoja ja kehittää jopa uusia. McKinseyn (2011) raportti otti tarkempaan tarkasteluun viisi heidän mielestään suurinta toimialaa, joita big data voi muuttaa merkittävästi. Nämä viisi ovat terveydenhuolto, julkinen sektori, vähittäiskauppa, teollisuus ja tietoliikenne. Nykyään vähittäiskaupan alalla louhitaan tietoa paljon, jotta saadaan tietoon asiakkaiden mieltymyksiä, ostotottumuksia ja kiinnostusten kohteita (Manyika ym., 2011). Mitä enemmän dataa saadaan, sitä parempi. Esimerkiksi Amazon käyttää dataa hyväkseen, jotta osaa paremmin ehdottaa muitakin kiinnostavia tuotteita asiakkailleen osiossa "Nämäkin tuotteet saattavat kiinnostaa sinua". Datalla voidaan myös esimerkiksi seurata asiakkaan paikkaa, käyttäytymistä kaupassa, myyjien tehokkuutta, segmentoida asiakkaita tai optimoida hintaa.

Big data mullistaa monia muitakin toimialoja kuin vähittäismyyntiä ja monia muitakin liiketoiminnan osa-alueita kuin tämän tutkielman aiheena oleva henkilöstönhallinta tai tässä alaluvussa mainittu markkinointi. Big datalla on potentiaalia luoda jopa uusia toimialoja, mutta keskitymme tässä tutkimuksessa seurantaan ja rekrytointiin.

3.2 Big datan käytöstä yrityksen sisällä

Henkilöstönhallinta- eli lyhyesti HR-sektorilla on hyötyä big datan käytöstä. Big datan käytöstä ei voida kuitenkaan puhua mainitsematta etiikkaa ja muita datan käyttöön liittyviä seikkoja. Säädöksiä kuten GDPR sekä maan lakien suhteen on oltava tarkkana, kun kerätään dataa yksityisistä henkilöistä. On myös tärkeää pohtia big datan käytön tarvetta työntekijöiden seurantaan yrityksissä ja pitää huoli, että työntekijät ymmärtävät miten big datan käyttö toimii sekä mitä datalla tehdään. Lisää aiheesta on luvussa 4.

HR-sektorin täytyy tehdä big datan käyttö kuitenkin oikein, jotta siitä on hyötyä eikä haittaa. Angrave, Charlwood, Kirkpatrick, Lawrence sekä Stuart (2016) varoittaa tutkimuksessaan, ettei monet big datan käyttöönottaneet yritykset ole sisäistäneet sen hyötyjä eivätkä siten käyttäneet big dataa strategisesti hyödykseen. Ongelma on myös heidän mukaansa siinä, ettei analyytikot ymmärrä työvoimaa. Siksi olisikin tärkeää kouluttaa henkilöstönhallinnon henkilökuntaa sekä palkata big datan asiantuntijoita, jos big dataa on tarkoitus käyttää henkilöstönhallinnon apuna.

3.3 Huomioitavaa big datan käytössä

Tietoturva on tärkeä osa yritystä ja sen järjestelmiä ja siten big datan tietoturvan on oltava kohdillaan. Big dataan liittyy paljon mahdollisuuksia tietoturvaan liittyen, mutta myös haasteita. Muun muassa ammattimaisen henkilöstön puute on monilla yrityksellä ongelmana (Alguliyev & Imamverdiyev, 2014). Toisaalta big datan analyysin keinoja voidaan käyttää myös itsessään tietoturvan parantamiseen keräämällä esimerkiksi tietoliikennedatata ja lokitietoja.

Big dataan liittyvästä juridiikasta voisi kirjoittaa hyvin paljon, mutta seuraavaksi mainitaan vain tärkeimmät huomioon otettavat. Uusi Euroopan tietosuojasetus eli GDPR on vaikuttanut yrityksiin jo pian vuoden ja totta kai datasta puhuttaessa on otettava huomioon sen tuomat rajoitukset. Liikenne- ja viestintäministeriö julkaisi 2013 julkaisun, jossa otti huomioon Suomen lakien rajoitettavia tekijöitä, mutta siinäkin jäi vielä avoimeksi miten mikäkin laki todella vaikuttaa. Julkaisussa mainittiin kuitenkin esimerkiksi henkilörekisteriin liittyvät rajoitukset ja tekijänoikeuslait. Suomen lain näkökulmasta dataa ei voida omistaa, mutta voidaan hallita (Seppälä, Juhanko & Mattila, 2018). Dataa hallitsee oletuksena se toimija, kenen laitteissa data sijaitsee. Erinäiset sopimukset voivat kuitenkin määritellä datan omistajuutta. Myös muun muassa elektroniseen valvontaan liittyy Suomessa lakeja, jotka rajoittavat yrityksiä ja suojaavat kansalaisten yksityisyyttä (Laki yksityisyyden suojasta työelämässä 759/2004). Lisää itsessään rekrytointiin ja seurantaan liittyvistä lakipykälistä ja asetuksista seuraavassa luvussa.

Big datan käyttö voi herättää myös huolia siitä, voiko datalla syrjiä tiettyjä ihmisjoukkoja. Tähän tietenkin vaikuttaa paljon, miten dataa käsitellään ja siten kuka sitä käsittelee. Jos datassa on vääristymiä, niin analyysikin menee silloin pieleen. On siis tärkeää pitää huolta datan oikeellisuudesta. Muun muassa Federal Trade Commissionin (2016) raportti kertoi big dataan liittyvässä työpajassa nousseen huolenaiheeksi, että big dataa käyttämällä voitaisiin estää matalatuloisia saamasta työpaikkoja. Raportissa suositellaankin seuraavia kysymyksiä mietittäväksi big dataa käyttäville yrityksille:

- Kuinka edustava datasettisi on?
- Ottaako datamalli huomioon harhat?
- Kuinka paikkaansa pitäviä tekemänne ennakoinnit big datan avulla ovat?
- Herättääkö big dataan turvautuminen eettisiä tai oikeudenmukaisuuteen liittyviä kysymyksiä?

Nämä seikat eivät ole big datan käyttöä estäviä asioita, mutta joihin tulisi kiinnittää huomiota, jotta big datasta saadaan täysi hyöty irti. Onnistumiseen vaikuttaa paljon henkilöstön ammattimaisuus, asiaan perehtyminen sekä täsmällisyys. Tämä kuitenkin pätee usein, kun ollaan ottamassa käyttöön uutta ja mullistavaa teknologiaa.

3.4 Big datan hyödyistä

Big dataa yritys voi hyödyntää useilla eri keinoilla. Big datan käytön tuomia hyötyjä voidaan myös lajitella monella eri tapaa. Voidaan tarkastella mihin big datan käytön vaiheeseen tai mihin henkilöstönhallinnan vaiheeseen hyödyt vaikuttavat. Voidaan myös lajitella hyötyjä arvon eli siten taloudellisen näkökulman mukaan.

Fosso ja Wamba (2015) tekivät tutkimuksessaan luokittelun viitekehyksen, jossa big datan tuoma arvo voidaan luokitella viiteen kategoriaan. He käyttivät apunaan McKinseyn suosittua raporttia big datasta (Manyika ym., 2011). Tässä tutkielmassa olen tutkinut McKinseyn raporttia paljon erityisesti teknologioiden ja tekniikoiden sekä hyötyjen osalta. Fosson ja Wamban (2015) viitekehyksen avulla voin luokitella big datan käytön tuomat hyödyt viiteen kategoriaan. Nämä kategoriat ovat: läpinäkyvyyden luominen, kokeilujen käyttöönotto tarpeiden löytämisen, vaihtelevuuden paljastamiseen ja suorituksen parantamiseen, toimintojen kustomointi segmentoinnilla, ihmisten päätöksenteon korvaaminen/tukeminen automaattisilla algoritmeilla sekä uusien liiketoimintamallien, tuotteiden tai palvelujen innovointi.

Barman ja Ahmedin (2015) jakoi henkilöstönhallinnan tärkeimmät osa-alueet suunnitteluun, rekrytointiin, valintaan, kouluttamiseen sekä suoritusten arviointiin. Voidaankin nähdä, että big data voi tuoda hyötyä henkilöstönhallinnon eri osa-alueille eri tavoilla. Näin ollen voidaan myös big datan käytön tuomat

hyödyt jakaa näiden osa-alueiden kautta. Toisaalta myös big datan työkalujen avulla voidaan lajitella mihin datan käsittelyn vaiheeseen mikäkin hyöty eniten liittyy. Vaiheet lajitellaan useimmin big datan hallintaan sekä big datan analyysiin (Philip Chen & Zhang, 2014).

Losh(2013) pohti kirjassaan big datan teknologioiden ominaisuuksia, käyttöönottoa sekä big datan käytön tuomia hyötyjä. Hyötyjä yritykselle tulee, kun teknologia tuo yritykselle arvoa. Yritykselle arvoa tuo tuottojen lisääminen, kustannusten vähentäminen, tuottavuuden lisääminen sekä riskien vähentäminen (Loshin, 2013). Taulukossa 2 on koottu eri tutkimusten näkemystä siitä, mihin big datan käytön tuomat hyödyt liittyvät.

TAULUKKO 2 : Big datan käytön hyödyt

Lähde	Big datan käytön hyödyt
Philip Chen & Zhang, 2014	Datan hallintaan liittyvät hyödyt sekä datan analyysiin liittyvät hyödyt
Fosso Wamba, Akter, Edwards, Chopin & Gnanzou, 2015	Big datan käytön hyödyt voivat liittyä läpinäkyvyyden luomiseen, kokeilujen käyttöönottoon, tarpeiden löytämiseen, vaihtelevuuden paljastamiseen ja suorituksen parantamiseen, toimintojen kustomointiin segmentoinnilla, ihmisten päätöksenteon korvaamiseen/tukemiseen automaattisilla algoritmeilla ja/tai uusien liiketoimintamallien, tuotteiden tai palvelujen innovointiin.
Barman & Ahmed, 2015	Henkilöstönhallinnan suunnitteluun, rekrytointiin, valintaan, kouluttamiseen ja suoritusten arviointiin liittyvät hyödyt
Loshin, 2013	Hyötyä tuo arvon tuottaminen eli tuottojen lisääminen, kustannusten vähentäminen, tuottavuuden lisääminen tai riskien vähentäminen

4 BIG DATAN KÄYTTÖ TYÖNTEKIJÖIDEN SEURANNASSA JA REKRYTOINNISSA

Big datan hallintaan ja analysointiin liittyy useita tekniikoita ja teknologioita, joiden avulla yritys voi löytää hyötyjä. Hyötyjä yrityksen sisällä erityisesti työntekijäresurssien tuottamasta ja työntekijöihin liittyvästä big datasta voidaan tuoda esiin luvussa kaksi läpikäydyillä työkaluilla.

Koska sosiaalisesta mediasta löytyy paljon rakenteetonta dataa niin potentiaalisista työnhakijoista kuin asiakkaiden mielipiteistä, tuo big datan hyödyntäminen rekrytoinnissa tutkimusten mukaan huomattavia etuja (Ikram, Su, Fiaz & Khadim, 2017). Keskityn tässä tutkielmassa kuitenkin hyötyihin, joita yritykset saavat big datan käytöstä yrityksen sisäisissä asioissa kuten työntekijöiden rekrytoinnissa ja seurannassa.

4.1 Työntekijöiden seuranta ja rekrytointi

Monissa tutkimuksissa, mitä aiemmin tai lähivuosina on tehty, painotetaan paljon yrityksen ja sen työntekijöiden välistä luottamusta sekä läpinäkyvyyttä ja yhteisten sääntöjen sopimista monitoroinnin ja seurannan yhteydessä. Organisaatioissa, joissa on vahva kontrollikulttuuri eli paljon työntekijöiden monitorointia, työntekijät kokivat vähemmän luottamusta organisaatioon ja tietojenkäsittelyyn (Chang, Yenching Liu & Lin, 2015). Tällöin koettiin enemmän rajaturbulenssia, eli työntekijöillä ja työnantajilla on erilainen käsitys siitä, missä menee raja työntekijän tiedon kanssa, jota voidaan tallentaa sekä käsitellä työpaikalla ja työntekijän yksityisen tiedon kanssa. Olisikin tärkeää, että yritys on selkeä ja läpinäkyvä seurannan suhteen, jotta työntekijöiden luottamus säilyy ja koetaan mahdollisimman vähän vastarintaa.

Westwood ja Cazier (2016) käsittelevät paperissaan työelämän balanssia. Heidän mukaansa työntekijä on tehokkaimmillaan ja voi hyvin, kun työelämän ja vapaa-ajan välillä on sopiva tasapaino. Tämä ei kuitenkaan tarkoita, että enemmän vapaa-aikaa on aina parempi. Sopivan balanssin löytäminen on yksilöllistä eikä siten kovin yksinkertaista.

Työntekijöillä on Shahin Iranin ja Sharifin (2017) mukaan suuri merkitys muutoksen onnistumiseen yrityksessä ja siten tietämällä työntekijöiden ajatuksia muutoksesta voi parantaa lopputulosta. Muutosvastarinta onkin iso tekijä uusien muutosten epäonnistumisessa. Muutosta on vaikeaa saada tapahtumaan, jos työntekijät eivät suostu tai osaa suhtautua muutokseen.

De Vries ja van Gelderc (2015) tutkivat seurannan vaikutusta ihmisiin, joilla on matala rehellisuus-nöyryys-luonteenpiirre eli ihmisiin, jotka helpommin ajautuvat tekemään moraalisesti vääriä asioita. He huomasivat, että kyseisiin ihmisiin ei vaikuta valvonnan määrä juurikaan. Nämä ihmiset siis tekevät moraalisesti vääriä asioita riippumatta valvonnan määrästä. Tämä merkitsee sitä, että

seuranta ei tee työntekijöistä sen rehellisempiä kuin he ovat, mutta seurannalla voidaan silti löytää ongelman lähteet ja siten estää ongelman pahenemista.

Paras keino estää epärehellistä käyttäytymistä työpaikalla on kiinnittää enemmän huomiota oikeanlaisten ihmisten rekrytointiin. Tämä on kuitenkin yksi rekrytoijien suurimpia haasteita, sillä 52% rekrytoinnin ammattilaisista kokee oikean henkilön löytämisen hakijoiden joukosta vaikeimmaksi tehtäväksi (Ideal, 2019). Rekrytoinnin yhteydessä ongelmana usein on kasvava hakemusten määrä ja miten saada tarvittava tieto esiin mahdollisimman vaivattomasti ja tehokkaasti hakemuksista (Amdouni & Ben abdessalem Karaa, 2010). Tähän hakemusten määrän kasvuun on osaltaan voinut vaikuttaa verkkohakemusten helppo tekeminen. Sen takia onkin yhä suurempi tarve suurten ja rakenteettomankin tiedon analysoimiseen.

Rekrytoinnin tutkimuksessa nousee paljon esille verkkorekrytointi, eli verkkoon laitettavat työhakemukset ja lomakkeet. Verkkorekrytointi ei ole kovin uusi asia yrityksille, mutta monet tutkimukset keskittyvätkin paljon kehittyviin maihin tai maihin, joissa on vielä monia ihmisiä, jotka eivät omista tietokonetta. Verkkorekrytoinnin tehokkuutta mitataan useimmiten kulujen säättämisellä, sopivan työntekijän löytämisellä, nopeudella sekä työntekijöiden pysyvyydellä (Zhang & Zhou, 2012).

Malik ja Mujtaba (2018) tutkivat verkkorekrytoinnin vaikutusta henkilöstönhallintaan Pakistanin yksityisissä yrityksissä. Tuloksissa selkeästi näkyy, että verkkorekrytointi vaikuttaa muun muassa henkilöstönhallinnan tehokkuuteen, työntekijöiden löytämiseen sekä kustannustehokkuuteen. He ohjeistavatkin verkkorekrytointia käyttävien kiinnittävän huomiota suunnitteluun ja toteutukseen, koulutukseen sekä käytettäviin sivustoihin ja tietokantoihin (Shaukat Malik & Assad, 2018). Rekrytoinnille voi siis olla paljon hyötyä verkkorekrytoinnista, jos se suunnitellaan ja toteutetaan tarkasti.

Yrityksen rekrytointiin vaikuttaa monet asiat kuten taloudelliset resurssit, brändi sekä aika. Beameryn blogin mukaan Glassdoorin tutkimuksessa 11% työntekijöistä sanoi hylkäävänsä työtarjouksen, jos yrityksellä on huono maine ja 86% vaihtaisi uuteen työpaikkaan yrityksessä, jossa on loistava brändi (Slater, 2016). Myös pitkä rekrytointiprosessi vaikuttaa, koska se kuluttaa yrityksen resursseja ja rekrytointiprosessi on kallista yrityksille.

Näistä tutkimuksista voidaan päätellä, että työntekijöiden seuranta harjoitetaan jo monissakin yrityksissä monella tapaa ja uusien teknologioiden kehitys on herättänyt uusia kysymyksiä vanhojen rinnalle. Suuri datamäärän kasvu isoissa yrityksissä herättää kysymyksen sopivista tietokannoista ja sopivista analysointityökaluista, myös sosiaalisen median datan käyttöä pohditaan niin käytännön kuin etiikan kautta. Seurannan etiikkaa ja työntekijöiden suostuvuutta seurannalle on pohdittu sekä akateemisesti että yrityksissä pitkään ja tutkimusten valossa näyttää siltä, että sopivan seurannan määrä riippuu yrityksestä. Rekrytoinnissa ei olla tutkimuksista päätellen vielä monissa yrityksissä otettu käyttöön uudempaa kuin verkkorekrytointi, mutta edelläkävijöitäkin on kuten myöhemmin mainittava jetBlue.

Uutta haastetta yrityksille on tuonut EU:ssa toimeen tullut uusi tietosuojasetus eli GDPR (Věra Jourová, 2016). Toki Suomen laki on sisältänyt jo pitkään henkilötietolain, joka suojaa Suomalaisten henkilötietoja. EU-asetus vaikuttaa vahvasti myös suuriin kansainvälisiin yrityksiin, koska EU-alueella toimivien yritysten on noudatettava asetusta. Myös muut lait on otettavat huomioon, kun käsitellään työntekijöiden tai työnhakijoiden dataa. Tietosuojavaltuutettu kertoi Ylen uutisessa vuonna 2015, että teknisestä valvonnasta on sovittava yhteistoimintamenettelyssä (Nurmi, 2015). Valvonnassa on siis otettava huomioon laki yksityisyyden suojasta työelämässä (759/2004) ja käytävä menettelyssä läpi käytännöt valvontaan liittyen sekä kirjoitettava kirjalliset säännöt.

Barma ja Ahmed (2015) kirjoittivat tutkimuksessaan, että henkilöstönhallinnon tärkeimmät tehtävät jakautuvat viiteen alueeseen. Nämä alueet ovat suunnittelu, rekrytointi, valinta, kouluttaminen sekä suorituksen arviointi. Big datan ja sen työkalujen käyttäminen voi vaikuttaa positiivisesti näihin kaikkiin osa-alueisiin seuraavassa alaluvussa kuvatuilla tavoilla.

4.2 Big datan käyttö työntekijöiden seurannassa ja rekrytoinnissa

4.2.1 Työntekijöiden seuranta big datan avulla

Työntekijät ovat yrityksen tärkeimpiä resursseja ja moni yrityksen menestykseen vaikuttava tekijä riippuu työntekijöistä. Nykyisessä tietoyhteiskunnassa työntekijöiden tietotaito on kilpailun ehto (Chien & Chen, 2008). Siksi onkin perusteltua seurata työntekijöidensä menestystä ja hyvinvointia ja tässä big data voi olla avuksi.

Kun big dataa on päätetty käyttää hyödyksi työntekijöiden seurannassa ja rekrytoinnissa, on hyvä miettiä sopivia työkaluja big datan hallintaan. Hadoop on monien suosima ja koko ajan kehittyvä projekti, jossa on useita erilaiseen käyttöön soveltuvia osia. Myös tietokantamalli voi tulla mietinnän alle, koska useissa yrityksissä on käytössä esimerkiksi perinteinen relationaalinen tietokanta, joka ei välttämättä suoriudu hyvin big datan suurista joukoista rakenteetonta dataa. NoSQL ja NewSQL ovat hyviä vaihtoehtoja relationaaliseen malliin tottuneille, jotka tarvitsevat kuitenkin lisää skaalautuvuutta.

Hadoop sekä tietokannat kuten NoSQL auttavat käyttämään laajempia datasettejä ja keräämään enemmän sekä rakenteellista että rakenteetonta dataa. On yleisesti tiedettyä, että suuremmilla datamäärillä saa tarkempaa analyysiäkin. Tämä on totta varsinkin tilastollisten menetelmien käytössä, jossa pienemmästä joukosta ei voida tehdä kovin tarkkoja johtopäätöksiä.

Bernard Marr mainitsi Forbesin artikkelissaan myös, että useat yritykset tekevät kyselyitä yrityksensä työntekijöille, joissa kysytään tyytyväisyyteen liittyviä kysymyksiä (Marr, 2018). Näitä työntekijäkyselyitä voidaan analysoida tehokkaammin big datan työkaluilla, kuten tekstianalyysillä tai koneoppimisella, joka voivat löytää toistuvia malleja tai kuvioita.

Asiakaspalvelun laadun tai työyhteisön ilmapiirin tilaa koskevaa dataa voi olla vaikeaa löytää ja analysoida pelkästään tekstianalyysin menetelmien tavoin, joten mielipideanalyysistä voi olla hyötyä. Näin voidaan saada esiin työntekijöiden tuntemuksia työpaikastaan ja työyhteisöstä. Myös sosiaalisen median datasta, joka on usein rakenteetonta, on mahdollista analysoida työntekijöiden tyytyväisyyttä työpaikkaan.

Big dataa voidaan käyttää erilaisten tekstien kuten raporttien analysoimiseen, säilömiseen sekä luomiseen. Big datan tekniikoiden avulla tekstistä, äänestä sekä videosta voidaan etsiä tarkoitusta tai tiettyjä muuttujia tehokkaasti. Esimerkkinä näiden tekniikoiden käytöstä toimii lääkäreiden pitkien potilasraporttien tiivistäminen (Simon, 2013). Simon (2013) mainitsee myös, että sähköposteista voidaan tekstianalyysin avulla löytää mahdollisia tärkeitä tietoja. Tällaisia tietoja voi olla esimerkiksi asiakaspalvelun laatu sekä henkilöstön keskinäisen ilmapiirin tila.

Toki sähköpostit eivät ole ainoa tekstinä oleva data ja tiedon lähde. Big datan tekniikoita voidaan mahdollisesti käyttää saamaan esimerkiksi suuressa yrityksessä tehdyn työntekijöiden tyytyväisyyskyselyn tulosten käsittelyyn. Tuloksia voi olla yhdessä yrityksen toimipaikassa satojakin, jolloin manuaalinen tekstin analysointi ei ole tehokasta. Tekstiä syntyy paljon työpaikan sisällä työntekijöihin liittyen, mutta myös työpaikan ulkopuolella, erityisesti sosiaalisessa mediassa. Sosiaalisen median dataa voisi hyödyntää esimerkiksi analysoimalla tageja tai yrityksen sosiaalisen median sivuja ja sinne kirjoitettuja kommentteja.

Ennakoimista voi tehdä työntekijöiden datalla analysoimalla dataa muun muassa koneoppimisella tai regressiota käyttäen (Manyika ym., 2011). Ennakoimisesta voi olla hyötyä esimerkiksi sairastumisten ennustamisessa, työhyvinvoinnin seuraamisessa ennakoimalla työntekijöiden stressiä, suoriutumisen ennakoimisessa.

Terveystilaan liittyvää dataa voidaan kerätä, tallentaa ja analysoida big datan keinoin (Gandomi & Haider, 2015). Esimerkiksi väsymyksen tarkkaileminen on monissa töissä tärkeää turvallisuuden ja hyvinvoinnin kannalta (Sedighi-Maman ym., 2017). Myös vaikeiden sairauksien löytäminen ajoissa voi auttaa työntekijää pääsemään nopeammin takaisin arkeensa ja siten työnsä pariin. Ennakoimalla voidaan mahdollisesti välttää tavallisten sairauksien leviäminen helposti työpaikalla.

Datan keräämistä työntekijöiltä puettavien teknologioiden avulla on jo kehitetty muutamilla eri aloilla ja huomattu hyötyjä puettavien teknologioiden avulla kerätyn datan analyysistä (Sedighi-Maman ym., 2017). Puettavien teknologioiden avulla voidaan muun muassa huomata väsymys työntekijöissä ja siten löytää keinoja ennaltaehkäistä ja vähentää väsymystä. Puettavista teknologioista saatavan datan avulla voidaan myös mahdollisesti vähentää sairastumisia, kroonisten sairauksien takia johtuvaa työkyvyttömyyttä sekä seurata työntekijöiden yleistä terveyttä. Big dataa voidaan käyttää myös hyväksi arvioimaan työn ja vapaa-ajan tasapainoa ja näin ylläpitämään työntekijän tyytyväisyyttä töihinsä sekä vapaa-ajan määrään (J. A. Westwood & J. A. Cazier, 2016).

Väsymistä havainnoi tutkimuksessaan Maman, Yazdi, Cavuoto ja Megahed (2017). Dataa oli kerätty puettavien teknologioiden avulla ja tutkimuksessa kerättyä dataa puhdistettiin ja pienennettiin, mutta big datan työkaluilla analysointi onnistuu isoilla datamäärillä. Dataa analysoitiin regression avulla (Sedighi-Maman ym., 2017). Työntekijän havaittu väsyminen voi myös kertoa työaikabalanssin epätasapainosta (J. A. Westwood & J. A. Cazier, 2016).

Ääntä analysoimalla voimme jopa havaita sairauksia (Gandomi & Haider, 2015). Gandomin ja Haiderin (2015) mukaan sairastumisia havainnoi tutkimuksessaan Hirschberg, Hjalmarsson ja Elhadad (2010). Hirschberg, Hjalmarsson ja Elhadad (2010) mukaan työntekijöiden puheluita analysoimalla voi yritys mahdollisesti havaita työntekijänsä mahdollisen sairauden, kuten syövän tai masennuksen. Työterveyslaitoksen mukaan Suomessa maksetaan sairauspäivärahaa eniten tuki- ja liikuntaelinsairauksiin (Oksanen, 2018). Varsinkin fyysisillä aloilla esiintyy paljon sairauspoissaoloja. Tähän voi liittyä huono ergonomia ja raskas työ ja tätä voisi ennaltaehkäistä seuraamalla työntekijöiden työtapoja esimerkiksi sensoreilla tai puettavilla teknologioilla. Analyysin perusteella voidaan päätellä parannuksia, joita olisi tarpeellista tehdä huonoihin työtapoihin, erityisesti niihin, jotka aiheuttavat tuki- ja liikuntaelinsairauksia.

Verkkohakujen analysoinnin avulla voidaan havainnoida moniakin asioita, kuten ennakoita sairastumisia tai havaita koulutuksen tarvetta. Google on harastanut verkkohakujen analysointia jo pitkän aikaa ja kehitti muun muassa Google Flu Trendsin, joka ennakoi flunssakausia (Simon, 2013). Vaikka Google Flu ei itsessään ollut kovin menestynyt ja kaatui loppuun ihmisten google-hakuihin olemattomista oireista, voi tästä kokeilusta oppia ja kehittää parempaa ennakoivaa tekniikkaa.

Big datan avulla voidaan myös tutkia yrityksen sisällä esimerkiksi työntekijöiden asenteita muutoksiin yrityksessä (Shah, Irani & Sharif, 2017). Työntekijöillä on Shahin Iranin ja Sharifin (2017) mukaan suuri merkitys muutoksen onnistumiseen yrityksessä ja siten tietämällä työntekijöiden ajatuksia muutoksesta voi parantamaan lopputulosta. Heidän tutkimuksensa mukaan varsinkin työtyytyväisyydellä, palkalla, sekä uskollisuuden ja indentifioinnin määrällä on merkitystä työntekijöiden muutosvalmiuteen. Big dataa voidaan siis käyttää yrityksen työntekijöiden muutosvalmiuden havainnointiin. Tällä voi olla merkittävää käyttöä yrityksille, joissa muutosvastarinta on ongelmana.

Työhyvinvointia voisi parantaa analysoimalla työntekijöiden keskenäisiä viestejä ja käyttäytymistä työpaikoilla. Toisaalta tässä voi tulla vastaan työntekijöiden yksityisen tiedon ja julkisen tiedon raja sekä heidän halukkuutensa seurantaan. Onnistumisesta on kyllä esimerkkejä, kuten Bank of America. Bank of America seurasi työntekijöidensä taukoja ja huomasi tehokkaimpien työntekijöidensä pitävän taukojaan samaan aikaan (Kuchler, 2014). Pankki muutti taukojen ajat kaikille samaan aikaan, jolloin työntekijöiden tulos parani ja stressi väheni.

Työntekijöiden suoritusta voidaan arvioida ja siten saada selville paremmin koulutuksen tarve sekä mahdolliset ylennykset (Yen, Bastani, Huang, Zhang & Yao, 2017) sekä työnhakijoiden ansioluetteloita käydä automaattisesti läpi

(Amdouni & Ben abdessalem Karaa, 2010). Tiedonlouhintaan liittyy koneoppimisen ja tilastotieteen menetelmien hyödyntäminen ja niitä voidaan hyödyntää muun muassa parhaiten menestyvien työntekijöiden ominaisuuksien tunnistamiseen (Manyika ym., 2011).

Tekstin analyysia ja luonnollisen kielen prosessoinnilla (NLP) voidaan mahdollisesti käyttää hyödyksi työntekijöiden monitoroinnissa. Tekstin analysoinnin ja NLP avulla voidaan analysoida muun muassa sähköposteja ja puheluiden transkriptioita (Simon, 2013). Tästä voi olla hyötyä työntekijöiden työn laadun varmistamisessa tai työmoraalin ylläpitämisessä. Äänianalyysin avulla on mahdollista analysoida puheluita ja säästää transkriptioiden tekeminen.

Erytisesti puhelimesta tapahtuvien asiakaspalvelutehtävien kuten myyntityön puheluiden analysointia voidaan näin helpottaa ja mahdollisesti parantaa työn laatua. Jotkin puhelinpalvelukeskuksista käyttävät äänen analyysia käydäkseen läpi miljoonia tunteja puheluita, jotta asiakastyytyväisyys paranee ja seurataksien työntekijöidensä suoriutumista (Gandomi & Haider, 2015). Big datan tekniikoiden, kuten luonnollisen kielen prosessoinnin ja muiden tekstin analyysin menetelmien, avulla voidaan muun muassa työntekijöiden kirjoittamien raporttien ja muiden pitkien tekstien pohjalta saada ymmärrettäviä ja tiivistettympiä tarvittaessa. Valvontakameroiden tallentamaa videota voi analysoida videoanalysoinnin keinoin ja näin voisi vähentää esimerkiksi rikollisuutta ja automatisoida valvontaa neuroverkkojen avulla.

Koska big datan teknologiat yhä kehittyvät ja erityisesti nanoteknologia sekä sensorit ovat vasta kehityksensä alussa, on tulevassa luultavasti paljon uusia käyttömahdollisuuksia luvassa. Nytkin tämä kehitys on mahdollistanut saatavan lisää käsiteltävää dataa jo olemassa olevista datan lähteistä (Simon, 2013). Nyt voimme myös tunnistaa yhä enemmän puhetta ja käyttää äänidataa hyväksemme.

4.2.2 Rekrytointi big datan avulla

Tekstianalyysia voidaan käyttää erilaisten tekstitiedostojen, jotka yleensä ovat rakenteetonta tietoa, ymmärtämiseen ja siitä merkityksen löytämiseen. Tekstianalyysin voi olla apuna hakemaan merkityksellisiä sanoja tai arvoja työhakemuksista, eli ansioluettelosta sekä hakemuskirjeistä. Tekstin louhinnan käyttöä ansioluettelosta tarpeellisen tiedon löytämiseen on tutkittu (Amdouni & Ben abdessalem Karaa, 2010). Samaa voisi soveltaa myös mahdollisesti hakemuskirjeisiin.

Ennakoimista voi tehdä big datan työkaluilla kuten koneoppimisella, regressiolla ja simulaatiolla. Ennakoimisesta voisi esimerkiksi rekrytoinnissa olla hyötyä, kun halutaan ennakoida työntekijän soveltuvuutta tai pysyvyyttä. Jet-Blue käytti big dataa hyväkseen löytääkseen työntekijän tarvittavia ominaisuuksia (Google, 2019).

Esimerkkejä big datan käytöstä yrityksen sisällä, erityisesti työvoimaan liittyviin päätöksiin, on monia. Google on aloittanut projektin re:Work, jolla Google yrittää lisätä datalähtöistä henkilöstönhallintaa. Re:Workin sivuille Google on

kerännyt ohjeita, blogeja, tapaustutkimuksia ja muuta tietoa datan käytöstä työntekijäresursseissa (Google, 2019). Re:Workin mukaan JetBlue aloitti big dataan liittyvät projektin 2013 ja aikoi big datan avulla parantaa työntekijöidensä taitotasoa sekä tehostaa rekrytointia. Keräämällä dataa työntekijöiltään asiakaslähtöisen puhelintyöskentelyn vaikeimmista tehtävistä sekä ominaisuuksista joita työntekijällä tässä tehtävässä tulisi olla, JetBlue löysi analyysin avulla työtehtäviin tarpeelliset ominaisuudet sekä loi rekrytointiprosessiin uuden tehtävän. Tässä tehtävässä työnhakijoita testattiin oleellisten ominaisuuksien osalta. Rekrytointitehtävässä pärjänneet pärjäisivät myös koulutuksessa paremmin ja yleisesti koulutuksen uuvuttavuus laski 25%.

Lukuisia kirjoja big datan hyödyntämisestä kirjoittanut Bernard Marr kirjoitti Forbesiin 2018 artikkelin, jossa listasi keinoja, jolla henkilöstöhallinto hyötyy big datasta rekrytoinnin suhteen. Marrin mukaan big data voi auttaa parantamaan yrityksen brändiä ja seurata sekä nykyisten että entisten työntekijöiden mielipiteitä (Marr, 2018). Yrityksen brändiä parantamalla voidaan lisätä pätevien hakijoiden kiinnostusta ja saada jopa muiden yritysten työntekijöitä vaihtamaan työpaikkaa. Big dataa voitaisiin käyttää hyödyksi jo työilmoitusta tehtäessä. Erilaisia datalähteitä analysoimalla tärkeiden avainsanojen löytäminen ja lupaavien hakijoiden huomiota herättävien tekstien voi helpottaa ja näin saada sopivia työnhakijoita.

Chien ja Chen (2008) teettivät tutkimuksen, jossa yrityksen työntekijöiden työssä suoriutumista verrattiin heidän työnhakijaprofiilinsa tietoihin, kuten työhistoriaan, koulutustaustaan ja demografisiin tietoihin. Tästä voitiin tehdä päätelmiä tarvittavasta koulutustasosta tiettyyn työtehtävään sekä arvioida työntekijän pitkäaikaisuutta. Mutta nyt kun ihmiset merkkäävät sosiaaliseen mediaan ja erityisesti LinkedIn-sivuston kaltaisiin paikkoihin työnhakijatietojaan, voidaan big datan keinoin nämäkin tiedot helpommin etsiä ja analysoida jotta yritys löytää sopivimmat työntekijät tehtäviinsä.

Tiedonlouhinnan avulla voidaan kerätä arvokasta dataa mahdollisista työntekijöistä ja tehdä parempia päätöksiä rekrytoinnissa (Chien & Chen, 2008). Yritykseen voi tulla satoja, tuhansiakin hakemuksia ja niiden käsittelyyn menee yrityksillä paljon aikaa ja resursseja. Muun muassa hakijoiden ansioluetteloita voidaan analysoida automaattisesti ja löytää tärkeitä avainsanoja tai tehdä hakemuksista tiivistelmiä (Amdouni & Ben abdessalem Karaa, 2010). Tekstin analysoinnilla sekä luonnollisen kielen prosessoinnilla voidaan käsitellä tekstiä nopeammin ja saada esimerkiksi yhteenvetoja hakemuksista, samalla tavalla kuin lääkäreiden raporteista voidaan käsitellä tiivistelmiä (Simon, 2013).

Vaikka big datan käytöllä voi olla monia kompastuskiviäkin, on osaavat asiantuntijat, tarvittavat teknologiat sekä halu hyötyä big datasta tärkeimmät ja tarpeeksi alkuun. Hyödyt big datan käytöstä voivat tuoda paljon säästöjä sekä tyytyväisyyttä yrityksen sisällä.

4.2.3 Big datan hyödyt työntekijöiden seurannassa ja rekrytoinnissa

Big datan työkaluja hyödyntäen saadaan sekä rakenteellisesta että rakenteettomasta datasta esille oleellista tietoa työntekijöistä, hakijoista tai itse yrityksestä. Tutkittuani kirjallisuutta voin päätellä, että oleellisimpia hyötyjä big datasta työntekijöiden seurantaan on muun muassa tarkempi analyysi, terveydentilan seuranta, työntekijöiden suoriutumisen seuranta, työntekijöiden tyytyväisyyden seuranta sekä kulujen vähentäminen.

On yleisesti tiedettyä, että suuri datamäärä parantaa varsinkin tilastollisissa menetelmissä tuloksen tarkkuutta, koska otannan on oltava tarpeeksi suuri edustaakseen joukkoa kattavasti. Terveydentilaan lasken kaikki sairauksista mielisairauksiin sekä stressiin ja väsymykseen. Terveydentilaan liittyvää dataa voidaan kerätä esimerkiksi puettavilla teknologioilla tai jopa analysoimalla puhetta. Suoritusta voidaan seurata, jotta työn laatu paranee tai ylennyksiä ja palkankorotuksia mietittäessä, dataa voidaan saada puhelusta, raporteista, projekteista. Puheluita jo analysoidaan laadun parantamiseksi joissain puhelinpalvelukeskuksissa (Gandomi & Haider, 2015), big data tarjoaa siihen tehokkaat keinot.

Työntekijöiden tyytyväisyyteen liitän läheisesti stressinhallinnan, työelämäbalanssin, tauot, palkan, urakehityksen sekä työilmapiirin. Näihin big data voi vaikuttaa stressin tasoa seuraamalla esimerkiksi kyselyillä tai puettavilla teknologioilla, taukoja voi seurata ja siten optimoida taukojen pituudet tai ajankohdat. Palkkakehitykseen ja urakehitykseen liittyen yritys voi seurata suoriutumista ja työtunteja ja työilmapiiriä voidaan seurata esimerkiksi mielipideanalyysin avulla.

Kuluja vähentää muun muassa tehokkuuden parantaminen työntekijöiden seurannan myötä saatujen analyysin tulosten perustella, sairauspoissaolojen vähentäminen terveystietojen analysoimisen kautta, työntekijöiden suorituksen seurannalla ja siten tehokkuutta parantamalla.

Rekrytointia big data hyödyttää auttamalla valitsemaan paremmin sopivia työntekijöitä, parantamalla yrityksen brändiä, tehostamalla hakemusten läpikäyntiä sekä rekrytointiprosessin parantamisessa. Sopivien työntekijöiden valitsemista ja tehokkuutta big data voi auttaa hakemusten automaattisella analyysillä, eli esimerkiksi luonnollisen kielen prosessoinnin keinoin tai algoritmeilla, toisaalta rekrytointiprosessia voidaan parantaa etsimällä tärkeitä työntekijän ominaisuuksia tai taitoja nykyisten työntekijöiden suorituksesta tai kyselyillä. Brändin kiinnostavuudesta saadaan dataa esimerkiksi sosiaalisesta mediasta tai kyselyillä.

Luvussa 2 jaettiin teknologiat datan hallintaan liittyviin sekä datan analysointiin liittyviin (Gandomi & Haider, 2015). Datan hallintaan liittyy datan kerääminen ja tallentaminen, datan louhinta, puhdistaminen ja huomautusten lisääminen sekä integraatio, datan laittaminen yhteen ja esittäminen. Datan analyysiin liittyy mallintaminen ja analyysi sekä tulkinta. Taulukossa 3 on lajiteltu teknologiat hallinnan tai analyysin mukaan ja kuvattu kyseisten teknologioiden tuomia hyötyjä.

TAULUKKO 3: Hyödyt datan käsittelyn vaiheen mukaan

Datan hallintaan liittyvä	Työkalu	Hyödyt
	Hadoop	Suurien ja monipuolisten datamäärien hallinta
	NoSQL, NewSQL	Skaalautuvuus, suurempien datamäärien hallinta
	Puettavat teknologiat	Sairauksien havaitseminen ja ennakointi, väsymyksen havaitseminen
	Sensorit	Taukojen optimointi
Datan analyysiin liittyvä	Koneoppiminen	Ennakoiminen
	Neuroverkot	Automatisoitu valvonta
	Tekstianalyysi	Raporttien ja muiden tekstien tiivistäminen
	Äänianalyysi	Puheluidenanalysointi, laadun parantaminen
	Videoanalyysi	Rikosten vähentäminen

Barman ja Ahmedin (2015) mukaan henkilöstöhallinnon tärkeimmät alueet ovat suunnittelu, rekrytointi, valinta, kouluttaminen sekä suorituksen arviointi. Näiden mukaan big datan tuomat hyödyt henkilöstöhallinnolle voidaan jakaa siis viiteen osaan, riippuen mille osalle henkilöstöhallintoa big data ja sen teknologiat tuovat hyötyä (Taulukko 4).

TAULUKKO 4 : Hyödyt henkilöstöhallinnon alueiden mukaan

Suunnittelu	Rekrytointi	Valinta	Koulutus	Suorituksen arviointi
Brändin parantaminen	Hakemusten analysointi	Pysyvyyden ennakointi	Koulutuksen parantaminen	Tarkempi analyysi
Taukojen optimointi	Työilmoitusten parantaminen	Sopivuuden ennakointi	Koulutuksen tarpeen arviointi	Laadun parantaminen suorituksen arvioinnilla

Arvon tuottamisen perusteella jaoteltuna voidaan lajitella big datan teknologioiden käytön hyödyt (Fosso Wamba, Akter, Edwards, Chopin & Gnanzou, 2015). Fosso ja Wamba tekivät tutkimuksessaan luokittelun viitekehyksen, jossa big datan tuoma arvo voidaan luokitella viiteen kategoriaan (Taulukko 5). Läpinäkyvyyden luominen tarkoittaa etsintään ja prosessoitiin kuluvan ajan lyhenemistä, kokeiluiden käyttöönotolla voidaan parantaa suoritusta, toimintoja voidaan kustomoida segmentoinnilla, päätöksentekoa korvata tai tukea automaattisilla algoritmeilla tai innovoida uutta. Taulukko 5 big datan mahdollisia käyttömahdollisuuksia on luokiteltu hyötyjen mukaan.

TAULUKKO 5: Hyödyt arvon tuoton perusteella viitekehyksellä

Läpinäkyvyyden luominen	Kokeilujen käyttöönotto tarpeiden löytämiseen, vaihtelevuuden paljastamiseen ja suorituksen parantamiseen	Toimintojen kustointi segmentoinnilla	Ihmisten päätösten korvaaminen/tukeminen automaattisilla algoritmeilla	Uusien liiketoimintamallien, tuotteiden tai palvelujen innovointi
Työnhakijoiden ansioluetteloiden analysointi	Puheluiden analyysi Raporttien ja vastaavien analyysi	Taukojen optimointi Vapaa-ajan ja työn tasapaino	Työnhakijoiden analysoiminen koneoppimisella, tekstin analysoimisella Ylennysten päättämisen helpottaminen suorituksen seurannalla	Työntekijäkselyiden automatisoitu analyysi

Kun on kyse big datan käytöstä työntekijöiden seurannassa sekä rekrytoinnissa, ei liiketoimintojen innovointi ole kovin suuressa roolissa. Eniten hyötyjä löytyy päätöksenteon korvaamisen sekä suorituksen parantamisen osalla.

Loshin (2013) pohti kirjassaan big dataa ja sen kontribuutiota yritykselle. Loshin mukaan tällöin täytyy tarkastella miten big data voi tuottaa arvoa yritykselle. Hän jakoi arvon tuottamisen neljään kategoriaan: tuottojen lisäämiseen, kustannusten vähentämiseen, tuottavuuden lisäämiseen sekä riskien vähentämiseen (Loshin, 2013). Tämä on järkevä lajittelu, sillä yrityksen päätavoitteita on arvon tuottaminen. Big datan käytön tuomat hyödyt voidaan siis lajitella sen mukaan, mitä arvoa big data tuottaa yritykselle (Taulukko 6).

TAULUKKO 6: Hyödyt arvon kautta

Tuottojen lisääminen	Kustannusten vähentäminen	vähentäminen	Tuottavuuden lisääminen	Riskien vähentäminen
Laadun parantaminen	Sairauspoissaolojen vähentäminen		Puheluiden analysoiminen	Väsymyksen ja stressin seuraaminen
Taukojen optimointi	Rikosten vähentäminen		Raporttien analysoiminen ja tiivistäminen	Rikosten vähentäminen
Puettavat teknologiat	Työnhakijoiden pitkäaikaisuuden ennakointi		Suoritusten seuraaminen	Sairauksien seuraaminen
Tekstin analyysi	Työntekijöiden sopivuuden ennakoiminen		Työntekijöiden sopivuuden seuraminen	Tehokkaampi työntekijöiden läpikäynti

5 YHTEENVETO

Big datan käytöstä on moninaista hyötyä henkilöstönhallinnassa. Keskityin erityisesti työntekijöiden seurantaan ja rekrytointiin ja tutkin kuinka big data voi hyödyttää näitä henkilöstönhallinnan alueita. Hyötyä big datan käytöstä henkilöstönhallinnassa on muun muassa päätöksenteossa, työntekijöiden tyytyväisyydessä, kustannusten vähentämisessä, työntekijöiden löytämisessä sekä tehokkuudessa.

Hyötyjä voidaan saavuttaa big datan työkaluilla, joilla voidaan kerätä, tallentaa, analysoida tai visualisoida. Suurin osa hyödyistä tulee esiin analyysin vaiheessa, mutta jos muut datan käsittelemisen vaiheet eivät ole kunnossa tai tehdään epäsopivilla työkaluilla, on sillä vaikutusta myös analyysiin.

Big dataan liittyvät teknologiat ja tekniikat voidaan jakaa datan hallintaan ja datan analysointiin liittyviin. Hallintaan liittyy datan keräämistä, tallentamista sekä esittelemistä ja datan analysointiin liittyy datan analysointi ja tulkinta. Hallinnan työkaluja on useita, mutta suosittuja on muun muassa ohjelmistoprojektin Hadoopin ohjelmistot, tietokannat kuten NoSQL ja NewSQL, tiedon keräämiseen puettavia teknologioita tai sensoreita. Analyysissä hyötyä on muun muassa tiedonlouhinnasta, tilastollisista menetelmistä, koneoppimisesta, luonnollisen kielen prosessoinnissa, teksti-, ääni- ja videoanalyysistä sekä neuroverkoista. Koska big data kasvaa koko ajan suuremmaksi, on työkaluja kehitettävä koko ajan eteenpäin ja tätä kehitystä on hyvä big datan hyödyntäjän seurata.

Työntekijöiden seuranta on yrityksissä tehty vuosia, mutta nykyinen kehitys teknologiassa tuo sekä mahdollisuuksia, että pakottaa seurannan kehittymään eteenpäin. Seurannan vaikutuksista on tehty paljon tutkimuksia ja huomattu, että seuranta ei välttämättä vähennä moraalisten ihmisten rikosten tekemistä. Seuranta voi silti vähentää todennäköisyyttä, että yritys joutuu rikollisuuden kohteeksi sekä auttaa selvittämään rikosten tekijät. Seurannan suhteen on myös oltava läpinäkyvä työntekijälle, että luotto yrityksen ja työntekijän välillä säilyy.

Rekrytointi on jokaisessa yrityksessä tärkeä osa toimintaa ja vaikuttaa yrityksen tulevaan menestykseen, koska työntekijöillä on suuri rooli yrityksen menestyksessä. Rekrytointiin kuluu yrityksillä vuosittain runsaasti rahaa ja pitkät rekrytointiprosessit ovat raskaita yrityksille. Verkkorekrytointi on muuttanut rekrytointia paljon ja suuri osa onkin nyt verkossa, mutta se myös voi osaltaan lisätä hakemusten läpikäynnin vaikeutta.

Hyötyjä big datan käytöstä yritykselle seurannan suhteen on tarkemman analyysin saaminen suuremmilla datamäärillä, työntekijöiden suoriutumisen seuraaminen, työntekijöiden terveydentilan seuraaminen sekä työntekijöiden tyytyväisyyden seuraaminen. Terveydentilaan liittyy myös stressin ja mielentilan seuranta. Suoriutumista voidaan seurata useilla eri keinoilla, esimerkiksi seuraamalla puheluita tai raportteja tehokkaammin. Tyytyväisyyteen voi kuulua tyytyväisyys yritykseen, että työyhteisöön tai esimerkiksi työtehtävään.

Rekrytoinnille big datasta on hyötyä eniten sopivien työntekijöiden löytämisessä ja valitsemisessa, yrityksen brändin parantamisessa, hakemusten läpikäynnin tehostamisessa sekä rekrytointiprosessin parantamisessa. Sopivien työntekijöiden löytäminen on tärkeää, jotta työntekijät ovat asiantuntevia sekä pitkäaikaisia yrityksessä. Brändillä on paljon merkitystä työntekijöiden pysyvyydessä sekä siinä, miten työnhakijat näkevät yrityksen. Rekrytointiprosessin parantamisesta on hyvänä esimerkkinä Jet Blue, joka paransi rekrytointiprosessiaan keräämällä dataa sopivan työntekijän ominaisuuksista omilta työntekijöiltään.

Hyödyt, joita big datan käyttö voi yritykselle tuoda, voidaan jaotella datan käsittelyn vaiheen mukaan, henkilöstönhallinnon vaiheen mukaan tai arvon tuottamisen mukaan. Erityisesti arvon tuottamisen mukaan jaottelu kuvaa hyvin big datan tuomia hyötyjä.

Tutkimusta tehdessä kävi hyvin selväksi, että yrityksen täytyy olla sitoutunut big datan käyttöön ja uskoa sen tuomaan etuun, jotta siitä voidaan hyötyä mahdollisimman paljon. Myös työntekijöiden luottamuksen säilyttämiseksi yrityksen kannattaa olla läpinäkyvä toimintatavoistaan ja keskustella avoimesti työntekijöiden kanssa aiheesta. Suomessa laki velvoittaa yrityksen käymään valvonnan käytännöt läpi työntekijöille sekä olemaan aina läpinäkyvä ja rehellinen valvonnan tapahtumisesta. Tämä voi parantaa työntekijöiden luottamusta seurantaan, kun laki velvoittaa työnantajaa.

Vaikuttavana tekijänä on myös ihmisten luonne ja oletus ihmisten rationaalisuudesta ja rehellisyydestä. Jos työntekijöiden seurannasta taikka rekrytoinnista saatava data on kaikille avointa, niin on mahdollista, että sitä käytetään hyväksi. Jos työntekijä tietää millaista dataa hänestä kerätään, voi työntekijällä olla mahdollisuus vaikuttaa miltä hänestä kerätty data vaikuttaa, silti parantamatta omaa suoritustaan. Toki datan voi myös oikeuttaa nähtäväksi vain tietyille henkilöille, mutta se voi vaikuttaa datan hyödyntämisen tehokkuuteen.

Kun kerätään demografisia tietoja kuten ikätietoja ja sukupuolitietoja ja niitä analysoidaan työntekijöiden ja työnhakijoiden arvioinnissa, voi syntyä kysymys syrjinnästä. Onkin tärkeä ottaa huomioon, ettei kerätty data eikä keinot jolla sitä analysoidaan, vaikuta negatiivisesti syrjivällä tavalla.

Nämä tiedon keruuta ja analysointia vaikeuttavat seikat eivät estä big datan käyttämistä seurantaan tai rekrytointiin, mutta ne on otettava huomioon big datan avulla tehtävää seurantaan tai rekrytointia suunniteltaessa. Tämäkin seikka puoltaa sitä, että big datan käytöstä vastaavan henkilökunnan on oltava ammattimaista ja huolellisia työssään.

Mahdollisia jatkotutkimusaiheita on monia. Big datan käyttöön henkilöstöhallinnossa liittyy paljon eettisiä kysymyksiä, joita olisi hyvä pohtia ja tutkia. Kun on kyse ihmisten datasta, jota voidaan käyttää heitä vastaan, on mahdollista että syntyy vastarintaa. Olisi hyvä tutkia missä koetaan menevän työntekijöiden yksityisen datan ja analyysiin käytettävän datan raja. Myös sitä miten big datan käyttöä syrjivästi voidaan estää olisi tärkeä jatkotutkimusaihe.

Käytettyjä teknologioita kuten Hadoopia ja Luonnollisen kielen prosessointia on jo tutkittu paljon, mutta niiden käyttöä eri konteksteissa tulisi tutkia lisää. Myös big datan käytön hyötyjä eri konteksteissa voisi tutkia, esimerkiksi

vaikuttaako big datan käyttö eri tavalla eri aloilla. Tässä tutkimuksessa kävin läpi big datan työkaluja sekä mahdollisia hyötyjä big datan käytöstä, mutta en miten suuresti big datan käyttö voi hyödyttää yritystä. Voiko esimerkiksi sähköpostien seuranta olla tehokkaampaa kuin puheluiden seuranta? Johdon halukkuutta big datan käyttöön henkilöstöhallinnassa olisi myös hyödyllistä tutkia, sekä big datan käytön toimivuutta henkilöstöhallinnassa muulla kuin kirjallisuuskatsauksella.

LÄHTEET

- Alguliyev, R. & Imamverdiyev, Y. (2014). Big data: Big promises for information security. *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)* (1-4). Astana, Kazakhstan, February 09, 2015.
- Amazon. (2019). What is a data lake? Haettu 10.1.2019 osoitteesta <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>
- Amdouni, S. & Ben abdessalem Karaa, W. (2010). Web-based recruiting. *ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010*. (1-7). Hammamet, Tunisia, September 27, 2010.
- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M. & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1-11.
- Apache. (2019a). Apache Hadoop. Haettu 10.1.2019 osoitteesta <http://hadoop.apache.org/>
- Apache. (2019b). Apache HBase. Haettu 10.1.2019 osoitteesta <http://hbase.apache.org/>
- Browne, R. (2018, 30. toukokuuta). 70% of people globally work remotely at least once a week, study says. *Cnbc*. Haettu 1.2.2019 osoitteesta <https://www.cnbc.com/2018/05/30/70-percent-of-people-globally-work-remotely-at-least-once-a-week-iwg-study.html>
- Chang, S. E., Yen-ching Liu, A. & Lin, S. (2015). Exploring privacy and trust for employee monitoring. *Industrial Management & Data Systems*, 115(1), 88-106.
- Chien, C. & Chen, L. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications*, 34(1), 280-290.
- Eilakaisla. (2019). Rekrytoinnin kustannukset -laskuri. Haettu 1.2.2019 osoitteesta <https://www.eilakaisla.fi/mita-uuden-tyontekijan-rekrytointi-mak-saa-laskuri>
- Ramirez, E., Brill, J., Ohlhausen, M.K. & McSweeney, T. (2016). *Big data: A tool for inclusion or exclusion?* Federal State Commission.

- Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246.
- Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- Google. (2019). Google re:Work. Haettu 10.12.2018 osoitteesta <https://re-work.withgoogle.com/>
- IDC. (2014). *The digital universe of opportunities*. International Data Corporation.
- Ideal. (2019). AI for recruiting: A definitive guide for HR professionals. Haettu 10.1.2019 osoitteesta <https://ideal.com/ai-recruiting/>
- Ikram, A., Su, Q., Fiaz, M. & Khadim, S. (2017). Big data in enterprise management: Transformation of traditional recruitment strategy. *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (414-419). Beijing, China, October 23, 2017.
- Westwood J. A. & Cazier J. A. (2016). Work-life optimization: Using big data and analytics to facilitate work-life balance. *2016 49th Hawaii International Conference on System Sciences (HICSS)* (1701-1709). Koloa, HI, USA, March 10, 2016.
- Kuchler, H. (2014). Data pioneers watching us work. *Financial Times*. Haettu 1.2.2019 osoitteesta <https://www.ft.com/content/d56004b0-9581-11e3-9fd6-00144feab7de>.
- Laki yksityisyyden suojasta työelämässä 759/2004*. Annettu Helsingissä 13.8.2004.
- Laney, D. (2001). *3-D data management: Controlling data volume, velocity, and variety*. META Group.
- Loshin, D. (2013). *Big data analytics : From strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Amsterdam: Elsevier/Morgan Kaufmann.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey.
- Mariotti, A., Robinson, S. & Esen, E. (2017). *2017 human capital benchmarking report*. SHRM.

- Marr, B. (2018). Data drive HR: How big data and analytics are transforming recruitment. *Forbes*. Haettu 1.2.2019 osoitteesta <https://www.forbes.com/sites/bernardmarr/2018/05/04/data-driven-hr-how-big-data-and-analytics-are-transforming-recruitment/>
- Miloslavskaya, N. & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, 88, 300-305.
- Nurmi, R. (2015). Tietosuojavaltuutettu: Työntekijän paikantamisesta sovittava yt-menettelyssä. Haettu 1.2.2019 osoitteesta <https://yle.fi/uutiset/3-8485643>
- Oksanen, T. (2018). *Sairauspoissaolot suomen suurissa kunnissa 2010-luvulla- ammatit erottelevat*. Työterveyslaitos.
- Philip Chen, C. L. & Zhang, C. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275, 314-347.
- Regalado, A. (2013). The data made me do it. *MIT Technology Review*. Haettu 19.12.2018 osoitteesta <https://www.technologyreview.com/s/514346/the-data-made-me-do-it/>
- Rigby, D. & Bilodeau, B. (2013). *Management tools & trends 2013*. Bain & Company.
- Rivera, J. (2015). Employee monitoring and ethical alignment. Haettu 1.2.2019 osoitteesta <https://www.gartner.com/smarterwithgartner/employee-monitoring-and-ethical-alignment/>
- SAP. (2012). Small and midsize companies look to make big gains with “Big data,” according to recent poll conducted on behalf of SAP. Haettu 10.12.2018 osoitteesta <https://news.sap.com/2012/06/small-and-midsize-companies-look-to-make-big-gains-with-big-data-according-to-recent-poll-conducted-on-behalf-of-sap/>
- Sedighi-Maman, Z., Alamdar Yazdi, M. A., Cavuoto, L. & Megahed, F. (2017). A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Applied Ergonomics*, 65(Volume 65), 515-529.
- Seppälä, T., Juhanko, J. & Mattila, J. (2018). *Data ownership and governance – Finnish law perspective*. Etna.
- Shah, N., Irani, Z. & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, 70, 366-378.

- Shaukat Malik, M. & Assad, u. M. (2018). Impact of E-recruitment on effectiveness of HR department in private sector of pakistan. *International Journal of Human Resource Studies*, 8(2), 80-94.
- Simon, P. (2013). *Too big to ignore : The business case for big data*. Hoboken, N.J.: John Wiley & Sons.
- Slater, B. (2016). 17 crucial recruiting statistics for 2016. Haettu 1.2.2019 osoitteesta <https://beamery.com/blog/recruiting-statistics-2016>
- Statista. (2019). Number of connected wearable devices worldwide from 2016 to 2021 (in millions). Haettu 1.2.2019 osoitteesta <https://www.statista.com/statistics/487291/global-connected-wearable-devices/>
- Takahashi, D. (2018). Smartphone users will top 3 billion in 2018, hit 3.8 billion by 2021. Haettu 1.2.2019 osoitteesta <https://venturebeat.com/2018/09/11/newzoo-smartphone-users-will-top-3-billion-in-2018-hit-3-8-billion-by-2021/>
- Věra Jourová. (2016). *EU:N tietosuojauudistus ja massadata*. Euroopan Unioni.
- Yen, I., Bastani, F., Huang, Y., Zhang, Z. & Yao, X. (2017). SaaS for automated job performance appraisals using service technologies and big data analytics. *2017 IEEE International Conference on Web Services (ICWS)*, (412-419). Honolulu, HI, USA, September 11, 2017.
- Zhang, W. & Zhou, C. (2012). How to enhance effectiveness of internet recruitment of enterprises. *International Journal of Human Resource Studies*, 8(2), 2623-2625.