



TAMPEREEN TEKNILLINEN YLIOPISTO  
TAMPERE UNIVERSITY OF TECHNOLOGY

TEPPO HJELT  
DATA-ANALYYTIKON KOMPETENSSIT

Tieteellisen kirjoittamisen harjoitustyö

Tarkastaja: professori Tarmo Lipping  
Aihe hyväksytty 7. kesäkuuta 2017

## TIIVISTELMÄ

**TEPPO HJELT:** Data-analyytikon kompetenssit

Tampereen teknillinen yliopisto

Tieteellisen kirjoittamisen harjoitustyö, 16 sivua

Heinäkuu 2017

Tietotekniikan ja johtamisen diplomi-insinöörin tutkinto-ohjelma

Pääaine: Ohjelmistotuotanto

Tarkastaja: professori Tarmo Lipping

Avainsanat: data-analyytikko, datatieteilijä, kompetenssit, osaaminen, ammattitaito, ominaisuudet

Data-analyytikko on melko uusi ammatti, josta yleisesti tiedetään aika vähän. Data-analyysista tulee lukijalle useimmiten mieleen big data ja big datasta taas melko ärsyttävät automaattiset mainokset vaikkapa Facebookissa. Data-analyysi liittyy toki vahvasti automaattisiin mainoksiinkin, mutta se on myös paljon muuta. Dataa on jo nyt valtavasti, ja sitä luodaan sekä tallennetaan koko ajan lisää. Data-analyytikon tehtävä on luoda sekavasta datasta merkityksellistä tietoa päätöksenteon tueksi – nykyään kutakuinkin kaikilla aloilla liiketaloudesta lääketieteen kautta yhteiskuntatieteisiin.

Tämän työn tarkoituksena on luoda katsaus data-analyytikon ammatissa vaadittaviin ominaisuuksiin sekä taitoihin tieteellisen kirjallisuuden, alakohtaisten blogien sekä työpaikkailmoitusten kautta ja luoda näin profiili tämän päivän data-analyytikosta.

## SISÄLLYSLUETTELO

1.	JOHDANTO .....	1
2.	KÄSITTEET .....	2
2.1	Data-analyytikko .....	2
2.2	Kompetenssi .....	5
3.	DATA-ANALYYTIKON KOMPETENSSIT KIRJALLISUUDESSA.....	6
3.1	Henkilökohtaiset ominaisuudet .....	6
3.2	Koulutustausta ja ammattitaito .....	6
3.3	Tekninen osaaminen ja työkalut.....	7
4.	DATA-ANALYYTIKON KOMPETENSSIT BLOGEISSA .....	9
5.	KATSAUS KOTIMAISIIN TYÖPAIKKAILMOITUKSIIN .....	11
6.	YHTEENVETO .....	14
	LÄHTEET .....	15

# 1. JOHDANTO

Data-analyytikko on vuosisadan seksikkäin ammatti (Davenport ja Patil, 2012). Vai onko tuo sittenkin vuosisadan lattein klisee? Joka tapauksessa meitä kiinnostaa, kuka voisi ostaa seuravan tuotteen, kuka on aikeissa poistua asiakaskunnastamme, millaisia palveluita asiakkaamme tarvitsevat, milloin koneemme kannattaa huoltaa, kuka on vaarassa sairastua ja mihin tautiin, mistä elokuvista pitäisimme, mitä musiikkia tykkäisimme kuunnella, miten liikenneuhkia saadaan vähennettyä, missä seuraava rikos tapahtuu jne. Data-analyytikon avulla voidaan nyt ja tulevaisuudessa ratkaista monia tärkeitä ongelmia. On se jonkin verran seksikästä.

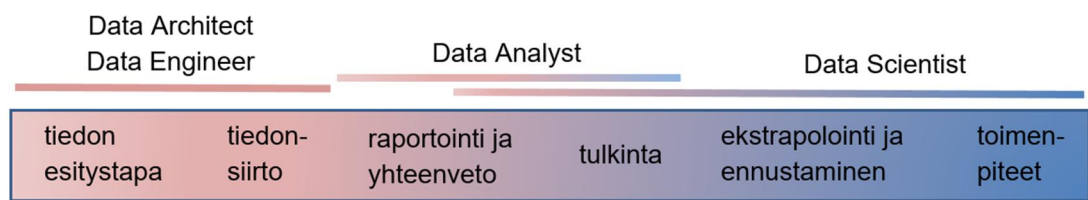
Työn tavoitteena on selvittää, mitä data-analyytikolta tänä päivänä vaaditaan. Vaatimuksia tarkastellaan tutkimuskirjallisuuden, blogien ja kotimaisten työpaikkailmoitusten kautta. Työssä etsitään sopivaa henkilökohtaisten ominaisuuksien, koulutustaustan, ammattitaidon ja teknisen osaamisen yhdistelmää. Eli mitkä ovat data-analyytikon tärkeimmät kompetenssit?

Johdannon jälkeen toisessa luvussa aukaistaan työn keskeiset käsitteet, data-analyytikko ja kompetenssi. Kolmannessa luvussa tarkastellaan tutkimuskirjallisuuden avulla, millaista osaamista ja millaisia ominaisuuksia on pidetty tärkeinä data-analyytikolle. Alakohtaiset blogit tuovat kuitenkin aiheeseen tuoreimman näkökulman top-listoineen, ja niihin keskitytään neljännessä luvussa. Viidennessä luvussa luodaan katsaus kotimaisiin työpaikkailmoituksiin, joilla yritykset data-analyytikkoja rekrytoivat ja tarkastellaan, ovatko ne vaatimuksiltaan yhteneviä kirjallisuuden ja blogien kanssa. Loppuyhteenvetossa kootaan tärkeimmät tulokset ja kerrotaan tiivistetysti lukijalle, millainen on tämän päivän data-analyytikko.

## 2. KÄSITTEET

### 2.1 Data-analyytikko

Data-analyytikko (*data analyst*) vai datatieteilijä (*data scientist*)? Ovatko ne eri ammattitejia vai kaksi eri nimitystä samasta ammatista? Kansainvälisissä teksteissä nämä kaksi käsitettä usein erotetaan toisistaan, mutta suomenkielisissä ei niinkään. Siksi käytän nyt eroja selvittäessäni englanninkielisiä termejä *data analyst* ja *data scientist*. Nimikkeiden eroja ja yhtäläisyyksiä ovat kirjoituksissaan pohtineet muun muassa Dalal (2016), Han Lee (2014) sekä Rivera ja Haverson (2014). Nimikekirjoa laajentavat vielä datan syntymisen ja keräämisen alkuvaiheissa tärkeässä osassa olevat *data engineer* ja *data architect*. Kuvassa 1 näkyy yksi tapa erotella, missä vaiheessa datan elinkaarta data-ammattilaisten roolit ovat vahvimmillaan.



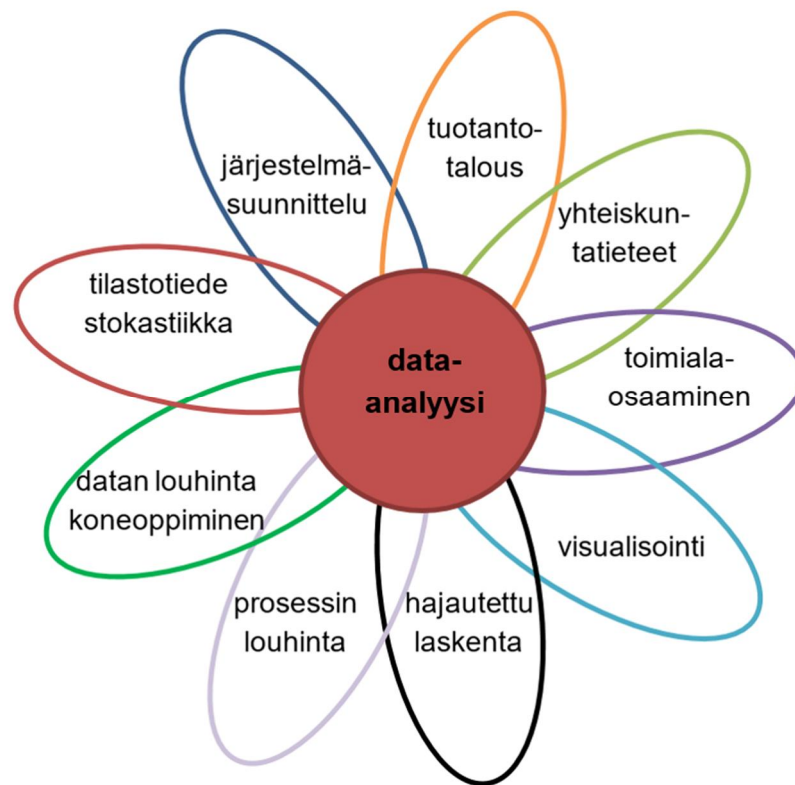
**Kuva 1.** Data-analyytikon roolit datan elinkaaren varrella (Rivera ja Haverson, 2014)

*Data architect* luo arkkitehtuurin datan esittämiseksi, tallentamiseksi ja tiedonsiirrolle. *Data engineer* huolehtii arkkitehtuurin toimivuudesta, ylläpidosta ja testaamisesta. *Data analyst* tulkitsee yleensä menneisyydessä ja nykyhetkessä kerättyä dataa luoden siitä raportteja ja yhteenvetoja päätöksenteon tueksi. Kysymykset, joihin raportit vastaavat, tulevat yleensä ulkoapäin – esimerkiksi yritysjohtolta. *Data analystin* pitää kyetä esittämään ja tulkitsemaan data yritysjohtolle ymmärrettävästi. *Data scientist* luo edellisen lisäksi datasta matemaattisia malleja, joiden perusteella voidaan tuottaa ennusteita tulevaisuutta varten ja reagoida niihin asianmukaisilla toimenpiteillä. *Data scientist* ei saa kysymyksiä välttämättä valmiina, vaan voi joutua myös luomaan ne itse. (Rivera ja Haverson, 2014; Dalal, 2016.) Molemmilla on kuitenkin yhteinen tavoite: isosta sekavasta datamäärästä pitää saada selvitettyä juuri se oleellinen ydin.

Suomalaisissa kirjoituksissakin data-analyytikko ja datatieteilijä joskus erotetaan toisistaan, esimerkiksi Pinola (2015) datatieteilijän kompetenssien määrittelyssä tutkimuksessa

pro gradu -tutkielmassaan. Esseessäni keskityn kuitenkin yleisesti data-analytiikan parissa työskentelevien asiantuntijoiden kompetensseihin ja tässä kontekstissa termiä data-analyttikko käytetään usein ammatillisena yleisnimikkeenä, joka sisältää myös datatieteilijän tehtävät. Työpaikkailmoituksen otsikossa näkeekin usein merkinnän ”*data-analyttikko/data scientist*”. Jatkossa siis data-analyttikko pitää sisällään sekä *data analystin* että *data scientistin* (joskus jopa *data engineerinkin*) ja on datan asiantuntija, joka analysoi, tulkitsee ja jalostaa dataa sitä tarvitsevien käyttöön.

Van der Aalst (2014) kokoaa havainnollisesti data-analyttikon profiilin esittämällä data-analyysin paikan tieteen kentällä (kuva 2). Data-analyysia voidaan pitää uutena tieteenalana, joka pohjautuu moneen eri tieteenalaan. Näin data-analyttikonkin pitää olla luova moniosaaja, jolla on vahva tilastotieteellinen ja tekninen osaaminen.

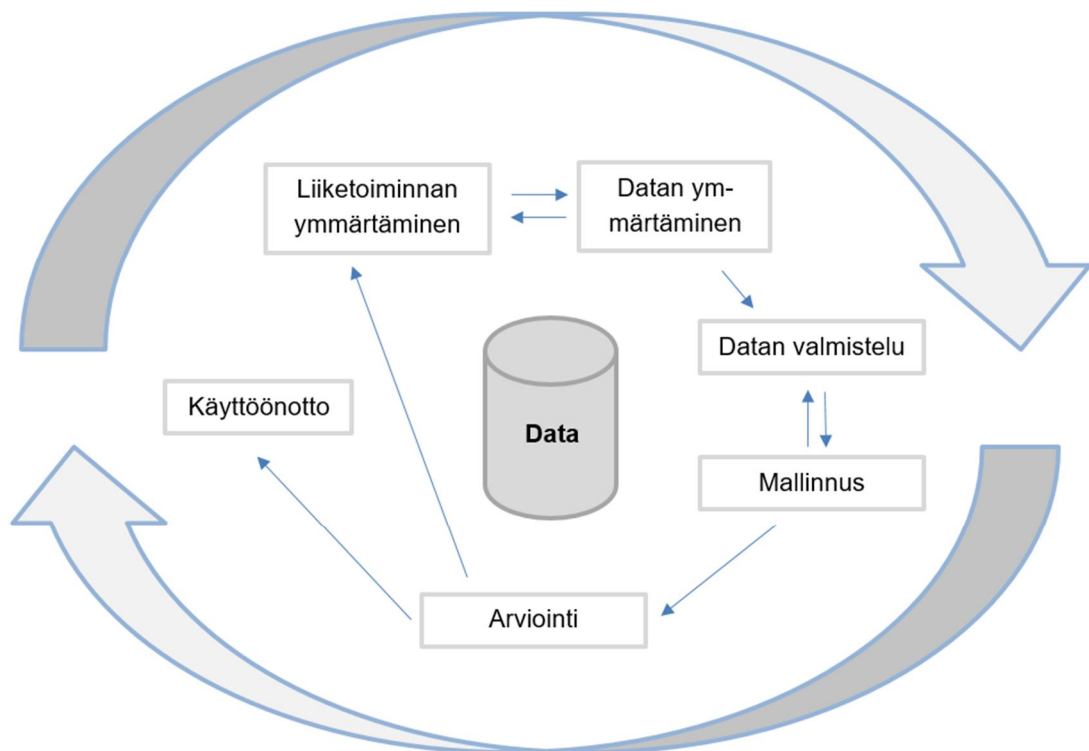


**Kuva 2.** Data-analyttikon profiili (van der Aalst, 2014, s. 22)

Kun data-analyttikon kohdalla puhutaan käsiteltävästä datasta, tarkoitetaan usein big dataa (toisinaan suomennettu massadataksi) eli hyvin suurta datamäärää, jota ei ole mahdollista analysoida perinteisin menetelmin tai perinteisillä analysointityökaluilla. Big datalla on kolme tärkeää ulottuvuutta: dataa on todella *paljon*, sitä kehittyi *nopeasti* (ja siihen pitää reagoida nopeasti) ja se on *monimuotoista* (tekstiä, kuvaa, ääntä ym.). Tästä ei

kuitenkaan pidä tehdä johtopäätöstä, että data-analytiikka on pelkkää big dataa – perinteisen ”small datan” merkitys ei ole kadonnut mihinkään. (Russom, 2011)

Data-analyttikon tehtävänä on yleensä louhia big datasta irti se oleellinen, jonka perusteella päätöksiä pystytään tekemään. Tämä tiedonlouhinnan prosessi noudattaa useimmiten kuvassa 3 näkyvää ns. CRISP-DM-mallia (Shearer, 2000). Liiketoiminnan ymmärrysvaiheessa määritellään kysymys, joka halutaan ratkaista. Datan ymmärrysvaiheessa selvitetään, millaista dataa on käytössä ja onko kysymys mahdollista ratkaista sen avulla. Datan valmisteluvaiheessa data siistitään ja saatetaan mallinnuksen vaatimaan muotoon. Mallinnuksessa data analysoidaan matemaattisin ja tilastollisin menetelmin ja pyritään saamaan aikaiseksi malli, joka ratkaisee kysymyksen parhaiten. Vaiheet voivat vaatia lukuisia iteraatioita. Arviointivaihe varmistaa, että on ratkaistu juuri alkuperäinen, oikea kysymys ja kertoo mallin virhemahdollisuuden. Lopuksi malli otetaan käyttöön ja jatketaan toiminnan kehittämistä. (Markkula, Syväniemi ja Suomela, 2015, ss. 95–96)



**Kuva 3.** Tiedonlouhinnan CRISP-DM-malli (Shearer, 2000).

Kuten prosessista voidaan huomata, eri vaiheet tarvitsevat erilaista osaamista. Johonkin tarvitaan vahvaa liiketoimintaosaamista, johonkin vahvaa järjestelmäosaamista, johonkin vahvaa matemaattista osaamista jne. Ei ole siis vaikea arvata, että data-analyysissa on usein kyse tiimityöskentelystä.

## 2.2 Kompetenssi

Kielitoimiston sanakirjan mukaan kompetenssi tarkoittaa pätevyyttä, kelpoisuutta tai viranomaisen toimivaltaa. Ammatillinen kompetenssi puolestaan tarkoittaa henkilön kykyä suoriutua ammattiin liittyvistä työtehtävistä (Ruohotie, 2005). Esseessä tarkastellaan ammatillisia kompetensseja seuraavista näkökulmista:

- Millaisia henkilökohtaisia ominaisuuksia data-analyytikolta odotetaan?
- Millaista pohjakoulutusta data-analyytikolta vaaditaan? Entä millaisia alakohtaisia ammattitaitovaatimuksia on (ohjelmointitaidot ym.)?
- Millaista teknistä osaamista ja työkaluosaaamista (ohjelmointikielet ym.) data-analyytikolla odotetaan olevan?

Pinola (2015) on pro gradu -tutkielmassaan käyttänyt Halvelkan ja Merhoutin (2009) IT-ammattilaisen kompetenssi viitekehystä, jossa liiketoimintaosaaminen on erotettu omaksi kategoriakseen henkilökohtaisten ominaisuuksien, ammatillisten taitojen ja teknisen osaamisen ohelle. Tämä viitekehys saattaa antaa käsityksen, että data-analyytikko toimisi avustamassa ainoastaan liiketoiminnan päätöksenteossa. Data-analyytikon osaaminen on avainasemassa myös monessa muussa analytiikkaa vaativassa päätöksenteossa ja linjauksessa, joten tutkin esseessäni liiketoimintaosaamista vain yhtenä osana data-analyytikon yleistä ammattitaitoa.



## 3. DATA-ANALYYTIKON KOMPETENSSIT KIRJALLISUUDESSA

### 3.1 Henkilökohtaiset ominaisuudet

Henkilökohtaisia ominaisuuksia ei voi saavuttaa sopivalla pohjakoulutuksella. Ne ovat osa identiteettiä, mutta ne voivat ajan kuluessa myös jalostua. Tietyt ominaisuudet voivat suunnata ihmisen kiinnostumaan esimerkiksi data-analytiikasta (Havelka ja Merhout, 2009).

Tärkeä data-analyytikolta vaadittava ominaisuus on halu, jopa intohimo, oppia ja rakentaa uutta (Davenport ja Patil, 2012). Uudella ja kehittyvällä alalla tämä on välttämätön ominaisuus. Vaikka käytettävät algoritmit saattavatkin olla joskus vanhoja ja yleisimmät ohjelmointikielet myös, pitää olla koko ajan valmis ottamaan käyttöön jotain uutta ja entistä nopeampaa. Tämä liittyy läheisesti myös ongelmanratkaisuhalukkuuteen. Vanhoilla itselle tutuilla menetelmillä ei löydy aina parasta ratkaisua ongelmiin. Tarvitaan halua kokeilla uutta ja soveltaa vanhoja menetelmiä uusilla alueilla.

Kansainvälisissä työpaikkailmoituksissa onkin perinteisesti pidetty tärkeimpänä henkilökohtaisena ominaisuutena juuri luovuutta. Nykyään kaikille aloille tärkeitä paineensietokyky ja joustavuus ovat tarpeellisia myös data-analyytikolle. Henkilökohtaiset ominaisuudet eivät ole työpaikkailmoituksissa juurikaan korostuneet. Vaatimukset ovat keskittyneet selkeästi tekniseen osaamiseen ja ammattitaitoon. (Pinola, 2015)

### 3.2 Koulutustausta ja ammattitaito

Tähän kategoriaan liittyvät ominaisuudet ja taidot ovat sellaisia, jotka yleensä opitaan koulutuksen ja kokemuksen kautta, mutta eivät liity suoranaisesti data-analyysin alaan. Näitä taitoja on mahdollisuus kehittää myös oman kiinnostuneisuuden kautta.

Data-analytiikan käyttö yhdistetään usein liiketoiminnan kehittämiseen. Liiketoimintaosaamista eli ymmärrystä yleisistä liiketoimintaprosesseista ja -käytänteistä pidetäänkin kirjallisuudessa yleensä data-analyytikolle tärkeänä taitona (Davenport ja Patil, 2012; Pinola, 2015). Data-analytiikka on kuitenkin ulottunut jo miltei kaikille: epidemioiden aikainen havaitseminen, luottokorttivarkaudet, sääennusteet, itse ajavat autot, Facebookin ystäväsuositukset, Amazonin kirjasuositukset, Netflixin elokuvasuositukset jne. (Granville, 2015). Topin ja Markuksen (2015) mukaan data-analytikkojen onkin tärkeää ymmärtää, mitä heidän analytiikkansa avulla tehdyt päätökset tai ennusteet voivat saada aikaan. Tulosten vääristymien arvioimisen lisäksi analytikolla pitää olla kykyä eettiseen pohdintaan.

Vaikka toimialakohtaista osaamista painotetaankin usein kirjallisuudessa, työpaikkailmoituksissa sitä on vaadittu Pinolan (2015) tutkimuksen mukaan melko vähän. Data-analytiikan ympärille kerätäänkin nykyään yhä enemmän isompia tiimejä, joissa data-analyttikoiden ja toimiala-asiantuntijoiden roolit yritetään tuoda yhteen ja saada täydentämään toisiaan (Viaene, 2013).

Varsinaisten data-analyttikkojen kouluttaminen on vielä niin uusi asia, että sitä ei oikein voida pohjakoulutuksena työhön edellyttää. Pohjakoulutusvaatimukset ovat yleensä keskittyneet luonnontieteisiin ja tekniikkaan, vähintään maisteritason koulutukseen. Esimerkiksi fyysikkojen on katsottu sopivan alalle, koska he ovat saattaneet työssään jo hoitaa vastaavia prosesseja datankeräämistyökalujen suunnittelusta lähtien. Data-analyttikolta odotetaan yleensä myös muutaman vuoden työkokemusta. (Davenport ja Patil, 2012; Pinola, 2015)

### 3.3 Tekninen osaaminen ja työkalut

Tekninen osaaminen tarkoittaa tässä lähinnä tietotekniikkaan ja data-analyysiin liittyvää alakohtaista osaamista. Se pitää siis sisällään myös eri ohjelmointikieliin sekä sovellusten, työkalujen ja työskentelyalustojen käyttöön liittyvän osaamisen. Tekninen osaaminen omaksutaan yleensä koulutuksen kautta, mutta se kehittyy huippuunsa vasta kokemuksen myötä. Havelkan ja Merhoutin (2009) IT-ammattilaisen kompetenssiivitekehyydessä tosin tekniseen osaamiseen ei ole sisällytetty yksittäisten työkalujen (esim. ohjelmointikielten) erityisosaamista, eikä niitä tutkimuskirjallisuudessa kompetenssien yhteydessä juurikaan käsitellä. Kirjallisuudessa keskitytään taitoihin, jotka eivät ole riippuvia työkalusta, eivätkä näin ollen ole niin herkkiä trendeille. Yksittäiset työkalut ovat kuitenkin kirjoitukseni tavoitteiden kannalta hyvin mielenkiintoisia ja niitä käsitellään enemmän seuraavassa luvussa blogitekstien yhteydessä – jonkin verran kuitenkin jo tässä.

Data-analyttikkona toimimisen perusedellytyksiä ovat tilastotieteelliset ja analyttiset taidot sekä koneoppimisen ja tiedonlouhinnan ymmärrys. Tämä pitää sisällään myös tärkeimpien big datan käsittelyyn liittyvien algoritmien osaamisen. Ohjelmointitaito on välttämätöntä data-analyttikolle. (Benda, 2012; Davenport ja Patil, 2012; Ohlhorst, 2013; Pinola, 2015)

Kriittisimpänä taitona pidetään usein kykyä tulkata datan merkityksellisyys sitä tarvitseville tavalla, jonka he ymmärtävät – ja mielellään vielä ymmärtävät helposti (Ohlhorst, 2013). Tämä on itse asiassa kaiken yhteen nivova taito ja edellyttää vahvaa osaamista kaikilla kompetenssialueilla. Yksi tärkeä osa datan merkityksellisyyden tulkkauksista on taito visualisoida dataa ymmärrettävästi.

Kaikki edellä mainittu näkyy hyvin myös data-analyttikon työnhakuprosessissa. Käydessään läpi kansainvälisiä työpaikkailmoituksia Pinola (2015, ss. 39–41) huomasi yli puo-

lissa ilmoituksista vaaditun tilastotieteellistä ja analyttistä osaamista, hyviä kommunikointitaitoja sekä osaamista koneoppimisesta. Liiketoiminta- ja ohjelmointitaitoja odotettiin n. 40 %:ssa ilmoituksista. Myös tiedonlouhintataitoja, algoritmien osaamista, ongelmanratkaisukykyä ja ennustavaa analysointia pidettiin tärkeinä. Data-analyttikko työskentelee usein osana isompaa tiimiä, joten tiimityöskentelyn pitää sujua hyvin.

Data-analyttikolta vaadittava ohjelmointiosaaminen eroaa normaalin ohjelmoijan vaatimuksista. Tärkeässä asemassa ovat ohjelmointikielien ja tekniikat, jotka sopivat tilastolliseen analysointiin ja joilla voidaan nopeuttaa isojen tietomäärien käsittelyä. Ohjelmointikielistä painotetaan usein R-kieltä, joka on kehitetty nimenomaan tilastotieteelliseen ohjelmointiin. R:n yksi vahvuus verrattuna muihin käytettyihin matematiikkaohjelmointikieliin on sen ilmaisuus – se on open source -ohjelmana vapaasti ladattavissa verkosta ja näin kuka tahansa voi myös kehittää ohjelmaa. (Granville, 2015; Pinola, 2015, ss. 39–40; *R-kieli*, 2017)

Useimmin mainittu ohjelmointikieli on kuitenkin Python. Python on monipuolinen ja sisältää paljon työkaluja big datan käsittelyyn. Pythonia pidetään myös melko helppona kielenä oppia. Myös muita korkeamman tason ohjelmointikieliä (etenkin Java) on mainittu tärkeinä työkaluina data-analyttikolle (Benda, 2012, s. 314; Pinola, 2015, ss. 39–40).

Tietokantaosaamisesta olennaista on varsinkin relaatiotietokantojen SQL-kyselykielen hyvä hallinta. Myös big datan käsittelemiseen ja pilvipalveluihin suunniteltujen tekniikoiden, kuten Hadoopin ja MapReducen, tunteminen on tärkeää. Kuten hyvän ohjelmoijan yleensä, data-analyttikonkin tulisi pystyä ohjelmoidessaan työskentelemään ilman graafista käyttöliittymää pelkän komentorivikäyttöliittymän avulla. Samoin erilaisten ohjelmointirajapintojen käytön pitäisi olla tuttua (Benda, 2012, s. 314; Granville, 2015).

## 4. DATA-ANALYYTIKON KOMPETENSSIT BLOGEISSA

Blogitekstit ovat erittäin mielenkiintoisia ajankohtaisuutensa vuoksi, jos halutaan saada selville tämän hetken trendejä data-analyytikon osaamisvaatimuksissa. Tässä osiossa läpikäytävät blogit on saatu googlaamalla hakulausekkeilla *data science skills* ja *data scientist skills*. Haut luonnollisesti sisälsivät paljon päällekkäisiä tuloksia. Mukaan on otettu yhdistettyjen hakutulosten 12 parasta tulosta. Taitojen luokittelussa on pyritty noudattamaan mahdollisimman paljon samanlaista jaottelua kuin teoriaosuuden kompetensseissakin. Tulokset näkyvät taulukossa 1.

*Taulukko 1. Data-analyytikon osaamisvaatimusten maininnat blogeissa.*

<b>Henkilökohtaiset taidot</b>	
Analyttiset taidot ja ongelmanratkaisutaidot, kyseenalaistaminen, uteliaisuus	12
<b>Ammattitaito</b>	
Visualisointi- ja esittämistaidot, kommunikointitaidot	11
Liiketoimintaosaaminen ja toimialaosaaminen	13
<b>Tekninen osaaminen</b>	
Tilastotiede, määrällinen analyysi, datan käsittely	11
Matematiikka, optimointi	7
Ohjelmointitaidot (yleisesti)	7
Tiedonlouhinta, koneoppiminen, algoritmiosaaminen	15
SQL (rakenteellisen tiedon käsittely)	9
Big data -työkalut (esim. Hadoop, Hive, Pig, Spark)	8
Python	9
R	7
Java	5

Taulukoon on otettu mukaan kaikki taidot, jotka saivat blogeissa enemmän kuin neljä mainintaa. Blogeissa korostetut kompetenssit noudattavat hyvin kirjallisuuden linjauksia, mutta ovat teknisen osaamisen suhteen yksityiskohtaisempia. Kaikissa blogeissa ei oltu

erikseen mainittu yleistä ohjelmointitaitoa, vaan oli suoraan listattu tärkeimpiä ohjelmointikieliä ja työkaluja, jotka data-analyytikon on syytä hallita. Yleisen ohjelmointiosaamisen saamat maininnat eivät siis anna siinä mielessä oikeaa kuvaa sen tärkeydestä. Jos ohjelmointitaitoon lasketaan mukaan ohjelmointityökalujen saamat maininnat, sen todellinen tärkeys nousee esille.

Mielenkiintoinen yksityiskohta blogiteksteissä oli iso painotus rakenteellisen tiedon käsittelyn osaamiseen. Data-analytiikkahan yhdistyy perinteisesti big datan kautta nimenomaan rakenteettomaan tietoon (Chen, Mao ja Liu, 2014, s. 186). SQL-osaaminen näyttäisi kuitenkin olevan data-analyytikolle useimmin mainittu yksittäinen osaamisvaatimus 15 maininnalla ja NoSQL-osaaminen tuli kaukana perässä kolmella maininnalla jääden kokonaan taulukon ulkopuolelle.

Toinen huomiota kiinnittävä seikka oli blogiteksteissä esille nostettu, mutta kirjallisuudessa vähälle huomiolle jäänyt tieteellinen lähestymistapa. Se ei aivan yltänyt taulukkoon asti, mutta sai kuitenkin neljä mainintaa. Data-analytikolta vaaditaan yleensä akateemista koulutusta, joten tieteellisen lähestymistavan ehkä ajatellaan tulevan sitä kautta. Saattaa olla, että tohtorin tutkinto on tulevaisuudessa varsin hyödyllinen data-analytikolle.

## 5. KATSAUS KOTIMAIISIIN TYÖPAIKKAILMOITUKSIIN

Tenuton (2016) mukaan LinkedIn-palvelun avointen työpaikkailmoitusten perusteella vuoden 2016 odotetuimmat työkaluihin liittyvät taidot data-analyytikoilta olivat suosituimmuusjärjestyksessä seuraavat:

1. SQL
2. Hadoop
3. Python
4. Java
5. R
6. Hive
7. MapReduce
8. NoSQL
9. Pig
10. SAS.

Tämä tukee kirjallisuuden ja blogien perusteella tehtyjä havaintoja, vaikka Hadoop-osaaminen onkin listalla yllättävän korkealla. Myös Hadoopin SQL-rajapinta Hive on mainittu LinkedInin työpaikkailmoituksissa hyvin usein. Blogeissa tällaisten työkalujen oli ehkä ajateltu sisältyvän muihin osaamisvaatimukseen, joko SQL-osaamiseen tai yleisiin big data -työkalujen käyttötaitoihin.

Kiinnostavinta on kuitenkin se, mitä osaamista data-analyytikolta vaaditaan kotimaisissa työpaikkailmoituksissa. Hakukoneena työpaikkailmoitusten kartoituksessa käytettiin ensin Duunitori-palvelua, joka lupaa listata ”*kaikki työpaikat yhdellä haulla*” (Duunitori, 2017). Listaa kuitenkin täydennettiin Monster.fi-palvelun ja TE-palvelujen työpaikkahauilla, sillä ne listasivat paikkoja, joita Duunitorin listalla ei ollut (Monster.fi, 2017, TE-palvelut, 2017). Työpaikkojen hakemisessa käytettiin hakusanoja *data-analyttikko*, *data analyst* ja *data scientist*. Hakutulokset sisältävät laajasti työpaikkoja datan elinkaaren kaikista ammattilaisrooleista eli data engineer, data analyst ja data scientist. Haku tehtiin 2.7.2017.

Kotimaisissa työpaikkailmoituksissa esiintyneet kompetenssit ovat melko lailla yhtäläiset kirjallisuudessa ja blogeissa esiintyneiden kanssa. Ilmoituksissa esitetään kuitenkin usein melko yksityiskohtaisia listoja työkaluista, joiden osaamista hakijoilla toivottaisiin olevan. Tämä teki kompetenssitaulukon täyttämistä haasteellista. Etenkin big datan kanssa työskentelemiseen suunnitellut työkalut korostuvat ilmoituksissa, joten niistä eniten mainitut on hyvä saada näkymään itsenäisinä. Taulukossa 2 onkin jätetty edellisen kappaleen taulukon luokka big data -työkalut kokonaan pois ja loppuun on listattu

erikseen yleisimmät big datan ja pilvipalvelujen työkalut (esim. Hadoop, Spark, AWS jne.).

**Taulukko 2.** Data-analyytikon osaamisvaatimusten maininnat työpaikkailmoituksissa.

<b>Henkilökohtaiset taidot</b>	
Analyttiset taidot ja ongelmanratkaisutaidot, kyseenalaistaminen, uteliaisuus	12
<b>Ammattitaito</b>	
Visualisointi- ja esittämistaidot, kommunikointitaidot	15
Liiketoimintaosaaminen ja toimialaosaaminen	8
Tiimityöskentelytaidot	7
<b>Tekninen osaaminen</b>	
Tilastotiede, määrällinen analyysi, datan käsittely	14
Matematiikka, optimointi	3
Ohjelmointitaidot (yleisesti)	8
Tiedonlouhinta, koneoppiminen, algoritmiosaaminen	10
SQL (rakenteellisen tiedon käsittely)	9
Python	9
R	7
Java	7
<b>Tekninen osaaminen - pilvi- ja big data -työkalut</b>	
Spark	8
Hadoop	7
Azure	5
Aws	4

Big data -työkaluista Hive-rajapinta mainittiin yllättäen vain kerran, vaikka se oli LinkedInin listauksessa sijalla kuusi. Ohjelmointikielistä taulukkoon päässeiden lisäksi vaadittiin muutamassa ilmoituksessa myös tilastolliseen ohjelmointiin suunnitellun SAS-kielen hallintaa. Data-analyytikon työskentely-ympäristö on kansainvälinen, joten kielitaitovaatimuksina on lähes poikkeuksetta englanti ja suomi – tässä järjestyksessä. Työpaikkailmoituksissa tiimityöskentelytaidot korostuivat huomattavasti enemmän kuin kirjallisuudessa ja blogeissa, joten se on otettu ammattitaitovaatimukseen omana kokonaisuutenaan.

Matematiikkaa ei mainittu vaatimuksissa kovinkaan usein, ainoastaan muutaman kerran pohjakoulutuksessa. Muut esiintyneet pohjakoulutusvaatimukset olivat tilastotiede, (tieto)tekniikka, fysiikka tai liiketalouspainotteisissa työpaikoissa taloustiede. Nämä koulutukset sisältävät runsaasti matematiikkaa, joten sen mainitsemista erikseen ei ehkä siksi katsottu tarpeelliseksi.



## 6. YHTEENVETO

Tieteellisten tutkimusten, ammatillisten blogien ja työpaikkailmoitusten tarjoaman tiedon perusteella tämän päivän data-analyytikolta vaadittu osaaminen ja ominaisuudet voidaan tiivistää seuraavasti. Data-analyytikolla pitää olla analyyttinen ote ongelmiin ja intohimo ratkaista ongelmat. Hän pystyy visualisoimaan ja esittämään ratkaisunsa niin, että vastaanottajan on helppo ymmärtää se. Data-analyytikolla on vahva tilastotieteellinen osaaminen ja kyky käsitellä isoja datamääriä – ja tarmoa siivota isoja datamääriä. Suurin osa data-analyytikon työajasta on nimittäin datan järjestelyä ja siivoamista (Press, 2016).

Tilastotieteellisen osaamisen ohella data-analyytikon pitää olla hyvä ohjelmoimaan. Ohjelmoinnissa tärkeimmät työkalut ovat SQL, Python, R ja Java. Yleisten ohjelmointitaitojen lisäksi vaaditaan vahvaa osaamista tiedonlouhinnasta, koneoppimisesta sekä algoritmeista. Koska data-analytiikka on nykyään paljolti big datan käsittelyä, on hyvä hallita myös siihen liittyviä tekniikoita, kuten Spark ja Hadoop.

Viaenen (2013) mukaan data-analytiikka on yhä enemmän tiivistä tiimityötä, jossa data-analytikot työskentelevät yhdessä toimiala-asiantuntijoiden kanssa. Tämä ei ole lainkaan vähentänyt töihin pyrkivän analyytikon toimiala- ja liiketoimintaosaamisen tarvetta – se näkyy yhä vahvasti työpaikkailmoituksissa. Ilmoituksissa näkyy kuitenkin myös se, että data-analyytikolla ei tarvitse olla kaikkia tärkeimpiä ominaisuuksia valmiina, vaan tiimit muodostuvat kunkin tiimin jäsenen omien vahvuuksien yhdistelmänä.

## LÄHTEET

van der Aalst, W. M. P. (2014) ”Data Scientist: The Engineer of the Future”, teoksessa *Enterprise Interoperability VI*. Cham: Springer International Publishing, ss. 13–26. doi: 10.1007/978-3-319-04948-9\_2.

Aalst, W. M. P. Van Der (2014) ”Data Scientist: The Engineer of the Future”, teoksessa Mertins, K., Bénaben, F., Poler, R., ja Bourrières, J. (toim.) *Proceedings of the I-ESA Conferences 7*, DOI: 10.1007/978-3-319-04948-9\_2. Cham: Springer International Publishing, ss. 13–26. doi: 10.1007/978-3-642-36796-0.

Benda, R. (2012) ”Science of Big Data: Background and Requirements”, teoksessa Pavelkova, D., Strouhal, J., Paskova, M., ja Suchacek, J. (toim.) *Advances in economics, risk management, political and law science : proceedings of the 1st WSEAS International Conference on Economics, Political and Law Science (EPLS '12)*. Athens: WSEAS, ss. 311–316.

Chen, M., Mao, S. ja Liu, Y. (2014) ”Big Data: A Survey”, *Mobile Networks and Applications*. Springer US, 19(2), ss. 171–209. doi: 10.1007/s11036-013-0489-0.

Dalal, K. (2016) *Data Analyst vs Data Scientist - What's the Difference?* Saatavissa: <https://www.simplilearn.com/data-analyst-vs-data-scientist-article> (Viitattu: 16. kesäkuuta 2017).

Davenport, T. H. ja Patil, D. J. (2012) ”Data scientist: the sexiest job of the 21st century: meet the people who can coax treasure out of messy, unstructured data”, *Harvard Business Review*, 90(October), ss. 70–77.

*Duunitori* (2017). Saatavissa: <https://duunitori.fi/> (Viitattu: 2. heinäkuuta 2017).

Granville, V. (2015) *Developing Analytic Talent: Becoming a Data Scientist*, John Wiley & Sons, Inc. Wiley. doi: 10.1017/CBO9781107415324.004.

Han Lee, C. (2014) *3 Data Careers Decoded and What It Means for You | Udacity*. Saatavissa: <http://blog.udacity.com/2014/12/data-analyst-vs-data-scientist-vs-data-engineer.html> (Viitattu: 16. kesäkuuta 2017).

Havelka, D. ja Merhout, J. (2009) ”Toward a theory of information technology professional competence”, *Journal of Computer Information Systems*, 50(2), ss. 106–116. doi: 10.1080/08874417.2009.11645389.

Markkula, T., Syväniemi, A. ja Suomela, S. (2015) *Analytiikkamatka : datasta tietoon ja tiedolla johtamiseen*.

*Monster.fi* (2017). Saatavissa: <https://www.monster.fi> (Viitattu: 2. heinäkuuta 2017).

Ohlhorst, F. (2013) *Big data analytics turning big data into big money*. John Wiley & Sons.

Pinola, T. (2015) *Datatieteilijän kompetensien määrittelyminen*. Jyväskylän yliopisto.

Press, G. (2016) *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says, Forbes*. Saatavissa: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#4a28fddb6f63> (Viitattu: 2. heinäkuuta 2017).

*R-kieli* (2017) *R-ohjelmointi.org*. Saatavissa: [http://www.r-ohjelmointi.org/?page\\_id=14](http://www.r-ohjelmointi.org/?page_id=14) (Viitattu: 26. kesäkuuta 2017).

Rivera, R. ja Haverson, A. (2014) *Data Scientist vs Data Analyst*. Saatavissa: <https://www.captechconsulting.com/blogs/data-scientist-vs-data-analyst> (Viitattu: 16. kesäkuuta 2017).

Ruohotie, P. (2005) ”Ammatillinen kompetenssi ja sen kehittäminen”, *Ammattikasvatuksen aikakauskirja*, 7(3).

Russom, P. (2011) *Big data analytics, TDWI Best Practices Report*. doi: 10.1109/ICCICT.2012.6398180.

Shearer, C. (2000) ”The CRISP-DM Model: The new blueprint for data mining”, *Journal of Data Warehousing*, 5(4).

*TE-palvelut* (2017). Saatavissa: <http://paikat.te-palvelut.fi> (Viitattu: 2. heinäkuuta 2017).

Tenuto, J. (2016) *What skills should data scientists have in 2016? | CrowdFlower, Crowdflower blog*. Saatavissa: <https://www.crowdfunder.com/what-skills-should-data-scientists-have-in-2016/> (Viitattu: 2. heinäkuuta 2017).

Topi, H. ja Markus, M. L. (2015) ”Educating Data Scientists in the Broader Implications of their Work”, *Journal of the Washington Academy of Sciences*, 101(3), ss. 39–48.

Viaene, S. (2013) ”Data scientists aren’t domain experts”, *IT Professional*, 15(6), ss. 12–17. doi: 10.1109/MITP.2013.93.