



## Data-analyysin asiantuntijoille riittää kysyntää

*Janne Liuttu*

toimitusjohtaja, Andumus Oy

### Johdanto

Andumus Oy on suomalainen asiantuntijayritys, jonka pelikenttänä on yritysten keräämien tietomassojen analysointi. Asiakkaitamme ovat suuret suomalaiset ja kansainväliset yritykset ja yhteisöt, ja palveluksessamme on kuusi enemmän tai vähemmän matemaattisen taustan omaavaa asiantuntijaa. Itse valmistuin vuonna 2012 Tampereen teknillisestä yliopistosta pääaineenani matematiikka, jonka lisäksi opiskelin teknillistä fysiikkaa sekä teollisuustaloutta. Andumuksen palveluksessa olen ollut syksystä 2011 asti, ensimmäiset vuodet analytikkona, ja toimitusjohtajan roolissa syksystä 2013 alkaen. Seuraavassa käyn läpi yleisiä trendejä, muutamia omia kokemuksiani tiedoista ja taidoista jotka ovat tarpeen tällä alalla, sekä muutamia vinkejä, joiden avulla on mahdollista erottua muiden työnhakijoiden joukosta.

### Yleinen trendi

Big data on tämän päivän muotisana. Tietoa muodostuu alati kasvavalla tahdilla, ja kaikkein konservatiivisimmatkin arviot ennustavat tallennetun datan määrän lisääntyvän 40 % vuosittain. Vuonna 2015 maailman datamäärän arvioidaan lähentelevän 10 zettatavua eli  $10^{22}$  tavua. Tämä kehitys on aiheuttanut kasvavan kysynnän asiantuntijoille, jotka pystyvät louhimaan näistä datamassoista hyödyllistä informaatiota. Nykyisin jo keskikokoisella suomalaisella yrityksellä on mahdollista saavuttaa merkittävää kilpailuetua tehokkaalla datan hyödyntämisellä, puhumattakaan Googlen, Facebookin ja Amazonin kaltaisista valtavista globaaleista toimijoista.

Suuren kysynnän vuoksi työmarkkinoilla on syntymässä laajeneva rako kysynnän ja tarjonnan välillä, ja pel-

kästään Yhdysvalloissa arvioidaan olevan vuoteen 2018 mennessä 140 000 – 190 000 työntekijän vaje datan analysointiin erikoistuneista asiantuntijoista. Näiden lisäksi tarvitaan jopa 1,5 miljoonaa uutta johtajaa, jotka pystyvät tekemään päätöksiä analysoidun datan perusteella. Kovimmista osajista on näin ollen kova kilpailu, ja näiden ihmisten palkat ovat myös sen mukaisia.

### Matemaattiset taidot

Matemaattiset taidot ovat luonnollisesti se perusta, jonka päälle muiden taitojen rakentaminen tapahtuu. Perusasiat tulee olla hyvin hallussa, ja peruskursseilta saatava kattaus esimerkiksi logiikkaa, algebraa ja analyysiä antaa tukevan pohjan, jolle omaa osaamista voi rakentaa. Perusasioiden hyvän hallitsemisen lisäksi suositeltavia ovat ennen kaikkea tilastotieteen ja tilastollisen analyysin opinnot, joiden lisäksi esimerkiksi mallinnus ja optimointi. Eksaktien ratkaisujen etsiminen on usein mahdotonta tai liian hidasta, joten myös numeerisen analyysin ja suurteholaskennan hallitseminen on eduksi.

Teoreettisen pohjan lisäksi kannattaa panostaa soveltaviin kursseihin, joissa pääsee hyödyntämään opittuja taitoja käytännön ongelmien ratkaisemiseen. Harjoitustyöt ovat usein melko työläitä, mutta näistä on yleensä myös eniten hyötyä myöhemmin työelämässä. Todellisten ongelmien ratkaiseminen vaatii soveltamista ja luovuutta, sillä selkeitä valmiita ratkaisuja ei aina ole olemassa. Minulle sanottiin aikanaan töihin mennessäni ohjeeksi, että ”Tilastotiede pitää olla niin hyvin hallussa, että tietää mitä sääntöjä voi rikkoa”. Tässä piilee totuuden siemen, sillä todelliset ilmiöt harvoin noudattavat normaali- tai ylipäätään mitään hyvin käyttäytyvää jakaumaa. Yleisestikin usein on hyö-

dyllisempää löytää äärellisessä ajassa jokin ratkaisu, kuin käyttää moninkertaisesti aikaa täydellisen ratkaisun etsimiseen.

## IT-taidot

Käytännön analyysityö tapahtuu nykyisin aina jonkinlaisella IT-järjestelmällä, joten ainakin jonkin näistä käyttö on hyvä hallita. Analyysiohjelmistoista yleisimmin käytetyt ovat SPSS ja SAS, ja näistä jompi kumpi löytyy yleensä myös yliopistojen tarjonnasta. Saatavilla on myös useita avoimen lähdekoodin ohjelmistoja kuten Rapidminer ja R, joita voi käyttää vapaasti. Eri ohjelmistot muistuttavat melko paljon toisiaan, ja yleensä jos hallitsee yhden, muiden käytön oppii muutamassa päivässä. Olennaista on tietää mitä tekee, sillä jokainen järjestelmä on lopulta yhtä hyvä kuin ihminen joka sitä käyttää.

Analyysiohjelmistojen lisäksi jonkin tasoinen tiedonhallinnan ja tietokantojen ymmärrys on hyväksi. SQL tulee tarpeeseen monessa paikassa, mutta myös yleisempi käsitys tietokannoista ja näiden toiminnasta on hyödyksi. Yritysten IT-infrastruktuurit koostuvat usein melko kirjavasta joukosta erilaisia järjestelmiä, ja on tarpeellista omata vähintäänkin periaatteen tasolla ymmärrys siitä, miten data liikkuu, mihin sitä käytetään, ja mitä rajoitteita missäkin vaiheessa on.

Ohjelmointitaidot eivät myöskään mene koskaan hukkaan, joten ainakin jonkin ohjelmointikielen hallinta on eduksi. Valmiita ratkaisupaketteja ei ole aina saatavilla, joten usein on tarpeen turvautua omaan koodiin. Tällä hetkellä Java-osaaminen on todennäköisesti helpointa valjastaa käytäntöön, mutta luonnollisesti mikä tahansa muukin kieli antaa hyvän pohjan jolla päästä eteenpäin.

## Kauppätieteet

Kauppätieteet ja talouden opinnot ovat usein matemaatikoille myrkyä, mutta nämä ovat jo pelkän yleisivistyksenkin kannalta erittäin hyödyllisiä. Pelkkä numeroiden murskaaminen itsessään tuottaa harvoin mitään hyödyllistä lopputulosta, joten on olennaisen tärkeää ymmärtää, mistä toiminnasta data todellisuudessa kertyy. Käytännössä tämä tarkoittaa niin tietojärjestelmien kuin liiketoimintaprosessienkin tuntemusta. Lisäksi viimeistään budjetin päällä istuu yleensä kaupallisen alan ihmisiä, joilla ei välttämättä ole syvällistä käsitystä teknisistä yksityiskohdista. Tällöin on arvokasta osata keskustella heidän kielellään, ja ymmärtää millä silmillä he katsovat suurempia kokonaisuuksia.

## Harrastuneisuus

Oma harrastuneisuus ei ole sinällään pakollista, mutta se on erinomainen keino erottautua muista kilpaili-

joista työmarkkinoilla. Internetistä on saatavilla ilmaiseksikin valtavat määrät aiheeseen liittyvää materiaalia, kunhan vain osaa etsiä oikeista paikoista. Esimerkiksi LinkedInissä on lukuisia ryhmiä, joissa alan ammattilaiset keskustelevat aiheeseen liittyvistä asioista. Valitettavasti useimpien näiden taso on nykyään melko kyseenalainen, ja keskustelu pyörii melko triviaalien asioiden tai oman yrityksen mainostamisen ympärillä. Parempaa materiaalia on saatavilla esimerkiksi Datasciencecentralin ja Kdnuggetsin kaltaisista palveluista, joita ylläpitävät alan arvostetuimmat ammattilaiset.

Oman harrastuneisuuden voi laittaa myös tuottamaan, sillä erilaisia aiheeseen liittyviä kilpailuja pyörii jatkuvasti, ja näihin voi yleensä kuka tahansa osallistua. Näistä suurimpien palkintosummat nousevat kymmeneen ja jopa satoihin tuhansiin dollariin, ja näiden aihepiirit ovat mitä moninaisimpia, vaihdellen esimerkiksi luottoriskien ennustamisesta aivovammojen diagnosointiin. Moniin näistä osallistuu myös suurten IT-yritysten tutkimusosastoja, jotka testaavat näin uusia menetelmiä. Kilpailujen kautta on erinomainen mahdollisuus päästä pyörittämään valmiiksi ratkaistujen esimerkkien sijaan aitoa dataa, jollaista yrityksissä käsitellään päivittäin.

## Viestintä- ja esiintymistaidot

Viimeisenä, ja todennäköisesti myös tärkeimpänä, ovat viestintä- ja esiintymistaidot. Tämä osa-alue on usein matemaatikoille haastava, ja ainakin omana opiskeluaikanaan tämän kehittämiseen kannustettiin harmillisen vähän. Jo työnhaussa näillä taidoilla on ratkaiseva merkitys, lähtien siitä miten oman hakemuksen saa erottumaan muista kilpailijoista, ja lopulta rekrytointipäätökset tasavahvojen kandidaattien kesken ratkeavat usein haastattelumenestyksen perusteella. Tämä ei kuitenkaan tarkoita, että olisi tarpeellista väkisin yrittää olla supliikki myyntimies, vaan tärkeintä on osata ottaa rennosti ja olla oma itsensä.

Itse työnteko tapahtuu käytännössä aina jonkinlaisen tiimin puitteissa, jolloin myös vähintäänkin välttävät vuorovaikutustaidot ovat ehdoton edellytys. Tulosten ja johtopäätösten selkeä esittäminen on tarpeen jatkuvasti, ja viimeistään siirryttäessä ravintoketjussa ylöspäin esimiesasemiin tämä korostuu entisestään. Vastavasti lopputuloksena on usein jonkinlainen toimenpidesuositus, joka edellyttää lopulta, että joku ihminen tekee jotain. Tällöin on tärkeää osata perustella ja selittää, mitä tehdään ja miksi.

## Loppusanat

Tässä on esiteltynä oma subjektiivinen näkemykseni muutamista ominaisuuksista, joista on hyötyä data-analyysin maailmassa. Yksi ihminen harvoin hallitsee

täysin kaikkia näitä osa-alueita, joten roolista riippuen hieman erilaiset yhdistelmät näistä ovat kulloinkin optimaalisia. Teknisemmässä roolissa matemaattiset ja IT-taidot korostuvat, kun taas kaupallisella puolella viestintä- ja esiintymistaidot ja ymmärrys reaalimaailman ilmiöistä ovat tärkeämpiä. Tätä kuvaa hyvin erään kollegani toteamus: "We are not in the business of analyzing data, we are just trying to understand the daily

life of people".

Kehitys on joka tapauksessa nopeaa, ja nykyiset käytännöt ja menetelmät vanhentuvat hyvinkin nopeasti. Muuttuvassa maailmassa parhaiten menestyvät ne, jotka osaavat mukautua muutokseen ja soveltaa osaamistaan uusiin kohteisiin. Näin ollen tärkein taito on aina lopulta kyky jatkuvaan uuden oppimiseen ja itsensä kehittämiseen.