

Niko Ruotsalainen

# ASIAKASLÄHTÖINEN ANALYTIIKKA- TUEN ARVIOINTI

# TIIVISTELMÄ

Niko Ruotsalainen: Asiakaslähtöinen analytiikkatyön arviointi  
Pro gradu -tutkielma  
Tampereen yliopisto  
Tietojenkäsittelytieteiden maisterin tutkinto-ohjelma  
Toukokuu 2023

---

Analytiikka on tärkeä osa organisaatioiden päivittäistä arkea ja dataa pyritään hyödyntämään erilaisten analyysien ja analytiikkatyökalujen avulla. Analytiikan tärkeyden takia sen elinkaarta on pyritty hallitsemaan erilaisin elinkaarimallien avulla ja vaikka alalla on useita erilaisia malleja elinkaarien ymmärtämiseen, niin niiden käyttö on vaihtelevaa ja riippuu paljon kyseessä olevasta organisaatiosta. Elinkaarimallit keskittyvät myös pitkälti erilaisten analytiikkaratkaisujen kehittämiseen ja tuotantoon viennin jälkeistä tukemisen vaihetta kuvattu ja tutkittu kirjallisuudessa vähäisesti.

Tässä tutkimuksessa tutkittiin analytiikan tukemista analytiikan elinkaarimallin CRISP-DM:n (The Cross Industry Standard for Process for Data Mining) laajennettuun versioon pohjaten. CRISP-DM laajennettu versio sisältää viimeisessä vaiheessaan erilaisia analytiikan tukemiseen liittyviä tehtäviä, jotka toimivat tutkimuksen teoreettisena viitekehyksenä. Teorian pohjalta muodostettiin haastattelurunko, jonka avulla haastateltiin eri organisaatioiden analytiikan kokonaisvaltaisia asiantuntijoita ja kerättiin kokemuksia ja huomioita analytiikan elinkaaresta ja erityisesti sen tukemisesta. Haastattelujen lisäksi tutkimuksessa esitellään tarpeellisessa laajuudessa analytiikkaan, sen tukemiseen ja elinkaarimalleihin liittyvä teoria ja olemassa oleva tutkimus.

Haastatteluiden perusteella tunnistettiin erilaisia tukemisen malleja, tukemiseen liittyviä haasteita, analytiikkaratkaisujen ajankohtaisuuteen liittyviä huomioita ja haasteita. Haastatteluissa korostui se, että onnistuneen tuen toimittamiseksi analytiikkatiimien kommunikaatio liiketoiminnan kanssa oli tärkeää. Haasteiksi nostettiin analytiikan tekijöiden vaihtuvuus, analytiikkaratkaisujen suuri määrä, liiketoiminnan omistajuuden puute suhteessa analytiikkaratkaisuihin ja resurssien vähäisyys. Analytiikan paremmaksi tukemiseksi nostettiin esille liiketoiminnan syvempi sitoutuminen analytiikkaratkaisujen tukemiseen ja ylläpitämiseen. Analytiikan liiketoimintahyötyjen arviointi oli haastateltavien organisaatioissa alkeellista. Tulosten vertailu valittuun elinkaarimalliin (CRISP-DM) paljasti sen, että mallin tarjoamat tehtävät kohtasivat osittain haastateltavien todellisuuden kanssa, mutta osittain mallin tarjoama jäi ohueksi.

Avainsanat: analytiikka, analytiikan elinkaari, CRISP-DM, analytiikan tukeminen

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

# ABSTRACT

Niko Ruotsalainen: Customer centric review of analytics support  
Master of Sciences Thesis  
Tampere University  
Master's Degree programme in Computer Science  
May 2023

---

Analytics is an important part of organizations' daily routines, and data is sought to be utilized through various analyses and analytic tools. Due to the importance of analytics, its lifecycle has been managed with the help of various lifecycle models. Although there are several different models for understanding lifecycles in the field, their usage varies and depends heavily on the organization in question. Lifecycle models also focus largely on the development and support phase of the analytics has been described and studied to a limited extend in the literature.

This study investigated the support of analytics based on an extended version of the analytics lifecycle model CRISP-DM (The Cross Industry Standard for Process for Data Mining). The extended version of CRISP-DM includes various tasks related to supporting analytics in its final phase, which serve as the theoretical framework for the study. Based on the theory, an interview frame was formed, which was used to interview analytics experts from different organizations, collecting experiences and insights on the analytics lifecycle and specifically its support. In addition to the interviews, the study presents the necessary extent of theory and existing research related to analytics, its support, and lifecycle models.

Based on the interviews, various support models, challenges related to support, observations on the quality over time of analytics solutions, and challenges of relevance of analytics were identified. The interviews highlighted the importance of effective communication between analytics teams and the business for delivering successful support. Challenges included turnover of analytics personnel, a large number of analytics solutions, a lack of ownership of analytics solutions in relation to the business, and limited resources. To improve support for analytics, deeper commitment from the business in supporting and maintaining analytics solutions was emphasized. Assessment of the business benefits of analytics was rudimentary in the interviewed organizations. Comparing the results to the chosen lifecycle model (CRISP-DM) revealed that the tasks provided by the model partly aligned with the reality of the interviewees, but partly fell short.

Keywords: analytics, analytics life cycle, CRISP-DM, analytics support

The originality of this publication is checked using Turnitin OriginalityCheck –application.

## ALKUSANAT

Tämä pro gradu -tutkielma on viimeinen askel yliopistourallani ja vaikka tutkielman tekeminen vaikutti uskomattoman isolta projektilta maisterivaiheen viimeisiä kursseja suorittaessa niin kun työskentelyyn pääsi kiinni, se sujui ja oli jopa joskus mukavaa. Vaikka meidän tietojenkäsittelijöiden valmistuminen on epävarminta koko yliopistokentällä, niin kyllä meistä osa pääsee maaliinkin. Ennen kiitoksia haluan pahoitella kaikille läheisilleni, työkavereilleni ja muille kohtaamille ihmisilleni, että en ole viime aikoina puhunut muusta kuin gradusta ja sen tekemisestä. Lupaan puhua jatkossa kaikesta muusta.

Haluan kiittää ensinnäkin ohjaajaani Mikko Ruohosta, siitä, että on tarjonnut mutkattomat puitteet valmistua ja päästä kiinni gradun tekemiseen, vaikka aikataulu olikin tiukka. Haluan kiittää myös toimeksiantajayritystä ja esimiestäni siitä, että työlle on saatu mahdollistettua (työ)aikaa ja olen työn viimeisissä vaiheissa voinut panostaa gradun tekemiseen niin paljon kuin olen vain ehtinyt tehdä. Iso kiitos kuuluu myös haastateltaville, jotka lähtivät ennakkoluulottomasti mukaan graduni tekemiseen. Kiitos myös koulutusasiantuntija Helille siitä, että kaikki Sisun tarjoamat haasteet on saatu selätettyä ja valmistumisen aikataulut ovat olleet joustavia. Kiitos kaikille ystäville ja perheelleni, että olen saanut keskittyä rauhassa tutkielman tekemiseen ja kotona (ja muualla) on annettu anteeksi opinnäytetyön aiheuttama muiden asioiden laiminlyönti.

Viimeinen kiitos kuuluu kaikille järjestöille ja ihmisille, joihin olen tutustunut niissä toimiessani ja jotka ovat olleet osa opintojani ja yliopistouraa, kiitos siis Luuppi, G-kerho, Akateeminen Olutseura, Tamy ja TREY. Ilman teitä en olisi viihtynyt yliopistolla näin pitkään.

Tampereella, 31.05.2023

Niko Ruotsalainen

## **1 Johdanto 1**

- 1.1 Tutkimusongelma – ja kysymykset 1
- 1.2 Tutkimuksen rakenne 2

## **2 Tutkimusmenetelmät 4**

- 2.1 Tutkimusprosessi 4
- 2.2 Tutkimusongelma 4
- 2.3 Aikaisempi kirjallisuus 5
- 2.4 Ongelman täsmennys 6
- 2.5 Tutkimusasetelma 7
  - 2.5.1 Laadullinen tutkimus 8
  - 2.5.2 Asiantuntijahaastattelu 8
- 2.6 Aineiston keruu ja analysointi 9
- 2.7 Johtopäätökset 12

## **3 Data ja analytiikka 14**

- 3.1 Data 14
- 3.2 Analytiikka 15
- 3.3 Visuaalinen analytiikka 18

## **4 Analytiikan elinkaari 21**

- 4.1 Elinkaaren mallit 21
- 4.2 CRISP-DM 22
- 4.3 Muut mallit 25
- 4.4 Analytiikkatuki 28

## **5 Asiakslähtöisyys 31**

## **6 Tutkimuksen tulokset 33**

- 6.1 Taustatiedot ja analytiikan sijoittuminen 33
- 6.2 Elinkaari 34
- 6.3 Tukeminen 36
- 6.4 Tukemisen haasteet 38
- 6.5 Ajankohtaisuus 41
- 6.6 Ajankohtaisuuden haasteet 42

6.7 Liiketoimintahyötyjen arviointi 43

6.8 Muut huomiot 44

**7 Pohdinta 45**

7.1 Tulokset suhteessa teoreettiseen viitekehukseen 45

7.2 Vastaukset tutkimuskysymyksiin ja huomiot analytiikkatuen toimittamiseen  
47

7.3 Tutkimuksen luotettavuus 48

7.4 Tulevaisuuden tutkimus 50

**8 Päätelmät 51**

**9 Viiteluettelo 53**

**10 Liite 1: Pohjustus tutkimuksesta 57**

**11 Liite 2: Asiantuntija-haastattelun runko 58**

## 1 Johdanto

Analytiikka on tärkeä osa organisaatioiden päivittäistä arkea enenevissä määrin ja dataa pyritään hyödyntämään erilaisten analyysien ja analytiikkatyökalujen avulla. Analytiikan elinkaarta on pyritty hallitsemaan erilaisin elinkaarimallien avulla, mutta vaikka alalla onkin vakiintuneita malleja elinkaaren ymmärtämiseen, niin usein analytiikkatiimit eivät käytä niitä (Aho et al., 2021). Erityisen ohutta kirjallisuus on analytiikan elinkaaren loppupäästä, jossa pyritään ymmärtämään, kuinka analytiikkaa tuetaan ja käytetään sen tuotantoon viennin jälkeen. Olemassa oleva kirjallisuus käsittelee lähinnä analytiikan hyväksyntää ja adoptioastetta (Bischoff et al., 2015).

Analytiikka on arvokasta organisaatiolle, mutta analytiikkaratkaisun tuotantoon viennin jälkeen sen tukeminen ei ole järjestelmällisesti määritelty. Mikäli tukea ei hoideta järjestelmällisesti, analytiikka kärsii usein huonosta laadusta, sen ajankohtaisuus heikkenee ja luottamus sen tarjoamaan informaatioon laskee. Tämän takia on tärkeää, että analytiikan tukeminen on hyvin määritelty ja on selkeää, kuinka analytiikka pidetään toimintakuntoisena myös sen kehittämisen jälkeen. Aikaisempien tutkimusten mukaan analytiikka on edelleen usein heikommin määriteltyä esimerkiksi ohjelmistokehitykseen verrattuna. Tiimien ja prosessien merkitystä analytiikan elinkaarelle on tutkittu säästeliäästi (Aho et al., 2021).

Tässä tutkimuksessa keskitytään elinkaaren loppuvaiheeseen, jossa analytiikkaa tuetaan sen tuotantoon viennin jälkeen. Tutkielman tarkoituksena on tuottaa tietoa siitä, miten analytiikkaa tuetaan tällä hetkellä ja miten tukiorganisaatiot ovat järjestyneet, mitä tuelta kaivataan ja mitä haasteita siihen liittyy. Lisäksi tarkastellaan sitä, miten analytiikan tukeminen vertautuu olemassa oleviin teoreettisiin elinkaarimalleihin ja onko käytännön ja teorian välillä eroja.

### 1.1 Tutkimusongelma – ja kysymykset

Tämän tutkielman tavoitteena on ymmärtää asiakaslähtöisestä näkökulmasta sitä, miten analytiikkaa tuetaan sen varsinaisen tuotantoon viennin jälkeen, mitä haasteita analytiikan tukemiseen liittyy ja miten analytiikkaa voitaisiin tukea paremmin. Tutkimusongelmaa selvitetään toimeksiantajayrityksen kontekstissa, sen asiakkaiden kokonaisvaltaisesti analytiikkaa ymmärättäviä ammattilaisia haastatellen. Haastateltavat on valittu siten, että heillä olisi kokonaisvaltaista ymmärrystä analytiikasta omassa organisaatiossaan. Ongelmaa käsitellään analytiikan elinkaarimalli CRISP-DM:n (Chapman et al., 2000) laajenutuksessa viitekehityksessä (Cochran, 2019).

Tutkimuksen tavoitteena on toimittaa toimeksiantajayritykselle huomioita analytiikkatuesta, joita voitaisiin myöhemmin käyttää tuen tarpeen parempana ymmärtämisenä ja mahdollisena pohjana tarjoaman kehittämisessä. Tarkoituksena ei ole kuitenkaan kehittää valmista mallia analytiikkatukeen.

Edellisten pohjalta tutkimuskysymykset ovat seuraavat:

1. Miten analytiikkaa tuetaan?
2. Mitä haasteita analytiikan tukemiseen liittyy?
3. Miten analytiikkaa voitaisiin tukea paremmin?

Näiden näkökulmien lisäksi on tarpeen verrata tuloksia laajennetussa CRISP-DM:n elinkaarimallin (Chapman et al., 2000) viimeisessä vaiheessa, joka liittyy analytiikan tukemiseen. Mahdollisia muita huomioita haastatteluista, jotka liittyvät analytiikan tukemiseen voidaan myös nostaa esille.

Tutkimusongelman selvittämisen rajoitteina ovat ne, että empiirisen osion haastattelukoko on suhteellisen pieni, ja että se on suunnattu suurehkoille asiakasorganisaatioille, joten voi olla, että huomioita ei voi soveltaa pienemmille asiakkaille. Lisäksi on mahdollista, että teoreettinen viitekehys ei ota kaikkia näkökulmia huomioon nopeasti kehittyvällä alalla, valittu viitekehys on vuodelta 2000 (Chapman et al.). Viitekehysten laajennettu versio on tosin päivitetty vuonna 2019 (Cochran).

## **1.2 Tutkimuksen rakenne**

Tutkielma koostuu useammasta luvusta, joista ensimmäiset käsittelevät tutkimuksen teoreettista osiota, tutkimusmetodeja ja teoreettista viitekehystä tutkimukselle. Teorian jälkeen tutkielma käsittelee tutkimuksen tuloksia, pohdintaa ja viimeisessä luvussa tutkielma päätetään.

Luvussa kaksi esitellään tutkimusmenetelmät päämenetelmänä ollen asiantuntijahaastattelu ja sen analysointi. Luvussa kolme esitellään tarpeellisella laajuudella data, analytiikka ja visuaalinen analytiikka analytiikan ollessa haastattelu tutkimuksen keskiössä. Luvussa neljä esitellään analytiikan elinkaari ja sen erilaiset mallit. Malleista tarkemmin esitellään CRISP-DM (Chapman et al., 2000) ja sen viimein osa-alue tarkemmin analytiikkatuen alla. Luvussa viisi esitellään asiakaslähtöisyyden konsepti lyhyesti, haastattelujen kohdistuessa toimeksiantajayrityksen asiakkaisiin. Luvussa kuusi käydään läpi tutkimuksen tulokset, esitellen haastattelujen tulokset. Luvussa seitsemän keskustellaan haastattelujen



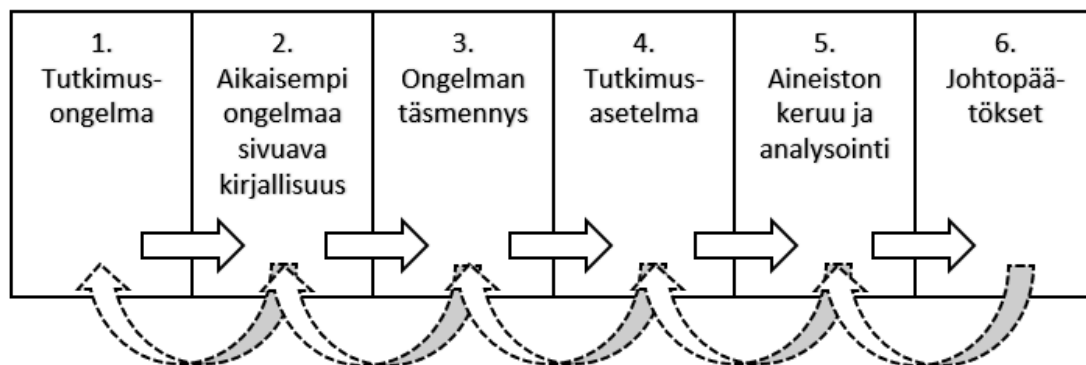
ja teoreettisen viitekehyksen suhteesta ja nostetaan esille niistä löytyviä huomioita. Luku kahdeksan päättää tutkimuksen ja kertaa tiiviisti tulokset.

## 2 Tutkimusmenetelmät

Tässä luvussa käydään läpi opinnäytetyössä käytetyt tutkimusmenetelmät. Tutkimuksen tarkoituksena on tarkastella analytiikkatukea, sen haasteita ja tuen parannusehdotuksia. Tutkimus on toteutettu laadullisena tutkimuksena Tiaisen (2014) tutkimusprosessia mukailten. Aineisto on kerätty asiantuntijahaastattelujen avulla käyttäen puolistrukturoitua haastattelumenetelmää. Tutkimukseen on käytetty teoriapohjaista aineiston keruuta, sillä haastattelujen kysymykset on muodostettu CRISP-DM:n (Chapman et al., 2000) laajennetun version (Cochran, 2019) pohjalta. Tutkimus on toteutettu toimeksiantajayritykselle, johon opinnäytetyöntekijä on työsuhteessa. Tutkimusta on ollut ohjaamassa yliopiston ohjaaja sekä toimeksiantajayrityksen edustaja.

### 2.1 Tutkimusprosessi

Tutkimusprosessi mukailee Tiaisen (2014) esittämää tutkimusprosessia, jossa on kuusi peräkkäistä vaihetta. Vaiheiden rajat eivät ole ehdottomia, vaan ne ovat osittain päällekkäisiä ja iteratiivisia kuten alla tutkimusprosessia kuvaavassa kuvassa on nähtävissä. Tämän luvun seuraavissa osissa käydään koko prosessi vaihe vaiheelta läpi, jotta lukijalla on mahdollisuus ymmärtää koko prosessi ja miksi erilaisiin valintoihin tutkimuksen kannalta on päädytty.



Kuva 1. Tutkimusprosessi mukailtuna Tiaisen (2014) esittämän taulukoinnin pohjalta.

### 2.2 Tutkimusongelma

Tutkimusongelman ideointi lähti liikkeelle tutkijan toimeksiantajayrityksen esimiehen ja silloisen yksikön johtajan kanssa, jossa todettiin, että analytiikan tukeminen varsinaisen projektin toimittamisen jälkeen ei ole niin suoraviivaista, kuten esim. IT-infran tuki sen pystyttämisen jälkeen ja tukeen ei ollut suoraa mallia. Tutkijalle annettiin tehtäväksi miet-

tiä aihetta enemmän ja tutustua pintapuolisesti alan tieteelliseen tai ammatilliseen materiaaliin. Aiheen pohtimisen jälkeen mukaan päätettiin tuoda asiakaskeinen näkökulma, sillä toimeksiantajan näkökulmasta asiakkaiden huomioidut aiheesta olisivat liiketoiminnallisesti mielenkiintoisia. Alustavaksi tutkimuskysymykseksi muotoutui vuoden 2022 keväällä otsikko ”Mitä asiakkaat odottavat analytiikkatueltä?” Aihe vahvistettiin toimeksiantajan toimesta.

Toimeksiantajan vahvistamisen jälkeen tutkimuskysymys esiteltiin yliopiston ohjaajalle sekä kirjallisesti että tapaamisen muodossa, jonka seurauksena työn otsikko rakenteellistettiin muotoon ”Analytiikkatueltä asiakaslähtöinen arviointi”. Tässä vaiheessa tutkimusongelmaksi käsiteltiin se, mitä analytiikkatueltä halutaan asiakkaan näkökulmasta ja onko sen toimittamiseen haluttuja malleja.

### **2.3 Aikaisempi kirjallisuus**

Tutkimuksen teoriaosuudessa on tarkoituksena ensin esitellä peruskäsitteistö liittyen analytiikkaan ja sitten esitellä analytiikan elinkaarta, sen malleja ja viimeisenä pureutua tarkemmin siihen, mitä analytiikkatueltä on. Tämän lisäksi asiakaslähtöisyyden konsepti esitellään lyhyesti. Kokonaisuudessaan aikaisemman kirjallisuuden ja tutkimuksen avulla luodaan tarpeeksi kattava teoreettinen kokonaiskuva (Tuomi & Sarajärvi, 2009).

Vaikka tutkija on kirjallisuuskatsauksen alkamisen vaiheessa työskennellyt analytiikan parissa noin vuoden verran ja IT-alalla useita vuosia, niin analytiikan ammatillinen tai tieteellinen kirjallisuus ei ollut erityisen tuttua, joten kirjallisuuden tutkiminen lähti liikelle sopivien hakusanojen löytämisellä erilaisten tietokantojen käyttöön. Tietokannat, jotka olivat eniten käytössä, olivat Tampereen yliopiston oma Andor, ProQuest, Trepo, IEE Xplore ja Science Direct. Kirjallisuuden hakemiseen käytettiin pääosin seuraavia hakusanoja:

- (Data) analytics support
- Evaluating analytics support
- Analytics support phase
- Support in Business Intelligence OR analytics
- Post-deployment AND analytics
- Analytics life cycle
- CRISP-DM deployment
- Analytics life cycle models
- Continuous use

Edellä mainittujen hakusanojen lisäksi hyödylliseksi yleislähteeksi nousi hakusanalla ”data” tehty haku Trepossa, josta löytyi aihetta sivuavia opinnäytetöitä ja artikkeleita, jotka omalta osaltaan johdattivat sopivien lähteiden äärelle.

Sopivan kirjallisuuden etsimistä jatkettiin pitkään ja tutkimusprosessin (Tiainen, 2014) iteratiivisen luonteen mukaisesti siihen palattiin myöhemmistä vaiheista. Ongelman täsmennyksen jälkeen, erityisesti erilaisten elinkaarimallien haku ja niiden tutkiminen vaativat lisää huomiota tutkijalta. Lisätäsmennystä vaati myös se, haettiinko termeillä ”visual analytics”, ”data analytics”, ”data science”, ”business intelligence” vai ”analytics”, sillä termejä käytetään arkikielessä ja ammattikielessä korvaamaan toisiaan (Davenport, 2012), jonka seurauksena tuloksia tuli isoja määriä ja läheskään kaikki eivät olleet relevantteja. Vaatikin paljon työtä, että tutkija sai määriteltyä, onko esimerkiksi analytiikalla ja data-analytiikalla mitään eroa (Vastaus: ei ole).

Haasteellista oli myös tasapainoilla ammatillisen kirjallisuuden ja tieteellisten julkaisujen välillä, sillä usein niiden tekijät olivat samoja ja ammatillinen kirjallisuus kattavampaa ja sitä löytyi määrällisesti enemmän. Katsauksessa pyrittiin kuitenkin minimoimaan erilaisten kaupallisten toimijoiden kuten Tableau tai Microsoft materiaalin määrä omista työkaluistaan, ja pysyä enemmän yleisellä tasolla, kuin tuottaa analytiikkatyökaluspesifiä tekstiä. Kaupallisista toimijoista ei voi kuitenkaan pysyä ”erossa”, sillä alalla yleisesti hyväksytyt elinkaarimallit ovat lähes poikkeuksetta kaupallisten toimijoiden tuottamia, esimerkiksi SAS Institutun muodostama SEMMA (2023) ja usean eri yrityksen yhteistyössä muodostama CRISP-DM (Chapman et al., 2000).

Löydetyn kirjallisuuden pohjalta pyrittiin luomaan riittävä, mutta kattava teoreettinen tausta työlle ja sen empiiriselle osiolle. Tutkielman haastattelut perustuvatkin löydettyyn kirjallisuuteen ja useasta lähteestä vahvistettuun elinkaarimalliin (Chapman et al., 2000).

## **2.4 Ongelman täsmennys**

Ongelman täsmentämiseksi ja tutkimuskysymysten tarkasti määrittämiseksi löydetystä kirjallisuudesta valittiin analytiikan elinkaarimalli CRISP-DM (Chapman et al., 2000), joka on elinkaarimalleista suosituin (Aho et al., 2021) ja lähinnä alalla olevaa standardia. Mallista löytyi myös laajennettu versio (Cochran, 2019), jossa mallin eri kohtia on eritelty vielä tarkemmin. Ylipäätään CRISP-DM oli malleista eniten tutkittu, määritelty ja keskusteltu, niin ammatillisen kuin tieteellisenkin kirjallisuuden puolella. Elinkaarimallin viimeisenä vaiheena on vaihe ”Deployment”, joka käsittää analytiikkaratkaisun tuotantoon viennin ja siihen myös liittyvän ratkaisun käytön tukemisen. Cochranin (2019) laajennetun mallin viimeinen vaihe valittiin teoreettiseksi viitekehyykseksi tutkimusongelmalle.

Viitekehystä käytettiin myöhemmin erityisesti asiantuntijahaastattelurungon pohjana. Tämän viitekehyksen avulla muodostettiin kolme tutkimuskysymystä:

- Miten analytiikkaa tuetaan?
- Mitä haasteita analytiikan tukemiseen liittyy?
- Miten analytiikkaa voitaisiin tukea paremmin?

Ongelman ja tutkimuskysymysten muotoon tehtiin aineistonkeruuta valmistellessa vielä pieniä tarkennuksia ja muutoksia, mutta ei merkittävässä määrin. Ongelman määrittelyyn palaaminen vahvistaa Tiaisén (2014) prosessimallin iteratiivista luonnetta.

Ongelman määrittelyn aikana muodostettiin myös työhypoteesi, jossa tutkija pyrki ennakkoimaan empiirisen perusteella nousevia huomioita. Työhypoteesi sisälsi seuraavat asiat:

- Analytiikan tukemisessa on haasteita
- Liiketoimintaymmärrys on keskiössä
- Analytiikan ajankohtaisuuteen reagoidaan reaktiivisesti
- Tuen tarve keskittyy loppukäyttäjien ongelmiin työkalujen kanssa

Hypoteesin sisältö muodostettiin tutkin kirjallisuuden ja tutkijan ammatillisen kokemuksen perusteella. Tutkimuksen empiirisen osion tekemisen aikaan tutkija oli ollut työsuhteessa toimeksiantajaan viiden ja puolen vuoden ajan, joista puolentoista vuoden ajan analytiikan parissa.

## **2.5 Tutkimusasetelma**

Tutkimusasetelma muodostettiin jo melko aikaisessa vaiheessa tutkimusta, valitsemalla toimeksiantajan asiakkaiden edustajat tutkimuksen lähteeksi, jolloin tutkimusongelman ratkaisun askeleet eivät tutkimuksen myöhäisemmässä vaiheessa muuttuneet juurikaan. Menetelmäksi valittiin laadullinen tutkimus ja tarkemmin asiantuntijahaastattelu. Menetelmät valittiin, sillä toimeksiantajan ja yliopiston ohjaajan kanssa koettiin mielekkäämmäksi ratkaisuksi, se että asiantuntijoiden haastatteluiden avulla saataisiin syvempi ymmärrys tutkimuskysymyksestä verrattuna esimerkiksi kyselylomakkeeseen tai määrälliseen tutkimusotteeseen. Lähempänä varsinaisten haastattelujen toteutumista tarkennettiin toimeksiantajan toimesta se, että valitut haastateltavat ovat tarpeeksi isoista asiakkuuksista, jotta tutkimuksen tuloksena syntyvät huomiot olisivat mielekkäitä. Pienempien asiakkaiden parissa koettiin ongelmaksi töiden vähyys ja se, että asiakkailla ei välttämättä ollut nimettyä henkilöä tekemässä tai hallitsemassa analytiikkaa. Asiantuntijoiden haas-

tattelujen avulla ja edeltä määritellyn kysymysrunгон avulla tavoiteltiin monialaista ymmärrystä annetun teoreettisen viitekehyksen ympärille. Seuraavissa alakappaleissa 2.5.1 ja 2.5.2 esitellään laadullisen tutkimuksen ja asiantuntijahaastattelun käsitteet.

### 2.5.1 Laadullinen tutkimus

Laadullinen tutkimus eli kvalitatiivinen tutkimus on kattokäsite useita eri tutkimusperinteitä yhdistävänä kattokäsitteenä ja käsitteen alle mahtuu useita eri metodeja. Laadullisen tutkimuksen tavoitteena on saada ymmärrystä tutkimuksen kohteiden näkökulmista tutkittavaan ilmiöön (Tuomi & Sarajärvi, 2009). Tuomi ja Sarajärvi (2009) viittaavat siihen, että laadullinen tutkimus viittaa laatuun, ei määrään, mutta Saaranen-Kauppinen ja Puusniekka (2006) muistuttavat, että termi voi olla myös harhaanjohtava ja antaa tulkitsijalle väärän kuvan ”paremmasta” tutkimuksesta verrattuna määrälliseen, eli kvantitatiiviseen tutkimukseen. Määrällinen tutkimus rakentuu usein systemaattisen analyysin päälle, jossa ollaan kiinnostuneita syy- ja seurasuhteista, luokittelusta ja sitä voidaan hallita usein valmiilla matemaattisilla malleilla (Hakala, 2008). Määrällisen tutkimuksen suorittaminen vaatii sen, että kerätty aineisto saatetaan taulukkomuotoon, josta eri tutkimuskohteille pyritään löytämään tilastollisia yhteyksiä.

Laadullisen tutkimuksen tavoitteena ei ole löytää tilastollisia todennäköisyyksiä ja niitä ei ole useinkaan tutkittavien kohteiden määrän vuoksi edes mahdollista tehdä, esimerkiksi haastattelun litteroinnin luoma tekstimäärä, joka estää käytännössä tilastollisesti merkittävän määrän haastattelun (Alasuutari, 2011). Laadullisen tutkimuksen analyysi ja kerääminen liittyvät usein toisiinsa, ja prosessi elää iteratiivisesti kuten Tiainen (2014) esittää. On myös mahdollista, että laadullisen tutkimuksen teoreettista viitekehystä ei kirjoiteta valmiiksi etukäteen kuten määrällisessä tutkimuksessa ja tutkimustehtävätkin voivat elää tutkimuksenkin aikana. Vaikka menetelmiä esitetäänkin usein vastakkaisina, niin niitä voidaan käyttää ja soveltaa samassa tutkimuksessa (Alasuutari, 2011).

### 2.5.2 Asiantuntijahaastattelu

Asiantuntijahaastattelu on asiantuntijan haastattelua, ja se lukeutuu aineistonkeruumenetelmän laadullisen tutkimuksen piiriin. Asiantuntija on henkilö, jolla on sellaista tietoa, jota maallikolla ei ole, tai on eriytynyttä ammattitaitoa pitkälle erikoistuneen ammattitaidon perusteella ja se usein määrittyy tapauskohtaisesti. Asiantuntijuus liittyy myös asiantuntijan kokemukseen, jonkin tietyn asian, instituution, ammatillisen tehtävän tai ilmiön kautta. Tutkimuksen kannalta ei ole yksiselitteistä, kuinka asiantuntija valitaan vaan se pitää tehdä tutkimuskohtaisesti. Asiantuntijoita haastatellaan usein puolistrukturoidun haastattelun tai sen muunnelmien avulla tai ryhmähaastattelulla.

Asiantuntijan saaminen haastatteluun ei ole aina yksinkertaista ja siihen saattaa liittyä hankaluuksia löytää haastateltava, tai saada haastateltavan taustaorganisaatiolta lupa haastatteluun. Voi olla kuitenkin, että vaikka kirjallisuudessa korostetaan haastateltavan saamisen vaikeutta, haastateltavat voivat olla avuliaita ja suostua mielellään haastatteluun. Onkin tärkeää tunnistaa tutkittavan ilmiön kannalta sopiva haastateltava.

Haastattelujen edetessä on tärkeää pitää mielessä, että mitä tutkitaan, ja asiantuntijoita haastatellessa on yleensä tehty paljon pohjatöitä, jotta tutkimuksen aiheena oleva ilmiö on jäsennetty hyvin tutkijan mielessä. Haastattelija voikin haastattelun edetessä tarvittaessa muokata kysymysrunkoaan parempien tulosten saamiseksi, samoin kuin haastateltava muokkaa puhettaan sen mukaan, kuinka hyvin hän kokee tutkijan ymmärtävän kyseessä olevan aihealueen. Asiantuntijoiden välistä vuorovaikutusta värittää ammattikieli ja vastavuoroisuus.

(Hyvärinen et al., 2017)

## **2.6 Aineiston keruu ja analysointi**

Aineiston kerääminen toteutettiin puolistrukturoidun haastattelumenetelmän avulla, jonka laadinnassa oli otettu huomioon asiantuntijahaastattelun konteksti (Hyvärinen et al., 2017) ja haastattelurungon (katso *liite 2*) muodostuksessa oli käytetty apuna CRISP-DM-viitekehystä (Chapman et al., 2000) laajennetussa muodossaan (Cochran, 2019). Haastatteluruko käytiin läpi yhdessä ensin toimeksiantajan edustajan kanssa, sitten yliopiston ohjaajan ja viimeisenä vielä kerran toimeksiantajan edustajan kanssa. Muutokset, jotka runkoon tehtiin toimeksiantajan ehdotuksesta oli Tiainenkin (2014) mainitsema ammattikielen käyttö tieteellisen kielen sijasta (esimerkiksi analytiikkamalli muotoiltiin muotoon analytiikkaratkaisu), yliopiston ohjaajan toimesta runkoon lisättiin kysymys miten analytiikan tukeminen on muuttunut aikojen saatossa.

Haastattelun voidaan katsoa olleen teorialähtöinen aineiston keruumenetelmä, sillä haastattelurungon pohjana oli CRISP-DM laajennettu malli (Cochran, 2019). Teorialähtöinen aineistonkeruu tarkoittaa sitä, että tutkijalla on valmis jokin teoriaan pohjautuva näkökulma tai hypoteesi (Tiainen, 2014), joka tässä tapauksessa oli CRISP-DM:n laajennettu malli.

Haastateltavat valittiin toimeksiantajien isompien asiakkaiden keskuudesta ja kukin haastateltavista oli useamman tuhannen ihmisen yrityksistä, joilla on oma datatiimi. Haastateltaviin otettiin joko suoraan sähköpostilla tai yhteyshenkilön kautta yhteyttä lyhyen saatetekstin kera, joka sisälsi haastattelupohjustuksen sisällön (*liite 1*) ja aikataulullisia ehdotuksia sekä ilmaisun haastateltavan asiantuntijuudesta ja tärkeydestä (Tiainen, 2014) ja

kiitokset mahdollisesta osallistumisesta. Kun haastattelut saatiin sovittua, heille lähetettiin pohjustus haastatteluun (*liite 1*). Pohjustuksessa kerrottiin lyhyesti haastattelun aihe, kesto, kieli, mihin tuloksia käytetään, miten haastattelut toteutetaan, miten haastattelun nauhoituksia käsitellään ja se, että tulokset käsitellään anonymisti. Haastateltavia valittiin kolme aikarajoitteellisista syistä ja siitä syistä, että toimeksiantajan datayksiköllä vastaavia asiakkuuksia oli rajallinen määrä.

<b>Haastateltava</b>	<b>Karkea tehtävä</b>	<b>Työkokemus analytiikan parissa (vuosina)</b>	<b>Organisaation karkea toimiala</b>
H1	BI erikoisasiantuntija	10	B2B
H2	Datatiimin johtaja	16	Prosessiteollisuus
H3	Vastaa data & analytics, MDM ja integraatioiden johtamisesta	7	Rakennusala

Taulukko 1. Haastateltavien tarkemmat kuvaukset

Haastattelut toteutettiin etänä Microsoftin Teamsin välityksellä kameroiden ollessa päällä jokaisessa haastattelussa aikavälillä 21.04. – 03.05.2023. Haastattelun tavoitteelliseksi kestoksi oli valittu tunti Tiaisen (2014) suosituksen mukaan. Haastattelut kestivät 52 minuutista yhteen tuntiin ja 18 minuuttiin. Ennen haastatteluja nauhoittamisen aloittamista haastateltavat tervehdettiin ja pyrittiin vähentämään jännitystä puhumalla aiheen vierestä, lisäksi ennen nauhoittamisen aloittamista haastateltavalle kerrattiin vielä seuraavat asiat:

- Haastattelun kesto noin yksi tunti
- Haastattelu tallennetaan
- Haastattelun tuloksia käytetään osana pro gradu tutkielmaa
- Tallenteet ja litteroinnit poistetaan gradun palauttamisen jälkeen
- Tulokset anonymisoidaan, jotta haastateltavan tai organisaation tunnistaminen ei ole mahdollista
- Kukaan mu ei käsittele haastattelumateriaaleja, paitsi tutkija

Kerrattujen asioiden jälkeen haastateltavalta kysyttiin vielä sanallisesti lupa aloittaa tallennus ja varsinainen haastattelu alkoi.

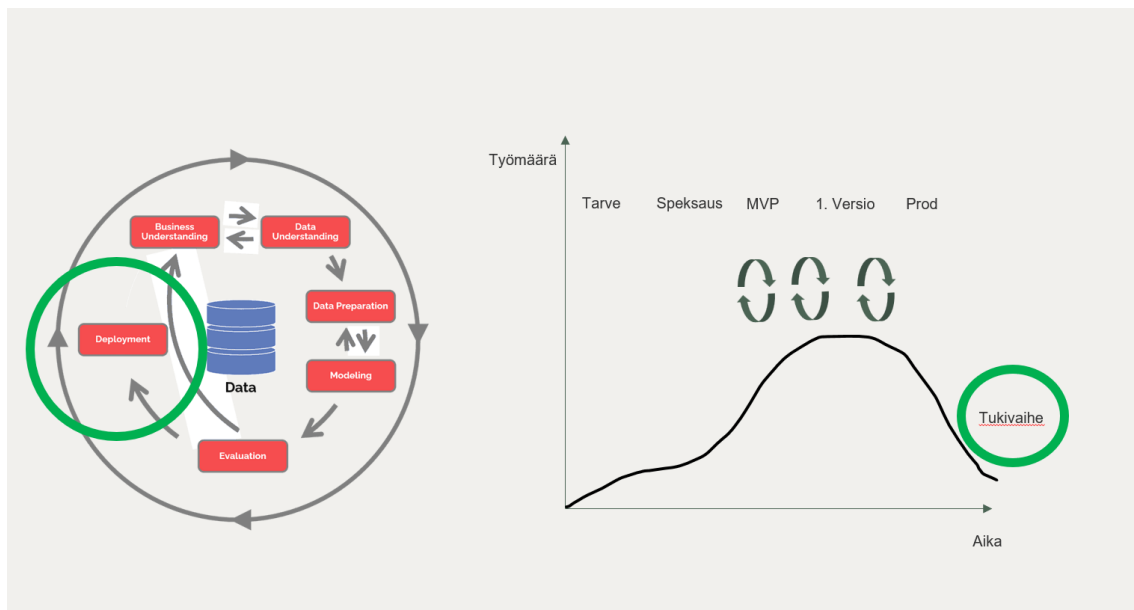
Haastattelun aluksi haastatteli esitteli analytiikan elinkaarimallin CRISP-DM:n (Chapman et al., 2000) ja toimeksiantajan edustajan kuvauksen analytiikkaprojektin elinkaaresta oheisen (kuva 2) Power Point-kuvan avulla. Kuvan tarkoituksena oli orientoida haastateltava siihen, että haastattelun pääpaino on analytiikkatuen piirissä ja esitellä



CRISP-DM:n malli haastateltavalle. Tämän lisäksi haastateltavalle näytettiin toimeksiantajan karkea yleistyks siitä, miten dataprojektin ympäristö rakentuu, kuva sisälsi järjestyksessä seuraavat kohdat, joissa data liikkuu:

1. Lähdejärjestelmät
2. Data-putki lähdejärjestelmästä tietovarastoon (Data Engineerin vastuualue)
3. Tietovarasto (Data Scientistin tai Analytics Engineerin alue datan käsittelyyn)
4. Visualisointityökalut (Analytics Consultantin vastuualue)
5. Loppukäyttäjät

Karkean yleistyksen tarkoituksena oli varmistaa termien yhteneväisyys haastateltavan kanssa, ennen haastattelun varsinaista alkamista. Kaikki haastateltavat kertoivat ymmärtävänsä mistä on kysymys ja olevansa samaa mieltä dataprojektin karkeasta yleistyksestä.



Kuva 2. Haastateltavan orientointi haastattelun aiheeseen CRISP-DM:n kuvan (Hotz, 2023) ja toimeksiantajaorganisaation vapaan analytiikkaprojektin elinkaaren hahmottelun avulla.

Haastattelun alkukysymyksiä pyrittiin selvittämään haastateltavien oleellisia taustatekijöitä kuten kokemusta analytiikan parissa, datatiimin kokoa ja sen sijoittumista organisaatiossa ja organisaation kokoa erilaisin mittarein. Taustakysymysten tehtävänä oli taustoittaa tutkimusta, mutta myös antaa haastateltaville ”helppoja” kysymyksiä vastattavaksi alkuun jään rikkomiseksi (Tiainen, 2014). Tämän jälkeen haastattelussa edettiin pääasiassa haastattelurungon mukaisesti.

Haastattelujen nauhoitteiden perusteella tallenteista muodostettiin litteroidut versiot, litterointien apuna käytettiin Microsoft Teamsin litterointiominaisuutta. Teamsin tekemään

litterointi oli yllättävän luotettavaa ja koneen tuottamaan litterointiin, piti tehdä yllättävän vähän muutoksia. Litteroinnin myötä tekstistä poistettiin mahdolliset yrityksiin ja henki-  
löihin viittaavat termit ja ne korvattiin anonyymeillä ilmaisuilla.

Laadullisessa tutkimuksen haastattelumenetelmässä usein tuloksena luokittelu, joka luokitellaan yhden tai kahden dimension mukaan. Hyvä luokittelu on ytimekäs, vahvaraken-  
teinen, kattava, laajennettavissa oleva ja kuvaava (Tiainen, 2014). Sisällönanalyysi voi-  
daan muodostaa deduktiivisesti, eli teorialähteisesti tai induktiivisesti eli aineistolähtöi-  
sesti. Teorialähtöisesti ohjattua analyysia ohjaa aikaisempi viitekehys. Teorialähtöinen  
analyysimenetelmä voi olla löyhä, tarjoten viitekehysten, mutta viitekehysten sisällä  
voidaan luoda luokituksia myös aineistolähtöisesti. Tämä lähestymistapa mahdollistaa  
sen, että aineistosta voidaan ottaa esille asiat, jotka kuuluvat viitekehukseen, mutta myös  
ne asiat, jäävät sen ulkopuolelle. (Tuomi & Sarajärvi, 2009)

Tässä tutkielmassa sisällönanalyysi tehdään teoriaperusteisen aineistonkeruumenetelmän  
pohjalta, joka tarkoittaa sitä, että vastauksia käsitellään haastattelurungon kysymysten  
pohjalta, jotka on muodostettu taas laajennetun CRISP-DM:n mallin pohjalta (Cochran,  
2019). Kysymysten sisällä taas muodostetaan alaluokkia aineistoperustaisesti, joka mah-  
dollistaa sen, että teoriakehysten ulkopuolelta tulevat huomiot on mahdollista nostaa mu-  
kaan tutkimuksen tuloksiin. Mikäli teoriaperusteisuuteen suhtauduttaisiin täysin ehdotto-  
masti, niin olisi vaarana, että osa tutkimuksen huomioista jäisi sen ulkopuolelle. Yhdistä-  
mällä teoria- ja aineistopohjaista menetelmää tutkimuksesta saadaan monipuolisempi.

Luokittelussa käytetään Tuomen ja Sarajärven Tuomi ja Sarajärvi (2009) esittämää pro-  
sessia, jossa haastattelut kuunnellaan ja litteroidaan, sisältöön perehdytään, etsitään pel-  
kistetyt ilmaukset, etsitään samankaltaisuuksia ja erilaisuuksia ilmauksista, luokitellaan  
ne ja tarvittaessa yhdistetään luokkia. Aineiston pieni määrä vaikeuttaa luokkien muo-  
dostamista, mutta se ei ole ongelma, sillä aineistopohjainen menetelmä tukee tässä teo-  
riaperusteista menetelmää eikä ole tutkimuksen ”pääroolissa”.

## **2.7 Johtopäätökset**

Johtopäätökset tutkimuksessa muodostetaan tulkitsemalla niitä suhteessa teoriaan, aiem-  
paan tutkimukseen ja valittuun viitekehukseen (Tiainen, 2014). Tutkimuksen tulosten tul-  
kinnassa otetaan huomioon tutkijan mahdollisesti tekemät virheet esimerkiksi haastatte-  
lujen aikana, jotka ovat mahdollisesti vaikuttaneet tuloksiin. Tutkimuksen tuloksia voi-  
daan arvioida myös sitä, että mikä niiden merkitys on tutkimuksen toimeksiantajalle, ja  
onko niillä merkitystä käytännön työhön. Johtopäätöksiä tehdessä on myös otettava huo-  
mioon tutkimuksen luotettavuus, eettisyys ja käytetyn kirjallisuuden arviointi. (Tuomi &  
Sarajärvi, 2009)

Tässä tutkimuksessa johtopäätöksissä arvioidaan miten tulokset suhtautuvat tutkimuskysymyksiin, teoreettiseen viitekehykseen (CRISP-DM) ja kuinka luotettavia ne ovat. Johtopäätöksissä nostetaan esille myös kirjallisuuden luotettavuus ja relevanttius ja kuinka niitä voidaan kehittää sen määrittelemisessä millaista analytiikkatukea asiakkaan mahdollisesti haluavat.

### 3 Data ja analytiikka

Tämän luvun tavoitteena on esitellä lukijalle tutkimuksen aihepiirin peruskäsitteet kirjallisuuskatsauksen perusteella tutkitusta tiedosta. Tämän luvun perustiedoilla luodaan perusta neljännen luvun käsitteille.

Tässä luvussa esitellään ensin se, mitä data käsitteenä tarkoittaa tämän tutkimuksen viitekehyksessä. Datan esittelyn jälkeen käsitellään analytiikkaa käsitteenä ja sen eri osa-alueet yleisesti hyväksytyyn jaon mukaan (Sharda et al., 2018). Tutkimuksen empiirisen kontekstin ymmärtämiseksi luvussa käsitellään tarkemmin myös visuaalinen analytiikka, joka liittyy kiinteästi moderniin analytiikkaan ammatillisessa mielessä (Kandel et al., 2012).

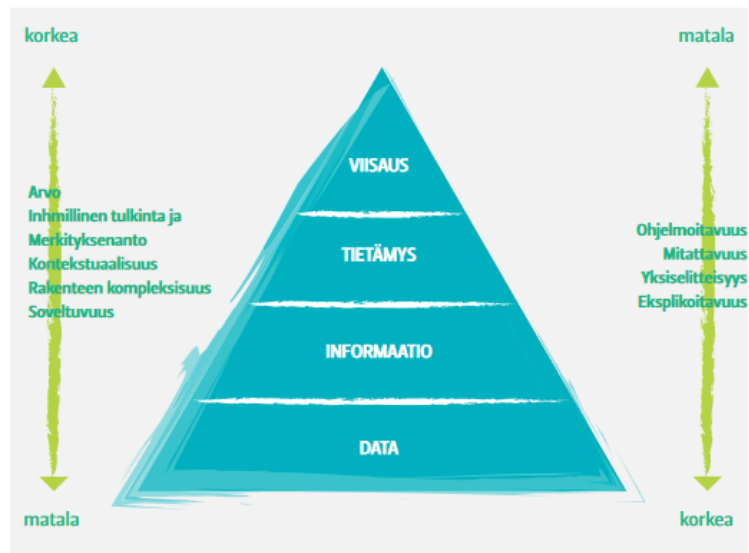
#### 3.1 Data

Data liitetään usein erilaisten termien eteen melko vapaasti tieteellisen keskustelun ulkopuolella, joka on ymmärrettävää sillä dataa ja se jalostettuja muotoja käytetään usein riskiin (Baskarada & Koronios, 2013), vaikka ne eivät tarkoita samaa asiaa. Datan, informaation, tietämyksen ja viisauden ymmärtämiseksi on kuitenkin muodostettu tieteellisessä kontekstissa yleisesti hyväksytty malli, johon tässäkin tutkimuksessa viitataan.

Data on yksittäisiä arvoja, joita ei ole vielä käsitelty, ei sinänsä tarkoita mitään, vaan on esimerkiksi vasta lista lukuja ilman kontekstia. Datasta saadaan jalostettua informaatiota antamalla sille konteksti esimerkiksi tulkitsemalla listan lukuja yrityksen myyntilukuina tai osana data-analytiikan prosessia.

Baskarada ja Koronios (2013) artikkelissaan kokoavat datalle erilaisia ominaisuuksia, kuten sen, että datalla ei ole arvoa tai tarkoitusta itsekseen ja, että data on vailla järjestystä ja omaa merkitystä. Datan on kuvattu myös olevan erilaisia asioita kuten, aktiviteetteja, tapahtumia ja asioita kuvaavia perustietoja.

Tieteellisessä kirjallisuudessa data jäsennetään usein hyväksytysti osaksi tiedon hierarkiaa, joka tarkoittaa sitä, että tiedolla on eri tasoja, joilla kaikilla on omat ominaispiirteensä ja paikkansa. Tässä luvussa viitataan erilaisiin tiedon tasoihin Tietojohdajan taskukirjassa suomennetuilla termeillä (Käpylä & Saloniemi, 2013).



Kuva 3. Tiedon pyramidi Rowleyn mallia mukailleen (Käpylä & Saloniemi, 2013)

Tiedon pyramidi (kuva 3) tai viisaushierarkia on akateemisesti yksi käytetyimmistä tiedon jäsennysmalleista, jossa esitetään, että datasta jalostetaan informaatiota, informaatiosta tietämystä ja tietämyksestä viisautta. Mitä korkeammalle pyramidissa edetään, sitä hankalampaa sen mittaaminen on. Vaikka tiedon pyramidi onkin akateemisesti vakiintunut malli, sitä käyttäessä on huomattava, että itse mallista on käyty melko vähän suoraa keskustelua ja, vaikka malli onkin vakiintunut, prosessit siitä kuinka data etenee pyramidissa ylöspäin kohti viisautta eivät ole niin vakiintuneita. (Rowley, 2007).

### 3.2 Analytiikka

Analytiikka on terminä nykyään hyvin käytetty ja sitä on liitetty moneen erilaiseen termiin ja analytiikka termiä käytetään usein rinnakkain tai korvaamaan sitä jossain määrin lähellä olevia konsepteja:

- data-analytiikka
- business intelligence
- decision support systems
- business analytics,

- web analytics
- big-data analytics
- Prescriptive analytics
- Operational analytics
- Real-time analytics

(Davenport, 2012; Nelson, 2018)

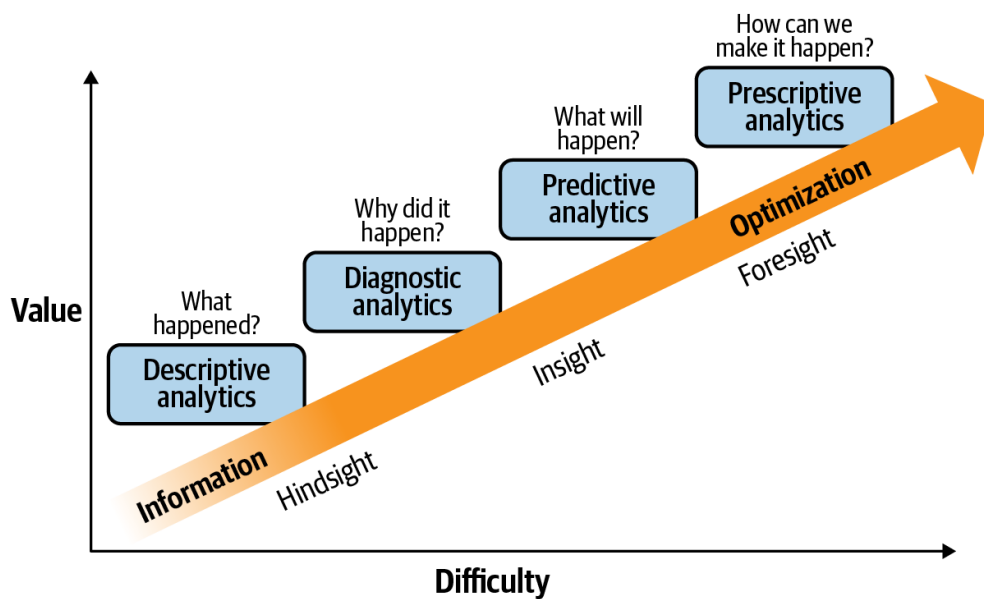
Vaikka edellä oleville käsitteille on olemassa tieteellisessäkin yhteisössä määrittelyjä, niin osa termeistä on tullut alalle teknologiatoimittajien tuotteiden mukana ja niiden hallitseminen alan ammattilaisten tai tieteenharjoittajien ulkopuolella voi olla haastavaa (Nelson, 2018). Termien käytössä historian saatossa on ollut vaihtelua, mutta nykyhetkellä käytössä on “analytics” (Davenport, 2012), joka Suomessa usein kääntyy arkikielessä data-analytiikkaan tai analytiikkaan.

Data-analytiikka ja analytiikka taas termeinä tarkoittavat samaa asiaa, tieteellisessä kirjallisuudessa termejä käytetään vaihtoehtoisina toisilleen ja niiden välille ei tehdä eroa (Balusamy et al., 2021). Kummankin termin harjoittamiseen tarvitaan dataa (Gartner, 2023). Data-analytiikan termin käytöllä voikin olla taustalla datan korostaminen halutussa kontekstissa.

Data-analytiikka tai analytiikka määritellään kokonaisvaltaiseksi, data-ohjautuvaksi ongelman ratkaisuksi (Nelson, 2018). Analytiikka käsittää datan laaja-alaisen hyödyntämisen, selittävien ja ennustavien mallien käytön, kvantitatiivista ja tilastollista analyysia (Davenport et al., 2007). Analytiikka ei ole teknologia vaan teknologia mahdollistaa analytiikan käytön. Data-analytiikka ja Business Intelligence ei välttämättä eroa käytännön kielessä mitenkään, mutta Nelson (2018) määrittelee BI:n strategiana jossa järjestelmällisesti pyritään lähestyä päätöksen tekoa tukemalla sitä datalla. BI:hin voidaan laskea mukaan datan kerääminen, se puhdistaminen, tiedon integrointi ja varastointi. Onkin vähintään maallikkokielessä näkökulmakysymys, sanotaanko analytiikan sisältävän BI:n vai toisin päin, mutta BI usein ei ehkä sisällä yhtä monimutkaisia menetelmiä kuten osalla analytiikan alakäsitteistä on. BI on usein menneisyyteen perustuvaa raportointia, joten voidaankin todeta, että tieteellisessä kontekstissa BI on yksi analytiikan alakäsitteistä. Analytiikkaa voidaan harjoittaa ilman olettamaa, että analytiikan tuloksilla tuetaan päätöksen tekoa tai suunnataan sitä johonkin suuntaan., joka on läsnä BI:n harjoittamisessa. Analytiikka on analytiikkaa, vaikka sillä ei haeta tukea päätöksiin tai toimintoihin (Davenport, 2012).

Analytiikka voidaan jakaa neljään eri yleisesti hyväksyttyyn neljään eri osa-alueeseen (kuva 4.) (Suomenkieliset käännökset yhdenmukaistettu Sitran julkaisusta (Takala, 2018)):

- Kuvaileva analytiikka (descriptive analytics)
- Diagnosoiva analytiikka (diagnostic analytics)
- Ennustava analytiikka (predictive analytics)
- Ohjaava analytiikka (prescriptive analytics)



Kuva 4. Gartnerin analyttisen kehittymisen malli (Finn et al., 2020).

Kuvaileva analytiikka kertoo mitä menneisyudessa tapahtui (Liebowitz, 2021) tai mitä juuri nyt tapahtuu (Gartner, 2023) ja onkin analytiikan perinteisen muoto. Kuvailevan analytiikan avulla voidaan koota dataa erilaisiin segmentteihin, jolloin datan luokittelu helpottuu. Esimerkkinä kuvailevan analytiikan avulla voidaan selvittää, paljonko tiettyä tuotetta on myyty viime vuonna tai montako uutta asiakasta on saatu hankittua vuoden alusta tähän päivään asti. Tämän tyyppinen analytiikka esitetään usein visualisoitu erilaisin mittarein tai raportein ja se ei ota kantaa siihen, miksi jokin asia on oikein tai väärin käyttäjän kannalta. (Liebowitz, 2021).

Diagnosoiva analytiikka vastaa kysymykseen miksi jokin asia tapahtui (Cochran, 2019; Gartner, 2023) ja sen avulla on mahdollista hakea vastauksia siihen, mitkä asiat korreloivat toisiinsa (Shi-Nash & Haroon, 2016). Diagnosoiva analytiikka voi kertoa vastauksen

esimerkiksi siihen, että miksi asiakas valitsi juuri tietyn tuotteen toisen tuotteen asemasta tai miksi matkustajat valitsevat raitiovaunun bussiliikenteen sijasta. Diagnosoiva analytiikka ja kuvaileva analytiikka koetaan yleensä osana taaksepäin katsovaa analytiikkaa (hindsight), koska ne koskevat suurelta osin dataa, joka on jo tapahtunut (Shi-Nash & Hardoon, 2016). (Gartner, 2023)

Ennustava analytiikka pyrkii nimensä mukaisesti päätellä sitä mitä tulevaisuudessa todennäköisesti tapahtuu (Sharda et al., 2018) eli vastaamaan kysymykseen ”Mitä voisi tapahtua?” (Takala, 2018). Tulevaisuuden päättelemiseksi tarvitaan kehittyneitä malleja, koneoppimista (Geng, 2017), datan louhintaa ja muita uudempia menetelmiä (Sharda et al., 2018). Ennustava analytiikka pohjautuu usein kuvailevan ja diagnosoivan analytiikan pohjalle, mutta pelkkä diagnosoiva analytiikka yksinään ei useinkaan riitä luotettavan ennusteen luomiseen, vaan siihen tarvitaan aiemmin mainittuja edistyneempiä tekniikoita onnistuneen mallin luomiseksi (Cochran, 2019). Ennustavalla analytiikalla voidaan pyrkiä päättelemään esimerkiksi sitä, että jos kuluttaja ostaa lipun junaan Tampereen ja Helsingin välille, niin mitä tuotteita hän kuluttaa ravintolavaunussa todennäköisemmin tai kuinka todennäköisesti käyttäjä valitsee Applen tietokoneen Lenovon kannettavan sijaan.

Ohjaava analytiikka pyrkii tarjoamaan tietoa siitä, miten haluttu lopputulos voitaisiin saavuttaa tai miten siihen voitaisiin vaikuttaa (Gartner, 2023) tai optimoimaan jonkin järjestelmän tai organisaation toimintaa (Sharda et al., 2018). Ohjaileva analytiikka vastaa siis kysymykseen, ”Mitä meidän pitäisi tehdä?” (Takala, 2018). Ohjaavan analytiikan pohjalta tehty analyysin muodosta on keskustelua, sillä osassa kirjallisuudesta (Liebowitz, 2021) painotetaan sitä, että tulos ei ole kyllä/ei-tyyppinen vastaus vaan useamminkin sarja toimenpiteitä tavoitteen saavuttamiseksi. Sharda, Delen ja Turban (2018) kuvailevat ohjaavan analytiikan tuloksia myös kyllä/ei-tyyppisiksi ja hinnan muutossuosituksiksi, useiden toimenpide suositusten lisäksi. Ohjaava analytiikka voi kuvata esimerkiksi sitä, että miten tuotteen hinnoittelua pitää muuttaa, jotta myynti lisääntyisi tai että kuinka pakkauskojoja muuttamalla voitaisiin parantaa yrityksen kannattavuutta.

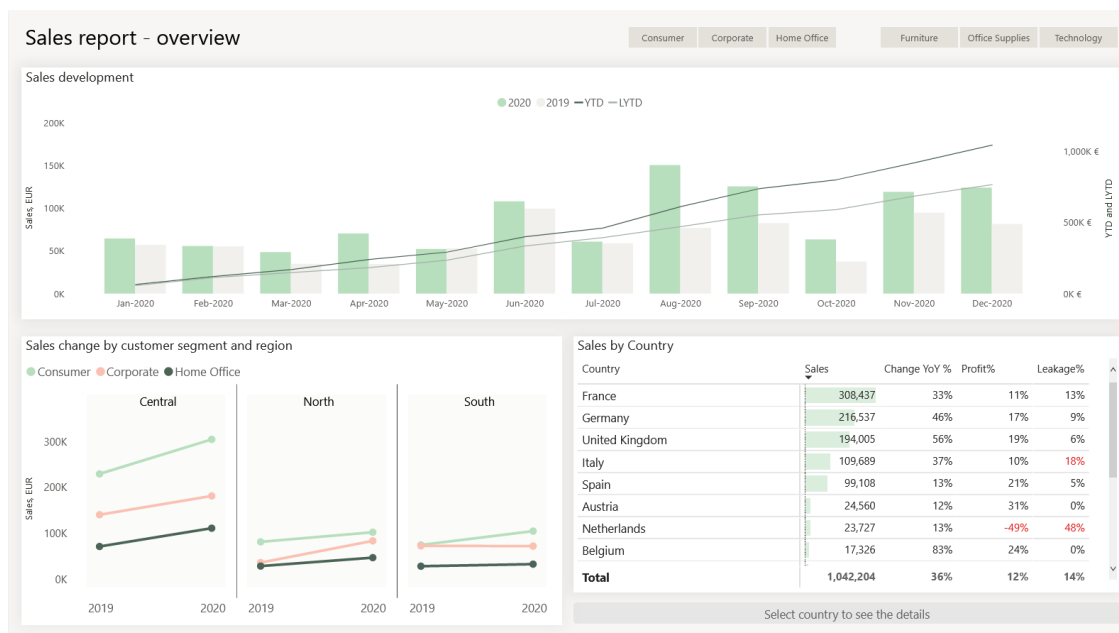
### **3.3 Visuaalinen analytiikka**

Visuaalinen analytiikka tai visuaalinen data-analytiikka on datan esittämistä erilaisin visuaalisointi tekniikoin (Sharda et al., 2018) ja visuaalisuus mahdollistaa usein suurien datamassojen esittämisen ymmärrettävässä muodossa (Geng, 2017). Gartnerin julkaiseman ”Magic Quadrant for Analytics and Business Intelligence Platforms” (Schlegel, 2023) visuaalisen analytiikan työkalujen markkinajohtajat määrittelevät visuaalisen analytiikan työkaluksi, joka mahdollistaa datan muuttamisen helposti ymmärrettävään ja muihin datalähteisiin yhdistettävään graafiseen muotoon (Microsoft, 2023b; Tableau, 2023). Sekä Microsoft (2023b) ja Tableau (2023) tunnistavat visuaalisen analytiikan ominaisuudeksi



sen, että visuaalisen representaation avulla datasta on mahdollista tunnistaa trendejä, poikkeuksia ja toistuvuuksia.

Visuaalisen analytiikan koetaan liittyvän erityisesti kuvailevan, diagnosoivan ja ennustavan analytiikan pariin (Sharda et al., 2018) ja muuttavan analytiikan matemaattista luonnetta enemmän kommunikointiin päin (Cochran, 2019). Voikin sanoa, että modernit visualisoinnin työkalut tarjoavat käyttöliittymän dataan, tarjoten mahdollisuuden analyttiseen päättelyyn ja datan käsittelyyn käyttäjätasoisemmassa ympäristössä (Nelson, 2018).



Kuva 5. Interaktiivinen näkymä Power BI-ohjelmasta demo-datalla (Microsoft, 2023a).

Tämän tutkimuksen empiirisen kontekstissa erilaiset visuaalisen analytiikan työkalut kuten Microsoft Power BI tai Tableau liittyvät kiinteästi visuaalisen analytiikan toteuttamiseen. Työkalut mahdollistavat erilaisten näkymien (dashboard) rakentamisen, joita käyttämällä käyttäjät voivat saada yhdellä näkymällä tiivistetyn näkymän datasta (Myers & Gregory, 2021). BI-työkalut ovat kehittyneet staattisesta raportoinnista dynaamisiin, interaktiivisiin näkymiin, joissa käyttäjät voivat suodattaa, tutkia ja etsiä syitä tapahtumille itsenäisesti datasta (Davenport et al., 2010; Myers & Gregory, 2021). Käytännössä työkalut mahdollistavat siis käyttäjän itsensä suorittaman data-analyysin visuaalisin keinoin.

Vaikka visuaaliseen analytiikkaan on liitettävissä ”itse-palvelu-analytiikka” (self-service), niin visuaalisuus itsessään ei tarkoita sitä, että datan perusteella on helppo tehdä päätelmiä, vaan esittämiseen pitää kiinnittää huomiota, sillä liian monimutkaiset keinot ovat loppukäyttäjille hankalia (Davenport et al., 2007). Visuaalista analytiikkaa tehtäessä

onkin tärkeää ottaa huomioon se, kenelle analytiikkaa esitetään (Cochran, 2019) ja usein loppukäyttäjät eivät ole teknisen tai matemaattisen alan ihmisiä, jolloin esitystavassa pitää ottaa huomioon se, miten monimutkaisia tuloksia esitetään ymmärrettävästi (Geng, 2017).

## 4 Analytiikan elinkaari

Tutkimuksen aiheena on analytiikan tukemisen arviointi asiakaslähtöisestä näkökulmasta ja kokonaisvaltaisen ymmärryksen takia tässä luvussa esitellään analytiikan koko elinkaari kuitenkin painottuen tukivaiheeseen. Analytiikka muistuttaa siinä mielessä muita IT-projekteja, että sillä on elinkaari ja projekteilla on alku -ja loppuvaiheet. Analytiikan ollessa nykyisessä laajuudessaan nuorempi ala verrattuna esimerkiksi ohjelmistokehitykseen, sen käytänteet eivät ole niin vakiintuneita ja tämän takia ei analytiikkaprojektissa ole aina selvää miten eri vaiheet toteutetaan – tai niiden toimitavat vaihtelevat aina tekijän mukaan. Alan kirjallisuudessa kehoitetaan käyttämään analytiikkakohtaisia elinkaari-malleja jo olemassa olevien IT-mallien sijaan (Davenport, 2012) ja perinteisiin IT-kehitysmalleihin verrattuna analytiikan alalla omaksutut mallit muistuttavat enemmän tutkimista (Provost & Fawcett, 2013).

Tässä luvussa käydään ensin läpi osiossa 4.1 erilaisia tapoja mallintaa analytiikan elinkaarta sillä, vaikka analytiikka ei ole niin vakiintunutta, niin sen elinkaarta on mallinnettu usealla eri tavalla. Osiossa 4.2 avataan tarkemmin CRISP-DM mallia, joka on malleista suosituin ja tämänkin tutkimuksen teoreettinen viitekehys perustuu sen pohjalle. Osiossa 4.3 käsitellään tarkemmin vaihtoehtoisia malleja CRISP-DM:lle ja osiossa 4.3 käsitellään elinkaaren lopussa olevaa tukivaihetta, joka on nostettu omaksi osiokseen sen ollessa tutkimuksen keskiössä.

### 4.1 Elinkaaren mallit

Analytiikan elinkaarta voidaan kuvata eri malleilla, mutta yleistä niille kirjallisuudessa on se, että ne tyypillisesti alkavat prosessina liiketoiminnan tarpeen määrittelynä, toisena datan käsittelyllä, sen mallintamisella ja lopulta analytiikkatuotteen tuotantoon vientinä, kuten Nelson (2018) esimerkiksi julkaisussaan esittää. Elinkaaren hallintaan on julkaisukohtaisten mallien lisäksi kuitenkin vakiintunut tiettyjä prosessimalleja, jotka analytiikkayhteisön parissa on hyväksytty (Cochran, 2019) ja niiden voidaan sanoa olevan alalla jo lähes standardeja (Azevedo & Santos, 2008).

Mallien nimeämiskäytäntö voi olla lukijaa hämäävää, sillä niiden yhteydessä käytetään termejä kuten tiedon louhinta (data mining), viitataan prosessien liittyvän vain data science osa-alueeseen tai viitataan niiden keskittyvän vain tulevaisuutta ennustavan analytiikan kehittämiseen. Mallit on kuitenkin otettu käyttöön analytiikan kokonaisvaltaisessa elinkaaren hallinnassa (Liebowitz, 2021) ja prosessimalleihin kuten CRISP-DM:n viitataan viitekehysenä koko analytiikan alueella, joka jäsentää tietämyksen tai informaation irrottamista raasta datasta järjestelmälliseksi hyvin määritellyiksi vaiheiksi

(Provost & Fawcett, 2013). Voikin tulkita, että mallien käytön nimeämisessä ja käytön suosittelussa näkyy samat haasteet kuin itse analytiikan nimeämisessä, sillä nimiä on monia ja ne voivat aiheuttaa haasteita mallien tulkitsijoille.

Analytiikan elinkaaren malleja on useita, mutta niistä suosituimmat ovat:

- CRISP-DM
- SEMMA
- KDD
- Oma malli tai oman organisaation malli tai ei mitään mallia

(Aho et al., 2021; Hotz, 2023; Mariscal et al., 2010)

Suosittujen mallien lisäksi seuraavissa kappaleissa esitellään ASUM-DM-malli, joka on yritys parantaa ja paikata CRISP-DM:n puutteita.

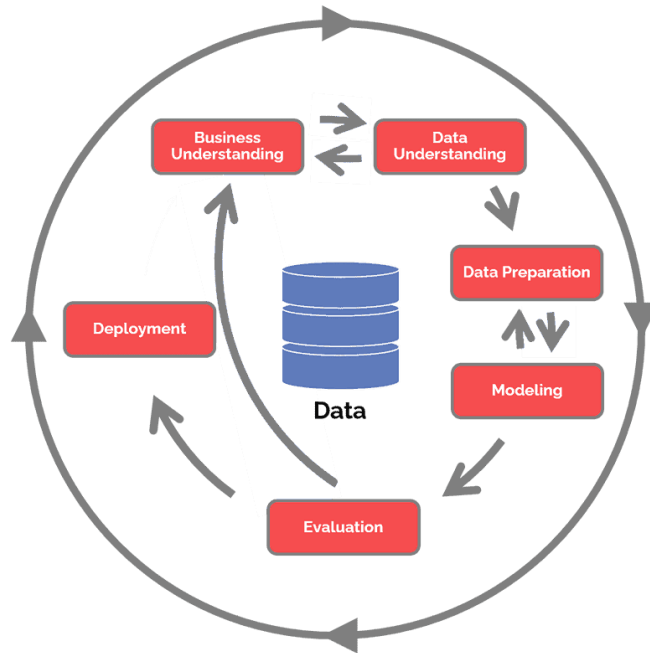
Käytettyjä malleja on useita, mutta on huomattava, että useat organisaatiot tai analytiikan parissa toimivat eivät käytä niitä ollenkaan, vaan korvaavat valmiit mallit omilla malleilla, organisaation malleilla tai ei millään ohjatulla toimintatavalla (Hotz, 2023; Mariscal et al., 2010). Näiden mallien lisäksi löytyy vaihtoehtoisia malleja, jotka ovat tutkimusten valossa niin pienessä käytössä, että tässä tutkimuksessa ei perehdytä niihin.

## **4.2 CRISP-DM**

CRISP-DM (The Cross Industry Standard for Process for Data Mining) on analytiikan prosessimalleista suosituin, ja sen kehittävät yhteistyössä joukko tiedonlouhinnasta kiinnostuneita organisaatiota, mukaan lukien Teradata, Daimler AG, NCR Corporation, OHRA ja SPSS (IBM). Kuten mallin nimikin kertoo, se kehitettiin alun perin erityisesti tiedon louhintaan, mutta analyttinen yhteisö on adoptoinut sen käyttöön kaikenlaisiin analytiikan projekteihin (Cochran, 2019).

CRISP-DM on vakiintuneista analytiikan elinkaarimalleista suosituin (Cochran, 2019; Mariscal et al., 2010) ja kirjallisuudessa siihen viitataan de facto-standardina (Mariscal et al., 2010) tai lähimpänä sitä mitä analytiikan alalla standardi on. Lisäksi se vaikuttaa olevan viitatuin malli myös tieteellisessä kirjallisuudessa (Aho et al., 2020). Mallista löytyvän runsaan kirjallisuuden, sen adoptioasteen ja ammattikirjallisuuden pohjalta CRISP-DM on valittu tämän tutkimuksen empiirisen haastatteluosuuden teoriakehykseksi.

CRISP-DM on, kuten kuvasta 6 on nähtävissä, iteroiva menetelmä ja vaikka prosessissa olisi päästy vaiheeseen ”Deployment” asti, niin projekteissa palataan usein ongelman määrittelyyn takaisin (Business Understanding), jotta ratkaisusta saataisiin parempi tai sen seurauksena, että data tai liiketoimintaprosessit ovat muuttuneet (Provost & Fawcett, 2013). Myös liikkuminen eri vaiheiden välillä on mallin mukaan sallittua ja usein analytiikkaprojektiin liittyy iteroitukierroksia eri vaiheiden välillä (Geng, 2017).



Kuva 6. CRISP-DM-malli Data Science Process Alliancen kuvaamana (Hotz, 2023).

CRISP-DM koostuu dokumentaationsa mukaan (Chapman et al., 2000) kuudesta eri vaiheesta, jotka ovat:

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Prosessin jokaisella vaiheella on määriteltyjä alatehtäviä, joilla prosessia ohjataan vielä tarkemmin. Seuraavassa osiossa alivaiheita ei eritellä erikseen omiksi tutkielman kappaleiksi, vaan ne avataan sanallisesti opinnäytetyön vaatimassa laajuudessa, joka mahdollistaa riittävän kokonaiskuvan saamisen empiirisen osion kannalta.

CRISP-DM malli alkaa vaiheella ”Business Understanding”, jonka tarkoituksena on ymmärtää liiketoiminnan tavoitteet analytiikkaratkaisulta ja määrittellä sille onnistumisen mittarit. Vaiheeseen kuuluu myös se, että siinä selvitetään projektille käytössä olevat resurssit, oletukset ja mahdolliset rajoitteet. Tässä vaiheessa listataan projektiin liittyvät riskit ja mahdolliset projektia vaikeuttavat tapahtumat ja niille laaditaan tarvittavat varasuunnitelmat. Viimeisenä ensimmäisessä vaiheessa projektille määritellään onnistumisen mittarit myös teknisessä sekä analyttisessä mielessä ja laaditaan projektisuunnitelma, jossa kuvataan edellä mainitut kohdat ja valitaan alustavat työkalut ja teknologiat projektille. CRISP-DM:n iteratiivisen luonteen vuoksi projektisuunnitelma elää prosessin eri vaiheissa ja siihen tulisikin palata vähintään eri vaiheiden välillä, jolloin projektin etenemistä voidaan arvioida ja suunnitelmaa tarkentaa. (Chapman et al., 2000)

Mallin toisessa vaiheessa ”Data Understanding” pyritään ymmärtämään dataa ja sillä yhdessä ensimmäisen vaiheen kanssa luodaan pohja koko projektille. Vaiheessa aluksi kerätään projektille tarpeellinen data, joka yleensä ladataan valittuun työkaluun, jolla dataa käsitellään. Käsittelyn jälkeen dataa tutkitaan ja sen ominaisuuksia pyritään kuvailemaan esimerkiksi sen määrää, taulujen kenttien nimiä tai muita pinnallisia ominaisuuksia. Dataan tutustumisen jälkeen dataa suositellaan tutkittavan visualisoinnin, kyselyiden (query) ja eri arvojen suhteiden tunnistamisella. Viimeisenä tässä vaiheessa tutkitaan datan laatua ja dokumentoidaan mahdolliset löydetty ongelmat, kuten virheelliset tai puuttuvat arvot. (Chapman et al., 2000)

Kolmannessa vaiheessa ”Data Preparation” valmistellaan dataa seuraavaa vaihetta varten ja vaiheen jälkeen datajoukon tulisi olla valmis. Datan valmistelu on usein analytiikka-projektin työläin ja pisin vaihe (Hotz, 2023). Vaiheessa kaksi valitusta datasta valitaan käsiteltäväksi sopivat datajoukot ja dokumentoidaan syyt niiden sisällyttämiselle projektiin ja dokumentoidaan myös syyt pois jätetylle datalle. Datan valitsemisen jälkeen se puhdistetaan, joka tarkoittaa edellisessä vaiheessa löydettyjen ongelmien korjaamista ja korjausten dokumentointia. Puhdistuksen jälkeen datajoukkoon rakennetaan uusia attribuutteja, jotka voivat olla esimerkiksi uusi kenttä tauluun kahden erillisen muuttujan laskuna, kuten painoindeksin laskeminen pituuden ja painon avulla. Vaiheeseen kuuluu myös se, että dataa yhdistetään useammasta lähteestä tarvittaessa tai sitä aggregoidaan tarvittavalle tasolle (esimerkiksi päiväkohtaista myyntiä kuukausitasolle). Yhdistäminen on tyypillisesti sitä, että samasta objektista olevia tietoja yhdistetään samaksi tauluksi eri tauluista. Viimeisenä datan valmisteluvaiheessa data formatoidaan sopivaan muotoon muuttamalla esim. tekstikenttä numerokentäksi tai dataan voidaan lisätä indeksisarake. (Chapman et al., 2000)

CRISP-DM:n neljäs vaihe on ”Modeling” eli tiedon louhinnan mallin valinta eli tarkka mallintamistekniikka, jolla haluttu tavoite saavutetaan, esimerkkinä regressiomalli. Mallin valinnan jälkeen mallia testataan ja testauksen tuloksia validoidaan ja testausprosessi dokumentoidaan. Mallintaminen ei kuitenkaan nykyaikaisilla työkaluilla ja CRISP-DM:n laajemmalla hyödyntämisellä tarkoita ainoastaan perinteisiä tiedon louhinnan menetelmiä, vaan esimerkiksi datan visualisointia ja yksinkertaisempia statistisia menetelmiä (Nelson, 2018). Testauksen päätyttyä malli suoritetaan varsinaiselle datajoukolle ja sitä tarvittaessa korjataan. Usein mallintamistekniikoita kokeillaan useita, ja vaiheen lopuksi niiden tuloksia vertaillaan, dokumentoidaan ja lopulta sopivin valitaan. (Chapman et al., 2000)

Viides vaihe on ”Evaluation”, jossa pyritään arvioimaan sitä, miten analytiikka kokonaisratkaisuna vastaa ensimmäisessä kohdassa määriteltyihin liiketoiminnan tarpeisiin. Jos teknisiä ratkaisuja on useampia, niin jokainen niistä arvioidaan erikseen. Myös koko kehitysprojekti arvioidaan kokonaisuutena ja pyritään tulkitsemaan, onko kaikki tarvittavat vaiheet suoritettu ja jos jotain on unohtunut, se pyritään korjaamaan. Löydösten perusteella päätetään, tarvitseeko ratkaisu iterointia vai onko se valmis julkaistavaksi. (Chapman et al., 2000)

Prosessimallin viimeisenä vaiheena on ”Deployment” eli analytiikan toimittaminen loppukäyttäjille. Ensimmäisenä tähän vaiheeseen kuuluu se, että mallin tuotantoon vienti suunnitellaan järjestelmällisesti. Analytiikkaratkaisun toiminnan varmistamiseksi sen valvonnan ja huollon suunnittelu tulee tehdä, jotta mahdolliset tuotannossa esiin nousevat ongelmat voidaan havaita. Mahdollisia ongelmia voi olla esimerkiksi liiketoimintalogiikan muuttuminen ja sen mahdolliset haitalliset vaikutukset analytiikkaratkaisun toimintaan, ratkaisuun tuleva uusi data tai muu tekijä, joka tekee ratkaisusta epäluotettavan. Tuotantoon viennin jälkeen projektista tehdään loppuraportti ja projektin arviointi, jossa arvioidaan asioita, jotka onnistuivat, epäonnistuivat ja liiketoiminnan tavoitteiden toteutumista. (Chapman et al., 2000)

### **4.3 Muut mallit**

Esitellyistä malleista KDD (Knowledge Discovery in Databases) (Saltz, 2015) esiintyi ensimmäisen 1980-luvulla (Mariscal et al., 2010) ja sillä on kuusi eri vaihetta (Taulukko 2). Ensimmäisessä vaiheessa (selection) valitaan sopiva datajoukko, josta on tarkoitus löytää tietämystä, toisena (pre-processing) dataa puhdistetaan ja varmistetaan datan yhtenäisyys, kolmantena dataa muokataan, karsitaan tai datalle muodostetaan dimensioita. Neljäntenä vaiheena datasta pyritään löytämään merkityksellisiä trendejä tai huomioita,

viidentenä ja viimeisenä tuloksia tulkitaan ja arvioidaan. Huomioitavaa on se, että KDD-prosessin nimettyihin vaiheisiin ei sisälly liiketoimintaymmärrystä, vaan malli olettaa, että tarvittava tieto on hankittu ennen varsinaista prosessia. (Azevedo & Santos, 2008)

SEMMA (Sample, Explore, Modify, Model, Asses) on SAS Instituutin vuosituhatosen alkuvaiheessa kehittämä menetelmä SAS Enterprise minerin tarpeisiin (Mariscal et al., 2010). SEMMAN dokumentaatiossa (Institute, 2023) prosessin ensimmäiseksi vaiheeksi määritellään ”Sample”, jossa tarkoituksena on prosessoida esimerkki datasta, jonka jälkeen vaiheessa ”Explore” tutkitaan datasta ennakkoon odotettuja yhteyksiä, löydetään odottamia trendejä tai muita poikkeamia. Kolmantena kohtana prosessissa on ”Modify”, joka tarkoittaa datan muokkaamista, siistimistä ja analysointia. Neljäntenä prosessissa on ”Model” vaihe, joka viittaa ennustavan mallin muodostamista erilaisten valmiiden mallien pohjalta, viidentenä ja viimeisenä malleja vertaillaan ja tulosten luotettavuutta arvioidaan. Menetelmänä SEMMA eroaa muista siten, että se on integroitu SAS Enterpriseen ja on epätodennäköistä, että sitä käytetään laajamittaisesti työkalun ulkopuolella (Mariscal et al., 2010).

Kolmas esiteltävä malli on ASUM-DM (Analytics Solutions Unified Method for Data Mining/Predictive Analytics), joka on CRISP-DM:n paranneltu ja laajennettu versio, jonka kehitti IBM vuonna 2015. ASUM-DM:n tarkoituksena on lisätä uusia vaiheita CRISP-DM:n päälle ja paikata sen puutteita. Kirjallisuudessa parannuksiksi mainitaan big datalle sopivuus, projektin hallinnan sisällyttäminen prosessiin, uudet nostot kommunikointiin ja tiedon hallintaan (Angée et al., 2018). Mallin dokumentaatiota ei ole löydettävissä IBM:n sivuilta ja siihen viitataan huomattavasti vähemmän kuin CRISP-DM:n kirjallisuudessa ja lähdeviitteet eivät johda enää valideihin verkkosivuihin (Ponsard et al., 2017), vaikuttaakin, että malli on haudattu lähivuosina. Vapaissa lähteissä kuten Data Science Alliancen (2023) verkkosivuilla ASUM-DM:n adoption arvellaan olevan minimaalista. ASUM-DM:n vaiheet ovat lyhyesti kuvailtuna seuraavat:

- ”Analyze”, eli liiketoimintaongelman ymmärtäminen
- ”Design” eli datan valmistelu
- ”Configure & Build” analytiikan mallintaminen, sen arviointi, testaaminen ja operatiivisen suunnittelun suunnittelu
- ”Deploy” eli QA-testaus, vastuunsiirto operatiiviselle, tuotantoon vienti
- ”Operations & Optimize”, ratkaisun valvonta, optimointi, käyttäjien tukeminen ja ratkaisun hallinta

(Ponsard et al., 2017)



KDD	SEMMA	CRISP-DM	ASUM-DM
<i>Pre KDD</i>	-----	Business Understanding	Analyze
Selection	Sample	Data Understanding	Design
Pre-processing	Explore		Configure & Build
Transformation	Modify	Data preparation	Deploy
Data mining	Model	Modeling	Operate & Optimize
Interpretation/Evaluation	Assesment	Evaluation	
<i>Post KDD</i>	----	Deployment	

Taulukko 2. Erot eri malleissa. Mukaelma Azevedon ja Santosin taulukosta (Azevedo & Santos) verrattuna ASUM-DM-mallin vaiheisiin (Ponsard et al., 2017)

Taulukossa 2. vertaillaan eri mallien vaiheita mukaillen Azevedon ja Santosin (2008) tekemää vertailua, jossa todettiin, että SEMMA:n ja KDD:n mainitut vaiheet ovat pitkälti vertailukelpoisia ja SEMMA:n erottaa KDD:sta sidos SAS Enterprise Minerin. CRISP-DM:n ja KDD:n askelissa on prosessimallissa taas selkeämpi ero ja äkkiseltään katsottuna, voidaan tulkita, että liiketoiminnan ymmärrys ja ratkaisun tuotantoon vienti puuttuu KDD:sta. Azevedo ja Santos (2008) kuitenkin argumentoivat, että KDD:hen sisältyy oletus siitä, että ennen prosessin soveltamista käyttäjällä tulisi olla toimialakohtaista ymmärrystä datasta. ASUM-DM:n ja muiden prosessimallien vertailu onkin hankalampaa, sillä aiheesta ei löydy juurikaan kirjallisuutta ja malli on tehty yhdistelmänä CRISP-DM:ää, perinteisiä projektinhallintamenetelmiä ja ketteriä menetelmiä (Ponsard et al., 2017). Tästä syystä taulukkoon on jätetty selkeä väli perinteisempien menetelmien ja ASUM-DM:n välille.

Mallien evoluutio kulminoituu kahteen taitepisteeseen, joista ensimmäinen on KDD:n kehittäminen. Kaikkien mallien voi katsoa kehittyneen KDD:stä, johon pohjautuvia malleja on useita (kuten SEMMA). Toisena taitepisteenä voidaan pitää CRISP-DM:n kehittämistä, joka pohjautuu KDD:hen ja sen ”jälkeläisiin”. CRISP-DM toimii KDD:n tavoin useamman mallin kehityspohjana (kuten ASUM-DM) ja vaikka CRISP-DM on suosituin malli edelleen, niin yrityksiä sen kehitykseen on ollut. (Mariscal et al., 2010)

#### 4.4 Analytiikkatuki

Analytiikkatuki tarkoittaa analytiikkaratkaisun tukemista ratkaisun tuotantoon viennin jälkeen ja tässä opinnäytetyössä se on keskiössä CRISP-DM:n viimeisen vaiheen ”Deployment” laajennettuna osana. Tuen roolia ja tehtäviä on hyödynnetty opinnäytteen empiirisessä osiossa haastattelurungon pohjana. Tämän osion tarkoituksena on esitellä siihen liittyvät osat kuten Cochran (2019) on ne teoksessaan esitellyt. Lisäksi kappaleen lopussa esitellään aiheesta löytynyttä muuta tutkimusta.

Osiossa 4.2. esiteltiin analytiikan elinkaarimalli CRISP-DM, jonka viimeinen vaihe ”Deployment” eli tuotantoon vienti sisälsi seuraavia asioita:

- Analytiikkaratkaisun tuotantoon vienti
- Analytiikkaratkaisun valvonta ja huolto
- Ratkaisun jatkuvan toiminnan varmistaminen muutoksista huolimatta
- Loppuraportti ja loppuarviointi

Cochran (2019) laajentaa CRISP-DM:n määritelmää kirjassaan vaihetta tuomalla siihen Certified Analytics Professionalin (CAP, 2023) määritelmän analytiikan ammattilaisen ammattitaidon sisällöstä JTA:n (Job Task Analysis) viitekehyksessä. Määritelmän alkuperäisenä tarkoituksena on tarjota alusta, jonka avulla yksittäisen analytiikan ammattilaisen osaaminen tietyllä analytiikan elinkaaren alueella voidaan kuvata tarkasti ja sertifioidusti. Cochran (2019) yhdistää sopivilta osin kumpaankin, mutta tässä osiossa esitellään vain laajennettu osa CRISP-DM:n viimeiseen vaiheeseen liittyen analytiikkatuki keskiössä. Kirjassaan Cochran (2019) nimeää nämä tehtävät ”Post deployment tasks” eli tuotantoon viennin jälkeiseksi tehtäviksi, mutta ne risteävät CRISP-DM:n ”deployment”-vaiheen tehtävien kanssa.

Laajennettu määritelmä käsittää seuraavat kohdat:

- Dokumentointi
- Laadun valvonta ja ajankohtaisuuden varmistaminen
- Käyttäjien koulutus
- Liiketoimintahyötyjen arviointi

Dokumentointi tarkoittaa sitä, että sen analytiikkaratkaisun toiminta dokumentoidaan siten, että sitä ylläpitävä taho kykenee pitämään ratkaisun toiminnassa. Isomman kokoluokan ratkaisuissa kehittävä tiimi ja ylläpitävä tiimi on yleensä erillisiä ja dokumentaation

määrittelyt syntyvät heidän yhteistyönään ja usein tuotantoon vientien yhteydessä kehitystiimiä tarvitaan vielä tukiorganisaation tukena (Cochran, 2019). CRISP-DM:n dokumentaatio huomioi myös sen, että ratkaisun päivittämisen prosessi tulisi dokumentoida ja siihen liittyvät erityiset tehtävät listata (Chapman et al., 2000).

Laadun valvonta ja ajankohtaisuuden varmistaminen tarkoittaa sitä, että analytiikkaratkaisun toimintaa seurataan sen luotettavuuden varmistamiseksi. Riippuen siitä mihin ratkaisu on tarkoitettu, sen tarkkuus ja sitä kautta luotettavuus voi laskea ajan kuluessa. Esimerkiksi ratkaisuun upotetut liiketoiminnan raja-arvot voivat muuttua ajan kuluessa ja niiden seurauksena ratkaisun tarjoama informaatio voi heiketä, ellei tilannetta korjata (Cochran, 2019). CRISP-DM:n dokumentaatio huomioi myös tämän mainitsemalla mahdollisen tarpeen ratkaisun päivittämiseksi, mutta ei mainitse aiheesta syvällisemmin (Chapman et al., 2000). Cochran (2019) kehottaakin kirjassaan valvomaan järjestelmällisesti ja tietyin aikajaksoin ratkaisun toimintaa, jotta mahdollinen suorituskyvyn tai tarkkuuden väheneminen huomataan ja voidaan raportoida analytiikasta tai datasta vastaavalle taholle ja ratkaisun toiminta voidaan korjata ajan tasaiseksi. Mikäli ratkaisusta ei pidetä huolta, niin on todennäköistä, että sen käyttö ja tarkkuus hiipuu, kun taustalla olevat liiketoimintaprosessit muuttuvat tarpeeksi.

Käyttäjien koulutuksella tarkoitetaan sitä, että analytiikkaratkaisun toiminta selitetään loppukäyttäjille ja käyttäjien toimintaa ratkaisun parissa seurataan. Analytiikkaratkaisun koulutusvastuu on usein IT-tiimillä tai analytiikkaratkaisun kehitystiimillä, joista jälkimmäisen on syvemmän ymmärryksensä takia tyypillisesti helpompi selittää ymmärrettävästi ratkaisun eri osa-alueet loppukäyttäjille. CRISP-DM:n dokumentaatio mainitsee koulutusta vastaavan tilaisuuden loppuraportin esittämisenä relevanteille loppukäyttäjille (Chapman et al., 2000). Mahdollisten koulutusten pitäjiä tulisi ymmärtää analytiikkaratkaisun toiminta täydellisesti, jotta loppukäyttäjille tulisi luottavainen olo ratkaisun toimivuudesta. Loppukäyttäjien koulutuksella vähennetään ylläpitokuormaa. (Cochran, 2019)

Liiketoimintahyötyjen arviointi tarkoittaa sitä, että ratkaisun tuotantoon viennin jälkeen arvioidaan sitä, mitä hyötyä analytiikkaratkaisusta on organisaatiolle. Hyötyen mittaaminen voi olla hankalaa, sillä analytiikasta ei ole aina osoitettavissa suoria hyötyjä. On kuitenkin tilanteita, joissa hyödyt voidaan osoittaa nopeastikin esimerkiksi ennustavien analytiikkamallien osalta, joka on auttanut varaston hallinnassa ja pienentänyt siihen liittyviä kuluja tai analytiikan avulla löydettyjen ongelmallisten työvaiheiden tunnistaminen ja niiden korjaaminen. Onkin tärkeää sitoa analytiikka siihen liittyvään liiketoiminnan KPI-lukuun, joka mahdollisesti paranee analytiikan seurauksena ja ymmärtää taustalla toimiva liiketoimintalogiikka. (Cochran, 2019)

Analytiikan tukemisesta ei ole erityisen paljon akateemista tutkimusta tarjolla, mutta tässä kappaleessa esitellään Bischoff et al. (2015) tekemä monimenetelmätutkimus, jossa viitataan tässäkin kappaleessa aikaisemmin viitattuihin tuen tehtäviin. Tutkimuksen mukaan BI-ratkaisujen (analytiikkaratkaisujen) tutkimus painottuu ratkaisujen hyväksyntään ja omaksuntaan, jatkuvan käytön jäädessä muutaman tutkimuksen varaan, jotka käyttävät lähinnä tietojärjestelmätieteen olemassa olevia tutkimusmalleja tai viitekehyksiä. Tutkimuksen tulokset kertovat, että analytiikkakontekstissa on tärkeää ottaa huomioon käyttäjät suunnittelussa, ja, että jo ratkaisun käyttöliittymän suunnitteluun on syytä kiinnittää huomioita. Ratkaisun hyödyllisyyden ja sen luotettavuuden välillä löydetään myös linkki, joka vahvistaa CRISP-DM:ssä (Chapman et al., 2000) mainittuja liiketoiminnan tavoitteiden saavutusta. Luottamukseen vaikuttaa myös informaation tai datan laatu ratkaisun sisällä. Tutkimuksessa nostetaan esille myös se, että ratkaisun jatkuvaan käyttöön vaikuttaa käyttäjätuki, organisaatiokohtaiset kulttuurit ja käyttäjien vaatimusten täyttäminen. Mitä enemmän käyttäjät ovat mukana ratkaisun kehittämisessä, sitä todennäköisemmin ratkaisuun luotetaan. Tutkimus ei itsessään tarjoa valmista ratkaisua analytiikan jatkuvaan käyttöön vaan tarjoaa pohjan tulevaisuuden tutkimukselle nostamalla edellä mainittuja asioita esiin ja toteamalla sen, että joiltakin osin tietojärjestelmätieteen perinteinen tutkimusote on sovellettavissa myös analytiikkaan. (Bischoff et al., 2015).

## 5 Asiakslähtöisyys

Tässä luvussa esitellään asiakslähtöisyys käsitteenä ja miten se peilautuu tutkielman teoreettiseen viitekehykseen. Asiakslähtöisyys on tässä tutkielmassa enemmänkin näkökulma empiriseen osioon, joten teoriaa käsitellään työn vaatimassa laajuudessa.

Asiakslähtöisyys määritellään tutkimuskirjallisuudessa organisaation taidoksi ottaa asiakkaat ja heidän tarpeensa huomioon ennakoiden ja reagoida niiden muutoksiin, oppia muutoksista ja tehdä niiden perusteella tarpeellisia muutoksia. Asiakslähtöisyyden tavoitteena on luoda arvoa asiakkaalle tuotteiden tai palveluiden avulla (Saarijärvi & Puustinen, 2020). Esimerkkinä asiakslähtöisestä toiminnasta on se, että organisaatio kehittää tuotteitaan yhteistyössä asiakkaiden kanssa, jolloin asiakkaiden tarpeet voidaan ottaa huomioon jo kehitysvaiheessa (Gummesson, 2008).

Asiakslähtöisyys terminä sisältää jo itsessään useita eri lähestymistapoja, ideoita, toimintatapoja ja strategioita, jotka ovat kehittyneet ajan kuluessa eri toimialojen pyrkimyksissä tuottaa asiakkaille arvoa. Kaikkien näiden eri tapojen yhteisenä tekijänä on se, että asiakslähtöisyyttä harjoittavan organisaation toiminta pyrkii lisäämään asiakaspysyvyyttä, vähentämään asiakaskatoa ja kilpailijalle vaihtoa ja nostamaan liikevaihtoa (Loshin & Reifer, 2013). Alan tutkimuksissa organisaation kykeneväisyys olla aidosti asiakslähtöinen on yhdistetty parempaan tulokseen liiketoiminnallisilla mittareilla (Saarijärvi & Puustinen, 2020), ja sitä on pidetty yhdysvaltalaisessa markkinointikulttuurissa organisaation hengissä säilymisen ja kannattavuuden lähtökohtana (Gummesson, 2008). Organisaatioiden kyky olla aidosti asiakslähtöinen on kuitenkin ollut kyseenalainen ja asiakslähtöisyyden vieminen käytäntöön on ollut hankalaa siihen liittyvien strategioiden ja työkalujen ollessa puutteellisia, joka on johtanut vajaisiin toteutuksiin (Loshin & Reifer, 2013; Saarijärvi & Puustinen, 2020).

Jo aiemmin mainittu tuotekehitys yhteistyössä asiakkaan kanssa on esimerkki siitä, kuinka organisaatio on asiakslähtöinen ja luo yhdessä arvoa (value co-creation) asiakkaan kanssa. Yhteistyöllä asiakas voidaan sitoa osaksi organisaation prosessia ja ”siirtää” passiivisesta asiakkaasta, aktiiviseksi osaksi organisaation resursseja, kuitenkin muistaen sen, että prosessin tarkoituksena luoda asiakkaalle lisäarvoa ja sitä kautta myös tuotetta tai palvelua tarjoavalle organisaatiolle (Saarijärvi, 2012). Arvon luominen yhdessä asiakkaan kanssa on keskiössä tämänkin opinnäytetyön empirisessä osiossa, jossa asiakkaan edustajia haastatellaan analytiikkatuen ilmiön paremmin ymmärtämiseksi.

Asiakaslähtöisyyden vastapuolena voidaan nähdä tuotelähtöinen organisaatio, jonka toimintaa ohjaavat tuotteet tai palvelut ja organisaation omat käytännöt. Tässä tilanteessa asiakkaat mukautuvat tarjontaan ja tuotelähtöinen organisaatio ei pyri toiminnallaan ymmärtämään jo olemassa olevia asiakkaitaan, vaan löytämään uusia asiakkaita, siinä missä asiakaslähtöinen pyrkii ymmärtämään jo olemassa olevia asiakkaitaan ja parantamaan arvontuotantoansa heille. Kummallakin lähtökohdalla on omat vahvuutensa ja tuotelähtöinen organisaatio kykenee tehostamaan ja yhdenmukaistamaan toimintaansa, kun taas asiakaslähtöinen lähestymistapa tarjoaa kilpailuetua tilanteessa, jossa organisaatioiden välinen kilpailu on kovempaa ja lähempänä asiakasta. (Saarijärvi & Puustinen, 2020)

## 6 Tutkimuksen tulokset

Tutkimus toteutettiin toimeksiantona IT-konsulttifirmalle, jonka tarkoituksena oli selvittää, miten sen asiakkaiden analytiikkatuki on järjestetty, mitä haasteita analytiikan tukemisessa on ja miten analytiikkaa voitaisiin tukea paremmin. Tutkimuksen aineisto kerättiin puolistrukturoiduilla haastattelulla valituilta toimeksiantajan asiakkaiden analytiikan ammattilaisilta. Tulokset on jäsennetty kysymyskohtaisesti, joiden sisällä on tehty tarvittaessa vastausten luokittelua. Tuloksia peilataan olemassa olevaan teoriaan luvussa seitsemän.

Ensimmäisenä tässä luvussa käsitellään vastaajien taustatietoja organisaation kannalta ja analytiikan sijoittumista organisaatiossa. Toisessa kysymyksessä selvitetään analytiikan elinkaarta ja siihen liittyviä käytäntöjä. Kolmannessa kysymyksessä käydään läpi sitä, miten analytiikkaa tuetaan ja miten tuki on järjestetty. Neljäs kysymys käsittelee tuen järjestämisen haasteita ja parannusehdotuksia. Viidennessä kohdassa selvitetään sitä, miten haastateltavat pitävät analytiikkaratkaisunsa ajankohtaisena ja toimintakuntoisena, ja kuudennessa kysymyksessä avataan ajankohtaisena pitämisen haasteita. Viimeisessä varsinaisessa kysymyksessä käsitellään sitä, miten analytiikan liiketoimintahyötyjä arvioidaan. Liiketoimintahyötyjen arvioinnin jälkeen kootaan muut mahdolliset haastatteluissa analytiikkatukeen liittyvät huomiot.

### 6.1 Taustatiedot ja analytiikan sijoittuminen

Haastattelujen aluksi osallistujilta kerättiin taustatietoja siitä, miten analytiikka on sijoittunut yritykseen ja millainen tausta haastateltavilla on analytiikan parissa. Taustatiedot tarjoavat lisäulottuvuutta muiden tulosten tulkintaan.

Kaikki haastateltavat henkilöt olivat töissä noin 5000 työntekijän suomalaisissa yrityksissä, jotka toimivat eri toimialoilla, prosessiteollisuudessa, rakennusteollisuudessa ja yritysten välisessä palveluteollisuudessa. Kaksi haastateltavista oli johtavassa asemassa ja yksi haastateltavista oli pitkällisellä asiantuntijauralla. Työkokemukset analytiikan parissa vaihtelivat seitsemän ja 16 vuoden välillä. Haastateltavien koulutukset poikkesivat keskenään hieman, kahden ollessa ekonomeja ja kolmannen ollessa diplomi-insinööri oman yrityksensä toimialalta. Koulutusjakaumasta ei voi tehdä näin pienellä otoksella, yleistyksiä, mutta oli mielenkiintoista huomata, että mukana ei ollut IT-alan koulutuksen saaneita, tai tiedolla johtamaan erikoistuneita tietojohdajia. Haastatteluista kävi kuitenkin ilmi se, että koettiin henkilökohtaisesti tärkeäksi olla linkkinä liiketoiminnan ja IT:n välillä ja, että kaksi haastateltavista ilmaisi hallitsevansa BI-työkalujen kehittäjäkäytön syvästi.

Analytiikan sijoittumisessa oli myös eroja, kahdessa organisaatiossa analytiikka sijoittui suoraan talouden alle, joista toisessa tiimi oli liikkunut jo kerran edestakaisin IT:n ja talouden välillä. Kolmannessa organisaatiossa analytiikka oli sijoittunut IT:n alle, mutta IT konsernin tukitoimintona sijaitti myös talouden alla. Voisikin olla mielenkiintoista tutkia tulevaisuudessa sitä, että onko analytiikan sijoittumisella merkitystä sen toteutumisen kannalta. Kirjallisuudessa kun painotetaan kovasti sen merkitystä IT:n ja liiketoiminnan välissä.

## 6.2 Elinkaari

Haastattelujen pohjalta voidaan todeta, että haastateltavat käsittelivät analytiikan elinkaarta erilaisin mallein. H3:sen analytiikkaorganisaatiossa oli käytössä oma yhdeksän vaiheinen malli, joka alkoi ketterästä ohjelmistokehityksestä tutulla ”DoR”-määrittelyllä (definition of ready), jolla käytännössä määriteltiin analytiikkatuotteen vaatimukset ja sen tärkeys ilmeneekin alla olevassa lainauksessa.

”Ilman DoR:ia (definition of ready) ei lähdetä tekemään mitään. DoR vastaa kysymykseen, milloin tuote on valmis mitä me tehdään - - Me on nähty niin monta kertaa, että speksi muuttuu matkan varrella ja se ei ole hyvä” – H3

H1:n organisaatiossa elinkaaren alku muodostui myös liiketoimintatarpeilla, ja vaatimukset tulivat yleensä nimetyiltä liiketoiminnan edustajilta tai muita epävirallisempia reittejä pitkin ennen kuin ne otettiin sprinttimalliseen kehitykseen. Sprintti tarkoittaa iteratiivista kehitystä, jossa tehdään muutaman viikon sykleillä tuotteeseen valmiita ominaisuuksia (Schwaber, 2023). Toisen haastateltavan organisaatiossa oli meneillään tilanne, että uusia analytiikkaratkaisuja ei tuotettu vaan keskityttiin tulevaan toiminnanohjausjärjestelmän (ERP) uusimiseen ja nykyistä analytiikkakatalogia pyrittiin pitämään yllä, liiketoiminnan jatkuvuuden varmistamiseksi. H2:sen organisaation haastattelussa tuli kuitenkin esille se, että ennen nykyistä tilannetta analytiikkaa oli rakennettu yhteistyössä IT:n, liiketoiminnan ja analytiikan tekijöiden kanssa viikoittaisten tapaamisten avulla, jossa oli saatu ratkaistua nopeastikin mahdollisia ongelmia ilman isompaa byrokratiaa ja prosessin aikana oli saatu automatisoitua BI-työkalun avulla useita satoja Excel-taulukko-työkirjoja. H2 muisti kuitenkin tapauksia, kun uutta analytiikkaa oli kehitetty katalogiin, joissa analytiikkaa oli kehitetty organisaatiossa äänekkäiden ihmisten toiveista heikommilla määrittelyillä ja varsinainen tuotos oli jäänyt käyttämättä, joka vahvistanee H3 väitettä määritysten tärkeydestä.

Kehitysprosesseissa H1 nosti esille jatkuvan kommunikaation liiketoiminnan kanssa, jotta tuote saadaan iteroitua sopivalle tasolle, kehittäjän aikaa ei mene hukkaan ja että



tuotteesta tulee varmasti toimiva. H3:n organisaation mallista löytyy myös validointi liiketoiminnan kanssa iteroiden, mutta H3 nosti esille erityisesti sen, että analytiikkaa tehdessä tulee ymmärtää datan luonne:

”Kaikkein kriittisintä on se mihin tarvitaan kovaa osaamista, että ymmärtää datan luonteen. Paljonko sitä dataa kertyy? - - Se toimii kolme päivää, mutta sitten kun pitäisi mennä vielä kolme vuotta eteenpäin” – H3

Datan määrällä on siis merkitystä analytiikan toiminnan kannalta ja tätä H3 peräänkuulluttikin haastattelussa, että taitava analytiikan ammattilainen näkee ratkaisussa pidemmälle ja osaa arvioida sen toimintaa pitkällä aikajänteellä ja ymmärtää voiko se edes toimia käytettävästi vuosienkin päästä nykyisilläkin työkaluilla.

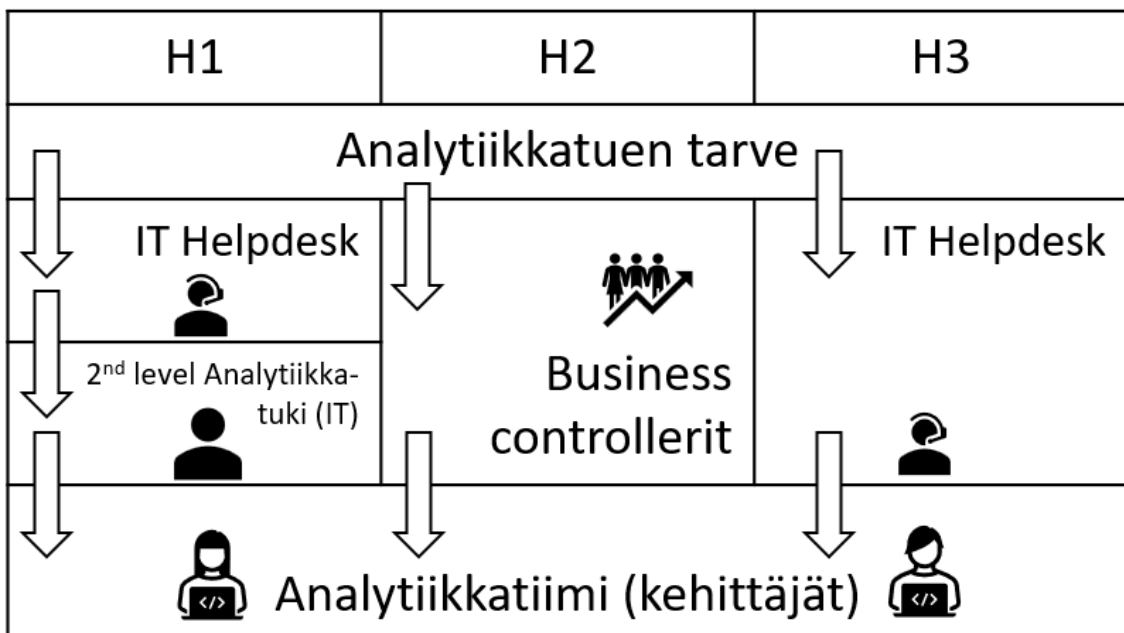
Käyttäjien kouluttamisen suhteen kaikki haastateltavat olivat samoilla linjoilla, että se kuuluu analytiikan elinkaareen. H1 nosti tässä vaiheessa liiketoiminnan vastuuta siinä, että sopivat ihmiset saadaan paikalle analytiikkaratkaisun esittelyä varten, jotta heille voidaan demonstroida sitä, miten analytiikka toimii, eli miten datalähteet päivittyvät, mistä data tulee, onko jotain mitä tiettyyn analytiikkaratkaisuun liittyvää erityistä, joka käyttäjän pitää tietää. H2 viittasi samanlaiseen prosessiin, jossa pyrittiin hakemaan sopivaa lähestymistapaa ratkaisun loppukäyttäjien määrän ja ratkaisun kompleksisuuden mukaan. Pienen korjauksen kommunikointi sähköpostilla koettiin hyväksi tavaksi ja laajemman kokonaisuuden ymmärtämiseksi ratkaisun demonstrointi ja käyttäjien kysymyksiin vastailu paikan päällä tai etäyhteyksien päästä koettiin hyväksi. Haasteena tuotantoon viennin yhteydessä oli kaikkien haastateltavien mukaan se, että tarvittavat loppukäyttäjät saadaan paikalle. H2 täydensi vielä loppuun onnistuneesta loppukäyttäjäkoulutuksesta:

”Pystytään sitten siinä loppuvaiheessa tuotantoon mennessä, niin kuin tarjoamaan mitä he kysyvät. Ja, että he pystyvät samaistumaan ja näkemään, että 3 kuukautta sitten minä kysyin teiltä tätä, ja mitä tuli putkesta ulos? Ja jos on niin kuin luvattu, niin sitten se on menestyksekkäs” – H2

Dokumentointi kuului myös oleellisena osana jokaisen haastateltavan analytiikan elinkaarta, mutta siinä koettiin olevan puutteita tai ongelmia. Koettiin, että dokumentaatio oli hajaantunut analytiikan eri järjestelmien osiin, osa dokumentaatiosta oli BI-työkalujen sisällä, osa datalähteissä, osa Word-dokumenteissa, sähköposteissa.

### 6.3 Tukeminen

Analytiikan tukemisen suhteen haastateltavista löytyi samankaltaisuuksia ja eroja. Kaikkien haastateltavien analytiikan kehittäjätiimit olivat karkeasti samankokoisia (noin viisi henkeä) ja viime kädessä analytiikan tukeminen oli kehittäjätiimien vastuulla. Siinä kuitenkin oli eroja, miten tuketilanteet etenivät haastateltavien organisaatioissa. Kuten alla olevassa kuvassa näkyy, niin kaikkien haastateltavien viralliset tukipyynnöt etenivät eri tavoin. H1 ja H3 organisaatioissa viralliset tukipyynnöt etenivät tikettijärjestelmän kautta, jossa kaikki tiketit käsiteltiin IT-vetoisen helpdeskin kautta. H3:n organisaatioissa tukipyynnöt etenivät suoraan helpdeskistä kehittäjille ja helpdeskin funktio oli lähinnä varmistaa pyyntöjen kohdistuminen oikealle työkalulle (esim. Power BI tai Tableau). Tikeitin pystyi kohdistamaan myös analytiikkaratkaisukohtaisesti pienelle osalle analytiikkaratkaisusta. H2:n organisaatioissa helpdeskin rooli oli samankaltainen eteenpäin ohjaava, mutta ennen kehitystiimiä organisaatioissa oli toisen tason tukiporras nimettynä analytiikkatyökaluille. Toisen tason tehtävänä oli ratkaista suurin osa työpyynnöistä (mm. pääsyt ja työkalutekniset ongelmat) ja ohjata sovitut asiat eteenpäin kehittäjille ja pyytää apua tarvittaessa vaativampien työpyyntöjen kanssa. Toisen tason tuki sijaitsi organisaation IT:ssä, kun taas analytiikkatiimi toimi talouden alla.



Kuva 7. Analytiikkatuen järjestäminen haastateltavien organisaatioissa.

H2:n organisaatioissa analytiikan tuki oli organisoitu erilaisista lähtökohdista. Analytiikan tukipyynnöt tulivat pääsääntöisesti liiketoimintayksikön business controllerille, jonka tehtävänä oli validoida tuki tai kehityspyyntöjen järkeenkäyppöys ottamatta kantaa tekniseen toteutukseen tai taustaan. Business controllerin eduiksi H2 luettelee sen, että tuntee

oman yksikkönsä ihmiset, osaa ottaa kantaa pääsyihin analytiikan suhteen, tuntee substanssin ja organisaatio on oppinut kysymään controllerilta apua. Alla olevassa lainauksessa H2 korostaa analytiikkatuen sijoittumista talouteen. Jos controlleri ei saa ratkaistua ongelmaa, niin pyyntö ohjautuu työpyynnöksi analytiikkatiimille.

”Kiehtoo niin kuin talousorganisaatiossa tämä asia, että se niin kuin toimit ikään kuin vähän välimiehenä, mutta kuitenkin sinä ymmärrät sen IT-maailman ja samalla sinä ymmärrät business controllerina sen loppukäyttäjän vaateen ja sen ison kuvan ja sitten siinä pystyt oikomaan” – H2

H1:n ja H2:n organisaatioissa käyttäjille oli tarjolla koottu hakemisto erilaisissa muodoissa, joista sai tukea analytiikkatyökalun käyttöön. H1:n organisaatiossa aika ajoittain uusille työntekijöille pidettiin koulutus ”Tableaun saloihin”, kun taas H3:n organisaation puolella analytiikan käytön koulutus nähtiin tässä tilanteessa enemmän liiketoiminnan tehtävänä resurssien ja olemassa olevien analytiikkatuotteiden määrän vuoksi.

Eroja löytyi myös siitä, että mikä koettiin yleisimmäksi tukipyynnöksi. H1:n organisaatiossa yleisin kysymys oli:

”Kysymys oikeastaan onkin, että mistä minä löytäisin tällaista dataa? Onko sellaista olemassa? Sen tyyppistä ihmettelyä ja sitten sitä yhdessä mietitään, että onko meillä jotain (analytiikkatuotetta) millä me saataisiin niin kuin vastattua sitten tuohon kysymykseen?” – H1

H1:n organisaatiosta on toki muistettava, iso osa työpyynnöistä jää toisen tason analytiikkatuotteelle ja kysymykset suoraan loppukäyttäjiltä tulevat liiketoiminnan edustajilta, joko virallisia tai epävirallisia reittejä pitkin. H2:n organisaatiossa yleisin työpyyntö on päästä käsiksi tiettyyn analytiikkatuotteeseen ja työpyyntöjä analytiikasta tulee ylipäätään melko vähän. H3:n organisaatiossa tukipyynnöt riippuvat siitä, mistä päin organisaatiota ne tulevat, esimerkkinä HR-yksikön henkilöt tuntevat datansa melko hyvin, kun taas toisaalla organisaatiossa kysymykset liittyvät liiketoimintasääntöjen ymmärtämiseen (tai ymmärtämättömyyteen). Yhteistä kaikilla haastateltavilla oli kuitenkin se, että työkaluista ja analytiikkanäkymistä oli pyritty tekemään niin helppoja ja itsensä selittäviä, että käyttäjien olisi mahdollisimman helppo käyttää ja ymmärtää niitä. Helppokäyttöisyyttä voi tukea esimerkiksi Tableaun tai Power BI:n sisäisillä infosuilla tai vihjeteksteillä (tooltip).

H1 näki tukemisen osaksi sen myös, että käyttäjille, jotka haluavat käyttää analytiikkaa edistyneemmin tarjotaan työkalut, järjestelmät ja lisenssit kuntoon datan käyttöä varten. Näissä tilanteissa datatiimi toimi myös sparrauskumppanina edistyneemmille käyttäjille

ja saattoi auttaa heitä suorittamaan omia analyysejään, kuten alla olevassa lainauksessa todetaankin:

”Koska se on kuitenkin koko organisaation etu mitä enemmän analytiikkaa pystytään hyödyntämään ja mitä enemmän sitten meidän datavarastossamme olevia tietoja saadaan käännettyä sille kulmalle, että niistä on liiketoiminnalle hyötyä ja sillä tavalla me tehdään tämmöistä (tuetaan edistyneempiä käyttäjiä)” – H1

Haastatellut olivat samaa mieltä siitä, että analytiikka on vielä tekijänsä näköistä, ja kaikki käytänteet eivät ole vakiintuneet tiimin kesken, jolloin kun kyseiseen tuotteeseen liittyvää työpyyntöä tulee, niin työpyynnön nopein ratkaisu tapahtuu välittämällä se analytiikka-tuotteen kehittäneelle taholle. H2 arvioikin, että tehokas tapa on, että kehittäjä tutkii mahdolliset ad-hoc-tarpeet ja muokkaukset ilman, että toisen kehittäjän täytyy syventyä raportin rakenteeseen. H2:n väite on linjassa sen kanssa, että aikaisemmin todettiin, että ratkaisujen dokumentointi on hajautunut eri puolille analytiikan kokonaisprosessia ja sen omaksuminen voi viedä aikansa.

#### **6.4 Tukemisen haasteet**

Tukemisen haasteista löytyi yhteneväisyyksiä haastateltavien välillä. H1 ja H3 korostivat sitä, että analytiikan tukemisessa ei ole ensiarvoisen tärkeää, että kaikkea liiketoiminnasta ymmärretään vielä ja että liiketoimintalogiikan voi omaksua analytiikkaratkaisua tutkiessa ja ajan kanssa. Liiketoimintalogiikan syvempi osaaminen oli heidän mielestään varattu uusien ratkaisujen kehittämiseen kuten H3 sen muotoilee:

”Tuen piiristä se on jollain tavalla vielä OK, että sinun ei tarvitse ymmärtää kaikkea kaikkea, vaan että vähän niin kuin pääsee opiskelemaan sillä raportilla. Mutta jos pitäisi tehdä devia sen suhteen niin sitten pitää ymmärtää aika paljon siitä asiasta, kun joku tulee kysymään” – H3

Yhteisiä haasteita oli myös tuen pysyvyys. H1:n organisaatiossa koettiin, että jos tuki on esimerkiksi järjestettyä ulkoiselta toimijalta, niin pelkkä tekninen kyvykkyys ei riitä, vaan tekijän tai tiimin pitää olla pysyvä, jotta ajan kanssa pääsee sisälle liiketoimintalogiikan toimintaan. H2:n ja H3:n organisaatioissa haasteena oli myös pysyvyys, joka näkyi analytiikan tekijöiden vaihtumisella, joka liittyy edellisessä kappaleessa esitettyyn tulokseen siitä, että analytiikka on tekijänsä näköistä. Tekijöiden vaihtuessa analytiikkatuotteisiin tulee enemmän ”tekijänsä näköistä jälkeä”.

H1 toi esille sen, että ideaalitulanteessa kehittäjät eivät olisi se taho, joka tarjoaa tukea, vaan se, että analytiikkatiimissä olisi erikseen roolitettu kehittäjät ja analytiikkatukea tarjoavat tahot. Nykyisessä mallissa kuitenkin, kun kehittäjät tarjoavat viimekädessä tuen, haasteeksi esitettiin se, että tuen pyytäminen pahimmillaan katkaisee kehitystyöputken. H1 peräänkuuluttikin selkeää mallia, kuinka tukipyynnöt etenevät kehittäjille. H2 ja H3 olivat enemmän sen kannalla, että paras tuki löytyy kehittäjiltä ja että työpyynnöt eivät työllistä kehittäjiä erityisen paljon:

”Joku sovellustuki niin siellä kehitetään ehkä 10–20% ajasta, 80-90% menee tukemiseen, niin se on täysin toisin päin meillä eli valtaosa ajasta menee siihen uuden kehittämiseen ja hyvin pieni prosentti menee siihen niin kuin sen opsin, servicen pyörittämiseen tai tukemiseen niin siksi minun mielestä devops-tiimi on ihan loistava kyllä.” – H3

Dokumentoinnin osalta kaikilla haastatelluilla oli haasteita. Dokumentointi oli osittain hyvää, mutta osittain hajautunutta, ohutta, ja siihen ei ollut täysin vakiintuneita käytänteitä. Dokumentoinnin ydinhaasteena oli se, että sitä oli ”dataputken” eri osissa. Esimerkiksi dokumentointia voi olla lähdejärjestelmän ja tietovaraston välisisissä ratkaisuisissa, tietovarastossa, BI-työkalussa (Tableau tai Power BI), sähköposteissa, dokumentaatioalustalla tai verkkolevyllä. Koettiin myös, että loppukäyttäjille oli haasteellista pitää ajankohtaista dokumentaatiota yllä, sillä ratkaisujen muutosten auki selittäminen erilliseen hakemistoon koettiin raskaaksi. Alla H1 ja H3 kuvailevat dokumentoinnin ongelmaa, ja sen alla H2 kuvailee ratkaisua ongelmaan.

”Meillä nyt on datatiimi ja ollaan niin kuin kehittynyt organisaatio jollain tavoilla, niin me on oltu niin pieniä ja ketteriä tässä kokonaisuudessa, että on vähän ehkä hypätty yli sen, että olisi mietitty dokumentoinnin tämmöiset periaatteet ja pääpiirteet. Missä muodossa dokumentoinnin pitää olla ja mitkä asiat tulee olla mihinkin. Eli meiltä puuttuu semmoinen dokumentoinnin viitekehys.” – H1

”Kun se sehän tapahtuu niin monessa paikassa; sitä on siellä lähdejärjestelmässä, sitä on siellä matkan varrella snowflakessa, sitten raportointityökaluissa ja ehkä jossain erillisessä sharepoint-listassa mistä haetaan väliin dataa. Se on niin hirveän vaikea hahmottaa, että mihin se pitäisi kirjoittaa. Ja me tehtiin sitä muutaman kerran ihan Wordiin ja PowerPointtiin otettiin vaikka videotakin, mutta me huomattiin se, kun mennään kaksi kuukautta eteenpäin, se on vanhentunut. Se on ihan hirveän vaikea ymmärtää, että mille tasolle sitä dokumentointia pitäisi tehdä.” – H3

”Hyvä dashboard on sellainen, että otat Matti Meikäläisen kadulta ja se intuitiivisesti näkee heti mihin kysymykseen tässä vastataan” – H2

Tuen toimittamisen haasteiksi H2 nostaa myös sen, että sopivia tekijöitä ei löydy aina helposti. Tekijän tulisi olla jotakin matemaattis-IT:n henkisen ja liiketoimintahenkisen väliltä, valmiina menemään epämukavuusalueelle oli se sitten IT:n tai liiketoiminnan puolella. Huomio on mielenkiintoinen sillä, kukaan haastatelluista ei varsinaisesti ole IT-taustainen eli voidaanko olettaa, että askel IT:n suuntaan on helpommin otettavissa?

Käyttäjien koulutuksen haasteina todettiin se, että datatiimin resurssit eivät riitä omaaloitteisesti kouluttamaan kaikkia käyttäjiä tai kokoamaan loppukäyttäjiä kokoon. Koulutusten järjestämiseen kaivattiin liiketoiminnan edustajan apua ja mielenkiintoa sen järjestämiseen. Koulutusta oli yleensä uusien tuotteiden käyttöönotossa, mutta koulutusten tarvetta pyrittiin vähentämään sillä, että analytiikkatuotteet olivat helppokäyttöisiä ja ymmärrettäviä.

Muina haasteina koettiin myös H3:n haastattelussa, että käyttäjät eivät ymmärrä analytiikkakokonaisuuksia tai työkalujen mahdollisuuksia ja se kuormittaa analytiikkatiimiä. Organisaation matala ymmärrys analytiikan suhteen lisää alkeellisten kysymysten määrää analytiikkatiimille. Toinen haaste, joka koettiin, oli H1:n ja H3:n haastatteluissa esille tullut analytiikan lähdejärjestelmien ongelmat (esim. ERP), jotka tulevat esille analytiikan loppukerroksessa eli BI-työkaluissa ja on haasteellista lähteä tutkimaan koko analytiikkaprosessia lopulta kuitenkin tullen lopputulokseen, että ongelma ei ole datatiimin vastuualueella. H3 korosti myös sitä, että monitoroinnin tekeminen on ollut haasteellista analytiikkaprosessin alusta loppuun ja sen seuraaminen, että missä vika ilmenee automaattisesti ei ole ollut helppoa ja siinä riittää vielä työtä:

”... dataputket, että ongelma näyttäytyy aina siellä loppupäässä, kun tulee Power BI tai Tableau tikettejä, että data on väärin niin, sitten siellä on joku lataus kaatunut. Se on semmoinen mikä meillä on isoin haaste ja on myös mikä on minusta tosi vaikeata ollut, sen monitoroinnin tekeminen ja automatisointi. On ollut niin kun koko sille end-to-end putkelle ollut ihan mahdotonta” – H3

## 6.5 Ajankohtaisuus

Analytiikkaratkaisujen ajantasaisuuden varmistamisen peruskäytänteet olivat kaikkien haastateltavien organisaatioissa suhteellisen samanlaiset. Valmiista ratkaisuista oli pyritty tekemään sellaisia, että esimerkiksi vuoden vaihtuminen ei riko ratkaisuja eikä mitään kovakoodattuja arvoja käytetä. Ylipäättään ratkaisuista oli pyritty saamaan sellaisia, että ne kestävät ajankulua ilman kehittäjien puuttumista, kuten H3 alla olevassa lainauksessa toteaa:

”Yhdessäkään ratkaisussa ei saa olla tiettyä manuaalisuutta eli kaikki pitää olla niin automaattisia, kuin vaan pystyy. Just tämän takia, että jos täppäät sinne jonkun yksittäisen vuosiluvun, joudut aina täppäämään sen sinne. Se on hölmöä. Ne pitää tehdä jollain muulla tavalla. Yritetään olla kierrättämättä teknistä velkaa mahdollisimman pitkään. Teknisen velan kerryttäminen ei aluksi tunnu pahalta, mutta parin vuoden päästä sinä teet pelkästään teknistä velkaa ja liiketoiminta kyseenalaistaa, miksi te olette näin epätehokkaita?” – H3

Ajankohtaisuuden ylläpitämisessä jokainen haastateltava korosti liiketoiminnan ja analytiikkatiimin kommunikaation tärkeyttä ja sitä, että mahdollisten liiketoimintaprosessien tai lähdejärjestelmien muutosten pitäisi kulkeutua analytiikkatiimille ennen muutosten syntyä ja optimitilanteessa muutoksia ei korjata jäljestä päin. Haastateltavien välillä oli pientä hajontaa sen osalta, kuinka hyvin he arvioivat tiedon kulkeutuvan tiimille. H2 arvioi, että omassa organisaatiossaan ymmärrys lähdejärjestelmien muutosten vaikutuksista analytiikkaan oli hyvällä tasolla:

”Ennustettava businessprosessi-managementti ja sen indikaatioiden end-to-end ymmärtäminen tämä tietous koko ajan lisääntyy firmassa” – H2

Haastateltavien mukaan ajankohtaisuuden valvontaan ei ollut olemassa automaattisia mekanismeja vaan, niitä korjattiin tarvittaessa liiketoiminnan aloitteesta tai loppukäyttäjien palautteen perusteella. H2:n organisaatiossa business controllerit, jotka toimivat myös tuen ensimmäisenä linjana katsoivat analytiikkaratkaisujensa perään ja taloudellisen raportoinnin toiminnasta vastasi analytiikkatiimi luonnollisesti ollessaan talousorganisaation sisällä. H3:n organisaatiossa lähes jokaisella analytiikkatuotteella oli omistaja liiketoiminnan puolella, joiden tehtävänä oli varmistaa, että tuote on ajantasainen ja nostaa mahdolliset ongelmat kehitystiimin käsiteltäväksi. H1:n organisaatiossa vain osalla analytiikkakokonaisuuksista oli yhteyshenkilö liiketoiminnan puolelta, mutta viime kädessä vastuu oli analytiikkatiimillä:

”Kyllä se varmaan meidän tiimimme on tätä nykyään. Ei ole ehkä yhtä oikeaa vastausta, varmaan liiketoiminnasta osoitetaan meidän suuntaamme ja IT:stä saatetaan osoittaa meidän suuntaamme kanssa.” – H1

Yhteistä haastateltavilla oli se, että analytiikka nojasi ydintietojen päälle, mutta sen ylläpitäminen ei ollut analytiikan kehittäjien vastuulla. H3 haastattelussa mainitsikin, että ydintietoja pidetään yhtenä tietolähteenä muiden joukossa. Toinen yhteinen asia oli se, että ratkaisujen laatuun ja ajankohtaisuuteen liittyviä seikkoja nostettiin esille tarvittaessa liiketoiminnan tai loppukäyttäjien toimesta, mikäli niissä huomattiin ongelmia.

## 6.6 Ajankohtaisuuden haasteet

Ajankohtaisuuden haasteina haastatteluissa nähtiin analytiikkaratkaisujen suuri määrä. Haastateltavien organisaatioissa erilaisten analytiikkanäkymien määrä vaihteli 180–1000 näkymän välillä, joka on suhteutettuna tiimien kokoon suuri määrä hallita. Haastateltavat totesivat, että ei ole resurssien kannalta mahdollista valvoa käsin sellaista määrää raportteja, vaan siihen tarvitaan enemmän liiketoiminnan panosta. Kaikki haastateltavat viittasivatkin siihen, että optimitilanteessa kaikilla ratkaisuilla olisi liiketoiminnan omistaja tai muu vastaava yhteyshenkilö, joka varmistaisi ratkaisujen ajantasaisuuden ja ymmärtäisi tiimin avustuksella syvemmin analytiikkaa, kuten H2 alla olevassa lainauksessa toteaa. H3 organisaatiossa useimmilla ratkaisuilla oli omistaja, mutta ihmisten vaihtaessa työpaikkaa tai toimenkuvaa jäi usein analytiikkatiimin kontrolle etsiä uusi omistaja tai patistaa liiketoimintaa tekemään sitä.

”Jokaisella raportilla pitäisi olla omistaja, myös se business-omistaja. - - Parannuskohta meille, että jos sinulla on se 800 raporttia, niin kuka on se Petri, Jaana tai Heikki, joka vastaa siitä, että raporttia käytetään. - - Ja, että heillä on jonkinlainen säännöllinen kädensija siihen (raporttiin) ja varmistavat, että se toimii ja kaikki on kunnossa” – H2

H1 mainitsee ajankohtaisuuden haasteeksi sen, että vanhojen ratkaisujen ylläpitäminen usein jää taka-alalle verrattuna uusien ratkaisujen kehittämistä, kun työtä priorisoidaan olemassa oleville resursseille. Priorisoinnin seurauksena kehittäjillä ei ole aikaa tarkastella olemassa olevia ratkaisuja niiden ajankohtaisuuden seuraamiseksi.

”Meillä on joihinkin liiketoimintoihin tämmöinen suhde tavallaan, että sieltä löytyy tämmöinen vastapeluri meille, jonka kanssa onkin sitten helppo käydä näitä ja miettiä niitä uudistustarpeita, happenemisen pelkoa nähdä ja seurata just sitä, että mikä toimii ja mikä ei enää toimi” – H1



Ajankohtaisuuden haasteena oli myös yleisesti se, että liiketoiminta ei aina viesti muutoksista lähdejärjestelmissä tai liiketoimintalogiikassa, niin se usein heijastui analytiikkaratkaisujen laadun heikkenemisenä tai jopa rikkoutumisella. H2 koki, että heidän organisaatiossaan tätä haastetta esiintyi melko vähän hyvän liiketoimintalogiikan hallinnan vuoksi, mutta, myönsi kuitenkin, että kaikkia muutoksia ei saada etukäteen kiinni. H1 mainitsi yllä olevassa lainauksessa, että osittain muutoksia saadaan etukäteen tietoon, mutta kaikista liiketoimintayksiköistä ei löydy yhteyshenkilöä, jolloin vaivaton kommunikointi analytiikan ajankohtaisuudesta ei onnistu.

### **6.7 Liiketoimintahyötyjen arviointi**

Liiketoimintahyötyjen arviointi oli haastateltavien kesken hyvin samankaltaista. Sitä ei juurikaan tehty. Kukaan haastateltavista ei tunnistanut omasta organisaatiostaan, että analytiikkaa olisi arvioitu sen liiketoimintahyötyjen kannalta, kaikki tosin olivat sitä mieltä, että analytiikka on arvostettu osa liiketoimintojaan kuten alla olevat lainaukset vahvistavat:

”En välttämättä edusta koko firman näkemystä, mutta me ei suoraan sitä seurata minun mielestäni, mutta meillä ymmärretään sen oleellisuus ja tärkeys muuta kautta, kuin sen liiketoimintahyödyn kautta. - - On olemassa vakaa usko siihen, että kun meillä on analytiikka tarjolla ihmisille ja he pääsevät katsomaan näitä asioita, niin se tehostaa asioita niin paljon, että siitä on myös liiketoimintahyötyjä.” – H1

”Me käytetään paljon analytiikkaa ja sitä arvostetaan. - - Minä sanoisin, että ollaan aika lapsenkengissä tuon (liiketoimintahyötyjen mittaamisen) kanssa.” – H2

”Tuon tyyppistä työtä (analytiikan liiketoimintahyötyjen mittaaminen) ei tee meillä kukaan. Ei ole vaan aikaa tehdä sitä, mutta se olisi ihan älyttömän mielenkiintoista tehdä.” – H3

H3 oli kertaluonteisesti arvioinut yksittäisten jo tehtyjen analytiikkaratkaisujen onnistumisia peilaten niiden määrittelyyn DoR:issa (definition of ready), joissa määritellään se arvo, jota liiketoiminta tavoittelee analytiikkaratkaisulta ja mitä ongelmaa ratkaisulla pyritään ratkaisemaan. Tällä yksittäisellä arviointikierroksella selvisi, että osa ratkaisuista ei ole saavuttanut tavoiteltua arvoa, tai se oli määritelty väärin, tai analytiikkaratkaisu oli otettu työstöön väärin perustein, kuten alla olevassa lainauksessa todetaan:

”Aloin kysellemään ihmisiltä, että miten tämä raportti on loppupeleissä muuttanut sinun arkeasi? Olisinkohan minä kysellyt kahdeksasta eri raportista ja olikohan kolmessa, että se oli totta? Viidessä oli, että ei ole totta, arvo ei ollut realisoitunut. Oli tavallaan joko väärin määritelty tai ei oikeasti tiedetty mikä se KPI on tai oli valehdeltu, että saadaan se tuote työstöön.”  
– H3

Haastateltavat mainitsivat, että analytiikan rahamääräisen hyödyn mittaaminen on hankalaa, sillä analytiikan peilattaessa lähinnä menneisyyttä on hankala osoittaa konkreettisia pisteitä, joissa analytiikka tarjoaa suoraa rahallista hyötyä. Tulevaisuuteen sijoittuva ohjellinen analytiikka voisi esimerkiksi tarjota indikaattoreita hinnoitteluun, jolloin rahalliset vaikutukset olivat selkeämpiä. Lisäksi tulevaisuuteen sijoittuvalla analytiikalla olisi mahdollista optimoida erilaisia prosesseja, kuten tuotantoa ja logistiikkaa. Yhteistä oli kuitenkin se, että kaikki haastateltavat totesivat jossain vaiheessa haastatteluja, että analytiikalla on automatisoitu monia ennen manuaalisia toimenpiteitä, joten voidaankin johdattaa, että vähintäänkin työtunteja Excelin parissa on saatu vähennettyä, mutta sitä ei ole varsinaisesti mitattu.

## 6.8 Muut huomiot

Haastatteluiden lopuksi haastateltavilta kysyttiin, että miten analytiikan tukeminen on muuttunut aikojen saatossa. Organisaatiot olivat siitä samanlaisia, että nykyaikaiset BI-työkalut, kuten Tableau tai Power BI oli otettu käyttöön organisaatiossa 2010-luvun puolivälin paikkeilla, joten kokemusta niiden tukemisesta oli jo useita vuosia. H1 mainitsi, että ennen nykyjärjestelmiä karkeasti, saadaksesi tukea vanhakantaisen analytiikan käyttöön, sait sitä, jos tunsit sopivan henkilön talousosastolta. H1:n ja H3:n organisaatioissa yhteistä oli se, että analytiikkatiimi oli perustettu ja eriytynyt pikkuhiljaa omaksi erilliseksi toimijakseen. H3 kuvailikin tiimin alkuaikoja seuraavasti:

”Lähdettiin kovaa liikkeelle ja tuli ihan ylimmästä johdosta kova kiinnostus siihen. Ja sitten siinä tuli vähän turhautumista, kun se otti niin paljon aikaa. - - Vähän niin kuin talon rakentaa, niin perusta pitää olla hyvin valettu ja se vie paljon aikaa, mutta yhtäkkiä se runko tulee siihen jo ihan muutamassa viikossa.” – H3

H2:n organisaatiossa analytiikan tekeminen oli joitakin vuosia ison ”Excel-ähkyn” purkamista, joka tarkoitti erittäin suuren määrän manuaalisten raporttien käsittelyä tai analyysien automatisointia nykyaikaisilla BI-työkaluilla. Sen lisäksi H2 mainitsi, että organisaatio on viimeisen vuosien aikana oppinut ymmärtämään analytiikkaa paremmin kokonaisuutena ja erilaisten liiketoiminnallisten muutosten aiheuttamat mahdolliset ongelmat ymmärretään nykyään entistä paremmin.

## 7 Pohdinta

Tässä luvussa pohditaan tutkimuksen tuloksia suhteessa aikaisempaan tietoon ja tulosten merkittävyyttä analytiikan tukemiseen käytännössä. Tutkimuksen tuloksia peilataan erityisesti haastattelurungon taustateorialla eli analytiikan elinkaarimallin CRISP-DM:n (Chapman et al., 2000) laajennetulla versiolla (Cochran, 2019). Teorian jälkeen pohditaan sitä, miten tutkimuksen tulokset vastaavat tutkimuskysymyksiin. Tulosten käsittelyn jälkeen käydään läpi tutkimuksen luotettavuus ja rajoitteet. Lopuksi tässä luvussa esitetään ajatuksia aiheen tulevaisuuden tutkinnasta.

### 7.1 Tulokset suhteessa teoreettiseen viitekehykseen

Tutkimusta lähdettiin tekemään CRISP-DM:n laajennetun version tuotantoon viennin jälkeisten tehtävien pohjalta (Cochran, 2019). Laajennettu määritelmä käsitti seuraavat kohdat: **dokumentointi, laadun valvonta ja ajankohtaisuus, käyttäjien koulutus ja liiketoimintahyötyjen arviointi.**

CRISP-DM ja sen laajennettu versio määrittelee **dokumentoinnin** sellaiseksi, että sitä ylläpitävä taho osaa ylläpitää analytiikkaratkaisua ja pitää sen ajankohtaisena dokumentaation avulla. Haastatteluissa tuli kuitenkin esille se, että vain monimutkaisimmista ratkaisuista on olemassa laajempaa dokumentaatiota ja että dokumentointi on usein pirstaloitunut pitkin analytiikan eri vaiheita. Vaiheilla tarkoitetaan tässä vaiheessa esimerkiksi BI-työkalua, tietovarastoa tai dataputkea. Lisäksi Cochran (2019) esittää, että analytiikkaa ylläpitää usein eri taho kuin niiden alkuperäiset kehittäjät, ja silti haastatteluissa tuli ilmi, että näin ei ollut käytännössä yhdessäkään haastateltavan organisaatiossa, jolloin dokumentaatiota ei samalla tavoin ”annettu” eri ryhmälle. Haastateltavien organisaatiossa oli dokumentoitu analytiikan yleisiä periaatteita ja esimerkiksi H3:n organisaatiossa manuaalista työtä vaativat ratkaisut olivat listattuna, joka taas tukee hieman teorian dokumentointivaadetta. Dokumentoinnin tarvetta oli kuitenkin jokaisen haastateltavan mukaan pyritty vähentämään sillä, että analytiikkaratkaisuista oli tehty mahdollisimman helppoja käyttää, jolloin sekä kehittäjien, että loppukäyttäjien oli helppo omaksua ratkaisu. Dokumentointia oli vähennetty myös sillä, että usein mahdollisissa ongelmatilanteissa analytiikkaratkaisun korjasi sen kehittänyt henkilö. Voisikin sanoa, että dokumentoinnin osalta Cochranin (2019) malli ei täydellisesti kohtaa todellisuutta, vaan tuen ollessa sama taho kehittäjien kanssa, dokumentointi jää helposti alkeelliselle tasolle ja että dokumentoinnin keskittämisessä nähdään haasteita.

**Analytiikan laadun valvonta ja ajankohtaisuuden varmistaminen** CRISP-DM:n laajennetussa (Cochran, 2019) tarkoittaa sitä, että analytiikkaratkaisu pysyy luotettavana

ajan kuluessa ja että sen valvontaan on olemassa mekanismeja ja keinoja huomata esimerkiksi ydintietojen tai liiketoimintalogiikoiden muutokset ja niiden vaikutukset analytiikkaan. Haastateltavien kesken yhteistä oli se, että analytiikassa ei varsinaisesti ollut mekaanisia valvontamekanismeja, vaan ratkaisuna analytiikan laadun varmistamiseksi oli kolme tapaa: liiketoimintalähtöinen omistaja analytiikkaratkaisuille valvoi niiden ajankohtaisuutta, analytiikkatiimi huomaa ratkaisujen laadun heikkenemisen ja sen, että loppukäyttäjien palaute viestii ratkaisujen laadun heikkenemisestä. Haastateltavien kesken yhteistä oli se että, ideaalitulanteessa jokaisella analytiikkaratkaisulla olisi nimetty liiketoiminnan omistaja tai henkilö, joka varmistaisi substanssiosaamisellaan ratkaisujen hyvän laadun ja ajankohtaisuuden. Verrattuna siis Cochranin malliin (2019) ajankohtaisuuden ja laadun valvonnan malleista puuttuu tietty järjestelmällisyys ja haastatteluissa liiketoiminnan merkitys korostui.

**Käyttäjien koulutus** elinkaarimallissa (Cochran, 2019) tarkoittaa sitä, että analytiikkaratkaisun toiminta esitellään loppukäyttäjille ja, että käyttäjien toimintaa seurataan ratkaisun parissa. Teoriassa mainittiin, että ratkaisun koulutusvastuu on usein analytiikka- tai IT-tiimillä ja, että koulutuksen avulla vähennetään tulevaisuuden ylläpitokuormaa. Kaikki haastateltavat tunnistivat loppukäyttäjien koulutuksen osana analytiikan elinkaarimallia. Käytännössä kuitenkin koulutuksiin tarvittiin usein liiketoiminnan apua analytiikkatiimin rajallisen resurssimäärän vuoksi. Analytiikkaratkaisujen tuotantoon viennin yhteydessä haastatteluissa tunnistettiin tuotteen lanseeraustilaisuudet, joissa tärkeimmille käyttäjille pyrittiin opettamaan analytiikkaratkaisun käyttö. Tärkeimpänä menetelmänä kuitenkin tulevaisuuden ylläpitokuorman vähentämiseen koettiin se, että itse analytiikanäkymät on tehty niin helposti ymmärrettäväksi, että niiden käyttöön ei tarvita erityistä koulutusta. Koulutuksen osalta Cochranin elinkaarimalli (2019) kohtasi tutkimusten tulosten kanssa hyvin, jos ei oteta huomioon IT:n osuutta käyttäjien koulutuksessa, sillä varsinaista IT-vetoista koulutusta ei tullut esille missään haastattelussa. Koulutuksen arvo tunnistettiin kuitenkin kuten elinkaarimallissa.

**Liiketoimintahyötyjen arvioiminen** tarkoittaa sitä, että analytiikkaratkaisun hyötyjä arvioidaan sen jälkeen, kun ratkaisu on viety tuotantoon. Hyötyjen mittaaminen ei ole aina suoraviivaista ja voi olla hankalaa mitata analytiikan arvoa. Tutkimusten tulosten mukaan, hyötyjä ei mitata tai niitä mitataan hyvin vähän. Haastateltavien organisaatiot tunnustavat analytiikan arvon, mutta sitä ei ole mitattu esimerkiksi euromääräisesti. Yksi haastateltavista oli kertaluontoisesti tutkinut analytiikan arvolupauksia ja verrannut niitä toteutuneisiin arvoihin. Voidaankin siis todeta, että elinkaarimallin (Cochran, 2019) mukaista hyötyjen tarkastelua ei suoriteta ja elinkaarimallin tämä vaihe ei toteudu todellisuudessa.

Kaiken kaikkiaan CRISP-DM:n (Chapman et al., 2000) laajennetun mallin (Cochran, 2019) toteutuminen tutkimuksen tulosten mukaan on osittaista. Tutkimuksessa sivuttiin koko elinkaarta, joten tutkimuksessa saatiin myös suppea kuva koko elinkaaresta. Pääpainon ollessa kuitenkin erityisesti tuotantoon viennin jälkeiseen aikaan tämän perusteella tutkimuksen tulokset ovat linjassa sen kanssa, että vaikka CRISP-DM on elinkaarimalleista suosituin (Mariscal et al., 2010), niin iso osa analytiikkatiimeistä ei käytä mitään elinkaarimallia tai seuraa omaa menetelmäänsä (Hotz, 2023). H2:n organisaatiossa oli esimerkiksi kehitetty oma yhdeksänportainen malli, jonka yhtenä osana CRISP-DM oli. Voidaankin todeta, että vaikka CRISP-DM osat ovat sovellettavissa erialaisiin analytiikkaympäristöihin, niin ne eivät toteudu jokaisessa organisaatiossa tämän tutkimuksen tulosten mukaan kuten Cochran (2019) ne esittelee.

## 7.2 Vastaukset tutkimuskysymyksiin ja huomiot analytiikkatuen toimittamiseen

Tutkimuksen ensimmäinen tutkimuskysymys oli se, **miten analytiikka tuetaan?** Analytiikan tukemiseen oli erilaisia malleja ja jokaisen haastatellun organisaatiossa tuki oli järjestetty eri tavoin. Suurimpana erona mallien välillä oli se, että H2:n organisaatiossa business controllerit käsitelivät analytiikan tukipyynnöitä H1 ja H3 organisaatioiden IT-helpdeskin sijaan. Tuen toimittamisen määrää pyrittiin vähentämään sillä, että tehdyt ratkaisut olivat mahdollisimman helppokäyttöisiä. H2 ja H3 kokivat, että on hyvä malli, kun kehittäjät tarjoavat tukea, sillä analytiikka on kehittäjänsä näköistä ja kehittäjillä on hyvä tieto ratkaisujen toiminnasta. H1 kritisoi hieman kuitenkin analytiikan tukemisen roolitamista kehittäjille sen häiritessä kehitystyötä. Kaikissa organisaatiossa kuitenkin kehittäjät vastasivat analytiikan tukemisesta viime kädessä. Analytiikan tukemiseen lukeutui mukaan myös se, että analytiikkatiimit pyrkivät olemaan aktiivisessa kommunikaatiossa liiketoiminnan kanssa, jotta mahdolliset liiketoimintalogiikoiden muutokset eivät vaikuttaisi negatiivisesti analytiikkaratkaisujen toimintaan.

**Analytiikan tukemisen haasteita** oli sen pysyvyys. Jos tuen toimittaja vaihtuu usein, niin sen seurauksena liiketoimintaymmärrys on heikompa, joka johtaa siihen, että tuen on hankalampi ymmärtää raportin sisältö. Analytiikan sisältöjen määrä nousi myös esille haasteena, sillä haastateltavien organisaatiossa oli iso määrä erilaisia analytiikkanäkymiä verrattuna analytiikkatiimin kokoon. Ratkaisujen dokumentointi koettiin haasteena ja dokumentoinnin hajanaisuus aiheutti päänvaivaa sen sijoittamisen suhteen. Lisäksi haastateltavat kokivat, että dokumentointi ei ollut niin hyvällä tasolla kuin voisi olla. Haasteena koettiin myös se, että datatiimin resurssit ovat rajalliset ja aikaa esimerkiksi loppukäyttäjien kouluttamiseen ei ole. Resurssit koettiin myös haasteena sen suhteen, että datatiimillä ei ollut mahdollisuutta olla niin tilannetietoinen organisaationsa muutoksista, että olisivat voineet aktiivisesti korjata muutosten mukaan analytiikkaratkaisuja, vaan tähän kaivattiin

liiketoiminnan apua. Yhtenä haasteena mainittiin myös se, että uusien analytiikkaratkaisujen kehitys oli usein vanhan ylläpidon yläpuolella prioriteettilistalla.

**Analytiikan parempaan tukemiseen** tutkimuksessa nousi erilaisia asioita. H1 esitti, että kehittäjät eivät lähtökohtaisesti tukisi analytiikan käyttöä, vaan analytiikkatiimiin esiteltäisiin erillinen rooli tukemista varten. H2 ja H3 olivat kuitenkin sitä mieltä, että kehittäjät ovat oiva tuen lähde. H2:n mukaan kuitenkin kokeneimmat kehittäjät olisi hyvä varata uusien analytiikkakokonaisuuksien määrittelyyn syvemmän liiketoimintaymmärryksensä takia. Parempaan tukemiseen esitettiin myös sitä, että jokaisella analytiikkaratkaisulla on oma omistaja tai vastaava taho liiketoiminnan puolelta, jolloin ratkaisut pysyisivät ajantasaisina yhdessä analytiikkatiimin ja liiketoiminnan yhteistyöllä. Tämän lisäksi analytiikan liiketoimintahyötyjen arviointiin ja mittaamiseen osoitettiin mielenkiintoa ja H1 totesikin, että olisi erittäin mielenkiintoista arvioida sitä, mikäli olisi aikaa.

Muita huomioita tuen toimittamisesta on se, että Davenportin (2012) esittämä väite siitä, että analytiikan elinkaaren hallitsemiseen suositellaan käytettäväksi jotain muita kuin IT-lähtöisiä malleja vaikuttaa olevan pätevä tämän tutkimuksen pohjalta. Tutkimuksessa selvisi, että yhdellä organisaatiosta on käytössä oma malli ja liiketoiminnan merkitys on hyvin oleellinen niin analytiikan kehityksen kuin sen tukemisenkin aikana. Analytiikan tukemisen mallien osalta talouslähtöinen tukimalli vaikutti haastattelujen perusteella toimivan sulavimmin. Lisäksi vaikutti, että haastateltavien organisaatiot keskittyivät analytiikkassaan erityisesti raportintiin eli menneisyyteen suuntaavaan analytiikkaan (Liebowitz, 2021) ja tulevaisuuteen päin suuntaavaan analytiikan tukeminen ei ollut analytiikkatiimien kehittämien ratkaisujen keskiössä.

### 7.3 Tutkimuksen luotettavuus

Tutkimuksen teoriataustaksi valittiin CRISP-DM:n laajennettu malli (Cochran, 2019), joka oli perusteltua, sillä CRISP-DM (Chapman et al., 2000) oli analytiikan elinkaarimalleista selvästi suosituin alan ammattilaisten (Mariscal et al., 2010) kuin tieteen harjoittajien keskuudessa (Aho et al., 2020). Haastattelujen teoreettiseksi viitekehykseksi oli kuitenkin valittu Cochranin (2019) laajennettu malli, jota ei oltu arvioitu erikseen. Arvioin kuitenkin, että Cochran yhdisteli CRISP-DM ja CAP:n JTA-mallia (2023) onnistuneesti ja perustellusti. On kuitenkin mahdollista, että mallien yhdistämisessä on ongelmallisuutta, joka ei tullut ilmi tässä tutkimuksessa.

Tutkimuksen teoriatausta on sekoitus tieteellistä, ammatillista ja kaupallista kirjallisuutta. Alan luonteen takia tutkimuksen keskeisin käsite (CRISP-DM) on esimerkiksi joukon kaupallisia toimijoita kehittämä viitekehys. Vaikka mukana on esimerkiksi analytiikkaohjelmistojen dokumentaatiota lähteinä, niin kirjallisuudessa ja taustateoriassa on pyritty

pitämään sopiva tasapaino erilaisten kirjallisten lähteiden välillä. On kuitenkin mahdollista, että osa kirjallisuudesta on jäänyt löytämättä, mutta en usko, että erityisen relevanttia aiheesta on jäänyt tutkimatta, vaikka varsinaista kirjallisuutta tai tutkimusta olikin melko vaikea löytää aiheella analytiikan tukeminen tai sen jatkuva käyttö.

Tutkimuksien menetelmien valinta tehtiin yhdessä toimeksiantajan ja yliopiston ohjaajan avustuksella, joten uskoisin, että lähestymistapana tutkittavaan ongelmaan se oli sopiva. Haastattelujen avulla saatiin ja haettiin syvempää ymmärrystä siitä, miten analytiikkatuki on järjestetty. Muilla menetelmillä kuin asiantuntijahaastatteluilla voisi mahdollisesti saada laajempia otoskokoja helpommin ja enemmän yleistettävää tutkimustulosta, mutta koen, että tämä oli tähän tarkoitukseen sopiva menetelmä.

Tutkimuksen haastattelu toteutettiin kolmen haastateltavan kanssa ja vaikka opinnäytetöiden keskeisin asia erityisesti laadullisen tutkimuksen osalta ei ole otoskoko (Tuomi & Sarajärvi, 2009), niin kolme haastateltavaa ei ole erityisen suuri otoskoko. Tutkimuksen tuloksia ei voi siis yleistää, mutta tuloksia on kuitenkin mahdollista luokitella ja tulokset olivatkin suhteellisen monipuolisia otoskokoan nähden. Haastattelujen tuloksia tulkittaessa on otettava huomioon myös se, että haastattelijana tunsin yhden haastateltavista etukäteen ammatillisesta yhteydestä ja on mahdollista, että se vaikutti haastattelutilanteeseen. Lisäksi olin toimeksiantajayrityksen palkollisena haastattelemassa toimeksiantajan asiakkaita, joka voi vaikuttaa myös haastattelutilanteeseen. Haastatteluiden yhdenmuokaistamiseksi haastattelijana yritin kuitenkin luoda puoliavoimen rungon haastatteluille, joka pohjasi aiemmin mainittuun analytiikan elinkaaren malliin. Teorialähtöisellä haastattelulla yritin myös vähentää omia ammatillisia ennakkokäsityksiäni aiheesta. Ennen haastatteluja haastateltavien kanssa pyrittiin myös ”synkronoimaan” ammatillinen kieli ja konteksti esittämällä elinkaarimalli sekä toimeksiantajan vapaasti muotoiltu analytiikkaprojektin elinkaari. Itse haastattelutilanteet etenivät ilman suurempia ongelmia, mutta yksi tilanteista venyi pidemmäksi ja yksi piti lopettaa kiireisempänä. Onkin mahdollista, että paremmalla aikataulutamisella haastatteluista olisi ollut mahdollista saada laajempia vastauksia.

Tutkimuksen osalta pitää ottaa myös se huomioon, että haastateltavat valittiin ”käsini”, joka tarkoittaa sitä, että haastateltavien roolit ja organisaatiot sopivat tiettyyn profiiliin ja löydökset eivät välttämättä ole yhteneväisiä esimerkiksi pienempien tai suurempien organisaatioiden kanssa. Kaikki haastattelut tehtiin myös suomalaisissa yrityksissä.

#### 7.4 Tulevaisuuden tutkimus

Tulevaisuuden tutkimukselle tutkimuksessa nousi monia mielenkiintoisia asioita. Ensinnäkin olisi mielenkiintoista tutkia H2:n organisaatiossa esiintynyttä talouslähtöistä analytiikan tukemista. Olisi mielenkiintoista verrata asiakastytyvääsyyttä esimerkiksi IT-lähtöisen ja talouslähtöisen analytiikkatuen välillä. Löytyisikö talouslähtöisestä mallista joitakin ongelmia, joita IT-lähtöisessä mallissa ei ole? Lisäksi olisi mielenkiintoista tietää lisää siitä, miten analytiikkatiimin sijoittuminen vaikuttaa sen toimintaan vai onko sillä mitään väliä, onko tiimi talouden vai IT:n alla?

Lisäksi toinen mielenkiintoinen tulevaisuuden tutkimus voisi olla se, että miten analytiikkaa tulisi dokumentoida? Tutkimuksessa tuli ilmi, että dokumentointi on hajaantunutta ja siinä nähdään haasteita. Miten dokumentointi kannattaisi tehdä, että siitä olisi eniten hyötyä kehittäjille? Onko analytiikan parissa olemassa olevia sovitteja käytänteitä, jotka helpottaisivat eri tahon kehittämien ratkaisujen ylläpitoa? Kuinka liiketoimintaymmärrys näyttäytyy teknisten taitojen rinnalla?

CRISP-DM (Chapman et al., 2000) on jo yli 20-vuotta vanha malli, joten olisi mielenkiintoista myös saada selville, että onko mallia päivitetty tai onko tarjolla uusia malleja? ASUM-DM:n ilmeisen hylkäämisen jälkeen kirjallisuus ei esitellyt selkeitä CRISP-DM:n seuraajia. Kuinka esimerkiksi nykyaikaisten BI-työkalujen ja analytiikkaprosessien ylläpitoa on tutkittu tai mietitty, vai ollaanko ainoastaan teknologiakehittäjien parhaiden käytäntöjen varassa?



## 8 Päätelmät

Analytiikka on nykyisessä laajuudessaan suhteellisen nuori tieteellinen ja ammatillinen suuntaus. Alan käsitteet ja terminologiat ovat vaihtuneet vuosien varrella ja vielä tälläkin hetkellä käsitteiden välillä on epäselvyyttä ja niitä käytetään ristiin. Analytiikkaratkaisut ovat tekijänsä näköisiä ja käytänteet niiden kehittämisessä ja tukemisessa eivät ole yhtä pitkälle kehittyneitä kuin esimerkiksi ohjelmistokehityksessä. Analytiikan hyväksynnästä ja sen adoptioasteesta on olemassa tutkimuksia, mutta varsinaisesta analytiikan tukemisestä tai sen jatkuvasta käytöstä on vähemmän materiaalia.

Tässä tutkielmassa tutkittiin analytiikan tukemista, analytiikan tukemisen haasteita ja analytiikan parempaa tukemista. Tutkimuksen aluksi tutkittiin aihepiirin teoreettinen tausta ja tutustuttiin analytiikan peruskäsitteisiin. Peruskäsitteiden jälkeen esiteltiin analytiikan elinkaarimallit ja erityisesti CRISP-DM (Chapman et al., 2000) ja sen laajennettu malli (Cochran, 2019), joiden avulla syvennyttiin analytiikan tukemiseen sen jälkeen, kun analytiikkaratkaisu on viety tuotantoon. Tutkimus toteutettiin toimeksiantona yritykselle, jonka asiakkaiden analytiikan asiantuntijoita tai johtajia haastateltiin laadullisen tutkimuksen menetelmien analytiikan elinkaaresta ja erityisesti analytiikan tukemisestä.

Analytiikkaa tutkimuksen tuloksena tukevat analytiikkatiimit, jotka sijaitsevat joko talouden tai IT:n sisällä. Tuen ensimmäisen tason malli vaihteli, osassa organisaatiossa tuki oli toimitettu IT-vetoisesti helpdeskin avulla, kun taas osassa tuki tarjottiin talousvetoisesti business controllerien toimesta. Talousvetoisen mallin etuna nähtiin se, että tuki ymmärtää liiketoiminnan ja tuntee analytiikan käyttäjät. Analytiikan tukemisen määrää pyrittiin kaikissa organisaatiossa vähentämään tekemällä ratkaisuihin mahdollisimman helppokäyttöisiä. Kaikessa analytiikan tekemisessä korostui se, että analytiikkatiimit olivat kommunikaatiossa liiketoiminnan kanssa niin uuden kehittämisessä kuin vanhan analytiikan ylläpitämisessä. Osa haastateltavista näki, että on hyvä malli, että analytiikan kehittäjät tukevat olemassa olevia analytiikkaratkaisuja ja osa taas näki sen työtä mahdollisesti häiritsevänä tekijänä.

Analytiikan tukemisessa haasteena oli tuen pysyvyys. Koettiin, että mikäli tuki vaihtuu usein, ongelmaksi nousee liiketoimintaymmärryksen puute. Nähtiin, että liiketoimintaymmärrys kasvaa tukemista tekemällä. Haasteena oli myös analytiikkanäkymien suuri määrä ja mahdollisten liiketoimintalogiikkamuutosten heijastuminen niihin. Ei koettu, että analytiikkatiimillä on tarpeeksi resursseja niiden valvontaan, vaan siihen kaivattiin avuksi liiketoimintaa. Resurssit olivat haasteena myös käyttäjien koulutuksessa, jossa apua odotettiin myös liiketoiminnan osalta.

Analytiikan parempaan tukemiseen nostettiin useampia kehityskohteita. Osa haastateltavista esitti analytiikkatiimiin erillisen tukiroolin perustamista, joka vapauttaisi kehittäjät tekemään kehittämistä. Esille nousi myös se, että jokaisella analytiikkaratkaisulla olisi hyvä olla liiketoiminnan puolelta omistaja tai muu vastaava taho, joka auttaisi analytiikkaratkaisujen laadun ja ajankohtaisuuden ylläpitämisessä. Liiketoiminta tuntisi oman substanssinsa ja siihen liittyvän analytiikan ja pystyisi siten pitämään ratkaisusta huolta.

Työssä käsitelty laajennettu elinkaarimalli CRISP-DM kohtasi osittain haastateltavien todellisuuden kanssa, mutta osittain mallin tarjoama pinta jäi ohueksi. Osa mallin sisällöstä ei ole tämän tutkimuksen kohdeorganisaation käytänteiden mukaisia ja vaikka mallin sisältöjen inspiroimia haastattelukysymyksiä ei sinänsä torjuttu haastatteluissa, niin on epätodennäköistä tulosten perusteella, että analytiikan elinkaarta tai edes sen tukemista tehtäisiin jatkossakaan mallin vaatimalla laajuudella. Jatkossa voisikin olla mielenkiintoista selvittää sitä, että mikä malli sopisi parhaiten nykyisille analytiikkaorganisaatioille vai olisiko mallissa päivitettävää?

## 9 Viiteluettelo

- Aho, T., Kilamo, T., Lwakatare, L., Mikkonen, T., Sievi-Korte, O., & Yaman, S. (2021). *Managing and Composing Teams in Data Science: An Empirical Study*. IEEE.
- Aho, T., Sievi-Korte, O., Kilamo, T., Yaman, S. G., & Mikkonen, T. (2020). *Demystifying Data Science Projects: a Look on the People and Process of Data Science Today*. Springer, Cham.
- Alasuutari, P. (2011). *Laadullinen tutkimus 2.0* (4. uud. p. ed.). Vastapaino.
- Alliance, D. S. P. (2023). *CRISP-DM*. Haettu 26.04.2023 osoitteesta <https://datascienceprocess.com/member-home-page/crisp-dm/>
- Angée, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J.-D., & Tabares-Betancur, M. S. (2018). Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects. In L. Uden, B. Hadzima, & I. H. Ting, *Knowledge Management in Organizations* Cham.
- Azevedo, A., & Santos, M. (2008). *KDD, semma and CRISP-DM: A parallel overview*.
- Balusamy, B., Abirami R, N., Kadry, S., & Gandomi, A. H. (2021). Big Data Analytics. In (1 ed., pp. 161-186). Wiley. <https://doi.org/10.1002/9781119701859.ch6>
- Baskarada, S., & Koronios, A. (2013). Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. *Australasian Journal of Information Systems*, 18(1). <https://doi.org/https://doi.org/10.3127/ajis.v18i1.748>
- Bischoff, S., Aier, S., Haki, M. K., & Winter, R. (2015). Understanding Continuous Use of Business Intelligence Systems: A Mixed Methods Investigation. *JITTA : Journal of Information Technology Theory and Application*, 16(2), 5-37.
- CAP. (2023). *Job Task Analysis*. Certified Analytics Professional. Haettu 27.04.2023 osoitteesta <https://www.certifiedanalytics.org/jta>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T. P., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.

- Cochran, J. J. (2019). *Infirms analytics body of knowledge* (1st edition ed.). Wiley.
- Davenport, T. (2012). *The Complete Guide to Business Analytics (Collection)* (1st edition ed.). Pearson.
- Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work : smarter decisions, better results*. Harvard Business Press.
- Davenport, T. H., Harris, J. G., & Paalosalo, M. (2007). *Analysoi ja voita : kilpailun uusi tie*. Talentum.
- Finn, J., Troughton, G., Baxter, P., Cookson, M., Cornish, K., Corrigan, J., Lye, Y., Sella, N., Pumphrey, S., Taylor, I., & Wagner, C. (2020). *Predictive Analytics for Healthcare* (1st edition ed.). O'Reilly Media, Inc.
- Gartner. (2023). *What Is Data and Analytics?* Gartner. Haettu 20.04.2023 osoitteesta <https://www.gartner.com/en/topics/data-and-analytics>
- Geng, H. (2017). *Internet of Things and Data Analytics Handbook*. John Wiley & Sons, Incorporated.  
<http://ebookcentral.proquest.com/lib/tampere/detail.action?docID=4771458>
- Gummesson, E. (2008). Customer centricity: reality or a wild goose chase? *European business review*, 20(4), 315-330. <https://doi.org/10.1108/09555340810886594>
- Hakala, J. T. (2008). *Uusi graduopas : melkein maisterin entistä ehompi niksikirja*. Gaudeamus Helsinki University Press.
- Hotz, N. (2023). *What is CRISP DM?* Datascience Process Alliance. Haettu 20.04.2023 osoitteesta <https://www.datascience-pm.com/crisp-dm-2/>
- Hyvärinen, M. K., Nikander, P., Ruusuvoori, J., & Aho, A. L. (2017). *Tutkimushaastattelun käsikirja*. Vastapaino.
- Institute, S. (2023). *Data Mining and SEMMA*. SAS Institute Inc. . Haettu 25.04.2023 osoitteesta <http://documentation.sas.com/doc/en/emcs/14.3/n0pejm83csbja4n1xueveo2uoujy.htm>
- Kandel, S., Paepcke, A., Hellerstein, J. M., & Heer, J. (2012). Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2917-2926.  
<https://doi.org/10.1109/TVCG.2012.219>

- Käpylä, J., & Saloniemi, H. (2013). *Tietojohdajan taskukirja. Tietojohdamisen näkökulmia aluekehittämiseen*. Tampereen teknillinen yliopisto, Tietojohdamisen tutkimuskeskus Novi.
- Liebowitz, J. (2021). *Data analytics and AI*. CRC Press, Taylor & Francis Group.
- Loshin, D., & Reifer, A. (2013). *Using information to develop a culture of customer centricity : customer centricity, analytics, and information utilization* (1st edition ed.). Morgan Kaufmann.
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2), 137-166.  
<https://doi.org/https://doi.org/10.1017/S0269888910000032>
- Microsoft. (2023a). Power BI demonäkymä. In *Microsoft Power BI*.
- Microsoft. (2023b). *What is data visualization?* . Microsoft. Haettu 22.04.2023 osoitteesta <https://powerbi.microsoft.com/en-us/data-visualization/>
- Myers, N. E., & Gregory, K. (2021). *Self-service data analytics and governance for managers*. John Wiley & Sons, Incorporated.
- Nelson, G. S. (2018). *The analytics lifecycle toolkit : a practical guide for an effective analytics capability* (1st edition ed.). Wiley.
- Ponsard, C., Touzani, M., & Majchrowski, A. (2017). Combining Process Guidance and Industrial Feedback for Successfully Deploying Big Data Projects. *Open Journal of Big Data (OJBD)*, 3, 26-41.
- Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly.
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163-180.  
<https://doi.org/10.1177/0165551506070706>
- Saaranen-Kauppinen, A., & Puusniekka, A. (2006). *KvaliMOTV - Menetelmäopetuksen tietovaranto [verkkójulkaisu]*. Haettu 03.05.2023 osoitteesta <https://www.fsd.tuni.fi/menetelmaopetus/kvali/>
- Saarijärvi, H. (2012). The mechanisms of value co-creation. *Journal of strategic marketing*, 20(5), 381-391. <https://doi.org/10.1080/0965254X.2012.671339>

- Saarijärvi, H., & Puustinen, P. (2020). *Strategiana asiakaskokemus : miksi, mitä, miten?* Docendo.
- Saltz, J. S. (2015, 29 Oct.-1 Nov. 2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. 2015 IEEE International Conference on Big Data (Big Data),
- Schlegel, K. S., Julian. (2023, 22.04.2023). Magic Quadrant for Analytics and Business Intelligence Platforms. <https://www.gartner.com/doc/reprints?id=1-2CF2LJQ8&ct=230130&st=sb>
- Schwaber, K. (2023). *About Scrum.org*. Haettu 09.05.2023 osoitteesta <https://www.scrum.org/about>
- Sharda, R., Delen, D., Turban, E., Aronson, J. E., Liang, T.-P., & King, D. (2018). *Business intelligence, analytics, and data science : a managerial perspective* (Fourth edition. ed.). Pearson.
- Shi-Nash, A., & Hardoon, D. R. (2016). DATA ANALYTICS AND PREDICTIVE ANALYTICS IN THE ERA OF BIG DATA. In (1 ed., pp. 329-345). Wiley. <https://doi.org/10.1002/9781119173601.ch19>
- Tableau. (2023). *What Is Data Visualization? Definition, Examples, And Learning Resources*. Tableau. Haettu 22.04.2023 osoitteesta <https://www.tableau.com/learn/articles/data-visualization>
- Takala, P. (2018). Data-analytiikka yhteiskunnallisen päätöksenteon tukena. In: Sitra.
- Tiainen, T. (2014). *Haastattelu tietojenkäsittelytieteen tutkimuksessa*. Tampereen yliopisto, informaatiotieteiden yksikkö.
- Tuomi, J., & Sarajärvi, A. (2009). *Laadullinen tutkimus ja sisällönanalyysi* (5., uud. laitos ed.). Tammi.

## 10 Liite 1: Pohjustus tutkimuksesta

### Asiakaskeskeinen analytiikkatuen arviointi

Tutkijan yhteystiedot:

Niko Ruotsalainen  
niko.ruotsalainen@tuni.fi  
0405563963  
Tampereen Yliopisto  
Informaatioteknologian ja viestinnän tiedekunta

Kiitos haastatteluun osallistumisesta!

Tutkimuksen tavoitteena on selvittää kokemuksia siitä, miten analytiikkatukea on järjestetty ja mitä haasteita siihen liittyy. Haastattelun aikana käydään läpi analytiikan elinkaarta CRISP-DM-mallia mukaillen.

Haastattelun kestona on noin tunti ja se käydään lähtökohtaisesti suomeksi ja sen tuloksia käytetään osana pro gradu tutkielmaani. Mikäli tutkielmaan, haastatteluihin tai sen sisältöihin tulee mitään kysyttävää, niin tutkijaan voi olla yhteydessä ennen tai jälkeen haastattelujen. Myös kesken haastattelun haastateltavan on mahdollista olla vastaamatta kysymyksiin tai pyytää tallentamisen lopettamista.

Haastattelu tallennetaan digitaalisesti Microsoft Teamsilla, jonka jälkeen haastattelu literoidaan tekstimuotoon. Tekstimuodossa olevaa dataa käsitellään anonymisti, eli henkilötiedot tai organisaation tunnistetiedot poistetaan tai muutetaan siten, että henkilöiden tai yritysten tunnistaminen ei ole mahdollista.

Haastattelumateriaalit ja tallenteet poistetaan gradun arvostelun jälkeen ja ne ovat ainoastaan tutkijan käytettävissä tutkimusprosessin ajan. Materiaalit säilytetään salasanasuojatussa sijainnissa.

Tutkimukseen voidaan sisällyttää suoria lainauksia haastatteluista ja niiden yhteydessä voidaan mainita taustatietoja, kuitenkin siten, että haastateltavan anonymiteetti säilyy.

## 11 Liite 2: Asiantuntija-haastattelun runko

### Taustatiedot

- **Kerrotko omasta kokemuksestasi data-analytiikan parissa ja roolistasi nykyisessä organisaatiossa?**
  - o Apukysymyksiä:
    - Alan kokemus vuosina, urapolku?
    - Nykyinen positio
    - Koulutus / tausta
  
- **Mikä on yrityksen toimiala?**
  
- **Miten analytiikka sijoittuu organisaatiossanne?**
  - o Apukysymyksiä
    - IT/Liiketoiminta-johtoinen/Data-vetoinen?
    - Hajautettu, keskitetty, konsultoiva, toimintopohjainen, center of excellence?
  
- **Minkä kokoisessa organisaatiossa toimit?**
  - o Apukysymyksiä
    - Kuinka iso datatiiminne on?
    - Käytättekö alihankkijoita tai kumppaneita?
    - Kuinka monta loppukäyttäjää analytiikalla on?

### Tukitehtävät

- **Kerrotko analytiikkaratkaisujen elinkaaresta ja erityisesti siitä vaiheesta kun analytiikkaratkaisu viedään tuotantoon?**
  - o Apukysymyksiä:
    - Millainen prosessi (onko aina samanlainen, onko systemaattista prosessia)?
    - Liittyykö analytiikan käyttöön koulutusta?
    - Mietitäänkö analytiikkaratkaisun elinkaarta tässä vaiheessa?
    - Onko tässä vaiheessa dokumentoitu ratkaisun toimintaa
  
- **Miten analytiikan käyttöä tuetaan tällä hetkellä?**
  - o Apukysymyksiä:
    - Miten tuki on järjestetty: Kuka tukee ja ketä? (Ulkoinen/sisäinen tuki, IT/liiketoiminta)
    - Toimittaako tukea samat henkilöt, jotka kehittävät analytiikkaa (development ja operations on sama)
    - Millaista tukea tarvitaan? (Mitä tehtäviä ja mihin tehtävät liittyvät (tekniikka, data, liiketoimintaymmärrys))
    - Millainen liiketoimintaymmärrys tuella on?



- **Miten haasteita analytiikan tukemisessa on tällä hetkellä?**
  - Apukysymyksiä:
    - Miten tuki tulisi järjestää? (Kuka tukee, miksi, ulkoinen/sisäinen tuki, IT/liiketoiminta)
      - Kehittäjien rooli?
      - Mikäli tuki on ulkoisesti hankittu, niin mikä olisi näkemyksesi mukaan järkevin malli?
    - Ketä tuettaisiin?
    - Tulisiko tuella olla enemmän ymmärrystä teknisistä, data- vai liiketoiminta lähtöisistä asioista? Vai kaikista näistä
    - Käyttäjien koulutus?
    - Dokumentointi?
  
- **Miten tällä hetkellä pidetään huolta siitä, että analytiikkaratkaisu pysyy ajankohtaisena?**
  - Apukysymyksiä:
    - Valvotaanko ratkaisua jotenkin? (Ollaanko varmoja, että luvut ovat oikein)
    - Nojaako analytiikka ydintietojen päälle ja jos, niin ylläpidetäänkö niitä
    - Miten ratkaisun toimintaa korjataan ajan kuluessa? (Aktiivisesti vai käyttäjien pyynnöstä)
    - Miten ratkaisun laadusta pidetään huolta?
    - Kenen vastuulla on pitää ratkaisu ajankohtaisena (IT/liiketoiminta)?
  
- **Mitä haasteita on analytiikkaratkaisun ajankohtaisena pitämisessä?**
  - Apukysymyksiä:
    - Kenen pitäisi omistaa ratkaisu?
    - Analytiikan tarpeet voivat muuttua, miten ratkaisua pitäisi seurata tai valvoa?
    - Miten ratkaisun laadusta pidetään huolta?
  
- **Miten analytiikkaratkaisun liiketoimintahyötyjä seurataan?**
- **Miten liiketoimintahyötyjä tulisi seurata?**
  - Onko analytiikkatuella tässä roolia?
  
- (Onko analytiikan tukeminen muuttunut aikojen saatossa)