

Lappeenrannan teknillinen yliopisto
School of Business and Management
Tietotekniikan koulutusohjelma

Kandidaatintyön loppuraportti

Markus Kyläheiko

**Big datan hyödyntäminen jalkapallojoukkueen urheilutoiminnan
tehostamisessa: case Suomen kakkosdivisioona**

Työn tarkastaja: Professori Jari Porras

Työn ohjaaja: Professori Jari Porras

TIIVISTELMÄ

Lappeenrannan teknillinen yliopisto
School of Business and Management
Tietotekniikan koulutusohjelma

Markus Kyläheiko

Big datan hyödyntäminen jalkapallojoukkueen urheilutoiminnan tehostamisessa: case Suomen kakkosdivisioona

Kandidaatintyö

2017

40 sivua, 4 kuvaa, 3 taulukkoa, 1 liite

Työn tarkastaja: Professori Jari Porras

Hakusanat: big data, data-analyysi, urheiluanalyysi, jalkapallo

Keywords: big data, data-analysis, sports analytics, football

Tämän kandidaatintyön tarkoituksena on tutkia, miten big dataa yleisesti käytetään, mistä dataa kerätään, mitä uhkia datan käytöstä mahdollisesti syntyy ja miten dataa voisi erityisesti Suomen kolmanneksi korkeimmalla sarjatasolla käyttää entistä tehokkaammin hyödyksi. Myös seurojen mielipiteitä datan käytöstä ja tarpeellisuudesta käydään työssä läpi.

Suomen kakkosdivisioonassa käytettyjä datan hyödyntämismenetelmiä, datan keräysmenetelmiä ja seurojen asenteita tutkittiin kyselyllä. Tuloksista käy ilmi, että seurojen asenteet dataa kohtaan olivat positiiviset ja dataa käytettiin moneen eri tarkoitukseen, kuten otteluanalysointiin, pelaajien ja vastustajan scouttaamiseen, loukkaantumisten ehkäisyyn ja hoitoon sekä pelaajien testaamiseen. Osalla seuroista oli myös seuran ulkopuolisia toimijoita apuna datan tuottamisessa, keräämisessä ja analysoinnissa.

Työn tuloksista voi päätellä, että datan käyttö Suomen kakkosdivisioonassa on välttämätöntä, mikäli siellä haluaa pärjätä. Ilman datan hyödyntämistä joukkueet jäävät muista seuroista jälkeen.

ABSTRACT

Lappeenranta University of Technology
School of Business and Management
Degree Program in Computer Science

Markus Kyläheiko

Exploitation of big data in improving a football team's sport performance: The case of the Finnish second division

Bachelor's thesis

2017

40 pages, 4 figures, 3 tables, 1 appendices

Examiner: Professor Jari Porras

Keywords: big data, data-analysis, sports analytics, football

The main goal of this bachelor's thesis is to examine how big data is used and collected, what kinds of risks and threats the usage of data may cause and how the usage of data could be more efficient especially in the Finnish second division. The opinions of the data users concerning the usage of data were also analysed by means of a questionnaire.

Different data exploitation methods, data collection methods and attitudes of the Finnish second division clubs were investigated by a questionnaire. It was noticed that the attitudes were positive and data was used for many different purposes, such as analysing matches, preventing injuries and testing players. Some of the clubs also used outside companies in producing, collecting and analysing the data.

Judging by the results, the use of data in the Finnish second division is necessary, if the team wants to outperform in the league. Without the data exploitation, a team gives a big advantage to the others.

SISÄLLYSLUETTELO

1	JOHDANTO	2
1.1	TAUSTA	2
1.2	TAVOITTEET JA RAJAUKSET	3
1.3	TUTKIMUSMENETELMÄT.....	4
1.4	TYÖN RAKENNE	5
2	KIRJALLISUUSKATSAUS	6
2.1	BIG DATA.....	7
2.2	BIG DATAN SOVELLUSALUEET	10
2.3	DATAN KERÄYSMENETELMÄT	13
2.3.1	<i>Videoteknologia</i>	13
2.3.2	<i>Puettavat laitteet</i>	14
2.3.3	<i>Laadullinen tieto</i>	15
2.4	DATAN HYÖDYNTÄMISTAVAT	15
2.4.1	<i>Scouttaaminen</i>	15
2.4.2	<i>Otteluiden ja harjoitusten analysointi</i>	16
2.5	DATAN AIHEUTTAMAT RISKIT JA ONGELMAT	17
3	CASE SUOMEN KAKKOSDIVISIOONA	20
3.1	KYSYMYKSET	20
3.2	VASTAUKSET	20
3.3	VASTAUSTEN ANALYSOINTI.....	26
3.4	JOUKKUEIDEN EROT	28
4	JOHTOPÄÄTÖKSET	30
	LÄHTEET	32
	LIITE1	

1 JOHDANTO

Johdannossa käydään läpi, mitä kandidaatintyössä käsitellään ja millä tavalla aihe on rajattu. Tässä kappaleessa on tarkoitus saada kokonaiskuva kandidaatintyöstä ja siitä, mitä työssä erityisesti tutkitaan ja minkälaisia tutkimusmenetelmiä käytetään. Kandidaatintyön aiheena on big datan hyödyntäminen jalkapallojoukkueen urheilutoiminnan kehittämässä ja case-vaiheessa käytetään Suomen kakkosdivisioonaa. Työ tehdään Lappeenrannan teknillisen yliopiston tietotekniikan osastolle.

1.1 Tausta

Opinnäytetyön aiheena on big datan hyödyntäminen jalkapallojoukkueen urheilutoiminnan tehostamisessa erityisesti Suomen kakkosdivisioonassa, eli Suomen kolmanneksi korkeimmalla sarjatasolla. Erityisen kiinnostava aihe on siksi, että datan määrä on lisääntynyt niin jalkapallossa kuin muissakin suurissa urheilulajeissa merkittävästi viime vuosina. Tässä työssä keskitytään siihen dataan, millä joukkueet pyrkivät parantamaan urheilullista menestystään. Toki big dataa käytetään jalkapalloseuroissa nykyään myös taloudellisen menestyksen tukemiseen, esimerkiksi markkinointia, brändin rakentamista, rahoitusta ja laskentatoimea kehitettäessä, mutta tämä puoli jätetään tässä työssä käsittelemättä. Fokus on puhtaasti urheilullisen menestyksen parantamisessa. Jalkapallossa datan hyödyntämistä on pidetty esimerkiksi baseballia tai koripalloa hankalampana, koska tarkkoja kaksinkamppailuja, kuten baseballissa syöttäjän ja lyöjän välillä, ei jalkapallossa varsinaisesti synny. Lisäksi kentän laajuutta, pelaajien suurta määrää ja pelin sujuvuutta on pidetty datan hyödyntämistä vaikeuttavina tekijöinä.

Big datalla tarkoitetaan kaikkea mahdollista saatavilla olevaa dataa, jota voidaan kerätä eri tavalla monista lähteistä. Dataa saadaan nykypäivänä jalkapallojoukkueiden käyttöön erilaisten mittarien, sensorien, videoanalysoinnin, GPS:n ja internetin kautta. Nykyään digitalisaation seurauksena dataa on niin paljon saatavilla, että seurojen tärkeimpiä tehtäviä onkin miettiä, mitä he kaikella saatavissa olevalla datalla tekevät ja miten sitä parhaiten hyödynnetään. (Rejec, 2016)

Nykypäivän jalkapallossa big dataa käytetään lähes kaikkeen, kuten pelaajien ja vastustajien scouttaamiseen, harjoitusten ja otteluiden tutkimiseen ja analysointiin sekä pelaajien pelikunnan ja palautumisen tarkkailuun. Pelaajascouttaamisella tarkoitetaan toimintaa, jossa pyritään esimerkiksi katsomalla tietyn pelaajan tilastoja ja peliesityksiä arvioimaan hänen tasoaan ja sopivuuttaan tiettyyn joukkueeseen ja vastustajaa scoutatessa pyritään arvioimaan vastustajan taktiikkaa, heikkouksia, vahvuuksia ja pelitapaa samalla tavalla videoiden ja tilastojen kautta. Vaikka Suomen alemmissa sarjoissa dataa ei ole läheskään yhtä paljon saatavilla kuin maailman suurimmissa liigoissa, käytetään sitä kuitenkin laajasti apuna seurojen toiminnan kehittämisessä etenkin ottelujen, harjoitusten ja mahdollisten hankintojen analysoinnissa. Lisäksi seuroilla on apuna dataa tuottavia yrityksiä, jotka esimerkiksi videoanalyysien avulla selvittävät lähes kaiken, mitä ottelusta on mahdollista annetun materiaalin avulla ylipäättänsä selvittää.

1.2 Tavoitteet ja rajaukset

Tämän työn tavoitteena on selvittää, millä tavoilla jalkapallossa käytetään big dataa apuna seurojen urheilutoiminnan kehittämisessä, mistä dataa saadaan, mitä uhkia data voi aiheuttaa ja miten sitä voidaan käyttää mahdollisimman tehokkaasti hyödyksi. Erityisesti tutkitaan niitä alueita, joissa dataa voidaan hyödyntää, kuten ottelut ja harjoitukset, pelaajien hankinta ja vastustajien arviointi. Tällä tavalla datan käyttöä tulee tutkittua tarkasti ja jokaiseen alueeseen keskittyään erikseen syvemmin. Työssä on tarkoituksena käydä ensin läpi eri menetelmiä, joita jalkapallomaailman huipulla käytetään ja mitä kautta kyseisiin menetelmiin on historiallisesti päädytty. Työssä käydään läpi aikaisempia tutkimuksia tähän aihepiiriin liittyen, jotta saadaan selkeä kuva siitä, kuinka pitkällä big datan hyödyntäminen suurissa jalkapalloseuroissa todellisuudessa tällä hetkellä on. Mahdollisia datan keruun ja analyysin aiheuttamia riskejä käydään opinnäytetyössä myös läpi.

Yleisemmän johdanto-osan jälkeen kandidaatintyössä käsitellään case-tyyppisesti, miten Suomen kolmanneksi korkeimman sarjataso seurat hyödyntävät dataa oman seuran urheilullisessa kehittämisessä. Tähän vaiheeseen valitaan viisi Suomen kakkosdivisioonassa pelaavaa joukkuetta, joille esitetään netin tai puhelimen välityksellä samanlaiset kysymyssarjat. Kysymyksien tarkoituksena on kattaa kaikki osa-alueet, kuten pelaajien ja vastustajan scouttaus sekä pelien ja harjoitusten analysointi. Tällä tavalla pyritään saamaan kokonaiskuva siitä, miten seurat nykyään käyttävät dataa, miten datan käyttö eroaa eri seurojen välillä ja mitä puutteita

datan hyödyntämisessä mahdollisesti on. Tietenkään tällä tasolla resurssit eivät ole läheskään maailman suurimpien seurojen tasolla ja se otetaan luonnollisesti työssä huomioon.

Big datan voi ymmärtää monella eri tavalla, mutta tässä työssä, sillä tarkoitetaan kaikkea saatavilla olevaa käsittelemätöntä dataa. Datan analysointi kuuluu jalkapalloseurojen tehtävään, jotta he saavat maksimaalisen hyödyn kaikesta saatavilla olevasta datasta. Termi big data on kuitenkin huomattavasti selkeämpi ymmärtää, kun sillä tarkoitetaan raakaa dataa, johon ei ole lisätty arvoa analysoinnin muodossa (Ylijoki & Porras, 2016, s. 77).

Opinnäytetyö rajataan siten, että yleiskatsaus tehdään maailman huipulla käytettyjen datan hyödyntämismenetelmien perusteella, mutta tarkempi yksityiskohtainen analysointi big datan hyödyntämisen nykytilasta tehdään siirtymällä casen avulla Suomen kolmanneksi korkeimmalle sarjatasolle. Tutkimuskysymykset ovatkin seuraavat:

1. Miten dataa kerätään jalkapallojoukkueissa?
2. Miten dataa käytetään hyväksi?
3. Mitä riskejä data saattaa aiheuttaa?
4. Miten dataa voisi käyttää mahdollisesti entistä tehokkaammin hyödyksi Suomen kolmanneksi korkeimmalla sarjatasolla.

1.3 Tutkimusmenetelmät

Opinnäytetyössä käytetään tutkimuskysymyksiin vastaamiseen erilaisia menetelmiä. Sellaisiin kysymyksiin, kuten miten dataa kerätään jalkapallojoukkueissa sekä miten dataa käytetään hyväksi ja mitä riskejä data mahdollisesti aiheuttaa, vastataan tutkimalla aikaisempia tutkimuksia ja muita kirjallisia lähteitä. Tämän lisäksi siirryttäessä työn case-osassa nimenomaan Suomen kakkosdivisioonan joukkueisiin, käytetään yhtenä tutkimusmenetelmänä nettikyselyjä tai puhelinhaastatteluja, riippuen osallistuvista henkilöistä. Nettikysely ja puhelinhaastattelu luodaan siten, että ne sopivat hyvin työn alkuosan kautta nousseisiin tutkimuskysymyksiin. Tavoitteena on saada selvät vastaukset viideltä Suomen kakkosdivisioonassa pelaavalta joukkueelta, joiden nimiä ei kuitenkaan kandidaatintyössä voida luottamussyistä paljastaa. Joukkueissa vastaamisen hoitavat joko päävalmentaja tai joku muu valmennustiimin jäsen.

Uusia datan käytön mahdollisuuksia mietitään yllä mainituiden menetelmien lisäksi myös oman pohdinnan ja ideoinnin kautta, vaikka se luonnollisesti jokaiseen kysymykseen vastaamiseen jollain tavalla kuuluukin. Tutkimuskysymykseen, miten dataa voisi käyttää mahdollisesti entistä tehokkaammin hyödyksi Suomen kolmanneksi korkeimmalla sarjatasolla, oma pohdinta liittyy kuitenkin poikkeuksellisen suuresti.

Kysymyssarja (kts. Liite 1), niin nettikyselyyn kuin puhelinhaastatteluunkin, luodaan siten, että se sisältää kahdeksan kysymystä, joiden tarkoituksena on kattaa opinnäytetyön pääkohdat. Kysymysten määrä pidetään suhteellisen alhaisena, jotta vastaajat jaksavat panostaa vastauksiin kunnolla ja viitsivät vastata niihin kattavasti. Vastauksia verrataan joukkueittain keskenään, jotta havaitaan, miten ne mahdollisesti eroavat toisistaan.

1.4 Työn rakenne

Luvussa 2 käsitellään sitä, mitä opinnäytetyön aiheesta on entuudestaan erilaisten kirjallisten julkaisujen kautta jo tiedossa. Kappaleen pitäisi auttaa ymmärtämään, millä tavalla dataa jalkapallossa käytetään ja mihin suuntaan datan hyödyntäminen on menossa.

Luku 3 käsittelee Suomen kakkosdivisioona seuroista tehtyä case-tutkimusta. Tässä luvussa käydään läpi, mitä tarkalleen tutkittiin, miksi tutkittiin ja mitä siitä opittiin. Tutkimuksen tuloksia käydään myös kattavasti läpi.

Luvussa 4 esitetään kandidaatintyön johtopäätökset ja vastaukset tutkimuskysymyksiin sekä käydään tarkemmin läpi, mitä työstä yleensä ja etenkin empiirisestä tutkimusosioista on opittu ja miten seurat voisivat entistä tehokkaammin hyödyntää big datan mahdollisuuksia.

2 KIRJALLISUUSKATSAUS

Jalkapallossa on tietenkin jo pitkään käytetty esimerkiksi vastustajaa koskevaa dataa esimerkiksi pelien scouttaamisen kautta (esimerkiksi lähettämällä joku katsomaan vastustajan pelejä ja mahdollisesti myös harjoituksia), mutta nykypäivänä dataa käytetään huomattavasti enemmän ja, mikä tärkeintä, systemaattisemmin kuin esimerkiksi kaksikymmentä vuotta sitten. Tässä ajassa datasta on tullut merkittävä osa jokaisen menestyvän jalkapallojoukkueen toimintaa, jota se on myös uuden teknologian kautta muuttanut (The Future of Sports, 2016). Datan oikeanlaisella hyödyntämisellä voidaan ainakin osittain kuroa mahdollisia eroja, kuten rahan puutetta, kiinni, minkä seurauksena dataa pyrkivät hyödyntämään nykypäivänä kaikki seurat mahdollisimman tehokkaasti. Tulevaisuudessa datan käyttö tulee lisääntymään entisestään, jolloin jokaisen seuran on pysyttävä kehityksessä mukana ja panostettava big data-analytiikkaan (Steinberg, 2015). On myös nähtävissä, että alalle on syntymässä big datan keruuseen ja analysointiin erikoistuneita yrityksiä, jotka suorittavat perusdata-analyysin, mutta koska tämä tieto on lähivuosina melko edullisesti ja helposti kaikkien saatavilla, ei varsinaista kilpailuetua voi sen varaan rakentaa. Selvää tietysti on, että jos tällä tasolla jää kilpailijoista jälkeen, tulokset voivat olla erittäin heikot. Pelkän perusanalyysin lisäksi tarvitaan siis erityiskykyä hyödyntää standardianalyysin tuloksia ja viedä niitä tehokkaasti niin pelaajascouttaukseen, valmennukseen, pelitaktiikkaan, harjoitteluun kuin loukkaantumisten estämiseen ja niistä toipumiseenkin.

Ensimmäisiä kertoja kerättyä dataa pyrittiin systemaattisesti hyödyntämään jalkapallossa 1950-luvulla Charles Reepin toimesta. Reep analysoi erityisesti maaleja edeltäviä syöttökuvioita, joista hän huomasi, että suurin osa maaleista tehtiin maksimissaan kolmen peräkkäisen onnistuneen syötön seurauksena. Reepiä onkin osittain syytetty brittiläisen suoraviivaisen pelitavan luomisesta, sillä hänen mukaansa pitkät onnistuneet syöttökuviot vähensivät maalin syntymisen todennäköisyyttä. (Pollard 2002, s. 853-855) Myöhemmin tämä yksinkertainen ajatus on vahvasti haastettu, mutta kaikesta huolimatta hän loi kuitenkin merkittävän ajatusmallin, jonka mukaan vanhat onnistuneet pelitapahtumat auttavat luomaan mallin siitä, mitä tulevaisuudessakin kannattaa onnistuakseen tehdä.

Vasta 1990-luvun loppupuolella videoita ja dataa alettiin kuitenkin käyttää enemmän apuna jalkapalloseurojen kehittämisessä. Silloin Prozone (STATS¹, 2017), Opta (Opta, 2017) ja muut niiden kaltaiset ohjelmistot alkoivat analysoida tarkemmin otteluja ja niissä syntyviä tilanteita. Noin kahdessakymmenessä vuodessa datan käyttö on yleistynyt todella nopeasti keskeiseksi osaksi jalkapalloa. Esimerkiksi yhdestä Saksan Bundesliiga-ottelusta otetaan nykyään automaattisesti reilusti yli 100 000 kuvaa, jotka tallennetaan välittömästi datapankkiin (Gehrmann, 2016). Datan käyttö on kuitenkin edelleen suhteellisen uutta jalkapallossa, joten kerätyn valtavan datamäärän käyttäminen mahdollisimman tehokkaasti hyödyksi aiheuttaa edelleen haasteita seuroille samalla, kun se tuottaa kilpailuetua analyysissään pidemmälle ja syvemmälle ehtineille.

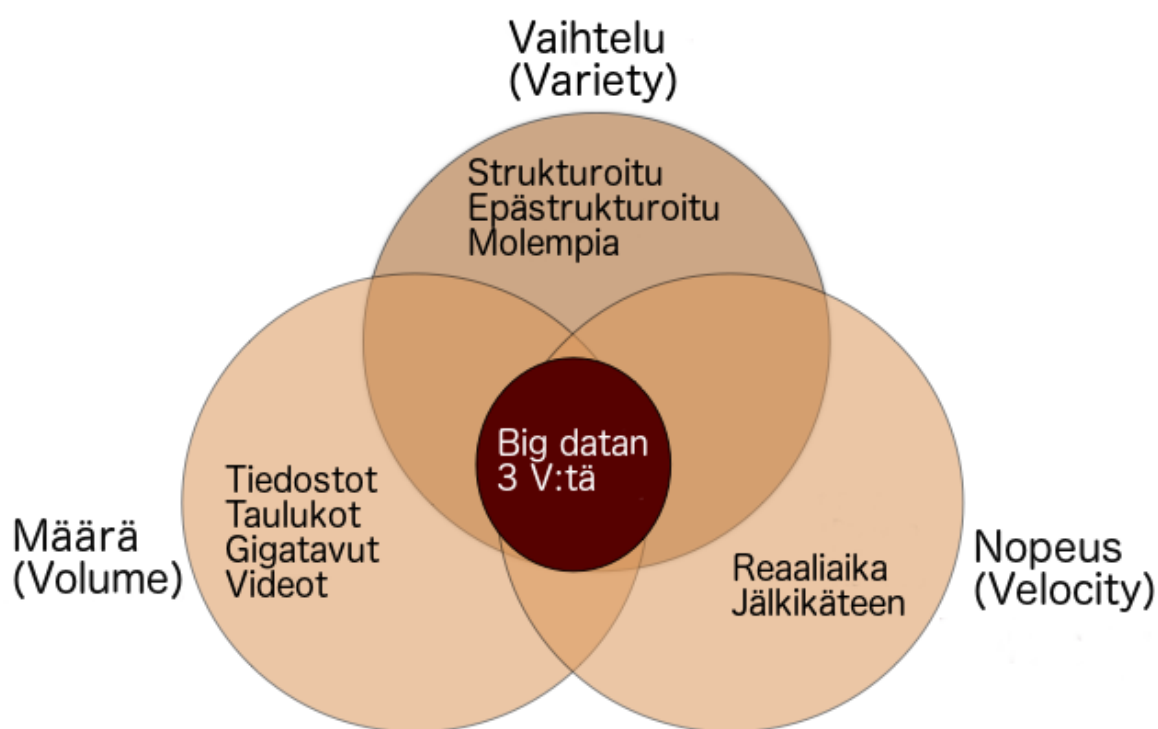
2.1 Big data

Kuten kuvasta 1 voi huomata, big datalla tarkoitetaan tietojoukkoa, jolta löytyy kolme eri ominaisuutta: määrä (volume), vaihtelu (variety) ja nopeus (velocity) (Muthanna, 2013). Yleisesti määrällä tarkoitetaan kaikkea sitä valtavaa datamäärää, joka liikkuu maailmassa tällä hetkellä teknologisten laitteiden avulla videoiden, taulukkojen, kuvien ja muiden vastaavien muodossa. Vaihtelu puolestaan viittaa suoraan datan rakenteeseen. Data voi nykypäivänä olla todella monessa eri muodossa, kuten teksti-, video-, GPS-, tai PDF-muodossa. Näistä esimerkeistä jäsennellyksi dataksi voidaan laskea esimerkiksi tekstit ja jäsen telemättömäksi dataksi videot. Datan nopeudella tarkoitetaan sen sijaan aikaa, joka datan käsittelyyn menee. Nykyisillä välineillä dataa virtaa järjestelmiin reaaliajassa, joten on erittäin tärkeää, että dataa käsitellään myös riittävän nopeasti. (Soubra, 2012)

Jalkapallossa määrällä tarkoitetaan kaikkea käytettävissä olevaa dataa. Datan määrä on yleensä erittäin suuri, jolloin datan käsittelyyn tarvitaan uusia tähän tarkoitukseen sopivia malleja ja muita analysointivälineitä. Tulevaisuudessa datan määrä tulee vain kasvamaan entisestään uusien tarkempien mittausten yleistyessä. Vaihtelu puolestaan tarkoittaa eri datatyyppejä ja lähteitä, joista dataa kerätään. Data voi esimerkiksi olla joko jäsennellyä, kuten laukaus- ja taklaustilastot, tai jäsen telemättöntä, kuten videot. Eri lähteitä, joista dataa kerätään, ovat esimerkiksi pelaajien pelipaikat, pelaajien kuntotasot sekä pelaajien taitotasot. Nopeudella viitataan datan luomis- ja käsittelynopeuteen. Osa datasta tulee reaaliajassa ja osa joudutaan

¹ STATS hankki Prozoneen vuonna 2015 omaan käyttöönsä (STATS, 2015).

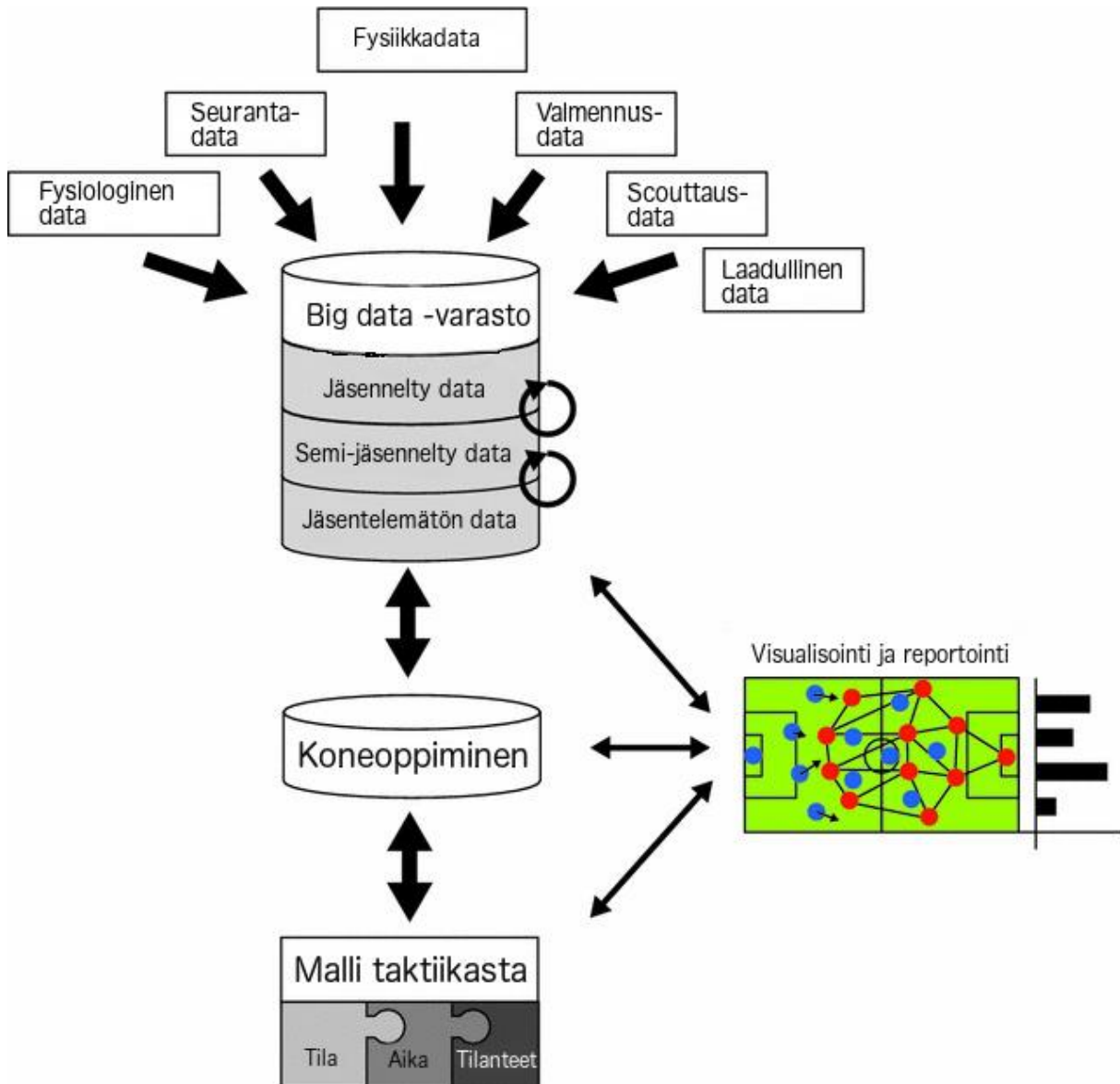
selvittämään jälkikätehen pelin jo päätyttyä. (Rein & Memmert, 2016) Jälkeenpäin datan luokitteluun on lisätty myös moninaisuus (variability) sekä arvo (value). Moninaisuudella viitataan datan arvon luokittelun laajentumiseen, eli big datalla pystytään keräämään tietoa niin pelaajien ajatuksista kuin mittareistakin. Arvolla puolestaan viitataan arvokkaaseen dataan, joka pitää löytää suuresta data joukosta. (Korhonen & Köykkä, 2016, s. 6) Käytännössä big data voidaan ajatella isoksi talletusnopeudeltaan suureksi monimuotoiseksi tietojoukoksi.



Kuva 1. Big datan 3 V:tä esitettynä Kavya Muthannan (2013) esittämän kuvan avulla.

Big data ei kuitenkaan itsessään tee jalkapallojoukkueesta suoraan muita parempaa. Erityisesti jalkapallojoukkueissa tarvitaan datan lisäksi taitavaa seuran johtamista, datan taitavaan analysointiin kykeneviä henkilöitä sekä sopivan tehokasta teknologiaa, joka pystyy organisoimaan ja hallitsemaan kerättyä dataa. Kun nämä kohdat yhdistetään riittävään data määrään, pystyvät joukkueet tekemään entistä tarkempia ennustuksia ja selvästi parempia päätöksiä kuin aikaisemmin. (McAfee & Brynjolfsson, 2012)

Robert Rein ja Daniel Memmert (Rein & Memmert, 2016) ovat ehdottaneet jalkapallossa käytettäväksi big data-järjestelmäksi kolmitasoisia mallia, joka on esitetty kuvassa 2. Sen tarkoituksena on luoda soveltuva pelitaktiikka kerätyn datan avulla.



Kuva 2. Reinin ja Memmertin (2016) luoma malli järjestelmästä, joka kehittää big datan avulla käytettävän taktiikan ja järjestelmän jalkapallojoukkueelle.

Mallin ensimmäisellä tasolla järjestelmä kerää dataa monesta eri lähteestä, kuten videoista, puettavista laitteista ja eritasoisista, osin subjektiivisista arvioista. Tämän jälkeen siirrytään toiselle tasolle, jossa hoidetaan datan tehokas varastointi ja käsittely. Tällä tasolla datan muodolla ei ole väliä, joten se voi olla jäsenneltyä (structured), jäsentelemätöntä (unstructured)

tai jotakin siltä väliltä (semi-structured). Viimeisellä eli kolmannella tasolla järjestelmä luo koneoppimisen menetelmien avulla mallin parhaiten soveltuvasta taktiikasta ja pelaajien rooleista, joiden pitäisi olla mahdollisimman toimivia annettuun tietoon perustuen. Haasteena järjestelmän luomiselle on suuri tutkimuksen ja asiantuntemuksen tarve, jotta järjestelmä ylipäätään pystytään luomaan.

2.2 Big datan sovellusalueet

Jalkapallossa kerättyä dataa käytetään moniin muihinkin eri tarkoituksiin kuin pelitaktiikan analyysiin ja kehittämiseen. Näistä tyypillisimpiä sovellusalueita ovat vedonlyönti sekä erilaiset jalkapallopelit, kuten FIFA ja Fantasy football. Näissä molemmissa sovellusalueissa käytetään oikeaa jalkapalloseuroista kerättyä dataa, jonka avulla vedonlyöntikertoimia sekä pelaajien hintoja ja taitoja määritetään kussakin erityistilanteessa. Tällä tavalla otteluista kerätyllä datalla on vaikutusta todella paljon muuhunkin, kuin suoraan seuran omaan toimintaan (Lopes-Gonzales & Griffiths, 2016).

Urheiluvedonlyönti on kehittynyt sen alkuajoista valtavasti. Nykyään vedonlyöntifirmat käyttävät big dataa urheiluun liittyvässä vedonlyönnissä etenkin kertoimien määrittämisessä. Kertoimet määräytyvät tarkasti dataa keräävien yritysten (esimerkiksi Optan) tuottamien tilastojen kuten ottelutilastojen, kotikenttäedun, kuntopuntarin, keskinäisten ottelujen, loukkaantumisten ja muiden poissaolojen tai joukkueen sisäisten ongelmien pohjalta. Tällöin on äärimmäisen tärkeää, että kaikki data otetaan mukaan päätöksentekoon, jotta firmat pystyvät toimimaan mahdollisimman kannattavasti eivätkä dataa myös entistä tehokkaammin hyödyntävät asiakkaat pysty voittamaan helppoa rahaa. (Delgado, 2017) Luonnollisesti tämä toimii myös toisinpäin, eivätkä ainakaan ammattimaiset vedonlyöjät suostu ilman vahvaa tilastollista näyttöä sijoittamaan rahojaan tarjottuihin vedonlyöntikohteisiin. Dataa kerätään myös vedonlyöntiyhtiön asiakkaista, jotta tiedetään, miten käyttäjät vedonlyöntisivuilla toimivat. Tällä tavalla vedonlyönnistä pystytään tekemään heille mahdollisimman helppoa ja sitä kautta kasvattamaan yrityksen voittoa. (Steve, 2014)

Analysoitavan dataa on siis valtavasti ja monessa eri muodossa, joten yritysten pitää käyttää itseoppivia algoritmeja, jotka luovat ja laskevat kertoimia jatkuvasti (The Economist, 2017). Luonnollisesti riskinä datan käyttämisessä on jälleen myös datan väärin tulkitseminen, jolloin kertoimet voivat mennä pahastikin pieleen. Esimerkiksi maalimääriä tarkastellessa on tärkeää

ottaa huomioon maalintekoyritysten määrä, jotta maalimäärän suuruutta tai pienuutta on helpompi ymmärtää. Tämän lisäksi urheiluun liittyvää sattumaa on hyvin vaikea ottaa huomioon vedonlyöntikertoimia suunnitellessa. Tähänkin tarkoitukseen on toki olemassa erilaisia tilastollisia menetelmiä (bayesiläinen analyysi, neuroverkot, simuloinnit, jne., kts. tarkemmin Brauchle, 2016). Mitä pidemmälle datan keruussa ja analyysissä päästään, sitä keskeisemmäksi muodostuu tiedon epätasaisen jakautuman (asymmetrian) hyödyntäminen. Vedonlyönnissä voittaa usein se, joka pystyy hyödyntämään vain harvoille tiedossa olevaa hiljaista tietoa (esimerkiksi salatuista vammoista). Seurojen itse keräämän salatun datan tullessa jollakin tavalla vedonlyöjien tai vedonlyöntiyhtiöiden käyttöön, olisi heillä mahdollisuus käyttää dataa omaksi edukseen (Delta & Matsuura, 2017). Siksi alalla onkin tarkat sisäpiirisäännökset. Tämä on kuitenkin yksi uhka, mikä etenkin vedonlyöntiyhtiöille on varsin tärkeä pelattaessa big dataa hyödyntäviä ammattimaisia vedonlyöjiä vastaan.

EA Sportsin FIFA on puolestaan jalkapallopelejä, jossa pelaajille on määritetty arvot välillä 1-99 todella monelle eri ominaisuudelle. Pelaajia on noin 18 000 kappaletta yli 700 joukkueessa, joten luonnollisesti arvoja päätettäessä etenkin alasarjajoukkueissa pelaaville pelaajille, täytyy ihmisten omaa päättelyä käyttää osana arvostelua. Pelaajien perustaso määräytyy sarjan mukaan, minkä jälkeen noin 9 000 datan kerääjää, joista useimmat ovat tietyn joukkueen kausikortin haltijoita ja pienempi osa palkattuja scoutteja, lähettävät EA:lle tietoja joukkueiden pelaajista EA:n luoman palvelun kautta. Lopulta EA saa valtavan määrän dataa, jonka jälkeen sen pitää vielä huomioida joukkueiden pelitavat virheellisyyksien estämiseksi. Esimerkiksi pallonhallintaan pelitapansa perustavat Manchester City sekä FC Barcelona syöttelevät pelien aikana todella paljon. Tämä pitää huomioida, jotta kyseisten joukkueiden pelaajat eivät saa liian hyviä arvoja syöttöihin liittyen vain joukkueen pelitavan ansiosta. Tiettyjen pelaajien arvoja pitää vastaavasti myös parantaa, jotta ne vastaavat pelaajan oikeita kykyjä kentällä. Esimerkiksi Bayern Münchenin Thomas Müllerillä ei ole varsinaisia huippuominaisuuksia, mutta hän tekee kuitenkin suuria maalimääriä ja on joukkueensa arvokkaimpia pelaajia. Pelissä tällaista on vaikea toteuttaa, joten Müllerin arvoja pitää parantaa, jotta hän on edes lähes yhtä hyödyllinen pelissä kuin oikeasti kentällä. (Lynch, 2016)

Erilaisissa fantasy football -peleissä, kuten Fantasy Premier League, pelaajien tehtävänä on valita itselleen noin viisitoista pelaajaa, joiden avulla he keräävät pisteitä. Pelaajat on hinnoiteltu eriarvoisiksi sen mukaan, kuinka paljon niiden odotetaan keräävän pisteitä kauden

aikana. Luonnollisesti aikaisemmillä kausilla on suuri vaikutus pelaajien pisteelykseen. Pisteitä pelaajat saavat maaleista, syötöistä sekä nollapeleistä ja pisteitä menetetään puolestaan keltaisista ja punaisista korteista sekä päästetyistä maaleista. (Billings et al., 2014) Tulevaisuudessa pisteely voi muuttua monipuolisemmaksi ja siihen voivat vaikuttaa jopa riistot ja sijoittuminen ja muut vaikeasti arvosteltavat tilastot. Joissain sovelluksissa tämä on jo käytössä, mutta se ei kuitenkaan ole vielä kovin yleistä. Pelaajat, jotka haluavat menestyä fantasy-peleissä mahdollisimman hyvin, hyödyntävät saatavilla olevaa dataa ja voivat käyttää apuna jopa tiettyjä algoritmeja (McCue, 2012). Tällä tavalla pelaajat saavat hyödyllistä infoa, kuten ketä heidän kannattaa kyseisellä kierroksella peluuttaa. Big datan ja sen hyödyntämisen yleistyessä entisestään, muuttuvat varmasti fantasy-pelitkin huomattavasti laajemmiksi ja tarkemmiksi. Fantasy-pelejä pidetäänkin jopa toivottuna vedonlyönnin korvaajana tulevaisuudessa, sillä ne todennäköisesti vähentäisivät sopupelien määrää, loisivat paremman kokemuksen faneille ja olisivat nuorille sopivampi tapa harrastaa kisailua urheilun välityksellä (The Future of Sports, 2016).

Taulukko 1. Big datan sovellusalueet vertailtuna keskenään.

	Vedonlyönti	Videopelit	Fantasy-pelit
Mistä dataa kerätään?	Yritykset (kuten Opta), käyttäjät	Yritykset (kuten Opta), Scoutit (subjektiiviset arviot), käyttäjät	Yritykset (kuten Opta), aikaisempien vuosien tilastot, käyttäjät
Mihin dataa käytetään?	Kertoimien määrittämiseen	Pelaajien arvojen määrittämiseen	Pelaajien hintojen määrittämiseen, pelaajien pisteelykseen, pelaajien valintaan
Mitä riskejä datan käyttö voi aiheuttaa?	Data voi aiheuttaa epäreilua vedonlyöntiä, jolloin yritykset tai pelaajat voivat kokea tulleensa huijatuiksi	-	Data voi tehdä pelistä epäreilua

Taulukosta 1 voi huomata, miten big datan käyttäminen eroaa sen sovellusalueissa. Dataa tuottavat yritykset ovat lisääntyneet hurjaa vauhtia, joten niiden avulla tapahtuu suurin osa datan hankkimisesta. Datan käyttäminen on video- ja fantasy-peleissä samantyylistä, eli arvojen ja pisteitten määrittämistä, mutta vedonlyönnissä dataa käytetään puolestaan kertoimien määrittämiseen. Suuria riskejä ei data videopelimaailmassa aiheuta, mutta fantasy-peleissä joukkueiden sisäisen datan avulla on mahdollista voittaa arvokkaitakin palkintoja. Vedonlyönnissä joukkueiden sisäpiireistä saadun datan käyttäminen omaksi hyödyksi on kuitenkin selkeästi suurin uhka etenkin vedonlyöntiyhtiöille.

2.3 Datan keräysmenetelmät

Jalkapallossa kerätään dataa useilla erilaisilla menetelmillä ja välineillä, jotka sopivat erilaisiin tilanteisiin. Jokaisella tiedonkeräysmenetelmällä on omat vahvuutensa ja heikkoutensa.

2.3.1 Videoteknologia

Videoteknologia on yksi merkittävimmistä apuvälineistä datan keräämisessä. Videoteknologia ei kuitenkaan tarkoita pelkästään kameraa ja sen avulla tuotettua videokuvaa, vaan myös ohjelmistoa, joka käsittelee videokuvaa. Eniten videoteknologiaa käytetään apuna tilastojen ja paikkatietojen tuottamisessa siten, että sen avulla saadaan tietoa sekä peleistä että harjoituksista. Yhdestä Bundesliiga-kaudesta syntyy videodataa noin 400 GB:tä, minkä lisäksi seurat käsittelevät itse keräämäänsä dataa harjoituksista, loukkaantumisista ja muista lähteistä (Rein & Memmert, 2016).

Käytettäessä videoteknologiaa datan keräämisen, koko pelialueen täytyy olla katettu kameroiden avulla. Tällä tavalla jokaista pelaajaa pystytään seuraamaan jatkuvasti ja tilastoista tulee tarkkoja. Tämä aiheuttaa kuitenkin myös haasteita videoteknologiaa tuottaville yrityksille, koska käytettävien kameroiden lukumäärä täytyy nostaa erittäin korkealle. Tämän lisäksi pelaajien tunnistaminen voi olla hankalaa samanlaisten peliasujen ja epäselvän kaukaa otetun kuvan takia.

Pelaajien paikkatietoja seurataan erilaisten tracking-algoritmien avulla. Nämä algoritmit ovat nykyään jo melko tarkkoja, mutta eivät kuitenkaan täydellisiä. Esimerkiksi kontaktitilanteet ja pelaajien sekoittaminen tuomariin tai katsojaan aiheuttavat haasteita algoritmeille. (D'orazio & Leo, 2010) Virhemahdollisuuden takia osa pelaajien seurannasta pitää edelleen tehdä

manuaalisesti, mikäli halutaan täysin varma tulos ja kuva siitä, miten tietty pelaaja on liikkunut tietyllä hetkellä (Lu et al., 2013).

2.3.2 Puettavat laitteet

Puettavilla laitteilla tarkoitetaan erilaisia välineitä, kuten mittareita ja vöitä, joita urheilijat käyttävät suoritustensa aikana. Helposti erilaisilla laitteilla mitattavia asioita ovat esimerkiksi syke, juostu matka, nopeus, kiihtyvyyys ja hapenottokyky.

Yleisiä otteluissa ja harjoituksissa käytettäviä laitteita ovat esimerkiksi älykellot ja sykevyöt. Tämän lisäksi dataa kerääviä siruja voidaan laittaa esimerkiksi jalkapallokenkiin, kuten Adidas teki miCoach-järjestelmän kanssa. Tärkeintä laitteiden käytettävyyteen liittyen kuitenkin on, että ne eivät häiritse käyttäjää millään tavalla. Nykyään myös vaatteisiin, kuten alushousuihin, on mahdollista liittää sensoreita, joiden avulla pystytään tarkkailemaan urheilijan lihasten aktivointia. Vieläkään puettavien laitteiden tarkkuus ei ole täydellinen, joten pelkästään niiden avulla toimiminen voi saattaa pelaajat myös suureen riskiin, koska esimerkiksi väärät arvot liittyen palautumiseen voivat johtaa vääranlaiseen harjoitteluun (The Future of Sports, 2016).

Erilaisia sensoreita, joita laitteissa käytetään, ovat muun muassa GPS, kiihtyvyyssanturi, kompassi ja sykesensori. Näiden lisäksi laitteisiin tarvitaan lähetin, joka toimii esimerkiksi bluetoothin välityksellä. GPS on seurantajärjestelmää edullisempi, mutta myös epätarkempi ratkaisu.

Harjoitusten lisäksi seurat voivat seurata pelaajien toimimista harjoitusten ja otteluiden ulkopuolella unen, askeleiden ja palautumisen kautta. Tämä voi kuitenkin joidenkin pelaajien mielestä olla jo liioittelua ja aiheuttaa ongelmia vapaa-ajan valvomisen seurauksena.

Suurin hyöty puettavilla laitteilla on kuitenkin harjoitusten yksilöllisessä suunnittelussa. Kun valmentajalla on ryhmänsä avulla tiedot pelaajien palautumisesta ja kunnosta tietyllä hetkellä, on heidän huomattavasti helpompi päättää, millä tavalla esimerkiksi tuleva viikko harjoitellaan. Tällä tavalla teknologia voi auttaa vähentämään liiallisesta rasituksesta syntyviä loukkaantumisia. (Johansen et al., 2015)

2.3.3 Laadullinen tieto

Videoteknologia sekä puettava teknologia ovat määrällisen ja objektiivisen tiedon keräystä varten käytettyjä menetelmiä, mutta laadullisella tiedollakin on yhä paikkansa urheiluseurojen toiminnassa. Laadullisella tiedolla tarkoitetaan tiettyjen henkilöiden subjektiivisia mielipiteitä, arvosteluja, pelaajien mielipiteitä sekä ajatuksia, joita löytyy internetistä ja etenkin sosiaalisesta mediasta. Pelaajat voivat tehdä esimerkiksi julkaisuja tai haastatteluja, joista voidaan nykytekniikan avulla päätellä pelaajan sen hetkisestä kunnosta, ajattelutavasta ja asenteesta erilaisia asioita.

2.4 Datan hyödyntämistavat

Nykyään dataa hyödynnetään jalkapallossa moniin eri tarkoituksiin, kuten pelaajien tarkkailuun, ottelujen ja harjoitusten analysointiin sekä vastustajien tarkkailuun. Jokaisella osaluueella pyritään olemaan parempia kuin kilpailevat seurat, jotta itselle saadaan pienikin etu vastustajia vastaan. Tärkeintä on löytää datajoukosta se osa, jolla on eniten arvoa ja hyödyntää sitä siten, että pystytään luomaan tai päättelemään jotakin arvokasta (Rudenko, 2017).

2.4.1 Scouttaaminen

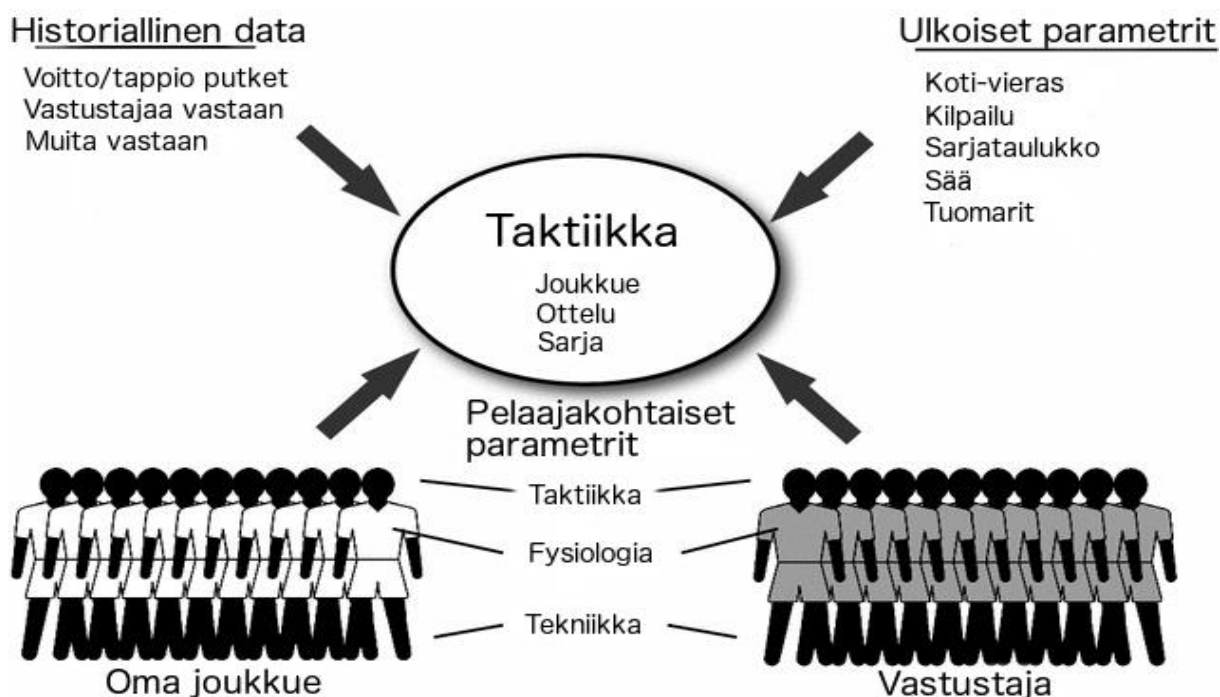
Pelaajien scouttaaminen liittyy vahvasti seuran taloudelliseen puoleen, sillä uusien hankintojen tehtävänä on tuottaa seuralle voittoa muutaman vuoden sisällä. Joukkueet pyrkivät aluksi selvittämään, minkälaista pelaajatyyppeä he tarvitsevat ja ketkä mahdollisista pelaajista täyttävät heidän vaatimuksensa. Mitä korkeamman tason pelaajaa seura etsii, sitä vähemmän kilpailijoita ja sitä suppeampi joukko mahdollisia kandidaatteja heillä on scoutattavanaan huippupelaajien kalliin hinnan ja korkean taitotason takia. Alemmilla sarjatasoilla sopivia pelaajia taas on puolestaan enemmän saatavilla, mutta kilpailijoiden määrä kasvaa samassa suhteessa samalla, kun epäonnistumisen riskit lisääntyvät pelaajien taitotason vaihdellessa. (Foster, 2016) Seuroilla on pelaajatyypin lisäksi muitakin ehtoja, joiden he odottavat täyttyvän. Esimerkiksi ikä, kansalaisuus, sopeutuvaisuus, potentiaali ja sopimustilanne voivat olla tietyissä tilanteissa erilaisten sääntöjen takia erittäin merkittäviä tekijöitä, jotka nostavat tai laskevat tietyn pelaajan hintaa (Yuan, 2013, s. 3-7). Pelaajien ja heidän ominaisuuksiensa valtavan määrän takia dataa on nykyään pelkästään pelaajista valtavasti tarjolla ja seurojen pitää hankkia uusia apuvälineitä tai henkilöitä käyttöönsä datan tulkintaa varten. Esimerkkinä tällaisen toiminnan keskeisyydestä mainittakoon, että vuonna 2012 Arsenal osti StatDNA-

nimisen firman, jonka tehtävänä on uusien sopivien pelaajien tunnistaminen, peleihin valmistautuminen ja ottelujen analysointi. (Hytner, 2014)

2.4.2 Otteluiden ja harjoitusten analysointi

Otteluja ja harjoituksia käydään tarkasti päivittäin läpi jokaisessa joukkueessa. Esimerkiksi Saksan A-maajoukkue käytti vuonna 2014 Brasiliassa järjestetyissä jalkapallon MM-kisoissa Adidaksen miCoach-järjestelmää, jonka avulla Saksa pystyi seuraamaan jokaisen pelaajan suoritusta, erityisesti sykkeitä, nopeutta, kestävyyttä, kiihtyvyyttä ja voimaa, jokaisessa harjoituksessa. Tällä tavalla valmennuksen oli helppo huomata, minkälaista harjoitusta kukin pelaaja tarvitsi vai oliko lepo palautumisen kannalta paras ratkaisu. Analyysin huipputasolle vienyt Saksa voittikin vuonna 2014 maailmanmestaruuden. Nykyään on fyysisten ominaisuuksien lisäksi mahdollista seurata jopa tiettyjä henkisiä ominaisuuksia erilaisten anturien avulla. (Hymers, 2014) Italialainen jalkapalloseura AC Milan käyttää puolestaan harjoituksissaan ohjelmistoa, joka havaitsee pelaajien liikkeiden ja toimintojen kautta myös mahdolliset loukkaantumiset, joita pelaajat eivät joko itse halua myöntää tai eivät edes huomaa (Schumaker et al., 2010, s. 56).

Otteluiden analysointi etu- ja jälkikäteen on yksi merkittävimmistä osa-alueista, joissa dataa käytetään nykyjalkapallossa hyödyksi. Kuten kuvasta 3 näkee, on taktiikkaa suunniteltaessa otettava huomioon monia erilaisia asioita, joihin melkein jokaiseen saa apua ja tietoa datan muodossa. Vastustajan joukkueen jokaisen pelaajan vahvuudet ja heikkoudet, niin fyysiset kuin henkiset, käydään läpi ja vastustajajoukkueen taktiikka opetellaan tarkasti, jotta siihen voidaan vastata mahdollisimman tehokkaasti. Joissakin tapauksissa jopa eri pelaajien harhautukset, kääntymis- ja syöttösuunnat jne. otetaan tarkasti huomioon, kun pelitaktiikkaa suunnitellaan. Seurat tutkivat myös aikaisempia otteluita ja huomioivat pelataanko ottelu joukkueen koti- vai vieraskentällä, minkälainen sää pelin aikana on ollut sekä miten toinen joukkue on pelannut tulevaa vastustajaa vastaan ja miten se on pärjännyt. Lisäksi valmentajan pitää miettiä, miten hän vastaa vastustajan vahvuuksiin ja iskee mahdollisiin heikkouksiin mahdollisimman onnistuneesti. (Rein & Memmert, 2016) Tässä kaikessa suunnittelussa käytetään apuna juuri StatDNA:n tapaisia ohjelmistoja, jotka sisältävät ja käsittelevät datan toivotulla tavalla.



Kuva 3. Reinin ja Memmertin (2016) hahmottelema kuva taktiikan luontiin vaikuttavista tekijöistä.

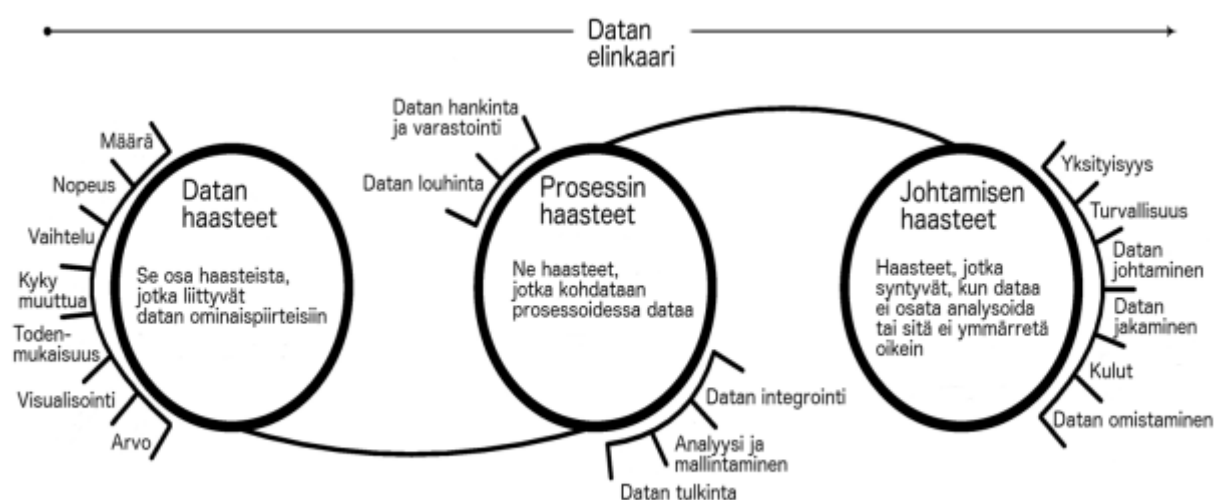
Otteluiden jälkeen pitää puolestaan miettiä, mikä onnistui ja missä jäi parantamisen varaa. Tätä varten StatDNA:n ja InStatin tapaiset ohjelmistot käyvät videoita läpi ja merkkäävät jokaisen kentällä käyneen pelaajan liikesuunnat, syöttösuunnat, pallonmenetykset, pallonriistot, keskimääräisen sijainnin kentällä, rikkeet ja kaiken muun, mitä jollakin tavalla vain voidaan mitata. Tällaisen analyysin kautta datan tulkinnasta vastaavat ryhmät miettivät, onnistuivatko ennen ottelua sovitut asiat sillä tavalla kuin oli sovittu ja mikäli eivät, mistä se johtui. Tällä tavalla seura pystyy seuraavalla kerralla suoriutumaan paremmin tekemällä vaadittuja muutoksia niin pelitapaan kuin pelaajien peluuttamiseenkin.

2.5 Datan aiheuttamat riskit ja ongelmat

Liiallinen datan käyttö voi kuitenkin mahdollisesti aiheuttaa myös haasteita joukkueelle, sillä kaikkia tärkeitä taitoja ei pysytä mittaamaan etenkin jalkapallon kaltaisessa lajissa, jossa yhteistyö monien pelaajien välillä suurella pelikentällä on poikkeuksellisen tärkeää. Ominaisuudet, kuten yhteistyökyky ja omistautuneisuus ovat esimerkiksi haastavia asioita tutkia datan avulla, joten niitä joudutaan (ainakin vielä) usein ratkomaan oman vaiston perusteella. Lisäksi tilastot saattavat ajoittain antaa väärän tai ainakin yksipuolisen kuvan tietystä pelaajasta. Tästä hyvänä esimerkkinä on laajasti käytetty WhoScored-datapankki, joka

painottaa erityisen paljon juuri maalitilanteiden rakentamista ja esimerkiksi pääpelivoittoja antaen näin usein harhaanjohtavan tai ainakin puutteellisen kuvan pelaajien arvosta joukkueelle. Toisena esimerkkinä voidaan mainita italialainen puolustaja Paolo Maldini, joka sai aikanaan kritiikkiä valmentajaltaan, koska hän osallistui kaksinkamppailuihin otteluiden aikana liian harvoin. Todellisuudessa Maldinin pelityyli vain oli erilainen kuin useammalla muulla puolustajalla ja hän pystyi peliä ennakoimalla välttämään kaksinkamppailut ja voittamaan tilanteet joukkueelleen. (Brauchle, 2016) Joukkueiden erilaiset pelityylit aiheuttavat myös haasteita taktiikkaa luotaessa. Toinen joukkue pärjää hyvin pitämällä palloa hallussaan, mutta toinen joukkue puolestaan hyötyy siitä, kun he saavat keskittyä puolustamiseen ja vastaiskuihin. Tämä onkin myös yksi syy, miksi jalkapallojoukkuetta on todella vaikea valmentaa ja johtaa pelkän datan avulla, sillä mahdollisia keinoja menestyä on niin monia ja ne voivat toimia erityyillisillä pelaajilla, erilaisella säällä tai erikokoisella kentällä niin monella eri tavalla.

Väärin tulkitsemisen lisäksi vaarana on, datan joutuminen väärin käsiin. Nykypäivänä kun dataa löytyy kaikesta aina pelaajien nukkumisesta harjoituksiin ja otteluihin asti, on erityisen tärkeää, että se pysyy aina oikeiden ihmisten käytössä. Datan avulla kilpailijat voisivat perustellusti pienentää tarjouksiaan pelaajista, joiden suoritukset eivät olekaan niin hyviä kuin on annettu ymmärtää. Esimerkiksi jos pelaaja kuntoutuu erityisen hitaasti loukkaantumisen jälkeen takaisin pelikuntoon, maksetaan hänestä vähemmän. Tämän lisäksi vastustajat voisivat saada yksityiskohtaista tietoa taktiikoista, jolloin heidän taktiikkansa olisivat helpommin suunniteltavissa tehokkaaksi.



Kuva 4. Sivarajahin et al. (2017) kuva big datan aiheuttamista haasteista.

Riskien lisäksi itse big datan ominaispiirteet voivat aiheuttaa haasteita niiden käsittelyssä. Datamäärät ovat luonnollisesti valtavia, jolloin niiden käsittely voi olla haastavaa. Tähän liittyen datan hyödyntämiseen sopivien mallien kehittäminen ja käyttäminen voi olla erityisen hankalaa, etenkin mikäli sopivaa valmista mallia ei ole saatavilla. Suurimpana yksittäisenä haasteena big dataa hyödyntäville yrityksille pidetäänkin vaadittavan infrastruktuurin korkeita kustannuksia, vaikka nykyään pilvipalveluiden avulla hintoja pystytäänkin painamaan hieman alaspäin.

Kuvasta 4 voi muun muassa havaita, kuinka big datan aiheuttamat haasteet muuttuvat erilaisiksi datan siirtyessä omassa elinkaarensa eteenpäin. Karkeasti big datan haasteet voidaan jakaa kolmeen kategoriaan, jotka ovat itse datan ominaisuuksien aiheuttamat haasteet, datan käsittelyyn liittyvät haasteet ja datan käytön hallintaan ja johtamiseen liittyvät haasteet. Ensimmäisessä kategoriassa esille nousevat jälleen jo aikaisemmin esille nostetut kolme V:tä (kts. s. 7-8), mutta myös monta muuta big datalle ominaista piirrettä. Haasteita aiheuttavat datan määrän, sen vaihtelun ja talletusnopeuden lisäksi muun muassa datan arvon määrittäminen, eli onko data arvokasta vai ei, sekä datan visualisointi. Visualisoinnin ollessa haasteellista voi olla vaikea ymmärtää, mihin kerätty data liittyy ja mitä sillä tarkoitetaan. Toisessa eli datan käsittelyvaiheessa puolestaan datan kerääminen, integrointi, muuntaminen ja oikean mallin valitseminen analyysia varten voivat aiheuttaa suuria vaikeuksia. Viimeisessä kategoriassa eli datan hallinnassa ja johtamisessa haasteita syntyy esimerkiksi kulujen hallinnasta, yksityisyyden sekä datan suojelemisesta. Nämä ovat kaikki ongelmia, joita kohdataan lähes jokaisessa liiketoiminnassa riippumatta siitä, käytetäänkö big dataa apuna vai ei. (Sivarajah et al., 2017)

3 CASE SUOMEN KAKKOSDIVISIOONA

Tässä case-tutkimuksessa käydään läpi viiden suomalaisen kakkosdivisioona-joukkueen datan hyödyntämismenetelmiä ja mielipiteitä datan hyödyllisyydestä ja sen käytön lisääntymisestä. Kyselyssä esitetään kahdeksan kysymystä, joihin osallistuvat joukkueet vastaavat joko kirjallisesti tai sanallisesti. Tutkimuksen tarkoituksena on verrata, miten toimintatavat ja mielipiteet eroavat saman sarjan sisällä. Joukkueiden nimiä ei kandidaatintyössä luottamuksellisuuden takia mainita, joten joukkueista käytetään nimiä Joukkue 1, Joukkue 2, Joukkue 3, Joukkue 4 ja Joukkue 5.

3.1 Kysymykset

Kysymyksiin vastaamisesta vastasivat pääasiassa seurojen valmentajat. Mukana oli kysymyksiä liittyen valmentajien asenteisiin dataan liittyen ja seuran toimintatapoihin.

Kysymykset, jotka kyselyssä esitettiin, olivat seuraavat:

1. Onko datan käytön lisääntyminen mielestänne hyvä vai huono asia ja miksi?
2. Millä tavalla dataa käytetään apuna joukkueenne urheilullisen toiminnan tehostamisessa? (esim. scouttaus, otteluanalysointi, harjoitukset, loukkaantumiset)
3. Mitä ylläolevista pidätte tärkeimpänä ja miksi?
4. Mitä konkreettisia vaikutuksia datan käyttö on joukkueelle aiheuttanut?
5. Minkälaisia mittareita tai laitteita käytätte datan keräämisessä?
6. Tekeekö seuran ulkopuolinen yritys yhteistyötä joukkueen kanssa datan analysoinnissa? Jos tekee, niin mikä ja miten?
7. Mitä riskejä ja ongelmia datan käyttöön mielestänne liittyy?
8. Millä tavalla uskotte datan käytön ja sen hyödyntämisen kehittyvän tulevina vuosina?

3.2 Vastaukset

Tässä kappaleessa käydään läpi seurojen vastauksia ylläoleviin kysymyksiin. Kysymykset ja niiden vastaukset käydään yksitellen läpi.

Onko datan käytön lisääntyminen mielestänne hyvä vai huono asia ja miksi?

Datan ja sen käytön lisääntymistä pidettiin hyvänä asiana jokaisessa vastauksessa. Data muun muassa vähentää mutu-mielipiteitä esittämällä tilalle faktoja, helpottaa pelin rakenteellisen

puolen, hyökkäyspelin ja puolustuspelin, analysointia, kehittää peliä eteenpäin, luo uusia ajatuksia sekä helpottaa pelaajien kuormituksen seuraamista ja harjoittelun rytmittämistä. Datan tulkintaa ja sen hyödyntämistä siten, että toiminta oikeasti kehittyy, pidettiin kuitenkin haastavana. Haastateltavien mielestä suurta datamäärää oli vaikea analysoida muun muassa osaamisen puutteen vuoksi.

Vastaajien mielipiteet liittyen datan käyttöön voidaan tiivistää kolmeen pääkohtaan:

- Vähentää mutu-mielipiteitä tuomalla tilalle faktoja
- Helpottaa pelin rakenteellisen puolen analysointia
- Datan käsittely on haastavaa

Millä tavalla dataa käytetään apuna joukkueenne urheilullisen toiminnan tehostamisessa?

Dataa käytettiin apuna useassa eri osa-alueessa. Jokaisessa vastauksessa nostettiin esille otteluanalysointi, eli omien otteluiden läpikäynti sekä pelaajien kehittäminen niin fyysisesti kuin taktisestikin. Otteluanalysoinnissa tärkeinä asioina pidettiin omien avainasioiden seuraamista ja maalipaikkojen ja niiden laadun arvioimista. Pelaajien kehittämisessä ja seuraamisessa data oli käytössä erilaisten fyysisten testien muodossa. Fyysisiä ominaisuuksia, joita testattiin, olivat nopeus, voima, kestävyys, kuormitus ja kiihdytysten, jarrutusten sekä spurttien määrä. Muita osa-alueita, joissa dataa käytettiin apuna, olivat vastustajan scouttaus, kuormituksen seuraaminen sekä uusien mahdollisten pelaajien scouttaus. Datan käyttöä pidettiin myös harjoittelua helpottavana tekijänä, sillä etenkin harjoituskaudella otteluanalyysin avulla pystyttiin arvioimaan, miten harjoitellut kuviot oli viikon aikana opittu. Datan avulla uusien asioiden oppimista oli huomattavasti helpompi tarkastella, etenkin videoiden avulla. Yksi seuroista olikin juuri hankkinut uuden videoanalyytikon tuottamaan videodataa seuralle.

Millä tavalla haastateltavat hyödynsivät dataa joukkueissaan:

- Avainasioiden seuraaminen ja pelaajien kehittäminen otteluanalysoinnin kautta
- Pelaajien kehittäminen fyysisten testien ja kuormituksen mittaamisen kautta
- Datan avulla asioiden oppimista on helppo tarkastella ja seurata

Mitä ylläolevista pidätte tärkeimpänä ja miksi?

Tähän kysymykseen esille nostettiin kaksi asiaa, otteluanalysointi sekä pelaajien/joukkueen kehittäminen. Nämä luonnollisesti liittyvät vahvasti toisiinsa, joten joukkueet olivat hyvin samoilla linjoilla. Otteluanalysointi auttaa havaitsemaan joukkueen kehittymisen, joka etenkin kauden ollessa kesken on erityisen tärkeää. Pelaajakehitystä ja joukkueen kehittämistä pidettiin myös tärkeänä joukkueen strategian toteuttamisessa. Pelaajien kehityksessä tärkeintä oli harjoitusten kuormittavuuden mittaaminen, jolloin pelaajan ja joukkueen harjoittelua on helppo muokata sopivaksi. Pelaajien yksittäisten suoritusten arviointiin dataa ei vielä ollut kovin paljoa käytetty. Pääosin tähän oli syynä ajan, vertailutilaston sekä mahdollisen osaamisen puute.

Mitä vastaajat pitivät tärkeimpinä datan käyttökohteina:

- Otteluanalysointi, koska se helpottaa joukkueen kehityksen seurantaan
- Pelaajien kehittäminen, koska se kuuluu joukkueen strategiaan

Mitä konkreettisia vaikutuksia datan käytöstä on joukkueelle syntynyt?

Konkreettisia vaikutuksia datalla on ollut etenkin tekniseen ja fyysiseen harjoitteluun sekä niiden kuormittavuuteen. Harjoitteluun on muun muassa tuotu mukaan useita pienempiä yksityiskohtia, kuten joukkueen tasapainon, heikkouksien ja vahvuuksien havainnointi. Datan avulla on pystytty myös ohjeistamaan yksittäisiä pelaajia aiempaa tarkemmin ja yksityiskohtaisemmin. Haastateltavien mukaan yksittäisten pelaajien ymmärrys omasta roolistaan osana joukkueen kokonaisuutta on kasvanut ja motivaatio suorittaa vaaditut tehtävät kunnolla on niin ikään lisääntynyt. Ymmärrys myös jalkapallon ulkopuolisiin asioihin, kuten rasitukseen, lepoon, ravintoon, ja vapaa-aikaan, on myös kehittynyt datan avulla. Sykeseurannan kautta on pystytty ehkäisemään loukkaantumisia ja pelisuunnitelmia on kehitetty paremmiksi datan avulla.

Konkreettiset datan käytön vaikutukset voidaan tiivistää kolmeen pääkohtaan:

- Harjoitteluun on tuotu pieniä yksityiskohtia mukaan
- Yksittäisen pelaajan ymmärrys pelistä ja urheilusta on kehittynyt
- Pelisuunnitelmaa on kehitetty toimivammaksi

Minkälaisia mittareita tai laitteita käytätte datan keräämisessä?

Erilaisia tapoja mitata ja kerätä dataa oli useita. Yleisimpiä olivat videointi, fyysiset mittarit (Firstbeat, Polar) sekä itsearviointi. Firstbeat oli käytössä pelaajien palautumisen tarkkailussa, jota se mittasi kahden pelaajaan kiinnitettävän sensorin avulla. Myös sovellukset (Quanter, Soccermeter, Wyscout) mainittiin kolmessa vastauksessa. Quanter perustuu pelaajien subjektiiviseen arvioon harjoitusten kuormittavuudesta ja omasta hyvinvoinnista, Soccermeter puolestaan auttaa keräämään dataa otteluista ja Wyscout sisältää valtavan määrän dataa pelaajista ja joukkueista. Tässä kyselyssä Wyscouttia käyttänyt seura käytti sitä etenkin pelaajien scouttauksessa. Myös kehonkoostumusmittauksia suoritettiin yhdessä joukkueessa. Taulukko 2 esittää haastateltavien joukkueiden datankeruun lähteet sekä data-analyysin painopisteet (joukkue/pelaaja). Huomioitavaa on, että viidessä seurassa toimitaan pitkälti samalla tyylillä, vaikka dataa tuottavissa ohjelmissa onkin selkeämpiä eroja. Suurin osa datan keräämisestä liittyy yksittäiseen pelaajaan, mikä on toki ymmärrettävää, sillä yksilöä on huomattavasti joukkuetta helpompi mitata monipuolisesti. Yllättävää oli, että dataa ei kuitenkaan vastausten perusteella käytetty kuin kahdessa joukkueessa (Joukkue 2 ja Joukkue 3) hyödyksi yksittäisten pelaajien arvioinnissa. Joukkue 1 antoi tähän hyvän perustelun: ”Emme ole vielä käyttäneet dataa pelaajien suoritusten arviointiin vaan se perustuu pitkälle subjektiivisiin näkemyksiin. Tämä johtuu pitkälle siitä, että aika ei yksinkertaisesti riitä, mutta myös meiltä ei löydy riittävästi vertailutilastoa eikä välttämättä myöskään osaamista, koska kukaan ei ole toistaiseksi keskittynyt pelkästään datan käyttöön.” Muissa joukkueissa yksittäisiä pelaajia arvioitiin subjektiivisesti videon perusteella.

Taulukko 2. Joukkueiden datankeruu ja data-analyysin painopisteet vertailtuna.

Joukkue	Datankeruu	Data-analyysin painopiste (joukkue/pelaaja)
Joukkue 1	Fyysiset testit (Polar), kuormitus (Quanter), Wyscout, videot	Pelaajille fyysiset testit, kuormitusmittaukset sekä Wyscout ja videolta analysoidaan joukkueen peliä
Joukkue 2	Fyysiset testit, kuormitus (Firstbeat), videot (InStat)	Pelaajille fyysiset testit, kuormitusmittaukset sekä otteluanalyysi ja videolta analysoidaan joukkueen peliä sekä yksittäisten pelaajien suorituksia
Joukkue 3	Fyysiset testit (Aikasyke/lääkäriasema), videot, itsearviointi (Excel)	Pelaajille fyysiset testit, kuormitusmittaukset, otteluanalyysi sekä itsearviointi ja videolta analysoidaan joukkueen peliä sekä yksittäisten pelaajien suorituksia
Joukkue 4	Fyysiset testit, videot (Soccermeter)	Pelaajille fyysiset testit sekä otteluanalyysi ja videolta analysoidaan joukkueen peliä
Joukkue 5	Fyysiset testit (Polar), kehonkoostumusmittaus, videot	Pelaajille fyysiset testit sekä kehonkoostumusmittaukset ja videolta analysoidaan joukkueen peliä

Tekeekö seuran ulkopuolinen yritys yhteistyötä joukkueen kanssa datan analysoinnissa?

Jos tekee, niin mikä ja miten?

Haastateltavat manitsivat Aikasykkeen, Lääkäriaseman ja InStatin yhteistyökumppaneikseen datan analysoinnissa. Aikasykkeen Firstbeat- menetelmää käytettiin apuna kahden seuran kuormituksen ja palautumisen mittaamisessa ja paikallinen lääkäriasema oli puolestaan yhden seuran apuna terveyden tarkkailussa. InStat tuotiin esille kahdessa vastauksessa, mutta käytössä se oli vain yhdellä. InStatin tehtävänä oli purkaa video pelin eri osa-alueille, purkaa yksittäisten pelaajien toiminnat otteluissa ja analysoida vastustajien peliä. Toinen seura oli puolestaan lopettanut InStatin käytön ja siirtänyt siihen kuluneet resurssit muualle kahdesta eri syystä. Ensimmäisen ongelma InStatissa oli se, että otteluanalyysin tekijä ei välttämättä näe tilanteita samalla tavalla kuin analyysiä haluava seura, erityisesti vastuuvallmentaja. Esimerkiksi 1vs1-tilanteet, kaksinkamppailut ja maalipaikat voidaan tulkita monella eri tavalla, jotka voivat erota toisistaan hyvinkin paljon. Vastauksessa mainittiin myös ongelman esiin tuominen myös

huippuseurojen edustajien puolesta. Toinen haaste InStatiin liittyen on sen raportin laajuus. Raportin purkaminen on haastavaa ja aikaa vievää, joten raportin kunnolliseen analysointiin tarvitsisi aina yhden päätoimisen henkilön. Nämä vastauksessa esitetyt ongelmat kuvaavat hyvin aikaisemmin tässäkin työssä esitettyjä datan analysoinnin haasteita. Tämä oli osittain datan käsittelyyn liittyvä haaste, mutta myös yleisesti hallinnallinen haaste, joka seuran piti pystyä ratkaisemaan.

Mitä hyötyä seuran ulkopuolisista yrityksistä on joukkueille:

- Joukkueet saavat käyttöönsä kuormituksen ja palautumisen mittaamiseen tarvittavia välineitä
- Yritykset tuottavat otteluanalyysin videoiden perusteella

Mitä riskejä ja ongelmia datan käyttöön mielestänne liittyy?

Suurimpana riskinä vastauksissa pidettiin datan väärin tulkintaa tai kokonaan tulkitsematta jättämistä ja niiden aiheuttamia ongelmia, jotka voivat pahimmillaan antaa täysin väärän kuvan seuran toiminnasta. Etenkin tärkeän datan erottaminen ja löytäminen suuresta datamäärästä voi olla haastavaa ja aikaa vievää. Myös datan hyödyntäminen oikein ja tehokkaasti voi olla seuroille haastavaa ja vaatia liian suuria resursseja. Näiden lisäksi datan suurena ongelmana pidettiin sitä, ettei sen avulla pystytä luotettavasti arvioimaan pelaajien henkisiä ominaisuuksia, joilla on kuitenkin merkittävä vaikutus pelaajan peliesityksiin. Datan luotettavuus aiheutti myös epäilyä, etenkin kun se tuotetaan kolmannen osapuolen kautta. Lisäksi datan ja subjektiivisten tuntemusten välinen ristiriita voi aiheuttaa valmennukselle haasteita.

Vastaajien kokemat riskit datan käyttöön liittyen voidaan tiivistää kahteen pääkohtaan:

- Dataa tulkitaan väärin
- Data ei ole luotettavaa, etenkin jos se tuotetaan kolmannen osapuolen kautta

Millä tavalla uskotte datan käytön ja sen hyödyntämisen kehittyvän tulevina vuosina?

Yhtenä tulevaisuuden suuntauksena haastateltavat pitivät sitä, että datan käytön lisääntyminen muuttaa valmennusta enemmän numeroiden tulkinnan suuntaan, mikä ei luonnollisesti ole jokaisen valmentajan mieleen. Vastaajat uskovat uusien mittarien ja mittausjärjestelmien lisääntyvän tulevaisuudessa ja datan tulevan helposti saataville myös pienemmille seuroille.

Vammojen ehkäisyssä ja ylikuormituksen seuraamisessa uskottiin tulevaisuudessa myös päästävän entistä korkeammalle tasolle. Tämän lisäksi uudet sovellukset voivat vastausten mukaan muun muassa analysoida suoraan, miten ketäkin yksilöä tai joukkuetta vastaan tulisi pelata. Ne voivat myös ennustaa datan pohjalta, miltä joukkueen syöttökuviot ja pelaajien perustoiminnot näyttävät animaatioin esitettyinä. Kuten yhdessä vastauksessa pohdittiin, tämän kaltaiset sovellukset johtavatkin helposti kysymykseen: ”Kummat lajia tulevaisuudessa suorittavat, ihmiset vai teknologia?”

Vastaajien mielipiteet datan käytön tulevaisuudesta voidaan tiivistää kolmeen pääkohtaan:

- Valmentaminen muuttuu enemmän numeroiden tulkitsemisen suuntaan
- Valmennusta auttavat sovellukset kehittyvät nopeasti ja vammojen ehkäisy helpottuu
- ”Kummat lajia tulevaisuudessa suorittavat, ihmiset vai teknologia?”

3.3 Vastausten analysointi

Vastauksissa oli tiettyjä eroavaisuuksia eri seurojen välillä, vaikka mielipiteet datan tarpeellisuudesta ja hyödyllisyydestä olivatkin samanlaisia. Erityisesti mielenkiintoisia kohtia tuli esille seurojen ulkopuolisia yrityksiä ja datan tulevaisuuden käyttökohteita tarkastellessa. Datan käyttö oli varsin monipuolista Suomen kakkosdivisioonassa ja todennäköisesti se tulee yleistymään entisestään seuraavien vuosien aikana.

Jalkapallotilastoja ja -analyysia tuottava InStat jakoi mielipiteitä haastattelussa. InStatia pidettiin kattavana ja hyödyllisenä ohjelmana, joka kuitenkin vaatii paljon aikaa ja resursseja toimiakseen kunnolla seuran eduksi. Tästä syystä etenkin Suomen kolmanneksi korkeimmalla sarjatasolla InStatin käyttö voi vielä olla liian haastavaa, sillä se vaatisi vastausten perusteella yhden osaavan henkilön vastaamaan analyysin purkamisesta ja analysoinnista, eikä tähän välttämättä ole monella seuralla tällä tasolla taloudellista mahdollisuutta. Lisäksi analysointia tekevän henkilön pelin ymmärtäminen voi erota suuresti valmentajan omasta ajatusmallista. Tilanteet, joissa pelaajat voittavat pallon itselleen ilman osumaa palloon voidaan merkata kaksinkamppailutappioiksi, vaikka todellisuudessa pelaaja voittaakin tilanteen ja pallon omalle joukkueelleen. Myös muut taktiikkaan kuuluvat suoritukset, kuten ristipallot voidaan merkata pallonmenetyksiksi, jotka laskevat pelaajan arvosanaa, vaikka ne voivat olla juuri joukkueen taktiikan mukaisia ja pelitapaan kuuluvia. InStat tai muut sen kaltaiset ohjelmat ovat korkealla

tasolla pelattaessa varmasti välttämättömiä, mutta tällä tasolla siihen kuluvat resurssit on mahdollista käyttää jopa tehokkaammin. Tärkeää kuitenkin on, että seurat huomaavat mikä datan analysointitapa heillä toimii tai ei toimi ja reagoivat ongelmiin mahdollisimman nopeasti ja järkevästi. Aivan kuten InStatin hylännyt joukkue teki havaittuaan paremman keinon analysoida videoita.

Asenteet ja mielipiteet datan käytön yleistymiseen liittyen olivat myös erittäin positiivisia, sillä datan käyttöä pidettiin välttämättömänä menestyäkseen kunnolla. Datan sanottiin muun muassa vähentävän mielipiteitä valmentajien keskuudessa esittämällä faktoja. Muutenkin datan koettiin helpottavan valmentajien työtä, vaikka sen tulkintaa ja suurten data määrien hyödyntämistä pidettiin haastavana. Nimenomaan datan hyödyntäminen, kuten InStatin yhteydessä mainittiin, onkin suurin haaste datan ollessa nykypäivänä varsin helposti saatavilla. Tästä syystä dataa ei pidä kerätä pelkämästä keräämisen ilosta, vaan sitä pitää oikeasti pystyä käyttämään seuran toimintaa edistävällä tavalla. Ajatukset datan tulevaisuudesta olivat myös mielenkiintoisia. Valmentajan työn muuttumista enemmän numeroita lukevaan ja ymmärtävään suuntaan pidettiin todennäköisenä. Missään vaiheessa, ainakaan pitkään aikaan, datan avulla ei kuitenkaan voida selvittää kaikkea. Tästä syystä valmentajan perinteiset taidot, kuten pelaajien mentaalisen puolen ymmärtäminen ja motivointi, pysyvät varmasti arvossaan. Ehkä valmentajien rooli tulee jossain vaiheessa olemaan enemmän motivointia ja reagointia pelitilanteisiin kuin itse taktiikan luomista.

Datan käytön haasteina pidettiin datan väärin tulkintaa, oikean datan löytämistä suuresta datajoukosta, pelaajien henkisten ominaisuuksien mittaamisen vaikeutta ja datan luotettavuutta, etenkin mikäli data tuotetaan kolmannen osapuolen kautta. Näitä kaikkia pidetään ongelmana yleisestikin dataa käytettäessä apuna urheiluliiketoiminnassa. Datan vääränlaisesta tulkinnasta ja väärään dataan tarttumisesta seurat pääsevät helposti eroon hankkimalla datan analysoinnista ja jalkapallosta yleisesti ymmärtävän seuran ideologiaan sopivan henkilön. Datan luotettavuudesta varmistuminen on puolestaan vaikeampi tehtävä, sillä varmistuakseen datan luotettavuudesta, pitäisi tilastot tarkistaa itse. Tämä taas vaatii valtavan määrän aikaa ja resursseja, minkä seurauksena se ei oikein ole mahdollista. Tästä syystä seurojen kannattaa valita käyttöönsä ohjelmia tai sovelluksia, jotka ovat keränneet hyvää palautetta ja ovat monella muullakin seuralla käytössä. Henkisten ominaisuuksien mittaaminen laitteiden avulla on tällä

hetkellä lähes mahdotonta. Kenties tulevaisuudessa sekin onnistuu, mutta etenkin alemmissa sarjoissa siihen menee varmasti todella pitkään.

Datan käytössä havaittiin haastateltavien joukkueiden osalta myös samoja haasteita kuin kuvassa 4 (kts. s.18) on esitetty. Jokaisesta kolmesta eri kategoriasta (datan ominaisuuksien aiheuttamat haasteet, datan käsittelyyn liittyvät haasteet ja datan käytön hallintaan ja johtamiseen liittyvät haasteet) löytyi haasteita, joita Suomen kakkosdivisioonassa oltiin kohdattu. Datan ominaisuuksista datan suuri koko ja myös datan arvon määrittäminen aiheuttivat ongelmia, sillä suuresta datajoukosta ei ollut helppoa löytää juuri sitä tietoa, josta seuralle on hyötyä. Datan käsittelyyn liittyvä datan analysointi oli myös joissakin seuroissa hankalaa, sillä esimerkiksi InStatin kohdalla laajan raportin hyödyntäminen kokonaan ei onnistunut kovin tehokkaasti. Datan hallintaan ja johtamiseen liittyi puolestaan dataa tuottava kolmas osapuoli, jonka luotettavuutta epäiltiin.

3.4 Joukkueiden erot

Joukkue 1 käyttää dataa todella kattavasti ja selvästi panostaa kyseiseen osa-alueeseen seuran toiminnan parantamiseksi. Kyseinen joukkue jätti muun muassa InStatin pois käytöstä sen vastatessa heikosti joukkueen tarpeisiin ja palkkasi yhden henkilön vastaamaan videoiden analysoinnista. Joukkueella on käytössä Wyscout-tili sekä Quanter-niminen sovellus, joiden avulla joukkue käy läpi uusia mahdollisia pelaajia ja nykyisten pelaajien hyvinvointia ja kuormitusta. Polarin välineillä seurataan myös pelaajien syketietoja, liikuttua matkaa ja spurttien sekä jarrutusten määrää.

Joukkue 2 käyttää dataa myös paljon. Joukkueella on käytössään hyväksi osoittautunut InStat-tili, joka vastaa tilastojen tuottamisesta. Dataa käytetään apuna otteluiden analysoinnissa (analysoidaan pelaajia sekä joukkuetta), pelaajien testaamisessa, vastustajan scouttaamisessa ja joukkueen sekä harjoittelun kehittämisessä. Datan oli havaittu kehittävän pelaajien pelin ymmärtämistä sekä heikkouksien ja vahvuuksien havainnointia. Aikasykkeen Firstbeat-mittausjärjestelmää käytetään kuormituksen ja palautumisen seurantaan.

Joukkue 3 käyttää dataa myös moneen eri tarkoitukseen. Pelaajien suorituksia ja oman sekä vastustajien joukkueita seurataan datan avulla, minkä seurauksena pelaajien ymmärrys omista tehtävistä ja urheilijan elämästä yleisesti on kehittynyt. Apuna joukkueella on Aikasyke sekä

paikallinen lääkäriasema, jotka auttavat terveyden ja suorituskyvyn mittaamisessa. Mahdollisesti jokin videoanalyysia tekevä henkilö tai yritys voisi auttaa seuraa entisestään saamaan entistä enemmän kuvatuista videoista irti.

Joukkue 4 käytti dataa hieman vähemmän kuin kolme ensimmäistä joukkuetta, eikä esimerkiksi seuran ulkopuolisia yrityksiä ollut käytössä lainkaan. Kuitenkin dataa kerätään ja analysoidaan seuran sisällä valmennustiimin kautta Soccermeterin avulla. Pelaajien fysiikkaa testataan myös kauden aikana ja siinä käytetään apuna erilaisia mittareita. Tälläkin joukkueella datasta voisi saada enemmän irti hankkimalla tietyn henkilön tai ryhmän vastaamaan pelkästään datan analysoinnista.

Joukkue 5 ei myöskään käyttänyt dataa aivan niin tehokkaasti hyödyksi kuin Joukkue 1, Joukkue 2 tai Joukkue 3. Dataa käytettiin perusasioihin, kuten otteluanalysointiin ja fyysisiin testeihin, mutta kuitenkin datan analysointi ei ollut erityisen kontrolloitua ja strukturoitua. Vastustajan ja oman joukkueen videoiden analysointi oman valmennustiimin kautta voi toki olla myös toimiva ratkaisu, mutta todennäköisesti tietyn erityisesti siihen osa-alueeseen keskittyvän henkilön tai ryhmän tuottama analyysi olisi kattavampi ja tarkempi.

Seuroilla oli pääosin samat tavoitteet datan hyödyntämiseen, mutta toteutustavat olivat erilaisia. Joukkueet 1, 2 ja osittain myös 3 käyttivät paljon ulkopuolista apua datan analysoinnissa ja dataa käytettiin järjestelmällisesti monella eri osa-alueella. Joukkueet 4 ja 5 toimivat puolestaan hieman yksinkertaisemmin, joten kyseiset seurat voisivat tulevaisuudessa panostaa datan analysointiin ja keräämiseen seuran toimintaa tehostaakseen. Joukkueet 2 ja 3 olivat ainoat, jotka käyttivät dataa apuna myös yksittäisten pelaajien pelisuorituksia analysoidessa. Suomen kolmanneksi korkeimmalla sarjatasolla dataa käytetään viiden seuran tarkastelun perusteella yllättävänkin kattavasti, vaikka vaihtelua eri joukkueiden välillä löytyy.

4 JOHTOPÄÄTÖKSET

Edellisessä luvussa käytiin Suomen kakkosdivisioonassa pelaavien joukkueiden valmennustiimien vastauksia läpi. Ennalta päätettyihin tutkimuskysymyksiin (miten dataa kerätään jalkapallojoukkueissa, miten dataa käytetään hyväksi, mitä riskejä data saattaa aiheuttaa sekä miten dataa voisi käyttää mahdollisesti entistä tehokkaammin hyödyksi Suomen kolmanneksi korkeimmalla sarjatasolla) saatiin tutkimuksen avulla vastauksia, kuten alla olevasta taulukosta 3 voi havaita. Etenkin erilaisten datan keräysmenetelmien määrä oli yllättävän korkea tälle sarjatasolle. Taulukosta voi myös huomata, että kaikki käytössä olevat datan keräysvälineet eivät ole teknisesti kovin vaativia, vaan myös itsearviointilomakkeet olivat edelleen käytössä. Mahdollisesti itsearvioinnilla pystyy analysoimaan myös pelaajien henkisiä ominaisuuksia.

Taulukko 3. Tutkimuskysymyksiin saatuja vastauksia.

Miten dataa kerätään?	Itsearviointilomake, Kehonkoostumusmittaus, video, Aikasyke (Firstbeat), InStat, Lääkäriasema, Wyscout, Fyysiset mittarit, esim. Polar, Soccermeter, Quanter
Miten dataa hyödynnetään?	Otteluanalysointi, pelaajascouttaus, harjoitusten seuraaminen, loukkaantumisten ennakointi ja hoitaminen, pelaajien testaaminen, pelaajien ymmärryksen kehittäminen, joukkueen heikkouksien ja vahvuuksien havainnointi, pelitavan kehittäminen
Mitä riskejä data voi aiheuttaa?	Data ei välttämättä ole luotettavaa, etenkin jos se tuotetaan kolmannen osapuolen kautta Oikean ja relevantin tiedon löytäminen Datan ja subjektiivisten tuntemusten ristiriita Data voi ohjata keskittymisen väärään paikkaan, mikäli se ymmärretään väärin Data ei auta henkisten ominaisuuksien arvioimisessa

Viimeiseen tutkimuskysymykseen, eli miten dataa voitaisiin käyttää mahdollisimman tehokkaasti tällä sarjatasolla hyödyksi, ei kuitenkaan saatu yhtä yksittäistä oikeaa ratkaisua. Kuitenkin vastausten perusteella voisi päätellä, että tärkeää on tehdä kattavaa otteluanalysointia (pelaajien ja joukkueen osalta), kehittää ja seurata pelaajia testien ja mittausten avulla ja arvioida uusia hankintoja erilaisten apuvälineiden (esimerkiksi Wyscout) avulla. Otteluanalysointia pitäisi pystyä tekemään siten, että pelitapaa sekä pelaajia saadaan kehitettyä. Tämä vaatii luonnollisesti datan oikeanlaista tulkitsemista ja analysointia, johon käytetään apuna esimerkiksi InStatin tapaista yritystä tai asiansa osaavaa videoanalyttikkoa.

Puolestaan muihin tutkimuskysymyksiin, eli datan keräysmenetelmiin, datan hyödyntämistapoihin ja mahdollisiin seurojen kokemuksiin riskeihin annetut vastaukset olivat tarkempia ja helpommin ymmärrettäviä niiden yksikäsitteisyyden vuoksi. Luonnollisesti jokaisella seuralla ei toimi samat välineet ja toimintatavat jo pelkästään erilaisten resurssien takia. Resurssien lisäksi jalkapallojoukkueiden peliajatukset ja toimintastrategiat ovat yleensä varsin erilaisia, minkä seurauksena toiset tarvitsevat ja haluavat enemmän tilastollista näyttöä omasta toiminnastaan. Hyvä osoitus tästä olikin muun muassa InStat, joka sopi haastattelujen perusteella hyvin toiselle joukkueelle, mutta toinen joukkue koki InStatin kuluviin resurssien menevän osittain hukkaan. Kenties tärkeintä onkin ottaa data rohkeasti käyttöön ja muokata ajan kuluessa toimintamallia parhaiten omalle joukkueelle sopivaksi.

Mahdolliset datan aiheuttamat riskit olivat myös sellaisia, joita oli odotettu. Nimenomaan tärkeän ja oikean tiedon löytäminen suuresta datajoukosta on nykypäivänä hankalaa, ei niinkään datan kerääminen. Tulevaisuudessa datan määrä tulee kuitenkin kasvamaan entisestään ja hyödyllisen tiedon etsiminen tulee olemaan entistä vaikeampaa. Viimeistään tässä vaiheessa, niin alemmilla kuin osittain korkeammillakin sarjatasoilla, täytyy joukkueiden pystyä entisestään panostamaan datan hyödyntämiseen pysyäkseen huipulla. Tämän työn perusteella voidaan sanoa, että ilman datan hyödyntämistä Suomen kakkosdivisioonassa hyvin pärjääminen on lähes mahdotonta.

LÄHTEET

1. Brauchle, Andreas. 2016. Anwendung von Data-Mining-Technologien zu statistischen Auswertungen und Vorhersagen im Fussball. [verkkodokumentti]. Hohenheimin yliopiston talous- ja yhteiskuntatieteiden laitos, [viitattu 28.6.2017]. Tietojärjestelmätien pro gradu -tutkielma. Saatavissa: <<https://github.com/brauchle/Masterarbeit-Data-Mining-Football>>
2. Billings, Andrew & Ruihley, Brody. 2014. The Fantasy Sports Industry: Games within Games. London. Routledge.
3. Delgado, Rick. 2017. How Big Data is Changing the Gambling World: Can data help turn the tables on the house? [verkkodokumentti]. [viitattu 6.9.2017]. Saatavissa <<https://channels.theinnovationenterprise.com/articles/how-big-data-is-changing-the-gambling-world>>
4. Delta, George & Matsuura, Jeffrey. 2017. Law of the Internet. New York. Wolters Kluwer. pp. 10-59.
5. D’Orazio, Tiziana & Leo, Marco. 2010. A review of vision-based systems for soccer video analysis. Pattern Recognition. Vol. 43, No. 8.
6. Foster, Richard. 2016. How football clubs calculate the cost of buying players in the transfer market. [verkkodokumentti]. [viitattu 26.6.2017]. Saatavissa <<https://www.theguardian.com/football/2016/apr/04/clubs-calculate-cost-transfer-market-leicester-southampton/>>
7. Gehrman, Sebastian. 2016. Wie moderne Datenanalyse den Fussball verändert. [verkkodokumentti]. [viitattu 28.6.2017]. Saatavissa <<https://www.bi-scout.com/die-vermessung-des-sports>>

8. Hymers, Matthew. 2014. How the adidas miCoach system has helped Germany in the World Cup? [verkkodokumentti]. [viitattu 26.6.2017]. Saatavissa <<https://www.sporttechie.com/how-the-adidas-micoach-system-has-helped-germany-in-the-world-cup/>>
9. Hytner, David. 2014. Arsenal's 'secret' signing: club buys £2m revolutionary data company. [verkkodokumentti]. [viitattu 26.6.2017]. Saatavissa <<https://www.theguardian.com/football/2014/oct/17/arsenal-place-trust-arsene-wenger-army-statdna-data-analysts/>>
10. Johansen, Håvard & Gurrin, Cathal & Johansen, Dag. 2015. Towards Consent-Based Lifelogging in Sport Analytic. In: Proc. of the International Conference on Multimedia Modeling, pp. 335-343.
11. Korhonen, Mari. & Köykkä, Kaisa. 2016. Big datan hyödyntäminen kestävässä liiketoiminnassa [verkkodokumentti]. Lappeenranta: Lappeenrannan teknillisen yliopiston tuotantotalouden laitos, [viitattu 28.6.2017]. Tuotantotalouden kandidaatintyö. Saatavissa <https://www.doria.fi/bitstream/handle/10024/123327/Kandidaatintyö_Korhonen_Köykkä.pdf?sequence=2>
12. Lopez-Gonzales, Hibai & Griffiths, Marks. 2016. Understanding the convergence of markets in online sports betting. International review for the sociology of sport. Vol. 1, No. 17.
13. Lu, Wei-Lwun & Ting, Jo-Anne & Little, James & Murphy, Kevin. 2013. Learning to Track and Identify Players from Broadcast Sports Videos. IEEE Trans. Pattern Anal. Mach. Intell. Vol. 25, No. 7.
14. Lynch, Gerald. 2016. This is how FIFA 17 player ratings are decided. [verkkodokumentti]. [viitattu 5.8.2017]. Saatavissa <<http://www.techradar.com/news/gaming/this-is-how-fifa-17-player-ratings-are-decided-1329441>>

15. McAfee, Andrew & Brynjolfsson, Bryn. 2012. Big Data: The Management Revolution. Harvard Business Review.

16. McCue, Matt. 2012. Academic algorithm takes on fantasy sports. [verkkodokumentti]. [viitattu 10.9.2017]. Saatavissa <http://www.espn.com/blog/playbook/fandom/post/_/id/6261/academic-algorithm-takes-on-fantasy-sports>

17. McHugh, Josh & Bronson, Po & Watters, Ethan. 2016. The Future of Sports. [verkkodokumentti]. [viitattu 10.9.2017]. Saatavissa <<http://futureof.org/wp-content/uploads/The-Future-of-Sports-2016-Report.pdf>>

18. Muthanna, Kavya. 2013. Big data: Why Enterprises need to start paying attention to their Data sooner? [verkkodokumentti]. [viitattu 30.8.2017]. Saatavissa <<https://kavyamuthanna.wordpress.com/2013/01/07/big-data-why-enterprises-need-to-start-paying-attention-to-their-data-sooner/>>

19. Opta. 2017. About Opta. [Optan www-sivuilla]. [viitattu 6.9.2017]. Saatavissa <<http://www.optasports.com/about/who-we-are/about-opta.aspx>>

20. Pollard, Richard. 2002. Charles Reep (1904-2002): pioneer of notational and performance analysis in football. Journal of Sports Sciences. Vol. 20, No. 10, pp. 853-855.

21. Rein, Robert. & Memmert, Daniel. 2016. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus. Vol. 5, No. 1.

22. Rejec, Jure. 2016. How Big Data is Changing the World of Soccer. [verkkodokumentti]. [viitattu 26.06.2017]. Saatavissa <<http://data-informed.com/how-big-data-is-changing-the-world-of-soccer/>>

23. Rudenko, Eugene. 2017. Big Data in Sports: Going for the Gold. [verkkodokumentti]. [viitattu 5.8.2017]. Saatavissa <<https://insidebigdata.com/2017/06/04/big-data-sports-going-gold/>>
24. Schumaker, Robert. & Solieman, Osama. & Hsinchun, Chen. 2010. Sports Data Mining. Springer US. Vol. 26, pp. 56.
25. Sivarajah, Uthayasankar & Kamal, Muhammad & Irani, Zahir & Weerakkody, Vishanth. 2017. Critical analysis of Big Data challenges and analytical methods. Vol. 70, pp. 263-286.
26. Soubra, Diya. 2012. The 3Vs that define big data. [verkkodokumentti]. [viitattu 6.9.2017]. Saatavissa <<http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>>
27. STATS. 2017. About STATS. [STATS:in www-sivuilla]. [viitattu 6.9.2017]. Saatavissa <<https://www.stats.com/about/>>
28. STATS. 2015. STATS Acquires Prozone. [STATS:in www-sivuilla]. [viitattu 10.9.2017]. Saatavissa <<https://www.stats.com/press-releases/stats-acquires-prozone/>>
29. Steinberg, Leigh. 2015. Changing the Game: The Rise of Sports Analytics. [verkkodokumentti]. [viitattu 6.9.2017]. Saatavissa <<https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/#36f2851e4c1f>>
30. Steve. 2014. How bookmakers track your every move & how to get around it (from an industry insider). Daily25. [viitattu 7.9.2017]. Saatavissa <<http://www.daily25.com/bookmakers-track-every-move-industry-insider/>>

31. The Economist. 2017. The Economist explains: How data changed gambling. [verkkodokumentti]. [viitattu 6.9.2017]. Saatavissa <<https://www.economist.com/blogs/economist-explains/2017/07/economist-explains-12>>
32. Ylijoki, Ossi. & Porras, Jari. 2016. Perspectives to Definition of Big Data: A Mapping Study and Discussion. Journal of Innovation Management. Vol. 4, No. 1, pp. 73-80.

LIITE 1.

Tässä liitteessä on viidelle Suomen kakkosdivisioonassa pelaavalle jalkapalloseuralle lähetetty kysymyssarja.

DATAN HYÖDYNTÄMINEN JALKAPALLOSEURAN URHEILUTOIMINNAN TEHOSTAMISESSA

Kysymykset:

1. Onko datan käytön lisääntyminen mielestänne hyvä vai huono asia ja miksi?
2. Millä tavalla dataa käytetään apuna joukkueenne urheilullisen toiminnan tehostamisessa? (esim. scouttaus, otteluanalysointi, harjoitukset, loukkaantumiset)
3. Mitä ylläolevista pidätte tärkeimpänä ja miksi?
4. Mitä konkreettisia vaikutuksia datan käytöstä on joukkueelle syntynyt?
5. Minkälaisia mittareita tai laitteita käytätte datan keräämisessä?
6. Tekeekö seuran ulkopuolinen yritys yhteistyötä joukkueen kanssa datan analysoinnissa? Jos tekee, niin mikä ja miten?
7. Mitä riskejä ja ongelmia datan käyttöön mielestänne liittyy?
8. Millä tavalla uskotte datan käytön ja sen hyödyntämisen kehittyvän tulevina vuosina?