

Severi Vielma

MASSADATA URHEILUANALYTIIKAN APUNA



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2017

TIIVISTELMÄ

Vielma, Severi

Massadata urheiluanalytiikan apuna

Jyväskylä: Jyväskylän yliopisto, 2017, 40 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja: Luoma, Eetu

Tämä kandidaatin tutkielma tutkii massadatan hyödyntämistä urheiluanalytiikassa. Vaikka massadata on jo yleistynyt käsite, sen hyödyntämismahdollisuudet urheiluanalytiikassa ovat vielä vähäisesti raportoitu aihe. Lähivuosina data-lähtöisen päätöksenteon hyödyntäminen on kasvanut ammattiurheilun toimialalla ja massadata tarjoaa urheiluorganisaatioille uusia mahdollisuuksia päätöksentekoon liittyen. Tämä tutkielma pyrkii vastaamaan siihen, miten massadataa voitaisiin hyödyntää urheiluanalytiikassa. Tutkielman toteutustapa on tieteellinen kirjallisuuskatsaus. Tulokseksi tutkielmassa saatiin, massadataa voitavan hyödyntää urheiluanalytiikassa monilla eri tavoilla, jotka liittyvät pelaajien ja otteluiden suoritusten analytiikkaan, urheiluliiketoiminnan analytiikkaan sekä pelaajien loukkaantumisten ehkäisemisen analytiikkaan.

Asiasanat: massadata, urheiluanalytiikka, analytiikka

ABSTRACT

Vielma, Severi

Utilizing Big Data in Sports Analytics

Jyväskylä: University of Jyväskylä, 2017, 40 p.

Information Systems, Bachelor's Thesis)

Supervisor: Luoma, Eetu

This bachelor's thesis aims to discover ways big data can be utilized in sports analytics. Regardless of big data being a talking point, it's fairly little reported topic in a sports analytics context. In the recent years data-driven decision making has become more common approach in sport organizations and big data enables new possibilities. This thesis aims to find out different ways big data can be utilized in sports analytics. The thesis has been carried out as a literature review. As a result the study found out that big data can be utilized in sports analytics in a many different ways that include analytics of players and team, sports business' analytics and analytics of players' injury prevention

Keywords: Big data, sports analytics, analytics

KUVIOT

KUVIO 1 Urheiluanalytiikan viitekehys (Alamar, 2013, s.102) **Virhe.**
Kirjanmerkkiä ei ole määritetty.

TAULUKOT

TAULUKKO 1 Otteluiden ja pelaajien suoritusten analytiikka (Davenport, 2014a)	22
TAULUKKO 2 Urheiluliiketoiminnan analytiikka (Davenport, 2014a).....	24
TAULUKKO 3 Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisyn analytiikka (Davenport, 2014a).....	26
TAULUKKO 4 Video-massadatan hyödyntäminen urheiluanalytiikassa.....	33
TAULUKKO 5 Biometrisen massadatan hyödyntäminen urheiluanalytiikassa	33
TAULUKKO 6 Sijainti-massadatan hyödyntäminen urheiluanalytiikassa	34
TAULUKKO 7 Massadatan hyödyntämiskeinot urheiluanalytiikassa	34

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
KUVIOT	4
TAULUKOT	4
SISÄLLYS.....	5
1 JOHDANTO.....	6
2 MASSADATA.....	8
2.1 Massadata käsitteenä.....	8
2.2 Massadatan ominaisuudet ja kolmen V:n malli	9
2.3 Massadatan mahdollisuudet.....	10
2.4 Massadatan haasteet.....	12
2.5 Massadatan yhteenveto	14
3 URHEILUANALYTIikka.....	15
3.1 Urheiluanalytiikan määritelmä ja tekniset komponentit.....	15
3.1.1 Datan hallinta.....	17
3.1.2 Analyttiset mallit.....	18
3.1.3 Tietojärjestelmät	18
3.2 Urheiluanalytiikan osa-alueet.....	19
3.2.1 Otteluiden ja pelaajien suoritusten analytiikka	20
3.2.2 Urheiluliiketoiminnan analytiikka	22
3.2.3 Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka	24
3.3 Urheiluanalytiikan yhteenveto	26
4 MASSADATA URHEILUANALYTIIKAN APUNA	27
4.1 Massadata urheiluanalytiikassa	27
4.2 Massadatan hyödyntäminen urheiluanalytiikassa.....	28
4.2.1 Videodata	28
4.2.2 Biometrinen data	29
4.2.3 Sijaintidata	30
5 JOHTOPÄÄTÖKSET	32
LÄHTEET	36

1 Johdanto

Lähivuosina tapahtunut massadatan tuleminen urheilumaailmaan avaa valtavia mahdollisuuksia urheiluorganisaatioille. Maailman katsotuimman urheiliigan, Englannin valioliigan, jokaisella urheiluseuralla on nykyään palkkalistoillaan data-analytiikkatiimi. Stadioneiden kattoon asennettujen kameroiden avulla yksittäisistä otteluista saadaan valtavat määrät dataa, mitä voidaan oikein tulkittuna hyödyntää. Pelaajien jokainen liike sekä kosketus palloon tallentuvat ja yksi ottelu tuottaakin noin 1,4 miljoonaa datapistettä. Tämän valtavan datamassan avulla urheiluseurojen analyttikot pyrkivät poimimaan yhdenmukaisuuksia ja tuottamaan seurajohtajille, valmentajille, pelaajille sekä muille urheiluorganisaatioissa työskenteleville henkilöille informaatiota urheiluanalytiikan avulla.

Huolimatta datan tuomista uusista mahdollisuuksista, sen hyödyntämiseen kohdistuu urheilumaailmassa suhteellisen paljon vastarintaa (Lee, 2016, s. 1). Urheilumaailma, jossa päätöksenteko on perinteisesti perustunut hyvin vahvasti intuitioon, on nyt kokemassa murrosta, jota datalähtöinen päätöksenteko mahdollistaa. Kirjoittajan henkilökohtaisena motivaationa aiheen tutkimukseen toimii tausta ammattuurheilun parista. Vuosien ajan suomalaisten jalkapalloseurojen toimintaa nähneenä olen huomionnut, miten vähän datalähtöistä päätöksentekoa käytetään seurojen toiminnassa, vaikka mahdollisuudet siihen olisivat olemassa. Tämä tutkielma onkin suunnattu urheiluorganisaatioiden päätöksentekijöille.

Tutkielmaan liittyy olennaisesti massadatan ja urheiluanalytiikan käsitteet. **Massadata** (engl. Big Data) on käsite, jolla kuvataan niin suuria ja monimuotoisia datamassoja, että niiden käsittelyyn, hallinnointiin ja analysointiin tarvitaan tavallista kehittyneempiä menetelmiä (Chen, Chiang & Storey, 2012, s. 1166). Massadatalle ei ole olemassa tarkkaa numeerista raja-arvoa, jonka ylittämällä jokin datamassa voitaisiin luokitella massadatakksi, vaan määritelmä vaihtelee sektoreittain, riippuen siitä, minkälaiset datamassat ja datan käsittelytyökalut ovat toimialalla yleisiä (Manyika ym. 2011, s. 1). Massadatan kolmena yleisenä ominaisuutena pidetään sen määrää, nopeutta ja monimuotoisuutta.

(McAfee, Brynjolfsson, & Davenport, 2012, s. 4-5; Russom, 2011, s. 6-7; Gandomi, & Haider, 2015, s. 13).

Urheiluanalytiikan (engl. Sports analytics) käsitteellä kuvataan yleisesti analytiikan toteuttamista ammattiuurheilun toimialalla (Tan, Hedman & Xiao 2017, s. 1). Urheiluanalytiikka on datan hallintaa, ennakoivien analyttisten mallien soveltamista tähän dataan sekä tietojärjestelmien käyttöä urheiluorganisaatioiden päätöksentekijöiden informointiin. Sen avulla pyritään auttamaan päätöksentekijöitä tekemään parempia, järkipäisiä, arvioita urheiluorganisaatioidensa sisällä (Davenport & Harris, 2007, s. 132). Urheiluanalytiikan tavoitteena on saavuttaa sitä hyödyntävälle organisaatiolle kilpailullista etua (Alamar, 2013, s. 4). Urheiluanalytiikka voidaan jakaa kolmeen osaan, jotka ovat pelaajien ja joukkueen suoritusten analytiikka, urheiluliiketoiminnan analytiikka sekä terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka (Davenport, 2014).

Tutkimuskysymyksenä tutkimuksessa on ”Miten massadataa voidaan hyödyntää urheiluanalytiikassa?”. Vastataksemme tutkimuskysymykseen, tule meidän ymmärtää kaksi tutkielman kannalta merkittävää kokonaisuutta, massadata ja urheiluanalytiikka. Tämän johdosta tutkimuksen kahtena apukysymyksenä ovat ”Mitä on massadata?” ja ”Mitä on urheiluanalytiikka?”. Apukysymykset auttavat myöhemmin tutkielmassa määrittelemään, mitä on massadata urheiluanalytiikan kontekstissa. Tämän määrittelemisen jälkeen tutkielman tavoitteena on tieteellisen kirjallisuuden avulla tutkia, millä eri keinoin tätä massadataa voidaan hyödyntää urheiluanalytiikassa. Tutkielma on suunnattu urheiluseurojen päätöksentekijöille.

Tutkimukseen kerätään tietoa pääasiallisesti Google Scholarin, ProQuestin sekä AISEL:n avulla. Mikäli koetaan tarpeelliseksi, saatetaan myös muita elektronisia kirjastoja, kuten IEEE Xplore, Scopus ja ACM Digital Library käyttää. Tiedonhaussa hakusanoina toimivat ”big data”, ”sports analytics”, ”big data analytics” sekä ”big data in sports”. Molempiin apukysymyksiin käytetään 15 lähdetä ja itse tutkimuskysymykseen 20 lähdetä. Koska urheiluanalytiikan näkökulma ei ole tieteellisesti vielä kovin paljon raportoitu, voidaan lähteiden määräästä poiketa perustellusti, jotta tutkielman taso pysyy laadukkaana. Artikkelit pyritään priorisoimaan niin, että tilanteessa, jossa tarjolla on useampi lähde samasta aiheesta, valitaan eniten siteeratuin lähde käyttöön tutkielmaan. Tutkielma tehdään Okolin ja Schbramin (2010) työohjeiden mukaan.

Tutkielman sisältö rakentuu niin, että ensimmäiseen lukuun sisältyy johdanto. Toinen kappale kertoo massadatasta ja siinä pyritään vastaamaan apukysymykseen ”Mitä on massadata?”. Kolmannen kappaleen aiheena on urheiluanalytiikka ja luvussa pyritään vastaamaan toiseen apukysymykseen ”Mitä on urheiluanalytiikka?”. Neljännessä kappaleessa määritellään massadata ammattiuurheilun toimialalla ja tarkkaillaan massadataa urheiluanalytiikan näkökulmasta. Viimeisessä kappaleessa esitetään tutkimuksen johtopäätökset. Kahdessa viimeisessä kappaleessa pyritään vastaamaan tutkimuskysymykseen ”Miten massadataa voidaan hyödyntää urheiluanalytiikassa?”.

2 Massadata

Vuodesta 2012 alkaen, noin 2.5 eksatavua dataa luodaan joka päivä ja luodun datan määrä tuplaantuu 40 kuukauden välein. Enemmän dataa kulkeutuu internetin välityksellä joka sekunti, kuin koko internetissä kokonaisuudessaan oli vielä 20 vuotta sitten (McAfee, Brynjolfsson & Davenport, 2012). Wu:n ym. (2014) selvityksen mukaan 90 prosenttia kaikesta saatavilla olevasta datatamme on luotu edellisen kahden vuoden aikana eikä kykymme luoda sitä ole koskaan ennen ollut mittakaavallisesti yhtä suurta kuin tänä päivänä. Selvitysten mukaan datan määrä lähitulevaisuudessa tulee jatkamaan kasvuaan (Chen, Mao, & Liu, 2014).

Liiketoimintaan liittyvät päätökset, jotka ennen tehtiin arvailun tai erittäin vaivalloisten manuaalisten mallinnusten avulla, voidaan nyt tehdä käyttäen datalähtöisiä matemaattisia malleja. Tämän kaltainen massadatan analysointi ohjaa nykyään lähes kaikkia yhteiskunnan osa-alueita, kuten teollisuutta, vähittäiskauppaa, biotieteitä, mobiilipalveluita ja luonnontieteitä (Gehrke ym. 2014). Tässä luvussa käsittelemme massadatan käsitettä, sitä määrittäviä ominaisuuksia sekä massadatan mahdollisuuksia ja haasteita, joita se esittää.

2.1 Massadata käsitteenä

Tieteellinen kirjallisuus ei ole vielä päässyt täyteen yhteisymmärrykseen massadatan määritelmästä. Eri henkilöillä on erilaisia mielipiteitä siitä, mitä termillä täsmällisesti tarkoitetaan. Yleisesti ottaen tieteellinen kirjallisuus kuitenkin määrittelee massadatan datamassoiksi, joita ei tavanomaisten tietojenkäsittelyohjelmistojen ja -laitteistoiden avulla pystytä havaitsemaan, käsittelemään, hallinnoimaan tai prosessoimaan riittävän nopeasti. (Chen, Mao & Liu 2014; Kessler, Armour, Espinosa & Money, 2013; Manyika ym. 2011). Tämä määritelmä on kuitenkin subjektiivinen ja siihen sisältyy vaihteleva käsitys siitä, kuinka suuri datamassan on oltava, jotta se voitaisiin luokitella massadataksi. Teknologian kehittyessä, myös näiden massadataksi luokiteltavien datamassojen koko tulee kasvamaan. Massadatan määritelmä vaihtelee sitä hyödyntävän yrityksen toimialan, koon sekä standardien mukaan. Esimerkiksi toimialan yleiset ohjelmistotyökalut vaikuttavat kyseisen alan käsitykseen siitä, mitä on massadata. Tästä syystä massadata voi sektoreittain vaihdella kymmenistä teratavuista lukuisiin petatavuihin (Manyika ym. 2011).

2.2 Massadatan ominaisuudet ja kolmen V:n malli

Koska numeeristen raja-arvojen esittäminen massadatalle on erittäin vaikeaa, on tieteellisessä kirjallisuudessa esitetty ominaisuuksia, joilla massadata voidaan erotella perinteisestä datasta (Abbasi, Sarker & Chiang 2016). Massadatan kolmena yleisimpänä ominaisuutena pidetään datan määrää, nopeutta ja monimuotoisuutta. Kolmen V:n malli on tieteellisessä kirjallisuudessa yleisin esitetyistä massadatan ominaisuuksien malleista. (McAfee, Brynjolfsson & Davenport, 2012; Russom, 2011; Gandomi, & Haider, 2015).

Kolmen V:n mallin ensimmäinen ominaisuus on **määrä (engl. volume)**. Massadatan määrä on ominaisuutena hyvin suhteellinen ja jatkuvasti muuttuva. Tietyn kokoluokan datamassa saatetaan tänään määrittellä massadatakksi, mutta se ei välttämättä ylitä kynnystä enää huomenna, koska tiedontallennusvälineet sekä kapasiteetit kehittyvät jatkuvasti. Lisäksi datan tyyppi määrittelee sen, mitä tarkoitetaan puhuttaessa ”isoista datamäärästä”. Kaksi saman kokoluokan datamassaa saattavat tarvita erilaisia datan hallintakeinoja datamassan tyypistä riippuen. Esimerkiksi taulukko- ja videodata vaativat hyvin erilaisia datanhallintamenetelmiä, vaikka ne määrällisesti olisivatkin yhtä suuria. Koska massadatan määritelmä on riippuvainen myös toimialasta, on epäkäytännöllistä määrittää jotakin raja-arvoa massadatan määrä-ominaisuudelle (Gandomi & Haider, 2015). Datan määrän kasvaessa tarvitaan reaaliaikaisia tekniikoita referoimaan ja suodattamaan mitä talletetaan, sillä ei ole taloudellisesti kannattavaa tallettaa raakaa dataa (Gehrke ym. 2014).

Datan **nopeudella (engl. velocity)** viitataan tahtiin, jolla dataa luodaan sekä siihen nopeuteen, missä se tulisi analysoida maksimaalisen hyödyn saavuttamiseksi. Digitaalisten laitteiden, kuten älypuhelimien ja sensorien leviäminen on johtanut ennennäkemättömään datan luomistahtiin. Massadatan nopeuden myötä kerätyn datan analysointi ja ennakoiva, faktapohjainen-suunnittelu, ovat nousseet erittäin tärkeään rooliin massadatan yleistyessä (Gandomi & Haider, 2015). Liiketoimintaan liittyvät päätökset tehdään nykyään nopeammin kuin koskaan ennen juurikin massadatan nopeuden takia (Abbasi, Sarker & Chiang 2016). Massadatan nopeus on monesti yrityksille tärkeämpää kuin sen määrä. Massadatasta saadun, reaaliaikaisen, informaation avulla yrityksen on mahdollista olla kilpailijoihinsa ketterämpi päätöksenteossaan. Massadatan nopeus onkin synnyttänyt siirtymää siinä, miten yritykset pyrkivät hyödyntämään informaatiota liiketoiminnassaan. Organisaatiot ovat entistä enemmän alkaneet hyödyntää reaaliaikaista analytiikkaa itse sen sijaan, että joku kolmas osapuoli tuottaisi niitä. Tämän hyötynä organisaatioille on, että aikaa säästyy muihin prosesseihin, koska aikaakuluttavista siirtymistä luovutaan. (Abbasi, Sarker & Chiang 2016). Massadatan määrän ja nopeuden kasvu tarkoittavat myös sitä, että organisaatioiden täytyy kehittää prosesseistaan, jotka liittyvät datan keräämiseen, analysointiin ja tulkitsemiseen jatkuvia (McAfee, Brynjolfsson & Davenport, 2012).

Monimuotoisuus (engl. variety) on kolmas V-mallin käsitteistä. Sillä tarkoitetaan valtavien, epäyhtenäisten, datamäärien moniulotteisuutta. Dataa keräävillä tahoilla on yleensä hyvinkin erilaisia kaavoja ja protokollia datan tallentamiseen. Tämä johtaa yleensä siihen, että tallennettu data on hyvin monimuotoista (Wu. X, Zhu, Wu. G. Q. & Ding, 2014).

Monet massadatan tärkeimmistä lähteistä ovat suhteellisen uusia. Perinteiset, strukturoidut, tietokannat, joihin yritykset ovat ennen tallentaneet tietoaan, soveltuvat heikosti massadatan tallentamiseen ja prosessointiin. Kun samanaikaisesti tietojenkäsittelyyn liittyvät kulut ovat tasaisesti laskeneet, tarkoittaa se yrityksille aiemmin kalliiden dataperäisten lähestymistapojen muuttuvan nyt entistä taloudellisemmiksi. Massadatasta saatavissa olevat datamassat ovat yleensä epästrukturoituja, mutta niistä on louhittavissa suuria määriä informaatiota (McAfee, Brynjolfsson & Davenport, 2012).

Näiden kolmen ominaisuuden lisäksi tieteellisessä kirjallisuudessa on esitetty myös muita ominaisuuksia massadatalle. Neljänneksi ominaisuudeksi monessa tutkimuksessa on määritelty massadatan **arvo (engl. value)** (Kaisler, Armour, Espinosa & Money 2013; Chen, Mao & Liu, 2014; Katal, Wazid & Goudar 2013). Kaislerin ym. (2013) mukaan datan arvo kuvaa sen käytettävyyttä päätöksenteossa. Arvo-ominaisuudella pyritään korostamaan massadatan sisältävää piilossa olevaa arvoa. Tällä datan arvolla pyritään lisäksi osoittamaan massadatan yksi kriittisimmistä ongelmista, eli sitä, miten kooltaan valtavista, nopeista ja monimuotoisista datamassoista, voidaan löytää arvoa (Chen, Mao & Liu, 2014). Neljänneksi ominaisuudeksi on arvon sijaan esitetty myös totuudenmukaisuutta (Abbasi, Sarker & Chiang 2016). Myös muita ominaisuuksia on tieteellisessä kirjallisuudessa esitetty. Kaisler ym. (2013) esittelevät massadatan viidenneksi ja kuudenneksi ominaisuudeksi datan monimutkaisuuden sekä vaihtelevuuden. Kolmen V:n malli on kuitenkin yleisesti käytetyin malli massadatan ominaisuuksia määriteltäessä.

Massadatan ominaisuuksista on syytä muistaa, etteivät ne ole riippuvaisia toisistaan, mutta kun yksi ominaisuuksista muuttuu, todennäköisyys siihen, että muutkin ominaisuudet muuttuvat, on suuri. Kolmen V:n käännekohtalla Gandomi ja Haider kuvaavat pistettä, jossa datan ominaisuudet skaalautuvat niin laajoiksi, että perinteiset datan käsittely- ja analyysimenetelmät muuttuvat riittämättömiksi. Tämä käännekohta on Gandomin ja Haiderin (2015) mukaan olemassa jokaisella organisaatiolla.

2.3 Massadatan mahdollisuudet

Liiketoiminnan digitalisoitumisen myötä uudet informaation lähteet sekä jatkuvasti halpenevat laitteistot tuovat meidät kohti uutta aikakautta, jossa lähes kaikesta liiketoiminnan kiinnostuksen kohteista löytyy suuria määriä digitaalista tietoa (McAfee, Brynjolfsson & Davenport, 2012). Massadata ei kuitenkaan itsessään tarjoa organisaatioille arvoa, vaan avainasemassa on organisaation kyvyk-

kyys tuottaa tästä datasta informaatiota, jolla voidaan auttaa päätöksentekoa tulevaisuutta ennakoimalla (Kim, 2015). Tehokkaalla massadatan hyödyntämisellä yritykset ja julkisen sektorin toimijat voivat lisätä organisaationsa tuottavuutta sekä kilpailukykyä. Massadatan hyödyntäminen liiketoiminnassa perustuu siihen, että se antaa organisaation päätöksentekijöille mahdollisuuden perustaa päätöksentekonsa todistettuun faktatietoon intuition sijaan (McAfee, Brynjolfsson & Davenport, 2012). Vaikka massadatassa on olemassa kohinaa, voidaan sopivien tilastollisten menetelmien avulla toteuttaa analyysejä hyvin tuloksin, huolimatta saatavissa olevan datan valtavista määristä (Gehrke ym. 2014).

Datasta on tullut Chenin, Maon ja Liun (2014) mukaan yrityksille yksi tärkeä tuotannontekijä, siinä missä esimerkiksi materialinen omaisuus tai henkilöresurssit ovat jo pitkään olleet. Multimedial, esineiden internetin ja sosiaalisen median kehittymisen myötä tulevat yritykset keräämään enemmän tietoa kuin koskaan ennen, minkä johdosta käytössä olevan datan määrä tulee vain kasvamaan jatkossa. Massadatalla on potentiaalia tuottaa yrityksille sekä kuluttajille merkittävää arvoa. Perusteellisesti analysoimalla massadatasta voidaan tunnistaa trendejä, jotka auttavat löytämään aiemmin tunnistamattomia yhteyksiä (Cole, Nelson & McDaniel 2015). Tämän tekeminen perinteisillä tietokantatekniikoilla olisi yrityksille hyvin vaikeaa, eikä kustannustehokasta (Shim, French, Guo & Jablonski 2015).

Oikein hyödynnettynä massadataa voidaan Chenin, Maon ja Liun (2014) mukaan käyttää useissa yrityksen eri liiketoimintaprosesseissa. Markkinoinnissaan yritykset voivat massadatan avulla ennustaa asiakaskäyttäytymistään ja näin löytää uusia liiketoimintamalleja. Kamiokan ja Tapanaisen (2014) mukaan etenkin markkinoinnin ja massadatan yhteys tulee tulevaisuudessa kasvamaan merkitykseltään. Myynnin suunnittelussa organisaatio voi massadatan avulla optimoida tuotteiden hintansa ja toiminnallisella tasolla yritykset voivat parantaa tehokkuuttaan esimerkiksi optimoimalla työvoimaansa ennakoimalla tarvittavaa henkilöstön määrää massadataa analysoiden. Toimitusketjuissa massadatan avulla voidaan suorittaa inventaarion ja logistiikan optimointia sekä sen avulla voidaan pienentää kuilua tuotteiden kysynnän ja tarjonnan välillä (Chen, Mao & Liu, 2014).

Yritykset jotka ovat alusta asti olleet digitaalisia, ovat jo massadatan käytössä hyvin kehittyneitä, mutta muiden toimialojen yrityksille massadatan mahdollinen käyttämätön potentiaali on kuitenkin huomattavasti suurempi. Tutkimuksissa on ilmennyt, että mitä enemmän yritykset luonnehtivat itseään datalähtöisiksi, sitä paremmin ne myös suoriutuvat taloudellisilta ja operatiivisilta tuloksiltaan. Etenkin yritykset, jotka kuuluvat toimialansa ylimpään kolmannekseen hyötyvät tästä dataan perustuvasta päätöksenteosta merkittävästi. McAfeen, Brynjolfssonin ja Davenportin (2012) mukaan näiden dataan perustuvaan päätöksentekoon tukeutuvien ylimmän kolmanneksen yritysten tuottavuus oli 5 % kilpailijoitansa korkeampi. Tämän lisäksi kyseisien yritysten voitollisuus oli 6 % kilpailijoihin korkeampi. Myös Kamiokan ja Tapanaisen

(2014) tutkimuksessa yritysten systemaattisella ja laajalla massadatan hyödyntämisellä oli suora, positiivinen, yhteys organisaation kilpailukykyyn.

Organisaatiot, jotka hyödynsivät massadataa tavallisen datan sijaan, erotuivat prosesseissaan muista yrityksistä kolmella merkittävällä tavalla. Massadataa hyödyntävät yritykset keskittyvät datan virtauksen seurantaan osakkeiden seurannan sijaan. Kyseiset yritykset luottivat liiketoimintaprosesseidensa päätöksenteossa datatieteilijöihin ennemmin kuin data-analyttikoihin. Kolmantena huomiona tutkimuksessa oli, että kyseiset yritykset ovat laajentamassa analytiikkaansa pelkästä IT-puolen toiminnasta yrityksen ydinliiketoimintapuoleen (Davenport, Barth & Bean, 2012).

Yritysten olisi päätöksenteossa huomioitava massadatan mahdollisuudet tulevaisuudessa, sillä kerätty data saattaa tulevaisuudessa muuttua arvokkaaksi tavoilla, joita emme vielä kykene ennakoimaan. Yritysten tulisi tämän lisäksi myös kehittää toimintaansa tapoja, jolla ne pystyisivät luomaan arvoa datasta, joka on puutteellisesti kerättyä. Tällä tavalla toimiminen nostaa organisaation sisällä tarvetta seurata datan lähteitä ja käsitellä sen epämääräisyyksiä ja virheitä (Gehrke ym. 2014).

2.4 Massadatan haasteet

Massadataan ja sen hyödyntämiseen liittyy mahdollisuuksien lisäksi myös haasteita. Ghasemaghaein (2007) raportti osoittaa, että keskimäärin vain joka neljäs yritys kokee saavansa kilpailullista etua massadatan keräämisestä ja sen analysoimisesta. Yrityksillä on tutkimuksien mukaan keskimääräisenä tavoitteenaan 3,5-kertainen tuotto jokaista massadataprojektiin sijoitettua dollaria kohden. Monet näistä projekteista kuitenkin tuottavat vain 1,5-kertaisia tuottoja, mikä kertoo siitä, etteivät yritykset saa massadatasta sellaisia tuottoja, joita ne projekteihin sijoitettaessa toivoisivat (Shim, 2015).

Massadatan kolme V:tä johtavat moniin haasteisiin, joita yritykset kohtaavat. Datan määrä ja nopeus asettavat haasteita massadatan hyödyntämiseen, sillä näiden ominaisuuksien myötä tavanomaiset päätöksenteon lähestymistavat eivät monesti ole enää soveliaita analyysiä tehdessä massadatan avulla. Siihen mennessä, kun organisaatio on saanut päätöksentekoon tarvittavan tiedon massadatasta, on uutta dataa jo saatavissa, mikä saattaa tehdä edellisestä päätöksestä vanhentuneen. Organisaatioiden tuleekin omaksua entistä yhtäjaksoisempia lähestymistapoja massadatan analysointiin ja sen perusteella tehtävään päätöksentekoon. Organisaatioille ei riitä pelkkä tietovirran seuranta, vaan niiden tulee myös olla valmiita tekemään päätöksiä saatujen tietojen perusteella. Tämän takia organisaatioiden tulisi vakiinnuttaa prosesseja, joiden avulla päätetään milloin tietyn päätöksen tekeminen tai toiminta on tarpeellinen. Tämä auttaa organisaatiota määrittämään päätöksentekoonsa liittyvät sidosryhmät, prosessit, kriteerit sekä aikavälit, joissa päätökset on tehtävä (Davenport, Barth & Bean, 2012). Massadatan monimuotoisuus taas johtaa monesti datan leviämiseen. Laajasti eri lähteistä levinnyt data tulee ensin kerätä, jotta muuta

analysointia voisi levinneen massadatan perusteella tehdä (Chen, Mao & Liu, 2014). Ghasemaghaein (2017) tutkimus osoittaa myös, että yksi organisaatioiden suurimmista haasteista massadatan hyödyntämisessä on ymmärtää, miten massadatasta saataisiin sen arvo hyödynnettyä. Jotta valtavista datamääristä saataisiin louhittua hyödyllistä informaatiota, on tiedonlouhinnan oltava tehokasta ja lähes reaaliaikaista toimintaa, sillä monesti yritysten on lähes mahdotonta tallentaa kaikkea saatavissa olevan dataa (Wu, X, Zhu, Wu, G. Q. & Ding, 2014).

Käytettävissä olevat teknologiat ovat yksi merkittävä tekijä, joka vaikuttaa siihen kuinka hyvin organisaatio onnistuu hyödyntämään massadataa liiketoiminnassaan. Yritysmailmassa on olemassa oleva tarve uusille teknisille innovaatioille ja tekniikoille, jotka auttavat käyttäjiä integroimaan, visualisoimaan ja hyödyntämään massadataa. Chen ym. (2014) selvityksen mukaan perinteiset tietokannat eivät pysty käsittelemään haasteita, joita massadata tietokannoille esittää. Tästä syystä joustavammat, suurempia datamääriä tukevat tietokannat, kuten NoSQL ovat muuttumassa suosituimmiksi ratkaisuksi massadatan tallettamiseen. Chenin ja Zhangin (2014) mukaan tarvitsemme kehittyneempiä arkkitehtuureja, tehokkaampia datalähtöisiä tekniikoita sekä edistyneempiä massadata-alustoja, jotta saisimme hyödynnettyä massadatasta kaiken sen potentiaalain.

Monet yritysten massadatan kautta kohtaamista haasteista liittyvät myös yritysten rakenteisiin ja toimintakulttuureihin. Organisaatioiden johtajilta puuttuu useasti ymmärrystä massadatan arvosta ja siitä, miten tämä arvo saataisiin hyödynnettyä (Manyika ym. 2011). Massadatasta saadun tiedon käyttäminen päätöksenteon olennaisena komponenttina vaatii organisaatiolta sekä sen johdolta paljon uusia kykyjä. Myös organisaation toimintakulttuurin tulee olla mahdollisimman vastaanottavainen. (Bughin, Chui & Manyika 2010). Käyttäjien positiivinen asenne, riittävät henkilöstöresurssit sekä hyvän tietoturvan omaava datahallinto olivat merkittäviä tekijöitä, joilla oli Kamiokan ja Tapanaisen (2014) tutkimuksen mukaan positiivinen vaikutus massadatan käytön laajuuteen yrityksissä. Tutkimus osoitti myös, että johdon osallistuminen sekä ymmärrys massadataan olivat suorassa yhteydessä organisaation siitä saamaan hyötyyn. Koko organisaation tulisi määrittää uudestaan ajatuksensa päätöksenteosta ja yritysten tulisi palkata massadataan liittyviin töihin henkilöitä, jotka ovat kykeneviä etsimään siitä kaavoja sekä muuttamaan niitä hyödyllisiksi liiketoimintainformaatioiksi (McAfee, Brynjolfsson & Davenport, 2012). Tieteellisten raporttien mukaan massadatan kanssa työskentelevällä henkilöstöllä tuleekin olla perustaidot IT-osaamisessa, mutta heidän on tämän lisäksi myös kyettävä pystymään soveltamaan näitä taitoja (Davenport, Barth & Bean, 2012; Mikalef ym. 2016).

Gehrken ym. (2014) mukaan yksi keskeinen massadatan haaste on, miten siitä saavutettu informaatio saadaan välitettyä päätöksentekijöille mahdollisimman ymmärrettävässä muodossa. Päätöksentekijöille tulee Gehrken mukaan tarjota saavutetut informaatiot tarpeeksi tiivistetyssä muodossa, jotta olennai-

simmat tiedot eivät huku informaation määrään. Tiedon visualisoiminen helposti ymmärrettävään muotoon on myös yksi tärkeä tekijä, sillä jos päätöksentekijöiltä menee raportin olennaisuudet ohi, ei yritys myöskään hyödy massadatasta mitenkään.

Tieteellisessä kirjallisuudessa on nostettu esiin myös joitakin muita massadatan haasteita kuten tietoturva, datan etsimisen haasteet, sen jakaminen sekä analysointi (Chen & Zhang, 2014). Myös datan omistajuus esittää haasteen, joka yritysten täytyy ratkoa (Kaisler ym. 2013). Mikäli tässä alaluvussa käsiteltyjä haasteita ei kyetä organisaatioissa ratkomaan, ei massadataa saada hyödynnettyä sen koko potentiaalin mukaan (Chen & Zhang, 2014).

2.5 Massadatan yhteenveto

Suurin osa ihmisistä ei ole tekemisissä suurien datamäärien kanssa jokapäiväisessä elämässään. Tämän ihmisen ja massadatan välisen henkilökohtaisen kokemuksen puuttuessa emme välttämättä myöskään ymmärrä mahdollisuuksia tai haasteita, joita se meille yksilö- ja organisaatiotasolla esittää (Kaisler ym. 2013). Edellisessä luvussa on pyritty tieteellisen kirjallisuuden avulla tutkimaan massadatan käsitettä, sekä sen ominaisuuksia tieteellisessä kirjallisuudessa yleisesti esiintyvän kolmen V:n mallin avulla. Kolme yleistä ominaisuutta, eli määrä, nopeus ja monimuotoisuus, erottavat massadatan perinteisestä datasta.

Ominaisuuksien jälkeen tutkimuksessa selvitettiin, mitä tieteellinen kirjallisuus kertoo massadatan mahdollisuuksista. Tieteelliset tutkimukset osoittavat, että etenkin dynaamisissa ja myllertävissä ympäristöissä suurin kilpailullinen etu yritysten välillä syntyy massadatasta ja sen analytiikasta. Oikein hyödynnettynä massadata ja sen analytiikka kehittävät merkittävästi yrityksen dynaamisuutta (Mikalef ym. 2016).

Lopuksi selvitettiin, millaisia haasteita massadataan liittyy. Tutkimukset osoittivat, etteivät yritykset monesti pääse tuottavuustavoitteisiinsa, joita ne massadataprosjekteilleen asettavat. Lukuisat esitetyistä ongelmista luovat yrityksille sekä organisaatioille mietittävää sen suhteen, miten massadata saataisiin valjastettua organisaation käyttöön mahdollisimman tehokkaasti.

3 Urheiluanalytiikka

Yksi siteeratuimmista analytiikan käsitteistä on peräisin Davenportilta ja Harrisilta (2007). Davenport ja Harris määrittävät analytiikan “datan, tilastollisten- ja määrällisten analysointien, havainnollistavien ja ennustavien mallien, sekä faktaperäisen johtamisen käyttämisenä päätöksenteossa ja muussa toimimisessa”. Analytiikka on kaksikon mukaan työkalu teknologioihin ja prosesseihin, jotka pyrkivät datan avulla ymmärtämään ja analysoimaan liiketoiminnallista suorituskykyä (Davenport & Harris 2007, s. 7).

Vaikka monet eri toimialat ovat tänä päivänä omaksumassa entistä analyttisempia lähestymistapoja päätöksentekoon, ei kuitenkaan millään muulla toimialalla ole samantapaista analyttistä virtausta alkamaisillaan, kun urheilun toimialalla (Davenport, 2014a). Tämä johtuu pääasiallisesti siitä, että teknologioiden kehittymisen johdosta ennen piilossa ollut informaatio on muuttunut ammattiurheilun toimialalla entistä helppopääsyisemmäksi ja julkisemmaksi (Xiao ym. 2017). Tässä luvussa käsittelemme urheiluanalytiikan tieteellisen määritelmän sekä siihen liittyvät keskeiset tekniset komponentit. Tämän jälkeen luvussa määritellään ja tutkitaan urheiluanalytiikan kolmea osa-aluetta Davenportin (2014a) jaottelun mukaan.

3.1 Urheiluanalytiikan määritelmä ja tekniset komponentit

Urheiluanalytiikalla tarkoitetaan Alamarin (2013, s. 5-6) mukaan rakenteellisen ja historiallisen datan hallitsemista, tulevaisuutta ennakoivien analyttisten mallien soveltamista tähän dataan sekä tietojärjestelmien käyttöä informoimaan urheiluseuran päätöksentekijöitä. Davenport (2014b) taas määrittelee urheiluanalytiikan data-analytiikan täsmälliseksi hyödyntämiseksi ammattiurheilun toimialalla. Urheiluanalytiikan tavoitteena on saavuttaa sitä hyödyntävälle urheiluorganisaatiolle kilpailullista etua. Tätä kilpailullista etua pyritään saavuttamaan säästämällä päätöksentekijöiden aikaa sekä tarjoamalla heille uusia näkemyksiä. Jos päätöksentekijöille pystytään tarjoamaan kaikki merkityksellinen informaatio helposti saavutettavalla tavalla, jää muutoin tiedon keräämiseen kulutettu aika analysointiin. Kun urheiluseurojen saavutettavissa oleva data lisääntyy ja monipuolistuu, kasvavat myös mahdollisuudet, että tämä data saataisiin muutettua uudeksi ja hyödylliseksi informaatioksi. (Alamar, 2013, s. 5-6).

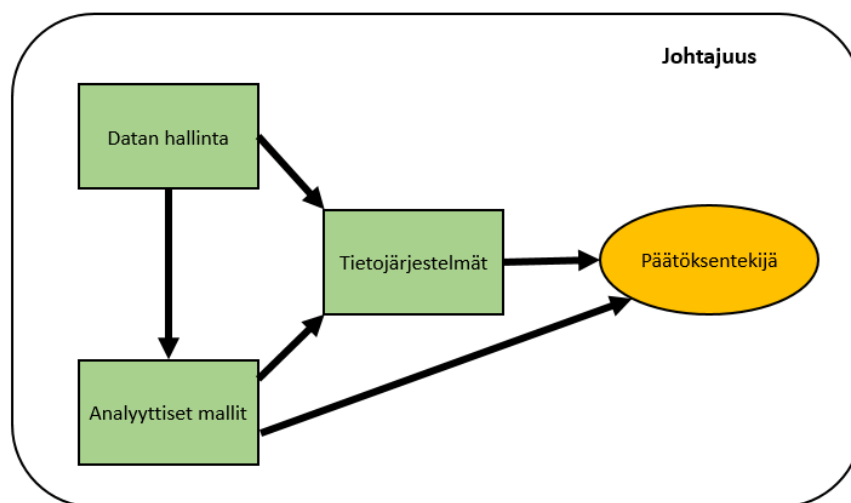
Sysäys kohti analyttisten työkalujen käyttöä johtuu urheilussa, kuten yleisesti muussakin liiketoiminnassa, kahdesta asiasta: tietotekniikan kehitymisestä sekä suurien datamassojen lisääntymisestä. Urheiluseurat, jotka investoivat analytiikkajärjestelmiinsä ja pitäytyvät aallonharjalla urheiluanalytiikan osalta, saavuttavat järjestelmällistä kilpailullista etua kilpailijoihinsa nähden. Urheiluseurojen tulisi teknologioihin ja menetelmätapoihin sijoittamisen lisäksi

investoida resurssejaan ymmärrykseen siitä, kuinka sisällyttää analytiikkaa organisaation päätöksentekoprosesseihin. Investoinnin määrät, organisaation pitkän aikavälin strategiat sekä lähestymistapa analytiikkaan ovat kaikki avaintekijöitä seuran urheiluanalytiikan hyödyntämisessä (Alamar, 2013, s. 1-3).

Urheiluseuroilla teetetyssä selvityksessä tuli ilmi, että urheiluseurojen tietotekniikka-puolen työntekijöiden näkemys seuran analyttisistä kyvykkyyksistä erosi merkittävästi muun henkilöstön näkemyksistä. Tietotekniikka-puolen työntekijöiden mielestä urheiluseuran analyttiset kyvykkyydet olivat selkeästi kilpailijoita parempia. Muu henkilöstö oli taas sitä mieltä, että kilpailijoiden kyvykkyys oli omaa organisaationsa suurempaa. Ristiriidasta voidaan päätellä, että joko seuran tietotekniikan työntekijät ovat hyvin optimistisia analytiikan käyttökyvystä tai sitten muun henkilöstön käsitys kyvykkyydestä oli riittämätöntä (Alamar, 2013, s. 2-3).

Urheiluanalytiikan komponentit ovat Alamarin viitekehysten (kuvio 1) mukaan datan hallinta, analyttiset mallit, tietojärjestelmät, päätöksenteko sekä johtaminen. viitekehysten mukaan urheiluanalytiikassa ensimmäisenä organisoidaan ja prosessoidaan datamassat, jonka jälkeen ne jaetaan analyttisten mallien ja tietojärjestelmien käytettäväksi. Analyttiset mallit käyttävät dataa joko organisaation yhteisen, standardoidun, mallin mukaan tai vastaavat johonkin päätöksentekijän esittämään täsmälliseen kysymykseen. Päätöksentekijä eroaakin kolmesta edellisestä komponentista, sillä edellä mainitut datan hallinta, analyttiset mallit ja tietojärjestelmät ovat teknisiä komponentteja. Nämä tekniset komponentit eivät kuitenkaan tee päätöksiä, vaan vastuu siitä on päätöksentekijällä (kuvio 1).

Viimeisenä komponenttina urheiluanalytiikassa on johtajuus. Johtajuuden avulla urheiluseuraan laaditaan tehokas, koko organisaation kattava, toimintasuunnitelma urheiluanalytiikan hyödyntämiseen, jonka jälkeen kaikilta seuran työntekijöiltä vaaditaan täyttä sitoutumista suunnitelmaan. (Alamar, 2013, s. 4-5). Johtajuuden avulla urheiluanalytiikka voidaan asettaa säännölliseksi osaksi urheiluseuran päätöksentekoprosesseja, jolloin seuran on mahdollista saavuttaa kilpailullista etua. (Alamar, 2013, s. 102-103).



KUVIO 1 Urheiluanalytiikan viitekehys (Alamar, 2013, s.102)

Seuraavissa alaluvuissa käydään tarkemmin läpi Alamarin (2013) viitekehyksen kolme keskeistä teknistä komponenttia datan hallinta, analyttiset mallit sekä tietojärjestelmät.

3.1.1 Datan hallinta

Datan määrän kasvaessa myös lähteet joista dataa kerätään laajenevat (Alamar, 2013, s. 6-7). Selvityksen mukaan lähes puolet urheiluseuroista keräävät dataa enemmästä kuin kuudesta eri lähteestä (Alamar, 2013, s. 14). Ulkoiset datan lähteet ovatkin yleistymässä, sillä monet urheilusarjat ja -liitot keräävät itse dataa, jonka lisäksi lähes kaikilla urheilulajeilla on kolmannen osapuolen datan toimittajia (Davenport 2014a). Aika, joka urheiluseuroilla kuluu eri lähteiden datan käsiksi pääsemiseen, voitaisiin käyttää päätöksentekoprosesseihin hyvin suunniteltujen tietojärjestelmien avulla. (Alamar, 2013, s. 14). Jotta datasta saataisiin hyödyllinen informaatio esiin, tulisi se järjestää sellaiseen tapaan, joka sallii kaikille seuran päätöksentekijöille vaivattoman pääsyn siihen eikä ole riippuvainen yksittäisistä henkilöistä (Alamar, 2013, s. 6-7). Selvitykseen vastanneista urheiluseuroista vain n. 31 % kertoi kaiken organisaation datan olevan keskitettyä. Datan keskittäminen on yksi avaintekijöistä tietojärjestelmien tehokkaassa hyödyntämisessä analytiikkaa ajatellen. Datahallinnon rooli urheiluanalytiikassa on järjestellä ja keskittää dataa sekä virtaviivaistaa datan tuloa urheiluseuran käyttöön (Alamar, 2013, s. 6-7).

Alamar (2013, s.24) nostaa kolmeksi datahallinnon periaatteeksi datan standardoinnin, keskittämisen ja integroimisen. Kyseiset kolme toimintaprosessia mahdollistavat urheiluseuran työntekijöille helpon ja oikea-aikaisen pääsyn informaatioon. Datan standardoinnilla tarkoitetaan yhteisien standardien luomista datan tallentamiseen. Keskittämisellä tarkoitetaan organisaation eri työntekijöiden keräämän data tallentamista yhteen paikkaan sen sijaan, että samaa data olisi useissa toisiltaan erillisissä tietokannoissa. Kun data on keskitetty, ei päätöksentekijöiden tarvitse kuluttaa aikaa datan etsimiseen. Sillä myös varmistetaan, että kaikilla organisaation työntekijöillä on sama tieto käytettävissä ilman eriäväisyyksiä.

Datan keskittämisen jälkeen käyttäjiä tulisi kouluttaa järjestelmän käyttämiseen. Kun urheiluorganisaation data on standardoitu ja keskitetty, voidaan se integroida. Integroinnin avulla organisaation eri osa-alueilla kerättyä dataa voidaan yhdistää muiden datamassojen kanssa, jolloin saatetaan havaita yhteyksiä, joita ei muuten välttämättä olisi havaittu (Alamar, 2013, s. 25–34). Mikäli tärkeintä dataa ei integroida, täytyy päätöksentekijöiden käyttää aikaa kyseisen tiedon keräämiseen, eivätkä seuran analyttikot myöskään kykene yhdistämään muista lähteistä saatua dataa itse keräämänsä datan kanssa (Alamar, 2013, s. 31). Yksi Euroopan menestyneimmistä jalkapalloseuroista, Bayern München, havaitsi kehittäessään analytiikkaprosesseitansa datan integroinnin puutteen tuottavan seuralle ongelmia. Seura onnistuikin datan integroinnin avulla ratkaisemaan datan hallinnalliset ongelmat ja kehittämään näin koko organisaation toimintaa tehokkaammaksi (Tan, Hedman & Xiao, 2017).

Seuran johdon tulisi määrittää Alamarin (2013, s. 31) mukaan selkeät suositukset siitä, miten ja minne dataa tallennetaan. Tämä voi olla vaikea vaihe prosessissa, sillä työntekijät todennäköisesti joutuvat muuttamaan vanhoja tapojaan. Kuitenkin vasta näiden muutosten jälkeen voidaan urheiluanalytiikkaan tehtyjen investointien odottaa tuottavan potentiaalista hyötyään. Datan hallinta onkin yksi hyvin tärkeä urheiluanalytiikan rakennuspalikka (Alamar, 2013, s. 6-7).

3.1.2 Analyttiset mallit

Analyttisillä malleilla on monia eri käyttötapoja, mutta niiden tarkoitus urheiluanalytiikassa on sama. Analyttisten mallien avulla raaka data muutetaan luotettavaksi ja käyttökelpoiseksi informaatioksi, jonka avulla voidaan tehdä perusteltuja johtopäätöksiä urheiluseuran päätöksenteon tueksi. Huolellinen analyysi käy läpi kaiken datan, löytää merkityksellisiä yhteyksiä siitä löytyvien muuttujien välillä ja sitten käyttää näitä yhteyksiä tarjotakseen ymmärrystä urheiluseuran tai sen pelaajien suorituksista (Alamar, 2013, s. 7-8). Monet seurat käyttävät analyttisiä malleja esimerkiksi auttamaan parhaiden mahdollisten pelaajien valitsemisessa urheiluliigojen järjestämissä varaustilaisuuksissa, joissa ammattilaisseurat hankkivat pelaajia aloittelijajoukkueista. (Davenport, 2014a) Hyödyntäessä erilaisia analyttisiä malleja tulisi seurojen muistaa, että moniin seurojen arvioitaviin asioihin liittyy lukuisia eri muuttujia urheilun toimialan ominaispiirteiden takia. Yksittäisen pelaajan ottelusuorituksiin liittyvät esimerkiksi pelaajan joukkuekaverit, vastustaja, käytetty pelijärjestelmät sekä pelaajan kyvykkyydet. Analyttisten mallien avulla pyritään eri lähtökohdista tulevista pelaajista tekemään vertailukelpoisia ottamalla nämä muuttujat huomioon. Kyky tunnistaa ja arvioida edellä mainittujen kaltaisia muuttujia on haaste, johon analyttiset mallit pyrkivät vastaamaan (Alamar, 2013, s. 7-8).

On kuitenkin tärkeä muistaa, että analyttisten mallien tarkoitus on tarjota informaatiota päätöksentekijöille, eikä tehdä päätöksiä. Analyttisten mallien avulla pyritään saamaan kaikki saatavilla oleva relevantti informaatio päätöksentekijöille käytettäväksi. Hyödyntäessä analyttisiä malleja päätöksentekijöiden tulisi nähdä mallit työkaluina, jotka saattavat auttaa vähentämään epävarmuuksia, auttamaan todistamaan informaatioita todeksi tai ristiriitaisissa tapauksissa johtamaan lisäkysymyksiin. Onnistuessaan analyttiset mallit voivat olla erittäin hyödyllisiä työkaluja päätöksentekijöille (Alamar, 2013, s. 7-8).

3.1.3 Tietojärjestelmät

Tietojärjestelmien tehtävä urheiluanalytiikassa on toimittaa datasta muokattu informaatio päätöksentekijöille mielekkäällä, tehokkaalla, yhtenäisellä sekä vuorovaikutteisella tavalla. Tietojärjestelmien tehtävä on myös järjestellä ja esitellä informaatiota, jotta päätöksentekijät voisivat käyttää enemmän aikaa tiedon analysoimiseen organisoinnin sijaan. Kun tietojärjestelmä on käyttöön otettu optimaalisella tavalla urheiluseuraan, tulisi jokaisella päätöksentekijällä

olla sama informaatio saatavilla. Näin päätöksentekijät analysoivat ja tuottavat omia päätelmiään varmasti samaan informaatioon perustuen. Tietojärjestelmät mahdollistavat myös sen, että päätöksentekijät kykenevät olemaan vuorovaikutuksessa informaation kanssa, jolloin valmentajien esittämiin täsmällisiin kysymyksiin saadaan vastattua (Alamar, 2013, s. 8-9).

Tietojärjestelmää suunniteltaessa, tulisi organisaatioilla olla riittävän hyvä ymmärrys nykyisistä järjestelmistään, datan lähteistään, erityyppisten informaatioiden käyttötavoista päätöksentekoprosesseissaan sekä siitä, miten päätöksentekijät ovat yleisesti vuorovaikutuksessa informaation kanssa organisaatiossa. Urheiluseuralla tulisi olla sen päätöksentekoprosesseista mahdollisimman selkeä kuva, jotta tietojärjestelmät suunniteltaisiin täsmällisesti tukemaan ja parantamaan näitä prosesseja (Alamar, 2013, s. 79). Oikein kriteerein rakennettu tietojärjestelmä on urheiluseuralle työkalu, joka sulauttaa urheiluseuran keräämän datan kyseisen organisaation strategian kanssa (Alamar, 2013, s. 89).

3.2 Urheiluanalytiikan osa-alueet

Urheiluseurojen eri osa-alueilta saatavilla olevan datan määrä on nykyään merkittävää. Erityyppistä dataa kerätään valtavalla skaalalla videodatasta fysiologiseen sensoridataan. Urheiluanalytiikkaa käsiteltäessä moni mieltää käsitteen pelaajien rekrytoimisen ja sopivien kentällisten valitsemisen kautta. Urheiluanalytiikka on silti paljon muutakin. Davenport (2014a) jakaa urheiluanalytiikan kolmeen osa-alueeseen, joita ovat:

1. Otteluiden ja pelaajien suoritusten analytiikka
2. Urheiluliiketoiminnan analytiikka
3. Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka

Seuraavissa alaluvuissa käsittelemme kyseiset kolme urheiluanalytiikan eri osa-aluetta kertoen tieteellistä kirjallisuutta käyttäen alan käytänteistä. Urheiluanalytiikan käytänteet ovat Davenportin (2014a) mukaisesti jaoteltu kahteen kategoriaan, kilpailuetua tuottavaan analytiikkaan sekä analytiikan tukikeinoihin. Tukikeinoilla tutkielmassa viitataan urheiluanalytiikan sovellutuksiin, jotka ovat muodostumassa tai jo muodostuneet yleisiksi käytänteiksi ammattiurheilun toimialalla. Näiden urheiluanalytiikan tukikeinojen tehokas hyödyntäminen tarjoaa organisaatiolle hyötyjä, mutta niiden avulla ei saavuteta yleensä merkittävää kilpailuetua, sillä konseptit ovat jo hyvin laajalle levinneitä (Davenport, 2014a).

3.2.1 Otteluiden ja pelaajien suoritusten analytiikka

Monet Pohjois-Amerikan urheiluseuroista ovat jo pitkään hyödyntäneet otteluiden ja pelaajien suoritusten analytiikkaa urheilusarjojen varaustilaisuuksissa. Analysoimalla tulokaspelaajien historiaa ja arvioimalla seuran nykyisen pelaajamateriaalin vahvuuksia sekä heikkouksia, seurat pyrkivät asettamaan varattavissa olevat pelaajat paremmuusjärjestykseen (Alamar, 2013, s.7-8; Davenport 2014a). Luokittelemiensä järjestysten avulla seurat pyrkivät varaustilaisuudessa valitsemaan seuraansa pelaajia, jotka kaikista suurimmalla todennäköisyydellä onnistuvat sopeutumaan ammattilaissarjoihin vastustajien parantuessa ja pelin vaatimustason noustessa (Alamar, 2013, s.7-8). Varaustilaisuuksia varten toteutettavat analyttiset lähestymistavat alkavat olla jo hyvin yleisiä, eivätkä ne enää tarjoa kilpailullista etua ellei seura pyri jatkuvasti etsimään uusia lähestymistapoja tähän pelaajien arvioittamiseen.

Urheiluanalytiikan tukikeinot vaihtelevat lajikohtaisesti. Esimerkiksi baseballissa otteluiden ja pelaajien analytiikan tukikeinoihin kuuluvat erilaiset lyöntimittarit, kun taas koripallossa tukikeinoja ovat pelaajakohtaiset plus/ miinus-tilastot, jotka kuvaavat miten joukkueen päästämät ja tekemät korit ovat jakautuneet analysoinnin kohteena olevan pelaajan ollessa kentällä (Ganeshapillai & Guttag, 2012; Davenport, 2014a). Joukkueet ovat koripallossa myös yleisesti hyvin tietoisia eri heittojen onnistumisodotteista. Yleinen käsitys lajissa on, että kaikista suurimmat onnistumisodotteet ovat kolmen pisteen viivan takaa lähteville heitoille sekä hyvin läheltä koria tapahtuville yrityksille (Davenport, 2014a). Samalla toimintaperiaatteella maaliodottamia on käytetty muidenkin urheilulajien analytiikassa. Esimerkiksi Macdonald (2012) raportoi jääkiekon maaliodottama-mallista. Maaliodottamat perustuvat historiallisiin tilastoihin eri alueilta suoritettujen maalintekoyritysten onnistumisprosentteihin. Tavoitteena näillä analytiikan malleilla on pyrkiä arvioimaan seurojen ja pelaajien peliesityksiä mahdollisimman objektiivisesti (Macdonald, 2012). Huolimatta siitä, että mallien avulla saavutettu tieto on jo yleistä, on esimerkiksi eri koripallojoukkueiden heittovalinnoissa suuria eroavaisuuksia keskenään. Jotkut seurat noudattavat heittojen onnistumisodotteita hyvinkin kurillisesti muihin seuroihin verrattuna. (Davenport, 2014a).

Otteluiden ja pelaajien suoritusten analytiikka on perinteisesti painottunut tilastoihin kuten syöttö- ja laukaisumääriin. Esimerkiksi jalkapallojoukkueet ovat perinteisesti keskittyneet pääasiallisesti siihen mitä kentällä tapahtui, sen sijaan, että painopisteenä olisi miksi joku tapahtuma tapahtui tai mitä seuraavaksi saattaa tapahtua (Davenport, 2014a). Tyypillisesti jalkapallon otteluiden ja pelaajien analytiikka on koostunut laukausmääristä, mutta tällöin jää tutkimatta mahdollisuus, että olennainen osa joukkueiden hyökkäämistä tapahtuu jo ennen maalipaikan luomista. Sama ilmiö ulottuu kaikkiin joukkueurheilulajeihin. Suurin ongelma asian ratkaisemisessa on kykenemättömyys arvioida analyttisesti yksilöiden taktisia päätöksiä. Otteluiden ja pelaajien suoritusten analytiikassa ei ole vielä yrityksistä huolimatta onnistuttu luomaan metodia, jonka

avulla kyettäisiin vertailemaan esimerkiksi ratkaisua yrittää maalintekoa, verrattuna syöttöön vartioinnista vapaalle joukkuekaverille (Cervone, D'Amour, Bornn & Goldsberry, 2014; Goldman & Rao, 2011).

Otteluiden ja pelaajien suoritusten analytiikan yksi lähes kaikkia urheilulajeja yhdistävä seikka on datan lähteiden lisääntyminen. Monet urheilusarjat ja -liitot, kuten koripallosarja NBA tai baseballsarja MLB, keräävät joukkueiden ja pelaajien suorituksista itse dataa. Hyvin yleistä on myös, että urheilusarjoilla on lisäksi ulkopuolinen datan tarjoaja, joka tuottaa avointa dataa kaikille seuroille ja näiden kannattajille. Nämä tilastot ovat Davenportin (2014a) mukaan kuitenkin perinteisesti olleet kuvailevia, hyvin yksinkertaisia, tilastoja. Muita otteluiden ja pelaajien analytiikan tukikeinoja ovat palkkojen ja kokoonpanojen optimointi, sekä otteluiden simuloiminen.

Kaikkien urheilulajien parissa merkittävin tällä urheiluanalytiikan osa-alueella kilpailuetua tuottava datan tyyppi on tällä hetkellä videodata. Videodata helpottaa merkittävästi urheilujoukkueiden ja pelaajien fyysisen sekä taktisen analyysin tekemistä (Zhu ym. 2009; Li ym. 2010; D'Orazio & Leo, 2010). Urheiluliigat ovat alkaneet kerätä, analysoida ja visualisoida otteluiden videodataa, jota on kerätty automaattisten kamerajärjestelmien avulla (Pileggi, Stolper, Boyle & Stasko, 2012; Davenport, 2014a). Videodatan käyttö urheiluanalytiikassa sisältää vielä monia haasteita. Yksi iso haaste sen hyödyntämisessä on D'Orazion ja Leon (2010) mukaan merkittävien tapahtumien louhinta datasta automaattisia videoleikkeiden tuottamista varten. Videoanalytiikan avulla pidetään kirjaa tilastollisista tapahtumista, kuten pallokosketuksista ja syötöistä. Monimutkaisemman analyysin tekemiseen, kuten esimerkiksi laskemiseen kuinka usein tietty pelaaja kuljettaa vasemmalle yrittäessään ohittaa puolustajaa, tarvitaan kuitenkin vielä manuaalista työtä (Davenport, 2014a). Erilaisten kamerajärjestelmien avulla tapahtuva pelaajien ja pelivälineiden seuranta on ollut aktiivinen tutkimusaihe viime vuosina ja aiheeseen on esitetty monia erilaisia lähestymistapoja. Vaikka optisten seurantamenetelmien tarkkuutta on onnistuttu kehittämään, tuottavat ne vielä virheitä ja epätarkkuuksia. Videodata on kuitenkin jo nyt yksi kilpailuetua kasvattava tekijä ja se tulee Halvorsenin ym. (2013) mukaan tulevaisuudessa muuttumaan entistä tärkeämmäksi osaksi otteluiden ja pelaajien analytiikkaa

Toinen merkittävästi kilpailuetua tällä hetkellä tuottava otteluiden ja pelaajien analytiikan osa-alue on sijaintia paikantavat sekä biometriikkaa hyödyntävät laitteet ja sensorit, kuten GPS-laitteet, radiotaajuuslaitteet, kiihtyvyyssanturit sekä muut biometriset sensorit. Näiden teknologioiden avulla pelaajien liikumista ja muuta fyysistä aktiivisuutta voidaan mitata. Monet Englannin Valioliigan jalkapallojoukkueista käyttävät GPS-laitteita ja kiihtyvyyssantureita harjoituksissaan. Laitteiden avulla saadaan kerättyä dataa harjoituksista, mutta laitteiden käyttö pelitilanteissa on vielä toistaiseksi kiellettyä. Sijaintitietoihin perustuvia laitteita käytetään tästä joukosta merkittävästi eniten. Näiden laitteiden yleisin käyttötarkoitus on Davenportin (2014a) mukaan pelaajien juoksumäärien ja -nopeuksien mittaaminen. Jotkut urheiluseurat ovat myös ottaneet

käyttöönensä itse luomiaan mittareita arvioimaan pelaajien ominaisuuksia (Mondello & Kamke 2014). Seurojen yksinoikeudella omaan käyttöönsä tuottama data on Davenportin (2014a) mukaan yksi kilpailuetua edistävistä tekijöistä. Urheiluseurat, jotka omaksuvat uusia teknologioita datan luomiseen säilyttävät uusista menetelmistä saadun kilpailuetunsa, vaikka teknologiat yleistyisivät myöhemmin muiden urheiluseurojen parissa. Jotkut yksittäiset urheiluseurat, kuten englantilainen jalkapalloseura Manchester City, ovat tehneet osasta datastaan julkista, jolloin seuran kannattajat ja muut siitä kiinnostuneet ihmiset pystyvät myös analysoimaan sitä (Mondello & Kamke, 2014; Davenport, 2014a). Tällainen kannattajien suorittama avoimen datan analytiikka on myös yksi kilpailuedun luojista.

Viimeiseksi kilpailuedun luojaksi Davenport (2014a) määrittää pelaajien sitouttamisen omien suorituksiensa analytiikkaan. Seuraavassa taulukossa on esitelty otteluiden ja pelaajien suoritusten analyttiset keinot jaoteltuina urheiluorganisaatioiden analytiikan tukikeinoihin sekä kilpailuetua tuottaviin tekijöihin (taulukko 1).

TAULUKKO 1 Otteluiden ja pelaajien suoritusten analytiikka (Davenport, 2014a)

Analytiikan tukikeinot	Kilpailuedun luojat
Ulkoisten datalähteiden käyttö	Videodatan analytiikka
Pelaajien kuvaileva analytiikka	Sijainti- ja biometrisen datan analytiikka
Kokoonpanojen optimointi	Faniien toteuttama avoimen datan analytiikka
Pelaajien arvioiminen varaustilaisuuksia varten	Pelaajien sitouttaminen analytiikkaan
Pelaajien palkkojen optimointi	Yksinoikeudellisen datan kerääminen ja analysointi
Otteluiden simulointi	
Ottelutaktiikoiden analytiikka	

3.2.2 Urheiluliiketoiminnan analytiikka

Toinen urheiluanalytiikan osa-alue on urheiluliiketoiminnan analytiikka. Urheiluliiketoimintaan katsotaan kuuluvan urheiluseurojen sekä -liigojen sponsorisopimukset, lipputulot, mediaoikeudet ja kannattajille myytävät fanituotteet (Shah, Kretzer & Mädche, 2015). Joukkuelajit kattavat urheiluliiketoiminnasta yli 80 % ja kaikista lajeista jalkapallo on ylivoimaisesti suurin, kattaen ammatturheilun toimialasta enemmän muut joukkuelajit yhteensä (Kearney, 2011). Vaikka analytiikan käyttö urheiluseurojen urheilullisilla puolilla on yleistynyt, analytiikan omaksuminen urheilun liiketoimintaan on vielä useasti rajoitunutta. Urheiluliiketoiminnan analytiikka ei kohtaa samanlaista innostuneisuutta kuin muut urheiluseurojen analytiikan osa-alueista (Mondello & Kamke, 2014; Troilo ym. 2016). Valtaosa urheiluorganisaatioista ei ole yhtä kehittyneitä

liiketoiminnan analytiikassa, kuin mitä johtavat, muiden alojen, yritykset ovat (Davenport, 2014a).

Urheiluliiketoiminnan analytiikka voi näkyä seuroissa eri tavoin. Lippujen hinnoittelumenetelmät, kannattaja-analyysit, myynninedistäminen sekä kannattajien sitouttaminen seuraan ovat kaikki urheiluliiketoiminnan analytiikan osaluokkia. Urheiluseurojen liiketoiminnan analytiikassa tukikeinoja ovat joustava lippujen hinnoittelu sekä kannattaja-analyysien tekeminen. Lippujen hinnoittelu on Davenportin (2014a) mukaan seuroille hyvä tapa ansaita ylimääräisiä myyntituloja analytiikan kautta. Nykyään melko yleinen lähestymistapa lippujen hinnoitteluun on edellä mainittu joustava hinnoittelu, jossa lippuja myydään eri otteluihin eri hinnoilla. Näiden joustavan hinnoittelun lippuhintojen muutokset ovat kuitenkin etukäteen määriteltyjä. Baseballsarja, Major League Baseballin kolmestakymmenestä seurasta 26 käytti vuonna 2014 jonkinasteista joustavaa ottelulippujen hinnoittelua (Davenport, 2014a).

Monimutkaisempi lippujen hinnoittelun lähestymistapa on dynaaminen hinnoittelu, jossa lippujen hinnat määrittyvät täysin kysynnän ja tarjonnan mukaan (Davenport, 2014a). Dynaamisessa hinnoittelussa urheiluorganisaatiot tekevät reaaliaikaisia mukautuksia lippujen hintoihin lukuisten tekijöiden mukaan, joita voivat olla muun muassa joukkueen edellisten otteluiden tulokset, loukkaantumistilanne, ottelun viikonpäivä tai sää. Dynaamisen hinnoittelun päätarkoitus on lisätä lippuostojen määriä sekä niistä saatavia tuottoja. Lähes puolet kaikista Pohjois-Amerikan isojen urheiluliigojen joukkueista käyttivät kotiotteluissaan dynaamista hinnoittelua (Harrison & Bukstein, 2016, s. 9).

Jotkut yksittäiset seurat ovat myös kokeilleet personalisoitujen internet-sisältöjen luomista seuran kotisivuilla käyville kannattajalle. Davenport (2014a) listaa internet-sisällön personalisoinnin yhdeksi kilpailuetua luovaksi tekijäksi urheiluliiketoiminnan analytiikassa. Muita kilpailuetua tuottavia tekijöitä Davenportin (2014a) mukaan ovat segmentointi ja optimointi, joiden tarkoituksena on kehittää kohdennettuja markkinointitapoja asiakkaiden historioihin sekä aiempiin ostoksiin perustuen. Esimerkiksi koripalloseura Orlando Magicilla on käytössään malli, joka kategorisoi seuran kausikorttiostajat sen perusteella, kuinka todennäköisesti asiakas kausikorttitilauksensa uusii. Kategorisoituaan asiakkaat, Orlando Magicin asiakaspalvelijat keskittyvät kauden loppumisen jälkeen asiakkaisiin, jotka eivät ole selkeästi kausikortin uudistamista vastaan tai sen puolesta (Harrison & Bukstein, 2016, s. 25–26; Davenport, 2014a). Saksalainen jalkapalloseura Bayern München on vienyt asiakkaiden kategorisointinsa vielä pidemmälle. Seura on luonut 80 erilaista kannattajaroolia sen otteluissa käyville ihmisille. Tämä mahdollistaa tehokkaamman kommunikoinnin organisaation ja sen eri asiakasryhmien välillä. Bayern München kerää ottelutapahtumistaan myös dataa, joka auttaa seuran tapahtumajärjestäjiä ottelupäivinä. Dataa kerätään seuran kotistadionilla Allianz Arenalla reaaliaikaisesti eri järjestelmistä kuten kassasysteemeistä, parkkihalleista ja sisäänkäyntiporteista. Tiedot siirtyvät tapahtumajärjestäjien mobiililaitteille, jossa seuran luoma sovellus tarjoaa työntekijöille tietoa esimerkiksi siitä, kuinka paljon ihmisiä stadionin sisällä

jo on. Mikäli stadionilla on sovellukseen mukaan huomattavasti keskiarvoa vähemmän ihmisiä, hälyttää se työntekijöitä selvittämään mikä katsojien saapumisesta stadionille hidastaa (Tan, Hedman & Xiao, 2017).

Langattoman verkon käyttäminen kannattajien toiminnan ymmärtämistä varten on Davenportin (2014a) mukaan yksi urheiluliiketoiminnan kilpailuedun luojusta. Esimerkiksi NBA-joukkue Phoenix Suns on tehnyt yhteistyötä teleoperaattori Verizon Wirelessin kanssa oppiakseen faniensa ennen otteluita sekä niiden jälkeen tapahtuvasta käytöksestä. Koska Verizon tietää asiakkaidensa puhelimien sijainnit, pystyy se informoimaan koripalloseuraa, missä sen asiakkaat ovat tiettyihin kellonaikoihin. Näiden sijaintitietojen perusteella yritys pystyy informoimaan urheiluorganisaatiota, jolloin seura voi esimerkiksi kohdentaa markkinointiaan optimaalisille asuinalueille (Davenport, 2014a). Seuraavassa taulukossa on esitelty urheiluliiketoiminnan analyttiset keinot tukikeinoihin sekä kilpailuetua tuottaviin tekijöihin jaoteltuina (taulukko 2).

TAULUKKO 2 Urheiluliiketoiminnan analytiikka (Davenport, 2014a)

Analytiikan tukikeinot	Kilpailuedun luojat
Muuttuva hinnoittelu	Dynaaminen hinnoittelu
Kannattajien tyytyväisyyden ja lojaaliuden analytiikka	Kannattajien segmentointi ja personalisointi
	Kannattajien sitouttaminen sosiaalisen median ja wifi:n avulla
	Markkinoinnin optimoiminen
	Mobiililaitteiden sijaintitietojen analytiikka

3.2.3 Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka

Kolmas urheiluanalytiikan kategoria on pelaajien terveyden edistämisen ja loukkaantumisten ehkäisyn analytiikka. Urheilujoukkueet ymmärtävät, että pelaajat ovat seurojen ylivoimaisesti kallein resurssi (Davenport, 2014b). On arvioitu, että Major League Baseball-seurojen pelaajat olivat vuonna 2012 n. 29 000 päivää loukkaantuneina ja loukkaantumisien hoitoon kului seuroilta noin 600 miljoonaa dollaria. Onkin hyvin luonnollista, että urheiluseurat haluavat analytiikallaan keskittyä myös siihen, että organisaation tärkeimmät resurssit, pelaajat, pysyvät terveisinä (Davenport, 2014a).

Urheiluseuroilla on ollut pelaajien loukkaantumishistorian seurantaan sekä kuntoutumisen dokumentointiin jo vuosikymmeniä erilaisia ohjelmistoja käytössä. Datamäärien kasvun johdosta pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikan painopiste on alkanut kääntyä loukkaantumisten seurannasta kohti niiden hallintaa sekä urheilijoiden pitkäaikaisen terveyden parantamista. Urheiluseurojen pelaajat eivät enää poistu louk-

kaantumisten hallintaprosesseista, vaan heidän terveyttään tarkkaillaan jatkuvasti, sillä raja loukkaantumisen, kuntoutuksen ja kilpailukyvyn välillä on ammattiurheilussa hyvin häilyvä. Loukkaantumisten hallinnassa onnistumisen määrittää pääasiallisesti se, miten nopeasti urheilija pääsee takaisin täyteen kuntoon (Hanisch, B. & Hanisch J. 2007). Pelaajien terveyden laaja-alainen monitorointi voi olla seuroille kallista, mutta Wilkersonin ja Guptan (2016) mukaan mobiililaitteiden käyttö pelaajien loukkaantumisriskien kartoituksessa voi nostaa tällä urheiluanalytiikan osa-alueella seurojen tehokkuutta.

Pelaajien loukkaantumisten ehkäiseminen ei ole yksinkertaista, sillä loukkaantumiset saattavat johtua monista eri tekijöistä ja niitä on useasti vaikea ennakoida. Eri seurojen valmentajat saattavat olla eri mieltä, miten seurojen tulisi lähestyä pelaajien loukkaantumisten ehkäisyä ja näin ollen seurojen fysiikkavalmentajat saattavat painottaa harjoittelussaan täysin eri asioita (Davenport, 2014a). Toinen merkittävä haaste pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikassa on datan määrä yhdistettynä ammattiurheilun kilpailutahtiin. Jotta pelaajan aika loukkaantuneena minimoitaisiin, tulisi hoitotoimeenpiteiden olla tiedossa mahdollisimman pian loukkaantumisen tapahtumisesta (Hanisch, B., & Hanisch, J. 2007).

Kyseisellä urheiluanalytiikan osa-alueella käytetään monesti samaa dataa otteluiden ja pelaajien suoritusten analytiikan kanssa. Pelaajan aktiivisuutta kuvailevan datan analytiikka on tukikeino organisaatioille pelaajien terveyden analytiikassa. Kilpailuetua terveyden edistämisen ja loukkaantumisten ehkäisyn analytiikassa tuottavat taas analyttiset keinot, joiden avulla kyetään havainnoimaan mahdollisia tulevia loukkaantumisia. Loukkaantumiseen mahdollisesti vaikuttavia tekijöitä voidaan tunnistaa esimerkiksi videodatasta pelaajan liikkumistapoja tutkien. Myös erilaisilla sijainnin ja aktiivisuuden mittareilla voidaan kerätä biometristä dataa, joka saattaa paljastaa pelaajista mahdollisia riskitekijöitä, kuten esimerkiksi lihasväsymystä tai heikkoja unitottumuksia. Myös biomekaaninen analytiikka on kilpailuetua urheiluseuroille luova analytiikan keino. Biomekaanisella analytiikalla mitataan muun muassa pelaajien neuromuskulaarista koordinaatiota, ruumiinasennon tasapainoja, epätasapainoja sekä pelaajan reaktioaikoja (Gruetzemacher, Gupta & Wilkerson, 2016).

Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikassa ollaan kuitenkin vielä alkutekijöissä. Datan lähteet kasvavat tällä osa-alueella hyvin nopeasti, minkä johdosta loukkaantumisia yksityiskohtaisesti ennustavat analytiikan keinot alkavat hiljalleen olla monille urheiluseuroille realistisia (Davenport, 2014a). Seuraavassa taulukossa on esitelty pelaajien terveyden edistämisen ja loukkaantumisten ehkäisyn analyttiset keinot jaoteltuina ne urheiluorganisaatioiden analytiikan tukikeinoihin sekä kilpailuetua tuottaviin keinoihin (taulukko 3).

TAULUKKO 3 Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisyn analytiikka (Davenport, 2014a)

Analytiikan tukikeinot	Kilpailuedun luojat
Pelaajien aktiivisuuden kuvaileva analytiikka	Loukkaantumisten analytiikka videodatan avulla Sijainti- ja biometrisen datan analytiikka Loukkaantumisten ennustava analytiikka Yksityiskohtainen biomekaaninen analytiikka

3.3 Urheiluanalytiikan yhteenveto

Urheiluanalytiikkaan kohdistuu sen tuomista mahdollisuuksista huolimatta paljon vastarintaa (Lee, 2016, s. 1). Ammattiurheilun toimialalla, analytiikalla on kuitenkin ollut käynnissä virtaus kohti entistä analyttisempää seurajohtamista (Davenport, 2014). Tämän kehittymisen ovat Alamarin (2013, s.1-3) mukaan mahdollistaneet teknologioiden kehittyminen sekä käytettävissä olevan datan kasvu.

Luvussa käytiin läpi Alamarin (2013) urheiluanalytiikan viitekehys, jonka komponentteja ovat datan hallinta, analyttiset mallit, tietojärjestelmät, päätöksentekijät sekä johtajuus. Tämän jälkeen esiteltiin viitekehysten kolme teknistä komponenttia ja selvitettiin, mikä niiden merkitys on urheiluorganisaatioiden analytiikassa. Lopuksi esiteltiin urheiluanalytiikan kolme osa-aluetta Davenportin (2014a) mukaan ja jaoteltiin urheiluanalytiikan yleiset käytänteet urheilu-seuroille kilpailuetua tuottaviin käytänteisiin sekä tukikeinoina toimiviin käytänteisiin.

4 Massadata urheiluanalytiikan apuna

Seuraavassa luvussa pyrimme selvittämään erilaisia tapoja, joiden avulla massadataa voitaisiin hyödyntää urheiluanalytiikassa. Koska massadatan käsite on riippuvainen sen kontekstista, pyrimme ensimmäisessä alaluvussa määrittelemään tieteellisen kirjallisuuden avulla, mitä on massadata ammattiurheilun toimialalla. Tämän jälkeen toisessa alaluvussa pyrimme selvittämään, millä eri keinoin tätä massadataa voidaan hyödyntää urheiluanalytiikassa.

4.1 Massadata urheiluanalytiikassa

Massadatan määrän ominaisuudella urheiluanalytiikassa tarkoitetaan toimialalla käytettävien datamäärien kokoja. Esimerkiksi yhden jalkapallo-ottelusta XML:llä koodatun sijaintitieto-tiedoston koko vaihtelee 86 ja 300 megatavun välillä. Näin ollen yhdeltä jalkapallokaudelta kerätyn sijainti- ja videodatan johdosta syntyy noin 400 gigatavua seurantadataa. Tämän lisäksi seurat keräävät myös muuta dataa, kuten fysiologista dataa sekä tapahtumadataa. (Rein & Memmert 2016). Tällaisen kokoluokan datamassat eivät tyypillisesti täyttäisi massadatan määritelmää (Pääkkönen & Pakkala 2015). Massadataksi luokiteltavien datamassojen määrä vaihtelee kuitenkin toimialalla käytettävien datan hallinta-standardien mukaan. Ammattiurheilun toimialalla tyypilliset dataratkaisut liittyvät monesti Excel-taulukoiden käyttöön, eivätkä nämä menetelmät skaalaudu hyvin edellä mainitun kaltaisen datan kanssa (Manyika, 2009; Rein & Memmert, 2016). Tästä syystä määrittelemmekin tähän ammattiurheilun kontekstiin massadatan määrän ominaisuuden täyttyvän erilaisen sijaintidatan, videodatan sekä fysiologisen sensoridatan määrissä.

Massadatan nopeuden ominaisuudella kuvaillaan sitä nopeutta, jolla uutta dataa syntyy sekä nopeutta, jossa se tulisi myös pystyä analysoida (Gandomi & Haider, 2015). Urheiluanalytiikassa datan nopeus esiintyy fysiologisen sensoridatan sekä sijaintitietoihin perustuvien datankeräämis-menettelmien reaaliaikaisessa datavirrassa (Rein & Memmert 2016). Myös aiemmin esiteltyjen automaattisten kamerajärjestelmien katsotaan täyttävän massadatan nopeusominaisuus tässä tutkielmassa. Tällaisten optisten seurantamenettelmien tuottama videodata on usein reaaliaikaista ja otteluiden analytiikassa urheilujoukkueiden valmentajien tulee Sacha ym. (2014) mukaan pyrkiä tekemään päätöksiä nopeasti, jopa kilpailuiden sisällä. Näin perustellen katsomme massadatan nopeusominaisuuden täyttyvän urheiluanalytiikassa fysiologisessa sensoridatassa, sijaintidatassa sekä videodatassa.

Viimeinen kolmesta massadatan yleisesti tunnistetuista ominaisuuksista on datan monimuotoisuus. Monimuotoisuudella tarkoitetaan datan moniulotteisuutta, jossa se esiintyy (Wu ym. 2014). Urheiluanalytiikassa dataa syntyy lukuisissa eri muodoissa. Muun muassa videodata, XML-data sekä notaatiodata

ovat esimerkkejä hyvin erilaisista datan tyypeistä, joita urheiluanalytiikassa esiintyy. Näiden lisäksi myös fysiologinen data sekä sijaintidata esiintyvät useissa eri datan muodoissa. (Rein & Memmert 2016). Näin ollen toteamme urheiluanalytiikassa massadataa olevan fysiologisen sensoridatan, sijaintidatan sekä videodatan.

4.2 Massadatan hyödyntäminen urheiluanalytiikassa

Selvittääksemme erilaisia keinoja, joilla massadataa voitaisiin hyödyntää ammattuurheilun toimialalla, määrittelimme urheiluanalytiikan massadatan kolmeen eri datan tyyppiin. Nämä eri tyypit, joissa massadata urheiluanalytiikassa esiintyy, ovat videodata, biometrinen data sekä sijaintidata. Seuraavaksi selvitämme, miten urheiluorganisaatiot voivat käytännössä hyödyntää tätä massadataa toiminnassaan kilpailuedun saavuttamiseksi.

4.2.1 Videodata

Ammattuurheilussa valmentajien rooli urheiluanalytiikassa on merkittävä, koska he vastaavat joukkueen valmistautumisesta otteluihin. Tästä syystä valmentajien on hyvin tärkeää analysoida oman joukkueensa ja vastustajaansa, jonka yksi oleellinen analysoinnin osa on otteluiden katsominen. Videodatan hyödyntäminen otteluiden analysoinnissa ja niihin valmistautumisessa on hyvin yleistä ammattuurheilussa (Davenport 2014a). Kokonaisten otteluiden katsominen ei kuitenkaan ole ajallisesti tehokasta ja osittain tästä syystä videodatan kerääminen sekä sen eri prosessien automaattisointi on hyvin ajankohtainen tutkimusaihe. Automaattisointi voi tapahtua eri videodatan prosessien vaiheissa, kuten esimerkiksi videoiden kuvaamisessa, leikkaamisessa tai niiden avulla luotujen tilastojen tuottamisessa (Zhu ym. 2009; D’Orazio & Leo; 2010 Li ym. 2010; Kim, Kwon & Li, 2011; Sacha ym. 2014; Han & Farin 2005). Etenkin joukkueurheilulajeissa lukuisten pelaajien yhtäaikainen optinen seuranta sisältää kuitenkin teknisiä haasteita, sillä pelaajat tekevät otteluissa paljon äkillisiä liikkeitä ja ovat fyysisessä kontaktissa toisiinsa (Xing ym. 2011).

Vaikka videodatasta automaattisesti tuotettujen videopätkien analysointi on vielä alkuvaiheessa, on Shah ym. (2015) mukaan videoista mahdollista tunnistaa ennalta määrättyjä tapahtumia. Näiden automaattisesti tuotettujen lyhyiden videopätkien tehtävä on etsiä videodatasta merkityksellisiä videoleikkeitä ja erotella ne erillisiksi videotiedostoiksi (Xing ym. 2011). Tämä mahdollistaa valmentajille sen, että heille jää aikaa videopätkien analysointiin niiden keräämisen sijaan (Alamar, 2013, s. 5-6).

Videodataa voidaan hyödyntää myös tilastojen luomisessa. Verrattuna ihmisten manuaalisesti tekemiin tilastoihin, on videodatan automaattinen analysointi huomattavasti käytännöllisempää (Davenport, 2014a; Shah ym. 2015).

Automaattisen analyysin avulla videodatasta voidaan tuottaa tilastoja alle sekuntien väliajoin ja esimerkiksi yksi johtavista urheiludatayhtiöistä, Opta Sports, kerää yhdestä jalkapallo-ottelusta 80 miljoonaa datapistettä. Videodatan avulla luotuja tilastoja ja videopätkiä valmentajat hyödyntävät päivittäisessä toiminnassaan tavoitteena kehittää yksittäisten pelaajien ja joukkueen suorituksia (Shah ym. 2015; Tan, Hedman & Xiao, 2017).

Videodataa keräävät optiset seurantamenetelmät ovat myös alkaneet mahdollistaa urheiluseuroille hiljalleen entistä syvällisempää analyysiä siitä, missä, miten ja miksi kentällä toteutuneet tapahtumat tapahtuivat (Xing ym. 2011). Esimerkiksi Major League Baseballin joukkue San Francisco Giants otti ensimmäisenä joukkueena käyttöönsä SportVision-yrityksen videodatapalvelun. Palvelusta saadun videodatan avulla seura pystyi seuraamaan ja määrittelemään useita, vaikeasti määriteltäviä, pelaajan arvonmittareita. Monet pelaajat myös itsenäisesti katsovat omia suorituksiaan ja analysoivat automaattisesti generoitujen videopätkien avulla omia suorituksiaan (Davenport, 2014a).

Videodataa voidaan myös käyttää arvioimaan pelaajien fyysistä kuormitusta tai liikemalleja, jotka saattavat johtaa loukkaantumisiin (Davenport 2014a). Yksi ensimmäisistä esimerkeistä suurten videodatamassojen hyödyntämisestä pelaajien terveyden edistämisessä ja loukkaantumisten ehkäisyssä oli jalkapallojoukkue AC Milanin rakentama MilanLab. MilanLab on seuran oma pelaajien terveyden edistämisen ja analytiikan tutkimuskeskus, jossa on otettu käyttöön lukuisia pelaajien fyysisen puolen mittaamisen teknologioita. AC Milan jokaisesta yksittäisestä pelaajasta kerätään noin 60 000 datapistettä, jonka lisäksi seura kerää pelaajistaan myös henkistä, biomekaanista sekä lihaksistollista dataa. Videodataa pelaajien suorituksista, kuten vertikaalisesta hypystä, kerätään kahden viikon välein eri laitteilla. (Davenport, 2014a). Seuran tietojärjestelmä on luotu hälyttämään, mikäli pelaajan data on odotetun vaihteluvälin ulkopuolella. Näitä datamassoja seura pyrkii analysoimaan ja käyttämään päätöksenteossaan, muun muassa ennustamalla tulevien loukkaantumisten todennäköisyyksiä. AC Milanin otettua MilanLab käyttöönsä, seura koki 90 % vähentymisen loukkaantumisten määrissä.

Myös muissa lajeissa tämän kaltaiset tieteelliset videodatan keräämiseen, analysointiin ja seuraamiseen perustuvat toimintamenetelmät ovat alkaneet muuttua entistä yleisemmäksi datan hallinnan teknologioiden kehittymisen myötä (Davenport, 2014a). Davenport (2014a) huomauttaa, että videodatan määrän ja laadun kasvamisen myötä urheiluanalytiikoille esiintyy entistä enemmän uusia osaamismahdollisuuksia liittyen tämän videodatan analysoimiseen.

4.2.2 Biometrinen data

Kansainvälinen jalkapalloliitto FIFA päätti hiljattain sallia pelaajien fysiologisen seurannan langattomien sensortechnologioiden avulla kilpailuissaan. Tämän johdosta tulevaisuudessa tullaan saavuttamaan entistä enemmän yksityiskohtaista dataa pelaajien fysiologisesta suoriutumuksesta otteluiden sisällä. Monet urheilujoukkueet ovat harjoituksissaan jo pitkään käyttäneet biometristä dataa

kerääviä sensoreita. Tämän datan avulla seurat valmentajat pystyvät määrittämään harjoittelun intensiteettiä ja näin ollen mukauttamaan harjoitteluprosesseja kohti haluttuja fyysisiä harjoitusintensiteettejä (Ehrmann ym. 2016; Rein & Memmert 2016; Carling ym. 2008).

Biometristä dataa mittaavien laitteiden avulla voidaan Davenportin (2014a) mukaan tehdä johtopäätöksiä kuten ”Pelaajien, joiden sykkeet ylittävät ottelutilanteessa 160 lyöntiä minuutissa ja pysyvät sen yläpuolella vähintään kahden perättäisen minuutin ajan, hidastuvat liikkeissään merkittävästi otteluiden viimeisellä viidenneksellä”. Tällaisten johtopäätösten luominen biometrisen sensoridatan avulla, auttaa valmentajia ratkaisemaan ongelmia, joiden syitä ei ilman biometrisen sensoridatan laaja-alaista analysointia olisi pystytty edes havaitsemaan.

Monet urheiluseurat ovat alkaneet myös käyttää pelaajillaan päälle puettavia sensoreita, jotka mittaavat pelaajien biomekaniikkaa. Näitä biometrisillä sensoreilla mitattavia asioita ovat muun muassa pelaajien sydämen toiminta, nukkumisen laatu ja verenpaine. Tämän kaltaiset mittaristot voivat parhaimmillaan auttaa seuran työntekijöitä paljastamaan malleja, jotka saattavat johtaa loukkautumisiin (Davenport, 2014a).

Davenport (2014a) uskoo myös, että lajeissa joissa pelaajien päihin kohdistuu säännöllisiä iskuja, tullaan pian näkemään sensoreita, jotka mittaavat näiden iskujen voimia. Kuten sijaintidatan sekä videodatan, myös biometrisen datan käyttökeinoja on vasta hiljattain alettu tunnustaa. Tällä hetkellä biometristä dataa käytetään pääosin pelaajien fyysisen aktiivisuuden mittaamiseksi, mutta on mahdollista, että tulevaisuudessa biometrisen sensoridatan avulla voidaan ymmärtää paremmin myös pelaajien välistä kentällä tapahtuvaa fyysistä vuorovai-
kutusta (Davenport, 2014a).

4.2.3 Sijaintidata

Urheiluseuroilla, jotka ovat ottaneet käyttöönsä sijaintidataa kerääviä laitteita, on kiistattomasti enemmän dataa, kuin mitä seurat ovat perinteisesti osanneet hyödyntää (Davenport, 2014a). Seuroilla on paljon mahdollisuuksia sijaintidatan keruussa, sillä markkinoilla on olemassa useilla eri seurantasysteemeillä toimivia sijaintidataa tuottavia laitteita. Näitä laitteita ovat esimerkiksi GPS-laitteet sekä radioaaltoja hyödyntävät sensoriteknologiat (Leser, Baca & Ogris, 2011). Laitteiden avulla voidaan mitata kentällä tapahtuvaa liikettä ja nopeuksia (Davenport, 2014a). Tutkimusten perusteella radioaaltoja hyödyntävät sensorit ovat GPS-laitteita tarkempia sijaintidatan tarjoamisessa eivätkä ne myöskään tarvitse suoraa näköyhteyttä mitattavaan kohteeseen kuten esimerkiksi videodatan avulla sijainnin seuraaminen tarvitsee (Duffield ym. 2010; Halvorsen ym. 2013).

Näin ollen uusien sijaintidataa keräävien teknologioiden avulla pelaajien suoritusanalyysiin avautuu jatkuvasti entistä parempia mahdollisuuksia (Davenport, 2014a). Esimerkiksi mittaamalla korkean intensiteetin juoksua ja

niiden määrää urheiluseura voi saada arvokasta dataa, sillä on tutkittu, että jalkapallossa kansainvälisellä tasolla pelaajat suorittavat enemmän korkean intensiteetin juoksuja kuin kansallisilla tasoilla (Anderson ym. 2010). Yleisesti sijaintidataa hyödyntävillä teknologioilla kerätään dataa muun muassa pelaajien liikutuista määristä, korkean intensiteetin juoksuista, kiihdytyksistä, jarrutuksista sekä pelaajien välisien kontaktien määristä (Davenport, 2014a). Vaikka sijaintidataa kerääviä laitteita käytetään useimmiten mittaamaan pelaajien aktiivisuutta kentällä, voidaan kuitenkin myös pelivälineen sijaintidataa kerätä ja analysoida. Keräämällä dataa pelivälineiden liikkeistä tarpeeksi pitkältä ajalta voidaan esimerkiksi mitata eri joukkueiden pelaamisen ominaispiirteitä. Näiden tietojen perusteella valmentajat voivat luoda pelisuunnitelmansa faktatietoon perustuen (Davenport, 2014a).

Sijaintidataa voidaan hyödyntää myös urheiluliiketoiminnassa. Aikaisemmin tutkielmassa kerrottiin Phoenix Sunsin yhteistyöstä teleoperaattori Verizon Wirelessin kanssa. Yhteistyökumppaninsa avulla Phoenix Suns pystyy seuraamaan otteluissaan käyviensä ihmisten sijaintidataa, jota analysoimalla seura pystyy tekemään liiketoimintansa kannalta arvokkaita johtopäätöksiä. Verizon pystyi sijaintidatansa avulla kertomaan Sunseille esimerkiksi, että peleihin osallistuvat katsojat ovat todennäköisimmin 25 ja 54 ikävuoden väliltä, tulevat yli 50 000 dollarin ansioiden kotitalouksista ja ovat yleisesti vielä kotona asuvien lasten vanhempia. Muita Sunseille selvinneitä asioita oli, että noin 22 % katsojista tulivat kaupungin ulkopuolelta ja Sunsin mainostamissa pikaruokaketjuissa oli pelien jälkeen n. 8 % enemmän ihmisiä kuin ilman mainostamista (Davenport, 2014a).

Tällainen liiketoimintainformaatio on luonnollisesti hyvin arvokasta urheiluorganisaatiolle e ja sen avulla voidaan tehostaa lukuisia eri liiketoimintaprosesseja. Isoja sijaintidatamassoja hyödyntämällä seurat voivatkin siis tehostaa monia eri urheiluanalytiikkansa osa-alueita.

5 Johtopäätökset

Tutkielmassa käsiteltiin tieteellisen kirjallisuuden avulla massadatan hyödyntämistä urheiluanalytiikassa. Tutkimuskysymykseen vastaamista varten määriteltiin ensin kaksi apukysymystä ”Mitä on massadata?” ja ”Mitä on urheiluanalytiikka?”, joiden avulla pyrittiin määrittelemään massadata urheiluanalytiikan kontekstissa. Urheiluanalytiikan massadatan määrittelemisen jälkeen pyrittiin tieteellisen kirjallisuuden avulla vastaamaan tutkimuskysymykseen ”Miten massadataa voidaan hyödyntää urheiluanalytiikassa?”.

Tuloksena ensimmäiseen apukysymykseen, ”Mitä on massadata?”, löydettiin massadataa olevan datamassat, joiden keräämiseen ja analysointiin eivät toimialan perinteiset datanhallintamenetelmät riitä. Toiseen apukysymykseen, ”Mitä on urheiluanalytiikka?”, löydettiin urheiluanalytiikan olevan rakenteellisen ja historiallisen datan hallitsemista, tulevaisuutta ennakoivien analyttisten mallien soveltamista tähän dataan sekä tietojärjestelmien käyttöä informoimaan urheiluseuran päätöksentekijöitä. Määriteltäessä massadataa urheiluanalytiikan kontekstissa, päädyttiin lähdemateriaalien pohjalta luokittelemaan ammattiurheilun toimialalla massadataksi videodata, biometrinen data sekä sijaintidata. Tämän jälkeen tutkittiin, miten näillä datan eri muodoilla voidaan tuottaa arvoa urheiluanalytiikassa.

Massadataksi luokiteltavan videodatan hyödyntämiskeinoja urheiluanalytiikassa löydettiin olevan automatisoitu tilastoiden tuottaminen sekä niiden avulla syvällisemmän analyysin tekeminen siitä, missä, miten ja miksi kentällä toteutuneet tapahtumat toteutuivat. Massadataksi luokiteltavaa videodataa voidaan urheiluanalytiikassa hyödyntää myös luomalla automatisoidusti leikkeitä, joita pelaajat kykenevät itse analysoimaan. Tämän lisäksi tutkimuksessa löydettiin videodataa kyettävän hyödyntämään urheiluanalytiikassa myös pelaajien fyysisen kuormituksen seurantaan (taulukko 4).

TAULUKKO 4 Video-massadatan hyödyntäminen urheiluanalytiikassa

	Hyödyntämisprosessi	Saatu hyöty
Otteluiden ja pelaajien suoritusten analytiikka	Automatisoitu tilastojen tuottaminen sekä syvälylysi, missä, miten ja miksi kentällä toteutuneet tapahtumat tapahtuivat	Menneiden otteluiden analysointi sekä tuleviin otteluihin valmistautuminen
	Automatisoinnin hyödyntäminen videoleikkeiden keräämisessä ja editoimisessa	Pelaajien itsenäinen omien suoritusten analysointi
Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka	Pelaajien fyysisen kuormituksen sekä liikemallien analysointi ja seuranta	Loukkaantumisten ehkäiseminen

Massadataksi luokiteltavaa biometristä dataa todettiin tutkielmassa hyödynnettävän otteluiden ja pelaajien suoritusten analytiikassa sekä pelaajien terveyden edistämisen ja loukkaantumisten ehkäisyn analytiikassa. Pelaajien fyysisestä suoriutumisesta kerättävän biometrisen datan avulla valmentajat voivat optimoida harjoittelunsa fyysisestä kuormituksesta. Toinen tutkimuksessa löydetty biometrisen massadatan hyödyntämiskeino oli erilaisten päälle puettavien biometristen sensoreiden hyödyntäminen. Näiden päälle puettavien biometristen sensoreiden avulla kyetään tunnistamaan loukkaantumisiin johtavia malleja ja näin ennaltaehkäisemään loukkaantumisia (taulukko 5).

TAULUKKO 5 Biometrisen massadatan hyödyntäminen urheiluanalytiikassa

	Hyödyntämisprosessi	Saatu hyöty
Otteluiden ja pelaajien suoritusten analytiikka	Fysiologisen datan kerääminen pelaajien fyysisestä suoriutumisesta	Harjoittelun fyysisen kuormittavuuden optimointi
Pelaajien terveyden edistämisen ja loukkaantumisten ehkäisemisen analytiikka	Päälle puettavien biometristen sensoreiden hyödyntäminen	Loukkaantumisten ehkäiseminen tunnistamalla loukkaantumisiin johtavia malleja

Massadataksi luokiteltavan sijaintidatan hyödyntämiskeinoja tutkimuksessa havaittiin olevan pelaajien nopeuksien mittaaminen ottelusuoritusten optimoinniksi sekä pelivälineen seuranta ja analysointi auttamaan päätöksentekijöitä ottelutaktiikoiden luomisessa havaitsemalla joukkueiden pelaamisen ominaispiirteitä (taulukko 6).

TAULUKKO 6 Sijainti-massadatan hyödyntäminen urheiluanalytiikassa

	Hyödyntämisprosessi	Saatu hyöty
Otteluiden ja pelaajien suoritusten analytiikka	Pelaajien nopeuksien mitaaminen	Ottelusuoritusten optimointi
	Pelivälineen seuranta ja analysointi	Ottelutaktiikoiden luominen havaitsemalla oman joukkueen sekä vastustajan pelaamisen ominaispiirteitä
Urheiluliiketoiminnan analytiikka	Otteluissa käyvien ihmisten sijaintitietojen kerääminen ja analysointi	Liiketoimintamallien arvioiminen ja markkinointiviestinnän optimointi

Tutkimuskysymykseen vastattaessa tutkielmassa päädyttiin lopputulokseen, että massadataa voidaan hyödyntää urheiluanalytiikassa useimmilla eri tavoilla, jotka liittyvät pelaajien ja joukkueiden suorituksien parantamiseen, terveyden edistämiseen sekä urheiluliiketoimintaan. Kaikki löydetyt massadatan hyödyntämiskeinot urheiluanalytiikassa esiteltiin taulukoiden ne datan tyyppin sekä urheiluanalytiikan osa-alueen perusteella (taulukko 7).

TAULUKKO 7 Massadatan hyödyntämiskeinot urheiluanalytiikassa

	Otteluiden ja pelaajien suoritusten analytiikka		Terveystiedon edistämisen ja loukkaantumisten ehkäisemisen analytiikka	Urheiluliiketoiminnan analytiikka
Videodata	Automaattisten tilastoiden tuottaminen sekä syvällisempi analyysi kenttätapahtumista	Automatisoinnin hyödyntäminen videoleikkeiden luomisessa	Pelaajien fyysisen kuormituksen sekä liikemallien analysointi ja seuranta	
Sijaintidata	Pelaajien nopeuksien mitaaminen	Pelivälineen seuranta ja analysointi		Otteluissa käyvien ihmisten sijaintitietojen kerääminen ja analysointi
Biometrinen data	Fysiologisen datan kerääminen pelaajien fyysisestä suoritusmuutuksesta		Päälle puettavien biometrinen sensoreiden hyödyntäminen	

Löydettyjen lopputulosten perusteella voidaan tehdä johtopäätös, että massadataa hyödynnetään kaikista eniten otteluiden ja pelaajien suoritusten analytiikan sekä terveyden edistämisen ja loukkaantumisten ehkäisyn osa-alueilla. Vähiten massadataa tutkimuksen mukaan hyödynnettiin urheiluliiketoiminnan analytiikassa. Tutkimuksessa selvisi myös massadatan mahdollistavan syvällisempiä urheiluanalytiikkaan liittyviä ratkaisuita, kuin mitä tavallinen data mahdollistaa.

Koska massadata-ilmio on urheiluanalytiikassa kuitenkin verrattain uusi, ei laaja-alaisia selvityksiä sen hyödyntämisestä vielä ole tieteellisessä kirjallisuudessa. Tutkielman aihepiirin tieteellisen kirjallisuuden perusteella voisi tehdä johtopäätöksen, että vaikka massadata on joillain toimialoilla jo hyvinkin tuttu käsite, on sen tietoinen hyödyntäminen urheiluanalytiikassa vielä alkutekijöissään. Toinen tekijä, mikä mitä todennäköisimmin vaikuttaa aiheen tieteellisten tutkimusten vähäisyyteen, on urheiluorganisaatioiden haluttomuus luopua kilpailuedustaan tuomalla omia urheiluanalytiikan käytänteitään julki.

Tutustuessa aiheen tieteelliseen kirjallisuuteen huomioitiin, että urheiluanalytiikan tieteelliset tutkimukset olivat pääasiallisesti urheiluorganisaation näkökulmasta tutkittuja. Mahdollinen jatkotutkimusaihe tutkielmaan voisikin olla massadatan hyödyntäminen urheiluanalytiikassa keskittyen yksilöurheilijan näkökulmaan. Muita mahdollisia jatkotutkimusaiheita voisi olla syvällisempi tutustuminen urheiluorganisaatioiden massadatan hyödyntämiseen sekä sen hyödyntämisen vaikutukseen urheiluorganisaation kilpailukykyyn.

LÄHTEET

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17(2)

Alamar, B. (2013). *Sports analytics: A guide for coaches, managers, and other decision makers*. Columbia University Press, s. 4

Andersson, H. Å., Randers, M. B., Heiner-Møller, A., Krstrup, P., & Mohr, M. (2010). Elite female soccer players perform more high-intensity running when playing in international games compared with domestic league games. *The Journal of Strength & Conditioning Research*, 24(4), 912-919.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly*, 56(1), 75-86.

Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). The role of motion analysis in elite soccer. *Sports medicine*, 38(10), 839-862

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4), 1166.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2014, February). POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data. In 8th Annual MIT sloan sports analytics conference, February (Vol. 28).

Cole, D., Nelson, J., & McDaniel, B. (2015). Benefits and Risks of Big Data. SAIS 2015 Proceedings.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.

Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43.

- Davenport, T. H. (2014a). Analytics in sports: The new science of winning. *International Institute for Analytics*, 2, 1-28.
- D’Orazio, T., & Leo, M. (2010). A review of vision-based systems for soccer video analysis. *Pattern recognition*, 43(8), 2911-2926.
- Duffield, R., Reid, M., Baker, J., & Spratford, W. (2010). Accuracy and reliability of GPS devices for measurement of movement patterns in confined spaces for court-based sports. *Journal of Science and Medicine in Sport*, 13(5), 523-525.
- Ehrmann, F. E., Duncan, C. S., Sindhusake, D., Franzsen, W. N., & Greene, D. A. (2016). GPS and injury prevention in professional soccer. *The Journal of Strength & Conditioning Research*, 30(2), 360-367.
- Farin, D., Han, J., & de With, P. H. (2005, July). Fast camera calibration for the analysis of sport sequences. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on* (pp. 4-pp). IEEE
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- Ganeshapillai, G., & Guttag, J. (2012, March). Predicting the next pitch. In *Sloan Sports Analytics Conference*.
- Ghasemaghaei, M. (2017). The Effects of Operational and Cognitive Compatibilities on the Big Data Analytics Usage: Firm Distinctive Value Creation
- Goldman, M., & Rao, J. M. (2011, March). Allocative and dynamic efficiency in NBA decision making. In *In Proceedings of the MIT Sloan Sports Analytics Conference* (pp. 4-5).
- Gruetzemacher, R., Gupta, A., & Wilkerson, G. B. (2016). Sports Injury Prevention Screen (SIPS): Design and Architecture of an Internet of Things (IoT) Based Analytics Health App. In *CONF-IRM* (p. 1
- Halvorsen, P., Sægrov, S., Mortensen, A., Kristensen, D. K., Eichhorn, A., Stenhaug, M., ... & Johansen, D. (2013, February). Bagadus: an integrated system for arena sports analytics: a soccer case study. In *Proceedings of the 4th ACM Multimedia Systems Conference* (pp. 48-59). ACM.
- Hanisch, B., & Hanisch, J. (2007). Injury management: The development and implementation of innovative software in an elite sporting club. *PACIS 2007 Proceedings*, 8.

- Harrison, C. K., & Bukstein, S. (Eds.). (2016). *Sport Business Analytics: Using Data to Increase Revenue and Improve Operational Efficiency*. CRC Press
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Rama-krishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 995-1004). IEEE.
- Kamioka, T., & Tapanainen, T. (2014). Organizational Use of Big Data and Competitive Advantage-Exploration of Antecedents. In *PACIS* (p. 372)
- Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *Contemporary Computing (IC3), 2013 Sixth International Conference on* (pp. 404-409). IEEE.
- Kearney, A. T. (2011). *The sports market. Major trends and challenges in an industry full of passion.*
- Kim, H. (2015). *Big Data: The Structure and Value of Big Data Analytics*
- Kim, H. C., Kwon, O., & Li, K. J. (2011, November). Spatial and spatiotemporal analysis of soccer. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 385-388). ACM.
- Lee, C. (2016). *A content analysis of sports analytics resistance.*
- Leser, R., Baca, A., & Ogris, G. (2011). Local positioning systems in (game) sports. *Sensors*, 11(10), 9778-9797.
- Macdonald, B. (2012, March). An expected goals model for evaluating NHL teams and players. In *Proceedings of the 2012 MIT Sloan Sports Analytics Conference*, <http://www.sloansportsconference.com>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity.*
- McAfee, A., Brynjolfsson, E., & Davenport, T. H. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- Mikalef, P., Pappas, I. O., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2016). Big Data and Strategy: A research Framework. In *MCIS* (p. 50)

Mondello, M., & Kamke, C. (2014). The Introduction and Application of Sports Analytics in Professional Sport Organizations. *Journal of Applied Sport Management*, 6(2).

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature re-view of information systems research.

Pileggi, H., Stolper, C. D., Boyle, J. M., & Stasko, J. T. (2012). Snapshot: Visualization to propel ice hockey analytics. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2819-2828.

Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big Data Research*, 2(4), 166-186.

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410.

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, 19, 40, s. 6-7.

Sacha, D., Stein, M., Schreck, T., Keim, D. A., & Deussen, O. (2014, October). Feature-driven visual analytics of soccer data. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (pp. 13-22). IEEE.

Shah, F. A., Kretzer, M., & Mädche, A. (2015). Designing an Analytics Platform for Professional Sports Teams.

Shim, J. P., French, A. M., Guo, C., & Jablonski, J. (2015). Big Data and Analytics: Issues, Solutions, and ROI. *CAIS*, 37, 39

Tan, F., Hedman, J., & Xiao, X. (2017). Beyond 'Moneyball' to Analytics Leadership in Sports: An Ecological Analysis of FC Bayern Munich's Digital Transformation, s. 1

Troilo, M., Bouchet, A., Urban, T. L., & Sutton, W. A. (2016). Perception, reality, and the adoption of business analytics: Evidence from North American professional sport organizations. *Omega*, 59, 72-83.

Wilkerson, G., & Gupta, A. (2016). Sports Injuries and Prevention Analytics: Conceptual Framework & Research Opportunities.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), 97-107.

Xiao, X., Hedman, J., Tan, F. T. C., Tan, C. W., Lim, E. T., Clemenson, T., ... & van Hilleberg, J. (2017). Sports Digitalization: An Overview and A Research Agenda. In ICIS 2017. Association for Information Systems. AIS Electronic Library (AISeL).

Xing, J., Ai, H., Liu, L., & Lao, S. (2011). Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling. *IEEE Transactions on Image Processing*, 20(6), 1652-1667.

Zhu, G., Xu, C., Huang, Q., Rui, Y., Jiang, S., Gao, W., & Yao, H. (2009). Event tactic analysis based on broadcast sports video. *IEEE Transactions on Multimedia*, 11(1), 49-67.