

Mattias Kitola

**URHEILUANALYTIIKAN ARVO
JOUKKUEURHEILUSSA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2020

TIIVISTELMÄ

Kitola, Mattias

Urheiluanalytiikan arvo joukkueurheilussa

Jyväskylä: Jyväskylän yliopisto, 2020, 38 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja(t): Kyppö, Jorma

Tämän kandidaatin tutkielman tarkoituksena on kirjallisuuskatsauksen avulla selvittää, mistä urheiluanalytiikasta syntyvä arvo muodostuu. Kirjallisuuskatsauksen ohella pohdintaluvussa tulen peilaamaan omiin kokemuksiini urheiluanalytiikan molemmilta puolilta. Urheiluanalytiikka käsittää historiallisen datan hallinnan, ennakoivien analyttisten mallien käytön ja tietojärjestelmien käytön päätöksenteossa. Urheiluanalytiikka on kilpailullisen edun saavuttamisen väline. Arvo muodostuu ajan säästämisestä ja uniikin informaation tuottamisesta. Data itse ei puhu ja päätöksentekijöiden merkitys urheiluanalytiikassa on merkittävä. Urheiluanalytiikan arvo voidaan nähdä pakottavana voimana saada oikeat henkilöt kysymään tarkempia kysymyksiä. Teknologiset haasteet ja kulttuuriset haasteet kuuluvat urheiluanalytiikan haasteisiin. Vanhan koulukunnan valmentajien asenteet voivat hidastaa urheiluanalytiikan evoluutiota. Tulee huomata, että urheiluanalytiikka ei ole liike vanhan koulukunnan valmentajia vastaan, vaan heiltä saatu palaute on yksi tärkeimpiä keinoja analyttisten metodien kehittämiseksi.

Asiasanat: urheiluanalytiikka, data-analyysi, data-analytiikka, moneyball

ABSTRACT

Kitola, Mattias

The value of sports analytics in team sports

Jyväskylä: University of Jyväskylä, 2020, 38 pp.

Information Systems, Bachelor's Thesis

Supervisor(s): Kyppö, Jorma

This bachelor's thesis will be conducted as a literature review. The purpose of this bachelor's thesis is to find out where is the value in sports analytics. In discussion section I will be reflecting my own experience in the field of sports analytics. Sports analytics consists of managing historical data, using predictive models, and using information systems in decision-making. Sports analytics is a way to gain competitive edge. Value consists of saving time and creating unique information. Data does not speak for itself and the importance of decision-makers are significant. You could argue that the value of sports analytics is to get the right people to ask more accurate questions. Sports analytics faces challenges from both technological and cultural standpoints. Attitudes from old-school-coaches can slow down the evolution of sports analytics. It is important to note that sports analytics is not against those old-school-coaches but the feedback from them is one of the most important tools to improve analytical methods.

Keywords: sports analytics, data analysis, data analytics, moneyball

TAULUKOT

TAULUKKO 1 Perinteisen data-analyysin ja massadata-analyysin eroista 14

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
TAULUKOT	4
SISÄLLYS.....	5
1 JOHDANTO.....	6
2 DATAN ANALYSOINTI	8
2.1 Data-analytiikan ja data-analyysin käsitteet.....	8
2.2 Massadata	9
2.2.1 Massadata-analyysi.....	10
2.2.2 Massadata-analytiikka.....	11
2.2.3 Data-analyysin ja massadata-analyysin eroista	11
2.3 Tiivistelmä datan analysoinnista	12
3 URHEILUANALYTIikka.....	14
3.1 Joukkueurheilu tuotteena	14
3.2 Urheiluanalytiikan käsite	15
3.2.1 Urheiluanalytiikan arvon muodostuminen	16
3.2.2 Urheiluanalytiikan haasteet.....	17
3.3 Teknologinen kehitys urheiluanalytiikassa	19
3.4 Tiivistelmä urheiluanalytiikasta	21
4 URHEILUANALYTIikka POHJOISAMERIKKALAISEN AMMATTILAISJOUKKUEURHEILUN KONTEKSTISSA	23
4.1 Pohjoisamerikkalaisen joukkueurheilukulttuurin luonne.....	23
4.2 Baseball ja urheiluanalytiikka	25
4.2.1 Moneyball.....	26
4.2.2 WAR.....	27
4.3 Tiivistelmä urheiluanalytiikasta pohjoisamerikkalaisessa ammattilaisjoukkueurheilukulttuurissa.....	28
5 POHDINTAA	30
LÄHTEET	36

1 JOHDANTO

Teknologinen kehitys on johtanut datan määrän kasvuun. Dataa analysoimalla voidaan löytää relevanttia tietoa ja säännönmukaisuuksia optimaalisen päätöksenteon ehdottamiseen (Runkler, 2016). Yritykset käyttävät datan analysointiin rahaa ja tiedonhallintajärjestelmien uskotaan muuttavan enemmän ja enemmän dataa ja informaatiota tietämykseksi ja tietämyksen tehokkaammaksi päätöksenteoksi (Intezari, Pauleen & Taskin, 2016).

Urheilu liiketoimintamuotona ei eroa tässä perinteisen liiketoiminnan muodoista. Urheiluanalytiikka tarkoittaa rakenteellisen historiallisen datan hallintaa, ennakoivien analyttisten sovellutusten käyttöä ja tietojärjestelmien käyttöä kilpailullisen edun saavuttamiseksi organisaatiolle (Alamar & Mehrotra, 2011a). Davenport (2014) jakaa urheiluanalytiikan kolmeen malliin: pelaajien ja joukkueen suoritusta mittaavaan analytiikkaan, organisaation liiketoimintaa mittaavaan analytiikkaan sekä pelaajien terveyttä ja loukkaantumisten ehkäisyä mittaavaan analytiikkaan.

Tässä tutkimuksessa rajataan organisaation liiketoimintaa mittaava analytiikka ja pelaajien terveyttä ja henkilökohtaista harjoittelua mittaava liikunta-tekniologia pois. Tutkimuksessa selvitetään urheiluanalytiikan merkitystä pelaajien ja joukkueen suorituksessa. Alamarin ja Mehrotran (2012) mukaan liian monet joukkueet eivät käytä data-analysejä toiminnassaan. Tähän lähden käänteisesti hakemaan vastausta, ja tutkimuskysymykseksi muodostui:

- Mistä urheiluanalytiikan arvo joukkueurheilussa muodostuu?

Tutkimus toteutettiin kirjallisuuskatsauksena. Tietokantoina pääosin käytettiin Google Scholaria, IEEE Xplore ja SPORTDiscusta. Hakusanat pääasiassa olivat "data analytics", "data analysis", "sports analytics", "moneyball" ja "sabermetrics".

Tutkimuksen päällimmäisenä tavoitteena on luoda pohjaa samasta aiheesta jatkettavalle pro gradu -tutkimukselle. Sekundäärinen tavoitteena on herättää keskustelua ja ajatuksia oman lajin, pesäpallon, piirissä ja sen urheiluorganisaatioissa mahdollisten uusien urheiluanalyttisten ideoiden oivaltamiseksi.

Tutkimus koostuu kolmesta käsittelyluvusta ja pohdintaosuudesta. Ensimmäisessä käsittelyluvussa käsitellään data analyysin, data-analytiikan ja massadatan käsitteitä, sekä data-analyysiprosessin vaiheita perustan luomiseksi. Toisessa käsittelyluvussa tutkitaan urheiluanalytiikka käsitteenä, sen arvoa ja haasteita sekä urheiluanalytiikan teknologista kehitystä. Viimeisessä käsittelyluvussa peilataan pohjoisamerikkalaista joukkueurheilukulttuuria, etenkin baseballia, tavoitteena löytää eroja ja yhteneväisyyksiä pesäpallon kanssa. Kirjallisuuskatsauksen luonteen takia vasta pohdintaluvussa peilataan aiemmasta tutkimuskirjallisuudesta löytynyttä tietoa pesäpallon kontekstiin. Aiemman tutkimuskirjallisuuden lisäksi tuon omia näkemyksiä urheiluanalytiikan molemmilta puolilta: yksityisen yrityksen, Titania pesisportaalin, urheiluanalytikonä datan tuottajana sekä Jyväskylän Kirittärien kaudella 2020 suomenmestaruuden voittaneen joukkueen urheiluanalytikkona.

2 DATAN ANALYSOINTI

Datan määrä kasvaa maailmassa valtavalla nopeudella. Runklerin (2016) mukaan tiedon määrä maailmassa tuplaantuu 20 kuukauden välein. Analysoimalla dataa voidaan löytää relevanttia tietoa ja säännönmukaisuuksia, jotka mahdollistavat uusien näkemysten luomisen, syiden ja seurauksien tunnistamisen sekä tulevaisuuden ennakoimisen ja optimaalisen päätöksenteon ehdottamisen (Runkler, 2016).

Informaation hierarkiamallissa kuvataan datan, informaation, tietämyksen ja viisauden suhdetta. Pyramidimallissa ihmisen merkitys kasvaa ylätasolla ja tietotekniikan merkitys alatasolla. Ylimpänä olevaan viisauden käsitteeseen liitetään inhimilliset ominaisuudet, kuten päätöksentekokyky, kokemus, arvot ja uskomukset. Ihmisen avaintekijöitä tietämyksen siirtämisessä päätöksentekoon on tietämys itsestä ja ympäristöstä, sekä kyky nähdä tilannetta monesta eri perspektiivistä. Päätöksenteko on keskeisiä arvoja, joita yritysjohtajalta tarvitaan organisaation johtamiseen. (Intezari ym., 2016).

Pohjaten datan merkitykseen liiketoiminnan päätöksenteossa, lähdin tutkimaan datan prosessointia tarkemmin. Datan käsitteitä etsiessä törmään käsitteisiin data-analytiikka ja data-analyysi samoissa asiayhteyksissä ja heräsi kysymys, mitä nämä käsitteet käytännössä tarkoittavat ja eroavatko ne toisistaan millään tavalla.

2.1 Data-analytiikan ja data-analyysin käsitteet

Aiemmasta tutkimuskirjallisuudesta huomaa, että data-analytiikkaa (*eng. data analytics*) ja data-analyysiä (*eng. data analysis*) käsitteinä käytetään samoissa asiayhteyksissä. Kuitenkin aiempaa tutkimuskirjallisuutta selvittäessä on huomattavissa käsitteiden välillä tiettyjä eroavaisuuksia. Sanakirjamäärittystä tutkiessa Cambridge Dictionaryn mukaan analyysi (*eng. analysis*) on organisoitu prosessi jonkin asian oppimiseksi, kun taas analytiikka (*eng. analytics*) samaisen

sivuston mukaan määritellään säännönmukaisuuksien löytämiseksi tietokoneella matemaattisia metodeja hyödyntäen (Cambridge Dictionary, 2020).

Data-analyysi on prosessi, jossa data valitaan, siistitään, suodatetaan, visualisoidaan ja analysoidaan, minkä jälkeen tuloksia tulkitaan ja arvioidaan (Runkler, 2016). Caon (2017) mukaan data-analyysissä hyödyllisen informaation löytämiseksi dataa prosessoidaan perinteisillä metodeilla, joita ovat tilastolliset mallit, matemaattiset mallit ja loogiset mallit. Myös Cheng, Jiang, Wang ja Tang (2016) tekevät jaottelun perinteiseen data-analyysiin. Perinteinen data-analyysi perustuu syy-seuraus-suhteeseen (Cheng ym., 2016) ja sen keskeinen tarkoitus on prosessin optimointi ja kilpailullisen aseman parantaminen (Runkler, 2016).

Runkler (2016) jakaa data-analyysiprojektin neljään vaiheeseen: valmisteluun, esiprosessointiin, analyysiin ja jälkiprosessointiin. Valmisteluvaiheessa luodaan prosessin suunnitelmat, kerätään ja valitaan data sekä luodaan ominaisuuksia. Esiprosessoinnissa dataa suodatetaan, siivotaan ja korjataan (Runkler, 2016). 80 prosenttia data-analyysiprosessista koostuu esiprosessointivaiheeseen, ja siinä tehtävään datan siivoamiseen ja valmisteluun (Donoho, 2015).

Runklerin (2016) data-analyysiprojektin kolmannessa vaiheessa dataa analysoidaan dataan soveltuvilla menetelmillä, kuten visualisointi hahmottamiskeinona, korrelaatiot kahden asian keskinäisten suhteiden tunnistamiseen, luokittelu, klusterointi ja ennakointi. Jälkiprosessoinnissa dataa tulkitaan, dokumentoidaan ja arvioidaan. Johdonmukaisuudet ovat merkittävässä roolissa datan tulkinnassa, kun pyritään tekemään luotettavia ennusteita tulevasta datasta. Datan analysointikeinoista etenkin visualisoinnilla on suuri merkitys, sillä ihmismieli ja silmä ovat tehokkaita työkaluja datan analysoinnissa. (Runkler, 2016).

Data-analytiikka käsittää perinteisen data-analyysin määritelmän, johon lisätään tietoteknisiäsovellutuksia, esimerkiksi tiedon louhinnan (eng. data mining) ja koneoppimisen (eng. machine learning) muodoissa (Cao, 2017). Tiedon louhinnassa piilossa olevaa tietoa etsitään datasta (Cao, 2017; Runkler, 2016). Myös Runkler (2016) tuo esille tietokonesovelluksien läsnäolon data-analytiikan käsitteessä, kuten myös toi Cambridge Dictionary:n sanakirjamääritelmä. Itse data-analytiikkaan Cao (2017) yhdistää kuvailevan analyysin, ennakoivan analyysin ja ohjaavan analyysin päätöksenteon tueksi. Näiden lisäksi data-analytiikan muotoihin kuuluvat syväoppiminen (eng. deep learning), jossa haetaan laajaa ymmärrystä siihen, miksi jotain tapahtuu, sekä kehittynyt analytiikka (eng. advanced analytics), joka määritellään teorioiksi, työkaluiksi, teknologioiksi ja prosesseiksi massadataan pohjautuvan ymmärryksen saamiseksi (Cao, 2017). Näihin ei tavallisen analyysin metodeilla päästä.

2.2 Massadata

Kehittyneet teknologiat lisäävät datan määrää ja nopeutta, millä uutta dataa syntyy. Massadatan (eng. Big data) onnistunut hyödyntäminen näkyy kilpailuedun saamisena ja tehokkuuden paranemisena, mikä näkyy yrityksen tuotta-

vuudessa, eli suorituksessa (Nguyen, 2018). Massadata voi alentaa kustannuksia, sytyttää liiketoiminta-ajatuksia ja paljastaa piilossa olevaa strategista tietoa (Kwon, Lee & Shin, 2014). Massadata nähdään yritykselle voimavarana (Shamsuddin & Hasan, 2015). Massadatalle ei ole kuitenkaan yleistä hyväksyttyä selitystä (Alsaig, Alagar & Ormandjieva, 2018).

Alsaig ym. (2018) mukaan massadatan kuvaaminen V-mallilla auttaa ymmärtämään massadatan luonteen ja he tutkimuksessaan etsivät tutkimuskirjallisuudessa käytettyjä eri V-termejä. Näistä käytetyimpiä massadataa kuvaavia määritelmiä olivat *variety*, *velocity* ja *volume* (Alsaig ym., 2018). *Variety*, monimuotoisuus, kuvaa, kuinka erilaisissa muodoissa dataa syntyy (Nguyen, 2018). Se kuvastaa, kuinka dataa tulee erilaisissa formaateissa, kuten kuvatiedostoina, äänitiedostoina ja tekstinä, sekä dataa tulee erilaisista lähteistä ja toimii erilaisissa tarkoituksissa (Alsaig ym., 2018). *Volume*, massa, kuvastaa jatkuvasti datan lisääntyvää kokonaismäärää (Nguyen, 2018). *Velocity*, nopeus, kuvastaa datan nopeutta ja dynaamisuutta (Nguyen, 2018; Alsaig ym., 2018). Nopeuden käsitteeseen kuuluu datan syntymisen nopeuden lisäksi nopeus, jolla dataa kerätään ja nopeus, jolla dataa prosessoidaan (Alsaig ym., 2018). Kirjallisuustutkimuksessaan Alsaig ym. (2018) listaavat näiden edellä mainittujen kolmen määritteen lisäksi käsitteet *value* ja *veracity* seuraavaksi käytetyimmiksi määritteiksi. *Value*, arvo, kuvastaa arvoa, mikä massadatasta voi syntyä yksilölle tai organisaatiolle ratkaisujen tekovaiheessa (Nguyen, 2018; Alsaig ym., 2018). *Veracity*, todenmukaisuus, -käsitteellä puhutaan datan totuudenmukaisuudesta perustuen siihen, että paljon dataa muodostuu nopeasti ja useassa eri formaatissa, jolloin sen laatu ja tarkkuus eivät kaikissa tilanteissa ole korkealla tasolla (Nguyen, 2018; Alsaig ym., 2018).

2.2.1 Massadata-analyysi

Massadatan aikakautena data-analyysi on kokenut muutoksia (Chen ym., 2016). Massadata ylittää perinteiset datan varastointi-, louhinta- ja analysointimenetelmät ja piilossa olevan tiedon löytämiseen tarvitaan uusia tekniikoita, arkkitehtuuria, algoritmeja ja analytiikkaa (Shamsuddin & Hassan, 2015). Massadatan kanssa tiedon löytäminen, tiedon louhinta ja yhteyksien löytäminen on data-analyysin keskeinen tarve (Chen ym., 2016).

Chen ym. (2016) kutsuvat massadatan aikakaudella tapahtuvaa data-analyysia massadata-analyysiksi. Massadata-analyysi nähdään lajitteluna, organisointina, varastointina, prosessointina, analysointina ja tulkintana (Chen ym., 2016). Prosessivaiheeseen edellä mainittu kuvaus ei siis eroa Runklerin (2016) perinteisen data-analyysin projektivaiheista. Chen ym. (2016) kuvaavat perinteisen datan analyysiä myös termillä ”pienen datan analyysi”. Massadata-analyysin ja ”pienen datan analyysin” välillä on havaittavissa eroavaisuuksia.

Lähtökohtaisesti Chen ym. (2016) kuvaavat, että perinteisen data-analyysin pienen datamäärän matemaattisten mallien on ollut vaikea sulautua suurempiin otantoihin ja datamääriin. Data on luonteeltaan erilaista, joten data-analyysissä on pystyttävä kehittämään datan keräämisen, datan prosessoinnin,

datan varastoinnin ja datan analysoinnin toimia tiedon ja arvokkaan datan löytämiseksi. Massadata-analyysiin ei oteta satunnaista otantaa, vaan otantana on kaikki saatavilla oleva data, kokonaisuus, ilmiön seuraussuhteiden löytämiseksi. Asioiden välisiin korrelaatioihin keskitytään, eikä niinkään niiden syihin. Tärkeää on tietää mitä tapahtuu, ei välttämättä tietää miksi jotain tapahtuu. Kolmantena erottavana tekijänä on tehokkuuden merkitys tarkkuuden saralla, sillä aika ja kustannus ovat tärkeämpiä arvoja kuin tarkkuus. Sama ilmiö mainittiin edellä veracity-termin ohessa. Vaikka analyysin tulos ei välttämättä ole tarkka, voi se tarjota enemmän vaihtoehtoja ja ratkaisuja. (Chen ym., 2016).

2.2.2 Massadata-analytiikka

Kun edellä on todettu massadatan luonteen vaativan tietoteknisiä sovellutuksia, aiemmin käytetyn Cambridge Dictionary:n määritelmä analyysin ja analytiikan välillä massadatan suhteen on haastava, sillä molemmat prosessit, analyysi ja analytiikka, vaativat tietoteknisiä sovelluksia ja matemaattisia malleja. Kwon ym. (2014) määrittelevät massadata-analytiikan teknologioiksi ja tekniikoiksi, kuten tiedonlouhinta, tietokantojen louhinta ja analyttiset menetelmät.

Massadata-analytiikka voi paljastaa malleja asiakkaan käyttäytymisestä. Massadata-analytiikalla yritys voi muun muassa monitoroida tuotteiden hyväksyntää markkinoilla ja ymmärtää paremmin oman markkinasegmenttinsä. Tavanomaisiin liiketoiminnan tekniikoihin nähden, massadata-analytiikka yleisesti prosessoi heterogeenistä dataa ulkoisista lähteistä, kuten sosiaalisesta mediasta irti saatavalla datalla, mikä voi paljastaa uusia mahdollisuuksia, strategioita ja kilpailuetuja. Tavanomaiset liiketoiminnan tekniikat käyttävät muodollista ja sisäistä datalähdettä. (Kwon ym., 2014).

2.2.3 Data-analyysin ja massadata-analyysin eroista

Chen ym. (2016) mukaan data-analyysin merkityksenä on datan kerääminen, objektiivinen arviointi, trendien ennakoiminen ja palaute päätöksenteon ohjaamiseen. Data-analyysi perustuu asioiden suhteisiin ja korrelaatioihin (Chen ym., 2016). Kuten edellä on todettu, data-analyysi ja massadata-analyysi eroavat toisistaan esimerkiksi datan koon suhteen, datan luonteen suhteen ja käytettävien menetelmien muodossa.

Chen ym. (2016) kuvaavat perinteisen data-analyysin käsittelevän suhteessa mega- ja gigatavuja, kun massadata-analyysissä liikutaan peta-, exa ja zettatavun tasoilla. ”Pienen datan” -analyysi perustuu rakenteelliseen dataan, kun massadata-analyysi rakenteettomaan, minkä vuoksi tarvitaan uusia datan prosessointi työkaluja. Perinteisen data-analyysin metodeja ovat loogiset menetelmät, systeemianalyysi, sosiologiset menetelmät, tilastolliset menetelmät ja matemaattiset menetelmät. Perinteinen data-analyysi on laadullinen ja määrällinen analyysi otannasta, jossa keskeistä on oikeiden menetelmien valitseminen otantaan nähden. Massadata-analyysissä oikean tiedonlouhinta-algoritmin löytäminen on keskeistä ja työkaluina toimivat tietokannat ja tietokoneet. Analyysissä eroa syn-

tyy myös täsmällisyyden suhteen, kun massadata-analyysi tuottaa reaaliaikais- ta tietoa, kun perinteinen data-analyysi perustuu enemmän menneeseen dataan. (Chen ym., 2016).

Kwon ym. (2014) tutkimuksessaan tekemä analyysin tulos paljastaa, että yritykset, jotka onnistuneesti ovat käyttäneet ulkoista dataa, esimerkiksi tutki- malla asiakkaiden vaatimuksia ja markkinoiden vaatimuksia, ovat taipuvai- sempia käyttämään massadata-analytiikkaa.

2.3 Tiivistelmä datan analysoinnista

Edellisessä luvussa on tarkasteltu data-analyysin, data-analytiikan sekä massa- datan käsitteitä. Sanakirjamääritelmässä data-analyysin ja data-analytiikan pro- sessien eroksi nousi matemaattisten metodien käyttäminen tietokoneen avulla. Jaottelu ei ole kuitenkaan näin yksiselitteinen, kuten Donoho (2015) yhdistää myös data-analyysin käsitteeseen tietotekniset sovellukset ja algoritmit. Jaotte- lun perustuen tietotekniikkaan, tekee se myös mahdolliseksi erottelun massa- data-analyysin ja massadata-analytiikan välille.

Data-analyysin merkityksenä on datan kerääminen, objektiivinen arviointi, trendien ennakoiminen ja palaute päätöksenteon ohjaamiseen (Chen ym., 2016). Datan määrän lisääntyessä teknologisen kehityksen myötä, myös datan analy- sointi on kehittynyt pienestä ja yksinkertaisesta data-analyysistä suuren datan data-analytiikkaan (Cao, 2017). Runklerin (2016) mukaan data-analyysiprosessi jakautuu neljään eri vaiheeseen: valmisteluvaiheeseen, esiprosessointivaihee- seen, analyysivaiheeseen ja jälkiprosessointiin. Massadata-analyysissä on näh- tävissä samoja prosessin vaiheita. Chen ym. (2016) tulkinnan mukaan voidaan puhua tavallisesta data-analyysistä, ”pienen datan” -analyysistä, tai perinteises- tä data-analyysistä verrattuna massadataan liittyvään massadata-analyysiin.

Massadatalle on vaikea luoda yhtä selkeää määritelmää. Massadatan luonnetta auttaa selvittämään yleisesti käytetty V-malli. Näistä käytetyimpiä massadatan määritteitä ovat massa (*eng. volume*), nopeus (*eng. velocity*) ja mo- nimuotoisuus (*eng. variety*) (Alsaig ym., 2018). Data on luonteeltaan erilaista, eli se tarvitsee erilaiset keinot data-analyysiin arvokkaan tiedon löytämiseksi (Chen ym., 2016).

Cao (2017) mukaan data-analytiikka on monitieteellinen ala määrälliseen ja laadulliseen datan tutkimiseen uusien johtopäätösten, näkemysten ja hypo- teesien luomiseksi. Dataan pohjautuva päätöksenteko on arvossaan, ja päätök- sentekijät pyrkivät jatkuvasti tuomaan dataa ja informaatiota jalustalle ongel- mien ratkaisemiseksi (Intezari ym., 2016). Tällä määritelmällä data-analytiikka koostuu monipuolisista data-analyysin metodeista yhdistettynä tulevaisuuteen suuntautuviin johtopäätöksiin. Määritelmässä on myös yhdistetty monitieteelli- syys, minkä myös Runkler (2016) mainitsee, puhuessaan data-analytiikasta mo- nitieteisenä kenttänä, joka on ottanut vaikutteita muun muassa tilastotieteistä.

Datan määrän kasvu, nopeus, heterogeenisuus ja todenmukaisuus tuovat haasteita data-analytikoille. Data usein sisältää virheitä, mitkä aiheuttavat

vääriä tulkintoja data-analyysissä (Runkler, 2016). Tieteilijöiden pitää oppia tunnistamaan dataa useista eri sensoreista, sen tulesa useissa eri formaateissa (Alsaig ym., 2018). Kwon ym. (2014) tutkimus paljastaa, että hyvällä datan laadunhallinnalla on merkitystä datan käytön hyötyihin, olivat lähteet sitten sisäisiä tai ulkoisia.

TAULUKKO 1 Perinteisen data-analyysin ja massadata-analyysin eroista (Chen ym., 2016; Kwon ym., 2014;

	Perinteinen data-analyysi	Massadata-analyysi
Otanta	Satunnainen otanta	Kaikki saatavilla oleva data
Keskittyminen	Syy-seuraus -suhde	Korrelaatiot
Arvot	Tarkkuus	Tehokkuus
Lähteet pääosin	Sisäinen	Ulkoinen
Datan kokoluokka	MB, GB	PB, EB, ZB
Rakenteellisuus	Rakenteellinen	Rakenteeton
Täsmällisyys	menneeseen perustuva	reaaliaikainen
Metodit	loogiset metodit, systeemi-analyysi, sosiologiset metodit, tilastolliset metodit, matemaattiset metodit	Tietokoneet, Machine Learning, Deep Learning, Advanced analytics, Data Mining
Keskeistä	Oikean metodin valitseminen	Oikean tiedonlouhinta-algoritmin valitseminen

3 URHEILUANALYTIikka

Kuten edellisessä luvussa todettiin Intezari ym., (2016) korostavat dataa ja informaatiota päätöksenteossa. Sama trendi on havaittavissa myös urheilussa. Millerin (2015) mukaan, voi jopa tulla aika, jolloin viisaimmat, parhaiten informoidut ja teknologisesti kehittyneimmät joukkueet ovat ne jotka voittavat. Datan ja sitä kautta tehtävän data-analyysin määrä ja mahdollisuudet kasvavat, mutta liian monet joukkueet eivät käytä niitä toiminnassaan (Alamar & Mehrotra, 2012).

Tässä käsittelykappaleessa tarkastellaan ensin joukkueurheilua tuotteena. Tiihosen (2015) urheilun alakäsitteiden määritelmässä puoliammattilaistaso kuuluu kilpaurheilun piiriin, kun ammattilaistaso huippu-urheilun piiriin. Tällä määritelmällä tässä tutkimuskehyksessä käsitellään, niin ultimaattista huippu-urheilua sisältäen pohjoisamerikkalaiset markkinat, kuten puoliammattilaistason joukkueurheilua. Selvimpänä esimerkkinä edelliseen on suomalainen pesäpallo. Pesäpallo näiden määritelmien mukaan kuuluu puoliammattilaistason ja sitä kautta pelkästään kansallisen kilpaurheilun piiriin, mutta tässä tutkimuksessa sisällytän pesäpallon myös osaksi huippu-urheilua, perustuen omaan kokemukseen urheilun ja kilpailun luonteesta. Ensin kuvaan joukkueurheilua tuotteena markkinan ja kilpailutilanteen ymmärtämiseksi, minkä jälkeen käsittelen urheiluanalytiikan käsitettä, sen arvoa ja haasteita, sekä teknologisen kehityksen vaikutusta urheiluanalytiikkaan.

3.1 Joukkueurheilu tuotteena

Ammattiurheilu eroaa osittain muista liiketoiminnan muodoista. Yritysmaailmassa taloustieteilijä voisi perustella maksimaalisen tuoton motiiviksi, mutta urheilumaailmassa omistajan motiivit eivät ole niin yksiselitteisiä (Miller, 2015). Seurajohtaja tai urheiluseuran omistaja voivat etsiä maksimaalista taloudellista tuottoa, maksimaalista voittojen määrää kilpailukaudella ja siitä seuraavaa me-

nestystä, tai sosiaalista pääomaa urheilutuotteen omistamisella (Miller, 2015). Perinteisissä liiketoimintayrityksissä dominoivan markkina-aseman saanti on tavoite, mutta urheilussa dominoivan aseman saanti voi heikentää kiinnostusta tuotetta kohtaan (Miller, 2015). Tämä tarkoittaa sitä, että pitää voittaa riittävän usein, muttei liikaa ja liian pitkällä aikavälillä ettei kiinnostus tuotetta kohtaan laske. Urheilumaailmassa myös asiakkaat, urheilun kuluttajat, ovat yhtä analyyttisiä kuin joukkueet: väitellään analyysimalleista, tilastoista ja eri päätöksenteosta koskien joukkueen toimintaa ja peluutusta (Davenport, 2014).

Urheilumarkkinaan kuuluvat sekä tuotteen myyjät että sen ostajat (Miller, 2015). Myyjiin lasketaan urheilutuotteen artistit, eli pelaajat ja joukkueet. Urheilumarkkinan ostajiin kuuluvat kuluttajat, eli urheilun fanit. Urheilutuote kamppailee viihdetuotteena ihmisten vapaa-ajasta muiden taiteen kenttien kanssa ja lisäksi keskenään kilpailukausien osuessa samalle ajanjaksolle (Miller, 2015). Tähän esimerkiksi Pohjois-Amerikassa haetaan ratkaisuja otteluaikoja siirtämällä, esimerkkinä vuoden 2020 viikolla 46 golfin The 2020 Masters Tournament sai amerikkalaisen jalkapallon pääsarjan NFL:n siirtämään normaalin sunnuntain 1:00 PM ET kierroksen valtaosan otteluista 4:05 PM ET aikatauluun, jotta suurin osa urheilun kuluttajista pystyisivät seuraamaan molempia lähe-tyksiä.

Data-analytiikalla on laajoja käyttöönnoton mahdollisuuksia urheiluorganisaation liiketoimintapuolella (Miller, 2015), mistä esimerkiksi Harrison ja Bukstein (2017) avaavat laajemmin kirjassaan ”*Sport Business Analytics: Using Data to Increase Revenue and Improve Operational Efficiency*”. Näkyvyyden kautta tulevat tulot ovat merkittäviä menestyvän urheiluseuran toiminnalle (Miller, 2015). Millerin (2015) listaukseen pohjaten, näkyvyyden käsitteeseen sisällän mediatu- lot, yhteistyökumppanuudet ja siihen liitettävät sponsorointitulot, mainonnan ja esimerkiksi oikeudet stadionin nimeen.

Huippu-urheilu eroaa tavallisesta yritysmaailmasta sillä, että se on jatku-vasti huomion keskipisteenä, ja vaatii median huomiota elääkseen (Miller, 2015). Joukkueet pyrkivät luomaan brändiä ja kehittämään positiivista mielikuvaa ja mainosta, samalla pyrkien tehdä yksittäisestä asiakkaasta joukkueen kannattaja, fani (Miller, 2015). Myös Davenport (2014) korostaa analytiikan käyttöä urhei- luorganisaation liiketoiminnassa fanien osallistamisella ja fanien lojaaliuden mittaamisella. Fanien tuntemuksia on kuitenkin vaika arvottaa, kun osa seuraa ja kannattaa urheilua, osa joukkuetta, ja osa yksittäistä pelaajaa (Miller, 2015).

3.2 Urheiluanalytiikan käsite

Alamarin ja Mehrotran (2011a) mukaan urheiluanalytiikka tarkoittaa rakenteellisen historiallisen datan hallintaa, ennakoivien analyyttisten sovellutusten käyttöä, jotka hyödyntävät kyseistä dataa ja tietojärjestelmien käyttöä helpottamaan päätöksentekijöitä saavuttamaan organisaatiolle kilpailullinen etu pelikentällä. Urheiluanalytiikassa merkitys on uusien mallien etsimisessä löytääkseen jotain, mitä muut eivät vielä ole löytäneet (Szymanski,

2020). Malleja hyödynnetään otteludataan avaintekijöiden tunnistamiseksi (Baumer, Jensen & Matthews, 2015). Tiivistäen, urheiluanalytiikka sisältää datan keräämisen, uusien mittausyksiköiden kehittämisen ja eri analyttisten mallien integroimisen päätöksentekoon (Alamar, 2013). Mallina prosessi on sama, kuin luvussa kaksi esitetty Runklerin (2016) data-analyysiprojektin malli.

Urheiluanalytiikka on kilpailullisen edun saavuttamisen väline, joka tekee voittamisen ja menestymisen todennäköisemmäksi (Miller, 2015). Szymanskin (2020) mukaan kilpailullinen etu muodostuu urheiluanalytiikalla suorituksesta saatavasta datasta ja siitä tehtävästä analyysistä, mikä johtaa ottelun ja kilpailun luonteen ja rakenteen parempaan ymmärrykseen.

Urheiluanalytiikassa merkittävä kysymys on se, kuinka mennään yksilön suorituksen taakse kuvaamaan hänen merkitystään joukkueelleen (Miller, 2015). Käsite, kuinka yksilön tai joukkueen suoritusta halutaan mitata, johtaa uusien mallien syntyyn. Urheiluanalytiikka tarjoaa malleja, numeroita ja analyysijä mittaamaan ihmisiä, jotka pelaavat peliä (Alamar, 2013). Urheiluanalytiikasta saatavilla kehittyneillä tilastoilla pelaajan kokonaissuoritusta pystytään arvioimaan paremmin (Baumer ym., 2015).

3.2.1 Urheiluanalytiikan arvon muodostuminen

Alamarin (2013) mukaan urheiluanalyttisilla ohjelmilla on kaksi tavoitetta: ajan säästäminen ja uniikin informaation tuottaminen. Ajan säästäminen näkyy konkreettisesti esimerkiksi valmentajien videotyössä valmistautuessa otteluun. Video on urheilussa datan keskeisin lähde (Davenport, 2014). Urheiluanalyttisilla ohjelmilla ottelut pystytään koostamaan helpommin purettavaan ja suodatettavaan muotoon, jolloin prosessi juuri halutun tiedon löytämiseksi ottelusta muuttuu nopeammaksi ja tehokkaammaksi. Nopeammat prosessit säästävät valmennukselta aikaa, mistä Alamarin (2013) mukaan kilpailullinen etu syntyy, kun valmentajalle tulee lisää aikaa keskustella pelaajien kanssa ja analysoida peliä.

Alamarin ja Mehrotran (2011a) urheiluanalytiikan määritelmässä tuotiin esille tietojärjestelmien osuus urheiluanalytiikassa. Tietojärjestelmien kehitys onkin merkittävässä osassa urheiluanalytiikan kehittymistä. Uusia tietojärjestelmiä on muodostunut, niin työpöytä-, mobiili-, kuin verkkoalustoille (Alamar & Mehrotra, 2011b).

Tiedon keskittäminen yhteen tietokantaan, eri lähteiden lukumäärän karsiminen ja koostaminen yhteen, on rakennuskivi tehokkaille tietojärjestelmille ja joukkueet jotka eivät ole näin tehneet, kuluttavat ajallista hyötyä, jota tietojärjestelmät voivat tarjota (Alamar, 2013). Tiedon keskittämiseen viitaten, myös Alamar ja Mehrotra (2011a) näkevät ongelmaksi eri päätöksentekijöiden pääsyn toisen päätöksentekijän tietoihin organisaation sisällä. Kun organisaation kaikki data on linkitetty yhteen, tieto löytyy nopeammin (Alamar, 2013). Tämä ongelma näkyy varsinkin isommissa organisaatioissa.

Tiedon keskittämisestä yhteen tietokantaan hyvänä esimerkkinä toimii pelaajien seuranta. Pelaajasta saadaan tietoa eri lähteistä eri valmentajilta ja pelaajatarkkailun ammattilaisilta, ja tiedon koostaminen yhteen voi olla haastavaa (Alamar, 2013). Tällaista tietoa ovat esimerkiksi pelaajan fysiologiset tiedot, pelaajan terveydelliset tiedot, pelaajan taidoista tehtävät raportit, pelaajan markkinallinen arvo ja pelaajan merkitys joukkueelle. Varsinkin isommissa organisaatioissa data ja tieto voivat sijaita eri työntekijöiden hallussa (Alamar & Mehrotra, 2011a). Koostettuna yhteen eri lähteistä tulleet tiedot, saadaan tilastot, videokoosteet, laajemmat vertailut ja useiden eri pelaajatarkkailijoiden raportit samaan paikkaan päätöksentekijän käsiteltäväksi (Alamar, 2013). Tiedon sijainti samassa paikassa ymmärrettävästi nopeuttaa prosesseja ja näin toimii ajan säästäjänä ja arvon tuottajana urheiluorganisaatioissa.

Suurin osa urheiluanalyttisestä toiminnasta koostuu kuvailevista analyysistä, ja toimintaa tulisi viedä enemmän ennakoivaan ja ohjaavaan analytiikkaan (Davenport, 2014). Alamarin ja Mehrotran (2011a) mukaan ennakoivat analyysit ovat prosesseja, joissa tilastollisia malleja sovelletaan dataan tavoitteena saada tietoa, mitä luultavasti tulevaisuudessa tulee tapahtumaan. Tällainen tieto urheilussa voi käsittää esimerkiksi urheilijan pelaajauran kestoa tai vastustajan vahvuuksien ja heikkouksien suhteuttamista oman joukkueen suorituskykyyn. Ennakoivalla analyysillä voidaan myös soveltaa pelaajan istuvuutta joukkueeseen, oli kyseessä sitten vapaa pelaaja (*eng. Free agent*), tai pohjoisamerikkalaiseen kulttuuriin kuuluva varattava pelaaja (*eng. Draft eligible*). Nämä analyysit voivat vaihdella yksinkertaisesta vertailusta monimutkaisiin tilastollisiin analyysihin. (Alamar & Mehrotra, 2011a).

Data itse ei puhu ja hyödylliset ennusteet eivät nouse tyhjästä (Miller, 2015). Urheiluanalytiikassa keskustelua herättää niin sanotun vanhan koulukunnan valmentajien toiminta analyttisen, teknologisen, kehityksen hidastajina. Davenport (2014) tuo myös kyseisen ilmiön ilmi todeten, että päätöksenteossa liian usein oma intuitio ja kokemus menevät datasta saatavan informaation edelle. Alamar (2013) korostaa, että urheiluanalytiikka ei ole vain vanhan koulukunnan valmentajia vastaan, vaan heiltä saatu palaute on yksi tärkeimpiä keinoja analyttisten metodien kehittämiseksi. Samalla kun analytiikka pyrkii auttamaan valmennusta päätöksenteossa, pyritään valmennukselta saamaan apua parempien analyttisten metodien ja mallien kehittämiseen. Tulee muistaa, että päätöksentekijät ovat tärkeimpiä urheiluanalytiikan asiakkaita (Alamar & Mehrotra, 2011a). Päätöksentekijät tässä yhteydessä käsittävät seurajohdon sekä valmennuksen. Tästä voidaan johtaa, että analytiikan arvo voidaan nähdä pakottavana voimana aina tarkempien ja tarempien kysymyksien esittämiseen (Alamar, 2013).

3.2.2 Urheiluanalytiikan haasteet

Edellisen perusteella on helppo todeta urheiluanalytiikan merkitys joukkueurheilulle, mutta urheiluanalyttinen kehitys olisi voinut olla

voimakkaampaa. Alamar (2013) nimeää joukkueiden analytiikan käytön esteiksi teknologiset haasteet ja johdon haasteet.

Alamarin ja Mehrotran (2011b) yksi keskeisin rakenteellinen este on ymmärrys, mistä urheiluanalytiikka ja -analyttinen toiminta pitää aloittaa. On olemassa useita ohjelmistotarjoajia ja on haastavaa löytää oikea väline kilpailuedun saavuttamiseksi, jos ei ole kokemusta analytiikan metodeista ja mahdollisuuksista (Alamar & Mehrotra, 2012). Lisähaasteen urheiluanalytiikkaohjelmien implementoinnissa organisaatioon tuo parhaiden käytänteiden puuttuminen (Alamar & Mehrotra, 2012). Koska kenttä on uusi, on olemassa myös organisaatioita, joilla ei ole mitään ymmärrystä analyttisestä kentästä ja analytiikan käytön tulos on sen mukaista (Alamar, 2013). Urheiluanalytiikka on kuitenkin tullut oppiaineeksi yliopistoihin ja erilaisten urheilun tutkimuskeskusten huomioon, mikä auttaa parhaiden käytänteiden syntymisessä ja lisää alan osaajien määrää, mistä organisaatiot voivat palkata urheiluanalytiikan osaajia (Alamar & Mehrotra, 2012).

Rakenteellinen este on edellä mainittu tekijöiden puute, sillä harvalla on akateeminen koulutus analytiikkaan yhdistettynä intohimoon urheilua ja urheilutuotetta kohtaan. Lajista, mihin analyttisiä sovellutuksia käyttää, tulee tekijöiden tietää tarpeeksi ollakseen vaarallisia vastustajan näkökulmasta. Ihmisiä on olemassa tarpeeksi rakentamaan matemaattisia malleja, mutta mallit eivät välttämättä vastaa lajin vaatimuksia. Vastaavasti urheiluorganisaation omistajilla ja päätöksentekijöillä ei ole tilastollista kokemusta, eivätkä tunnista erilaisten matemaattisten mallien mahdollisuuksia, ongelmia, puutoksia ja vajaavaisuuksia. (Alamar & Mehrotra, 2011b; Alamar&Mehrotra, 2012).

Muutosvastarinnan vaikutus urheiluanalytiikan jalostamisessa organisaation toimintaan tulee tässäkin liiketoiminnan alalajissa, urheiluliiketoiminnassa, huomioida. Bovey ja Hede (2001) tutkimuksessaan osoittavat tunteiden vaikutuksen irrationaalisuuteen ja sitä kautta muutoksen vastarintaan. Urheilumaailmassa monet valmentajat, päätöksentekijät, ovat kasvaneet lajiin vuosien pelaamisen ja valmentamisen tukemana (Alamar & Mehrotra, 2012). Heillä henkilökohtainen kokemus ja tunne määrittelevät, miten lajia kuuluu pelata ja näin ollen datasta tulevaa informaatiota ei pidetä luotettavana, varsinkaan kun se tulee henkilöiltä, joilla ei ole kokemusta lajin ytimestä (Alamar & Mehrotra, 2012). Davenport (2014) nosti myös tämän ongelman esille, jossa intuitio menee informaation edelle päätöksenteossa.

Urheilun ja analytiikan välille rakennettava yhteinen kieli on elintärkeää urheiluanalytiikkaprosessien implementoinnissa organisaatioon (Alamar & Mehrotra, 2012). Kuten edellä on todettu, tämä käsittää molemmat puolet, niin vaikeaselkoisen analyttisen puheen, kuten analytiikkaosaston ymmärryksen urheilun vaatimuksista. Alamarin ja Mehrotran (2012) mukaan urheiluanalytiikan evoluution nopeuteen vaikuttaa se, miten urheilujohtajat kokevat analytiikan sijoituksena kilpailullisen edun saavuttamiseksi ja uskoa tähän ei tule, elleivät analyttikot osaa välittää viestiänsä oikein.

3.3 Teknologinen kehitys urheiluanalytiikassa

20 vuoden aikana kehitystä on tapahtunut niin informaatioteknologiassa, tietotekniikassa, urheilulähetyksissä, kuten monitoroinnissa, kun ennen käytössä olivat vain tilastot ja raaka otteluvideo (Alamar & Mehrotra, 2011a). Kehittyneet teknologiat ovat mahdollistaneet tietojen irti saamisen pelistä paremmin ja suorituksen analysoimisen korkealla tasolla (Szymanski, 2020). Lisäksi internet nopeuttaa eri tahojen raporttien kautta tapahtuvaa viestin välittymistä (Alamar & Mehrotra, 2011a).

Videon jälkeen toinen keskeinen datalähde urheilussa on tila- ja liikeseensorit (Davenport, 2014). Monitorointitekniologia mahdollistaa yksityiskohtaisen informaation pelaajille ja joukkueenjohdolle (Glickman & Stern, 2016). Nykyiset tilasensorit tunnistavat pelaajan ja pelivälineen useita kertoja sekunnissa reaaliaikaisesti ja näitä lähteitä analytiikot hyödyntävät pelin sisällä mittaamiseen ja mallien tekemiseen, mitä hyödynnetään päätöksenteossa (Miller, 2015).

Tietojärjestelmät lisäävät tehokkuutta ja kehittyneet tietojärjestelmät muuttavat valmentajien valmistautumista otteluun (Alamar, 2013). Analyttiset järjestelmät voivat automaattisesti tunnistaa vastustajan suorituskyvyn muutoksen positiiviseen tai negatiiviseen suuntaan (Alamar, 2013). Tällä saavutetaan tiukassa aikataulussa suoritettavaa laajempaa tarkastelua. Esimerkkinä valmentajan on helppoa nähdä, jos joukkue on voittanut viimeisestä seitsemästä pelistään yhden, mutta vaikeampi käydä jokainen ottelu läpi löytääkseen syy, miksi kyseiseen lopputulokseen on päädytty (Alamar, 2013). Jos tiukassa aikarajoituksessa otannaksi saadaan edelliset kaksi ottelua, voi neljässä edellisessä ottelussa tapahtuneet säännönmukaisuudet jäädä huomioimatta. Isompi otanta mahdollistaa yhdistävien tekijöiden löytämisen tietyille joukkueelle hävinneiden ja voittaneiden joukkueiden välillä (Alamar, 2013). Lisäksi tietojärjestelmät mahdollistavat visualisoinnin ja interaktiivisen analyysin useista eri lähteistä alustan integroidessa eri lähteistä tullutta dataa (Alamar & Mehrotra, 2011a).

Datan määrä päätöksentekijöille on kasvanut eksponentiaalisesti 15 vuoden aikana ja sitä tulee useista eri lähteistä, useissa eri muodoissa (Alamar & Mehrotra, 2011a). Jo pelkästään liikkeentunnistusteknologian mukaan tulo lisää informaation määrän ottelussa muutamasta sadasta rivistä aina miljoonaan riviin (Alamar & Mehrotra, 2011a). Tilasensorien lisäksi erilainen kannettava, ihmisessä kiinni oleva, monitorointitekniologia lisää datan ja informaation määrää huomattavasti (Passfield & Hopker, 2017). Saatavilla olevan datan määrän ja tietokoneiden tehokkuuden kasvaessa, myös data-analyysien määrä kasvaa (Szymanski, 2020). Datan määrän kasvaessa tulee huomioida riittävä IT-osaaminen, sillä modernit teknologiat tarvitsevat laskentatehoa (Davenport, 2014).

Datan lisääntynyt määrä, nopeus millä sitä syntyy ja arvo päätöksenteossa, huomioituna aiemmin esitettyyn datan ja massadatan määritelmään, voidaan todeta joukkueurheilun huipputasolla hyödyntävän massadataa urheiluanalytiikassaan. Kun datamäärä on tarpeeksi suuri, voidaan siitä rakentaa malleja, jotka mahdollistavat ennakoinnin, optimoinnin ja päätöksenteon ehdottamisen

(Davenport, 2014). Kuten edellisessä käsittelyluvussa on todettu, massadatan käyttö kasvattaa otantaa, ja isompi otanta edistää säännönmukaisuuksien löytämistä (Szymanski, 2020). Suurten datajoukkojen ymmärtäminen säännönmukaisuuksien löytämiseksi vaatii kuitenkin urheiluanalyttisten metodien kehittämistä (Passfield & Hopker, 2017).

Hyvä datanhallinta on keskeistä, sillä monimutkaiset datajoukot sisältävät usein virheitä (Alamar, 2013) sekä datasta johdetaan suoraan tietojärjestelmiin ja analyttisiin malleihin, ja näin ollen heikkolaatuinen data heikentää analytiikan arvoa ja organisaation sijoitusta urheiluanalytiikkaan kilpailullisen edun katoamisen muodossa (Alamar, 2013; Alamar & Mehrotra, 2011a). Datanhallinnalla Alamar ja Mehrotra (2011a) käsittävät kaikki prosessit, jotka yhdistetään datan hankintaan, todentamiseen ja tallentamiseen. Esimerkkejä näistä prosesseista ovat datan standardointi, keskittäminen ja integrointi, joihin on tarjolla ulkoisesti ostettavia järjestelmiä (Alamar, 2013). Teknologinen kehitys mahdollistaa virheentunnistus prosessin, jolla voidaan tunnistaa epä johdonmukaisuuksia datasta (Alamar, 2013) ja näin ollen edistää edellä mainittua datanhallintaa.

Uusien monitorointi teknologioiden lisäksi historiallista dataa edellisistä otteluista on pystytty paketoimaan talteen (Alamar & Mehrotra, 2011a). Tämä toimii otteluiden simuloimiseen erilaisia tekoälyn ja koneoppimisen sovellutuksia hyödyntäen. Koneoppimisen ja tiedon louhinnan käytöistä urheiluanalytiikassa avaa Chazan-Pantzas (2020) tarkemmin väitöskirjassaan, jossa hän tutki joukkueen suorituksen ja pelaajien yksilösuoritusten ennakoimista. Koneoppimisella pystytään ehkäisemään loukkaantumisia, arvioimaan pelaajan urheilullista ja markkinallista arvoa sekä ennakoimaan joukkueen suorituskykyä (Chazan-Pantzas, 2020).

Simuloinnissa matemaattisilla malleilla voidaan määrittää parasta kokoonpanoa otteluun maksimaalisen odotusarvollisen tuloksen saavuttamiseksi (Miller, 2015). Historiallisen datan yhdistäminen ennakoiviin malleihin on oleellista, sillä se muuttaa menneen raakadatan informaatioksi (Alamar & Mehrotra, 2011b) ja ennakoivat mallit toimivat tässä suhteessa kilpailullisen edun määrittelijöinä (Miller, 2015). Ennakoivista malleista saatava informaatio voidaan liittää Alamarin (2013) urheiluanalyttisen toiminnan tavoitteista uniikkiin informaatioon.

Massadatan hyödyntäminen otannan kasvattajana mahdollistaa tuloksien ennakoinnin (Szymanski, 2020). Näitä ennakoivia malleja ovat esimerkiksi Monte Carlo -simulaatio, Poisson-prosessi ja Maseyn-metodi (Szymanski, 2020). Devlin ja Treloar (2018) näkevät tuloksien ennakoinnissa ja sitä kautta joukkueiden arvottamisessa kolmena tunnetuimpana metodina Maseyn-, Colleyn- ja Markovin metodit. Metodit eroavat toisistaan joukkueiden arvioinnissa vastustajan tason suhteen (Balreira, Miceli & Tegtmeyer, 2014). Vastustajien taso voi heitellä arviointiperusteissa, jossa voittamalla heikkotasaisen vastustajan annetaan vähemmän arvoa, kuin voittamalla kovatasaisen vastustajan. Julkisesti saatavilla olevat ennakoivat mallit tuloksien ennakoinnin kannalta eivät kui-

tenkaan ole tarpeeksi hyviä, esimerkiksi vedonlyöntivälittäjien voittamiseen (Szymanski, 2020).

Myös Valero (2016) tutki ottelutulosten ennakointia baseballissa 10 vuoden historiadataa ja tiedonlouhintaa hyödyntäen. Ennakoivilla malleillaan Valero (2016) pääsi hieman alle 60 % tarkkuuteen, eli kuitenkin arvaamista parempaan tarkkuuteen, muttei hänen mukaansa riittävään tarkkuuteen vedonlyöntikontekstissa. Valeron (2016) ja Szymanskin (2020) näkemyksistä kuitenkin tulee huomioida, että mikäli sinulla on hallussasi malli, jolla voitat urheiluvedonlyöntivälittäjät, tuottaa tämä tieto sinulle rahaa eikä sitä kannata julkistaa. Mallin rakentaminen on kuitenkin mahdollista, kuten Chazan-Pantzalis (2020) toteaa väitöskirjassaan. Kuitenkin Chazan-Pantzalin (2020) rakentama malli pääsi Englannin Valioliigassa 57 % tarkkuuteen, eli samoihin tarkkuuksiin, kuin Valero (2016) tutkimuksessaan.

Ulkoisia palveluita ottelutulosten ennakointiin, ja täten vedonlyöntivälittäjien voittamiseen on olemassa. Esimerkiksi Sharp Football Analysis ja heidän ennakoiva mallinsa lupaa 60 % ennakointitarkkuuden ottelun pistemääriin, ja myy tätä palvelua kesken kauden yli 500 dollarin kausihinnalla (Sharpfootballanalysis, 2020).

3.4 Tiivistelmä urheiluanalytiikasta

Urheiluanalytiikka käsittää rakenteellisen historiallisen datan hallinnan, ennakoivien analyttisten sovellusten käytön ja tietojärjestelmien käytön päätöksenteon helpottamisessa (Alamar & Mehrotra, 2011a). Analyttisistä malleista pystytään saavuttamaan ihmisten mittaamista paremmilla yksiköillä ja ymmärtämään paremmin niin sanotusti tuloksen taakse. Urheiluanalytiikka siirtää päätöksenteon olettamuksesta tietoon pohjautuvaan päätöksentekoon. Urheiluanalytiikalla on mahdollista saavuttaa kilpailullinen etu (Szymanski, 2020; Miller, 2015; Alamar, 2013).

Alamarin (2013) mukaan urheiluanalytiikan arvo konkretisoituu ajan säästämiseen ja uniikin informaation tuottamiseen. Ajan säästäminen näkyy parhaiten valmentajan valmistautumisessa otteluun, jolloin otannat saadaan suuremmiksi riittävän hyvien johtopäätösten saavuttamiseksi. Säästynyt aika pystytään käyttämään muihin toimiin, kuten pelaajien henkilökohtaiseen kehittämiseen. Uniikki informaatio muodostuu uusista teknologisista innovaatioista, kuten monitorointiteknologioista sekä historiadataa hyödyntävistä ennakoivista analyyseistä.

Aiempaa tutkimuskirjallisuutta lukiessa selkeästi korostuu, että data itsessään ei tee paremmaksi, vaan se vaatii sekä oikean asiantuntijan datan käsittelyyn, että oikean päätöksentekijän tarvittavien johtopäätösten rakentamiseen kohdelajin saralla. Alamarin (2013) mukaan päätöksentekijöiden tärkein ominaisuus analyttisten ohjelmien kehittämiselle on kyky ja halukkuus kysyä kysymyksiä, niin koskien analytiikkaa, kuin päätöksiä mitä niistä tulisi johtaa. Ajatusmaailma, jossa analytiikasta on tuotu yksi rasite lisää, ja jolle pitää löytää

aikaa, on este analytiikan adaptaatiossa ja aiheuttaa konflikteja organisaation sisällä (Alamar & Mehrotra, 2012).

Vaikka analyttikot usein tarjoavatkin hyvän näkemyksen urheilusta, ei heidän tietämyksensä lähtökohtaisesti ole vertailtavissa päävalmentajan tietämykseen, jolloin päätöksenteon lisäarvo muodostuu yhdessä päävalmentajan näkemyksestä ja hyvän kysymyksenasettelun seurauksena muodostuneesta analyttisestä lisäoivalluksesta (Alamar, 2013). Kommunikaation puutos analyttikon ja päätöksentekijän välillä mahdollisesti heikentää saatavilla olevaa lopputulosta. Alamar ja Mehrotra (2011a) toteavat, että edelleen liikaa aikaa käytetään siihen, kun analyttikko yrittää löytää vastauksia kysymyksiin, jotka eivät ole päätöksentekijälle tärkeitä.

Datan muodon ja luonteen muuttuminen on pakottanut palkkaamaan uusia ammattilaisia organisaatioihin (Alamar & Mehrotra, 2011b). Uusien ammattilaisten palkkaaminen voi tarkoittaa yhteistyökuviota analyttisten ohjelmien palveluntarjoajan kanssa, esimerkkinä MLB:ssä San Francisco Giants, tai hyödyntämällä urheilun intohimoisia faneja, joilla varmasti on osaaminen tiettyihin analyttisiin malleihin olemassa olevasta datasta (Davenport, 2014). Tärkeää on saada eri osastojen päätöksentekijät ja analyttikot ymmärtämään yhteistyön merkitys yhteisen tavoitteen saavuttamiseksi (Alamar & Mehrotra, 2012). Johdon haasteiden lisäksi analyttiikkaan sisältyy teknologisia haasteita ja organisaatiokulttuurisia haasteita.

Kuten Philadelphia 76ers koripallotoiminnan johto, analytiikan saralla menestynyt, Daryl Morey on todennut, kilpailullinen etu syntyy omista järjestelmistä ja datasta mitä kellään muulla ei ole (Davenport, 2014). Täten urheiluanalytiikassa tulee huomioida tietty salamyhkäisyys, kun puhutaan tiedosta, mikä päävalmentajilla on ja mitkä siirretään pelaajille. Jos on olemassa omaa joukkuetta edesauttava tieto, ei sitä luonnollisesti kannata paljastaa. Tämä rohkaisee salassa pitämiseen ja samalla heikentää urheiluanalytiikan julkisen tietämyksen kehittymistä (Szymanski, 2020). Toisaalta tulee huomioida, että ammattilaisurheilussa valmentajat ja toimitusjohtajat vaihtuvat usein, mikä tarkoittaa analyttisten ajatusten liikkumista (Davenport, 2014).

4 URHEILUANALYTIikka POHJOISAMERIKKALAISEN AMMATTILAIJSJOUKKUEURHEILUN KONTEKSTISSA

Pohjois-Amerikassa mielenkiinto ja toiminta urheiluanalytiikkaa kohtaan ovat räjähtäneet, kun vaikka 2005 vain muutamat koripallon ammattilaissarja NBA:n joukkueet miettivät analytiikan sisällyttämistä joukkueen strategiaan (Alamar & Mehrotra, 2011a). Michael Lewisin vuonna 2003 julkaisema teos ”Moneyball” toimi katalysaattorina niin kehittyneiden tilastoanalytiikan mallien hyödyntämiseen organisaatiossa sekä yleisenä julkisena kiinnostuksen herättäjänä urheiluanalytiikkaa kohtaan (Baumer & Zimbalist, 2014). Joukkueet selvästi panostavat analytiikkaan palkkaamalla henkilöstä ja luomalla kehittyneitä tietojärjestelmiä (Alamar, 2013).

Tässä luvussa käsitellään pohjoisamerikkalaista ammattilaisjoukkueurheilua. Tämä käsittää niin sanotut isot sarjat (*eng. major leagues*): amerikkalaisen jalkapallon NFL, baseballin MLB, koripallon NBA ja jääkiekon NHL. Pohjoisamerikkalaisen joukkueurheilukulttuurin tarkastelun jälkeen käsitellään tarkemmin urheiluanalytiikkaa baseballissa sen ollessa kehittyneintä urheilussa (Miller, 2015). Luvussa tarkastellaan niin sanottua ”moneyball”-ilmiötä, kehittyneitä henkilökohtaisen suorittamisen tilastoja ja lisäksi aiempaa tutkimusta, miten urheiluanalytiikkaa hyödyntävät joukkueet ovat menestyneet kilpailussa.

4.1 Pohjoisamerikkalaisen joukkueurheilukulttuurin luonne

Urheilumarkkinan arvo Pohjois-Amerikassa kasvaa. Miller (2015) listaa vuonna 2015 kolmeksi rahallisella arvolla mitattuna suurimmaksi NFL joukkueeksi Dallas Cowboysin 3,2 miljardilla, New England Patriotsin 2,6 miljardilla ja Washington Football Teamin (2015 Washington Redskins) 2,4 miljardilla. Moskowitz (2020) verkkoartikkelissaan listaa samojen seurojen arvoiksi vuonna 2019 Washington Football Team 3,4 miljardia, New England Patriots 4,1 miljardia ja Dallas Cowboys 5,5 miljardia. Joukkueiden arvoissa huomataan miljardista yli

kahteen miljardiin kasvanutta arvoa. Tämän perusteella voidaan todeta ainakin rahallisella arvolla mitattujen kärkiseurojen arvon kasvavan vuosi vuodelta. Yritysten arvon kasvaessa myös pelaajakompensaatiot ovat kasvaneet räjähdysmäisesti, mikä on saanut päätöksentekijät etsimään uusia malleja laajemman informaation saamiseksi arvioitavista pelaajista (Baumer & Zimbalist, 2014).

Suurin osa huippu-urheilusarjoista Pohjois-Amerikassa sisältää palkkakattojärjestelmän (Miller, 2015). Vaihtoehtoja tälle kuitenkin löytyy. Esimerkiksi MLB:ssä on käytössä Competitive Balance Tax -järjestelmä. Toiselta nimeltään luxury tax (MLB, 2020), on järjestelmä, jossa rajakynnyksen ylittävät joukkueet maksavat veroa jokaisesta dollarista millä ylittävät rajakynnyksen, veron kasvaessa suuremmaksi jokaisesta toistuvasta vuodesta, kun kynnyks ylitetään.

Palkkakatolla pyritään tasaamaan kilpailu, kun huippu-urheiluorganisaatiot kilpailevat toistensa kanssa pelaajista, eli työvoimasta. Palkkakaton tasainen nousu tarkoittaa myös huippupelaajien maksimaalisen ansion kasvua, mikä voidaan huomata esimerkiksi NFL pelinrakentajien tuoreita pelaajasopimuksia tarkasteltaessa. Kärkipelaajat vievät prosentuaalisesti ison osuuden palkkakatosta. Tämä herättää kysymyksen siitä, miten kustannukset pelaajiin vaikuttavat menestykseen (Miller, 2015), mihin Alamar ja Mehrotra (2012) ovat todenneet, että ainakin MLB:ssä joukkueen sisällä olevat palkkaerot korreloivat negatiivisesti joukkueen voittoprosenttiin. Tästä voidaan johtaa kysymys, onko seurojen järkevää panostaa isolla rahalla vapaiden pelaajien markkinoiden alkaessa (*eng. Free Agent Frenzy*) vai rakentaa joukkuetta ja tulevaisuutta varausjärjestelmän kautta.

Joukkueen rakentamisen suhteessa kehittyneistä analytiikoista on hyötyä. Voidaan mitata kokeneen pelaajan suorituskykyä suhteessa kasvavaan ikään (Miller, 2015) ja pystytään ennakoimaan pelaajan roolia ja merkitystä joukkueelle, ja sitä kautta rakentamaan oikean arvoista pelaajasopimustarjousta (Alamar, 2013). Lisäksi, että pystytään mittaamaan yksittäisen pelaajan suoritusta, pystytään ennakoimaan miten hän todennäköisesti tulee toimimaan vieressä olevien pelaajien kanssa. Pelaajakemiat tuovat kuitenkin haasteen, sillä saman pelaajan henkilökohtainen arvo voi olla toisen pelaajan kanssa suurempi, kuin mitä toisen pelaajan kanssa (Miller, 2015). Kehittynyt monitorointitekniologia kuitenkin tarjoaa mahdollisuuksia huomioida pelaajien välillä olevaa kemiaa (Glickman & Stern, 2016).

Varausjärjestelmässä yksittäiseen varattavaan pelaajaan, varsinkin ensimmäisellä kierroksella varattuun, kohdistuu kovat pelilliset odotukset sekä taloudellinen investointi. Kuten edellisessä luvussa todettiin, Alamar ja Mehrotra (2011a) kuvaavat ennakoivat analyysit tilastollisiksi malleiksi, joilla pyritään saamaan tietoa, mitä tulevaisuudessa tulee tapahtumaan. Ennakoivien mallien käyttö organisaatiossa yleensä aloitetaan varausjärjestelmän yhteydessä, sillä ison osan historiadatasta ollen julkisesti saatavilla, ei mallien rakentamiseen tarvitse kuluttaa valtavasti aikaa (Alamar & Mehrotra, 2011b). Prosessissa tilastotijat työskentelevät pelaajatarkkailijoiden kanssa erilaisten mallien rakentamiseksi (Szymanski, 2020).

Käytännössä tässä pelaajan arviointiprosessissa ennakoivilla analyyseillä ollaan varaamassa pelaajaa A ja verrataan hänen yliopisto- ja junioriprofiiliansa saman tyyppiseen pelaajaan B, joka on jo kilpaillut ammattilaistasolla. Tästä pyritään rakentamaan malleja, miten pelaaja A tulee menestymään organisaatiossa ja ammattilaissarjassa (Alamar & Mehrotra, 2011b). Varaustilaisuus on Pohjoisamerikkalaisessa kulttuurissa muutenkin toiminut uusien tietojärjestelmien käyttöönoton laukaisijana, kun on tuotu erilaisia interaktiivisia raportointityökaluja vanhojen pelaajataulujen (*eng. big board*) korvaamiseksi (Alamar & Mehrotra, 2011b).

4.2 Baseball ja urheiluanalytiikka

Baseballissa käytettävä analytiikka on kehittyneintä urheilussa (Miller, 2015) ja muissa lajeissa analyttinen kehitys on ollut hitaampaa (Szymanski, 2020). Jo vuonna 2006 SportVision alkoi kerätä liikedataa jokaisesta MLB:ssä syötetystä syötöstä (Alamar & Mehrotra, 2011b). Toisaalta baseballissa analyttinen kehitys näkyy aina juniorisarjoissa asti, mitä voidaan pohtia, viekö se juniorisarjat kilpailullisempaan tuloskeskeiseen muotoon, eikä pelaajaa ja ihmistä kehittävään muotoon (Sanderson & Baerg, 2018).

Eri urheilulajeissa on tunnistettavissa keskenään yhtenäisiä ja erottavia piirteitä. Baseball ja golf vaativat äärimmäistä tarkkuutta osua palloon, amerikkalainen jalkapallo ja jääkiekko nousevat esiin fyysisyydellä sekä jalkapallo ja koripallo vaativat aerobista kestävyyttä (Miller, 2015). Millerin (2015) ja Szymanskin (2020) mukaan baseball hajoaa muista joukkueurheilun muodoista sen koostuessa summasta yksilökohtaisista, näkyvimpänä lyöjä vastaan syöttäjä.

Baseballin ollessa vähemmän joukkueorientoitunut, kuin muut joukkueurheilulajit, on data-analytiikan käyttöönotto paremmin jalostettavissa (Elitzur, 2020). Dynaamisissa lajeissa tällaisia selkeästi eroteltavia yksilö vastaan yksilö tilanteita ovat esimerkiksi rangaistuslaukaukset. Vaikka dynaamisen pelin sisällä on poimittavissa yksittäisiä tilanteita, koostuvat ne kuitenkin vuorovaikutuksesta, jossa tilanteeseen on edetty pelaajan X tehdessä jotain, ja pelaajan Y vastaavasti reagoiden pelaajan X toimintaan. Tilastollinen analyysi on monimutkaisempaa, kun on enemmän joukkueen sisäisiä interaktioita (Elitzur, 2020). Dynaamisen vuorovaikutuksen mittaamiseen oletettavasti kaikki tarvittava tieto on jo olemassa kentällä, mutta vielä ei ole tarvittavaa ymmärrystä tiedon irti saamiseksi (Szymanski, 2020).

Pelin koostuminen selkeistä erillisistä yksilötilanteista johtaa siihen, että suoritukset ja suorituksen tekijät pystytään helposti erottelemaan ja pystytään toteuttamaan erottelevaa tilastointia, mikä auttaa tunnistamaan pelaajien kyvykkyyksiä (Baumer ym., 2015). Urheiluanalytiikka mahdollistaa yksittäisten suoritusten objektiivisen mittaamisen (Sanderson & Baerg, 2018). Tapahtumat (syötöt) ovat yksittäisiä suorituksia, eikä pelissä ole jatkuvaa kelloa tai aikarajoitetta. Baseballissa mahdollisia mitattavia yksittäisiä asioita ovat syötön sijainti ja

voimakkuus, lyönnin osumatodennäköisyys historiadatalla, lyönnin voimakkuus sensoriteknologialla sekä lyönnin sijainti lyöntikartalla (*eng. heat map*).

Myös Szymanski (2020) tuo esiin dynaamisen pelin eron baseballiin, kun dynaamisissa lajeissa peli on jatkuvassa liikkeessä, eikä tapahdu selkeästi yksittäinen suoritus kerrallaan. Lisäksi, että tapahtumat ovat yksittäisiä, myös erilaisia mahdollisia tilanteita on vähemmän. Baseballissa erilaisia tilanteita on 24, johon Szymanski (2020) ottaa vertaukseksi amerikkalaisen jalkapallon, jossa johtopäätöksien saamiseen pitäisi olla tarvittava otanta tilanteista, kuten esimerkiksi tilanne ”3&11, 4min peliaikaa jäljellä, alle yhden pallonhallintavuoron (*eng. possession*) johtoasema”. Vastaavasti baseballissa tilanteet rakentuvat muun muassa seuraaviin muotoihin, kuten 2-tilanne yhdellä palolla ja ajolähtö kahdella palolla.

4.2.1 Moneyball

Michael Lewisin teos *Moneyball* (2003), josta myös Bennett Miller on ohjannut vuonna 2011 samaa nimeä kantavan elokuvan, muutti asenteita urheiluanalytiikkaa kohtaan. Teos kertoo tositapahtumiin perustuen, dramatisointia sisältäen, miten Oakland Athleticsin general manager Billy Beane otti dataanalytiikkaa käyttöönsä kilpailullisen edun saavuttamiseksi. Teoksessa nostetaan esille perinteisen pelaajatarkkailun virheellisyyttä, kun perinteiset tilastot eivät riitä kuvaamaan lajin vaatimuksia. Millerin (2015) mukaan *moneyball*-teoksen teesi ja samalla urheiluanalytiikan pohjimmainen eetos ovat, että myös pienen markkinan joukkueet voivat voittaa ja menestyä kilpailussa käyttämällä rahansa viisaasti.

Lewisin teoksen mukaan Beane näki pelaajan etenemisen pesälle tärkeämpänä tilastona, kuin lyöntitehokkuutta ja tätä kautta sai hänen mielestään alipalkatuilla ja aliarvostetuilla pelaajilla rakennettua joukkueen, joka sarjan yhdellä alhaisimmista palkkarakenteista onnistui etenemään MLB:n pudotuspeleihin kausina 2000–2003. Suorituksen arvoa korostaa se, että isoista sarjoista baseballissa on kaikkein vaikein päästä pudotuspeleihin (Elitzur, 2020). MLB:ssä vain kolmannes pääsee pudotuspeleihin, kun NFL:ssä sarjajärjestelmäuudistuksen myötä yli 40 % joukkueista jatkaa pudotuspeleihin ja NBA:ssa sekä NHL:ssä luku on noin 50 %.

Beane perusteli *on-base-percentage* (OBP/OBA) -tilaston merkittävämpänä voiton määrittäjänä, kuin käytössä ollut *slugging percentage* (SLG)-tilasto, joka huomioi osumat, muttei ”vapaataipalia”. Teoksen julkaisun jälkeen OBP:n merkitys pelaajasopimusten arvoissa nousi ja ohitti SLG:n merkityksen, ja tätä tapahtumaa kutsutaan termillä ”*moneyball moment*”, kun huomattiin, että dataanalytiikka voi muuttaa joukkueen suoritusta ja ajattelutapaa baseballissa. Muissa lajeissa ei ole koettu vastaavaa ilmiötä. (Szymanski, 2020).

Lewisin teos ei kuitenkaan selviä ilman kritiikkiä. Baumerin ja Zimbalistin (2014) mukaan kunnia ei pelkästään kuulunut Billy Beanelle, vaan suuri merkitys oli urheiluanalytikko Paul DePodestan (elokuvassa nimellä Peter Brand) analytiikoissa ja ajatusmalleissa. Lisäksi, mikäli Billy Beane pohjasi päätöksen-

tekonsa pelkästään tilastoihin, etenkin OBP:hen, hänen pelaajakauppansa Jeremy Giambin ja Carlos Peñan kohdalla eivät olleet perusteltavissa, kun vaihdossa tuli takaisin pienemmän OBP-arvon pelaajia. Analyyttisiä malleja ja henkilöitä oli organisaatioissa jo 1900-luvulla, mutta Lewisin teoksen mukaan he eivät kuitenkaan olleet samalla tavalla päätöksenteon keskiössä, kuten Billy Beane ja Paul DePodesta. (Baumer & Zimbalist, 2014).

Moneyball terminä kuvastaa organisaatioita, jotka käyttävät data-analytiikkaa hyödyksi toiminnassaan, ja hyöty nähdään kehittyneenä pelaajahankintana ja ottelun hallintana (Elitzur, 2020). Kehittyneistä baseball analytiikoista käytetään myös nimeä *sabermetrics*, tai *SABRmetrics* (the Society for American Baseball Research) (Elitzur, 2020).

Elitzur (2020) tutki, että menestyvätkö niin sanotut moneyball-joukkueet paremmin kuin muut joukkueet ja onko kilpailullinen etu ylläpidettävissä pidemmällä aikavälillä. Tutkimus osoittaa, että moneyball joukkueilla on ollut kilpailullinen etu kustannukset suhteutettuna suoritukseen urheiluanalytiikan adaptaation myötä, mutta etu on vähitellen hävinnyt, kun data-analyttiset keinot ja konseptit ovat paljastuneet muulle markkinalle (Elitzur, 2020; Davenport, 2014). Tulos viittaa edellisessä käsittelyluvussa mainittuun Szymanskin (2020) väitteeseen urheiluanalyttisen tiedon pitämisessä salassa. Tästä voidaan tehdä väite, että tiedon tullessa julkiseksi sen arvo vähitellen katoaa. Sama ilmiö on myös huomattavissa Billy Beanen ja hänen mentorinsa Sandy Aldersonin menestyksessä. Beane ja Alderson menestyivät paremmin ennen, kuin teos Moneyball julkaistiin 2003 (Elitzur, 2020).

4.2.2 WAR

WAR, eli *wins above replacement*, tai *wins above replacement player* tilaston konsepti on ollut yksi urheiluanalytiikan ja sabermetricsin suuria saavutuksia (Baumer ym., 2015). Joukkueurheilussa pelaajan arvon määrittelemine joukkueen kokonaisuudelle on haasteellista, toisin sanoen miten jakaa voittojen merkitys pelaajien kesken (Baumer ym., 2015). WAR kuvastaa pelin jokaisen osa-alueen merkityksen pelaajan kokonaisuoritukseen, oli kyseessä sitten sisäpelisuoritus, tai ulkopelisuoritus (*eng. Fielding*) (Baumer ym., 2015). Tämä pelillinen merkitys suhteutetaan joukkueen voittoihin. Käytännössä WAR kuvastaa pelaajan merkityksen voittoina suhteessa perustason (*eng. replacement*) pelaajaan (Miller, 2015). Baumer ym. (2015) näkevät WAR-tilaston olevan osittain verrattavissa +/- tilastoon, jossa perusajatuksena on selvittää pelaajan suhdetta pisteiden tekemiseen ja pisteiden päästämiseen. Kuitenkin tulee huomioida, että +/- tilasto ei ole suhteessa joukkueen voittoihin.

Elitzur (2020) tutkimuksessaan toteaa, että joukkueen pelaajien yhteen laskettu WAR, *TeamWAR*, korreloi vahvasti joukkueen voitto% kanssa. Tästä voidaan todeta WAR tilastolla olevan selkeä merkitys menestyksen indikaattorina. Myös Baumer ja Zimbalist (2014) toteavat WAR tilaston yhtenä edellä mainitun Oakland Athleticsin 2000-luvun alkupuolella tapahtuneen menestyksen selittävästä tekijöistä.

WAR-tilastossa on useita hyödyllisiä ominaisuuksia. Baumerin ym. (2015) mukaan se auttaa määrittämään pelaajan arvon oikein, mitä pystytään hyödyntämään esimerkiksi sopimustarjouksia tehdessä tai vertailtaessa eri vapaita pelaajia keskenään. Millerin (2015) mukaan yksinkertaiset verrattavat tilastot ovat suositeltavia, koska niitä on helpompi selittää ymmärrettävästi faneille, valmentajille ja omistajille. WAR on tässä suhteessa helposti ymmärrettävissä, sillä esimerkiksi pelaajan X loukkaantuessa joukkue tulee voittamaan hänen WAR arvonsa verran vähemmän pelejä täydessä runkosarjassa. Mitä isompi arvo on, sitä merkittävämmästä pelaajasta on kyse. MLB-kauden 2019 arvokkain pelaaja oli ESPN:n mukaan Los Angeles Dodgersin Cody Bellinger WAR-arvon ollessa 9.0 (ESPN, 2020). Eli hän toi joukkueelleen 9 voittoa enemmän, kuin perustason pelaaja olisi tuonut.

Saadakseen WAR-tilasto toimivaksi, vaatii se tarkat tilastot sekä eiluonnollisten pelitilanteiden huomioimisen. WAR heikkouksia ovat epävarmuusestimoiminnin puute ja heikko toistettavuus (Baumer ym., 2015). WAR ei läpäise avoimuustestiä, ja siitä on eri versioita, kuten fWAR, rWAR ja WARP (Baumer ym., 2015), ja näiden laskentametodit eroavat esimerkiksi perustason pelaajan määrittämisen saralla (Miller, 2015). Tämä johtaa siihen, että erilaiset WAR-tilastot eivät ole keskenään vertailtavissa. On olemassa kuitenkin myös julkisia laskentamalleja, kuten openWAR vertailtavuuden edistämiseksi (Miller, 2015; Baumer ym., 2015).

WAR-tilaston konseptissa on poimittavissa monia tässäkin tutkimuksessa esille tuotuja urheiluanalytiikan piirteitä. Lähtökohtana on uuden mittaamis-
muodon määrittäminen uniikin informaation saamiseksi, ja sitä kautta pelaajan todellisen arvon määrittämiseksi. Mittaamisessa käytetään erilaisia metodeja kokonaisvaltaisen tuloksen saamiseksi. Mitatut tulokset pelaajan suoritukselta suhteutetaan voitettuihin otteluihin, eli tavoitteellisuus tulokseen kulkee mukana.

4.3 Tiivistelmä urheiluanalytiikasta pohjoisamerikkalaisessa ammattilaisjoukkueurheilukulttuurissa

Pohjoisamerikkalaisessa ammattilaistason joukkueurheilussa urheiluorganisaatioissa on useita eri päätöksentekijöitä, kuten joukkueen omistaja, manageri (*eng. general manager*), pää- ja apuvalmentajat sekä pelaajatarkkailun koordinaattorit. Urheiluanalytiikka tarjoaa mahdollisuudet parempaan dataan, parempiin työkaluihin ja parempiin prosesseihin, jotka auttavat tehokkaaseen miljardiyritysten päätöksentekoon ja miljonääriyöntekijöiden hallintaan (Alamar & Mehrotra, 2012). Kuten luvussa 3 on todettu, isoissa organisaatioissa tiedon keskittäminen ja kommunikaatio analytikkojen ja päätöksentekijöiden välillä on keskeistä. Etenkin pelaajien tarkkailusta syntyy tietoa, joka voi sijaita useissa eri sijainneissa. Urheiluanalytiikassa merkityksellistä on, kuinka saadaan tärkeä, relevantti, tieto ja oikeat analyttiset mallit niiden henkilöiden käsiin, jotka ovat oikeassa asemassa saadakseen siitä hyötyä irti (Alamar & Mehrotra, 2011a).

Michael Lewisin teos "Moneyball" (2003) herätti kiinnostuksen urheiluanalytiikkaa kohtaan, niin julkisella, kuin yksityisellä tasolla. Teoksessa Oakland Athleticsin general manager Billy Beane nähdään urheiluanalytiikan innovaattorina. Teos on kuitenkin saanut kritiikkiä ja esimerkiksi Baumer ja Zimbalist (2014) väittävät, ettei Billy Beane toiminut urheiluanalytiikan innovaattorina, kuitenkin samalla he tunnustavat hänen merkityksensä kehittyneiden analyytisten tilastojen käytön leviämässä ja laajentumisessa. 2000-luvun alun Oakland Athleticsin ohella Chazan-Pantzalis (2020) nostaa 2015 New York Metsin menestyksekkään urheiluanalytiikan implementoinnin, heidän päästessään World Series -loppuotteluun huolimatta siitä, että olivat keventäneen palkkarakennettaan 200 miljoonan USD verran kahdessa vuodessa.

Viitaten organisaatioiden kokoon ja henkilöstön määrään, myös Paul DePodesta on todennut, että osa heidän käyttämistään mittauksista Oakland Athleticsissa 2000-luvun alussa syntyi täysin analytikoiden ja pelaajatarkkailijoiden yhteisistä keskusteluista (Baumer & Zimbalist, 2014). Toisin kuten "Moneyball" -elokuvassa oli esitetty ja dramatisoitu totuutta draaman tuottamiseksi niin sanotun uuden koulukunnan ja vanhan koulukunnan välille.

Pohjoisamerikkalaisessa kulttuurissa baseball on toiminut analytiikan edelläkävijänä (Miller, 2015). Baseball on joukkuelaji, joka koostuu useista yksilökohtaisista, näkyvimpänä lyöjä vastaan syöttäjä (Szymanski, 2020). Yksittäisten tilanteiden ja suoritusten helppo eroteltavuus on johtanut siihen, että analytiikkaa on helpompi käyttöönottaa ja on helpompi kehittää uusia tilastollisia malleja (Elitzur, 2020). Erotteleva tilastointi auttaa tunnistamaan paremmin pelaajien kyvykkyyksiä (Baumer ym., 2015). Yksilön eri kyvykkyyksiä kokoava ja kyvykkyyksien merkitystä joukkueen voittoihin kertova tilasto on WAR, eli wins-above-replacement. WAR tilaston konsepti on ollut yksi urheiluanalytiikan ja sabermetricsin suuria saavutuksia (Baumer ym., 2015).

5 POHDINTAA

Datan määrä kasvaa maailmassa valtavalla nopeudella. Perinteiset liiketoimintayritykset pyrkivät käyttämään dataa ja informaatiota kilpailuedun saavuttamiseen, eikä urheilu liiketoiminnan alalajina poikkea tästä. Kilpailullisen edun saavuttamisen väline on käsitteenä keskeinen, koska huippu-urheilussa kilpailu on äärimmäisen tiukkaa. Ymmärretään, että data itsessään ei tee autuaaksi (Miller, 2015), mutta oikein käytettynä se voi tuoda juuri sen ratkaisevan edun voiton tavoittelussa (Szymanski, 2020; Miller, 2015; Alamar, 2013). Ensin tulee kuitenkin ymmärtää, kuinka datan kanssa tehdään töitä, mistä data syntyy ja kuinka se prosessoidaan data-analyysiin (Miller, 2015).

Urheiluanalytiikka toimii kilpailullisen edun saavuttamisen välineenä. Data-analyysin urheiluanalytiikassakin voi nähdä pohjautuvat Runklerin (2016) data-analyysiprosessin malliin, jossa on valmisteluvaihe, esiprosessointivaihe, analyysivaihe ja jälkiprosessointivaihe. Urheilussa tiedon arvo ja määrä kasvavat ja tietoa syntyy erilaisissa muodoissa. Tähän ovat johtaneet tilasensorit sekä teknologiset kehitykset historiallisen datan hyödyntämisessä. Määriteltäessä dataa kyseisten ominaisuuksien kautta voidaan puhua Alsaig ym. (2018) määritelmän mukaan urheilussakin hyödynnettävän massadataa.

Davenport (2014) jakaa urheiluanalytiikan joukkueen suoritusta mittaavaan analytiikkaan, urheiluorganisaation liiketoimintaa mittaavaan analytiikkaan sekä yksilön harjoittelua ja kuntoutumista edistävään analytiikkaan. Näistä tarkasteltaessa urheiluanalytiikkaa joukkueen suorituksen mittarina Alamar (2013) näkee kaksi maalia: ajan säästäminen ja uniikin informaation tuottaminen.

Tietojärjestelmien kehitys on toiminut tärkeänä teknologisenä kehityksenä ajan säästämisen tavoitteen näkökulmasta. Tiedon keskittäminen ja erilaiset ulkoiset palvelut mahdollistavat otteluvideoiden nopeamman koonnin, mikä säästää valmennukselta aikaa. Tämä säästynyt aika pystytään käyttämään esimerkiksi pelaajien kanssa henkilökohtaiseen keskusteluun tai kokonaisuuden hallintaan.

Uniikkia informaatiota pystytään synnyttämään erilaisilla sensoreilla tai erilaisilla malleilla. Sensorit pystyvät mittaamaan reaaliaikaisesti niin pelaajan

kuin pelivälineen sijaintia ja tuomaan siitä tietoa analyytikolla, joka johtaa tiedon eteenpäin päätöksentekijälle. Kehittyneillä numeerisilla tilastoilla urheilijan kokonaissuoritusta pystytään arvioimaan paremmin (Baumer ym., 2015).

Toisaalta näen uniikin informaation olevan kontekstisidonnaista sen suhteen, millä mittareilla päätöksentekijä haluaa, että hallussa olevaa dataa mitataan. Kuten Szymanski (2020) toteaa, kaikki tieto on jo olemassa kentällä, mutta vielä ei ole tarvittavaa ymmärrystä tiedon irti saamiseksi. Kun samaa peliä lähtee mittaamaan uudesta oivaltavasta näkökulmasta voi saada uniikin informaation, jota vastustajalla ei ole.

Esimerkiksi urheiluanalytikko Warren Sharp (2020) The Ringer NFL Show -podcastissa toteaa, että virallisten tilastojen käyttämisen sijaan hän pyrkii uusilla tilastoilla saamaan paremman käsityksen ottelun luonteesta ja hyödyntämään löydettyä tietoa omassa toiminnassaan. Sharp (2020) toteaa, että hän ei käytä tavallisia virallisilta kanavilta saatavia tilastoja, kuten YPA (*yards-per-attempt*). Sen sijaan hän käyttää uusia kehittyneempiä tilastoja, kuten EDSR (*early-down-success-rate*) onnistumisen todennäköisyyden määrittämisessä. YPA määrittelee amerikkalaisessa jalkapallossa saavutetut jaardit yritystä kohden, kun EDSR määrittelee tehokkuuden ensimmäisillä ja toisilla yrityksillä, mikä sulkee pois mahdollisen epävalidin tilaston.

Historiadan hyödyntäminen ja sen koostaminen mahdollistavat erilaiset koneoppimisen ja tiedonlouhinnan sovellutukset, joilla pystytään ennakoimaan tulevaa, kuten pelaajan todennäköistä suorituskykyä tai joukkueen todennäköistä menestystä sarjassa. Reaaliaikainen informaatio, millä todennäköisyydellä tapahtuu jotain, jos valmentaja tekee tietyn tyyllisen ratkaisun, voi olla ohjaamassa ja vaikuttamassa päätöksentekoon kesken ottelun. Kun todennäköisyyksien valossa pystytään ennustamaan tulevaa, osataan tehdä informaation pohjautuvia ratkaisuja sulkien vaihtoehtoista pois ne, jotka odotusarvollisesti heikentävät oman joukkueen todennäköisyyttä voittoa. Vastaavasti historiadataa pystytään hyödyntämään vastustajan näkökulmasta, kun osataan ennakoida millä todennäköisyydellä vastustajan valmentaja tulee tekemään pelin sisällä minkäkin ratkaisun.

On tärkeää huomioida, että data itse ei puhu ja hyödylliset ennusteet eivät nouse tyhjästä (Miller, 2015). Tämä tarkoittaa analyttisten osaajien, kuten valmentajien, pelillisten osaajien yhteistyötä. Valmennus pystyy omalla kokemuksellaan ja näkemyksellään vaikuttamaan juuri edellä mainittujen uusien oivaltavien mittareiden syntymiseen. Edellisestä voidaan Alamarin (2013) tavoin johtaa, että yksi urheiluanalytiikan arvon muodoista on saada oikeat henkilöt kysymään tarkempia kysymyksiä. Mikäli organisaation kulttuuri on sellainen, jossa urheilujohtajat eivät koe analytiikkaa sijoituksena kilpailullisen edun saavuttamiseksi, vaikuttaa se negatiivisesti urheiluanalytiikan evoluution nopeuteen (Alamar & Mehrotra, 2012).

Pohdittaessa suomalaisen pesäpallon tämänhetkistä urheiluanalyttista tilaa sekä urheiluanalyttisia mahdollisuuksia, tulee ensimmäisenä todeta tietty salamyhkäisyys koskien urheiluanalytiikkaa. Kuten Daryl Morey on todennut, kilpailullinen etu syntyy omista järjestelmistä ja datasta, jota kellään muulla ei

ole. On siis huomioitava mahdollisuus, että on olemassa urheiluanalyttisiä oivalluksia ja teknologioita, joita ei ole julkisella tutkimuksella tiedossa. Tämä tulee huomioida edellä, että pesäpallossa voi olla joukkueilla käytössä erilaisia analytiikan malleja, jotka eivät julkisesti ole tiedossa.

Kuten Davenport (2014) on todennut, video on urheilussa datan keskeisin lähde. Lähtökohtaisesti se, että videot ovat pääsarjassa kaikkien saatavilla on pesäpalloliitolta ensimmäinen merkittävä tuki urheiluanalytiikan suhteen. Kilpailusarjan tuen on myös Davenport (2014) nostanut esille.

Otteluvideon nopeampaan käsittelyyn on tarjolla yksityisiä palveluntarjoajia. Ennen palvelun tarjosi TJP Sports Consulting Oy ja nykyisin ulkoisena palveluna toimii E-tesis. Urheiluanalytiikan arvon määrittelyssä ajan säästäminen on nostettu keskeiseksi arvoksi, ja e-tesis sovelluksena on erinomainen tähän tarkoitukseen. Pelaajien ja tilanteiden suodattaminen säästää aikaa niin valmennukselta kuin pelaajilta valmistautuessa otteluun, ja tämä aika voidaan käyttää esimerkiksi muuhun harjoitteluun ja palautumiseen tai edelleen vielä tarkempaan valmistautumiseen.

E-tesisistä uniikin informaation tuottajana lähdän aiempaan tutkimuskirjallisuuteen pohjaten haastamaan. E-tesis ei konseptina ole mitenkään salainen, ja tällä perusteella Elitzurin (2020) ja Davenportin (2014) mukaan kilpailullista etua ei synny suhteessa muihin kyseistä palvelua käyttäviin joukkueisiin. Toki ero syntyy joukkueisiin, joilla ei palvelua ole ja täten osa joukkueista, joilla kyseistä palvelua ei ole antaa lähtökohtaisesti jo kilpailullista etua vastustajilleen. Edellisen näen myös uniikin informaation kontekstisidonnaisuudella.

Kuten edellä totean uniikin informaation olevan kontekstisidonnaista, niin tässäkin tulee huomioida, että osa varmasti saa samasta sovelluksesta paremman kilpailuedun mitä toiset. Tätä vahvistaen Alamar (2013) totesi olevan organisaatioita, joilla ei ole ymmärrystä analyttisistä mahdollisuuksista, ja käytön tulos on sen mukaista. Ajan säästämisen näkökulmasta E-tesis alustana pystyisi enemmän tuottamaan numeerista tietoa, jolloin voidaan rakentaa tilastollisia malleja, joita viralliset tilastopalvelut eivät tarjoa. Mielenkiintoisena esimerkkinä tästä olisi esimerkiksi tunnistaa pelaaja, joka korkealla prosentilla purkaa palon jälkeistä ykköstilannetta ensimmäisellä lyönnillä ja tätä pelaajaa pystyy valmennus käyttämään jokerina haastavan tilanteen purkamiseen jatkuvuuden saamiseksi. Toki tämä vaatii palveluntarjoajalta resursseja joko työvoimaan tai laskentatehoon, että kyseisiä numeroita pystytään tuottamaan henkilövoimin tai automaattisesti otteluvideolta.

Uniikin informaation hyödyntämisenä näen esimerkiksi Jyväskylän Kiritärien kaudella 2020 käyttämät lyöntikartat. Kyseistä palvelua ei tarjottu virallisten tilastojien puolelta, eikä ulkoisilta palveluntarjoajilta. Näin ollen omilla ohjelmistoilla rakennetusta uniikista informaatiosta muotoutui urheiluanalytiikan kautta kilpailullisen edun väline. Toki edellisessä tulee tunnistaa, että meillä ei ole 100 % tarkkuutta siitä, minkälaisia ohjelmistoja muut joukkueet ovat käyttäneet ja siten väite muodostuiko kilpailuetua, voidaan kyseenalaistaa.

Tutkimuksen tavoitteena oli vertailla analytiikkaa pesäpallon saralla myös pohjoisamerikkalaiseen baseballiin. Lähtökohtaisesti, kun verrataan suomalais-

ta urheilua pohjoisamerikkalaiseen urheiluun, tulee tunnistaa pienemmät markkinat, pienemmät pelaajamassat, pienemmät yritysten taloudelliset arvot, sekä pienemmät henkilöstömäärät.

Pesäpallo ja baseball sopivat hyvin yhteen molempien ollessa yksilökohtaamisista koostuva joukkuepeli. Tämä mahdollistaa data-analytiikan helpomman käyttöönoton (Elitzur, 2020). Baseballiin nähden voidaan verrata, että pesäpallo enemmän koostuu kolmesta yksilökohtaamisesta lyöjä vastaan syöttäjä vastaan etenijä. Kuitenkin peli koostuu erilaisista yksittäisistä tilanteista ilman dynaamista pelikelloa, joten suoritukset ja suorituksen tekijät pystytään helposti erottelemaan ja pystytään toteuttamaan erottelevaa tilastointia pelaajien kyvykkyyksien tunnistamiseen (Baumer ym., 2015).

Erottelevaa tilastointia tapahtuu liiton puolesta, kun erilaisia tilanteita tilastoidaan eri kategorioihin. Kyvyt ja mahdollisuudet olisivat kuitenkin paljon laajempaan tilastointiin, kun esimerkiksi lukkareita ja muita ulkopelaajia ei virallisen tahon puolesta tilastoida ollenkaan. Lukkarien tilastointi tuo mielenkiintoisen eron, kun baseballissa syöttäjistä tullutta dataa on tilastoitu jo vuodesta 2006 alkaen liikedatan muodossa (Alamar & Mehrotra, 2011b). Baseballissa tällä hetkellä liikkeentunnistuksella tunnistetaan mihin tulleista syötöistä lyöjä saa parhaan tuloksen, ja on äärimmäisen mielenkiintoista tutkia, kun vastaavaa teknologiaa saadaan suomalaiseseen pesäpalloon, mistä lautasen kohdasta pääkotiuttajat lyövät eniten juoksuja, ja mistä vastaavasti vähiten.

Billy Beanen perustelemaa erottelevaa tilastointia esimerkiksi OBP:n ja SLG:n suhteen voidaan jalostaa pesäpalloon. Sen sijaan, että tutkitaan pelkääntään esimerkiksi ykköstilanteen purkamista kärjen vientinä, pystytään tilastoa laajentamaan EDSR mallisesti, montako kertaa onnistuminen tapahtuu ensimmäisellä tai toisella lyönnillä, eikä viimeisellä milloin takapalon riski kasvaa. Mallista voidaan johtaa joukkueita, jotka purkavat esimerkiksi palon jälkeisen tilanteen kovalla prosentilla ja joukkueita, jotka huonolla prosentilla osaavat puolustaa palon jälkeistä tilannetta. Lähtökohtia uuteen tilastointiin on monia, mitkä pohjautuvat yhteistyöhön ja oivallukseen analyytikon ja päätöksentekijän välillä (Alamar, 2013).

Pohjoisamerikkalaiseen baseballiin kuuluu muitakin mittausajatuksia, joita tulisi suomalaisessa pesäpallossa tutkia tarkemmin. Baumer ja Zimbalist (2014) toteavat, että pelaajan tulee varastaa pesänvälejä vain, jos onnistumisprosentti on vähintään 65, pienemmän prosentin heikentäessä juoksun lyömisen todennäköisyyttä kyseisen sisävuoron aikana. Vastaavaa varastamista tapahtuu pesäpallossa ja olisi mielenkiintoista dataan pohjaten nähdä sen hyöty, vai heikentääkö se keskimääräisesti joukkueen tuloksentekeä. Tässäkin kyse on odotusarvosta, millä todennäköisyydellä ratkaisut edesauttavat joukkueen pääsyä voittoon ja tällä tiedolla pyritään karsimaan ne ratkaisut, jotka vievät joukkueita kauemmas voitosta.

WAR-tilaston konsepti on ollut yksi urheiluanalytiikan ja sabermetricsin suuria saavutuksia (Baumer ym., 2015). Vastaavan tilaston soveltamisen näen haastavana suomalaisen pesäpallon konseptissa. Runkosarjan ottelumäärä on niin pieni, jopa viisi kertaa pienempi, että todennäköisesti lopputulokset pelaajien

jien voiton merkityksessä keskenään muodostuvat kymmenyksiin. Lisäksi etenemisnopeuden merkitys pesäpallossa on valtava, samalla kun toisaalta pelataan juoksunlyöntipeliä. Mallin rakentaminen ja näiden kahden kyvykkyyden arvottaminen on haasteellista.

Elitzur (2020) toteaa tutkimuksessaan, että joukkueen pelaajien yhteen laskettu TeamWAR korreloi vahvasti joukkueen voittoprosentin kanssa. Jos kuvitteellisesta WAR-tilastosta rakennettaisiin paras mahdollinen 12-pelaajan joukkue pesäpallossa, koostuisiko se 12 lyöjästä, 12 etenijästä vai missä suhteessa se tulisi muodostumaan? Lyöntivoima ja etenemisvoima kuitenkin ovat ne, jotka muovaavat ulkokentän koostumuksen. Toinen on osittain pelattavissa pois, mutta molemmat yhdessä ovat haastava yhdistelmä.

Pohjoisamerikkalaiseen urheilukulttuuriin kuuluvat ominaispiirteet voidaan sulkea pois suomalaisesta kontekstista. Varaustilaisuutta ei ole ja näin olen analytiikkaa varaustilaisuutta varten ei ole olemassa. Nuorien pelaajien (*eng. prospects*) seuraamista tapahtuu toki suomalaisessa kulttuurissakin ja tietystä määrin olisi mielenkiintoista nähdä, miten tilastot juniorien superpesiksessä korreloivat menestykseen aikuistason superpesiksessä. Tietenkin tämä vaatii laajempien tilastojen tekemistä.

Vapaiden pelaajien hankkimista puolestaan tapahtuu Suomessa pesäpallon kontekstissa. Pelaajamassat ovat pienempiä ja joissain tilanteissa uskallan jopa argumentoida, että pelaajahankintaa määrittelee enemmän pelaajan pelipaikka, kuin henkilökohtainen menestys. Käytännössä kehittyneillä analytiikoilla pystyttäisiin rakentamaan ennakoivia malleja, joilla muuttaa historiadata informaatioksi (Alamar & Mehrotra, 2011b). Vapaiden pelaajien suoritusta pystytään kehittyneillä analytiikoilla ennakoimaan ja pyrkiä siten käyttämään pelaajahankintoihin varattu budjetti oikein.

Alamar (2013) nimeää urheiluorganisaatioiden analytiikan käyttöönoton esteiksi teknologiset haasteet sekä johdon haasteet. Suomalaisessa kontekstissa teknologiset haasteet ovat varmasti ongelma. Davenport (2014) argumentoi, että kustannukset teknologiaan ovat este pohjoisamerikkalaisessakin kulttuurissa, ja tekijöitä on vaikea saada, koska organisaatioissa pidetään ”vain” sataa työntekijää. Tämä suhteutettuna suomalaisiin urheiluorganisaatioihin kokoluokkiin on ymmärrettävää, että urheiluanalyttinen kehitys on hitaampaa.

Vaikka teknologiset haasteet ovat kiistattomia, organisatorisella tasolla näen mahdollisuuksia edistää urheiluanalyttistä kehitystä myös Suomessa. Tekijöiden puute on haaste (Alamar & Mehrotra, 2011b), varsinkin kun pienestä massasta joudutaan löytämään samassa paketissa niin analytiikan osajia, kuin riittävää kohdelajin tuntemusta ollakseen oikeasti uhka vastustajan näkökulmasta.

Pesäpallo on maakuntien peli ja perinteet ovat voimakkaat. Riskinä on, että valmentajan henkilökohtainen kokemus ja tunne määrittelevät lajia enemmän, kuin datasta tuleva informaatio, varsinkin kun informaatio tulee sellaisilta henkilöiltä, jotka eivät ole kokeneet samaa lajin ytimeä kuin valmentaja (Alamar & Mehrotra, 2012). Mikäli analytiikkaa ei koeta sijoituksena, eikä siihen

uskota, ei analytiikan evoluution kehittyminen edisty (Alamar & Mehrotra, 2012).

Kuitenkin, jos ollaan valmiita rakentamaan yhteinen kieli urheilun ja urheiluanalytiikan välillä, on se mahdollista myös pienemmän budjetin organisaatioissa. Davenportin (2014) mukaan vaihtoehtoja tässä ovat avun hakeminen analyttisten palveluiden tarjoajilta tai omien intohimoisten kannattajien kartoittaminen ja selvittäminen, onko siellä apuja auttamaan analyysien tekemisessä olemassa olevasta datasta tai yksinkertaisten uusien tilastoiden kirjaamisessa. Ulkoisten palveluntarjoajien kanssa voi miettiä ratkaisuja sponsorointiyhteistyön kautta (Davenport, 2014).

Käytyäni keskusteluja eri päättävissä elimissä toimivien tahojen kanssa, on oletettavaa, että liikkeenmonitorointitekniikka pesäpallon kontekstissa ei ole muutaman vuoden säteellä realistinen. Tästä syystä urheiluanalyttistä kehitystä pesäpallon saralla tulisi edistää niin ajan säästämisen, kuin uusien mallien saralla. Olemassa olevista tilastointimenetelmistä on löydettävissä malleja, miten mitata suoritusta uudesta näkökulmasta. Muun muassa ulkopelaaminen kokee murrosta ja olisi mielenkiintoista nähdä tilastointia, miten passiivinen ja aktiivinen ulkopelaaminen näkyvät juoksujen lyömisessä ja tuleeko juoksua enemmän mahdollisesti suoritusvirheistä vai sijoittumisvirheistä kuvioita vaihdettaessa.

Pesäpallon koostuessa yksilökohtaisista baseballin tavoin on helppo todeta, että urheiluanalytiikka voi pesäpallossakin toimia kilpailullisen edun saavuttamisen välineenä tiukassa kilpailussa. Keskeisiä keinoja on useampia. Ensimmäinen on valmentajan ajan säästäminen kokonaisuuden hallintaan, otannan kasvattamiseen ja pelaajien kehittämiseen. Ajan säästämiseen on jo olemassa erinomaisia ulkoisen palveluntarjoajan tarjoamia ohjelmistoja. Toinen keskeinen keino on uniikin informaation tuottaminen erilaisilla teknologioilla ja oivaltavilla ajatusmalleilla tarkastella perinteisiä tilastoja. Kuitenkin kuten tutkimuksessa on todettu, urheiluanalyttinen evoluutio vaatii yhteistyötä ja oikeaa asennetta niin päätöksentekijöiltä, kuin urheiluanalyttiseltä väeltä.

Teknologinen kehitys ja kustannukset huomioiden oma näkemykseni on se, että erilaisten liikkeenseurantateknologioiden tai historiadataa hyödyntävien ennakoivien mallien käyttöönotto ei ole vielä lähivuosina tehtävälle omalle pro gradu -tutkielmalle tarpeeksi käsillä olevia aiheita. Jatkotutkimusmahdollisuuksien voisi ajatella koskevan urheiluanalytiikan implementointia organisaatioon ja sen peilaamista organisaation tulokseen, tai uuden oivaltavan tilaston käyttöönoton seuraamista pelaajien kyvykkyyden määrittämiseen.

LÄHTEET

- Alamar, B. C. (2013). *Sports analytics: A guide for coaches, managers, and other decision makers*. Columbia University Press.
- Alamar, B., & Mehrotra, V. (2011a). Beyond 'Moneyball': The rapidly evolving world of sports analytics, Part I. *Analytics Magazine*.
- Alamar, B., & Mehrotra, V. (2011b). Sports analytics, part 2: The role of predictive analytics, organizational structures and information systems in professional sports. *Analytics Magazine*.
- Alamar, B. C. & Mehrotra, V. (2012). Analytics & Sports, part III: Improving resource allocation with portfolio decision analysis. *Analytics Magazine*.
- Alsaig, A., Alagar, V., & Ormandjieva, O. (2018, August). A critical analysis of the V-model of big data. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 1809-1813). IEEE.
- Balreira, E. C., Miceli, B. K., & Tegtmeyer, T. (2014). An Oracle method to predict NFL games. *Journal of Quantitative Analysis in Sports*, 10(2), 183-196.
- Baumer, B. S., Jensen, S. T., & Matthews, G. J. (2015). openWAR: An open source system for evaluating overall player performance in major league baseball. *Journal of Quantitative Analysis in Sports*, 11(2), 69-84.
- Baumer, B., & Zimbalist, A. (2014). *The sabermetric revolution: Assessing the growth of analytics in baseball*. University of Pennsylvania Press.
- Bovey, W. H., & Hede, A. (2001). Resistance to organizational change: the role of cognitive and affective processes. *Leadership & Organization development journal*.
- Cambridge Dictionary. (08.11.2020). ANALYSIS | meaning in the Cambridge English Dictionary. Haettu osoitteesta <https://dictionary.cambridge.org/dictionary/english/analysis>
- Cambridge Dictionary. (08.11.2020). ANALYTICS | meaning in the Cambridge English Dictionary. Haettu osoitteesta <https://dictionary.cambridge.org/dictionary/english/analytics>
- Cao, L. (2017). Data science: a comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3), 1-42.

- Chazan-Pantzalis, V. (2020). Sports Analytics Algorithms for Performance Prediction.
- Chen, J., Jiang, Q., Wang, Y., & Tang, J. (2016, March). Study of data analysis model based on big data technology. In *2016 IEEE International Conference on Big Data Analysis (ICBDA)* (pp. 1-6). IEEE.
- Davenport, T. H. (2014). Analytics in sports: The new science of winning. *International Institute for Analytics*, 2, 1-28.
- Devlin, S., & Treloar, T. (2018). A network diffusion ranking family that includes the methods of Markov, Massey, and Colley. *Journal of Quantitative Analysis in Sports*, 14(3), 91-101.
- Donoho, D. (2015, September). 50 years of Data Science. In *Princeton NJ, Tukey Centennial Workshop* (pp. 1-41).
- Elitzur, R. (2020). Data analytics effects in major league baseball. *Omega*, 90, 102001.
- ESPN. (27.11.2020). MLB Wins Above Replacement – WAR – Major League Baseball – ESPN. Haettu osoitteesta http://www.espn.com/mlb/war/leaders/_/type/seasonal/year/2019
- Glickman, M. E., & Stern, H. S. (2017). Estimating team strength in the NFL. *Handbook of Statistical Methods and Analyses in Sports*, 113-136.
- Harrison, C. K. & Bukstein, S. (2017). Sport Business Analytics: Using Data to Increase Revenue and Improve Operational Efficiency. CRC Press, Taylor & Francis Group.
- Intezari, A., Pauleen, D. J., & Taskin, N. (2016, January). The DIKW hierarchy and management decision-making. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 4193-4201). IEEE.
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, 34(3), 387-394.
- Major League Baseball. (20.11.2020). What is a Competitive Balance Tax? Haettu osoitteesta <http://m.mlb.com/glossary/transactions/competitive-balance-tax>
- Miller, T. W. (2015). Sports analytics and data science: winning the game with methods and models. FT Press.
- Moskowitz, D. (20.11.2020). 10 Most Valuable NFL Teams 2019: Cowboys Lead the Pack. Haettu osoitteesta

<https://www.investopedia.com/articles/personal-finance/022315/5-mostvaluable-nfl-franchises.asp>

Nguyen, T. L. (2018, December). A framework for five big v's of big data and organizational culture in firms. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 5411-5413). IEEE.

Passfield, L., & Hopker, J. G. (2017). A mine of information: can sports analytics provide wisdom from your data?. *International journal of sports physiology and performance*, 12(7), 851-855.

Runkler, T. A. (2016). Data Preprocessing. In *Data Analytics* (pp. 23-36). Springer Vieweg, Wiesbaden.

Sanderson, J., & Baerg, A. (2020). Youth baseball and data analytics: Quantifying risk management and producing neoliberal responsible citizenship through the GameChanger app. *Communication & Sport*, 8(1), 72-91.

Shamsuddin, S. M., & Hasan, S. (2015, October). Data science vs big data@ UTM big data centre. In *2015 International Conference on Science in Information Technology (ICSITech)* (pp. 1-4). IEEE.

Sharp Football Analysis. (26.11.2020). Sharp Football: Anlysis & NFL Betting Tips from Warren Sharp. Haettu osoitteesta <https://www.sharpfootballanalysis.com/>

Szymanski, S. (2020). Sport Analytics: Science or Alchemy?. *Kinesiology Review*, 9(1), 57-63.

Tiihonen, A. (2015). Liikuntakulttuurin käsitteitä tiedetoimittajille. Haettu osoitteesta https://www.miksiliikun.fi/wp-content/uploads/2012/08/Liikuntakulttuurin_k%C3%A4sitteit%C3%A4_Tiihonen2015.pdf

The Ringer NFL Show. (25.11.2020). *Favorites on Thanksgiving and Sunday Matchups | The Ringer NFL Show* [Podcast] Haettu 25.11.2020 osoitteesta <https://open.spotify.com/episode/1iXTY9AiZAsdHZNLmrHRqZ>

Valero, C. S. (2016). Predicting Win-Loss outcomes in MLB regular season games–A comparative study using data mining methods. *International Journal of Computer Science in Sport*, 15(2), 91-112.