

Tuukka Salminen

**DATA-ANALYTIIKAN HYÖDYNTÄMISEN
MAHDOLLISUUDET JALKAPALLON TAKTISESSA
ANALYYSISSÄ**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2020

TIIVISTELMÄ

Salminen, Tuukka

Data-analytiikan hyödyntäminen jalkapallon taktisessa analyysissä

Jyväskylä: Jyväskylän yliopisto, 2020, 26 s.

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja: Marttiin, Pentti

Tämä kandidaatintutkielma tutkii data-analytiikan hyödyntämisen mahdollisuuksia jalkapallon taktisessa analyysissä. Jalkapallo-otteluista kerätyn datan määrä on kasvanut suureksi. Kuten monissa muissa urheilulajeissa, myös jalkapallossa on alettu omaksumaan data-analytiikan käyttöä. Tämä tarjoaa uusia mahdollisuuksia taktiseen analyysiin. Tutkielma pyrkii vastaamaan siihen, miten data-analytiikkaa on mahdollista hyödyntää jalkapallon taktisessa analyysissä. Tutkielma toteutettiin kirjallisuuskatsauksena. Tulokseksi saatiin, että soveltamalla analytiikkaa video- ja paikkadataan, voidaan saavuttaa kilpailuetua tarkemman ja tehokkaamman taktisen analyysin muodossa.

Asiasanat: data-analytiikka, massadata, urheiluanalytiikka, jalkapallo, taktinen analyysi

ABSTRACT

Salminen, Tuukka

Utilizing data analytics in tactical analysis of football

Jyväskylä: University of Jyväskylä, 2020, 26 pp.

Information Systems, Bachelor's thesis

Supervisor: Marttiin, Pentti

This bachelor's thesis aims to discover ways in which data analytics can be utilized in tactical analysis of football. The amount of data collected from football matches has grown to be large. Among with many other sports, football has also started to adapt the use of data analytics. This provides new opportunities to tactical analysis. This thesis aims to answer how data analytics could be utilized in tactical analysis of football. The thesis has been carried out as a literature review. As a result of the study was found that by applying analytics to video data and positional data, competitive edge can be found via more precise and efficient tactical analysis.

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

KUVIOT

KUVIO 1 Massadatateknologiapino jalkapallon taktiseen analyysiin (Rein & Memmert, 2016) 20

TAULUKOT

TAULUKKO 1 Data-analytiikan termejä (Cao, 2014) 9

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT

TAULUKOT

1	JOHDANTO.....	6
2	DATAN HYÖDYNTÄMINEN.....	8
	2.1 Data-analytiikka.....	8
	2.2 Massadata	9
3	URHEILUANALYTIikka.....	12
	3.1 Urheiluanalytiikan määritelmä ja komponentit.....	12
	3.1.1 Datan hallinta.....	13
	3.1.2 Analyttiset mallit.....	13
	3.1.3 Tietojärjestelmät	14
	3.2 Pelaajien ja otteluiden suoritusten analytiikka.....	14
4	DATA-ANALYTIIKAN HYÖDYNTÄMINEN JALKAPALLON TAKTISESSA ANALYYSISSÄ.....	16
	4.1 Taktiikat	16
	4.2 Perinteinen taktinen analyysi ja siihen liittyvät haasteet.....	17
	4.3 Data-analytiikka jalkapallon taktisessa analyysissä	17
	4.3.1 Videodata	18
	4.3.2 Paikkadata	18
	4.3.3 Massadata.....	19
5	JOHTOPÄÄTÖKSET	21

1 JOHDANTO

Data-analytiikan käyttö on viime vuosina yleistynyt monilla toimialoilla. Millään muulla toimialalla ei ole kuitenkaan samankaltaista analyttistä oma-aloitteisuutta kuin huippu-urheilussa. Esimerkiksi yhdestä jalkapallo-ottelusta voidaan kerätä pallon ja pelaajien seurannan avulla noin 3,1 miljoonaa datapistettä. Datan määrän kasvu luo myös mahdollisuuksia jalkapallon taktiseen analyysiin ja analytiikan avulla joukkueiden taktisesta käytöksestä voidaan saada huomattavasti tarkempaa informaatiota kymmeniä kertoja perinteisiä menetelmiä nopeammin.

Tutkielman tutkimuskysymys on ”Kuinka data-analytiikkaa voidaan hyödyntää jalkapallon taktisessa analyysissä?”. Jotta tutkimuskysymykseen voitaisiin vastata, tulee ymmärtää aihealueen käsitteistöä. Ennen siirtymistä spesifisemmin data-analytiikan käyttöön jalkapallon kontekstissa, määritellään tutkielmassa käsitteet data-analytiikka, massadata ja urheiluanalytiikka tieteelliseen kirjallisuuteen perustuen.

Data-analytiikalla tarkoitetaan yleisesti isoja datajoukkoja hyödyntäväksi analytiikaksi, jota käytetään päätöksenteon tukena (Runkler, 2016). **Massadata** on valtavia, monimuotoisia datamassoja, joita pystytään tallentamaan ja käsittelemään nopeasti (Chen, 2014). **Urheiluanalytiikka** taas on analyttisten mallien soveltamista dataan päätöksenteon tueksi ja kilpailuedun luomiseksi urheilun saralla (Alamar, 2013, s. 3).

Tutkielmaan kerättiin tietoa pääasiassa Google Scholarista ja ProQuestista. Pääasiallisina hakusanoina toimivat esimerkiksi ”data analytics”, ”big data” ja ”football tactical analysis”. Käsitteiden määrittelyyn ja tutkimuskysymykseen vastaamiseen käytetään yhteensä noin 35-40 lähdettä.

Tutkielma rakentuu siten, että johdantoluvun jälkeen toisessa luvussa määritellään tutkielman kannalta olennaiset yleisemmät käsitteet data-analytiikka ja massadata. Tämän jälkeen kolmannessa luvussa määritellään urheiluanalytiikka eli data-analytiikka urheilun kontekstissa. Neljännessä luvussa käsitellään data-analytiikan hyödyntämistä jalkapallon taktisessa analyysissä. Se sisältää katsauksen jalkapallon taktikoiden määrittelyyn sekä taktisen analyysin perinteisiin menetelmiin ennen siirtymistä video- ja paikkadatan analy-

tiikkaan sekä massadatan hyödyntämiseen jalkapallossa. Viidennessä eli viimeisessä kappaleessa esitetään tutkimuksen johtopäätökset sekä tehdään tutkielmasta yhteenveto. Tutkimuskysymykseen pyritään vastaamaan kahdessa viimeisessä kappaleessa.

2 DATAN HYÖDYNTÄMINEN

On arvioitu, että maailmassa oleva informaatio tuplaantuu noin 20 kuukauden välein (Frawley, Piatetsky-Shapiro, & Matheus, 1992) Koska informaatio on tarkoituksenmukaista dataa, jolla on konteksti, myös datan määrä kasvaa valtavasti. Tärkeitä datalähteitä ovat esimerkiksi liiketoiminta- ja teollisuusprosessit, tietokannat, kuvat ja videot sekä biomedikaalinen data (Runkler, 2016). Datamäärä on synnyttänyt valtavia taloudellisia mahdollisuuksia, mutta tuo mukanaan myös haasteita (Cao, 2017).

2.1 Data-analytiikka

Data-analytiikka on yläkäsite sen erilaisia asioita toteuttaville osa-alueille. Yleisesti sillä tarkoitetaan eri keinoin toteutettavaa analytiikkaa datalle, jota käytetään päätöksenteon tukena, esimerkiksi liiketoiminnan päätöksenteossa ja teollisuudessa tuotannon optimoinnissa.

Runkler (2016) määrittelee data-analytiikan seuraavasti: Data-analytiikka on tietokoneiden soveltamista isojen datajoukkojen analysoimiseksi ja käytettäväksi päätöksenteon apuna. Se on hyvin monitieteellinen ala, joka on omaksunut osia monilta muilta aloilta. Näitä aloja ovat esimerkiksi tilastotiede, koneoppiminen, hahmontunnistus (eng. pattern recognition), systeemiteoria, operaatiotutkimus sekä tekoäly.

Cao (2017) määrittelee data-analytiikan puolestaan teorioiksi, teknologioiksi, työkaluiksi ja prosesseiksi, jotka mahdollistavat syvällisen ymmärryksen datasta. Data-analytiikalla on myös erilaisia ominaisuuksia ja osa-alueita (taulukko 1). Data-analytiikan käyttötarkoituksen kannalta onkin olennaista, millaista analytiikkaa harjoitetaan. Esimerkiksi kuvaileva analytiikka kuvailee dataa eri statistiikkojen avulla ja tuottaa näin tietoa, kun taas ohjaileva analytiikka tuottaa suosituksia päätöksentekoon, ja ennakoiva analytiikka luo ennusteita ja tuo esiin syitä niiden takaa. Näistä viimeksi mainitussa on myös usein

taustalla kehittynyttä analytiikkaa eli massadataan sovellettavaa data-analytiikkaa.

Termi	Kuvaus
Kehittynyt analytiikka	Massadataan sovellettava data-analytiikka, samaa lopputulosta ei voida saavuttaa perinteisen data-analyysin avulla.
Kuvaileva analytiikka	Kuvaillaan dataa eri statistikkojen avulla, jotta siitä saataisiin tietoa.
Ennakoiva analytiikka	Ennakoidaan tulevia tapahtumia, tuodaan esille syitä ennusteiden takaa, usein kehittynyttä analytiikkaa.
Ohjaileva analytiikka	Optimoidaan saatua tietoa ja annetaan suosituksia päätöksentekoon.
Eksplisiittinen analytiikka	Kuvailevaa analytiikkaa toteuttava raporttoimalla, kuvailevalla analyysillä, varoittamalla ja ennustamalla.
Implisiittinen analytiikka	Syvää analytiikkaa toteuttava esim. ennakoivalla mallintamisella, optimoinnilla, ohjailevalla analytiikalla ja tuomalla hyödynnettävää tietoa.
Syvä analytiikka	Analytiikkaa, jonka avulla voidaan saada syvällistä ymmärrystä minkä takia ja kuinka asiat ovat tapahtuneet, tapahtuvat tai ovat tapahtumassa. Tätä ei voida saavuttaa kuvailevalla analytiikalla.

TAULUKKO 1 Data-analytiikan termejä (Cao, 2014)

2.2 Massadata

Chenin (2014) mukaan massadatalla (eng. big data) on valtaviin datamääriin lisäksi myös muita ominaisuuksia, jotka erottavat sen pelkästään valtavista datamassoista. Massadatalla ei ole yhtä yleisesti tunnustettua määritelmää, vaan eri tahoilla, kuten yrityksillä, tutkijoilla ja data-analytikoilla on eri mielipiteitä sen määritelmästä ja se määritellään usein sen piirteiden kautta (Chen, 2014; Baro, Degoul, Beuscart, Chazard, 2015; Romanillos, Eterna, de Kruijf, 2016). Yleisesti massadata tarkoittaa kuitenkin datajoukkoja, joita ei olisi mahdollista hahmottaa, hankkia, hallita ja prosessoida perinteisillä ohjelmistoilla ja laitteistoilla siedettävässä ajassa.

Apache Hadoop määritteli massadatan vuonna 2010 seuraavasti (Chen, 2014): Massadata on datajoukkoja, joita ei voida saada haltuun, hallita tai prosessoida yleisillä tietokoneilla hyväksyttävässä ikkunassa. Manyika ym. (2011) puolestaan määrittelivät massadatan edellä esitetyn määritelmän pohjalta ”datajoukoiksi, joita ei voitaisi hankkia, tallentaa, hallita ja analysoida perinteisillä tietokantaohjelmistoilla”. Heidän mukaansa massadata on innovaation, kilpailun ja produktiivisuuden seuraava eturintama. Massadatan määritelmään mahduttavien datajoukkojen kokoa ei myöskään määritellä tarkasti, vaan se on subjektiivista ja siihen liittyviä tekijöitä ovat esimerkiksi toimiala ja siellä yleensä käytettävien datajoukkojen koko sekä yleisesti saatavilla olevat ohjelmistot. Datajoukkojen koon oletetaan myös kasvavan teknologian kehittyessä. Kuitenkin, vaikka tarkkoja rajanvetoja koon suhteen ei ole, monilla toimialoilla massadataksi laskettavat datajoukot vaihtelevat joistain kymmenistä terabiteistä tuhansiin terabiteihin (Manyika ym. 2011). Koska massadataa ei voida määritellä pelkästään datajoukkojen koon perusteella, käytetään sen ominaisuuksien kuvaamiseen erilaisia malleja. Yleisin tähän tarkoitukseen tieteellisessä kirjallisuudessa käytetty malli on kolmen V:n malli (Laney, 2001; Meijer, 2011; Zikopoulos ym. 2011; McAfee & Brynjolfsson, 2012; Chen, 2014; Aggarwal, Bhatnagar, & Mishra, 2017), joka on alun perin kehitetty kuvaamaan datamäärien kasvun mahdollisuuksia ja haasteita.

Ensimmäinen ominaisuus kolmen V:n mallissa on määrä (eng. volume). Vaikka valtavan suuret datajoukot eivät välttämättä ole massadataa, massadatan datajoukot kuitenkin ovat valtavan suuria ja datan määrä lisääntyy jatkuvasti, suurella volyyymilla. Toinen ominaisuus on nopeus (eng. velocity). Datamäärien koon lisäksi datan luomisen ja hyödyntämisen nopeus on tärkeää, jotta sitä seuraava taloudellinen hyöty saataisiin maksimoitua. Viimeisenä massadatan ominaisuutena on monimuotoisuus (eng. variety). Se tarkoittaa, että kerättävä data ei ole pelkästään perinteistä jäsennettyä dataa, vaan mukana on esimerkiksi teksti-, kuva- ja videodataa, sensoreista kerättyä dataa tai matkapuhelinten GPS-signaaleja (Laney, 2001).

Kolmen V:n malliin on esitetty massadatan neljänneksi ominaisuudeksi arvoa (eng. value) (Chen, 2014; Kaisler, Armour, Espinosa & Money, 2013; Katal, Wazid, & Goudar, 2013) sekä viidenneksi ominaisuudeksi luotettavuutta (eng. veracity) (Emani, Cullot, & Nicolle, 2015). Esimerkiksi massadata-jätti IDC kuvasi artikkelissaan vuonna 2011 massadataa ”teknologioiden ja arkkitehtuurien uudeksi sukupolveksi, joka on suunniteltu erottamaan taloudellisesti arvoa hyvin suurista, monimuotoisista datamassoista hyödyntämällä nopeaa tiedon keräämistä, tutkimista, ja analysoimista”. Gantzin ja Reinselin (2011) mukaan massadatan arvoa voidaan arvioida kolmen eri kriteerin avulla: tuottaako se hyödyllisempää tietoa, parantaako se tiedon tarkkuutta tai parantaako se vastauksen oikea-aikaisuutta. Emanin ym. (2015) mukaan esimerkiksi epäselvyydet, duplikaatit ja vaillinaisuus ovat syitä, jotka voivat aiheuttaa massadatassa epävarmuutta. Tämän takia sen tulee olla luotettavaa. Vaikkakaan sen luotettavuutta ei pystyittäisi todistamaan, sille voidaan määritellä todennäköisyyksiä.

Edellä esitettyjen määritelmien perusteella voidaan todeta, että data-analytiikka on yläkäsite ja massadata osaltaan toteuttaa data-analytiikan tiettyjä toimintoja. Massadatan määritelmästä on eri mielipiteitä eikä täyttä yksimielisyyttä ole löydetty. Siksi on mahdotonta määritellä täysin tarkasti, milloin jokin on massadataa hyödyntävää kehittynyttä analytiikkaa.

3 URHEILUANALYTIikka

Data-analytiikka on melko uusi ja kehittyvä työkalu esimerkiksi yritysmaailmassa. Myös urheilun saralla sitä otetaan enemmän ja enemmän käyttöön. Se mahdollistaa isojen datamäärien hyödyntämistä päätöksenteon tukena, jolloin voidaan saavuttaa kilpailuetua. Analytiikkaan ja tiedonkeruuseen panostavat urheilujoukkueet saavatkin kilpailuetua muihin verrattuna (Alamar, 2013, s. 1-3).

Urheiluanalytiikka jakaa määritelmällisesti data-analytiikan ominaisuudet, mutta myös massadatan ominaisuuksia. Vaikka urheilussa esiintyvät datamäärät ovat Pakkalan ja Pääkkösen (2015) määritelmää pienempiä (Rein & Memmert, 2016), voidaan argumentoida urheiluanalytiikan toteuttavan sitä, sillä arviossa on otettava huomioon toimialaspesifit ohjelmistot ja se, kuinka suuria datajoukot yleisesti ovat (Manyika ym., 2011). Seuraavassa luvussa määritellään urheiluanalytiikka ja tarkastellaan sen eri komponentteja ja osa-alueita.

3.1 Urheiluanalytiikan määritelmä ja komponentit

Alamarin (2013, s. 3) mukaan urheiluanalytiikka on ”strukturoidun, historiallisen datan hallintaa, ennustavien analyttisten mallien soveltamista tuon datan hyödyntämiseksi, tietojärjestelmien käyttöä, joiden avulla päätöksentekijöitä voidaan informoida ja mahdollistaa heidän organisaatioidensa kilpailuedun saavuttaminen omalla pelikentällään”. Urheiluanalytiikan ensimmäinen päätavoite on tuoda tarvittu tieto tehokkaasti saataville. Tällä tavoin päätöksentekijöiden aika voidaan käyttää tiedon analysointiin sen keräämisen sijaan. Toinen päätavoite on tuottaa hyödyllistä tietoa käyttämällä analyttisiä malleja. Yhdistämällä tieto asiantuntijoiden näkemyksiin ja raportteihin, voidaan saada tarkempia arvioita halutuista asioista, kuten pelaajien ominaisuuksista. Nämä tavoitteet yhdistämällä voidaan saavuttaa kilpailuetua suhteessa muihin (Alamar, 2013, s. 5).

Urheiluanalytiikan neljä komponenttia ovat datan hallinta, analyttiset mallit, tietojärjestelmät ja johtaminen. Datan hallinnan tehtävänä on tuottaa dataa analyttisiin malleihin ja tietojärjestelmiin. Analyttiset mallit käyttävät dataa tuottaakseen tietoa tietojärjestelmiin tai vastatakseen päätöksentekijöiden spesifeihin kysymyksiin. Tietojärjestelmien tehtävänä on tarjota tieto päätöksentekijöille tehokkaalla ja käyttökelpoisella tavalla. Johtamista tarvitaan, jotta voidaan luoda tehokas analytiikkastrategia, sillä ilman sitä urheiluanalytiikasta ei saada sen maksimaalista hyötyä irti (Alamar, 2013, s. 3-5).

3.1.1 Datan hallinta

Datan lähteitä on paljon ja lähteiden määrä kasvaa jatkuvasti. Myös datan määrä kasvaa eksponentiaalisesti. Näiden lisäksi datatyyppejä on erilaisia; kvantitatiivista sekä kvalitatiivista, kuten tarkkailu- ja lääketieteellisiä raportteja sekä videodataa. Tämän takia on tärkeää, että data ja tieto eivät ole jakautuneet yksittäisille ihmisille. Datan hallinnan tehtävänä on varmistaa, että datan tulo sisään ja sen prosessointi eri toiminnoissa saadaan organisoitua, keskitettyä ja virtaviivaistettua. Tällöin tieto saadaan yhteen tietojärjestelmään, jolloin kaikki sen kanssa toimivat henkilöt pääsevät siihen käsiksi, eikä heidän tarvitse käyttää aikaansa sen keräämiseen esimerkiksi muilta henkilöiltä (Alamar, 2013, s. 5-6).

3.1.2 Analyttiset mallit

Analyttisten mallien keskeisin tehtävä on muuttaa raakadata luotettavaksi ja käyttökelpoiseksi tiedoksi. Ne ottavat kaiken datan, etsivät sieltä tarkoituksenmukaisia yhteyksiä eri muuttujien välillä ja käyttävät niitä tarkoituksenmukaisen tiedon tuottamiseksi esimerkiksi pelaajan tai joukkueen nykyisistä tai tulevista suorituksista. Tavoitteena analyttisillä malleilla on tukea päätöksentekoa tuottamalla monipuolisempaa ja tarkempaa tietoa. Päätöksentekijöiden vastuulle kuitenkin jää punnita, kuinka olennaista tämä tieto on minkäkin päätöksen kannalta.

Esimerkkinä analyttisen mallin hyödyntämisestä voidaan käyttää tilannetta, jossa päätetään, hankitaanko jokin pelaaja vai ei. Tällöin käytetään pelaajasta kerättyä tietoa kartoittamaan tämän vahvuuksia. On kuitenkin otettava huomioon myös muita tekijöitä; kuinka pelaajan ominaisuudet sopivat yhteen muiden joukkueen pelaajien kanssa? Miten ne vertautuvat vastustajiin? Pärjääkö pelaaja kovemmalla tasolla, mikäli hän tulee heikommasta sarjasta? Analyttisten mallien avulla pelaajaa pystytään vertailemaan esimerkiksi muihin samankaltaisista lähtökohdista tullessiin pelaajiin ja tekemään arvioita onnistumisen todennäköisyydestä (Alamar, 2013, s. 6-7).

3.1.3 Tietojärjestelmät

Tietojärjestelmät tarjoavat datasta erotettua tietoa päätöksentekijöille tarkoituksenmukaisesti, tehokkaasti, yhdenmukaisesti ja interaktiivisesti. Ne organisovat ja esittävät tietoa, jotta se on saatavilla keskitetysti. Tällöin päätöksentekijä voi käyttää aikansa tiedon analysointiin sen keräämisen sijasta. Yhdenmukaisuus tarkoittaa sitä, että kaikilla päätöksentekijöillä on käytössään sama informaatio. Tällöin esimerkiksi tulevaa vastustajaa varten valmistautuessa ei ihmisten tarvitse neuvotella keskenään vastustajan vahvuuksista ja heikkouksista, vaan aika voidaan käyttää suunnittelemaan sitä, millä tavoin niitä voidaan hyödyntää. Tietojärjestelmän interaktiivisuus taas tarkoittaa sitä, että eri syöte-tiedoilla päätöksentekijä saa tietoa erilaisista skenaarioista, kuten pelaajien keskinäisistä yksi vastaan yksi -tilanteista. Tietojärjestelmä pystyy siis eri muuttujia hyväksikäyttämällä tuottamaan arvioita, joita valmentajat voivat hyödyntää omassa päätöksenteossaan (Alamar, 2013, s. 8-9).

3.2 Pelaajien ja otteluiden suoritusten analytiikka

Useimmat ihmiset ajattelevat analytiikan urheilussa tarkoittavan joukkueen ja pelaajien suoriutumisen parantamista (Davenport, 2014). Vaikkakin se on yksi urheiluanalytiikan osa-alueista, se ei ole ainut. Davenport (2014) jaottelee urheiluanalytiikan osa-alueet kolmeen; pelaajien ja otteluiden suoritusten analytiikkaan, urheiluliiketoiminnan analytiikkaan ja pelaajien terveyden ja loukkaantumisten analytiikkaan. Tämän tutkielman rajauksen vuoksi käsitellään kuitenkin vain pelaajien ja otteluiden suoritusten analytiikkaa.

Yksi urheiluanalytiikan tavoitteista on pelaajien ja joukkueiden suoritusten parantaminen. Esimerkkeinä urheiluanalytiikan hyödyistä ovat oikeiden pelaajien hankinta, oikeiden pelaajien peluuttaminen ja oikeiden ratkaisujen teko kentällä. Vaikka urheiluanalytiikan yleisenä tavoitteena on tuottaa kilpailuetua, on myös osa-alueita, joissa siihen ei enää pystytä. Tällöin analytiikan hyödyntämisestä on tullut normi, jota kaikki käyttävät, joten ilman uusia lähestymistapoja ei kilpailuetua saada. Näitä osa-alueita kutsutaan tässä tutkielmassa analytiikan tukikeinoiksi. Hyvällä tukikeinojen toteutuksella voidaan kuitenkin saavuttaa pieniä hyötyjä ja ilman tukikeinojen käyttöä annetaan muille etumatkaa (Davenport, 2014).

Esimerkkinä analytiikan tukikeinosta on analytiikan käyttö varaustilaisuuksissa. Analytiikan avulla pyritään asettamaan varattavat pelaajat järjestykseen, jotta joukkueeseen varattaisiin suurimmalla todennäköisyydellä ammattilaisrjassa menestyvä pelaaja (Alamar, s. 7-8; Davenport, 2014). Tämä on kuitenkin yksi yleisimmistä analytiikan tukikeinoista ja jokainen muukin joukkue hyödyntää sitä. Ei myöskään ole todisteita siitä, että varaustilaisuuksien osalta

joukkueiden sisäinen analytiikka johtaisi parempiin tuloksiin kuin kolmansien osapuolten tekemät pelaajalistaukset (Davenport, 2014).

Esimerkiksi NBA:ssa ja NHL:ssä käytetään erilaisia malleja, jotka kertovat pelaajien panoksesta missäkin tilanteessa. NBA:ssa tämä tarkoittaa kentällisten optimointia eli pelaajien panosta suhteessa vastustajaan ja muihin kentällä oleviin omiin pelaajiin, niin hyökkäyksessä kuin puolustuksessa (Davenport, 2014; Kang, 2018). NHL:ssä pelaajien panosta mittaamaan on kehitetty esimerkiksi erilaisia laukauksien perusteella maaliodottamia luovia malleja, kuten Corsi- ja Fenwick-luokitus. (Macdonald, 2012).

Urheiluanalytiikan kilpailuetua luovat käyttötavat ovat pääsääntöisesti vielä varhaisessa omaksumisen vaiheessa. Lajista riippumatta, näistä tärkein on videodatan analytiikka (Davenport, 2014). Videodatan analytiikkaa voidaan hyödyntää esimerkiksi jalkapallossa koostamalla automaattisesti ottelun tärkeimmät tapahtumat, jolloin pitkistä videosta saadaan halutut kohdat tallennettua tietokantaan ja etsittyä niitä tietokannasta sisällön perusteella. Toinen esimerkki analytiikan käyttötavasta on korkean tason analyysin saaminen automaattisesti ja vähällä ihmisen avustuksella. Korkean tason analyysin avulla voidaan tarkastella esimerkiksi joukkueiden taktiikoita ja pelitapoja, sekä yksittäisten pelaajien ominaisuuksia ja rooleja (D'orazio & Leo, 2010).

Toinen aikaisessa omaksumisen vaiheessa oleva kilpailuedun luoja on paikka- sekä biometrisen datan käyttö (Davenport, 2014). Sen avulla voidaan esimerkiksi tunnistaa automaattisesti samankaltaisia hyökkäysten tapahtumaketjuja (Hirano & Tsumoto, 2005). Aikaisessa omaksumisvaiheessa olevien tapojen lisäksi kilpailuedun luojaan kuuluvat esimerkiksi joukkueen sisäinen, ulkopuolisilta suljettu pelaajadata sekä pelaajien osallistaminen analytiikan hyödyntämiseen (Davenport, 2014).

4 DATA-ANALYTIIKAN HYÖDYNTÄMINEN JALKAPALLON TAKTISESSA ANALYYSISSÄ

Usein käsiteltäessä data-analytiikan hyödyntämistä jalkapallossa, yhdistetään siihen massadata. Vaikka myös jalkapallon datamäärät ovat kaukana Pakkalan ja Pääkkösen (2015) useamman terabitin määritelmästä, käsitellään jalkapallossa toteutettavaa analytiikkaa usein massadatan analytiikkana (Rein & Memmert, 2016). Koska eri tutkimukset ja artikkelit käsittelevät jalkapallon data-analytiikkaa eri termein, käytetään tässä tutkielman osassa yleisempää käsitettä analytiikka. Massadataa spesifisemmin käsiteltäessä käytetään kuitenkin käsitettä massadata.

4.1 Taktiikat

Taktiikat ovat keskeinen komponentti menestykseen modernissa huipputason jalkapallossa (Rein & Memmert, 2016). Cambridgen sanakirja määrittelee taktiikan ”suunnitelluksi tavaksi tehdä jotakin”. Oikean taktiikan valinta on erittäin tärkeää ennen peliä tapahtuvassa valmistautumisessa ottelun voittamiseksi (Carling, Williams, Reilly, 2005; Yiannakos & Armatas, 2006; Sampaio & Macas, 2012).

Gargantan (2009) mukaan taktiikka määrittää sen, miten joukkue käyttää tilaa, aikaa ja tekoja ottelun voittamiseksi. Tilan käyttämisellä tarkoitetaan sen hallintaa, missä tapahtumat tapahtuvat ja mitä aluetta joukkue haluaa miehittää sekä hyökätessä että puolustettaessa. Aika tarkoittaa sitä, kuinka usein tietyt tapahtumat tapahtuvat, kuinka kauan ne kestävät (esim. pallonhallinta) tai kuinka nopeasti ne toteutetaan. Teot ovat yksittäisten pelaajien tekemiä valintoja – syöttöjä, laukauksia, katkoja. Taktisen suorituksen onnistuminen riippuukin siis yksittäisten pelaajien tai joukkueiden ajasta ja paikasta riippuvien suoritusten laadusta (Memmert, Lemmink, & Sampaio, 2017). Taktiikat eivät myöskään ole muuttumattomia, vaan niitä pitää adaptoida joukkueiden välisen vuo-

rovaikutuksen perusteella (Garganta 2009; Grehaigne ym. 1997). Jalkapallon taktiikat ulottuvat siis yksilötasolta joukkue-tason päätöksentekoprosesseihin.

4.2 Perinteinen taktinen analyysi ja siihen liittyvät haasteet

Perinteisesti jalkapallon taktinen analyysi on tehty käyttämällä dataa, joka on kerätty manuaalisesti ottelua seuraamalla ja eri muuttujia tarkkailemalla (Mackenzie & Cushion 2013). Aiemmin taktinen analyysi on ollut riippuvainen kerätyistä datasta muodostetuista keskiarvoisista statistiikoista ja luvuista (Hughes & Bartlett, 2002). Käytettyjä indikaattoreita ovat olleet esimerkiksi erilaiset syöttöihin liittyvät muuttujat (Hughes & Franks, 2005; Liu, Lago-Penas & Sampaio, 2015), pallonhallinta (Collet, 2013), pallon takaisin saaminen (Vogelbein, Nopp & Hokelmann, 2014) ja pelityyli (Tenga, Holme, Ronglan, & Bahr, 2010).

Manuaalisella seurannalla kerätyistä datasta puuttuu kuitenkin usein luotettavuus ja objektiivisuus (James, Mellalieu, & Hollely, 2002), sen tekeminen tarkasti on erittäin aikaa vievää (Carling, Bloomfield, Nelsen, & Reilly, 2008) ja se vaatii eksperttien osaamista ja tietotaitoa (Memmert ym. 2017). Myöskään relevanteista, seurattavista muuttujista ei ole yleistä konsensusta (Sarmiento ym. 2014). Tämän lisäksi, kun taktinen analyysi tehdään hyödyntämällä perinteisin keinoin kerättyä dataa eli kuvailevia, keskiarvoisia statistiikkoja, sen selitysteho jää heikoksi (Hughes & Bartlett 2002; Tenga ym. 2010; Mackenzie & Cushion 2013).

4.3 Data-analytiikka jalkapallon taktisessa analyysissä

Analytiikan fokus jalkapallossa on ollut pitkälti kuvailevissa statistiikoissa, mutta painotus on muuttumassa kohti ennustavaa ja vikaa etsivää. Sen sijaan, että analysoitaisiin mitä tapahtui, analysoidaan mitä tulee tapahtumaan ja miksi (Davenport, 2014). Aiemmin analytiikan hyödyntämistä on vaikeuttanut saatavilla olevan relevantin datan vähäisyys (Rein & Memmert, 2016). Teknologian kehittymisen myötä on kuitenkin mahdollistettu esimerkiksi tarkan paikkadatan (Memmert ym. 2017) ja fysiologisen datan (Davenport, 2014; Rein & Memmert, 2016) kerääminen kaikilta pelaajilta. Tämä on johtanut jopa päinvastaiseen ongelmaan, sillä nykyisin dataa on saatavilla niin paljon, ettei tiedetä, miten sitä tulisi hallita ja hyödyntää. Esimerkiksi yhdestä ottelusta saadaan kerättyä kaikkien 22 pelaajan ja pallon paikkojen seurannalla yhteensä noin 3,1 miljoonaa datapistettä (Rein & Memmert, 2016).

Nykyisin voidaan automaattisesti kerätä otteluista paikkatietoa. Tämän tiedon avulla pystytään purkamaan jalkapallo-ottelun kompleksinen ja dynaaminen luonne erilaisiksi visualisoinneiksi (Janetzko ym. 2014; Memmert ym.

2017). Myös videodatasta voidaan analytiikan avulla saada tietoa joukkueiden taktisesta käytöksestä. TV-lähetysten ja ympäri kenttää asennettavien kameroiden avulla voidaan koostaa korkean tason analyysijä joukkueiden taktiikoista (D'orazio & Leo, 2010). Seuraavissa kappaleissa käsitellään tarkemmin video- ja paikkadatan hyödyntämistä.

4.3.1 Videodata

Jalkapallon taktisen analyysin kannalta videodatan analytiikkaa voidaan hyödyntää erilaisten yhteenvetojen ja koontien luomiseen sekä korkean tason analyysin tekemiseen automatisoidusti. Yhteenvetojen avulla pitkät otteluvideot voidaan jakaa tärkeimpiin tapahtumiin ja tallentaa tietokantaan, josta niitä voidaan etsiä avainsanoja hyödyntäen (D'orazio & Leo, 2010). Tällä tavoin voidaan helpottaa valmentajien tekemää taktista analyysia.

Videodatasta pystytään myös automatisoidusti luomaan korkean tason analyysijä. Tällöin tv-kameran kuvastriimin perusteella pystytään vastaamaan esimerkiksi seuraaviin kysymyksiin (D'orazio & Leo, 2010):

- Mitkä ovat joukkueiden hyökkäyspelille luonteenomaisia piirteitä?
- Mitkä ovat yksittäisten pelaajien vahvuudet ja heikkoudet?
- Mitkä ovat pelaajien roolit?
- Sopivatko pelaajien ominaisuudet määriteltyihin rooleihin?
- Onnistuvatko pelaajat suorituksissaan?
- Miten joukkue rakentaa maalintekopaikkansa?
- Mitkä ovat kunkin pelaajan taidot?
- Mikä on joukkueen taktinen muodostelma?

Korkean tason analyysiä tehdessä seurantaohjelmiston tehtävä on havaita ja seurata liikkuvia objekteja. Näitä ovat pallo, pelaajat ja tuomari. Myös useamman ihmisen havaitseminen ja seuraaminen, näiden etäisyyksien palloon seuraaminen sekä useamman kuvakulman liikerata-analyysi ovat tarpeen korkean tason analyysiä tehtäessä. Tällä tavoin on myös mahdollista tehdä reaaliaikaista videon prosessointia, jolloin tietoa voidaan saada valmentajille pelin aikana ja valmentajat voivat muokata taktiikoita kesken pelin (D'orazio & Leo, 2010).

4.3.2 Paikkadata

Paikkadata on jalkapallon kontekstissa yleensä spatio-temporaalista, eli sijaintitietojen lisäksi siihen liittyy yleensä myös aikatietoja, eli missä vaiheessa ottelua pelaaja on ollut tietyssä paikassa. Pelaajien video- ja GPS-seurannan yleistettyä paikkadatan kerääminen on mahdollistunut tiheämmäksi ja tarkemmaksi (Janetzko ym. 2014). Pelaajien seurannan myötä analytiikan avulla luodut peliraportit ovat myös yleistyneet (Rein & Memmert, 2016) ja jalkapallo-ottelun dynaaminen ja kompleksinen luonne pystytään kuvaamaan ja visualisoimaan (Janetzko ym. 2014; Memmert ym. 2017). Seurantajärjestelmien avulla paikkada-

tan käsittely on nopeutunut valtavasti ja analyysiin käytetty aika on pudonnut entisen manuaalisen analyysin 6-8 tunnista joihinkin minuutteihin (Memmert ym. 2017).

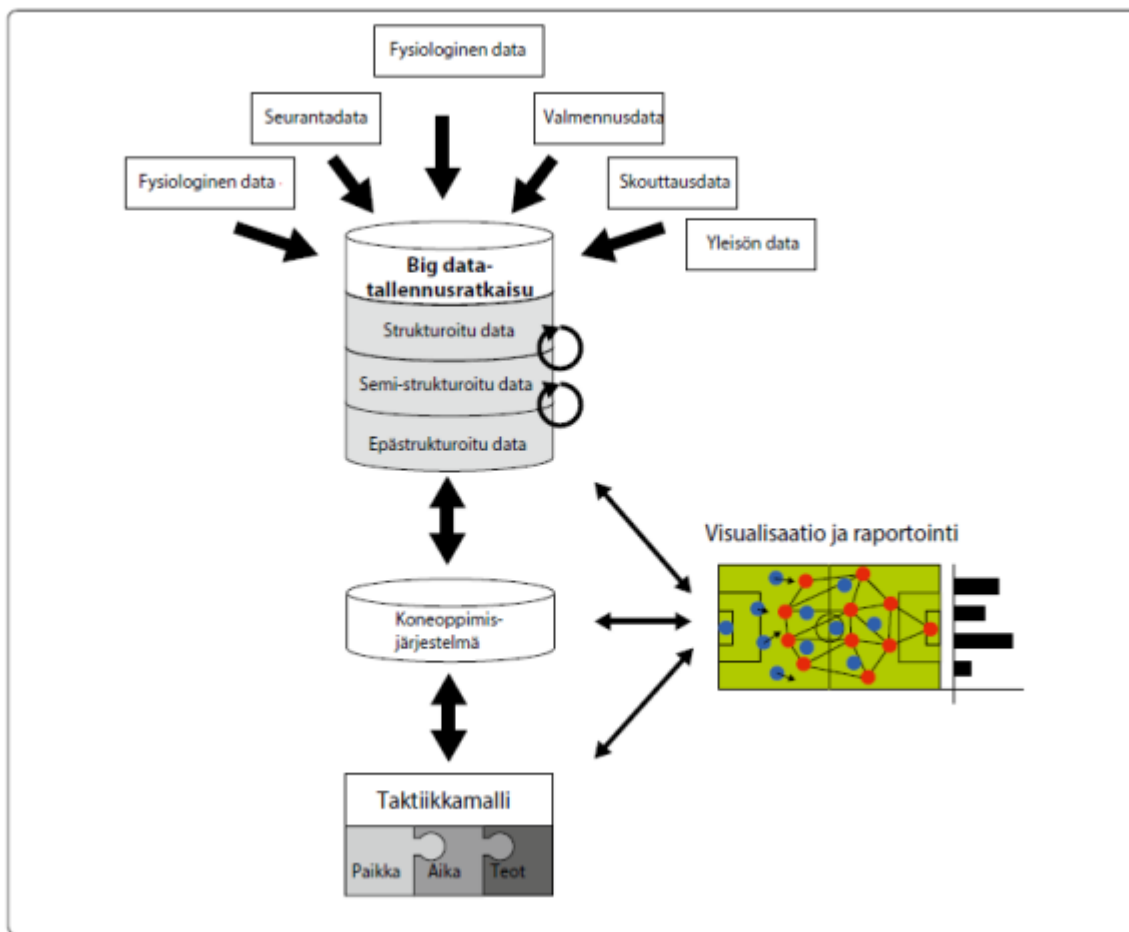
Taktiset järjestelmät jalkapallo-ottelussa ovat osaltaan vastuussa pelaajien tuottavuuden ja hyödyn parantumisesta tai huonontumisesta. Paikkadatan avulla voidaan luoda kollektiivisia suorituskyvyn indikaattoreita, jotka pystyvät kuvaamaan ja ymmärtämään näiden kompleksisten, epälineaaristen ja kaoottisten organisaatioiden, eli kentällä olevien joukkueiden, dynamiikkaa. Mittaamalla pelaajien nopeuksia ja etäisyyksiä oman pelipaikkansa geometriseen keskipisteeseen, voidaan luoda arvioita pelaajien välisestä koordinaatiosta ja esittää ennusteita pelaajien liikkumiskäyttäytymisestä. Näiden perusteella voidaan arvioida yksittäisen pelaajan kokonaiskontribuutioita joukkueelle (Memmert ym. 2017).

Käyttämällä pelaajien pelipaikkakohtaisia geometrisiä keskipisteitä, voidaan tarkastella joukkueen ja joukkueen eri linjojen (puolustus, keskikenttä, hyökkäys) sijoittumista ja toimintaa kentällä. Näiden avulla pystytään kuvaamaan joukkueiden hyökkäys- ja puolustuspelaamista (Olthof, Frencken, & Lemmink, 2015). Kun tarkastellaan näiden välisiä etäisyyksiä esimerkiksi ennen maalintekotilanteita, voidaan saada ymmärrystä maalintekoprosessin toiminnasta tai vastaavasti vastustajan maalintekoprosessin estämisestä. Erityisesti linjojen välisten etäisyyksien vaihtelun on havaittu olevan suurempaa 10-20 sekuntia ennen maalintekotilannetta (Memmert ym. 2017).

4.3.3 Massadata

Yhteisesti hyväksytyjen seurattavien muuttujien lisäksi yleisen teoreettisen mallin puuttuminen on yksi este joukkuetaktiikoiden tutkimisessa (Garganta, 2009; Mackenzie & Cushion, 2013). Massadatan on yhdessä koneoppimisen kanssa ehdotettu olevan myös ratkaisu teoreettisen mallin rakentamiseen (Rein & Memmert, 2016). Jalkapallon kontekstissa kolmen V:n malli voisi tuoda ratkaisuja myös seuraavissa asioissa (Rein & Memmert, 2016):

- Volyyymi: Dataa voidaan tallentaa strukturoidusti ja tarjota käyttäjälle pääsy siihen helposti.
- Monimuotoisuus: Massadata tarjoaa ratkaisuja erilaisen datan – strukturoidun, semi-strukturoidun ja strukturoimattoman – yhdistämiseen.
- Nopeus: Dataan päästään käsiksi nopeasti ja sitä voidaan käyttää esimerkiksi reaaliaikaisesti ottelun aikana taktiikan muokkaamiseen.



KUVIO 1 Massadatateknologiapino jalkapallon taktiseen analyysiin (Rein & Memmert, 2016)

Jokainen massadatan avainkonsepti tarjoaa siis spesifejä ratkaisuja huipputason jalkapallon taktiseen analyysiin (kuvio 1). Massadatan analytiikan hyödyntäminen vaatii kuitenkin huomattavan määrän asiantuntemusta onnistuakseen. Aiheen tutkimus vaatii myös monitieteellisyyttä, sillä urheilututkijat tarvitsevat oman tieteenalaspesifin osaamisensa lisäksi tietojenkäsittelytieteilijöiden teknisiä taitoja (Rein & Memmert, 2016).

5 JOHTOPÄÄTÖKSET

Tutkielmassa käsiteltiin tieteellisen kirjallisuuden avulla data-analytiikan hyödyntämistä huipputason jalkapallon taktisessa analyysissä. Tutkimuskysymykseen vastaamiseksi määriteltiin käsitteet **data-analytiikka**, **massadata** sekä **urheiluanalytiikka**. Tämän jälkeen pyrittiin vastaamaan tutkimuskysymykseen ”Kuinka data-analytiikkaa voidaan hyödyntää jalkapallon taktisessa analyysissä?”.

Data-analytiikka määriteltiin yleisesti analytiikaksi, jonka avulla isoista datajoukoista saadaan informaatiota käytettäväksi päätöksenteon tukena. Massadata puolestaan määriteltiin valtaviksi datamassoiksi, joiden käsittelyyn perinteiset tietokoneet eivät riitä ja jotka ovat monimuotoisia sekä niiden käsittely on nopeaa. Kolmas käsite, urheiluanalytiikka, määriteltiin olevan ennustavien analyttisten mallien soveltamista historialliseen dataan päätöksentekijöiden informoimiseksi ja kilpailuedun saavuttamiseksi urheilun kontekstissa.

Jotta data-analytiikasta voidaan jalkapallojoukkueessa saada kaikki sen potentiaali irti, tarvitaan sen käyttöön organisaation tuki. Tarvitaan myös toimiva datan prosessointiputki ja tietojärjestelmä, johon data voidaan tallentaa ja josta se saadaan nopeasti hyödynnettäväksi. Kun eri lähteistä saatava erityyppinen data voidaan keskittää yhteen paikkaan, on jokaisella dataa käsittelevällä ihmisellä pääsy samaan dataan. Tällöin dataa voidaan hyödyntää tehokkaasti. Massadata tarjoaa jalkapallon kontekstissa näihin ongelmiin erityisiä ratkaisuja. Sen avulla suuria määriä dataa erityyppistä dataa voidaan tallentaa ja käyttää nopeasti. Sen käyttö vaatii kuitenkin myös todella suuren määrän asiantuntemusta onnistuakseen.

Tutkielmassa käsiteltiin tieteellisen kirjallisuuden pohjalta myös jalkapallon taktiikoita ja perinteisiä menetelmiä niiden analysointiin. Taktiikka on keskeinen tekijä joukkueen menestyksessä ja myös vastuussa pelaajien tuottavuuden parantumisesta tai huonontumisesta. Se määrittää sen, kuinka joukkue käyttää tilaa, aikaa ja tekoja jalkapallo-ottelussa. Perinteisesti taktiseen analyysiin käytetty data on kerätty manuaalisesti ottelua seuraamalla. Tällä tavoin kerätty data on kuitenkin usein epäluotettavaa, sen kerääminen on erittäin hidasta ja sen selitysteho on heikkoa.

Data-analytiikkaa hyödynnettäessä saatu informaatio on objektiivisempaa ja tarkempaa, sillä se ei perustu ihmisten subjektiivisiin näkemyksiin. Se tekee datan keräämisestä myös valtavasti nopeampaa. Tärkeitä datalähteitä ovat video- ja paikkadata. Näiden pohjalta analytiikan avulla voidaan luoda visualisointeja ja raportteja jalkapallo-otteluista. Näin voidaan analysoida joukkueiden taktista käytöstä, kuten millä tavoin joukkue hyökkää ja luo maalipaikkoja tai mikä on sen taktinen muodostelma. Myös yksittäisten pelaajien kontribuutioita joukkueelle voidaan arvioida tällä tavoin.

Käsitellyn kirjallisuuden pohjalta voidaan tehdä johtopäätös, että hyödyntämällä data-analytiikkaa video- ja paikkadataan voidaan saada huomattavia määriä kilpailuetua suhteessa muihin joukkueisiin taktisessa analyysissä. Kilpailuetua syntyy tällöin sekä informaation tarkkuudessa eli taktinen analyysi on tarkempaa, että tehokkuudessa, jolloin analyysiä tekevien ihmisten ei tarvitse kuluttaa aikaansa esimerkiksi datan keräämiseen. On kuitenkin huomattava, että data-analytiikan adaptoiminen ei ole helppoa, vaan se vaatii organisatiotason muutoksia, huomattavia rahallisia panostuksia ja asiantuntemusta.

Jalkapalloseurojen massadatan käytöstä taktisessa analyysissä ei ole tehty tieteellistä tutkimusta. Aiheeseen liittyvä konkreettinen tieto perustuu lähinnä muutamiin lehtiartikkeleihin, jotka käsittelevät pintapuolisesti yksittäisten seurojen massadatan käyttöä. Oletettavasti seurat eivät siis halua paljastaa tapoja, joilla ne saavat massadatasta kilpailuetua. Jatkotutkimusaiheena voisikin selvittää massadatan käyttöä jalkapalloseurojen taktisessa analyysissä. Seurojen saavutettavuuden vuoksi tutkimus voisi olla järkevä tehdä Suomen huipputason jalkapalloseuroilla.

LÄHTEET

- Alamar, B. (2013). Sports analytics: A guide for coaches, managers, and other decision makers. Columbia University Press, s. 4
- Aggarwal, V. B., Bhatnagar, V., & Mishra, D. K. (2017). *Big data analytics*. Singapore: Springer. doi:10.1007/978-981-10-6620-7 Retrieved from [https://ebookcentral.proquest.com/lib/\[SITE_ID\]/detail.action?docID=5087714](https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=5087714)
- Cao, L. (2017). Data science. *ACM Computing Surveys (CSUR)*, 50(3), 1-42. doi:10.1145/3076253
- Cambridge Dictionary (n. d.). *Cambridge University Press* Haettu osoitteesta <https://dictionary.cambridge.org/dictionary/english/tactic>
- Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). *The role of motion analysis in elite soccer* Springer International Publishing. doi:10.2165/00007256-200838100-00004
- Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in european and international football, 2007-2010. *Journal of Sports Sciences*, 31(2), 123-136. doi:10.1080/02640414.2012.727455
- Emilie Baro, Samuel Degoul, Régis Beuscart, & Emmanuel Chazard. (2015). Toward a literature-driven definition of big data in healthcare. *BioMed Research International*, 2015, 639021-9. doi:10.1155/2015/639021
- Carling C, Williams AM, Reilly T (2005a) From technical and tactical performance analysis to training drills Handbook of soccer match analysis: a systematic approach to improving performance. Routledge, London
- Gantz J, Reinsel D (2011) Extracting value from chaos. IDC iView, pp 1-12
- Davenport, T. H. (2014a). Analytics in sports: The new science of winning. International Institute for Analytics, 2, 1-28.
- Emani, C., Cullot, N., & Nicolle, C. (2015). *Understandable big data: A survey* Elsevier. Retrieved from https://www.openaire.eu/search/publication?articleId=od_1631::e973924fea2f58ca0fbf8a04b5ec8daa
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, 13(3), 57.

- Garganta, J. (2009). Trends of tactical performance analysis in team sports: Bridging the gap between research, training and competition. *Revista Portuguesa De Ciências do Desporto*, 9(1), 81-89. doi:10.5628/rpcd.09.01.81
- Grehaigne, J., Bouthier, D., & David, B. (1997). Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences*, 15(2), 137-149. doi:10.1080/026404197367416
- Hirano, S., & Tsumoto, S. (2005). Grouping of soccer game records by multiscale comparison technique and rough clustering. doi:10.1109/ICHIS.2005.53 Retrieved from <https://ieeexplore.ieee.org/document/1587780>
- Hughes, Mike & Franks, Ian. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of sports sciences*. 23. 509-14. 10.1080/02640410410001716779.
- James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, & Angela Hung Byers. (2011). *Big data: The next frontier for innovation, competition, and productivity*. ().McKinsey Global Institute.
- James, N., Mellalieu, S. D., & Hollely, C. (2002). Analysis of strategies in soccer as a function of european and domestic competition. *International Journal of Performance Analysis in Sport*, 2(1), 85-103. doi:10.1080/24748668.2002.11868263
- Janetzko, H., Sacha, D., Stein, M., Schreck, T., Keim, D. A., & Deussen, O. (Oct 2014). Feature-driven visual analytics of soccer data. doi:10.1109/VAST.2014.7042477 Retrieved from <https://ieeexplore.ieee.org/document/7042477>
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (Jan 2013). Big data: Issues and challenges moving forward. doi:10.1109/HICSS.2013.645 Retrieved from <https://ieeexplore.ieee.org/document/6479953>
- Kang, K. (2018). *Estimation of NBA players' offense/defense ratings through shrinkage estimation* doi:10.1184/r1/6685616.v1
- Katal, A., Wazid, M., & Goudar, R. H. (Aug 2013). Big data: Issues, challenges, tools and good practices. doi:10.1109/IC3.2013.6612229 Retrieved from <https://ieeexplore.ieee.org/document/6612229>
- Laney D (2001) 3-d data management: controlling data volume, velocity and variety. META Group Research Note, 6 February
- Liu H, Gomez MÁ, Lago-Peñas C, Sampaio J. Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*. 2015 ;33(12):1205-1213. DOI: 10.1080/02640414.2015.1022578.

- Macdonald, B. (2012). Adjusted plus-minus for NHL players using ridge regression with goals, shots, fenwick, and corsi. *Journal of Quantitative Analysis in Sports*, 8(3) doi:10.1515/1559-0410.1447
- Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of Sports Sciences*, 31(6), 639-676. doi:10.1080/02640414.2012.746720
- McAfee, A., & Brynjolfsson, E. (2012, Oct). Big data: The management revolution. *Harvard Business Review*, 90, 60. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23074865>
- Meijer, E. (2011, Oct 1.). The world according to LINQ. *Communications of the ACM*, 54, 45-51. doi:10.1145/2001269.2001285 Retrieved from <http://dl.acm.org/citation.cfm?id=2001285>
- Memmert, D., Lemmink, Koen A P M, & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1), 1-10. doi:10.1007/s40279-016-0562-5
- Olthof, S. B. H., Frencken, W. G. P., & Lemmink, K. A. P. M. (2015). The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Human Movement Science*, 41(June), 92-102. doi:10.1016/j.humov.2015.02.004
- 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. doi:10.1016/j.bdr.2015.01.001
- Rampinini, E., Coutts, A. J., Castagna, C., Sassi, R., & Impellizzeri, F. M. (2007). Variation in top level soccer match performance. *International Journal of Sports Medicine*, 28(12), 1018-1024. doi:10.1055/s-2007-965158
- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1-13. doi:10.1186/s40064-016-3108-2
- Romanillos, G. (2016). Big data and cycling. *Transport Reviews*, 36(1), 114-133. Retrieved from <http://www.econis.eu/PPNSET?PPN=851382096>
- Runkler, T. A. (2016). *Data analytics* (2nd ed. 2016 ed.). Wiesbaden: Springer Gabler. in Springer Fachmedien Wiesbaden GmbH. doi:10.1007/978-3-658-14075-5 Retrieved from [https://ebookcentral.proquest.com/lib/\[SITE_ID\]/detail.action?docID=5585267](https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=5585267)

- Sampaio, J., & Maças, V. (2012). Measuring tactical behaviour in football. *International Journal of Sports Medicine*, 33(5), 395-401. doi:10.1055/s-0031-1301320
- Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). Match analysis in football: A systematic review. *Journal of Sports Sciences: Science and Medicine in Football*, 32(20), 1831-1843. doi:10.1080/02640414.2014.898852
- Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from norwegian professional soccer matches. *Journal of Sports Sciences*, 28(3), 245-255. doi:10.1080/02640410903502766
- Vogelbein M, Nopp S, Hokelmann A (2014) Defensive transition in soccer – are prompt possession regains a measure of success? A quantitative analysis of German Fussball-Bundesliga 2010/2011
- Yiannakos, A., & Armatas, V. (2006). Evaluation of the goal scoring patterns in european championship in portugal 2004. *International Journal of Performance Analysis in Sport*, 6(1), 178-188. doi:10.1080/24748668.2006.11868366
- Zikopoulos P, Eaton C et al (2011) Understanding big data: analytics for enterprise class hadoop and streaming data. McGraw Hill Osborne Media