

Rasmus Fors

BIG DATA ANALYTIIKAN HYÖDYNTÄMINEN JALKAPALLOSSA



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2022

TIIVISTELMÄ

Fors, Rasmus

Big data analytiikan hyödyntäminen jalkapallossa

Jyväskylä: Jyväskylän yliopisto, 2022, 25 s

Tietojärjestelmätiede, kandidaatin tutkielma

Ohjaaja: Kyppö, Jorma

Tässä kandidaatin tutkielmassa tutkitaan Big datan käyttöä jalkapallon pelillisten elementtien kehitystyökaluna. Big data on ollut kuuma puheenaihe monilla aloilla viimeisen vuosikymmenen aikana. Tämä on myös heijastunut urheilun maailmaan. Big data analytiikan tehokasta käyttöä varten vaaditaan kuitenkin merkittäviä investointeja ja osaamista. Oikein toteutettuna kuitenkin Big data voi tarjota jalkapallo joukkueille kilpailullisen etuaseman suhteessa vastustajiin esimerkiksi tehostamalla pelaajien ominaisuuksia, pienentämällä loukkaantumiseriskejä ja tehostamalla taktista analyysiä. Tutkielma on toteutettu kirjallisuuskatsauksena. Tutkielmassa huomattiin Big datan käytön olevan melko pientä verrattuna muihin isoihin urheilulajeihin, mutta sitä Big dataa implementoimalla on jo saatu merkittäviä urheilullisia saavutuksia.

Asiasanat: Big data, urheiluanalytiikka, jalkapallo, Big data analytiikka

ABSTRACT

Fors, Rasmus

The usage of big data analytics in football

Jyväskylä: University of Jyväskylä, 2022, 25 pp.

Information Systems, Bachelor's thesis

Supervisor: Kyppö, Jorma

This bachelor's thesis is a literature review of how big data is being used as a tool to enhance footballers and football teams skills and methodologies. Big data has been a hot topic in various fields the past decade, even in the sports industry. Efficient use of big data however requires significant investments and know how. When implemented correctly big data can provide football teams a competitive advantage against rivalry teams. Big data can provide better tools to develop the players skills, help minimize injuries and enhancing tactical analysis. The usage of big data is not as common in football when compared to other sports, however the usage of big data has already provided significant achievements to those who have implemented it correctly.

Keywords: Big data, sports analytics, football, soccer, big data analytics

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KUVIOT JA TAULUKOT

1	JOHDANTO.....	5
2	BIG DATA.....	8
2.1	Big datan määritelmiä.....	8
2.2	Big datan mahdollisuudet.....	9
2.3	Big datan haasteet.....	10
2.4	Big data analytiikka.....	11
3	URHEILUANALYTIikka.....	13
4	BIG DATA JA URHEILUANALYTIikka JALKAPALLOSSA.....	17
4.1	Datan kerääminen.....	17
4.1.1	Videodata.....	18
4.1.2	Biometrinen data.....	18
4.1.3	Sijaintidata.....	18
4.2	Big datan hyödyntäminen jalkapallossa.....	19
5	YHTEENVETO.....	21
	LÄHTEET.....	23

1 JOHDANTO

Digitalisaation seurauksena ympärillämme oleva maailma ja kulttuuri on muovautunut ennennäkemätöntä vauhtia. Toimintamme on siirtynyt yhä enemmässä määrin digitaalisille alustoille ja tämä vuorovaikutus on jo lisännyt ja tulee lisäämään suurenevia datamassoja. Valtavia, järjestelemättömiä ja monimuotoisia tietomassoja kutsutaan Big dataksi (Suom. Massadata). Big datan määritelmä on varsin löyhä ja täysin yhtenäistä määritelmää ei ole vielä kyetty kehittämään (Gandomi & Haider, 2015). Datamäärille ja -muodoille on tapahtunut räjähdysmäinen kasvu, jolle nähdään valtavasti liiketoimintamahdollisuuksia lähes kaikilla aloilla jossain muodossa. Big data mullistus koskettaa myös urheilumaailmaa, jonka perinteisesti ei nähdä liittyvän informaatioteknologiaan juuri lainkaan. Big datan yhdistäminen urheiluun arvioitu kasvavan vuoteen 2024 mennessä 5.2 miljardin dollarin toimialaksi, mikä kertoo sen kasvavasta osuudesta urheilumaailmassa. Big data ratkaisuihin liittyy kuitenkin paljon haasteita. Näitä haasteita ovat esimerkiksi osajien ja laskentatehon puutteesta sekä johdon tietämättömyydestä Big datan mahdollisuuksista (McAfee & Brynjolfsson, 2012). Urheilussa Big dataan liitettäviä hyötyjä ovat esimerkiksi loukkaantumisriskien minimointi, paremmat mahdollisuudet kehittää pelaajien ominaisuuksia, paremmat pelaajahankinnat ja urheiluliiketoiminnan tehostaminen. Aiemmassa kirjallisuudessa on myös pohdittu, kuinka perinteiset, intuitiivisuuteen pohjautuvat valmennustavat, ovat ristiriidassa dataohjautuviin menetelmiin ja millaisia haasteita valmennustyö kohtaa lähitulevaisuudessa.

Data-analytiikka työ jalkapallon kontekstissa nähdään monesti alkaneen 1950-luvulla T. C. Reepin ponnisteluiden kautta. Hänen havaintojensa perusteella huomattiin, että alle kolmen syötön syöttökuvioista syntyi useimmiten maaleja (Apostolou & Tjortjis, 2019). Tämä johti etenkin englantilaisessa jalkapallossa pelityyliin, jossa suosittiin pitkiä syöttöjä, jotta voitaisiin edetä hyökkäyssuunnassa mahdollisimman nopeasti. Pitkiä palloja suosivaa pelityyliä voidaan vielä tänäkin päivänä havaita englantilaisessa jalkapallossa. Nykypäivänä data-analytiikkamenetelmät ovat kuitenkin huomattavasti kehittyneemmät ja datasta voidaan huomata paljon muutakin kuin pelkästään numeroita.

Kandidaatintutkielmassani tutkin, kuinka Big data on vaikuttanut jalkapallossa toteutettavaan urheiluanalytiikkaan ja millaisia hyötyjä Big datasta voidaan valjastaa urheiluasuoritusten kehittämisen näkökulmasta. Urheiluanalytiikka itsessään on toistaiseksi verrattain vähäisesti tutkittu aihe, mutta aihe nähdään olevan kuuma urheiluliiketoiminnan lähitulevaisuudessa (Morgulev ym., 2018).

Tutkielman tutkimuskysymys on ”Miten Big dataa käytetään jalkapallossa pelaaja- ja joukkuesuoritusten kehittämiseen?”. Jotta voidaan vastata tutkimuskysymykseen, käytetään apukysymyksenä ”miten ja millaista dataa kerätään jalkapallo data-analytiikkaa varten.

Tutkielman teon kannalta löytyy 3 keskeistä käsitettä, jotka käydään seuraavaksi läpi lyhyesti. Valtavia, järjestelemättömiä ja monimuotoisia tietomassoja kutsutaan **Big dataksi** (Suom. Massadata). Big datan määritelmä on varsin löyhä ja täysin universaalia määritelmää ei ole vielä kyetty kehittelemään (Gandomi & Haider, 2015). Big dataa kuvataan yleensä kuitenkin yksittäisen liiketoiminnan näkökulmasta, mikä tekee käsitteestä erittäin kontekstisidonnaisen. Yleisin määritelmä Big datasta on ns. 3 v:n määritelmä, jossa Big dataa luonnehditaan sitä kuvaavien kolmen v-kirjaimisen ominaisuuden kautta. Nämä ominaisuudet ovat määrä (engl. Volume), nopeus (engl. Velocity), monimuotoisuus (engl. Variety). (Gandomi & Haider, 2015; Venkatram & Geetha, 2017). Määrä ominaisuus kuvaa datan valtavaa määrää. Nopeus kuvaa datamäärän kasvuvauhtia ja käsittelyvauhtia. Monimuotoisuus niputtaa yhteen eri datamuodot, kuten videot, xml tiedostot ja kuvat yhden kattokäsitteen alle.

Urheiluanalytiikka tarkoittaa koneoppimismetodien implementointia urheiluun saavuttaakseen hyödyllisiä lopputuloksia urheiluasuoritusten ja urheiluliiketoiminnan näkökulmista. Nämä löydökset voivat vaikuttaa yksittäisen pelaajan suoritukseen tai vaikuttaa joukkueetasolla esimerkiksi joukkueen valitsemaan taktiikkaan. Urheiluanalytiikkaa hyödynnetään muutenkin kuin vaikuttamalla urheiluasuorituksiin. Esimerkiksi pelaajien markkinahintoja ja loukkaantumisriskien tunnistamista. Alamarin mukaan urheiluanalytiikalla tarkoitetaan strukturoidun historiallisen datan hallintaa ja ennakoivien analyttisten mallien laadintaa, mitkä käyttävät tätä dataa, sekä tietojärjestelmien käyttöä, jotka viestivät informaatiosta päättäjille. (Alamar, 2013) Jakautuu kolmeen alaosaan: Datan hallinta, ennakoivat mallit ja tietojärjestelmät. Ohjelmien tarkoituksena on auttaa organisaatioiden päätöksentekijöitä saadakseen kilpailullisen edun.

Luotettavan kuvan aikaansaamiseksi tarvitaan holistinen kokoelma dataa, jotta voidaan tunnistaa kaikki relevantit parametrit analytiikkatyötä varten. (Apostolou & Tjortjis, 2019). Viime vuosien muutokset FIFA:n laatimissa säännöissä, jotka koskevat langattomia, elektronisia monitorointi järjestelmiä ovat avanneet ovet ennennäkemättömän suureen data määrään, joka mahdollistaa laaja-alaisemman analytiikka työn. Muutoksen seurauksena voidaan päästä käsiksi paitsi pelaajan fysiologisiin tietoihin kuten sydämensykkeeseen ja pulssiin mutta myös esimerkiksi psykologisiin tekijöihin kuten fanien vaikutuksiin pelaajasuorituksiin. (Apostolou & Tjortjis, 2019)

Big data-analytiikka tarkoittaa sitä työprosessia, jossa tilastotieteellisiä menetelmiä hyödyntäen Big datasta kerätään ja analysoidaan kiinnostavaa

informaatiota saadakse liiketoimintahyötyjä (Elgendy & Elragal, 2016). Datamäärien kasvaessa tehtävän työn kompleksisuus kuitenkin kasvaa merkittävästi ja samalla myös vaadittu laskentateho ja tallennuskapasiteetti (Russom, 2011).

Tutkielma on toteutettu kirjallisuuskatsauksena. Tutkielmassa käytettyjä lähteitä on haettu Google Scholar tietokantaa käyttäen hakusanoilla "Big data", "Big data analytics", "football," "soccer". Artikkelit valikoituivat niiden sisällön, viittausmäärän ja julkaisuvuoden mukaan.

Tutkielman rakentuu johdannosta, kolmesta sisältöluvusta ja yhteenvedosta. Ensimmäinen sisältöluke käsittelee Big dataa ja Big data-analytiikkaa. Toisessa sisältökuvussa käsitellään urheiluanalytiikkaa. Viimeisessä sisältöluvussa yhdistetään Big data ja urheiluanalytiikka ja tarkastellaan, miten jalkapallossa voidaan hyödyntää Big data analytiikkaa. Viimeisessä sisältöluvussa pyritään vastaamaan tutkimuskysymykseen. Viimeisessä luvussa kootaan yhteen tutkielman keskeiset havainnot ja pohditaan jatkotutkimuskohteita.

2 BIG DATA

2.1 Big datan määritelmiä

Big datalla ei ole yhtä ainoaa universaalia määrittelytapaa. Kyseessä on monitulkintainen ja kontekstisidonnainen termi, jonka ympärillä pyörii valtavasti kysymyksiä ja ihmetystä. Yhteistä kaikille näkemyksille on kuitenkin ajatus valtavista järjestelemättömistä tietomassoista, joita ei voida perinteisin keinoin hallita (Power, 2014).

Kaikista yleisin ja tunnettu Big datan määrittelytapa on kolmen v:n malli. Mallissa Big datan ominaisuuksia on jaoteltu kolmeen Big dataa kuvaavaan attribuuttiin. Määrä (Volume) kuvaa Big datan valtavaa datamäärää. Yksi tapa ilmoittaa datan koko on esimerkiksi teratavuina. Kuitenkaan pelkkä suuri määrä dataa ei tee Big datasta suurta.

Toinen attribuutti, nopeus (Velocity), kuvaa toisaalta uuden datan syntynepeutta, mutta myös sitä vauhtia, missä dataa tulisi käsitellä ja analysoida hedelmällisten liiketoimintatuloksien aikaansaamiseksi. Big datalle tyypillistä on datan määrän nopea kasvu, joten myös käsittelyvauhdin tulisi olla nopeaa maailman liiketoimintahyödyn aikaansaamiseksi. Datan nopeampi käsittely mahdollistavat organisaatioille saavuttaa kilpailullista etua suhteessa kilpailijoihin. Datan reaaliaikainen käsittely on myös erityisen tärkeää, koska uutta dataa syntyy jatkuvasti nopealla tahdilla (McAfee & Brynjolfsson, 2012). Esimerkiksi jalkapallon kontekstissa Saksan maajoukkue onnistui hyödyntämään reaaliaikaista Big data-analytiikkaa vuoden 2014 jalkapallon maailmanmestaruuskilpailuissa ja onnistui tehostamaan syöttökuvioitaan pitkin kisoja, mikä osaltaan auttoi Saksan marssia maailmanmestaruuteen (Tan ym., 2017).

Viimeinen 3 v:n attribuuteista on monimuotoisuus (Variety), joka kuvaa datan lähteen, luonteen ja tyyppien monipuolisuutta. Erilaisia datan muotoja ovat esimerkiksi kuvat, videot, tekstit ja xml-dokumentit. Big data voi olla strukturoitua, strukturoimatonta tai semi-strukturoitua. Strukturoidulla datalla viitataan

dataan, joka on standardoidussa muodossa ja toteuttaa jotakin tietomallia. Suuri osa Big datasta on strukturoimatonta, joka tarkoittaa rakenteellisen tietomallin puuttumista datasta. Kuvat ja videot ovat esimerkkejä strukturoimattomasta datasta. Semi-strukturoidussa datassa on rakenteellisen datan elementtejä, mutta ei kuitenkaan noudata mitään tietomallia (Rusu ym., 2013).

Kuten Big data nimityksestä huomaa kyseessä on jotain suurta, mutta mikä lopulta tekee datasta suurta? Gandomi ja Haider (2015) esittävät Big datalle vaatimuksen nk. ”tipping pointin”, jossa perinteiset menetelmät eivät enää kykene käsittelemään datajoukkiota sen suuren koon ja kompleksisuuden vuoksi. Heidän määritelmässään korostuvatkin Big datan eri dimensioiden vuorovaikutus. Heidän mukaansa Big datan volyyymiin vaikuttavat myös datan käsittelynopeus (Velocity) ja datan luonne (variety). Gandomi ja Haiderin näkemyksen mukaan eri dimensioiden vuorovaikutus korostuu Big dataa luonnehtiessa. Kahden erillisen rakenteeltaan erilaisen, mutta samankokoisen datajoukkion käsittely voi vaatia eri määrän aikaa ja teknologioita, näin voidaan huomata Big datan olevan hyvin kontekstisidonnainen nimitys.

Ajan saatossa on syntynyt muitakin Big dataa luonnehdittavia kuvaavia attribuutteja kuten arvo (value) ja todenmukaisuus (veracity). Arvo attribuutilla tarkoitetaan Big datasta potentiaalisesti saatavaa liiketoimintahyötyä. Todenmukaisuudella viitataan Big datan oikeellisuuteen ja tarkkuuteen (Nguyen, 2018). Vaihtelevuus (Variability) kuvaa datan syntyneopeuden vaihtelevuutta ja tendenssiä syntyä eri tahtiin eri aikoina vaihtelevin määrin. Vaihtelevuuteen lukeutuu myös datan alkulähteiden muutokset, mitkä vaikeuttavat datan integrointia ja datan muuttamista oikeaan muotoon (Gandomi & Haider, 2015; Power, 2014). Viiden V:n mallissa Big dataan sisältyvät määrän, nopeuden ja monipuolisuuden lisäksi myös arvo ja todenmukaisuus.

Kirjallisuudessa nousee toistuvasti myös esille Big datan näkemisenä prosessina pelkän datajoukon sijasta. Big data prosessiin liitetään toisaalta edellä esitelty datan informatiivinen puoli, mutta yhdistetään datan käsittelyyn liitettävät analysointi ja datan visualisointi operaatiot sekä datan käsittelyyn vaadittavat teknologiat (Gandomi & Haider, 2015; Labrinidis & Jagadish, 2012; Venkatram & Geetha, 2017). Big data prosessiin kuuluu myös datan hankinta, datan muuntaminen sopivaan muotoon ja datan analysointi sekä visualisointi (Gandomi & Haider, 2015).

2.2 Big datan mahdollisuudet

Big data on terminä valtavan hehkutuksen alaisena, missä korostetaan Big dataa liiketoimintaa uudistavana tekijänä. Big datan data ei kuitenkaan itsessään ole arvokasta, vaan arvo piilee siihen tehtävistä datanhallinta operaatioista ja niissä päätöksissä, jotka syntyvät Big dataan nojautuen (Gandomi & Haider, 2015).

Big data prosessilla nähdään olevan 3 keskeistä hyötyä yleisellä tasolla (Venkatram & Geetha, 2017). Ensimmäinen on kustannusten alentaminen. Big dataan soveltuvilla teknologioilla tehostetaan datan käsittelyä, sekä voidaan tunnistaa informaatiota parempien päätöksen tekemiseen. Toiseksi Big data-analytiikka oikein toteutettuna on nopeampaa ja päätöksiä voidaan tehdä reaaliaikaiseen dataan pohjautuen. Reaaliaikainen data-analytiikka mahdollistaa myös mahdollisuuden reagoida muutoksiin nopeasti ja tehokkaasti (Elgendy & Elragal, 2016). Kolmanneksi Big data analytiikan kautta voidaan paremmin ymmärtää asiakkaiden tarpeita ja nyansseja asiakasmielityksistä. Näin organisaatioiden tuotteita voidaan tuottaa tehokkaammin täsmäämään asiakastarpeiden kanssa (Thakkar & Shah, 2021). Big dataan liitettävät hyödyt ovat kuitenkin erittäin kontekstisidonnaisia ja niiden yleistäminen on varsin vaikeaa.

2.3 Big datan haasteet

Big data on käsitteenä hyvin trendikäs, mutta siihen liittyy paljon epävarmuutta ja riskejä. Devlin (Devlin ym., 2012) kuvaa, kuinka Big dataa käsitteenä ei täysin ymmärretä tai millaisia implementointivaatimuksia siihen liittyy. Koska Big dataa ei ymmärretä ei siitä voida myöskään hyötyä maksimaalisesti. Buhlin mukaan Big data termi on suuren hypetyksen syövyttämä ja antaa väärän kuvan liiketoiminnallisena ihmelääkkeenä (Buhl ym., 2013). Organisaatioiden ylin johto on usein pettynyt Big datan tuomiin hyötyihin, koska tulokset eivät vastaa johdon ylisuuria odotuksia. Big data termi itsessään ei kerro mitään, vaan konteksti määrittelee sen sisällön ja vaadittavat teknologiat Big datan tehokkaaseen hallintaan. Buhlin (Buhl ym., 2013) mukaan peräti joka toinen Big data projekti keskeytetään ennen valmistumista. Big datan implementointiin liittyy monia haasteita ja Big dataa käyttöönottaessa tulisikin muokata liiketoimintaprosesseja Big dataa tukeviksi. Jollakin aloilla tämä on tuottanut merkittäviä ongelmia Big datan integrointiin (Brynjolfsson & McElheran, 2016). Big dataan liittyy aina merkittävä määrä roskadataa, oikeiden suodattimien ja mittareiden löytäminen onkin yksi avaintekijä, jotta organisaatio ei hukkuisi dataan (Gandomi & Haider, 2015).

Big dataan liitetään yleensä automaattisesti alentuneet kustannukset ja hinnat, mutta toisaalta big data prosessoinnissa käsitellään suurempia data määriä, haalitaan uusia datalähteitä, mikä johtaa kalliimpaan datan hallintaan, jotka sitten heijastuvat tuotteiden ja palveluiden hinnoissa (McAfee & Brynjolfsson, 2012).

Yksi suurimmista ongelmista Big datan kanssa on osaajapula. Osaajia on vaikea löytää ja ovat yrityksille kalliita palkata (Davenport, 2014; McAfee & Brynjolfsson, 2012). Osaajapula liittyy sekä itse Big dataan, että vaadittavien teknologioiden hallintaan.

Big datan luonteeseen liitettävät V:t ovat myös itsessään haasteellisia. Dataa voi olla todella paljon, eri muodoissa ja sitä tulee jatkuvalla syötöllä lisää. Haasteena on löytää relevanttia dataa kaiken seasta reaaliaikaisesti ja suodattaa

roskadata pois. Toisaalta myös roskadatan poistamiseen liittyy aina riskejä, sillä nykyinen roskadata voi olla tulevaisuudessa liiketoiminnan kannalta tärkeää dataa.

2.4 Big data analytiikka

Big data analytiikka on työtä, jossa kehittyneitä analytiikkatyökaluja kohdistetaan Big datajoukkoihin arvokkaan informaation löytämiseksi (Elgendy & Elragal, 2016). Koska Big data on luonteeltaan niin monimutkaista, etteivät perinteiset analysointimenetelmät sovellu aineistojen käsittelyyn on jouduttu kehittämään uusia menetelmiä Big datan hallintaan ja analysointiin (Russom, 2011). Houtmeyersin ym. (Houtmeyers ym., 2021) mukaan Big data-analytiikka on iteraatiivinen prosessi, jossa löydetään arvokasta tietoa parempien ja nopeampien päätöksien vuoksi. Rehmanin (Rehman ym., 2016) mukaan Big data-analytiikassa on kyse toimintakelpoisten informaatiokaavojen (eng. *information patterns*) tunnistamisesta Big datasta. Big data analytiikan tarkoituksena on tarjota päätöksentekijöille dataan pohjautuvaa informaatiota ja näin tehostaa päätöksentekoprosessia nopeuttamalla ja tehostamalla päätöksiä (Elgendy & Elragal, 2016). Urheilussa Big data-analytiikka voi olla kustannustehokastapa vähävaraiselle joukkueella parantaa kilpailullista asemaansa (Thakkar & Shah, 2021). Tästä tunnetuin esimerkki on baseball joukkue Oakland Athleticsin data-analytiikkatyö, millä pienibudjettinen joukkue löysi huomattomia timantteja ja kykeni haastamaan sarjan kärkipään joukkueita dataohjautuvuuden kautta.

Venkatram ja Geetha (Venkatram & Geetha, 2017) esittävät Big data-analytiikalle neljä erilaista muotoa, jotka tuottavat datan pohjalta erilaista informaatiota päätöksentekijöille. Ohjaileva analytiikka kertoo, millaisia päätöksiä tulisi tehdä. Ennakoiva analytiikka luo datan pohjalta skenaarioita todennäköisiä lopputulemia ja painotuksia eri lopputulemien toteutumiselle. Ennakoivassa analytiikassa ei ennusteta tulevaisuutta vaan tunnistetaan erilaisia kaavamaisuuksia datan pohjalta. Diagnostinen analytiikka on historiallisen datan tarkastelua, jonka pohjalta voidaan havaita mitä on tapahtunut ja miksi. Kuvaileva analytiikka kuvaa organisaation nykytilannetta datasyötteen perusteella.

Rehman (2016) esittävät big data-analytiikan 6 vaiheisena prosessina. Ensimmäinen vaihe on datan hankinta. Dataa kerätään useista eri lähteistä ja oikeiden suodattimien asettaminen haulle onkin tärkeää, jotta saadaan kerättyä kiinnostavaa dataa eikä irrelevanttia dataa. Tietovirtojen optimointi onkin vaiheelle hyvin tärkeää. Erilaiset datatyypit vaativat erilaisia datanhallintamenetelmiä, mikä vaikeuttaa prosessin läpiviemistä merkittävästi (Gandomi & Haider, 2015) Toinen vaihe on datan valmistelu vaihe, jossa data muutetaan oikeaan muotoon analytiikkatyötä varten. Tavoitteena on parantaa datan luotettavuutta ja laatua. Datan luotettavuutta parannetaan esimerkiksi poistamalla anomalioita, tunnistamalla poikkeuksia, vähennetään datan dimensioita ja käsitellään puuttuvia arvoja. Rehmanin ym. mukaan Big datan valmisteluvaihe on prosessin tärkein vaihe ja vaikuttaa merkittävästi analytiikkatyön nopeuteen ja tehokkuuteen.

Seuraavassa vaiheessa luodaan oppimisen malli datan analysointia varten. Mallin luontia varten apuna käytetään valmentavaa datajoukkoa, joka on luonteeltaan samantyyppinen kuin itse analysoitava aineisto. Neljännessä vaiheessa arvioidaan opetusmallin toimivuus ja soveltuvuus oikealle datajoukolle, sekä mahdollisesti muutetaan sitä. Viides vaihe on jalkauttamisvaihe, jossa applikaatiot tunnistavat informaatiokaavoja Big datasta. Viimeinen vaihe on monitorointi vaihe, jossa tehostetaan Big data-analytiikan prosesseja tulevaa dataa varten ja liiketoiminnan tarpeita varten.

3 URHEILUANALYTIikka

Urheiluanalytiikka on termi, joka on Big datan aikakaudella suuressa huudossa urheilumaailmassa. Termillä kuvataan ilmiötä, jossa data-analytiikan menetelmien ja prosessien kohdistamista urheilusuorituksiin ja urheiluliiketoiminnan kehittämiseksi (Davenport, 2014). Urheilumaailma ja urheiluorganisaation johtaminen on perinteisesti pohjautunut vahvasti intuitiivisiin päätöksiin. Urheiluanalytiikka on suoraan ristiriidassa intuitioon pohjautuvan päätöksenteon kanssa. Tästä huolimatta urheiluanalytiikan suosio on voimakkaassa nosteessa ja monilla huippu-urheiluorganisaatioilla on omat data-analytiikka yksikkönsä (Patel ym., 2020; Tan ym., 2017). Kuumen nosteen takana on ennen näkymättömien, valtavien datamassojen näkyväksi tuleminen ja parantuneet data-analytiikan mahdollisuudet parantuneen laskentatehon vuoksi (Alamar, 2013).

Patelin mukaan urheiluanalytiikka puolestaan tarkoittaa ”Historiallisen datan hallinnointia, ennakoivien analyttisten mallien käyttämistä ja tietojärjestelmien soveltamista urheilulliseen dataan päätöksenteon parantamiseksi ja kilpailullisen edun aikaansaamiseksi” (Patel ym., 2020). Patel korostaa myöskin näkemystä, jonka mukaan urheiluanalytiikka ei ole pelkästään suoritusten kehittämistä, vaan kokonaisvaltaisempi liiketoiminnan muoto esimerkiksi kannattajien osallistuttamiseksi organisaation toimintaan ja paremman urheilukokemuksen tuottaminen.

Davenport jakaa urheiluanalytiikan osa-alueet kolmeen alakategoriaan; 1) joukkueen ja pelaajien suoritusten analyysi 2) urheiluliiketoiminnan analytiikka ja 3) terveydentilan ja vammojen ennakoivaa analytiikkaan. Vastaavasti Hoegen ym. (Hoegen ym., 2020) mukaan urheiluanalytiikka voidaan jakaa 7 alakategoriaan: 1) Kannattajien osallistuttaminen 2) Joukkueen suoritusten analysointi 3) Kilpailijoiden analysointi 4) Pelaajien ja joukkueen arviointi 5) johtaminen 6) Pelaajien terveydentila ja turvallisuus sekä 7) Videolähetysten hallinnointiin. Molemmissa näkemyksissä on samoja piirteitä erityisesti urheilusuoritusten kehittämisen näkökulmasta, mutta Hoegen ym. näkemyksessä korostuu holistisempi näkemys urheiluanalytiikasta osana liiketoiminnan tehoastajana.

Urheiluanalytiikan termin määritelmä ei ole täysin yksiselitteinen, ja mitä kaikkea siihen liitetään, on hyvin vaihtelevaa. Kuten Big datankin kohdalla,

urheiluanalytiikalle ei ole olemassa universaalia määrittelytapaa. Kaikissa määritelmässä kuitenkin korostuu kuitenkin sekä urheiluasuoritusten kehittäminen että päätöksenteon parantaminen. Urheiluanalytiikalla pyritään säästämään myös valmentajien aikaa ja rahaa automatisoimalla esimerkiksi videomateriaalin analysointia. Vaikka urheiluanalytiikan menetelmät voivat parantaa päätöksiä ja säästää kustannuksia, niin lyhyellä aikavälillä menetelmien implementointi kasvattaa urheiluorganisaation kustannuksia ja ylimmän johdon vakuuttaminen menetelmien soveltuvuudesta on usein ongelmia lisäkustannuksien vuoksi (Tan ym., 2017).

Alamarin mukaan urheiluanalytiikka on strukturoidun historiallisen datan hallinta ja ennakoivien analyttisten mallien laadinta, jotka käyttävät tätä dataa sekä tietojärjestelmien käyttöä, jotka viestivät informaatiosta päättäjille. Alamarin (Alamar, 2013) mukaan urheiluanalytiikan tavoitteena on ajankäytön tehostaminen sekä uniikin informaation tuottaminen, jolla saavutetaan kilpailullista etua. Alamarin näkemyksen mukaan urheiluanalytiikkaohjelma koostuu kolmesta alaosasta: Datan hallinta, ennakoivat mallit ja tietojärjestelmät. Ohjelmien tarkoituksena on auttaa organisaatioiden päätöksentekijöitä saadakseen kilpailullisen edun. Alamarin viitekehyksen mukaan datanhallinta funktio tarjoaa datan ennakoiville malleille ja tietojärjestelmille. Analyttiset mallit puolestaan käyttävät datan standardoidulla tavalla vieden tiedot tietojärjestelmiin tai etsien vastauksia ennalta määrättyihin kysymyksiin päätöksentekijöille. Tietojärjestelmien tehtävänä on näyttää toimintapohjainen informaatio päätöksentekijöille (Alamar, 2013).

Alamar jakaa urheiluanalytiikan tuomat hyödyt kolmeen osaan. Ensimmäinen osa on valmennuksen kehittäminen. Jalkapalloseurojen valmentajat etsivät aina yliotteita kilpailijoidensa nähden. Yksi kilpailuetua tuottava elin on analytiikka. Alamarin mukaan kun kaikki analytiikkaan vaadittava data on integroitu yhteen järjestelmään, on valmentajan helppoa löytää kaikki relevantti tieto otte- lutaapahtumiin liittyen ja kehittää joukkueensa taktiikkaa (Alamar, 2013). Esimerkiksi jos joukkue on 6 ottelun tappioputkessa, voi analytiikkajärjestelmä automaattisesti tunnistaa vastustajan kaikkien maaliin johtaneiden hyökkäyksien tulleen vasemmalta laidalta. Järjestelmä säästää näin valmentajan arvokasta aikaa joukkueensa taktiikkansa hiomiseen, ettei vasemmalta laidalta enää vuotaisi hyökkäyksiä.

Toinen Alamarin nostama pelaajien arviointi urheiluanalytiikan hyöty liittyy pelaajien suoritusten arviointiin. Pelaajan arviointiin liittyvät tiedot tulevat usein eri lähteistä kuten videoista, terveystieteistä ja kykyjenetsijöiden artikkeleista. Kun kaikki tieto on koottu yhden tietojärjestelmän alle, voivat valmentajat tehokkaasti arvioida eri pelaajien ominaisuuksia ja soveltuvuutta joukkueeseen ja sen taktiikoiden toteuttamiseen (Alamar, 2013).

Kolmas Alamarin nostama hyöty liittyy pelaajien suoritusten kehittämiseen. Kehittyneen analytiikkajärjestelmän avulla valmentajat voivat tunnistaa pelaajien kehityskohteita tehokkaasti ja asettaa erilaisia täsmätoimia kehittämistyötä varten. Järjestelmän avulla voidaan myös seurata kehitystä ja muuttaa kehittämisprosesseja reaaliaikaisesti. Kun tiedot pelaajan ominaisuuksista, terveydestä

ja otteluista on yhden järjestelmän alle voidaan tietoja yhdistellä tehokkaasti, saadakseen maksimaallisen hyödyn järjestelmästä ja eniten irti pelaajasta (Alamar, 2013).

Yhden jalkapallo ottelun aikana syntyy valtavat määrää dataa. Davis ym. nostaa esille 3 eri urheiluanalytiikan osalta keskeistä datan muotoa, jotka ovat tapahtumavirta data, optinen jäljitysdata ja sensoridata. Tapahtumavirta datan, joka kuvaa aikaa sekä sijainteja tiettyjen ottelutapahtumien syntyhetkillä, kuten esimerkiksi laukauksia maalia kohtia tai torjuntia. Optinen jäljitys data tallentaa sijaintitietoa pelaajista ja pallosta useampia kertoja sekunnissa erilaisilla optisilla jäljitysjärjestelmillä. Sensoridataksi luokitellaan esimerkiksi pelaajien sydämen syke ja kiihtyvyys, joita seurataan tiivistii urheilutapahtumien aikana (Davis ym., 2019).

Tärkeimpiä urheilijoiden suoritusten kehittämisen näkökulmasta saatavilla olevia dataa ovat biografista, video, ottelutilastot, lääketieteelliset asiakirjat pelaajista ja kykyjenmetsästyraportit. Kerätty data on standardisoitu, keskitetty, integroitu ja analysoitu erilaisilla mittareilla, jotta voidaan saavuttaa kilpailullista etua vastustajiin nähden (Morgulev ym., 2018).

Ottelutapahtumien analytiikka vaatii huomattavan määrän erilaisia mittareita ja analysointityö on monimutkaista. Yksittäisiin ottelutapahtumiin vaikuttavat monesti monet eri tekijät, joiden painoarvoa voi olla hyvinkin vaikea arvioida. Tällaisia mittareita ovat esimerkiksi eri pelaajat ottelutilanteessa, katsojien lukumäärä, ottelun tärkeys, jäljellä oleva peliaika, lähellä olevien pelaajien pelitavat ja joukkueen strategia. Näiden mittarien yhteensovittaminen vaatii merkittävän määrän työtä. Ottelutapahtumat jotka ovat keskenään enemmän samankaltaisia on helpompaa analysoida, kuten esimerkiksi rangaistuspotkut, koska muuttujia on huomattavasti vähemmän (Morgulev ym., 2018).

Alamarin ja Morgulevin mukaan on tärkeää integroida eri datalähteistä saatavaa informaatiota datanhallintajärjestelmään. Näin voidaan saavuttaa parempi kokonaiskuva pelaajan suorituksista ja huomattavasti mitkä mittarit vaikuttavat toisiinsa (Alamar, 2013; Morgulev ym., 2018). Vaaditut investoinnit näihin teknologisiin järjestelmiin datan integrointiin varten ovat kuitenkin merkittävät ja vähentävät ylimmän päättäjien halukkuutta investoida niihin (Morgulev ym., 2018; Tan ym., 2017).

Hoege ym. (Hoege ym., 2020) keskustelevat urheiluanalytiikkaan liittyvistä haasteista. Heidän mukaansa ensimmäinen haaste liittyy tietojen julkisuuteen. Toiset seurat puhkuvat intoa löydöksistään ja analytiikkatyökalujen käytöstä kehittääkseen entisestään analytiikan merkitystä urheilussa. Toiset urheiluseurat puolestaan haluavat pitää kaiken tiedon itsellään, jotta he eivät menettäisi kilpailullista etuaan.

Toinen ongelma liittyy urheilijoiden biometrisen datan käyttöön ja urheilijan turvallisuuteen. Epäpätevien mittareiden käyttö voi johtaa huonoihin ratkaisuihin urheilijoiden terveyden näkökulmasta. Juoksija Mary Cainin tapauksessa mittarit kertoivat urheilijan painon olleen suurin ongelma kehittymisen näkökulmasta. Tämä johti urheilijan käyttämään lääkkeitä painonpudottamiseen aiheuttaen vakavia fysiologisia ja henkisiä vammoja (Hoege ym., 2020).

Kolmas ongelma käsittelee huolia, kuinka automatisaatio ja analytiikka voivat potentiaalisesti syrjäyttää inhimillisen osaamisen urheilijoiden suoritusten kehittämisessä. Hooge ym. kuitenkin korostavat analytiikkatyökalujen ja inhimillisen osaamisen yhdistämisen merkityksellisyyden. Valmentajatyö vaatii urheilijoiden kokonaisvaltaisen tuntemuksen aina ihmissuhteista ja henkisen tilan tuntemuksesta urheilullisiin ominaisuuksiin.

4 BIG DATA JA URHEILUANALYTIikka JALKAPALLOSSA

Perinteisesti jalkapallon analytiikassa haasteena on ollut olemassa olevan datan vähäisyys ja riittämättömät keinot datan analysointiin (Rein & Memmert, 2016). Nykyään Big datan aikakaudella kuitenkin yhä enemmän dataa kerätään ammatillisjalkapallossa analyttisiin toimenpiteisiin erilaisin menetelmin. Pelaajien sijainteja ja liikkeittä monitoroidaan yhä tarkemmin teknologisten edistysaskeleiden seurauksena (Goes ym., 2021; Rein & Memmert, 2016). Pelaajien sijaintitieto on tärkeä elementti pelaajien suoritusten ja joukkueiden valitsemien taktiikoiden kehittämisessä. Jalkapallo kuitenkin eroaa monista muista urheilulajeista siten, että pisteiden tekotilanteita (maalit) on merkittävästi vähemmän kuten esimerkiksi jääkiekossa tai koripallossa. Näin ollen jalkapalloanalyysi keskittyy enemmän taktisiin tekijöihin, kuten sijaintitietoihin, syöttökuvioihin, pallonhallintaan ja mentaalisiin tekijöihin (Morgulev ym., 2018).

4.1 Datan kerääminen

Davenport erittelee kolme tärkeää datanlähdettä jalkapallon analyysin näkökulmasta. Nämä ovat sijaintitieto, videodata ja biometrinen data. Näitä datalähteitä käytetään erilaisiin käyttötarkoituksiin, mutta niiden tietoja myös yhdistellään paremman kokonaiskuvansaamiseksi. Davenport mainitsee erityshaasteeksi, kuinka sama tieto voidaan havaita useasta eri lähteestä, mikä aiheuttaa valtavan tietotulvan ja turhaakin dataa. Esimerkiksi videodatasta ja sijaintidatasta voidaan molemmista tulkinta kuinka paljon yksittäinen pelaaja on juossut ottelun aikana (Davenport, 2014). Tietovirtojen optimointi onkin yksi keskeisimmistä haasteista urheiluanalytiikassa, mutta myös data-analytiikan menestykseen vaikuttava tekijä.

4.1.1 Videodata

Ensimmäinen merkittävä datan lähde jalkapallon pelillisen kehittämisen näkökulmasta on videodata. Videodatalla on merkittävä rooli otteluihin valmistautumisessa. Valmentajien tehtävänä on analysoida oman joukkueen pelaamista, että vastustajien. Manuaalisesti tehtynä, tämä kuitenkin vie valtavasti aikaa, joten prosessi pyritään automatisoimaan (Alamar, 2013). Videonhakujärjestelmiä onkin kehitetty vastatakseen tähän ongelmaan. Videodatasta pyritään automaattikalla tunnistamaan erilaisia tilanteita, joiden pohjalta generoidaan tilastoja ottelusta (Davenport, 2014). Videodatan pohjalta voidaan kehittää omia taktiikoita, etsiä heikkouksia vastustajista ja näin muokata omia valintoja ja taktiikoita eri otteluiden tarpeisiin sopiviksi (Probst ym., 2018). Parhaimmillaan tämä kaikki voidaan tehdä reaaliaikaisesti ja tehdä muutoksia pelin aikana. Videodatan pohjalta voidaan myös arvioida pelaajien kuormitusta ja loukkaantumistilanteita (Davenport, 2014). Kerätäkseen videodataa käytetään erilaisia optisia seuranta-menelmiä. Esimerkiksi Saksalaisjoukkue Bayern Munchen kerää jokaisessa ottelussa noin 60 miljoonaa datapistettä. Joukkue kerää videodataa stadionin kattoon asennetuilla kameroilla. Tämän lisäksi joukkue kerää lisää materiaalia kuvausryhmällä, jotka kuvaavat jokaisen ottelun eri näkökulmista (Tan ym., 2017).

4.1.2 Biometrinen data

Toinen merkittävä datatyyppe on biometrinen data. Fifa on lähivuosina sallinut joukkueiden käytettävään langattomia sensoreita myös otteluiden aikana (Apostolou & Tjortjis, 2019). Nämä sensorit mittaavat erilaisia fysiologisia tietoja pelaajista. Näitä sensoreita on ollut jo pidempään harjoituksissakin käytössä, mutta kuormitus peleissä on erilaista ja antaa valmentajille merkittävästi lisätietoa pelaajien suorituksista (Rein & Memmert, 2016). Pelaajista mitataan esimerkiksi sydämen sykettä, verenpainetta, unen laatua ja ravinnon laatua. Eri tietoja, kuten pelaajien terveystietoja, biometrisen datan ja sijaintidatan yhdistämällä valmentajilla on käsissään täysin uusia mahdollisuuksia vaikuttaa pelaajiensa ja joukkueensa suorituskykyyn (Davenport, 2014).

Biometrisen datan mittaaminen mahdollistaa myös erilaisten iskujen voimakkuuksien mittaamista. Näin voidaan vaikuttaa tehokkaammin loukkaantumistilanteiden syntymiseen, että nopeammin arvioida iskun vaikutuksia pelaajien suorituskykyyn (Davenport, 2014; Rossi ym., 2018). Biometristä dataa tutkimalla voidaan luokitella väsymystä laaja-alaisemmin. Näin valmentajilla on parempi tieto siitä, miten pelaajia tulisi peluuttaa parhaimman lopputuloksen aikaansaamiseksi (van der Westhuizen & van der Haar, 2018).

4.1.3 Sijaintidata

Davenportin mukaan urheiluseuroilla on jo enemmän sijaintidataa, kuin mitä pystyvät tehokkaasti käsittelemään (Davenport, 2014). Seurat käyttävät esimerkiksi GPS-laitteita sekä radioalloilla toimivia sensorilaitteita mitatakseen pelikentän etäisyyksiä, liikkeitä, nopeuksia ja liikkeiden toistuvuuksia (Rein &

Memmert, 2016). Sijaintidatan pohjalta laaditaan erilaisia tiheyspintakarttoja, havaitakseen merkittävimmät tilanteet ja alueet ottelun näkökulmasta.

4.2 Big datan hyödyntäminen jalkapallossa

Analytiikka jalkapallossa pyöri pitkään tilastojen kuvailussa. Nykyään ollaan siirtymässä kohti ennakoivaa ja diagnosoivaa analytiikkaa. Sen sijaan, että kerrotaan mitä on tapahtunut, pyritään vastamaan siihen mitä tulee tapahtumaan ja miksi näin on käynyt (Davenport, 2014). Taktisella käyttämisellä tarkoitetaan miten joukkue kykenee spatiaalisella sijoittautumisella ajan saatossa saavuttamaan yhteisen päämäärän (ottelun voittaminen) vuoro vaikuttaessaan toistensa ja vastustajan, sekä otteluolosuhteiden kanssa (Goes ym., 2021). Taktinen käyttäytyminen kuitenkin muuttuu otteluolosuhteiden muuttuessa (Davenport, 2014)

Davis (Davis ym., 2019) luokittelee 4 eri kategoriaa joissa voidaan hyödyntää Big data-analytiikka jalkapallo-organisaation pelillisen kehittämisen näkökulmasta. Ensimmäinen näistä on pelaajien arviointi. Pelaajiin kohdistetaan erilaisia suoritusten mittauksia eri ajanjaksoina, esimerkiksi ottelukohtaisesti tai pitkän kautta. Pelaajan ominaisuuksia arvioidaan esimerkiksi numeerisesti ja mahdollisuuksia suoriutua erilaisista pelillisistä haasteista, kuten pallonriistoista otelua kohden tai laukauksia maalia kohti. Näitä mittareita voidaan verrata toisiin pelaajiin ja voidaan käyttää esimerkiksi joukkueen rakentamisessa (Davis ym., 2019).

Toinen Davisin mainitsema kategoria on ottelujen kohokohtien tunnistaminen. Näitä huippukohtia hyödyntää monet eri sidosryhmät, kuten journalistit ja fanit, mutta ennen kaikkea valmentajat hyödyntävät näitä kehittääkseen joukkueen toimintaa ja muuttaakseen ottelukohtaisia taktiikoita parantaakseen voiton mahdollisuuksiaan muita joukkueita vastaan.

Kolmas Davisin kategoria on pelaajien hankinta. Kun vertaillaan eri pelaajien ominaisuuksia erilaisissa ottelujen konteksteissa kykyjenetsijät voivat löytää pelaajia omaan organisaatioon, jotka toteuttavat haluttuja taktiikoita paremmin kuin nykyiset pelaajat. Toisaalta voidaan myös arvioida pelaajien hyötyjä kustannuksiin nähden (Bransen ym., 2019).

Neljäs Davisin mainitsema kategoria on taktinen analyysi. Ymmärtämällä pelaajien henkisiä ominaisuuksia paremmin, voidaan pelaajien roolituksia muuttaa sopimaan paremmin joukkueen tarpeita. Yksi keskeinen esimerkki on pelaajan paineensietokyky. Pelaaja, jolla on huonompi paineensietokyky suoriutuu heikommin omaan perustasoonsa nähden korkean panoksen peleissä (Bransen ym., 2019). Ymmärtämällä tämä joukkue voi tehdä muutoksia joukkueen kokoonpanoon ja taktiikoihin parantaakseen voiton mahdollisuuksia. Joukkueiden parempi valmius taktiikkojen muutoksiin onkin keskeinen Big data-analytiikan tuomista hyödyistä. Bransen ym. huomasivat rajaheittojen merkitysten olevan

hyvin alhainen pelien alkuvaiheessa, koska pelin aiheuttamat paineet eivät vielä vaikuttaneet tarpeeksi pelaajien suorituksiin ja valintoihin. Liverpoolin valmentaja Jürgen Klopp nimesi joukkueelleen oman rajaheittovalmentajan, joka on tehnyt joukkueesta erittäin vaarallisen rajaheittjoukkueen Englannin valioliigassa todella korkealla vaarallisten tilanteiden luontiprosentilla (Bransen ym., 2019). Kehittyneellä taktisella analyysillä valmentajilla on parempi käsitys sekä oman että vastustajajoukkueen vahvuuksista ja heikkouksista. Näin ollen valmentaja voi paremmin vastata pelin vaatimiin tilanneolosuhteisiin ja toteuttaa joukkueen valitseman taktiikan parhaiten (Decroos ym., 2019).

Urheiluvammojen ennaltaehkäisy on yksi urheiluanalytiikan osa-alue (Rossi ym., 2018). Urheiluvammat luovat urheiluseuralle merkittävästi kustannuksia. Kustannukset muodostuvat itse loukkaantumisen hoidosta, mutta myös pelaajan poisjäänti pelaavasta kokoonpanosta voi aiheuttaa tappioita ja alentaa organisaation muita tulonlähteitä, kuten esimerkiksi lippujen ja fanituotteiden myyntituloja (Rossi ym., 2018). Big data on mahdollistanut urheiluvammojen paremman ennakoimisen monitoroimalla pelaajien kuormitusta ja liikkeitä ottelujen ja harjoittelun aikana. Lisäksi voidaan mitata erilaisten iskujen voimakkuuksia ja vaikutuksia pelaajan fysiologiaan (Davenport, 2014; Rossi ym., 2018). Rossin havainnosta organisaatiot, jotka ovat käyttäneet ennakoivaa analytiikka urheilijoiden kuormitusten seuraamisessa ovat tehneet merkittäviä kustannussäästöjä ja parantaneet sijoituksiaan sarjataulukossa. Pelaajan terveydentilan ennakoiminen tarkentuu entisestään, jos saatavilla on usean kauden biometrinen ja sijaintidata sekä pelaajan terveystiedot (Rossi ym., 2018). Urheiluorganisaatio AC Milan otti käyttöönsä tutkimusyksikön MilanLabin vuonna 2002 tarkoituksenaan monitoroida pelaajien terveydentiloja. MilanLabin havaintojen ansiosta AC Milan on pystynyt vähentämään loukkaantumisia 90% vuodesta 2003 lähtien (Davenport, 2014). Tämä on auttanut joukkuetta kamppailemaan Euroopan huijilla vuosi vuodelta.

5 YHTEENVETO

Tutkielmassa käsiteltiin tieteellistä kirjallisuutta hyödyntäen Big datan käyttöä jalkapallossa pelillisen kehittämisen näkökulmasta. Lisäksi tarkasteltiin miten ja millaista dataa jalkapallo-organisaatiot keräävät edellä mainittuun tarkoitukseen. Tutkimuskysymykseen vastaamiseen määriteltiin käsitteet Big data, Big data-analytiikka ja urheiluanalytiikka. Näitä käsitteitä käyttäen vastattiin tutkimuskysymykseen ”Miten Big dataa käytetään jalkapallossa pelaaja- ja joukkuesuoritusten kehittämiseen?”

Big data määriteltiin valtavaksi, järjestelemättömiksi ja monimuotoisiksi tietomassoiksi, joita ei voida käsitellä perinteisin menetelmin. Big data-analytiikalla kuvataan sitä työprosessia, jossa tilastotieteellisiä menetelmiä hyödyntäen Big datasta kerätään ja analysoidaan kiinnostavaa informaatiota saadakseen liiketoimintahyötyjä. Urheiluanalytiikka tarkoittaa koneoppimismetodien implementointia urheiluun saavuttaakseen urheiluorganisaatiolle hyödyllisiä lopputuloksia. Nämä löydökset voivat vaikuttaa yksittäisen pelaajan suoritukseen tai vaikuttaa joukkueetasolla esimerkiksi valittuun taktiikkaan, sekä yleisesti urheiluorganisaation liiketoimintaan.

Tutkielman toisessa luvussa käsiteltiin Big dataa ilmiönä, sen mahdollisuuksia ja haasteita yleisellä tasolla sekä Big data analytiikkaa. Big datan määrittelyyn käytettiin avuksi kolmen v:n mallia, jossa Big datan ominaisuuksia kuvataan 3 v kirjaimisen sanan avulla: Määrä (volume), Monipuolisuus (variety) ja Nopeus (velocity). Määrä kuvaa datan suurta määrää. Monipuolisuudella viitataan datamassojen datatyyprien monipuolisuuteen. Käsiteltävä data voi esimerkiksi olla tekstinä, kuvina tai videoina. Nopeudella kuvataan sekä datan kasvunopeutta, että nopeutta missä dataa tulisi käsitellä. Luvussa nostettiin myös esille muita Big dataan liitettäviä ominaisuuksia kuten totuus (Veracity) ja arvo (Value). Big data-analytiikasta nostettiin esille käsitteen määritelmä, sekä erilaisia analytiikan muotoja, sekä Big data-analytiikan prosessia.

Tutkielman kolmannessa luvussa käsiteltiin urheiluanalytiikkaa. Luvussa avattiin urheiluanalytiikan käsite ja kerrottiin sen erilaisista alakohdista sekä urheiluanalytiikan tavoitteista. Termillä kuvataan ilmiötä, jossa data-analytiikan menetelmien ja prosessien kohdistamista urheilusuorituksiin ja

urheiluliiketoiminnan kehittämiseksi. Urheiluanalytiikalla pyritään parantamaan päätöksentekoa, säästää valmentajien aikaa, sekä parantaa joukkueen pelisuorituksia.

Tutkielman neljännessä luvussa otettiin kiinni tutkimuskysymykseen. Ensiksi käytiin läpi, kuinka jalkapallo organisaatiot keräävät dataa. Keskeisiä datanlähteitä on kolme. Biometristä dataa kertoo pelaajan fysiologisesta tilasta mitaushetkillä, esimerkkeinä sydämensyke ja verenpaine. Biometristä dataa kerätään langattomien sensortehteknologien avulla. Sijaintidata kertoo pelaajien ja pelivälineen liikkeitä pelikentällä. Sijaintidataa voidaan kerätä esimerkiksi GPS laitteiden avulla ja toissijaisesti videodatasta. Videodataa kerätään ensisijaisesti optisia menetelmiä hyödyntäen.

Viimeiseksi käytiin läpi, kuinka Big data vaikuttaa jalkapalloseuran pelilliseen toimintaan. Löydöksiä voidaan tiivistää viiteen pääkohtaan. Ensimmäinen kohta on pelaajien arviointi. Pelaajan eri ominaisuuksia arvioidaan eri mittausdatojen perusteella. Tätä dataa analysoimalla voidaan pelaajan ominaisuuksia sovittaa pelaavan joukkueen kokoonpanoon sopivaksi. Ottelujen kohokohtien tunnistaminen on toinen pääkohta. Ottelujen kohokohtien tunnistaminen auttaa valmentajia säästämään aikaa ja taktisten analyysin toteuttamista. Big data analytiikka edistää myös pelaajahankinta prosessia. Kun pelaajista on kattavat tiedot, voidaan tietoja vertailla keskenään, jotta seura löytää joukkueeseen sopivimmat pelaajat. Neljäs kohta on taktisen analyysin kehittäminen. Kun valmentajilla on paremmat tiedot pelaajien ominaisuuksista, voidaan joukkueen taktiikkaa muuttaa pelaajien ominaisuuksiin nähden sopivaksi ja saada kaikki irti pelaajista ja maksimoida voiton mahdollisuus. Viimeinen kohdistaa on urheiluvammojen ennaltaehkäiseminen. Big dataa hyödyntämällä voidaan paremmin mitata kehon rasitusta erilaisissa urheilusuorituksissa ja mitata erilaisten iskujen voimakkuutta. Näin voidaan ennakoita paremmin urheiluvammojen syntymistä ja pienentää toipumisajanjaksoa. Urheiluvammojen minimointi vähentää myös urheiluorganisaation kustannuksia.

Vaikka Big data on kuuma aihe urheilun maailmassa, Big datan ja jalkapallon välistä suhdetta on tutkittu verrattain vähän ja tutkimukset ovat vasta alkuvaiheessa. Artikkeleista voi myös päätellä, että urheiluorganisaatiot eivät halua paljastaa käyttämiään menetelmiä tarkemmin läpi, etteivät menettäisi kilpailullista etuaan. Jatkotutkimusaiheena voisikin olla kuinka implementoida Big data analytiikkajärjestelmä urheiluorganisaatioon ottaen urheiluorganisaation erityistarpeet ja -haasteet huomioon.

LÄHTEET

- Alamar, B. C. (2013). *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers*. Columbia University Press.
- Apostolou, K., & Tjortjis, C. (2019). Sports Analytics algorithms for performance prediction. *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–4.
<https://doi.org/10.1109/IISA.2019.8900754>
- Bransen, L., Robberechts, P., Haaren, J. V., & Davis, J. (2019). Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. *SciSports*, 25.
- Brynjolfsson, E., & McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, 106(5), 133–139.
<https://doi.org/10.1257/aer.p20161016>
- Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big Data: A Fashionable Topic with(out) Sustainable Relevance for Research and Practice? *Business & Information Systems Engineering*, 5(2), 65–69.
<https://doi.org/10.1007/s12599-013-0249-5>
- Davenport, T. (2014). *Analytics in sports: The new science of winning* [Data set].
<https://doi.org/10.1287/8943f842-86f8-4d42-9a64-9a7cd07b31f5>
- Davis, J., Bransen, L., Decroos, T., Robberechts, P., & Haaren, J. V. (2019, marraskuuta 14). *Assessing the Performances of Soccer Players* | SpringerLink.
https://link.springer.com/chapter/10.1007/978-3-030-35048-2_1
- Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals: Valuing Player Actions in Soccer. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1851–1861. <https://doi.org/10.1145/3292500.3330758>
- Devlin, B., Rogers, S., & Myers, J. (2012, marraskuuta). Big Data Comes of Age: An EMA and 9sight White Paper. *9sight Consulting: Insight beyond Business Intelligence*. <http://www.9sight.com/2012/11/big-data-survey-2012/>
- Elgendy, N., & Elragal, A. (2016). *Big Data Analytics in Support of the Decision Making Process* | Elsevier Enhanced Reader.
<https://doi.org/10.1016/j.procs.2016.09.251>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S., & Lemmink, K. A. P. M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic

- review. *European Journal of Sport Science*, 21(4), 481–496.
<https://doi.org/10.1080/17461391.2020.1747552>
- Hoegel, J., Lansing, M., Nelson, S., Ungerleider, D., Iyer, R., Rhodes, C., Metzger, B., Worcester, P., Chandra, A., Leonard, J., Kreitzer, R., & Scherer, W. (2020). An Interdisciplinary Approach to Sports Analytics in a University Setting. *2020 Systems and Information Engineering Design Symposium (SIEDS)*, 1–6. <https://doi.org/10.1109/SIEDS49339.2020.9106647>
- Houtmeyers, K. C., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, 16(11), 1719–1723. <https://doi.org/10.1123/ijsp.2020-0958>
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033. <https://doi.org/10.14778/2367502.2367572>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60–66, 68, 128.
- Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213–222. <https://doi.org/10.1007/s41060-017-0093-7>
- Nguyen, T. L. (2018). A Framework for Five Big V's of Big Data and Organizational Culture in Firms. *2018 IEEE International Conference on Big Data (Big Data)*, 5411–5413. <https://doi.org/10.1109/BigData.2018.8622377>
- Patel, D., Shah, D., & Shah, M. (2020). The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. *Annals of Data Science*, 7(1), 1–16. <https://doi.org/10.1007/s40745-019-00239-y>
- Power, D. J. (2014). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, 23(2), 222–228. <https://doi.org/10.1080/12460125.2014.888848>
- Probst, L., Rauschenbach, F., Schuldt, H., Seidenschwarz, P., & Rumo, M. (2018). Integrated Real-Time Data Stream Analysis and Sketch-Based Video Retrieval in Team Sports. *2018 IEEE International Conference on Big Data (Big Data)*, 548–555. <https://doi.org/10.1109/BigData.2018.8622592>
- Rehman, M. H. ur, Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, 36(6, Part A), 917–928. <https://doi.org/10.1016/j.ijinfomgt.2016.05.013>
- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410. <https://doi.org/10.1186/s40064-016-3108-2>

- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*, 13(7), e0201264. <https://doi.org/10.1371/journal.pone.0201264>
- Russom, P. (2011). Big Data Analytics. *BIG DATA ANALYTICS*, 38.
- Rusu, O., Halcu, I., Grigoriu, O., Neculoiu, G., Sandulescu, V., Marinescu, M., & Marinescu, V. (2013). Converting unstructured and semi-structured data into knowledge. *2013 11th RoEduNet International Conference*, 1-4. <https://doi.org/10.1109/RoEduNet.2013.6511736>
- Tan, F. T. C., Hedman, J., & Xiao, X. (2017). Beyond 'Moneyball' to Analytics Leadership in Sports: An Ecological Analysis of FC Bayern Munich's Digital Transformation. *AMCIS 2017 Proceedings: A Tradition of Innovation*, 10. <https://research.cbs.dk/en/publications/beyond-moneyball-to-analytics-leadership-in-sports-an-ecological->
- Thakkar, P., & Shah, M. (2021). An Assessment of Football Through the Lens of Data Science. *Annals of Data Science*, 8(4), 823-836. <https://doi.org/10.1007/s40745-021-00323-2>
- van der Westhuizen, E. J., & van der Haar, D. T. (2018). A wearable device-based framework for determining player effectiveness on the football pitch. *Proceedings of the 6th International Conference on Information and Education Technology*, 226-231. <https://doi.org/10.1145/3178158.3178203>
- Venkatram, K., & Geetha, M. A. (2017). Review on Big Data & Analytics - Concepts, Philosophy, Process and Applications. *Cybernetics and Information Technologies*, 17(2), 3-27. <https://doi.org/10.1515/cait-2017-0013>